NOTICE

AVIS

Canada

SPEECH ENHANCEMENT FOR MOBILE TELEPHONY

by

Martie Meryle Goulding

B.A.Sc. (Elec. Eng.), University of British Columbia, 1987

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE (ENGINEERING SCIENCE)

in the School

of

Engineering Science

© Martie Meryle Goulding 1988
Simon Fraser University

December 1988

Canada

APPROVAL

NAME: Martie Meryle Goulding

DEGREE: Master of Applied Science (Engineering Science)

TITLE OF THESIS: Speech Enhancement for Mobile Telephony

EXAMINING COMMITTEE:

Chairman: Dr. Vladimir Cuperman

Dr. John S. Bird
Senior Supervisor

Dr. James K. Cavers

Dr. Paul Ho

DATE APPROVED: ___1994 02 23_____

ii

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Speech Enhancement For Mobile Telephony

Author:

(signature) _ℓ_

Martin M. Goulding

(name)

1989 04 07

(date)

ABSTRACT

The properties of noise fields in automobile interiors are discussed with a view towards speech enhancement for the purpose of voice-activated mobile telephony. The limitations on performance of adaptive noise cancellation and adaptive beamforming are explained in the context of the spatial correlation properties of the noise field. A simple Delay-Equalized Near-Field array of directional microphones is analyzed and found to be effective for increasing the signal-to-noise density ratio and reducing the reverberant distortion of the speech, without introducing any further distortion. Experimental results and comparisons with the predictions based on a computer image model for reverberant enclosures are presented.

For mom and dad, with love.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. MOTIVATION FOR RESEARCH

"Whenever you use a cellular phone while driving, realize that you may be endangering yourself, your passengers and other motorists."[1] Thus warns the American Automobile Association's Potomac division in its recommendations for mobile telephone users.

While the cellular telephone market has grown into a multi-billion dollar business, the fundamental problem of user safety has not been satisfactorily resolved. There are at least two activities involved in use of a conventional cellular telephone that may undermine the driver's ability to drive safely. First, the user must hold the handset, thus removing one of his hands from the steering wheel. This problem has been somewhat alleviated by hands-free cellular phones which use a remote microphone and speaker in place of the conventional handset; however, performance of current systems is marginal at best, as anybody who has used such a system can attest to.

The second and more pressing problem is entering the phone numbers. Although users are warned to refrain from placing calls until the vehicle is stopped, the sight of motorists busy dialing, distracted from their driving is too familiar. Most experts agree that the solution is a cellular phone that responds to verbal commands. To quote one, "The hands-free, voice-activated phone will represent the ultimate in safety."[2]

Such a phone would require computer based recognition of speech that is both masked by high levels of ambient noise and distorted by reverberation. This thesis is an investigation of various signal processing techniques with the purpose of providing the highest quality speech for computer based recognition. The actual speech recognition results are beyond the scope of the thesis, and are reported elsewhere.[3] The research was supported by Glenayre Electronics Ltd. and the Natural Science and Engineering Research Council (NSERC).

1.2. BACKGROUND AND RESEARCH METHODOLOGY

A literature search conducted at the beginning of the project revealed two previous investigations into speech enhancement for mobile telephony in automobiles.[4,5] Both papers reported results for Adaptive Noise Cancellation (ANC) techniques in car interiors, but their conclusions were somewhat contradictory. Goubran and Hafez[4] reported 10 dB gains in Signal-to-Noise Ratio (SNR) from using the Least Mean Squares (LMS) method of ANC, while Savoji[5] claimed that LMS was ineffective, but that an adaptive lattice structure achieved gains of up to 13 dB. The gains reported from either paper were significant, and worth investigating.

Rather than dwelling initially on the question of ANC's effectiveness, we chose to focus on the problem's acoustical properties, since they ultimately determine the performance of any enhancement algorithm. Thorough investigation of the automobile interior's noise field led to serious doubt concerning the effectiveness of ANC. Subsequent experimentation supported the conclusion that ANC is unsuitable for the automobile environment.

Based on the observations of the noise field properties, another technique for enhancing speech was developed. Experimental results confirm the effectiveness of this method, in both reducing the noise level and the effects of reverberation on the speech.

1.3. OUTLINE OF THESIS

One of the primary considerations for any speech enhancement algorithm must be the acoustical environment it is employed in. The second two chapters of this thesis discuss relevant properties of the the acoustics of an automobile interior.

The effects of reverberation are considered in chapter 2. These effects are very important, since both speech and noise suffer the reverberation process that shapes the spectral and correlation characterisitcs of the acoustic field. Chapter 2 also introduces an image model[6] for describing reverberation in rectangular enclosures. This model, enhanced to include directionality in the microphones, is used in chapter 2 and again in chapter 5 to investigate point-to-point acoustic transfer functions between the source and

any number of microphones. The acoustic transfer functions and therefore the received speech energy are highly sensitive to changes in the positions of the speaker and microphone, as well as the structural configuration of the room. The usefulness of the reverberant energy for speech recognizers is expected to be limited by this sensitivity, since the reverberant speech can change drastically with minute changes in the acoustical environment.

The relevant properties of the noise field for the interior of an automobile are described in chapter 3. The noise power spectral density in the automobile interior is presented for a number of conditions (car speed, fan level, and state of windows). The noise is seen to be concentrated at very low frequencies, with the majority of the noise energy below 500 Hz. The coherence, a measure of the noise field's spatial correlation, is also introduced in chapter 3 and found to be small at all but the lowest frequencies.

Chapter 4 describes and investigates Adaptive Noise Cancelling (ANC), a speech enhancement algorithm that employs multiple inputs, the primary input consisting of speech plus noise and any number of reference inputs consisisting of noise alone. Effective ANC requires significant coherence between the noise in the primary and secondary inputs, and since the coherence of the noise field in the car is known from chapter 3 to be low, it is no surprise that ANC is found to be ineffective at all but very low frequencies, outside the range of interest for speech. Note that this low frequency gain explains the significant SNR gains reported previously because SNR was defined as the ratio of total signal power to total noise power, which included the large irrelevant low frequency components. Two recent papers from Europe (published after this research was done)[7,8] reported similar conclusions regarding the poor performance of ANC in an automobile. However, while one[8] hinted that an array of some form should be investigated, neither paper offered a solution.

In chapter 5, we consider an alternative model for the acoustical environment of the automoble interior, and from it derive our solution, the delay-equalized near-field array. Basically, the array delays the inputs and adds them in a coherent manner so that the speech is re-inforced. Such an array has at least five advantages:

1) It achieves gains of approximately N, the number of microphones.

2) By depending on low coherence rather than high coherence for its gain, it is effective across the entire speech frequency band.

3) It decreases reverberation in the speech, which should help the speech recognition.

4) It doesn't introduce any distortion into the speech.

5) The system is very simple to implement. For each element (microphone) in the array, only a very short digital filter (interpolator) is required.

The major limitation of the array is that the microphones must be separated by distances of at least 15 cm, thus limiting the value of N attainable within the confines of the vehicle.

Chapter 6 discusses conclusions from the research.

This research has led to a better understanding of the acoustical field in the vehicle, which has in turn given us the ability to predict the performance of candidate speech enhancement schemes. We have also developed a solution that has significant potential for speech enhancement for mobile telephony.

2. REVERBERATION EFFECTS

2.1. INTRODUCTION

Affecting both speech and noise, reverberation must be understood before any meaningful study of a vehicle interior's acoustics can be undertaken. Both the spectral and correlation characteristics of the acoustic field are governed by the reverberation.

The Audio Dictionary[9] defines reverberation as "the remainder of sound that exists in a room after the source of the sound is stopped." Conceptually, it may be pictured as a series of echoes from the walls and other objects in the room that interfere with each other so that distinct echoes are not discernible. In general, reverberation is very sensitive to the relative positions of the source, receiver, and objects in the room; this dependence results in effects that are highly unpredictable, and are expected to degrade the performance of speech recognizers.

This chapter gives a short analysis of reverberation and its effects on acoustic signals. We begin with a short discussion of the data acquisition hardware and its limitations in section 2.2. Section 2.3 compares the two main methods of analysis for reverberation - a frequency domain approach involving modes of vibration and a time domain approach involving approximate image models. Due to its simplicity and superior accuracy[10], the time domain method is chosen for analysis and modelling.

Section 2.4 describes an approximate image method[6] that was implemented in Fortran to model point-to-point acoustic transfer functions for rectangular reverberant enclosures. The model was enhanced in two ways: first, as suggested by Peterson,[11] low-passed impulses were used to ensure correct phase characteristics (especially important when modelling the responses of an array of microphones as is done in chapter 5). Secondly, a method for introducing directionality into the microphone responses was developed and incorporated into the model.

In section 2.5, recorded impulse responses are discussed. Since exact modelling of the acoustic field inside a vehicle is expected to be difficult,[12, 13] our objective was not to

create a perfect model, but rather one that would behave similarly in terms of the parameters required to test speech enhancement schemes. Candidates passing the criterion of usefulness in the modelled environment could then be experimentally verified, which in turn would verify the essential characteristics of the model.

Finally, the effects of reverberation on the frequency response of a room are discussed in section 2.6.

2.2. DATA ACQUISITION METHODS

The data acquisition hardware for all measurements is displayed in figure 2.1. The acoustical signal was converted to a low-level electrical signal by the microphone(s). The pre-amplifier/amplifier/mixer raised the signal to levels compatible with the tape-recorder (and mixed two or more microphone signals as required).

In our laboratory, the recorder played back the signals through an analog anti-aliasing filter (with cutoff frequency at 5 kHz) into our data acquisition board. With the board we oversampled the signals at 36 kHz, to provide a larger transition bandwidth for the board's digital filter (a 128 point low-pass response with cutoff at 4 kHz) to operate. The board decimated the filter output by a factor of 4, resulting in an effective sampling rate of 9 kHz. The oversampling and decimation scheme allowed us to benefit from the linear phase and consistency of a digital filter without having to store an excessive number of samples.

Two types of microphones were used. The omnidirectional were Realistic brand electret condenser microphones. The directional microphones were gradient microphones with a cardioid pattern of reception in the axial plane and a circle in the circumferential plane. The frequency responses for both the directional and omnidirectional microphones are given in figure 2.2. For the directional microphone, there is a 6 dB/octave fall off below 1 kHz which is typical for the far-field response of gradient microphones.[14]

The cardioid directionality pattern is shown in figure 2.3 and is described by the equation (where θ is measured from the axis of the microphone)

$$\text{Gain}(\theta) = \tfrac{1}{2}(1 + \cos\theta) \tag{2.1}$$
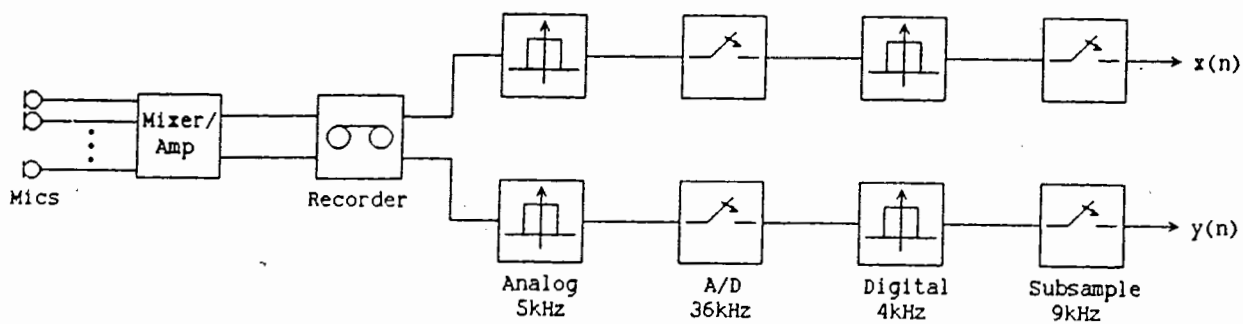
Figure 2.1 Data Acquisition Hardware
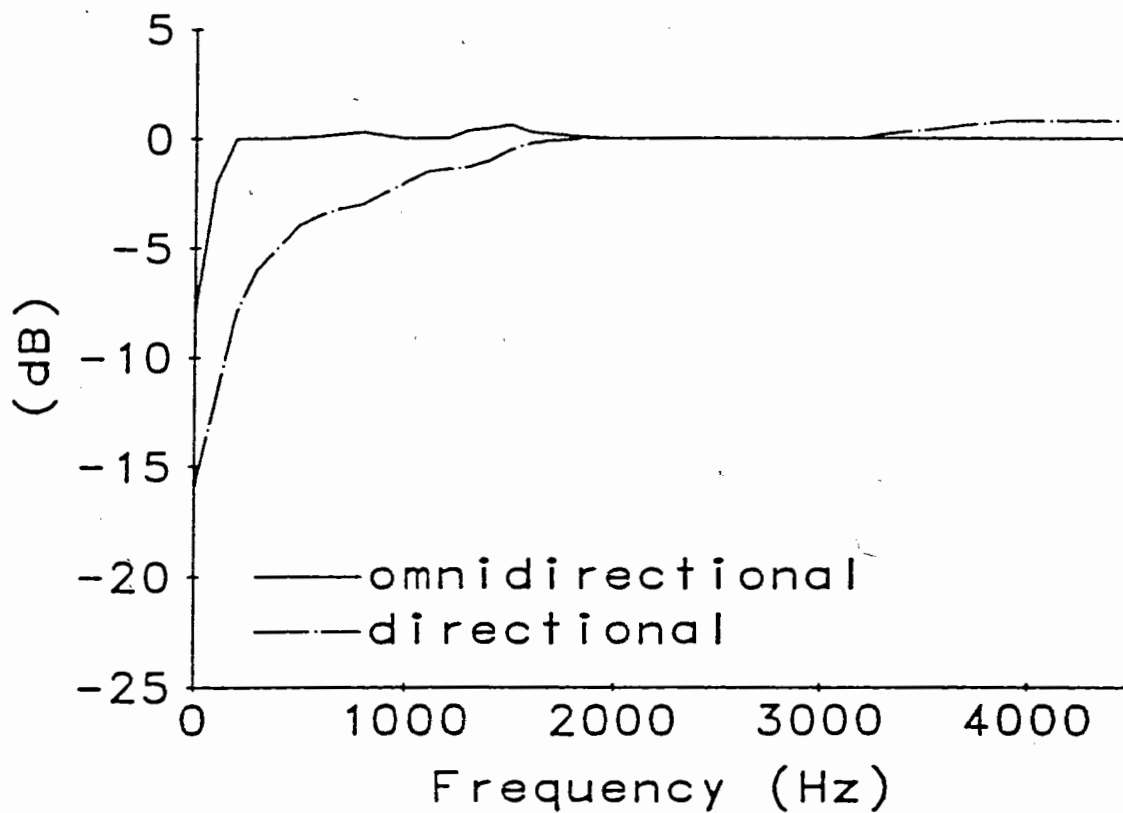
Figure 2.2 Frequency Response of Microphones.
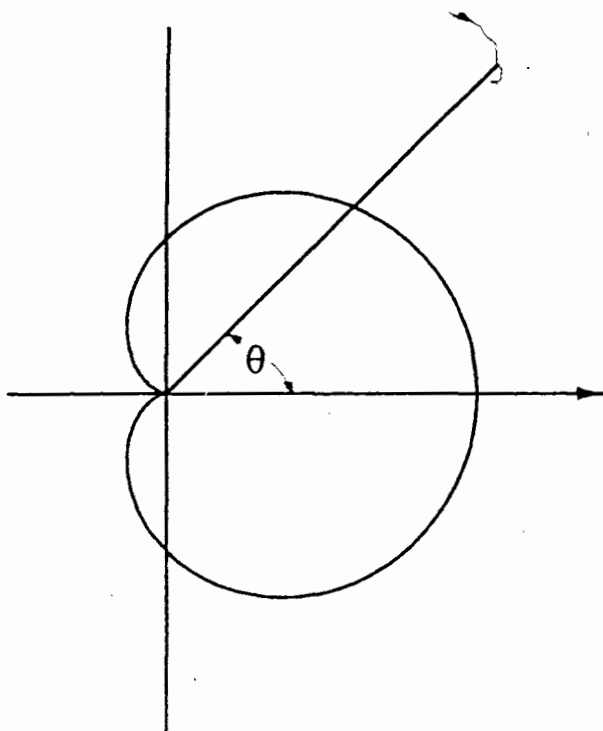
Figure 2.3 Directionality Pattern for Cardioid Response.

To quantify the distortion of the recording equipment, the circuit shown in figure 2.4 was used. The same signal was fed into two channels of the mixer, and the difference between the two outputs measured on the oscilloscope. The distortion of the mixer/amplifier, defined as the difference in amplitude between the outputs divided by the amplitude of the input, was measured and found to be less than 0.5% over the frequency range of interest. (Sine waves at various frequencies from 100 Hz to 4.5 kHz were used for the measurements, and pseudo-white noise was used to confirm the broadband performance.) The distortion of the entire data acquisition system, about 1.5%, was measured by recording the mixer output and comparing the subsequent outputs of the recorder.

To measure the cross-channel interference of the recorder, the outputs were measured after recording with an input only on one channel. The interference, defined as the output of the channel without an input divided by the output of the channel with the input, was less than 1%.

The frequency responses of the tape-recorder, tapes and amplifier were measured by observing the output of the recorder when sine waves of constant amplitude and varying frequency were applied. The response was flat well beyond the frequencies of interest;
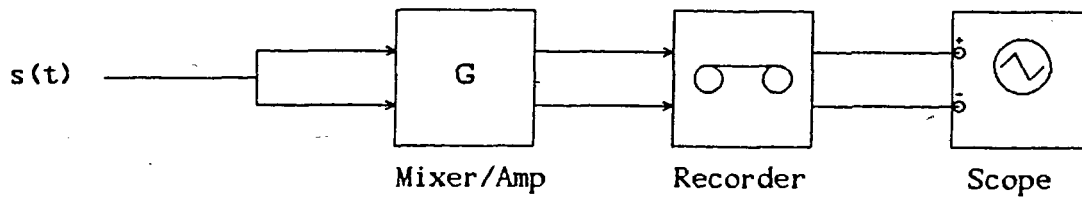
Figure 2.4 Circuit to Test Fidelity of Recording Equipment.

the tape recorder had a low-pass cutoff frequency of approximately 13kHz, which made the analog anti-aliasing filter redundant as only frequencies above 18 kHz would be aliased.

The digital low-pass filter was created by applying a 127 point Hamming window to an ideal brick-wall filter.[15] Figure 2.5 shows the filter's frequency response. The attenuation at the Nyquist frequency of 4.5 kHz (after decimating in time) is approximately 60 dB.

2.3. ANALYSIS OF REVERBERATION

In analysing the effects of reverberation, two distinct approaches can be taken.[10] The first is to consider the frequency response of the room by solving the wave equation

$$\nabla^2 p = \left[\frac{1}{c^2}\right]\frac{\partial^2 p}{\partial t^2} \tag{2.2}$$

subject to the boundary conditions imposed by the walls of the room. Here p is the sound pressure and c is the speed of sound. To solve equation (2.2), the technique of separation of variables[16] is usually employed. This results in a solution of the form

$$p(\omega,x,y,z,t) = A(x)B(y)C(z)T(t) \tag{2.3}$$

Such a solution will involve a summation of normal modes of vibration, which occur at frequencies

$$\omega_n = \pi c = \sqrt{(n_x/L_x)^2 + (n_y/L_y)^2 + (n_z/Lz)^2} \tag{2.4}$$
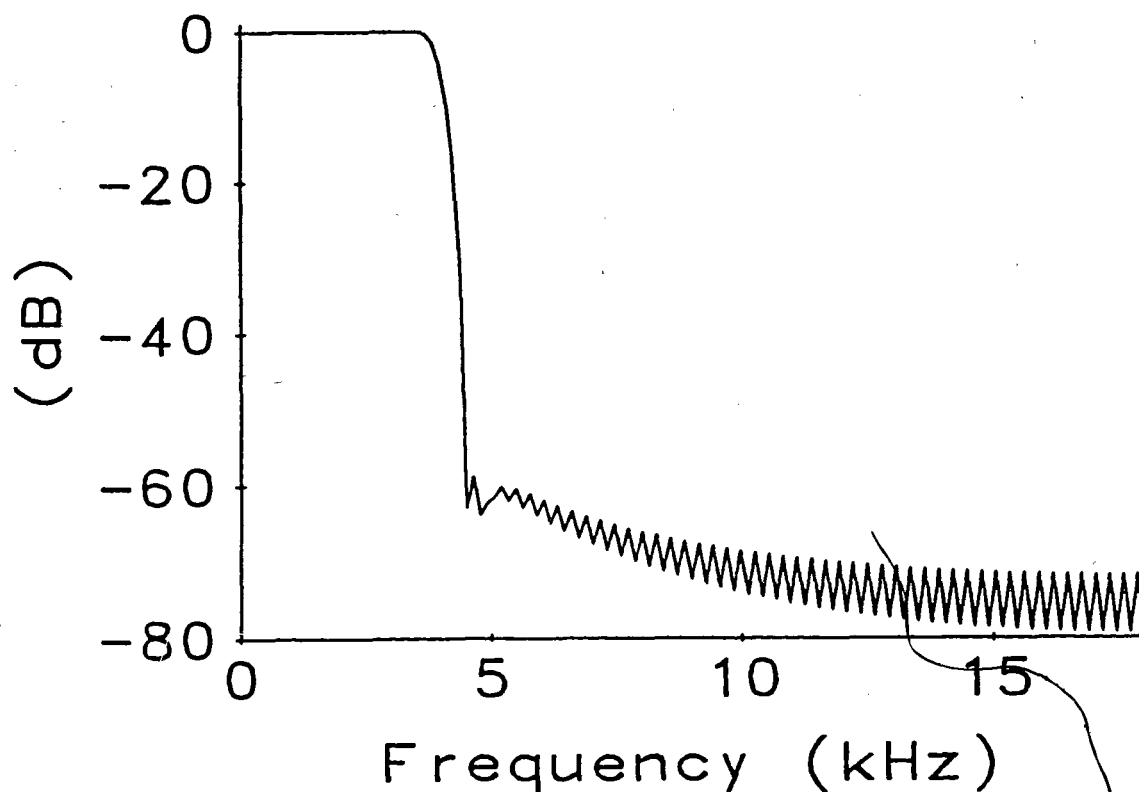
Figure 2.5 Low Pass Filter Frequency Response.

where $n_x, n_y, n_z = 0,1,2, \cdots$, and the room (assumed to be rectangular) is of size L_x by L_y by L_z.

In general, to derive meaningful results, the modes for all frequencies of interest (100 Hz - 4 kHz) must be calculated, plus corrections for modes outside this range. To determine the effect of each mode, a set of transcedental equations must be solved to find the pole location, and the residue of the pole must be evaluated to find the mode gain. It has been shown[17] that between 300 and 350 modes are required to simulate the low-frequency response to a room. Berman[10] points out that these modes don't really exist in the form of standing waves; rather they are mathematical abstractions in the same way that we describe the frequency components of a square wave by taking the Fourier Transform.

The second, and more intuitive approach is to study time domain behavior, in which sound propagates out from the source and is reflected between the walls until the sound energy dies away. This approach has the advantage of being easier to visualize and model, and provides a better explanation of the behavior of enclosures.[10]

The analysis of the effects of reverberation on acoustical signals based on the time domain approach can be found in Morse and Ingard.[18] The basic assumptions behind the analysis are:

1) In analogy to the light rays of geometrical optics, we consider may sound as a collection of acoustic rays.

2) The acoustic energy density W is uniformly distributed throughout the enclosure.

3) The acoustical absorption of the walls of the enclosure may be adequately represented by a single absorption coefficient $\alpha(\vec{r}_s)$, defined to be the average fraction of incident acoustic power absorbed by the wall near the point $\vec{r}_s$. $\alpha(\vec{r}_s)$ is averaged over the hemisphere of all directions of incidence under the assumption of isotropic distribution of incidental rays. If the individual walls are assumed to have constant $\alpha(\vec{r}_s)$, then we may define the absorption of the room, a, by the equation

$$a = \sum_i \alpha_i S_i \qquad (2.5)$$

where α_i and S_i are, respectively, the absorption coefficient and area of wall i. The unit for a is the metric sabin which has the dimensions of inverse volume.

If conservation of acoustical energy is applied, the rate of change of the total energy is equal to the difference between the power introduced into the enclosure by the source $F(t)$, and the power absorbed by the walls, $aI(t)$, where $I(t)$ is the intensity of the sound incident on the walls. Since the energy density w is assumed constant, the total power is merely the product of w and the volume V. The solution is

$$I(t) = \frac{c}{4V} e^{-act/4V} \int_{-\infty}^{t} e^{ac\tau/4V} F(\tau)\, d\tau \qquad (2.6)$$

The intensity at a given time depends on the power output $F(t)$ for the previous $4V/ac$ seconds, but depends very little on $F(t)$ before that time. If F(t) varies widely in a short time compared with $4V/ac$, the intensity will follow $F(t)$ and the sound will be 'blurred'.

If $F(t)$ is an impulse occuring at time $t = 0^-$, the intensity will be

$$I(t) = I_0 e^{-\frac{act}{4V}} \tag{2.7}$$

Thus the reverberation (and therefore the distortion) increases as the enclosure's volume increases, and as the absorption decreases.

The effect of reverberation on the room's frequency response is difficult to see from either equation (2.3) or (2.6). As will be seen from the model of section 2.2, typically the overall response is quite flat, with a number of notches in it.

In section 2.3, experimental results are given for the reverberation in automobiles. We now consider the problem of modelling a reverberant enlcosure.

2.4. MODELLING REVERBERATION - THE IMAGE MODEL

As discussed in section 2.1, the computational expense of using modes to simulate reverberation effects is prohibitively high. By extending the assumptions of geometrical acoustics, we can derive a model for reverberant enclosures that can be implemented on a computer.[6]

We seek to find the transfer functions between the source and microphone, both assumed to be points. For simplicity, we assume the enclosure is rectangular (clearly this is a very gross approximation to a vehicle enclosure). In investigating a computer model for the automobile, we are not attempting to simulate the exact reverberation response of the automobile interior. Rather, we are seeking principles that can be inferred from our rectangular model and extended to the practical application. This will become important when investigating the performance of our array solution in chapter 5.

Using the notation of Allen and Berkley,[6] let

p = pressure

X = source location(x,y,z)

X' = receiver location(x',y',z')

$R = |X - X'|$

Then if the source is an acoustic pressure impulse originating at X, the sound pressure received at point X' is

$$p(t,X,X') = \frac{\delta(t-\frac{R}{c})}{4\pi R}$$ (2.8)

The denominator reflects the fact that received pressure is inversely proportional to the distance between source and receiver.

Now consider looking at a point light source placed in front of a mirror as in figure 2.6(a). To the observer at X' (the asterisk), there are now two apparent light sources (the plus signs), the original at X and its image, which appears to be on the other side of the mirror at the same distance from the mirror as the actual source. The same principle applies to acoustic systems. A rigid body (the wall) is a perfect reflector of acoustic energy, so if the source is a pressure impulse, the received pressure is

$$p(t,X,Y) = \frac{\delta\left[t-\frac{|X-X'|}{c}\right]}{4\pi|X-X'|} + \frac{\delta\left[t-\frac{|X+X'|}{c}\right]}{4\pi|X+X'|}$$ (2.9)

If now we consider the case of a rectangular room with perfectly rigid walls, the source now has six initial images, one for each wall. In turn, each of these images produces six more images which produce their own six images and so on and so on... The images appear at positions $(x \pm x' + 2nL_x, y \pm y' + 2lL_y, z \pm z' + 2mL_z)$.

Figure 2.6b shows a two-dimensional projection of the image positions for a source in a cubical room. The solid square represents the original room and the asterisk (*)

represents the microphone. The minus sign (-) is the original source and the plus signs (+) are its image sources, one in each of the "virtual" rooms (each wall and virtual wall acts as a reflector). Note that the actual image space is three dimensional.

Defining $\vec{R_r} = (2nL_x, 2lL_y, 2mL_z)$ and $\vec{R_p} = (x-x'+2qx', y-y'+2iy', z-z'+2kz')$, and considering all images, the pressure at the receiver is given by

$$p(t,X,X') = \sum_p \sum_r \frac{\delta\left[t - \frac{\left[|\vec{R_p}+\vec{R_r}|\right]}{c}\right]}{4\pi|\vec{R_p}+\vec{R_r}|} \qquad (2.10)$$

where the sum over r is over the eight combinations of $q,i,k = 0,1$ and the sum over p is infinite, but truncated by practical modelling constraints.

(a)

(b)
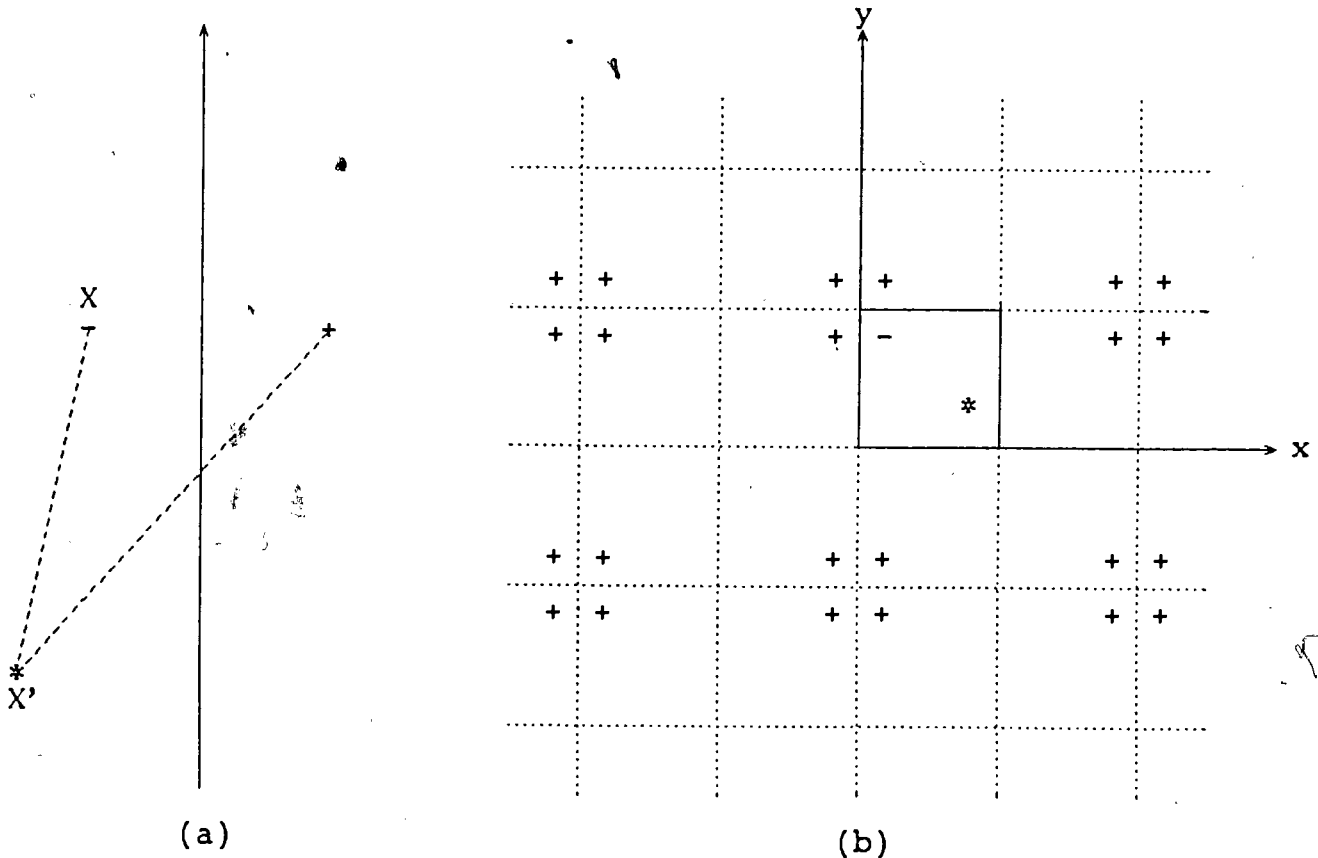
Figure 2.6 Location of Image Sources for (a) a single wall and (b) part of a 2-dimensional slice through a room (the image volume is 3-dimensional and infinite).

2.4.1. Effect of Lossy Walls

The effect of absorbent walls on the image model is not presently understood. In fact, the effects on even a single image are very complicated.[19] After Allen and Berkley[6] we assume the approximate point image remains valid for lossy walls.

The absorbing properties of the walls are assumed to be represented with sufficient accuracy by an angle independent pressure wall reflection coefficient β_i. (Note $\beta_i = (1-\alpha_i)^{\frac{1}{2}}$ where α_i is the absorption coefficient of wall i.) Furthermore, β_i is assumed to be independent of frequency. When the β_i are introduced into equation (2.10), the impulse response becomes[6]

$$p(t,X,X') = \sum_{p=0}^{1} \sum_{r=-\infty}^{\infty} \beta_{x_1}^{|n-q|} \beta_{x_2}^{|n|} \beta_{y_1}^{|l-j|} \beta_{y_2}^{|l|} \beta_{z_1}^{|m-k|} \beta_{z_1}^{|m|} \frac{\delta\left[t - \frac{[|\vec{R}_p + \vec{R}_r|]}{c}\right]}{4\pi|\vec{R}_p + \vec{R}_r|} \tag{2.11}$$

2.4.2. Low Pass Enhancement

Equation (2.11) describes the pressure impulse response as a summation of image arrivals ("echoes") of known amplitudes. In the original paper by Allen and Berkley,[6] the arrival time of each echo was quantized to the nearest sampling instant. The impulse response was the sum of these echoes. While this method is simple, and has been successfully applied to a number of problems,[20,21] the following argument from Peterson[11] shows simple quantization results in a form of aliasing, and doesn't preserve correct interchannel phase when simulating the response of multiple microphones to a single impulse.

Consider the acoustic system that we are attempting to model (as in figure 2.1). The sampled impulse response of the room is recorded by passing an acoustic impulse through the room to the microphone. Before sampling, the microphone signal must be passed through a low-pass filter with cut-off frequency less than half the sampling frequency in order to satisfy Nyquist's criterion for no aliasing. For a single impulse (or echo), the impulse response of the filter spreads the response over several samples.

Conversely, the sampling theorem says that only band-limited impulses can be unambiguously represented in discrete-time. The model must represent the room response cascaded with a continuous real-time low-pass filter. For each echo in equation (2.11), the exact arrival time should be determined, and then the output of a continuous-time low-pass filter sampled at appropriate instants. It was decided to use the same filter as Peterson: a 41 point Hanning Window applied to a sinc function:

$$h(t) = [1 + \cos(2\pi t/T_W)]f_c \, \text{sinc}(2f_c t) \quad \text{for} \quad -\frac{T_W}{2} < t < \frac{T_W}{2} \tag{2.12}$$

where $\text{sinc}(x) = \dfrac{\sin \pi x}{\pi x}$, $T_W = 40/9$ kHz = 4.44 msec, the window duration, and $f_c = 4$ kHz, the cutoff frequency of the filter.

Due to a large d.c. residual from the model[6] it was also necessary to high-pass filter the response. Both the modelled and experimental impulse responses were passed through a digital high-pass filter with cutoff at 100 Hz.

2.4.3. Directionality of Source and Microphones

While other researchers have expressed an interest in enhancing the image model by incorporating directionality into the source and receiver,[20] none have appeared to have done so. Such an enhancement is presented in the following discussion.

Consider the case of a cardioid microphone. To an echo arriving from direction $\vec{R} = \vec{R}_p + \vec{R}_r = (x_r, y_r, z_r)$, a microphone with axis along $\vec{A} = (x_a, y_a, z_a)$, has response given by equation (2.1)

$$\text{Micresponse}(\theta) = \tfrac{1}{2}(1 + \cos\theta) \tag{2.13}$$

where θ is the angle between vectors $\vec{R}$ and $\vec{A}$. But from linear algebra theory,[22] the angle between any two vectors $\vec{X}$ and $\vec{Y}$ in space is given by

$$\cos\theta = \frac{\vec{X} \cdot \vec{Y}}{|\vec{X}||\vec{Y}|} \tag{2.14}$$

Thus to model a cardioid response, for every echo in equation (2.11), $\vec{R}$ and $\vec{A}$ are

determined and normalized so that equation (2.1) becomes

$$\text{Micresponse} = \tfrac{1}{2}(1 + x_r x_a + y_r y_a + z_r z_a) \tag{2.15}$$

This technique will work for any directionality pattern that is a function of cosines; fortunately, most directional microphones have such patterns. The one limitation of the method lies in the implicit assumption that the directionality pattern is independent of frequency (but we have previously assumed that the β_i are also frequency independent).

For our application, omnidirectional sources were employed so directionality was only applied to the microphones. Directional sources can be handled in a similar manner. Appendix 3 gives a Fortran listing for the program implementing the image model, with the enhancements of low-passed impulses and directional microphones included.

2.5. EXPERIMENTAL AND MODELLED IMPULSES

The major problem associated with determining the acoustic transfer functions is finding a practical acoustical impulse source. Two types of sources have been used in the literature: starter's pistols[4] and electric spark generators.[10] Neither method was satisfactory for our purposes, because we were attempting to record near-field responses. Starter's pistols (cap guns in our case) were found to be too loud, and also to produce an impulse that was too long in duration. When an electric spark generator was used, the microphones responded to the large electric fields present, rather than the acoustic field. For our recordings, we used the centres from Christmas Crackers (small fire-cracker like devices.)

Figures 2.7 and 2.9 show a typical impulse from a Christmas Cracker recorded in a near-anechoic environment (in a field) by a directional and omnidirectional microphone. The spectra of the impulses, shown in figures 2.8 and 2.10 are hardly flat, but at least there are no significant nulls. The effects of the low-pass filter (cutoff at 4kHz) and the high-pass filter (cutoff at 100Hz) are evident. The directional gradient microphone doesn't exhibit the expected low pass response below 1 kHz[14] due to the fact that the impulses were in the near-field, some 20 cm away from the microphones. In the near field, where the sound arrives as spherical rather than plane waves, the response of the

microphone changes,[14] and no longer attenuates low frequency sound.

While there was some variation in the impulse shapes (the First Law of Acoustics is "Never repeat a measurement!"[9]), it is important to note that in all cases, the relevant acoustical properties (decay rate, performance of array in chapter 5) were consistent among trials.

To confirm the validity of the image model in a small enclosure, recordings of the impulse response were made in a rectangular cement tank with a wooden top, and dimensions 2.4 by 1.2 by 0.8 m. The microphones were at position (0.96,0.53,0.38) while the impulse was at (0.75,0.53,0.38), where the co-ordinates are measured in meters, relative to an origin at one of the bottom corners of the tank. The enclosure being sealed, there was concern that the omnidirectional microphones might have trouble with the large pressure when the impulse was ignited. The directional microphones, responding to the pressure gradients rather than absolute pressure were expected to perform better.

The impulse responses for both a directional and omnidirectional microphone (recorded simultaneously) are shown in figures 2.11 and 2.12. The acoustical energy remains significant for at least 400 msec.

When attempting to model the responses, the non-perfect nature of the impulse sources must be accounted for, by convolving the initial modelled responses from equation (2.11) with the isolated impulse responses. When reflection values of $\beta_i = 0.995$ for the floor and walls (concrete) and $\beta_i = 0.95$ for the ceiling (wood),[23] were used, the resulting modelled impulse responses for the directional and omnidirectional microphones are shown in figures 2.13 and 2.14.

The model does appear to predict the reverberant tail for the directional microphone reasonably well (comparing figures 2.11 and 2.13), but there appears to be much more reverberant energy in the experimental omnidirectional impulse than in the modelled. This disparity may have been caused by an attenuation of the first arrival, since the plots were all normalized with respect to the peak amplitude of the response (which occurs at the first arrival). Figure 2.14 seems to support this explanation, since a short time after the initial pulse, the response becomes similar to the modelled response. The apparent

attenuation of the initial arrival, which appeared consistently over several trials, is probably attributable to the pressure overload effect we were concerned about.

Figure 2.15 shows the impulse response recorded by a directional microphone in an automobile. The reverberant energy appears to last for about 30 msec, which is longer than the 25 msec reported by Goubran and Hafez[4] possibly due to the duration of our impulse source.

One expects that the image model may have a more difficult time modelling the acoustical impulse response in a car interior than a rectangular cement tank; afterall, the car interior is hardly rectangular, and the reflection coefficients aren't constant along the "walls". Unfortunately, the expectation is true. The modelled impulse response is given in figure 2.16. The dimensions used for the model are 1.1 X 0.8 X 1.0 m, much smaller than the physical dimensions of the car. If one considers the actual acoustical environment where the recording was made, where there are effectively three reflective walls (the windows and windshield) and three absorbent walls (the seat or floor, the ceiling and the seat back), it is no surprise that a rectangle with dimensions the actual width of the car, the height from the celing to the seat (not floor) and length from the windshield to the front seat most closely models the actual car.
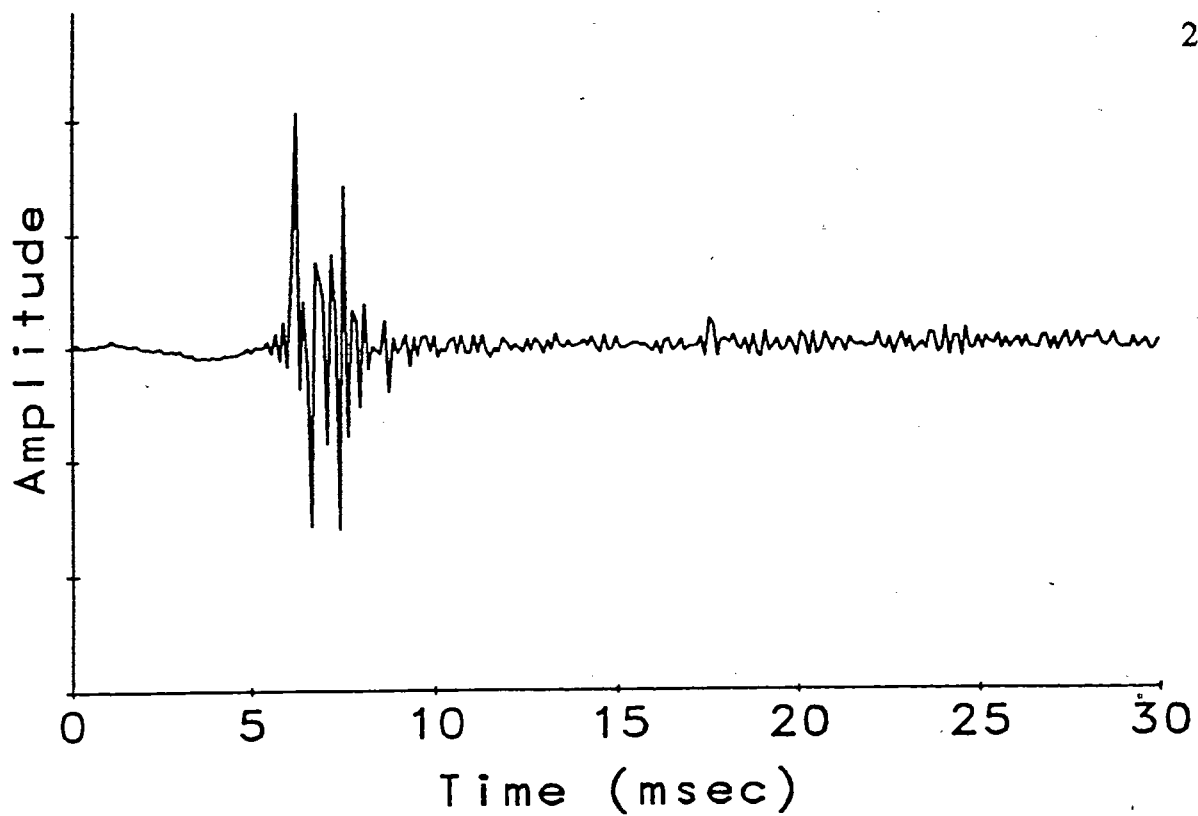
Figure 2.7 Isolated Impulse Recorded by Directional Micro-
phone
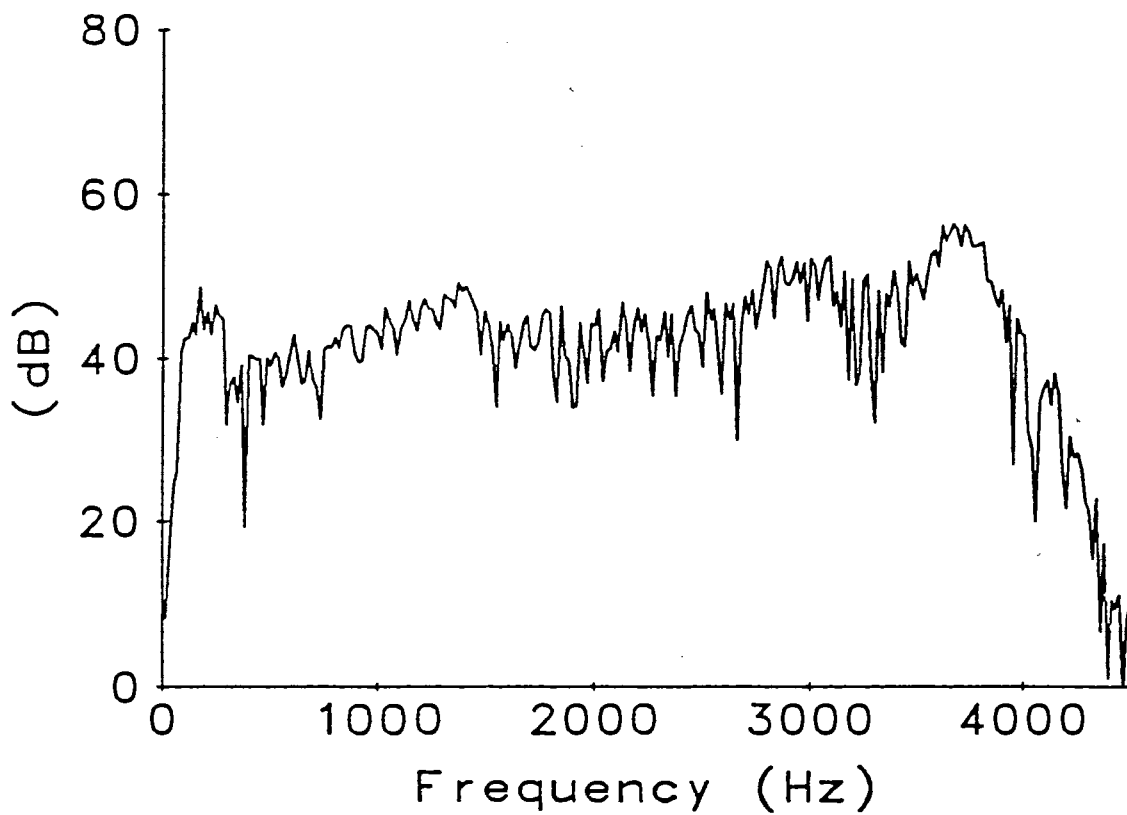
Figure 2.8 Spectrum for Isolated Impulse Recorded by Direc-
tional Microphone
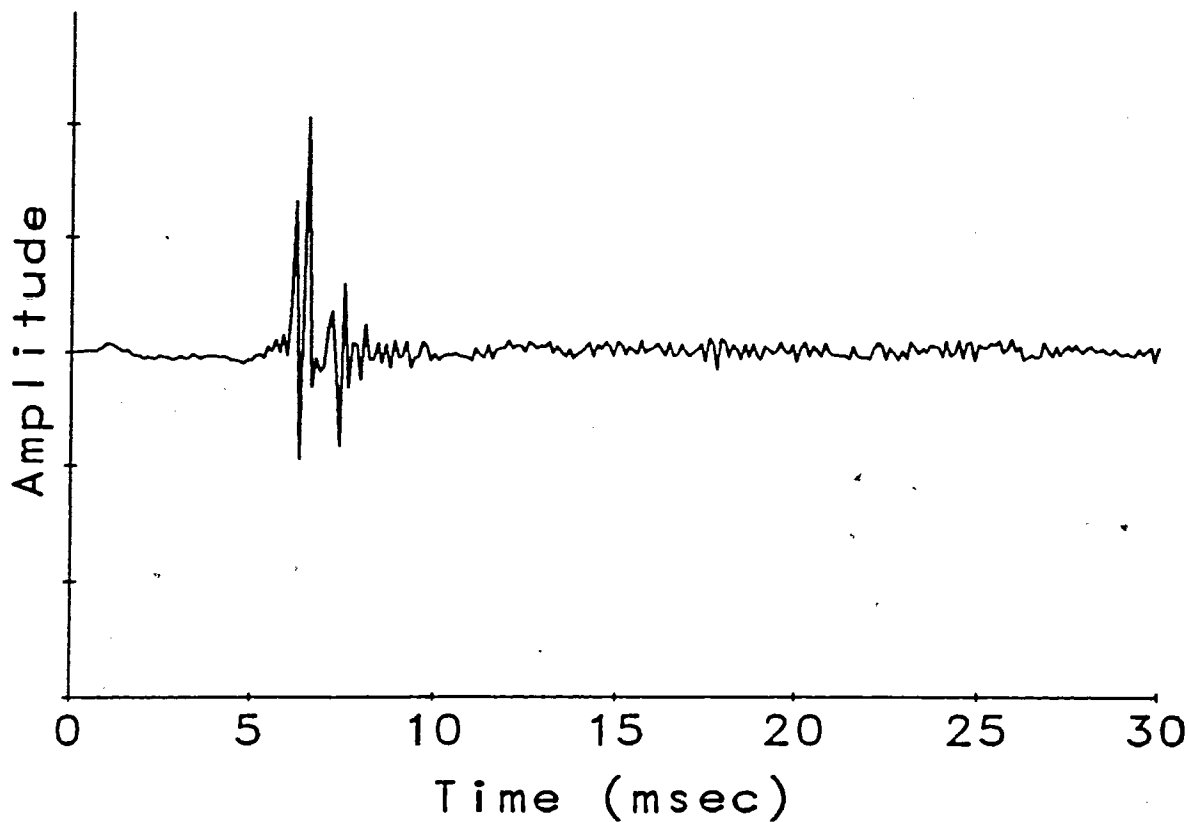
Figure 2.9 Isolated Impulse Recorded by Omnidirectional Microphone
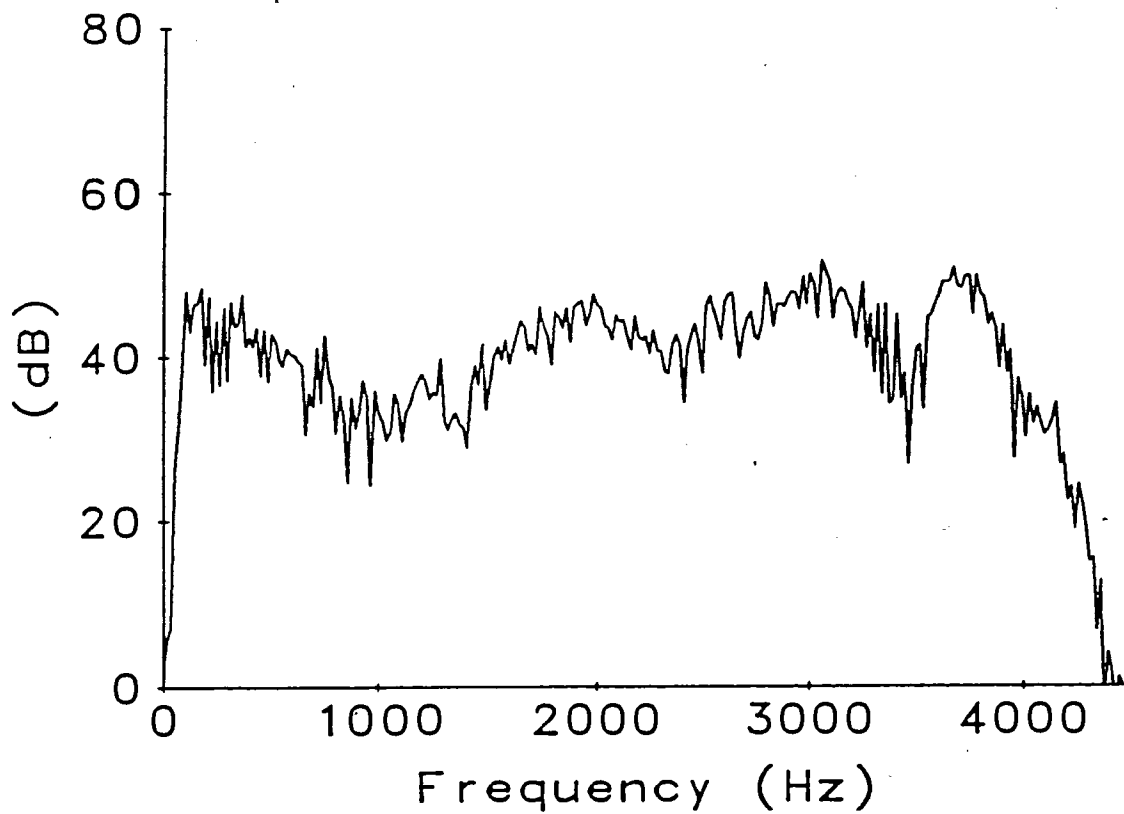
Figure 2.10 Spectrum for Isolated Impulse Recorded by Omnidirectional Microphone
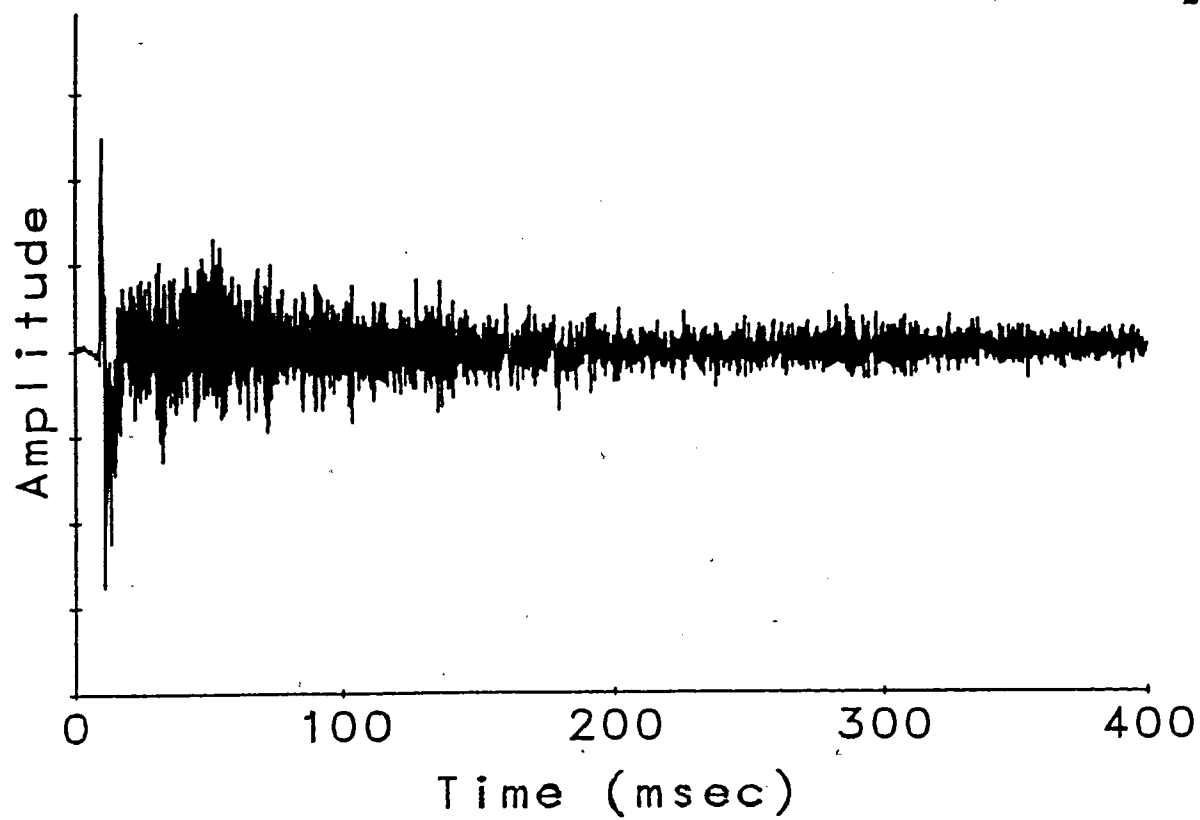
Figure 2.11 Impulse Response Recorded by Directional Microphone in Cement Tank
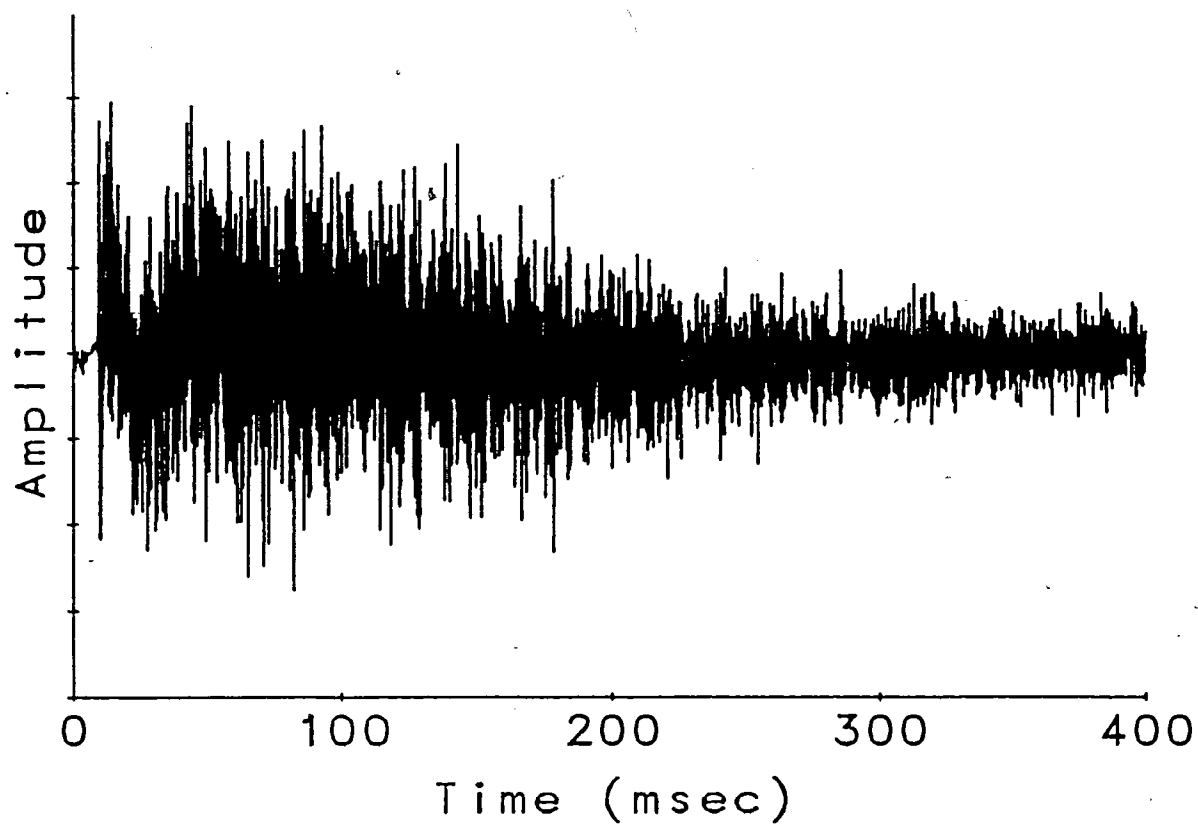
Figure 2.12 Impulse Response Recorded by Omnidirectional Microphone in Cement Tank
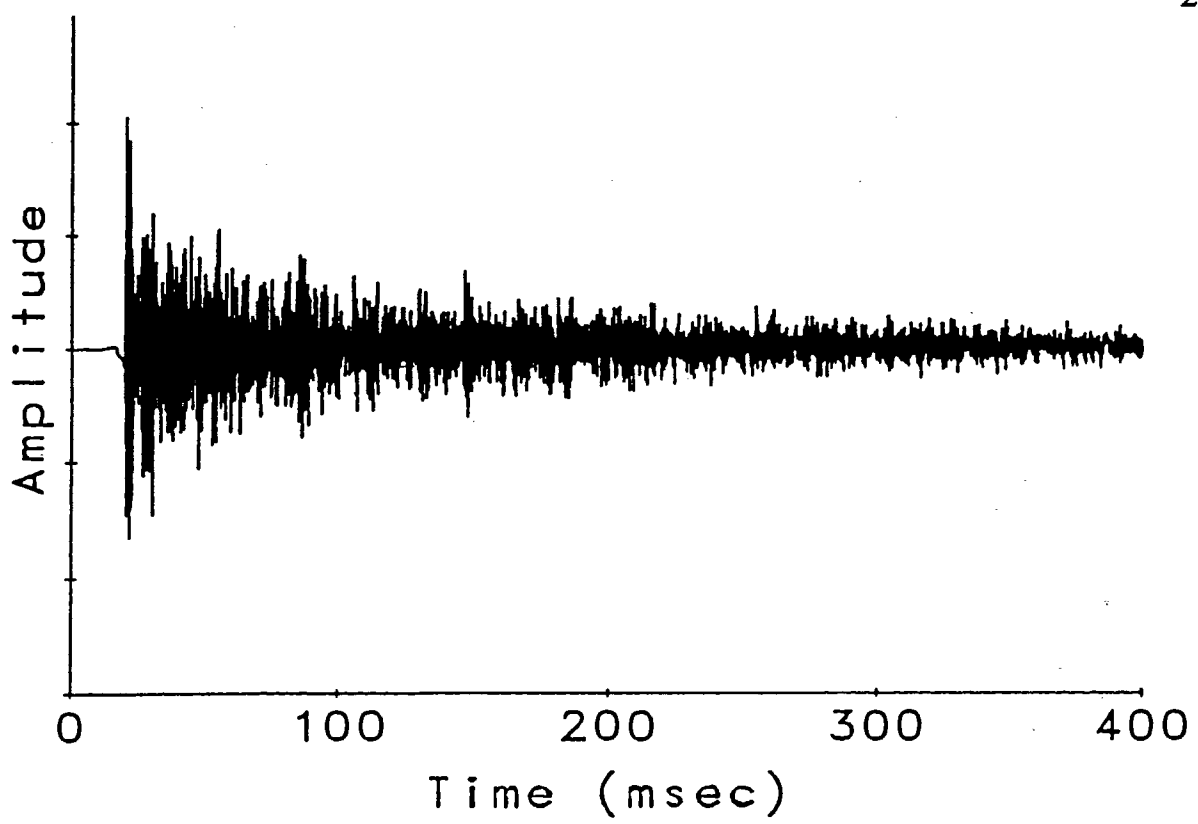
Figure 2.13 Modelled Impulse Response for Directional Microphone in Cement Tank
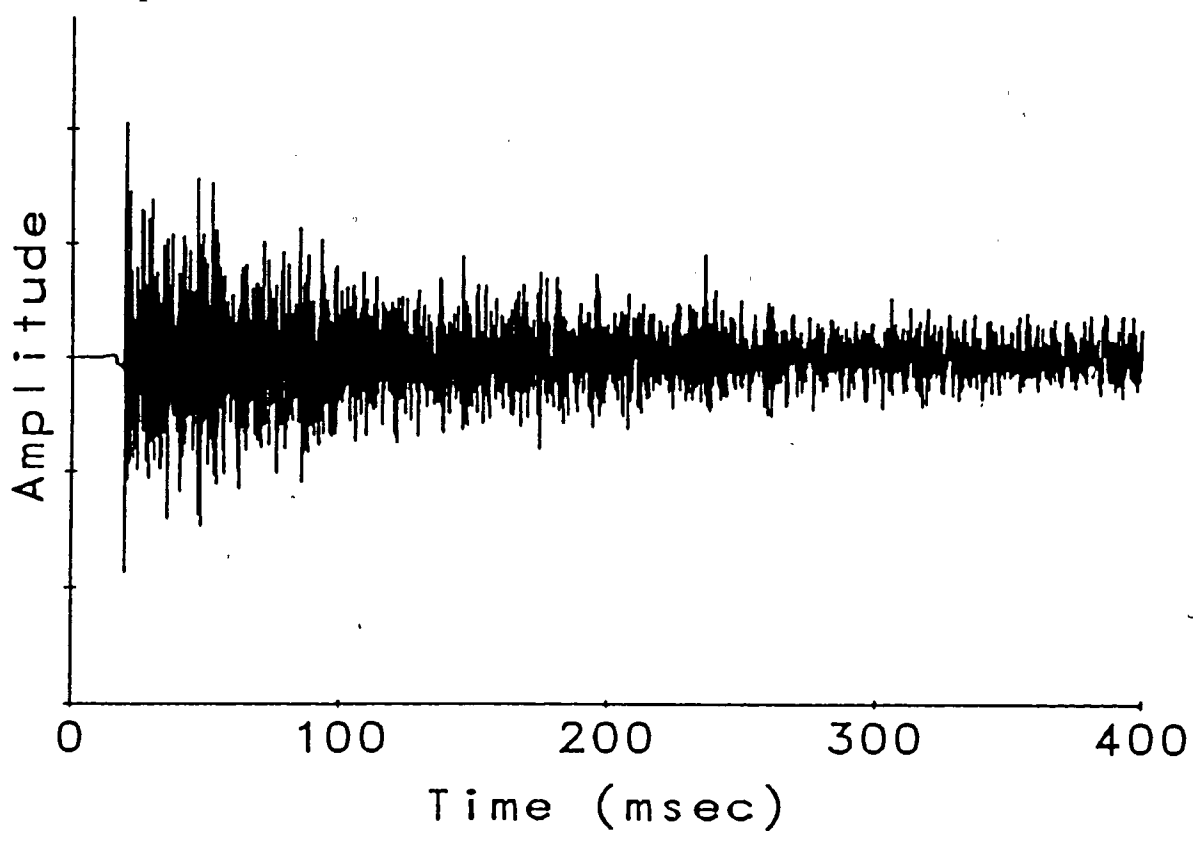
Figure 2.14 Modelled Impulse Response for Omnidirectional Microphone in Cement Tank
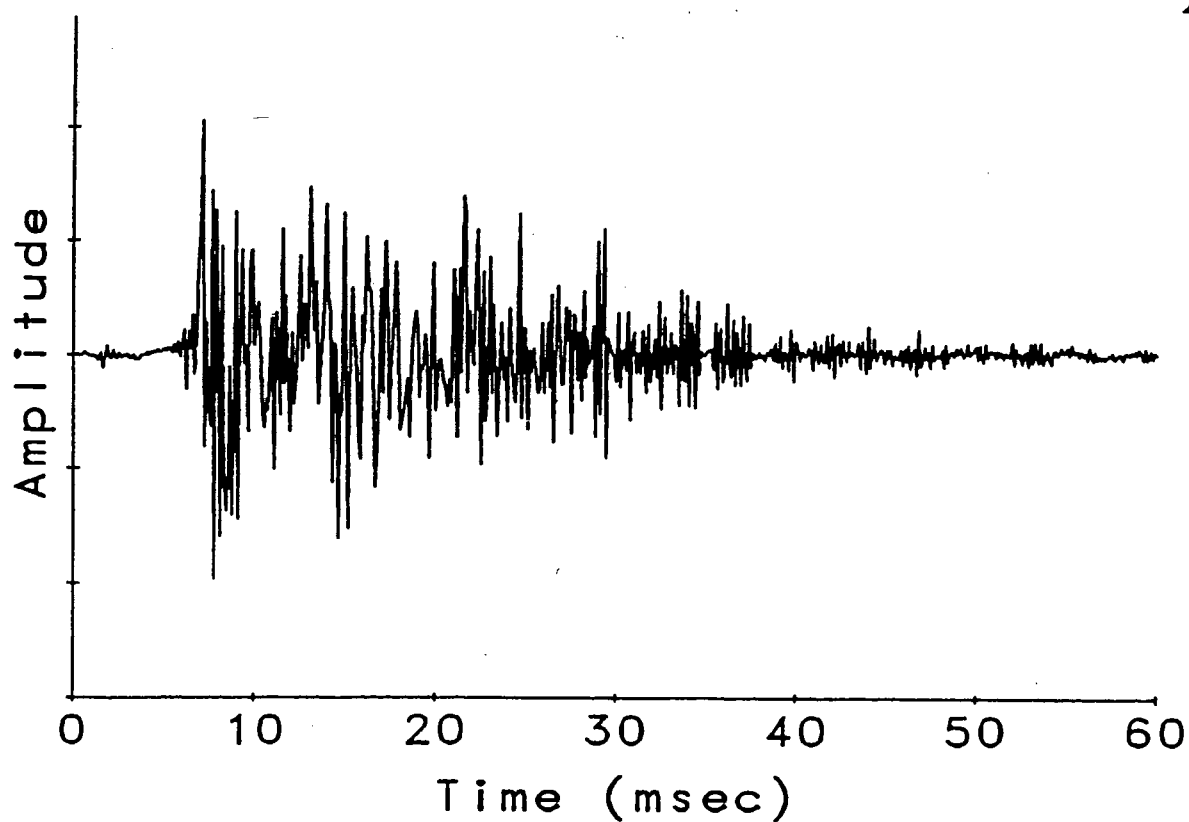
Figure 2.15 Impulse Response Recorded by Directional Microphone in Automobile
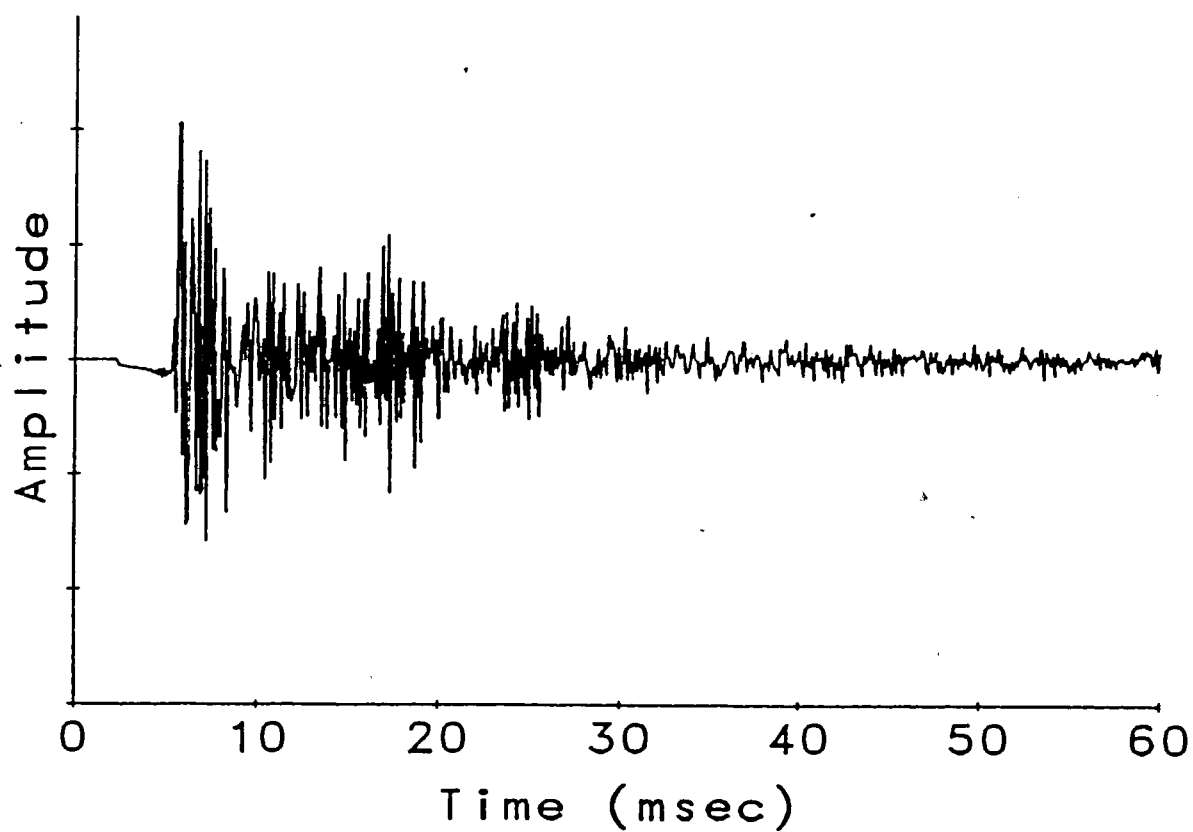
Figure 2.16 Modelled Impulse Response for Directional Microphone in Automobile

2.6. FREQUENCY RESPONSE

Due to the non-flat characteristics of the isolated impulse (figures 2.9 and 2.10), we cannnot expect to predict the frequency response by studying the actual recorded impulse responses. To get a feel for the frequency response of a reverbant enclosure, we took the Fourier Transform of the modelled impulses (before convolution with the isolated impulses). For example, the spectra for the modelled impulses of the cement tank (figures 2.13 and 2.14) are shown in figures 2.17 and 2.18. The spectra are moderately flat over the pass band from 100 to 4000 Hz, but exhibit a number of notches. The deviation from a flat response (which we would see if there were no reverberation present) is a measure of the distortion introduced by the reverberation. As a figure of merit, we took the variance of the impulse frequency response measured in dB, over the passband and called it the 'Spectral Flatness'. Comparing figures 2.17 and 2.18, the response for the directional microphone has a spectral flatness of 29.1 compared with 48.6 for that for the omnidirectional. Figure 2.19 shows the frequency response of the model for the car. Its spectral flatness of 24.7 reflects the lesser degee of reverberation.

As expected, we have been able to create a model of the acoustical environment of the automobile that is only approximate. With this model, it is possible to evaluate various signal enhancement techniques (as will be done with our array solution in chapter 5). From a speech enhancement point of view, reverberation has two main effects:

1) The speech is distorted, as evidenced by figure 2.17. In general, the distortion is very strongly dependent on the locations of the speaker, microphone and the physical layout of the car interior. Therefore, the usefulness of reverberant energy for speech recognizers is expected to be limited, due to the large changes in the reverberant speech that result from minute changes in the acoustical environment.

2) A given noise source has several images, which cause the microphone to receive noise from several directions. In the case of the car, where there are already many sources, the reverberation will cause the noise field to become isotropic. Chapter 3 will next consider the noise field in further detail.
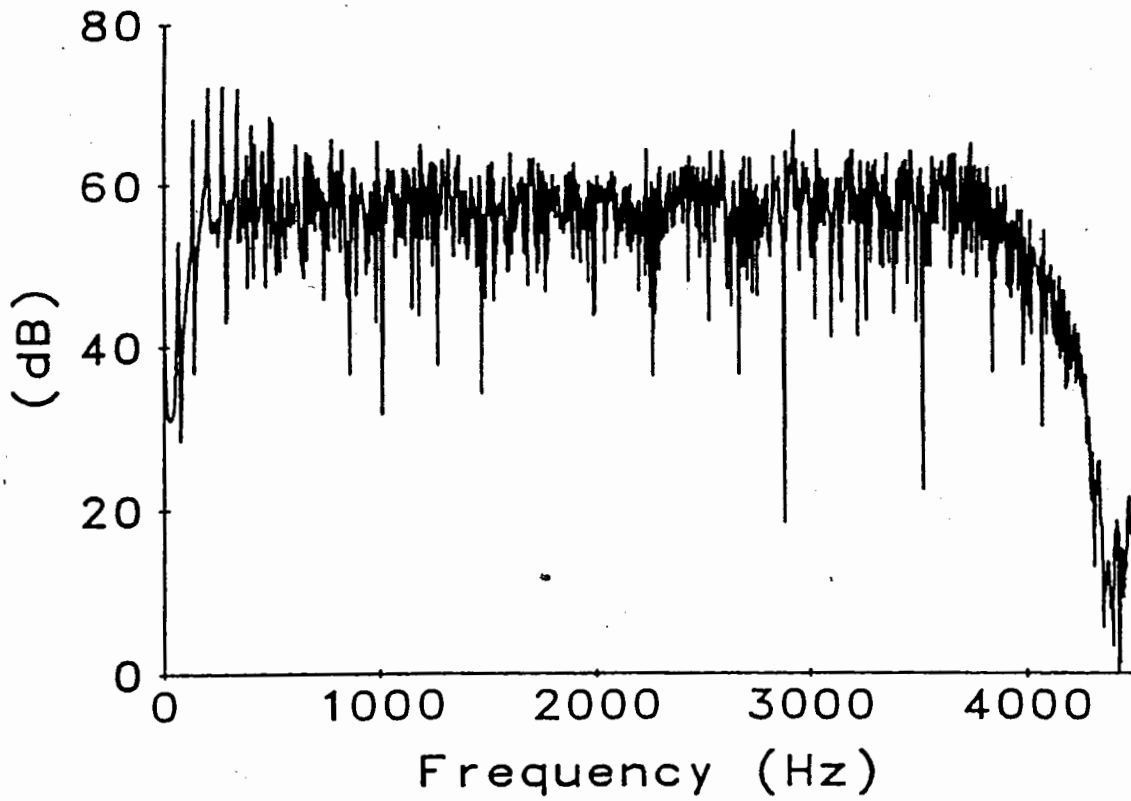
Figure 2.17 Frequency Response of Model for Directional Mic in Cement Tank
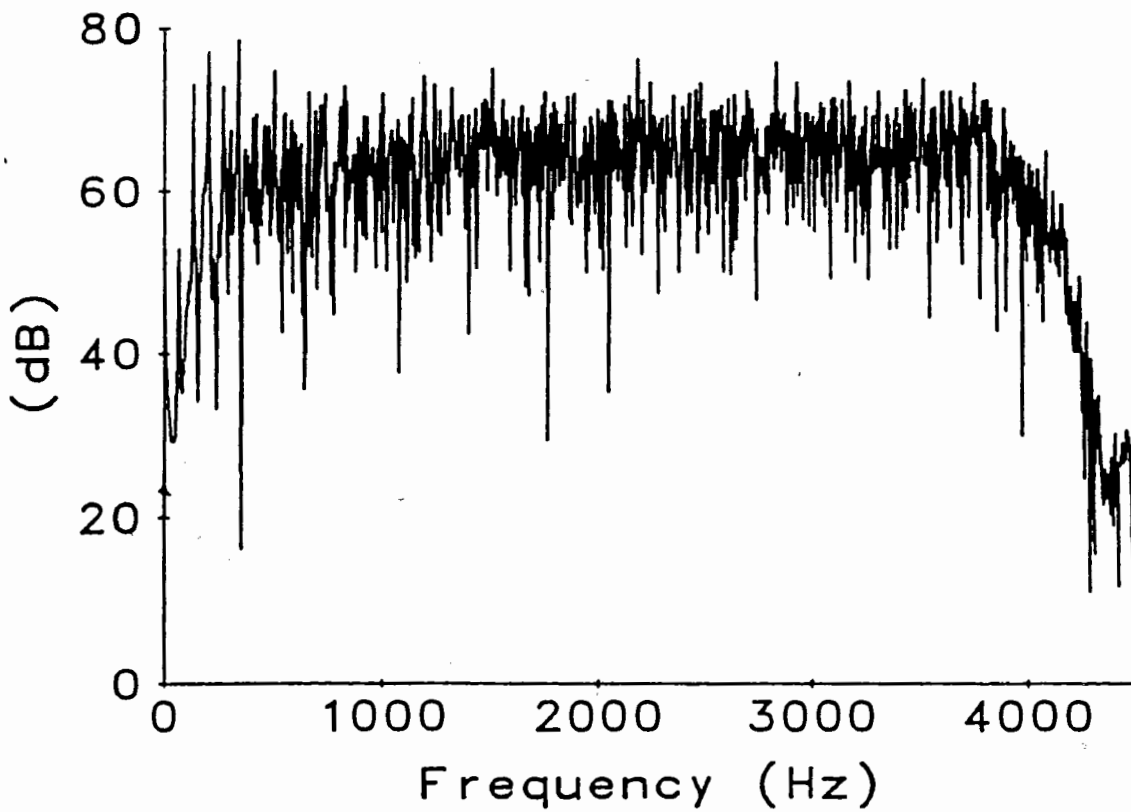
Figure 2.18 Frequency Response of Model for Omni-directional Mic in Cement Tank
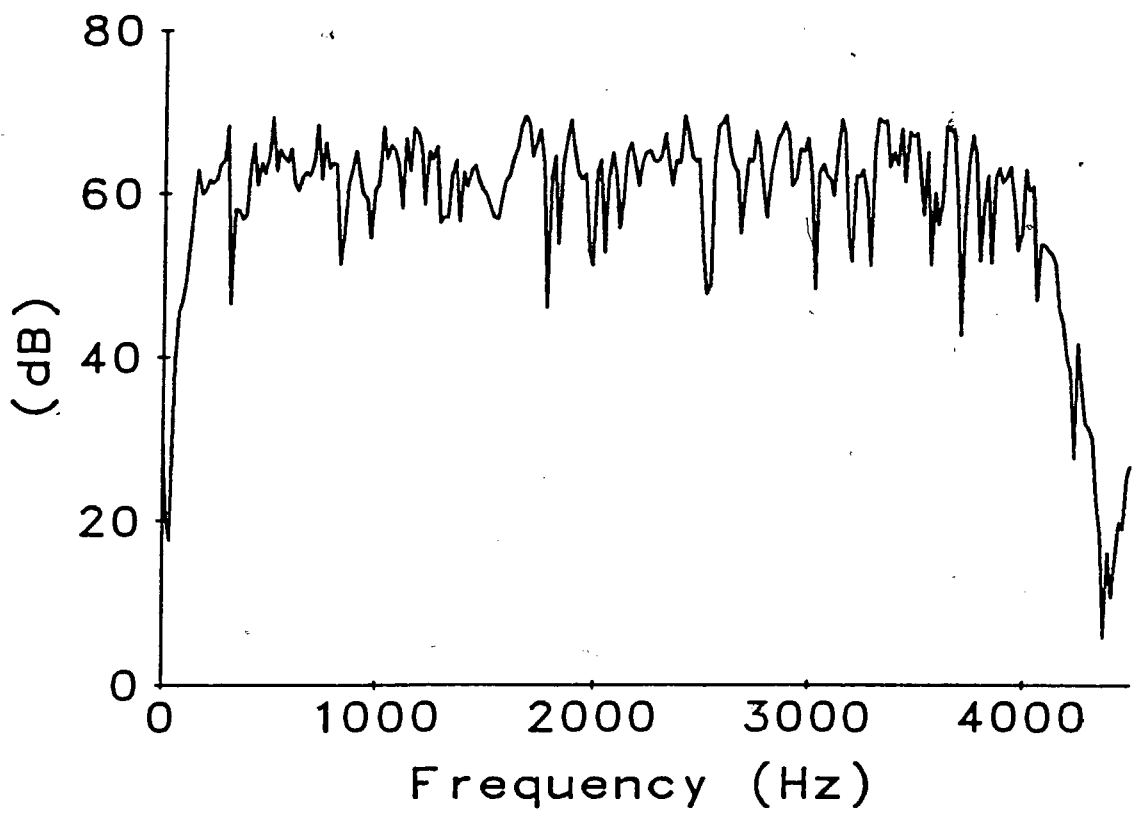
Figure 2.19 Frequency Response of Model for Car

3. THE NOISE ENVIRONMENT

In chapter 2 we considered reverberation and its effect on speech. Before considering speech enhancement techniques, we must investigate and understand the other half of the problem, namely the noise. The power spectral density of the noise in an automobile interior is investigated under a number of conditions in section 3.1. The spatial correlation properties of the noise field are investigated in section 3.2 with the coherence function. We will see that both the spectral power density and the coherence properties of the noise field will determine the effectiveness of any enhancement algorithms.

3.1. SPECTRAL PROPERTIES

This section describes some of the spectral properties of the noise field found in automobiles. We begin with spectral estimation methods in section 3.1.1, and give results in section 3.1.2.

3.1.1. Spectral Estimation

Welch's method of averaged modified periodograms[24] was employed to develop estimates of the noise spectra. While other methods, including the autoregressive method, and the maximum entropy method[25] are known to give greater spectal resolution, in this application where the spectra are relatively smooth, Welch's method is preferred due to its simplicity.

Consider estimating the power spectral density of signal $x(n)$, $S_x(e^{j\omega})$ by Welch's method. First we extract an interval of data of length 4096, and split it into 15 segments of length 512, with a 256 point overlap of between successive segments. The individual segments are then windowed, and the Fast Fourier Transform applied to each. Finally the average of the magnitude-squared of the transforms is calculated and normalized, resulting in the spectral estimate. Specifically:

$$S_x(e^{j\omega}) = \frac{1}{15} \sum_{k=0}^{14} I^{(i)}(e^{j\omega}) \qquad (3.1)$$

with $J^{(i)}$ = the magnitude of the square of the Fourier Transform of windowed interval i:

$$J^{(i)}(e^{j\omega}) = \frac{1}{512U}\left|\sum_{n=0}^{511} x^{(i)}(n)w(n)e^{-j\omega n}\right|^2 \tag{3.2}$$

where $x^{(i)}(n)$ is the nth sample in the ith segment and U is the normalizing factor for the energy in the window:

$$U = \frac{1}{512}\sum_{n=0}^{511} w^2(n) \tag{3.3}$$

A Hamming window was chosen for w(n),

$$w(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{511}\right] \quad \text{for } 0 \le n \le 511 \tag{3.4}$$

as a compromise between bias and resolution. The bias and resolution of a spectral estimate are a function of the window used. For small bias, low sidelobes in the window are required, and for increased resolution a narrow main sidelobe is required. A Hamming window gives lower sidelobes than, say a Hanning window, but a narrower main sidelobe than a Blackman.

Estimation of cross-spectra is a straightforward extension of the above. To estimate $S_{xy}(e^{j\omega})$, the cross-spectrum of x and y, we again extract intervals of length 4096 for both x and y, subdivide them into segments, window the segments and take Fast Fourier Transforms. Rather than take the magnitude squared of the transforms, we take the magnitude of the product of the transform of x and the complex conjugate of the transform of y:

$$S_{xy}(e^{j\omega}) = \frac{1}{15}\sum_{i=0}^{14} K^{(i)}(e^{j\omega}) \tag{3.5}$$

with

$$K^{(i)} = \frac{1}{512U}\left[\sum_{n=0}^{511} x^{(i)}(n)w(n)e^{-j\omega n}\right]\left[\sum_{m=0}^{511} y^{(i)}(m)w(m)e^{-j\omega m}\right]^* \tag{3.6}$$

$w(n)$ and U are defined as above, and $()^*$ denotes the complex conjugate of $()$.

The estimate will cross-correlate signals with delays on the order of $512/9000 = 57$ msec (actually less than 57 msec given the window). Since (from chapter 2) the reverberant energy is known to be negligible by 50 msec, the window length is sufficient.

3.1.2. Spectral Density of Noise Field in the Automobile Interior

While speech enhancement in the automotive environment has been attempted, [4,5,7,8] most investigations have dealt with a stationary automobile and an idling engine. Therefore, a study of the noise field under various conditions was performed. The responses of the omnidirectional and directional microphones are compared in section 3.1.2.1. As a result of their superior performance, directional microphones were employed to investigate the effect of various car conditions (including speed, fan level and state of windows). The results of this investigation are discussed in section 3.1.2.2.

3.1.2.1 Effect of Microphones

To study the difference between the omnidirectional and directional microphones, one of each was placed in a non-reverberant environment, with as little separation as possible between the two. A source placed equidistant from the microphones and along the axis of maximum gain for the directional microphone was employed to determine the difference between the responses of the two microphones. The microphones were then placed in the car, again with as little distance between them as possible. With the car travelling 50 km/h, the fan off, the windows closed, and the microphones mounted on the car visor pointed towards the driver, the noise field was recorded simultaneously by the two microphones. The difference between the two responses measured previously was used to normalize the recorded noise from the car. The resulting responses are displayed in figure 3.1. In this and all subsequent spectral density plots, the absolute dB values are relative to an aribtrary level, but the relative levels of any two curves on the same plot represent the true difference. Examining the figure reveals that at least two effects are present.
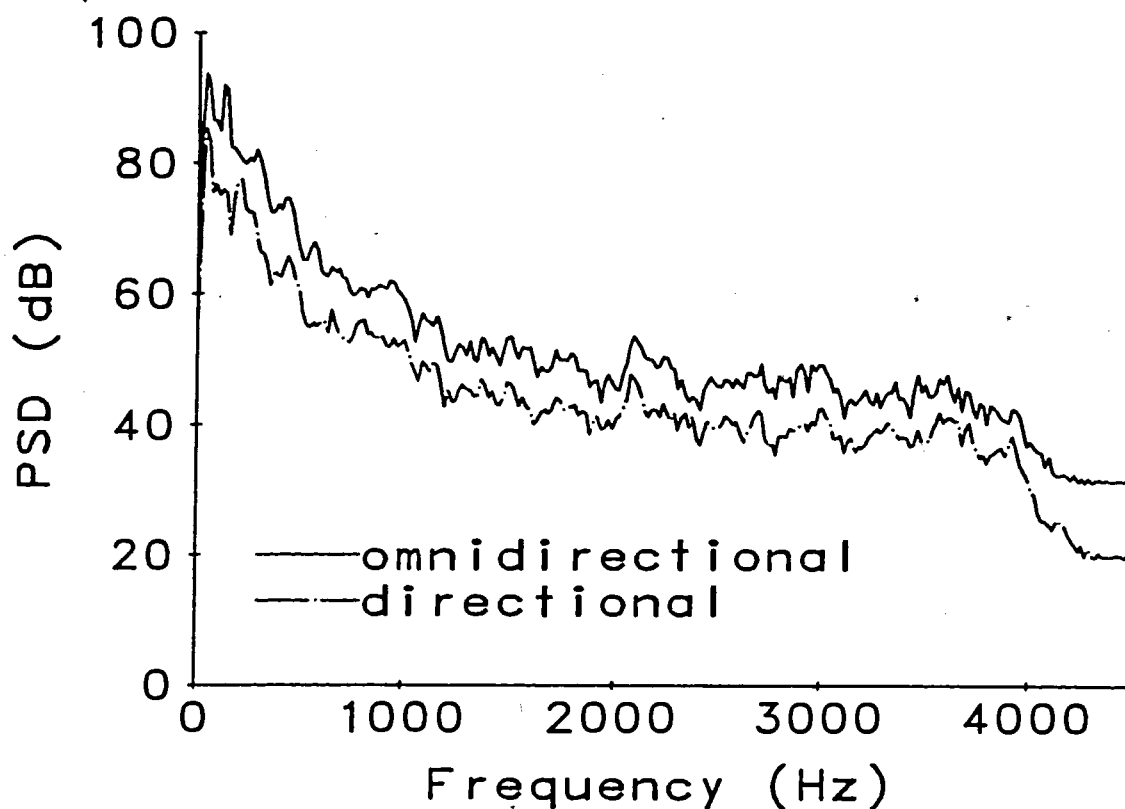
Figure 3.1 Comparison of Omnidirectional and Directional Microphone Responses.

First, at all frequencies the directional microphone receives less noise than the omnidirectional. We can predict the gain expected from the microphone's directionality from an appropriate model for the noise field. In an automobile, where there are many noise sources exciting a reverberant enclosure, the number of virtual (image) sources becomes very large. If we assume that the images are so plentiful that there are uncorrelated plane waves of constant amplitude travelling in every direction, the noise field is called a diffuse field.[26]

For a cardioid mic, a wave arriving from direction θ is received with power $(\frac{1}{2}[1+\cos\theta])^2$. The noise power received by the mic in a diffuse field can be calculated by averaging over all directions, since the uncorrelated plane waves are uniformly distributed. Then the power received by the cardioid microphone is

$$P_{card} = \frac{1}{4\pi} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} (\frac{1}{2}[1+\cos\theta])^2 \sin\theta \, d\theta \, d\phi \tag{3.7}$$

$$=\frac{1}{8}\int_0^\pi (1+\cos\theta)^2\sin\theta\, d\theta$$

$$=\frac{1}{8}\int_0^\pi (1+2\cos\theta+\cos^2\theta)\sin\theta\, d\theta$$

$$=\frac{1}{8}\left[-\cos\theta-\cos^2\theta-\frac{1}{3}\cos^3\theta\right]_0^\pi$$

$$=\frac{1}{3} \tag{3.8}$$

For the omnidirectional mic, the gain is $\equiv 1$ for all directions, so that

$$P_{omni}=1 \tag{3.9}$$

Therefore, we expect the directional mic to have a power level about $\frac{1}{3}$ (4.8 dB) below that of the omnidirectional mic.

The second effect noticeable from figure 3.1 is that the difference between the directional and omnidirectional responses increases at low frequencies. To see this more clearly, consider figure 3.2 where the difference between the two curves of figure 3.1 is plotted. We expect this gain at low frequency because of the low frequency roll-off characteristic of gradient microphones (see figure 2.2) Since the noise seems to be concentrated at lower frequencies, this filtering by the directional microphone is highly desirable.

The peaks and valleys of Figure 3.2 also give evidence for the claim that reverberation is strongly dependent on position of the microphones. The frequency responses of the two microphones being relatively smooth (see figure 2.2), one would expect figure 3.2 to also be smooth since the two microphones were close together. The effects of reverberation, however cause the valleys and peaks in the plot.

These two reasons explain why directional microphones should be used in speech enhancement systems in noisy reverberant environments; therefore, the following spectra were all recorded with cardioid microphones. The omnidirectional microphones were

relegated to a role of confirming theoretical predictions where their flat spectral response and non-directional characteristics proved useful.
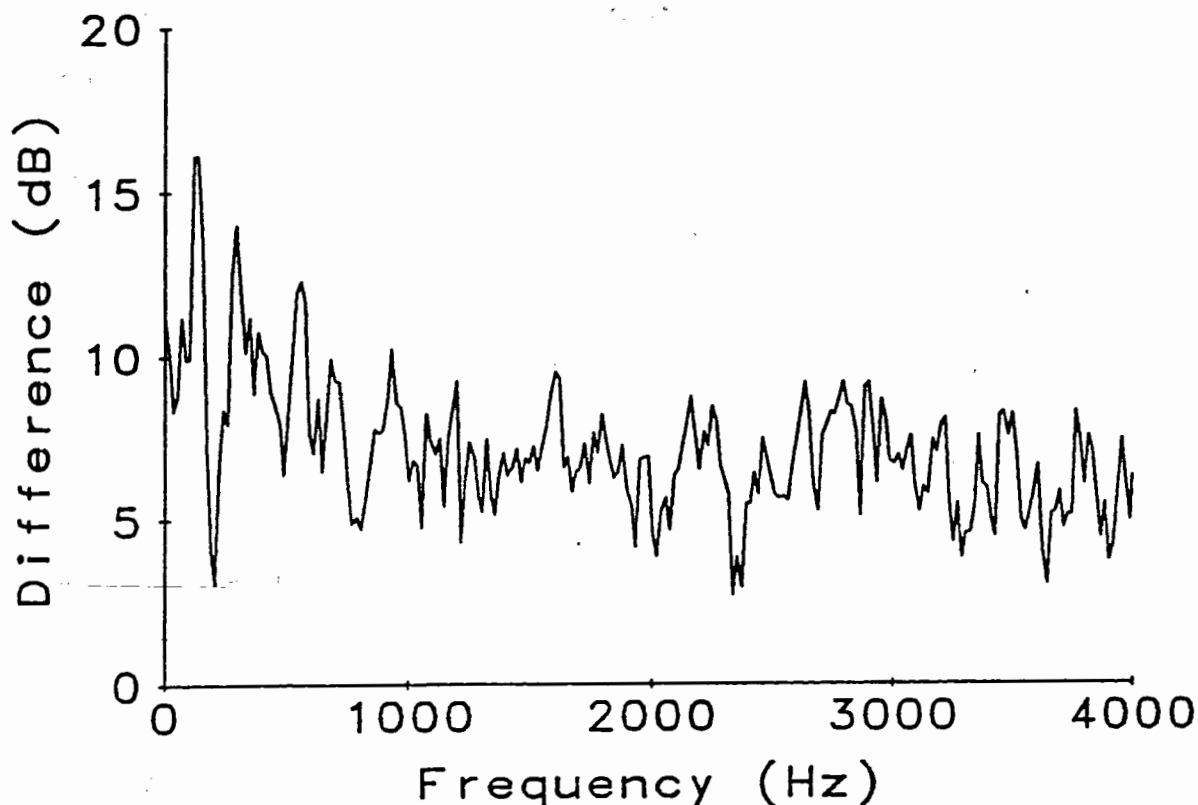
Figure 3.2 Difference Between Omnidirectional and Directional Microphone Responses.

3.1.2.2. Effect of Car Conditions

Recordings were made of the noise in the interior of a mid-sized American car (a 1985 Buick Century) under various conditions: at idle, 50 km/h and 100 km/h, with the fan off, low and high, and with the windows open and closed. All the 50 km/h tests were done on the same road, while all the 100 km/h tests were done on another. (Unfortunately, no single road was available to do both the 50 km/h and 100 km/h tests.) The roads were dry, and relatively smooth. Using the methods of section 3.1.1, spectral estimates were made for each of the recordings. The subsequent spectral plots are all plotted relative to an arbitrary level (which is the same for all the plots.)

To get an idea of the absolute sound levels, consider table 3.1, where the A-weighted sound levels (relative to 0 dB = 0.0002 μbar) are given. By approximating the

human ear's high-pass frequency response, the A-weighting gives a closer indication of perceived loudness than the uniform C-weighting. This high-pass filtering also results in much lower loudness scores than from C-weighting, since the majority of the energy is at low frequencies and is therefore being reduced. For example, at 100 km/h with the fan on high, the C-weighted sound level was 93 dB-C, compared to the A-weighted value of 72 dB-A.

Table 3.1 Absolute Sound Levels in Automobile With Windows Closed

Speed of Car (km/h)	Fan Level	Sound Level (dB-A)
Idle	Off	< 50
Idle	Low	< 50
Idle	Hi	< 50
50	Off	61
50	Low	63
50	Hi	66
100	Off	69
100	Low	69
100	High	72

Figures 3.3 - 3.5 show the effect of car speed on the noise Power Spectral Density (PSD), with the fan at various levels. With the fan off, the effect of speed is quite discernible. At 100 km/h in particular, the car is much noisier, and higher frequency components are evident. The fan produces a noise floor (at all but the lowest frequencies) that increases with fan speed. Even at the high fan speeed, the floor was still below that for the 100 km/h noise level. This would seem to indicate that at high speeds, the road noise (tires, air hitting windshield etc.) dominates, while at lower speeds, the state of the fan becomes important. At all speeds, the engine noise and vibration govern the low

frequency noise.

Figures 3.6 - 3.8 confirm these conclusions. At idle and 50 km/h, the fan doesn't appear to affect the low frequencies but drastically increases the higher frequencies. At 100 km/h, the fan increases the noise in only the very high frequencies.

The effect of opening the windows with speed 50 km/h is seen in figures 3.9 and 3.10. With the fan off, opening the window is similar to turning the fan on low, except that it boosts low as well as high frequency noise. With the fan on high, seen in figure 3.10, opening the window has negligible effect. Unfortunately, at 100 km/h the wind through the open window caused the gradient microphones to create "popping" noise. However, it is expected that the window noise will dominate at high speed (as the fan couldn't be heard in the car).

In all situations, the noise is dominated by low frequency components. From a speech enhancement point of view, this is good news since the power spectral density of speech is bandpass.[27] Some of the low frequency noise can be filtered out with a simple high-pass filter without degrading the speech provided the cutoff of the filter is sufficiently low. The cutoff frequency of the filter can be as high as 600 Hz for some types of speech recognizers.[3]

The speech will be affected most by the higher frequency components of the noise, which are determined by the speed of the car and the fan, and the condition of the windows. Ideally, one would only operate a cellular phone under the quietest conditions of low speed, no fan and the windows closed (but then again, ideally, drivers wouldn't attempt to dial their phones while driving either!).
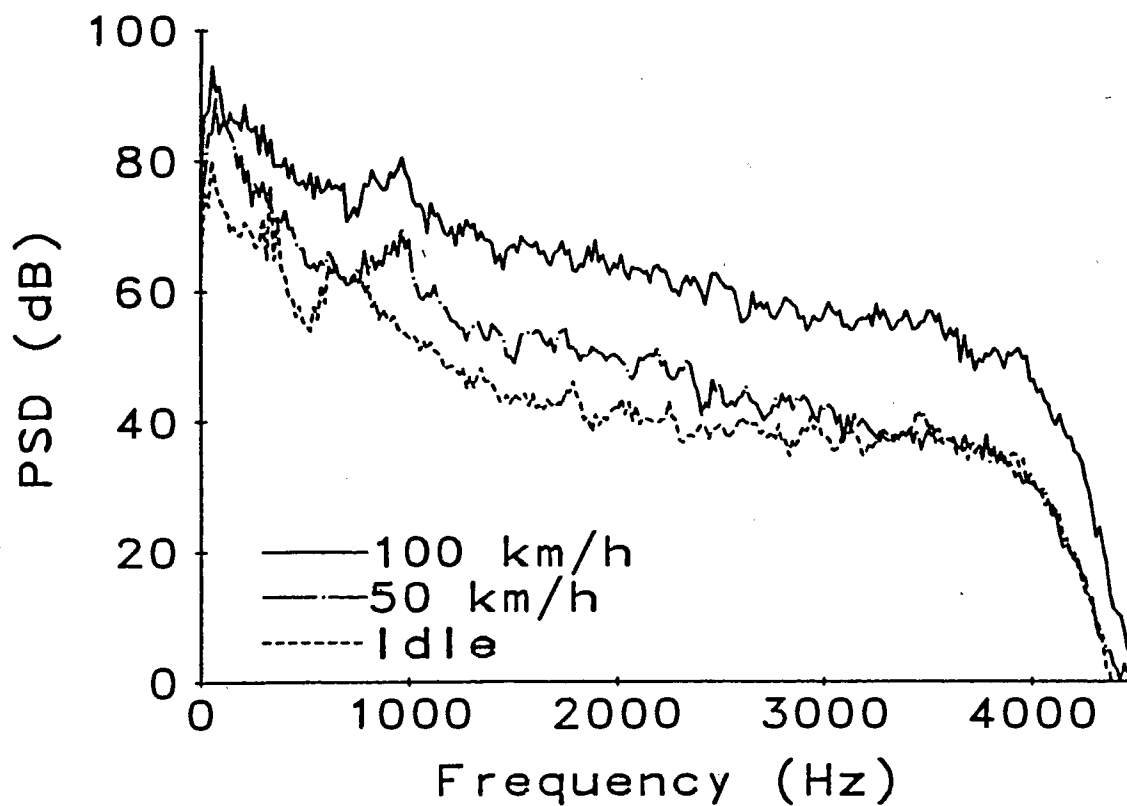
Figure 3.3 Effect of Speed on Noise PSD With Fan Off and Windows Closed.
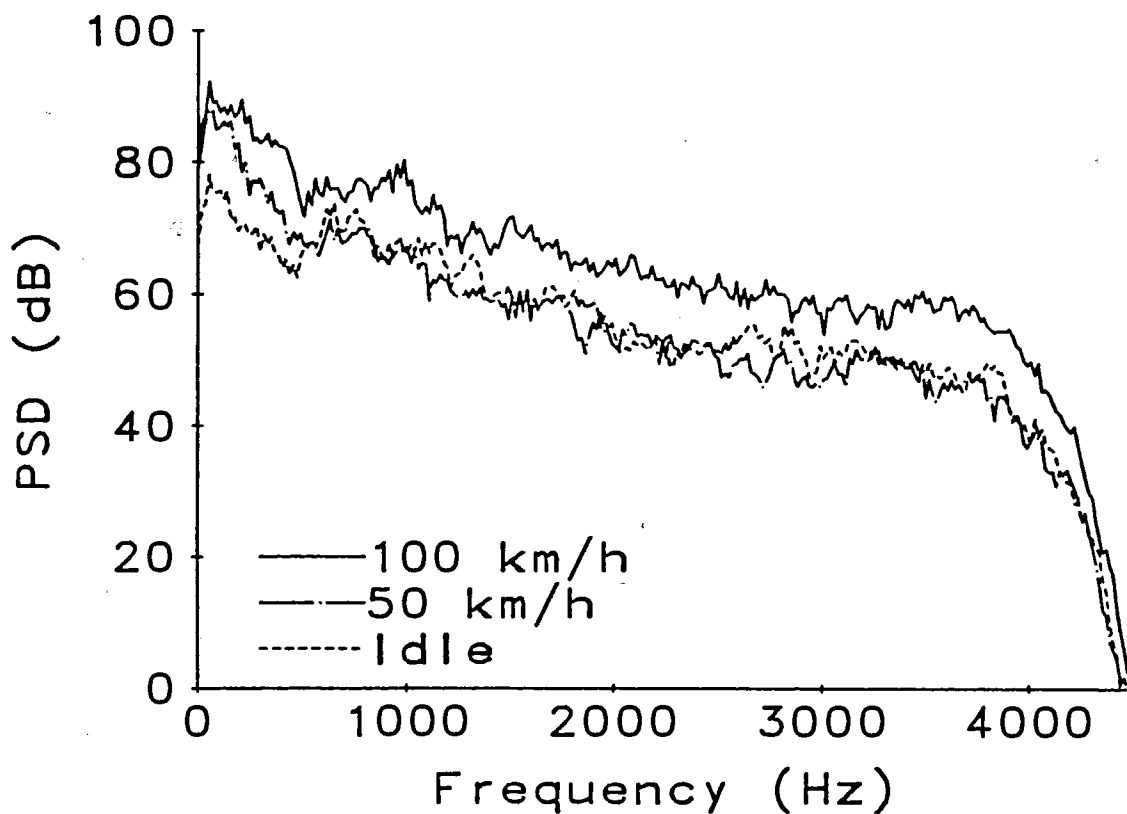
Figure 3.4 Effect of Speed on Noise PSD With Fan On Low and Windows Closed.
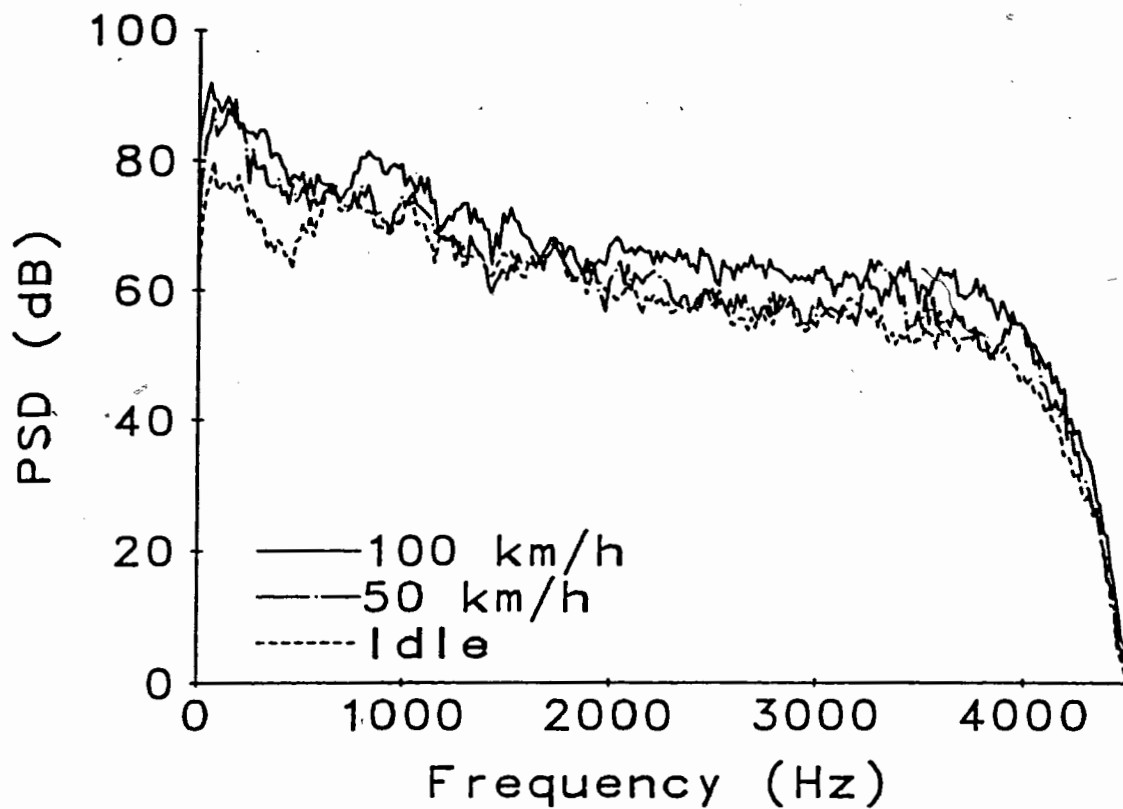
Figure 3.5 Effect of Speed on Noise PSD With Fan On High and Windows Closed.
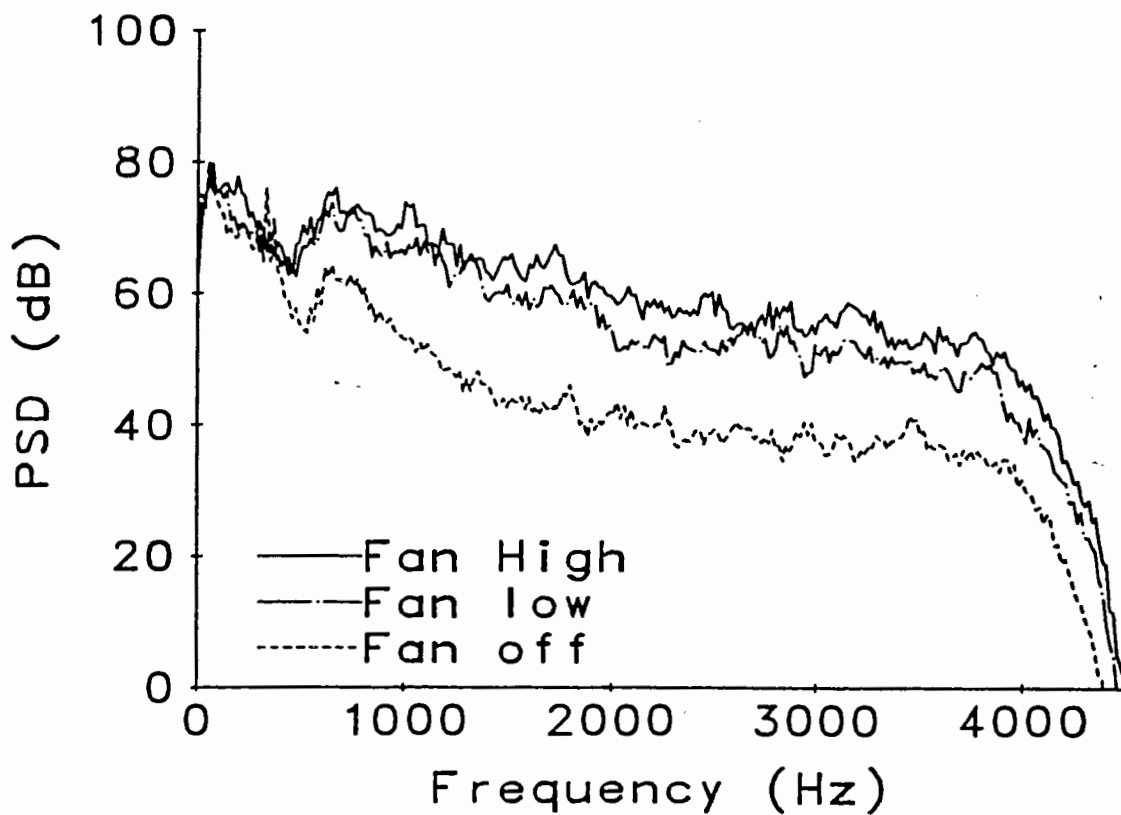
Figure 3.6 Effect of Fan on Noise PSD With Engine Idle and Windows Closed.
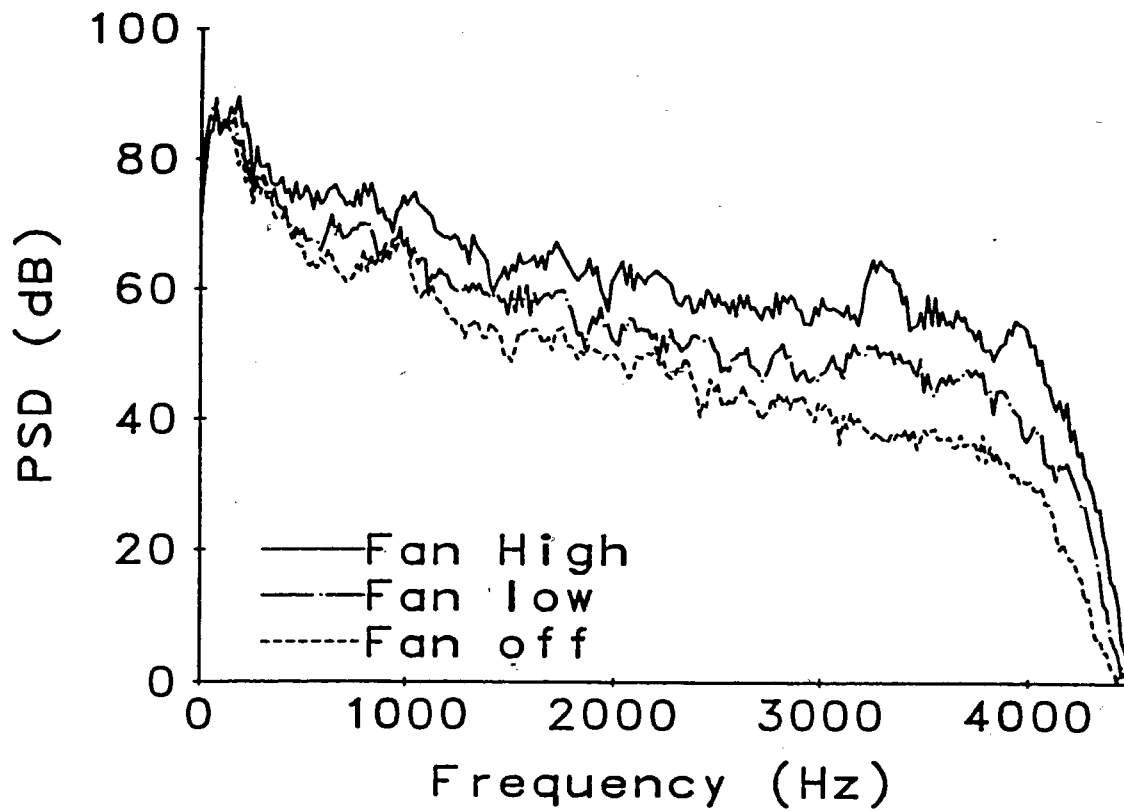
Figure 3.7 Effect of Fan on Noise PSD at 50 km/h with Windows Closed.
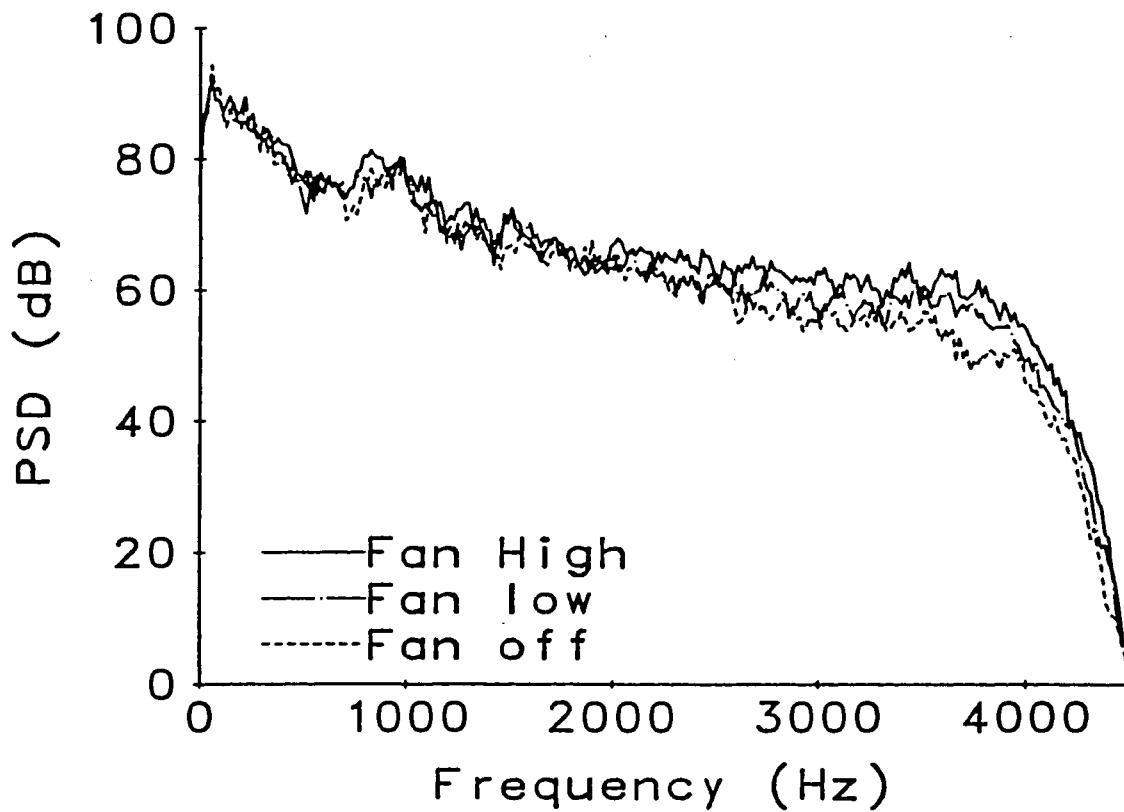
Figure 3.8 Effect of Fan on Noise PSD at 100 km/h with Windows Closed.
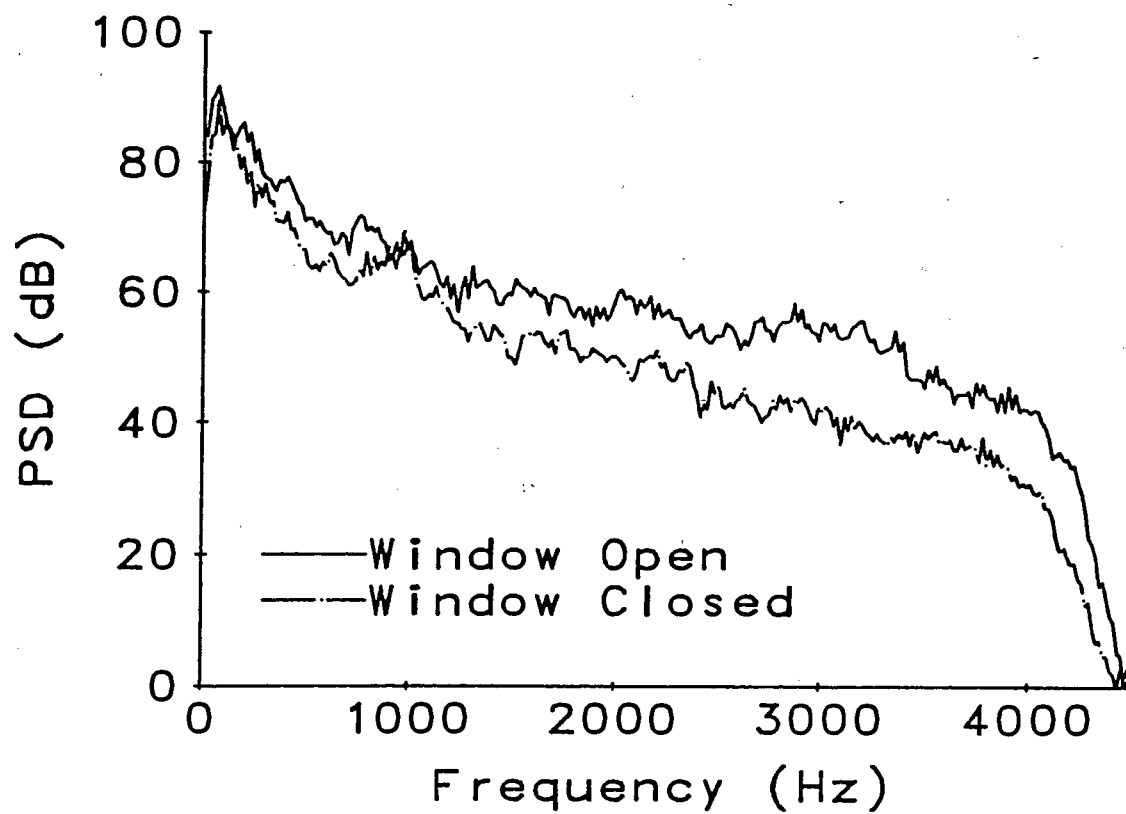
Figure 3.9 Effect of Windows on Noise PSD at 50 km/h with Fan Off.
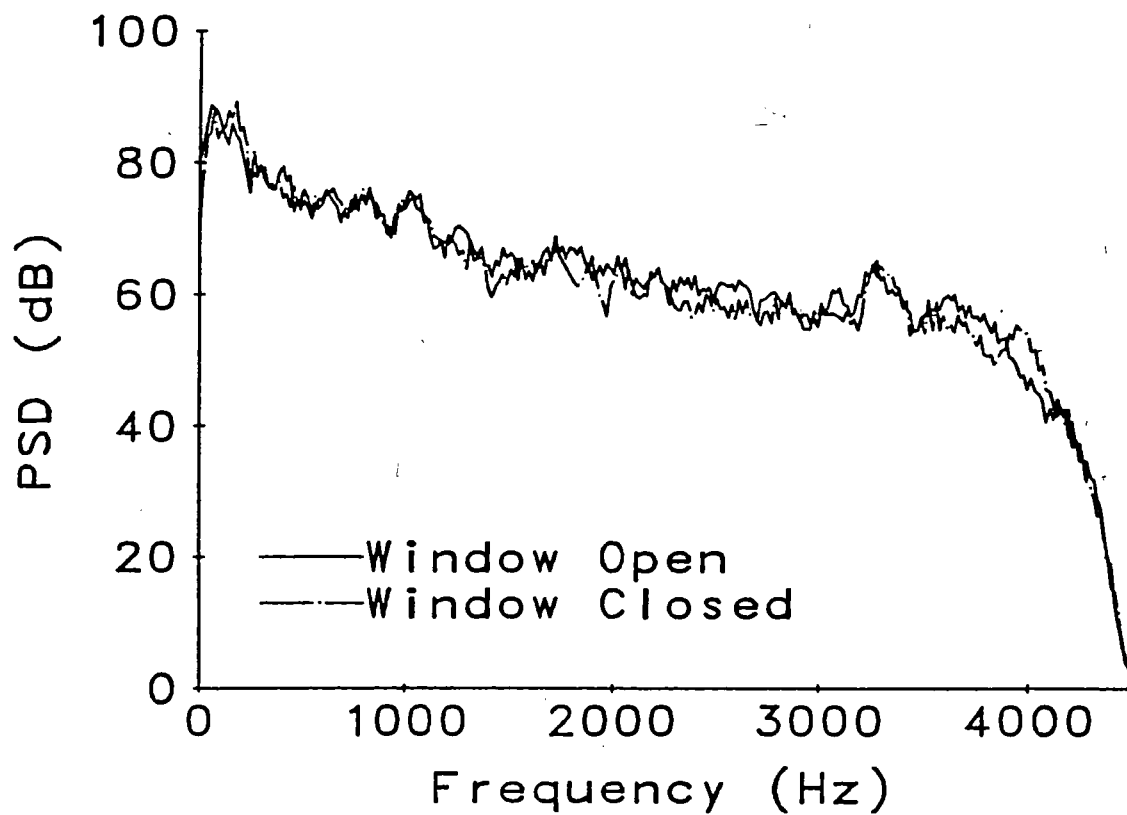
Figure 3.10 Effect of Windows on Noise PSD at 50 km/h with Fan On High.

3.2. COHERENCE

The spectrum of the noise, while important, doesn't give sufficient information to allow predictions to be made concerning the expected performance of various speech enhancement algorithms. The coherence function[26] is a measure of the correlation between two signals, and does enable such predictions to be made. The coherence between two signals $x(n)$ and $y(n)$ is defined by:

$$\gamma_{xy}(e^{j\omega}) = \frac{S_{xy}(e^{j\omega})}{[S_x(e^{j\omega})S_y(e^{j\omega})]^{\frac{1}{2}}}$$

(3.10)

Generally, the cross-spectral density, $S_{xy}(e^{j\omega})$ and therefore the coherence $\gamma_{xy}(e^{j\omega})$ is a complex function. The square of the magnitude of the coherence (MSC for Magnitude Squared Coherence), often mistakenly called the coherence in literature refering to Adaptive Noise Cancellation is defined by

$$|\gamma_{xy}(e^{j\omega})|^2 = \frac{|S_{xy}(e^{j\omega})|^2}{S_x(e^{j\omega})S_y(e^{j\omega})}$$

(3.11)

As we will have use for both the coherence and MSC functions, we will maintain the distinction.

To better understand coherence, consider a linear filter acting on a signal x to give an output y as in figure 3.11.

$$x(t) \longrightarrow \boxed{h(t)} \longrightarrow y(t)$$
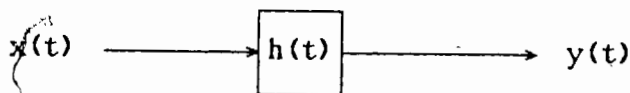
Figure 3.11 Coherence from Linear Filter.

The frequency domain representation of the output, $Y(e^{j\omega})$ is given by

$$Y(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega})$$

(3.12)

The output power spectral density, $S_y(e^{j\omega})$ is

$$S_y(e^{j\omega}) = |H(e^{j\omega})|^2 S_x(e^{j\omega}) \tag{3.13}$$

while the cross-spectral density between x and y, $S_{xy}(e^{j\omega})$ is

$$S_{xy}(e^{j\omega}) = H^*(e^{j\omega}) S_x(e^{j\omega}) \tag{3.14}$$

The resulting MSC is therefore

$$|\gamma_{xy}(e^{j\omega})|^2 = \frac{|H^*(e^{j\omega}) S_x(e^{j\omega})|^2}{S_x(e^{j\omega})|H(e^{j\omega})|^2 S_x(e^{j\omega})} \tag{3.15}$$

$$\equiv 1 \text{ since } S_x(e^{j\omega}) \text{ is real.} \tag{3.16}$$

The magnitude of the coherence is a measure of the proportion of $S_x(e^{j\omega})$ related to $S_y(e^{j\omega})$ by a linear filter.

The coherence function is a very valuable tool in giving theoretical limits about how well signal enhancement schemes utilizing more than one microphone will perform. The performances of both ANC and our delay-equalized array are relatively simple functions of the noise-field coherence.

3.2.1. Coherence of Diffuse Noise Field

When many noise sources are exciting a highly reverberant environment, as is the case in an automobile, the noise field can often be modelled as a diffuse field consisting of a series of uncorrelated plane waves travelling in all directions. We will now derive the coherence between the signals received by two omnidirectional microphones separated by a spacing of d, in such a field. In section 3.2.1.2, we will consider the effect of directional microphones on the coherence.

3.2.1.1. Omnidirectional Microphones

Figure 3.12 shows the geometry to be considered. The separation d, is assumed without loss of generality to lie along the z axis. For a plane wave from angle θ, there is a delay of $\frac{d}{c}\cos\theta$ between the arrival of the wave at x and at y, where c is the speed of sound (≈ 341 m/s at room temperature).
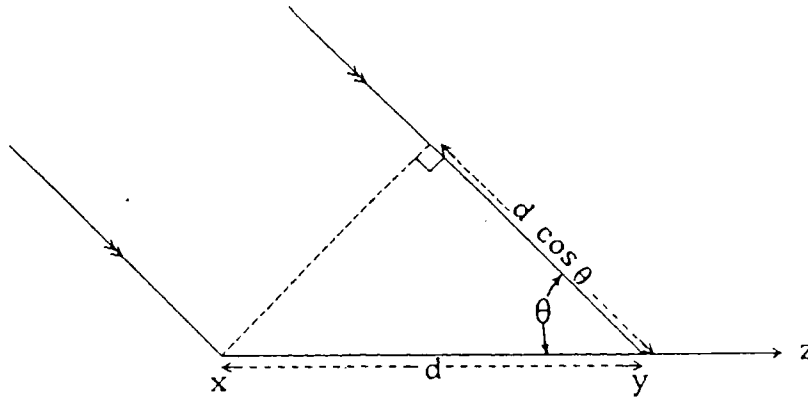
Figure 3.12 Geometry to Derive Coherence of Diffuse Field.

For clarity, the derivation will be made in the continuous domain. First,

$$y(t) = x(t - \frac{d}{c}\cos\theta) \tag{3.17}$$

Then by definition

$$S_y(\omega) = F\left\{ R_y(\tau) \right\} \tag{3.18}$$

$$= F\left\{ E\left\{ y(t)y(t+\tau) \right\} \right\}$$

$$= F\left\{ E\left\{ x(t - \frac{d}{c}\cos\theta)x(t - \frac{d}{c}\cos\theta+\tau) \right\} \right\}$$

$$= F\left\{ R_x(\tau) \right\}$$

$$= S_x(\omega) \tag{3.19}$$

and

$$S_{xy}(\omega) = F\left\{ E\left\{ x(t)y(t+\tau) \right\} \right\} \tag{3.20}$$

$$= E\left\{ F\left\{ x(t)y(t+\tau) \right\} \right\}$$

where F denotes the Fourier Transform and E the Expectation operator, both of which are linear and commutative so that their order can be exchanged. Then

$$S_{xy}(\omega) = E\left\{ F\left\{ x(t)x(t+\tau - \frac{d}{c}\cos\theta) \right\} \right\} \tag{3.21}$$

$$= E\left\{ X(\omega) * X(\omega) e^{j\omega(\tau - \frac{d}{c}\cos\theta)} \right\} \tag{3.22}$$

$$= E\left\{ X(\omega) * X(\omega) e^{j\omega\tau} \right\} E\left\{ e^{-jkd\cos\theta} \right\} \tag{3.23}$$

where the asterisk (*) denotes the convolution operation, and $k = \dfrac{\omega}{c}$ is the magnitude of the wave vector. The first term is merely $S_x(\omega)$, so that the coherence is given by

$$\gamma_{xy}(\omega) = \frac{S_x(\omega) E\left\{ e^{-jkd\cos\theta} \right\}}{[S_x(\omega)S_x(\omega)]^{\frac{1}{2}}} \tag{3.24}$$

$$= E\left\{ e^{-jkd\cos\theta} \right\} \tag{3.25}$$

Now, to evaluate the expectation, we average over the spherical co-ordinates θ and ϕ:

$$E\left\{e^{-jkd\cos\theta}\right\} = \frac{1}{4\pi} \int\limits_{\phi=0}^{2\pi} \int\limits_{\theta=0}^{\pi} e^{-jkd\cos\theta}\sin\theta \, d\theta \, d\phi \tag{3.26}$$

$$= \frac{1}{2}\int\limits_{0}^{\pi} e^{-jkd\cos\theta}\sin\theta \, d\theta \tag{3.27}$$

Using the substitution

$$x = kd\cos\theta \quad \text{so that} \quad dx = -kd\sin\theta \, d\theta \tag{3.28}$$

$$\gamma_{xy}(\omega) = -\frac{1}{2kd} \int\limits_{x=kd}^{-kd} e^{-jx} dx \tag{3.29}$$

$$= \text{sa}(\omega d/c) \tag{3.30}$$

$$\text{sa}(x) = \frac{\sin(x)}{x} \tag{3.31}$$

Thus, the coherence between two signals received by omnidirectional microphones with separation d in a diffuse field is real and is given by:

$$\gamma_{xy}(\omega) = \frac{\sin(\omega d/c)}{\omega d/c} \tag{3.32}$$

The coherence of a diffuse field falls off with increased spacing between the points, and also with increased frequency. This makes intuitive sense, since in both cases the distance between the points, when measured in wavelengths, is larger.

3.2.1.2. Effect of Directionality in Microphones

When directional microphones are used, the coherence between the two received signals becomes a function of the microphone patterns. If the two microphones have directionality patterns $m_x(\phi,\theta)$ and $m_y(\phi,\theta)$, then

$$S_{xy}(\omega) = F\{E\{m_x(\phi,\theta)x(t)m_y(\phi,\theta)y(t+\tau)\}\} \tag{3.33}$$

and

$$S_x(\omega) = F\{E\{m_x^2(\phi,\theta)x(t)x(t+\tau)\}\}$$ (3.34)

(with a similar expression for $S_y(\omega)$). Equation (3.24) modified for directional microphones becomes

$$\gamma_{xy}(\omega) = \frac{E\left\{m_x(\phi,\theta)m_y(\phi,\theta)\,e^{-jkd\cos\theta}\right\}}{\left[E\{m_x^2(\phi,\theta)\}E\{m_y^2(\phi,\theta)\}\right]^{1/2}}$$ (3.35)

In appendix 1, this is evaluated for two cardioid microphones with axes along unit vectors $(x_1 y_1 z_1)$ and $(x_2 y_2 z_2)$, and separated by a distance d which is assumed, without loss of generality, to be along the z axis. The result is

$$\gamma_{xy}(\omega) = \frac{3}{4}\left[\text{sa}(kd) + \left[x_1 x_2 + y_1 y_2\right]\left[\frac{\sin(kd)}{(kd)^3} - \frac{\cos(kd)}{(kd)^2}\right] + z_1 z_2\left[\text{sa}(kd) + \frac{2\cos(kd)}{(kd)^2} - \frac{2\sin(kd)}{(kd)^3}\right]\right.$$

$$\left. + j\left[z_1 + z_2\right]\left[\frac{\cos(kd)}{kd} - \frac{\sin(kd)}{(kd)^2}\right]\right]$$ (3.36)

The transformation for mapping a general $\vec{d}$ to the z axis (so (3.36) can be used) is given in Appendix 2.

Figure 3.13 compares the theoretical coherence for various combinations of microphones with separation 15 cm in a diffuse noise field. The arrows in the legend refer to the direction that the directional microphones are pointing. For example, the two arrows > > indicate that the two directonal microphones are both pointed in the same direction, along the line separating them, while ^ v indicates that the two microphones are pointed in opposite directions, perpendicular to their line of separation.

We see that for cardioid microphones in a diffuse field, the relative axes of the microphones affects the coherence significantly, especially at low frequencies. For higher frequencies, we expect low coherence regardless of the microphone configuration. When the microphones are pointed in different directions, low coherence is expected at all frequencies. In section 3.2.2 the actual MSC of noise recorded in the automobile interior is discussed.
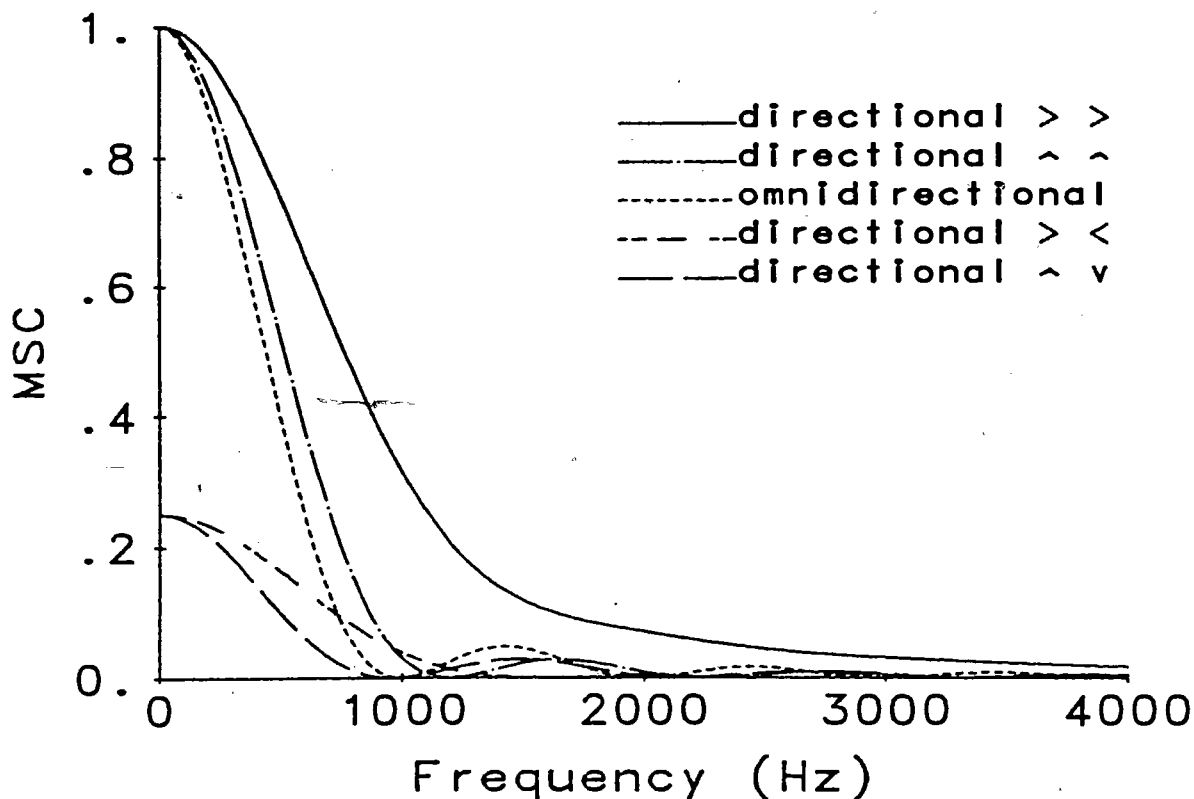
Figure 3.13 Theoretical MSC for Two Microphones With Separation $d=$ 15 cm. The arrows in the legend refer to the axis of the directional microphones relative to their line of separation.

3.2.2. Coherence in Automobile Interior

The coherence of noise fields in an automobile was measured with the microphones in the various configurations of figure 3.13 to investigate how accurately a diffuse field models the noise field in the car. Equation (3.10) and the techniques of section 3.1.1 were used to calculate the estimates. Figures 3.14 - 3.18 show the experimental MSC along with that predicted for a diffuse noise field by equation (3.36).

The experimental MSC is higher than that predicted in all cases, especially for frequencies above 1kHz. While variance in the MSC estimate might contribute to this, a more likely cause is the existence of directionality in the noise field.

The disparity between the diffuse and actual MSC is most pronounced in figure 3.18, where the two microphones were placed along the car visor, 15 cm apart, one facing the window and one facing the back of the car. Since the window is a good acoustic

reflector (reflection coefficient, β> .96 at 1 kHz[23]), the microphone facing the window sees a delayed version of any wave travelling from the back of the car. The situation is similar to figure 3.16, where the two microphones are both facing backwards (since delay is linear and therefore doesn't affect coherence).

The peaks and valleys in the experimental measurements are due to variance in the spectral estimates behind the measurements, and also due to the fact that the measurements were only made once. If the coherence were measured several times with the same configurations, it is expected that the peaks and valleys would average out.

While the diffuse model doesn't match perfectly, the fact that it does predict the low-frequency fall-off in MSC reasonably well makes it useful. In general, extreme values of MSC are preferred for speech enhancement algorithms. It will be shown in chapter 5 that lower values of coherence can be exploited by a non-adaptive array, while larger values are necessary for Adaptive Noise Cancellation. In chapter 4, we will consider ANC, and show that values of MSC in excess of 0.7 are required for meaningful performance. In view of our experimental MSC results, we do not recommend ANC for speech enhancement.
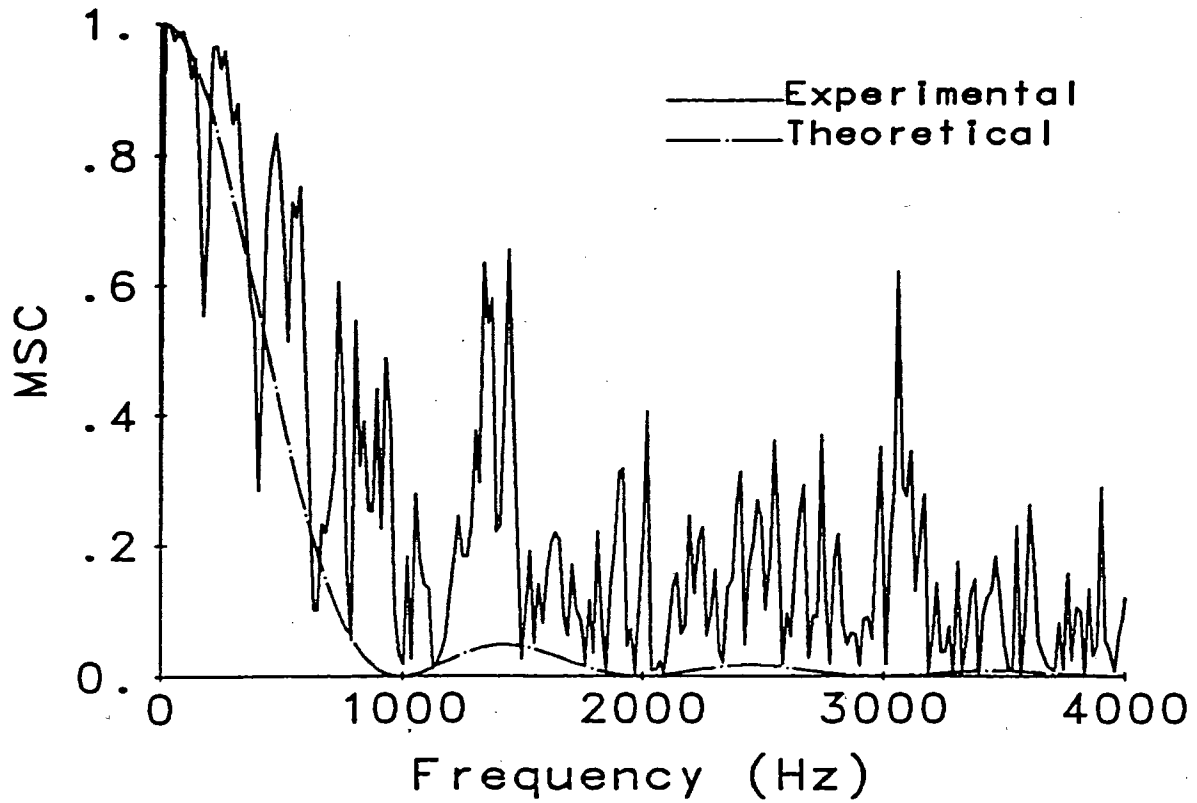
Figure 3.14 MSC for Two Omnidirectional Mics with Separation 15 cm
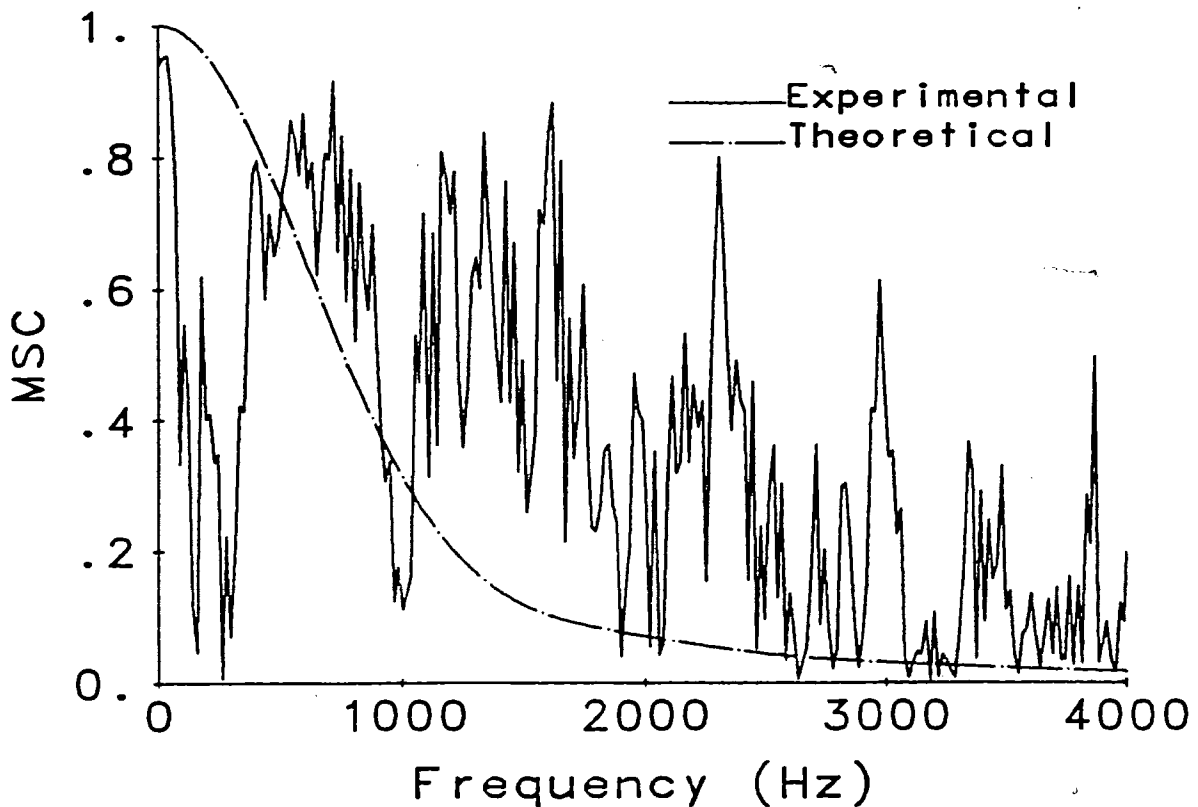
Figure 3.15 MSC for Two Directional Mics Separated by 15 cm Both Pointing in Same Direction Along Axis of Separation.
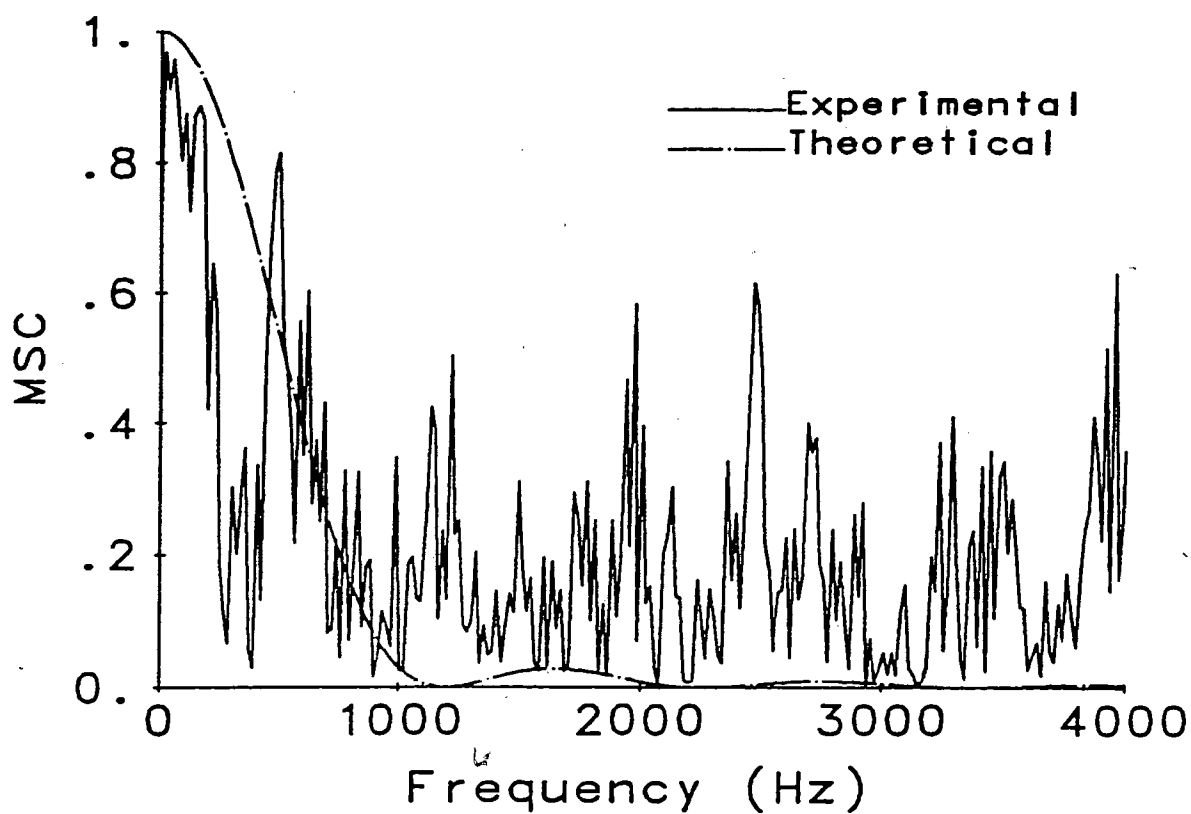
Figure 3.16 MSC for Two Directional Mics Separated by 15 cm Both Pointing in Same Direction Perpendicular to Axis of Separation.
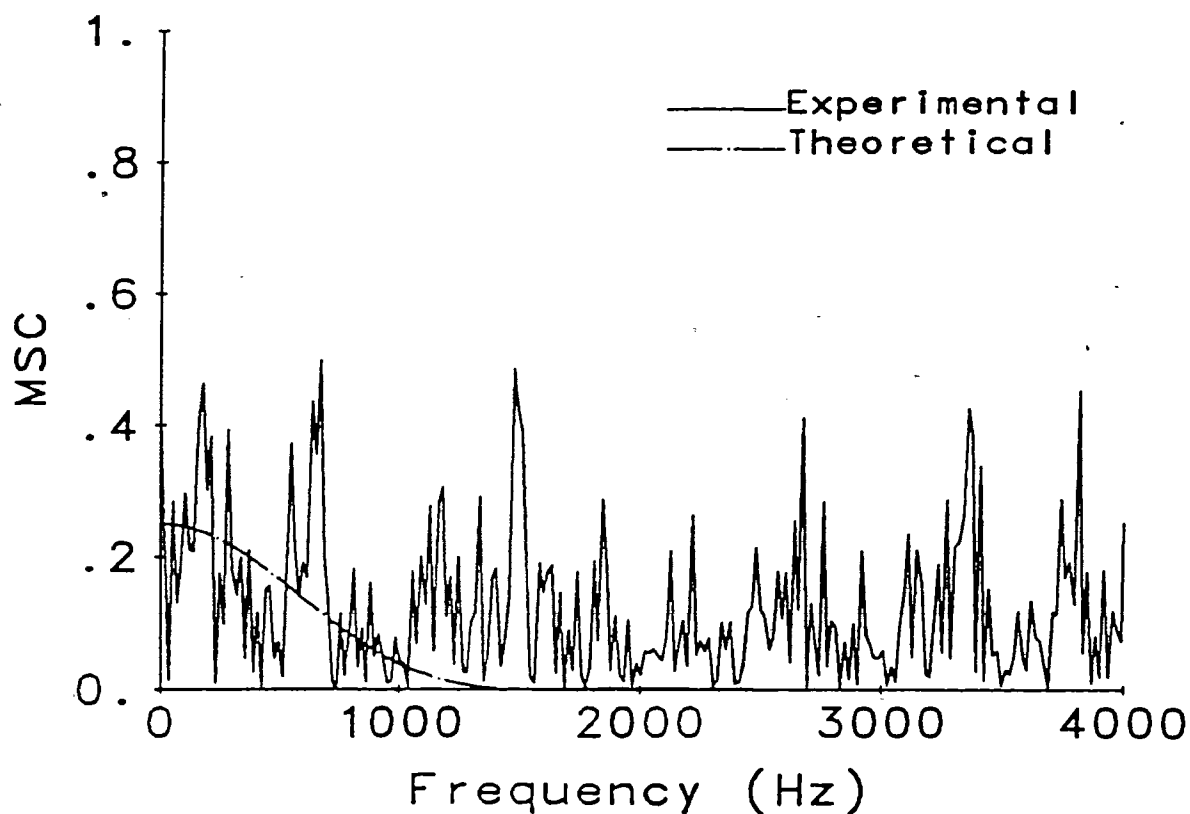
Figure 3.17 MSC for Two Directional Mics Separated by 15 cm Pointing in Opposite Directions Along Axis of Separation.
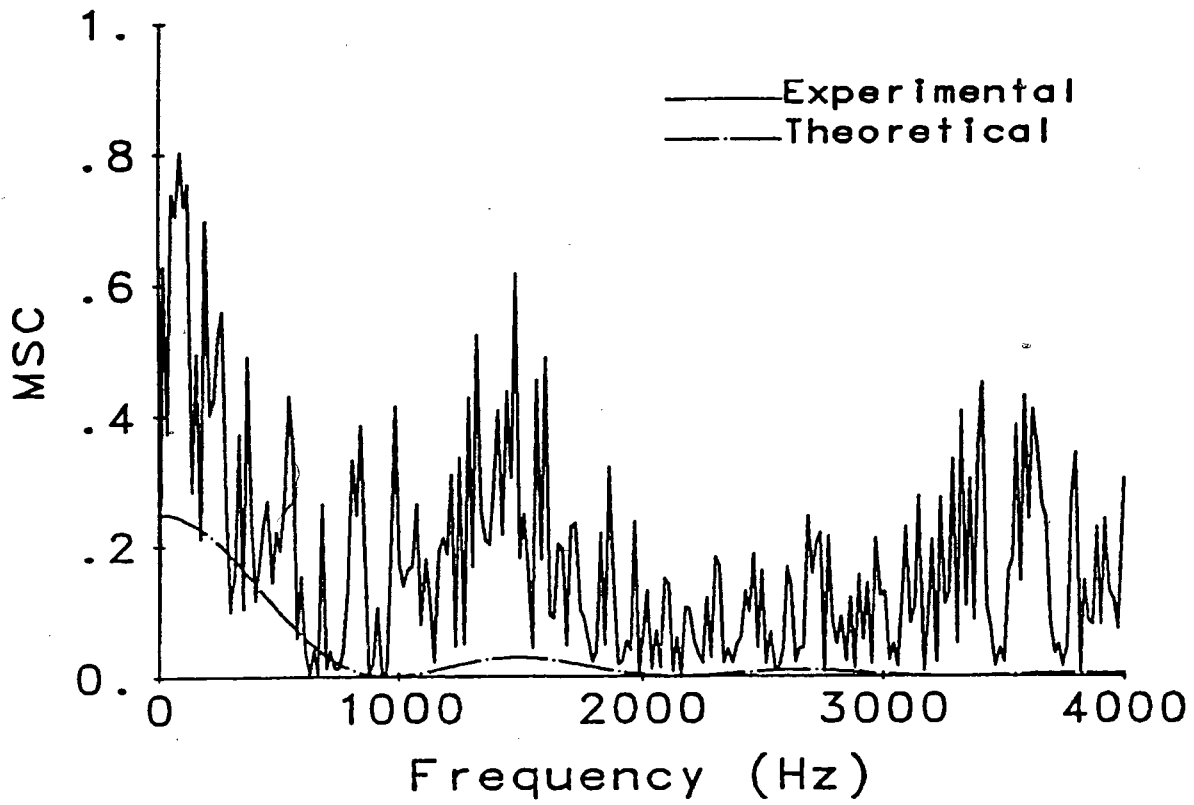
Figure 3.18 MSC for Two Directional Mics Separated by 15 cm, Pointing in Opposite Directions Perpendicular to Axis of Separation.

4. ADAPTIVE NOISE CANCELLATION: PROMISE AND PRACTICE

When the statistics of the speech and noise are both known and stationary, optimal linear time invariant filters can be used to enhance noisy speech. In the automobile, where the noise field is non-stationary, a logical alternative is adaptive filtering.

Adaptive noise cancellation[7,28] is a scheme that has been investigated quite thoroughly for the automotive environment[4,5,8] as well as for aircraft[29,30,31] and underwater for divers.[32] The basic idea is to use multiple inputs: a primary signal, assumed to consist of speech plus noise, and any number of reference signals, assumed to consist only of noise that is correlated to that in the primary signal (i.e. no speech). The reference signals are adaptively filtered and then subtracted from the primary signal, resulting in a cleaner speech signal.

Section 4.1 describes the theoretical model assumed for effective adaptive noise cancellation. The optimal unrestrained (Wiener) filter is then derived. In section 4.2, some violations of the theoretical model are discussed, and shown to render adaptive noise cancellation ineffective in the automobile environment. To test the predictions of section 4.2, an adaptive noise canceller, based on Widrow's Least-Mean-Square (LMS) algorithm,[33] was implemented and investigated. The LMS algorithm is discussed in section 4.3. The experimental results of applying LMS to ANC in an automobile are given in section 4.4, and compared to the results from other researchers.

4.1. THEORY OF ADAPTIVE NOISE CANCELLATION

This section describes the theory behind ANC. The basic model is described in section 4.1.1, while the optimal (Wiener) filter and its performance are covered in section 4.1.2.

4.1.1. Basic Model of ANC

The basic model for adaptive noise cancellation is shown in figure 4.1. The reference signal, $x(t)$ is assumed to consist of noise with no speech. The primary input, $p(t)$ is composed of the speech signal $s(t)$ plus $d(t)$, the result of a transfer function H acting on $x(t)$. For adaptive noise cancellation, H is assumed to be at least approximately linear. We assume s is uncorrelated with x and therefore also with d and y. The goal of ANC is to approximate H by $\hat{H}$ so that $y(t)$ becomes a good approximation of $d(t)$. Then when $y(t)$ is subtracted from $p(t)$, the output $z(t)$ is a good estimated of $s(t)$.
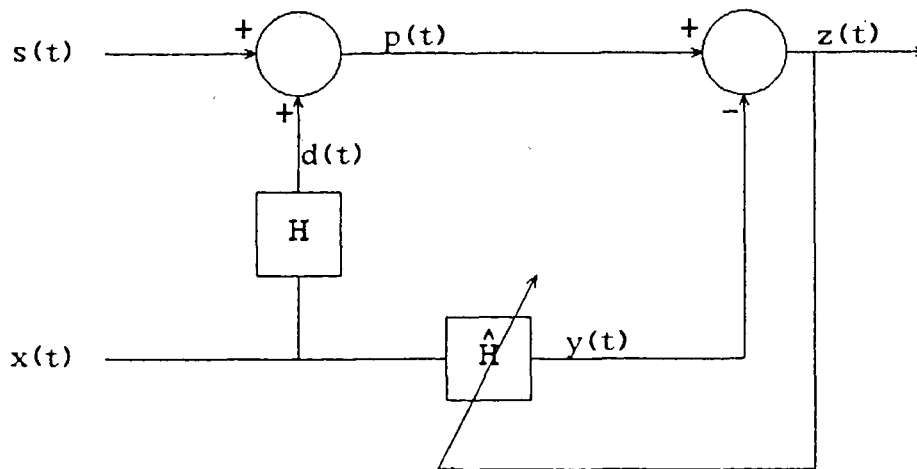
Figure 4.1 Model for Adaptive Noise Cancellation

The estimate $\hat{H}$ is usually found by attempting to minimize the total output power. Consider the output noise power:

$$E\{(z-s)^2\} = E\{z^2\} - 2E\{sz\} + E\{s^2\}$$

$$= E\{z^2\} - 2E\{s(s+d-y)\} + E\{s^2\}$$

$$= E\{z^2\} - E\{s^2\} \tag{4.1}$$

since s is uncorrelated with d and y.

Now, s and therefore $E\{s^2\}$ remain unaffected by the adaptive filter $\hat{H}$. Therefore minimizing the total output power, $E\{z^2\}$ is equivalent to minimizing the output noise power $E\{(z-s)^2\}$. The output noise power can also be expressed as a function of the error in the noise estimate $y-d$.

$$E\{(z-s)^2\} = E\{(s+d-y-s)^2\}$$

$$= E\{(d-y)^2\} \tag{4.2}$$

Thus, minimization of total output power is equivalent to minimizing (in the mean square sense) the error in both $z-s$ and $d-y$. Intuitively this makes sense, since if $\hat{H}=H$, then $d=y$ so that $z=s$ and we have perfect cancellation. (Note if H is time invariant and measureable, we could set $\hat{H}=H$ and be finished.)

Consider the other extreme, where x and d are uncorrelated. Then x is also uncorrelated with y, and the output noise power becomes

$$E\{(z-s)^2\} = E\{(d-y)^2\}$$

$$= E\{d^2\}+E\{y^2\} \tag{4.3}$$

Minimization implies that $E\{y^2\}=0$, that is $\hat{H}\equiv0$. The output is then $z=s+d$, so that no cancellation occurs.

4.1.2. Wiener Filtering: The Ultimate Performance

If we assume that H is time invariant, and that s and x are stationary, we can derive the optimal (Wiener) filter. [28] Figure 4.2 illustrates this filter.

Since $s(t)$ is uncorrelated with the noise $x(t)$ and $d(t)$, we may neglect it, and consider only the noise cancellation. We seek to minimize the function

$$\xi = E\{e^2\} \tag{4.4}$$

$$= E\{(d-y)^2\} \tag{4.5}$$

But y is the output of the filter $\hat{H}$ with input x:
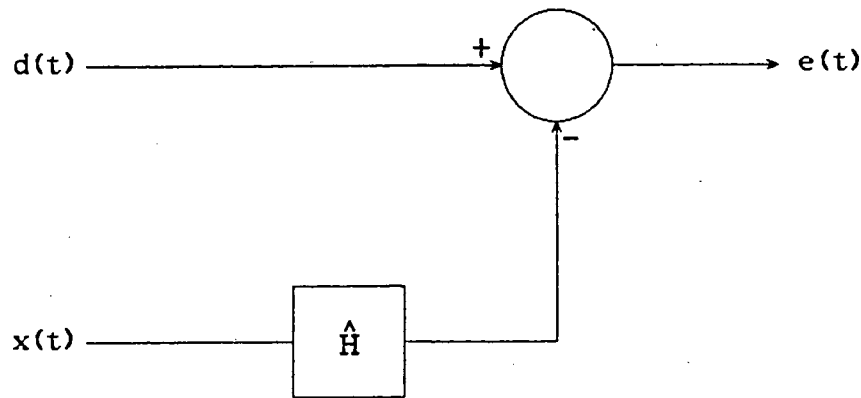
Figure 4.2 Optimal (Wiener) Filter

$$y(n) = \sum_m \hat{h}(m)x(n-m) \tag{4.6}$$

Therefore

$$\xi = E\{(d - \sum_m \hat{h}(m)x(n-m))^2\}$$

$$= E\{d^2\} + \sum_m \sum_l \hat{h}(m)\hat{h}(l)E\{x(n-m)x(n-l)\} - 2\sum_m \hat{h}(m)E\{d(n)x(n-m)\} \tag{4.7}$$

But the expectations can be recognized as correlation functions:

$$= R_d(0) + \sum_m \sum_l \hat{h}(m)\hat{h}(l)R_x(m-l) - 2\sum_m \hat{h}(m)R_{xd}(m) \tag{4.8}$$

We seek a mimimum by setting to zero the partial derivatives of ξ with respect to the filter weights:

$$\frac{\partial \xi}{\partial \hat{h}(m)} = 2\sum_l \hat{h}(l)R_x(m-l) - 2R_{xd}(m) \equiv 0 \tag{4.9}$$

Solving this,

$$\sum_l \hat{h}(l)R_x(m-l) = R_{xd}(m) \tag{4.10}$$

If the summation index l runs from $-\infty$ to $+\infty$, the sum becomes a discrete convolution

and

$$\hat{h}(m) * R_x(m) = R_{xd}(m) \tag{4.11}$$

Taking Z transforms:

$$\hat{H}(z)S_x(z) = S_{xd}(z) \tag{4.12}$$

and therefore, the optimal filter is given by

$$\hat{H}_{opt}(z) = \frac{S_{xd}(z)}{S_x(z)} \tag{4.13}$$

This is the optimal unconstrained Wiener filter. The filter is assumed to be as long and non-causal as necessary (doubly infinite). Under these ideal conditions, the output noise power spectral density is given by

$$S_e(e^{j\omega}) = S_d(e^{j\omega}) - |\hat{H}_{opt}(e^{j\omega})|^2 S_x(e^{j\omega}) \tag{4.14}$$

$$= S_d(e^{j\omega}) - \left| \frac{S_{xd}(e^{j\omega})}{S_x(e^{j\omega})} \right|^2 S_x(e^{j\omega})$$

The ratio of the original noise power in the primary signal, $S_d(e^{j\omega})$ to the output noise power $S_e(e^{j\omega})$ gives the filter's noise cancellation power:

$$\text{Cancellation}(e^{j\omega}) = \frac{S_d(e^{j\omega})}{S_d(e^{j\omega}) - \left| \dfrac{S_{xd}(e^{j\omega})}{S_x(e^{j\omega})} \right|^2 S_x(e^{j\omega})}$$

$$= \frac{1}{\left[1 - \dfrac{|S_{xd}(e^{j\omega})|^2}{S_x(e^{j\omega})S_d(e^{j\omega})} \right]} \tag{4.15}$$

But from equation (3.11),

$$|\gamma_{xd}(e^{j\omega})|^2 = \frac{|S_{xd}(e^{j\omega})|^2}{S_x(e^{j\omega}) S_d(e^{j\omega})} \tag{4.16}$$

so the cancellation can be expressed as a function of MSC:

$$\text{Cancellation } (\gamma^2_{xd}(e^{j\omega})) = \frac{1}{1 - |\gamma_{xd}(e^{j\omega})|^2}$$ (4.17)

Expressed in dB:

$$\text{Cancellation}(\gamma^2_{xd}(\omega)) = -10\log(1 - |\gamma_{xd}\omega|^2) \quad \text{dB}$$ (4.18)

Figure 4.3 shows a plot of cancellation power as a function of MSC. As can be seen, significant coherence is required for even modest cancellation. (Values of $|\gamma_{xy}(\omega)|^2$ near 0.7 are required for even 5 dB of attenuation.)

Equation (4.18) and figure 4.3 give the maximum cancellation possible by a linear filter with no constraints on filter length or causality. In general, we expect practical filters to have poorer performance.
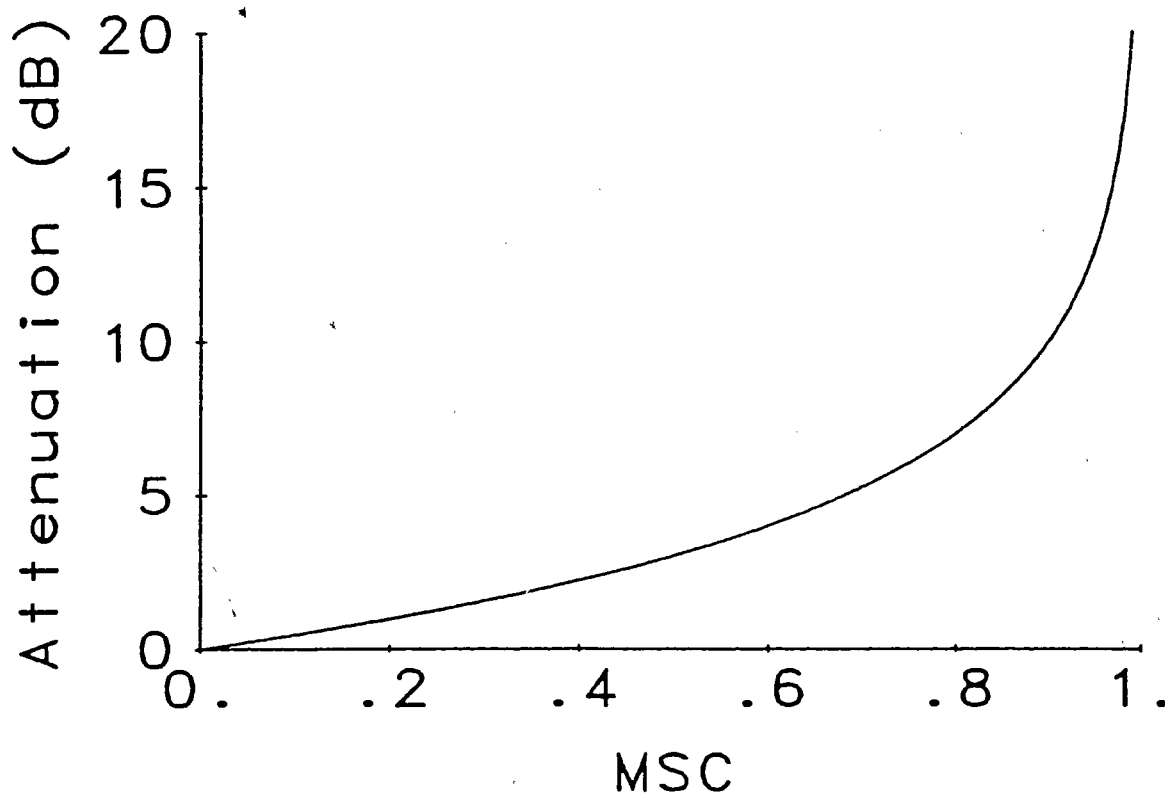
Figure 4.3 Optimal Filter Cancellation as Function of MSC

4.2. VIOLATIONS OF THE MODEL

The basic two assumptions of the model of figure 4.1 are:

1. The noise in the reference and primary microphone are highly correlated. Equivalently, there must be significant coherence between the noise in the two signals.

2. There is speech signal in only the primary microphone. (Any speech in the reference microphone will tend to cancel speech at the output).

It is seen that there is a tradeoff between these two assumptions. To achieve speech isolation, we will want to separate the primary and reference microphone, and point them in different directions but from chapter 3, we realize that these two actions will lower the coherence.

The model of figure 4.1 assumes that the noise field is excited by only one noise source, $x(t)$. We now consider the problem of how multiple noise sources reduce the noise field coherence and limit ANC's effectiveness.

4.2.1. Coherence of Multiple Sources

From equation (4.18), we recall that the theoretical maximum attenuation obtainable is a function of MSC. For a simple example of the problem of cancelling noise fields composed of multiple sources, consider figure 4.4, where we have only two sources. Quite obviously, the adaptive filter $\hat{H}$ cannot simultaneously be both $H_1(z)$ and $H_2(z)$, so cancellation cannot be perfect.

Assume that $n_1(n)$ and $n_2(n)$ are uncorrelated white noise, with equal power spectral densities $S_{n_1}(z) = S_{n_2}(z) = S_n(z)$ and that $H_1(z)$ and $H_2(z)$ are pure delays, so that

$$H_1(z) = z^{-\tau_1} \text{ and } H_2(z) = z^{-\tau_2} \tag{4.19}$$

Then, the cross-correlation of x and d becomes

$$R_{xd}(m) = E\{[r_1(n) + r_2(n)][n_1(n+m) + n_2(n+m)]\}$$

$$= R_{r_1 n_1}(m) + R_{r_2 n_2}(m) \tag{4.20}$$
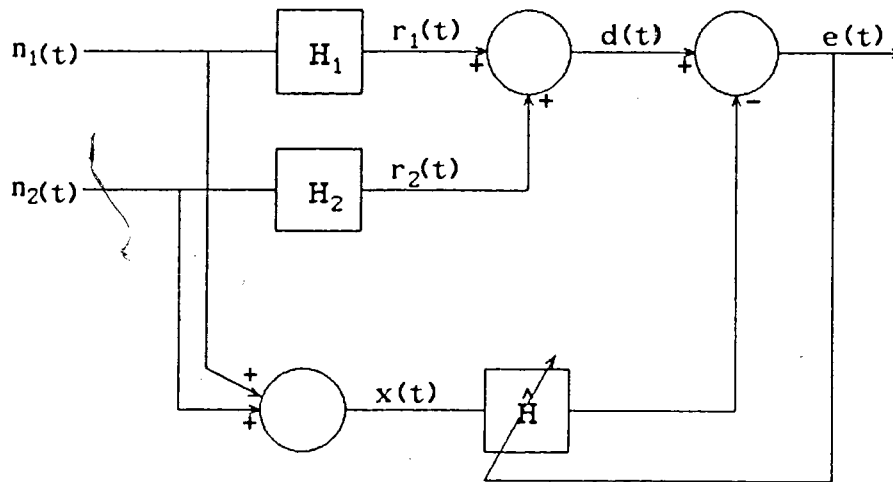
Figure 4.4 Coherence for Two Point Sources

Taking the Z Transform yields

$$S_{xd}(z) = S_{r_1 n_1}(z) + S_{r_2 n_2}(z)$$

$$= H_1(z)S_{n_1}(z) + H_2(z)S_{n_2}(z)$$

$$= (z^{-\tau_1} + z^{-\tau_2})S_n(z) \qquad (4.21)$$

Now

$$S_x(z) = |H_1(z)|^2 S_{n_1}(z) + |H_2(z)|^2 S_{n_2}(z)$$

$$= (|z^{-\tau_1}|^2 + |z^{-\tau_2}|^2)S_n(z) \qquad (4.22)$$

and

$$S_d(z) = S_{n_1}(z) + S_{n_2}(z)$$

$$= 2S_n(z) \qquad (4.23)$$

Then the MSC is given by

$$\gamma_{xd}^2(z) = \frac{\left| z^{-\tau_1} + z^{-\tau_2} \right|^2}{|z^{-2\tau_1}| + |z^{-2\tau_2}|} \qquad (4.24)$$

Evaluated along the unit circle, where $z = e^{j\omega}$, we have

$$\gamma_{xd}^2(e^{j\omega}) = \frac{\left| e^{-j\omega\tau_1} + e^{-j\omega\tau_2} \right|^2}{(2)(2)}$$

$$= \frac{1}{4}[(e^{-j\omega\tau_1} + e^{-j\omega\tau_2})(e^{j\omega\tau_1} + e^{j\omega\tau_2})]$$

$$= \tfrac{1}{2}[1 + \cos(\omega(\tau_1+\tau_2))] \tag{4.25}$$

The coherence has nulls and therefore no cancellation when

$$\omega(\tau_1+\tau_2) = (2n+1)\pi$$

or in terms of the cyclic frequency, when

$$f = \frac{2n+1}{2(\tau_1 + \tau_2)} \tag{4.26}$$

for any integer n.

Even in the simple case of two sources, there are frequencies where no cancellation occurs. While a case may be made for using more than one adaptive filter for adaptive noise cancellation,[28] there remains a problem. Such a technique would require as many filters as there are sources, which for the case of an automobile would be impractical.

When the number of noise sources increases to infinity, we approach a diffuse noise field. Chapter 2 gave the coherence of such a field. Recall that for significant coherence at all but the lowest frequencies, the two microphones must be very close together. In the case of directional microphones, the two must point in the same direction.

This brings us to the basic problem of adaptive noise cancellation in automobiles. In order to achieve significant coherence, the microphones must be close together and be pointing in the same direction. In a vehicle, such a configuration will never receive speech in one microphone without receiving it in the other. The effect of speech spill-over into the reference signal is now addressed.

4.2.2. Effect of Speech in Reference Signal

When the noise in the reference and primary signals are perfectly correlated and $\hat{H}$ is the unrestrained optimal Wiener filter, it can be shown[28] that the output signal-to-noise density ratio, $\rho_{out}(e^{j\omega})$ is the reciprocal of the signal to noise density ratio at the reference input, $\rho_{ref}(e^{j\omega})$:

$$\rho_{out}(e^{j\omega}) = \frac{1}{\rho_{ref}(e^{j\omega})}. \tag{4.27}$$

This is the so called power inversion process. For optimal performance, we require a low signal-to-noise density ratio at the reference input; however, in attempting to achieve this by separating the microphones or pointing them in different directions, we reduce the coherence in the noise received by the two microphones, and thereby reduce the noise-cancelling potential.

4.2.3. Non-Minimum Phase Properties

A more general model for ANC than that of figure 4.1 is given by the figure 4.5. In figure 4.1 it was implicitly assumed that we have access to $x(t)$, the noise source. In general, the reference noise will not be $x(t)$, but rather $x'(t)$, the output of some tranformation $H_2(\omega)$ when $x(t)$ is input.

Figure 4.6 shows an equivalent model, generated by multiplying the transfer functions in figure 4.5 by $H_2^{-1}(\omega)$. The optimal adaptive filter is now given by $\hat{H}(\omega) = H_1(\omega)H_2^{-1}(\omega)$. However, $H_2(\omega)$ will not, in general, be mimimum phase.[21,34] Then $H_2^{-1}(\omega)$ will have to be non-causal and doubly-infinite to be stable.

Practically, small amounts of delay can be introduced into $\hat{H}(\omega)$ to introduce some non-causality. However, the filter's finite length may cause problems.

We will now consider an adaptive filter based on the Least Mean Square Algorithm. This filter was used to test our predictons regarding the ineffectiveness of ANC.
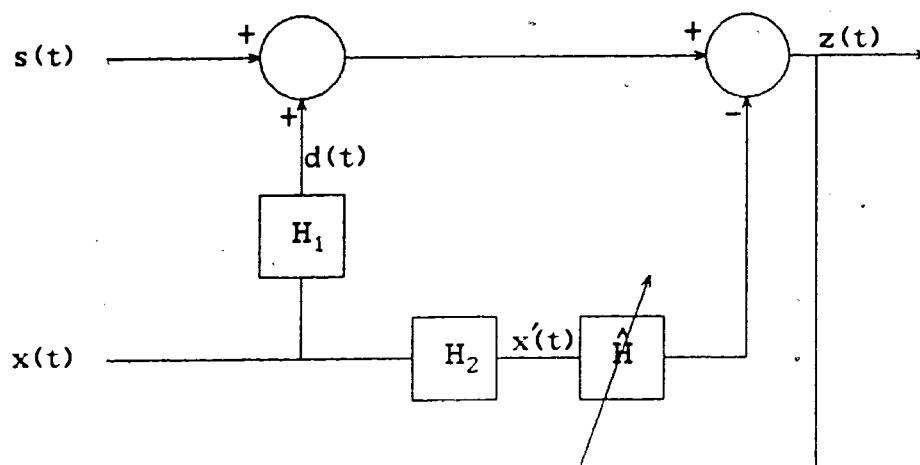
Figure 4.5 More General Model for ANC
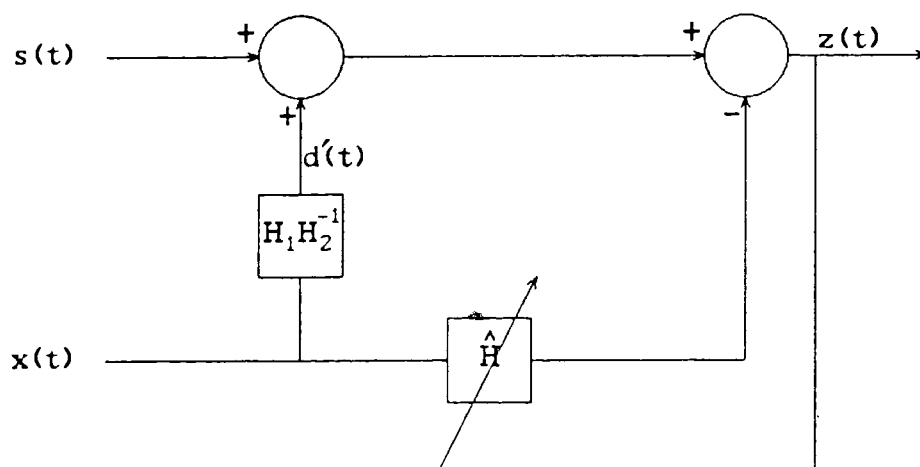
Figure 4.6 Equivalent Model for ANC

4.3. THE LEAST-MEAN-SQUARE ALGORITHM

The most common form for the adaptive filter is a finite impulse response (FIR) filter with variable coefficients. The Widrow-Hoff LMS [28] algorithm is a very popular and simple method of varying the coefficients of the FIR filter.

Recognizing that the limits of the indices in equations (4.6) to (4.10) are finite, we define the vector quantities

$$\vec{X_j} = [x(j), x(j-1), \cdots x(j-L+1)]^T$$

$$\vec{H} = [\hat{h}(0), \hat{h}(1), \cdots, \hat{h}(L-1)]^T \tag{4.28}$$

where the ^{T} denotes the transpose operator. We further define the matrix R as having as the $(i,j)^{th}$ element $R_x(i,j)$ or equivalently

$$R = E\{\vec{X_j}\vec{X_j^T}\} \tag{4.29}$$

Finally, we define the vector $\vec{P}$ as the vector of cross-correlations $R_{xd}(i)$ or equivalently

$$\vec{P} = E\{d_j\vec{X_j}\} \tag{4.30}$$

Then we can re-write equations (4.4) - (4.13) in terms of these vectors and matrices

$$y(j) = \vec{X_j^T}\vec{H} \tag{4.31}$$

$$e(j) = d(j) - \vec{X_j^T}\vec{H} \tag{4.32}$$

so that

$$\xi = E\{e^2\}$$

$$= E\{d^2\} + \vec{H}^T R\vec{H} - 2\vec{P}^T\vec{H} \tag{4.33}$$

Consider taking the gradient of ξ and setting it to zero as before:

$$\vec{\nabla}\xi = \frac{\partial \xi}{\partial \vec{H}}$$

$$= R\vec{H} - 2\vec{P} = 0 \tag{4.34}$$

Solving this, we have the optimal constrained filter weight (sometimes called the Wiener weight vector[28]):

$$\vec{H}_{opt} = R^{-1}\vec{P} \tag{4.35}$$

Note the similarity of equations (4.13) and (4.35), the difference being the length of the optimal filter.

Next consider the problem of minimizing the output power, σ_z^2, which is given by: Since

$$\sigma_z^2 = E\{z^2\} = E\{(z-s)^2\} + E\{s^2\}$$

$$= E\{(d-y)^2\}+E\{s^2\}$$

$$= \xi+E\{s^2\} \tag{4.36}$$

From equation (4.33), the function ξ and therefore the total output power is a hyper-parabolic surface in h space. Since R is positive definite, the surface has one global minimum.

One approach to find the minimum is the method of steepest descent. From vector calculus, recall that the gradient gives the direction of steepest increase for a function. In the steepest descent method, for a given $\vec{H}_k=$ the current vector of filter weights at time k, the gradient of the output power function is calculated, and a small multiple of the negative of the gradient is added to $\vec{H}_k$. Thus

$$\vec{H}_{k+1} = \vec{H}_k - \mu \vec{\nabla}_k \sigma_z^2 \tag{4.37}$$

where μ is a small constant and $\vec{\nabla}_k \sigma_z^2$ is the gradient of the output power function at time k. From the above, the gradient is given by

$$\vec{\nabla}_k \sigma_z^2 = \frac{\partial E\{z^2(k)\}}{\partial \vec{H}_k}$$

$$= \frac{\partial \xi(k)}{\partial \vec{H}_k} \tag{4.38}$$

since $E\{s^2\}$ is not a function of $\vec{H}_k$.

4.3.1. The Stochastic Gradient Approximation

The Least Mean Square algorithm makes the approximation that the expected value of e^2 can be replaced by its current value:

$$\frac{\partial \xi(k)}{\partial \vec{H}_k} = \frac{\partial E\{e^2(k)\}}{\partial \vec{H}_k}$$

$$\simeq \frac{\partial e^2(k)}{\partial \vec{H}_k}$$

$$= 2e(k) \frac{\partial e(k)}{\partial \vec{H}_k}$$

$$= 2e(k)\vec{X}_k \tag{4.39}$$

The filter update equation then becomes

$$\vec{H}_{k+1} = \vec{H}_k + 2\mu e(k)\vec{X}_k \tag{4.40}$$

Because this very simple equation requires no matrix inversion and no knowledge of second order statistics, it can be implemented easily using conventional digital signal processing hardware.[35,36]

4.3.2. Convergence Properties of Least Mean Square

While no proof of unconditional convergence has yet been published for the Least Mean Square algorithm, it can be shown [37] that for stationary d and x, $\vec{H}_k \rightarrow H_{opt}$ as $k \rightarrow \infty$ if

$$0 < \mu < \frac{1}{\lambda_{max}} \tag{4.41}$$

where λ_{max} is the largest eigenvalue of the autocorrelation matrix R. In general, it is too difficult to calculate the eigenvalues explicitly, and a more restrictive bound is used, by recalling that R is positive definite or at worst positive semi-definite so that

$$\lambda_{max} \le \sum_{i=1}^{L} \lambda_i \tag{4.42}$$

and

$$\sum_{i=1}^{L} \lambda_i = \mathrm{tr}(R) \tag{4.43}$$

where $\mathrm{tr}(R)$ is the trace of R. When the reference signal is assumed to be stationary,

$$\mathrm{tr}(R) = LE\{x^2\} \tag{4.44}$$

so that the conditions for convergence become

$$0 < \mu < \frac{1}{LE\{x^2\}} \qquad (4.45)$$

where $E\{x^2\}$ is the power in the reference signal.

While limited by equation (4.45), the choice of μ is not simple. A large μ will speed up convergence, but too large a value will cause excessive error in the steady state, as the filter will tend to oscillate about the optimal H_{opt}. One suggestion is to use an adaptive value for μ that is large when the error is large and becomes smaller as the error reduces (as $H \rightarrow H_{opt}$). For the stationary case, this is a viable alternative; however, for non-stationary noise, the filter will not be able to keep up with the changes in H_{opt} if μ is too small. Widrow[37] gives an approximate analysis of the dependence of steady-state error and convergence rate as a function of μ for the stationary environment.

4.4. EXPERIMENTAL RESULTS

To investigate the ability of equation (4.18) to predict the attenuation achievable by a practical adaptive filter, the LMS algorithm was used to attempt ANC on the data anlayzed at the end of chapter 3. Various filter lengths, values for μ, and amounts of non-causality were tried. The best results were achieved with a filter length of 450 and a delay of 50 samples introduced into the reference input. The value for μ used was ten times smaller than the value that gave instability in the algorithm. The resulting improvement, calculated as the ratio of the power in the original noise signal $d(k)$ over that in the error signal $e(k)$ is shown in table 4.1 and figures 4.8 - 4.12.

The SNR improvements are deceptive, as examining figures 4.7 through 4.11 reveals. While the SNR is improved in all cases, the gain is a result of cancellation at low frequency; in fact the adaptive filter actually adds noise at frequencies above 500 Hz. The SNR is a measure of signal energy, which is heavily concentrated at low frequencies. Since ANC performs well at low frequencies, the SNR gain is impressive, but actually a high-pass filter would perform as well without adding noise at the higher frequencies.

The omnidirectional microphones achieve a larger SNR gain due to their flat frequency response and therefore larger low frequency energy. (Since the directional

Table 4.1 SNR Improvement of ANC

Configuration	MSC Diagram #	ANC Diagram #	SNR Gain (dB)
omni	3.14 •	4.7	16.9
> >	3.15	4.8	4.4
^ ^	3.16	4.9	7.8
> <	3.17	4.10	2.0
^ v	3.18	4.11	7.9

microphones have a low-pass response, they also cancel this energy, but this cancellation doesn't appear as ANC gain.)[30] Note that if LMS were attempted after high pass filtering, no gains would be expected due to the low coherence at higher frequencies.

The increased noise at higher frequencies is a consequence of using the LMS algorithm.[34] Since the adaptive filter $\hat{H}$ is a linear combination of previous error outputs (equation 4.40), the filter does a poor job of estimating the optimal filter over the larger frequencies where the signal energy is lower.

While other algorithms[38,39] might not add the error at higher frequencies, they wouldn't be expected to improve on the optimal (Wiener) performance. Dal Degan and Prati[8] attempted used an adaptive lattice filter to attempt ANC in an automobile. Their results indicate that a lattice filter performs better at higher frequencies by not adding any noise (but not reducing any either), with similar attenuation to ours at lower frequencies.
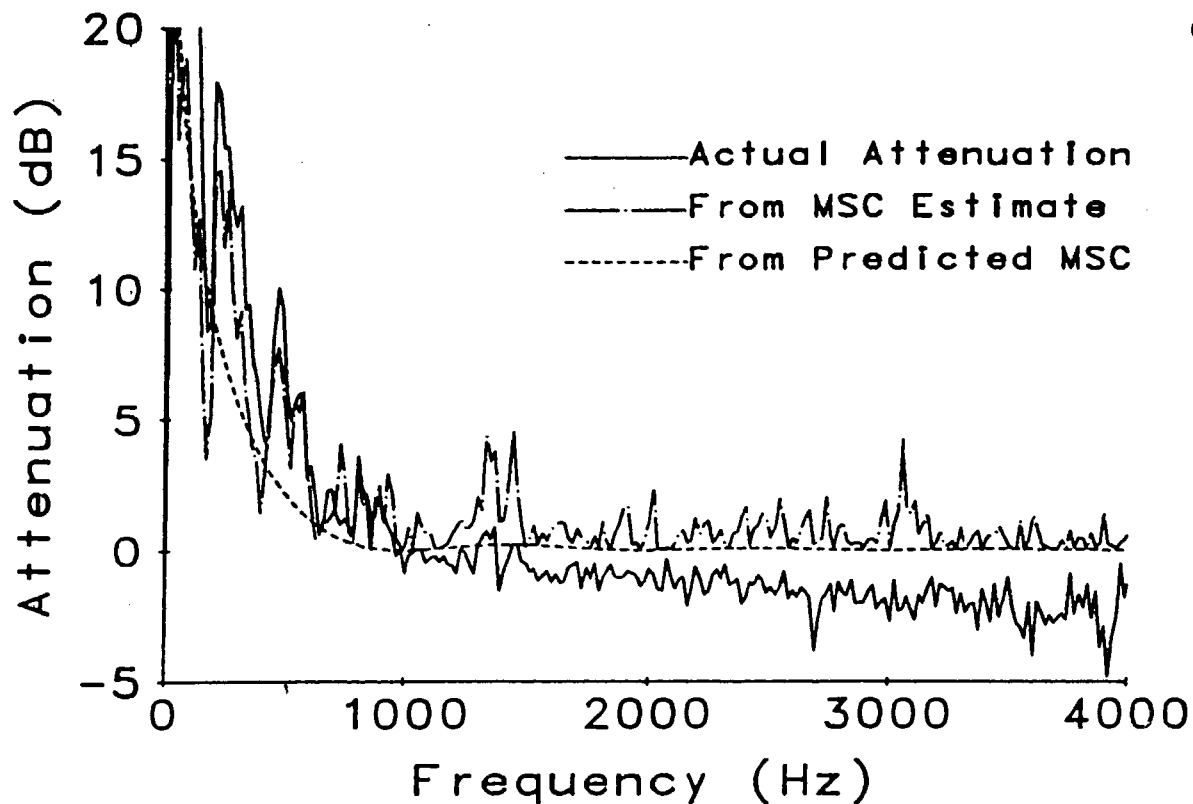
Figure 4.7 ANC Performance for Two Omnidirectional Microphones Separated by 15cm
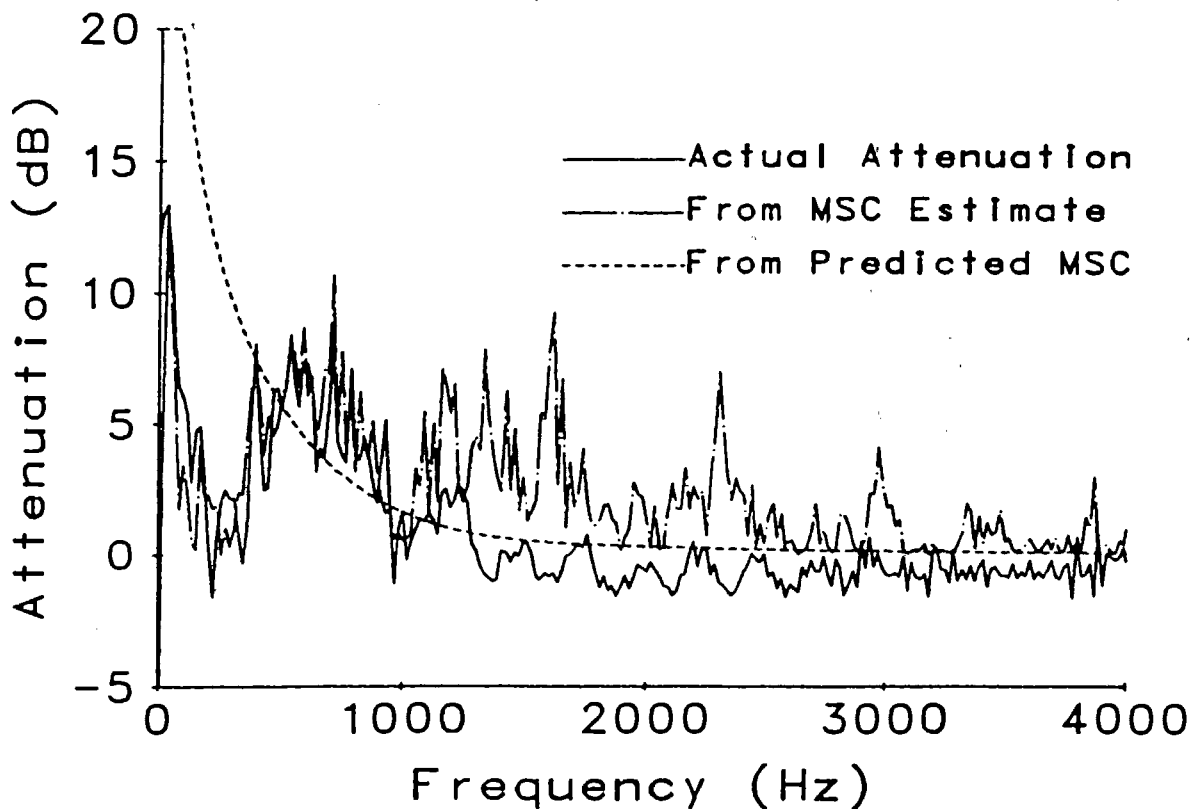
Figure 4.8 ANC Performance for Two Directional Microphones Separated by 15cm Both Pointing in Same Direction Along Axis of Separation
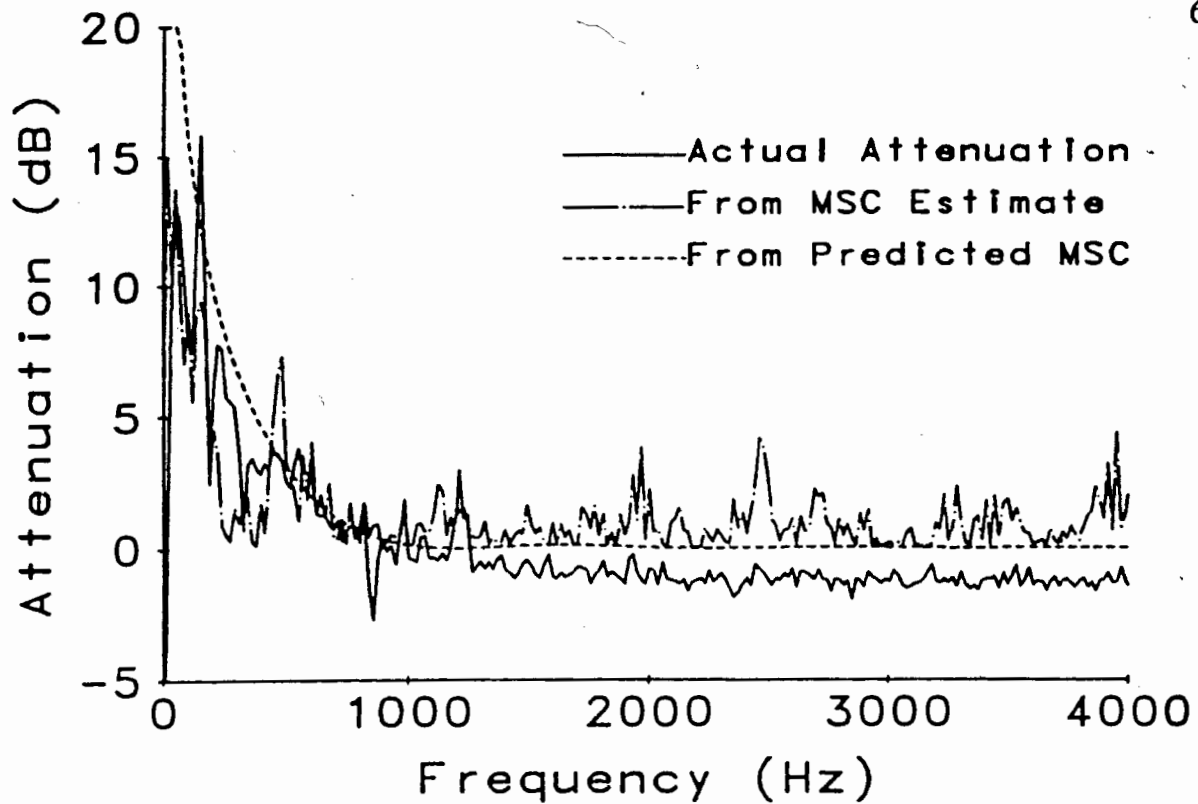
Figure 4.9 ANC Performance for Two Directional Microphones Separated by 15cm Both Pointing in Same Direction Perpendicular to Axis of Separation
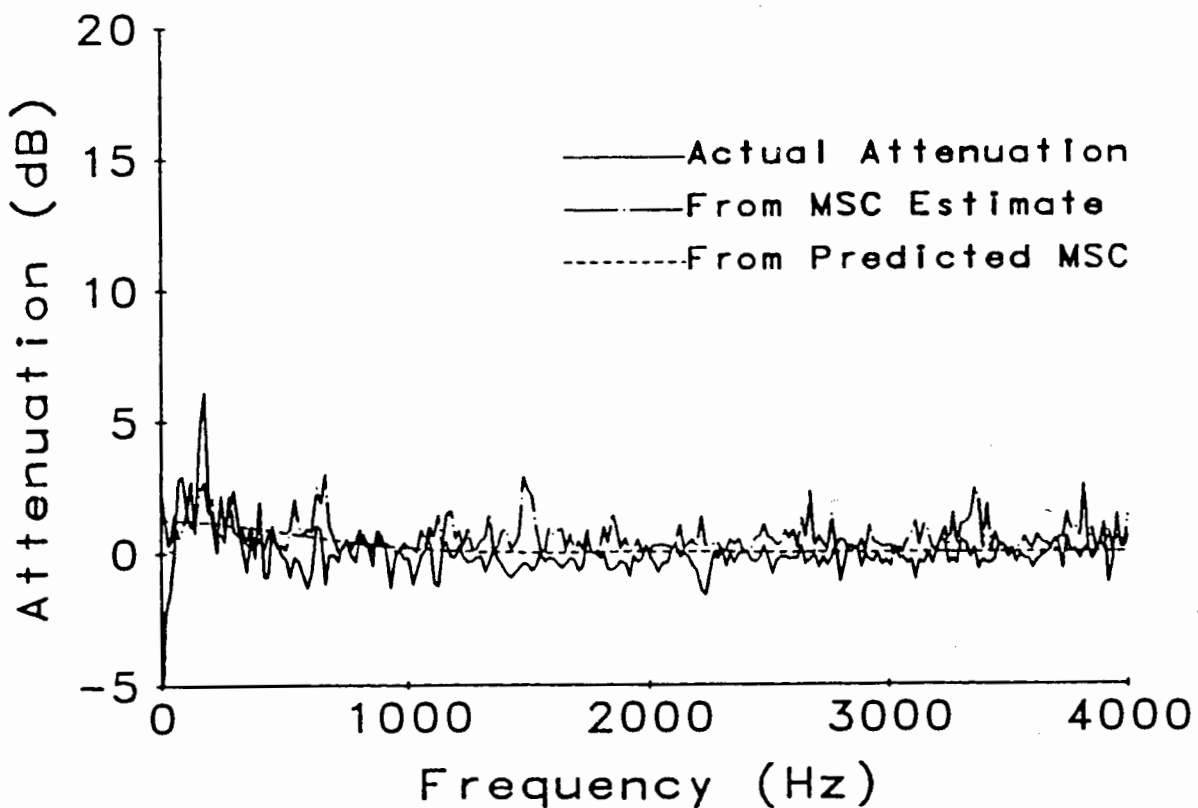
Figure 4.10 ANC Performance for Two Directional Microphones Separated by 15cm Pointing in Opposite Directions Along Axis of Separation
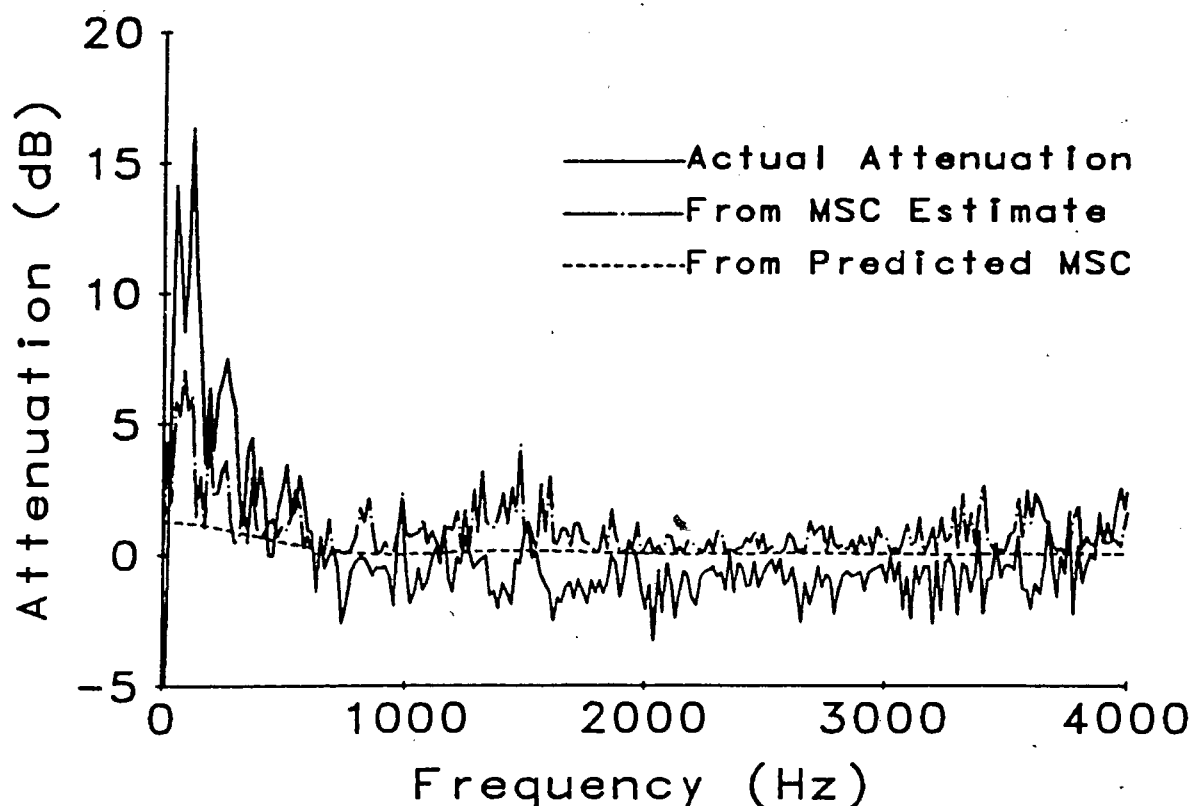
Figure 4.11 ANC Performance for Two Directional Microphones Separated by 15cm Pointing in Opposite Directions Perpendicular to Axis of Separation

4.4.1. Comparison With Other Researchers

There have been at least four other investigations involving ANC in the automobile interior. Two dealt with the case of an idling engine, and achieved relatively good ANC results, while the other two concerned a vehicle in motion, and reported results comparable to ours.

Goubran and Hafez[4] achieved 10 dB of noise cancellation from ANC with a 128 tap LMS filter with the car stopped and the engine idling. To achieve speech isolation, they placed the reference microphone "outside the cabin of the car, close to the engine compartment." While this is feasible for a stationary vehicle, the coherence between the noise in the vehicle interior and outside is not expected to be large when the car is in motion. As will be seen in chapter 5, when considering the delay-equalized array, the coherence in the noise field in the car is largest when the car is stationary, with the engine idling. This is because the noise field for a stationary vehicle is produced by a single dominant source, the engine. In addition, the noise is low in frequency and therefore has a relatively large coherence.[8]

Savoji[5] achieved between 7 and 13 dB by using a variable length adaptive lattice filter. In an attempt to achieve speech isolation, he placed a directional reference microphone under the dash facing the engine block and added speech digitally after recording noise alone. When we attempted a similar configuration, there was significant speech in the reference microphone.

Armbruster, Czarnach and Vary[7] arrived at similar conclusions to ours regarding the effectiveness of ANC in an automobile. They achieved 6 dB of (low-frequency) attenuation, without any speech isolation.

In the most recent paper, Dal Degan and Prati also[8] arrived at similar conclusions. In a related study, Andresdottir and Schafer[20] found ANC to be ineffective in simulated reverberant environments.

One might think that an acoustic barrier introduced between the primary and reference microphones might help. Such is the case in fighter jet cockpits, where the pilot wears a helmet. After much study[30,31,40] in the aircraft environment, the conclusion seems to be that ANC is ineffective due to the diffuse noise field.

Another ANC technique is shown in figure 4.12. The sum and difference signals of the original primary and reference inputs are used as the new primary and reference inputs to a standard ANC filter. The reference consists of noise only, since the speech is subtracted out assuming that the speech is received coherently at both inputs. Such a filter is called a Griffiths-Jim beamformer.[41,42] The condition for effective performance is a very low (ideally one) number of noise sources, spatially isolated, so that the beamformer can place a null in the direction of the noise. In the car, where there are many noise sources and reverberation placing image noise sources in every direction, this condition is not satsified. In a study with simulated rooms, Peterson[42] found the Griffiths-Jim beamformer to be ineffective when there was a single noise source in a reverberant room. When the sources are spatially diffuse, the situation would be even worse.

We have seen why ANC cannot be effective in an automobile interior. Experimental results from use of the LMS algorithm have confirmed this expectation. Other techniques involving adaptive beamforming are similarly of no use. A new approach to speech enhancement is required. Chapter 5 will introduce our solution: a delay-equalized

near-field beamformer.
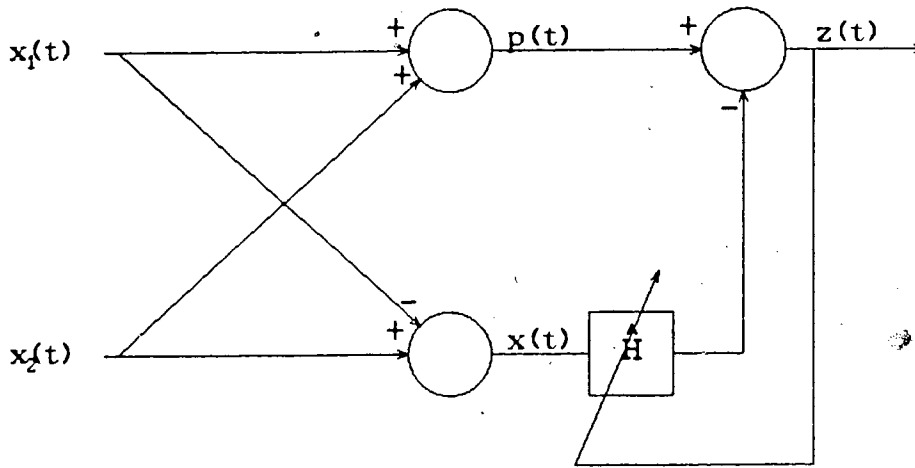
Figure 4.12 Griffiths-Jim Beamformer

5. DELAY-EQUALIZED NEAR-FIELD BEAMFORMER

In chapter 4, we saw that Adaptive Beamforming is ineffective in the automotive environment, due to the large number of noise sources and reverberation. We now consider ways of enhancing the speech rather than concentrating on cancelling the noise. This chapter introduces a new model which incorporates a numer of signals each composed of speech plus noise, with the noise in each being mutually uncorrelated. A simple method of combining the signals to coherently add the speech is derived and shown to have robust performance in improving SNR, as well as reducing the reverberation in the speech. This method is similar to the non-adaptive beamformer of Andresdottir and Schafer[20] except that we have a near-field situation where we cannot assume that the speech signals are plane waves.

In section 5.1, we consider the new model and show that optimal performance of the array can be achieved by simple delaying and scaling each of the signals so that the speech is coherently re-inforced. The array's signal enhancement ability is then analyzed and found to be a function of the coherence of the noise between the noise in the individual signals. Section 5.2 gives experimental results for the noise reduction measured by using the array in a vechicle, and compares these with theoretical results based on a diffuse field approximation. The effect of the array on the speech is considered in section 5.3, where we see that the reverberation of the speech is significantly reduced. In section 5.4, we consider some practical design problems of the array. Finally, section 5.5 draws some conclusions.

5.1. A NEW MODEL: A NEW APPROACH

Figure 5.1 illustrates a signal model more likely to be valid in the acutal vehicular environment than the model required for ANC. The inputs $x_i(t)$, consist of speech $s(t)$ passed through a sum of transfer functions $g_i(t) + r_i(t)$ and corrupted by noises $n_i(t)$. The inputs are then filtered by functions $h_i(t)$ and summed giving $v(t)$. The direct-arrival path between the speech source and microphone i is given by $g_i(t)$, where

$$g_i(t) = \frac{\delta(t-\tau_i)}{c\tau_i}$$

(5.1)

Here τ_i is the time for the speech to travel to microphone i, c is the speed of sound, and $\delta(t)$ is the dirac-delta impulse function. The denominator, $c\tau_i$, gives the attenuation of the speech over the path, since acoustic signal strength falls off inversely with distance. The function $r_i(t)$ describes the transfer function generating the reverberant tail of the speech in microphone i. We assume[43] that when the speech passes through the $r_i(t)$ the outputs become decorrelated. That is, the reverberant speech tails for each microphone are uncorrelated with the others. The noises $n_i(t)$ are also assumed to be uncorrelated with each other and with $s(t)$, and to have equal power spectral densities.
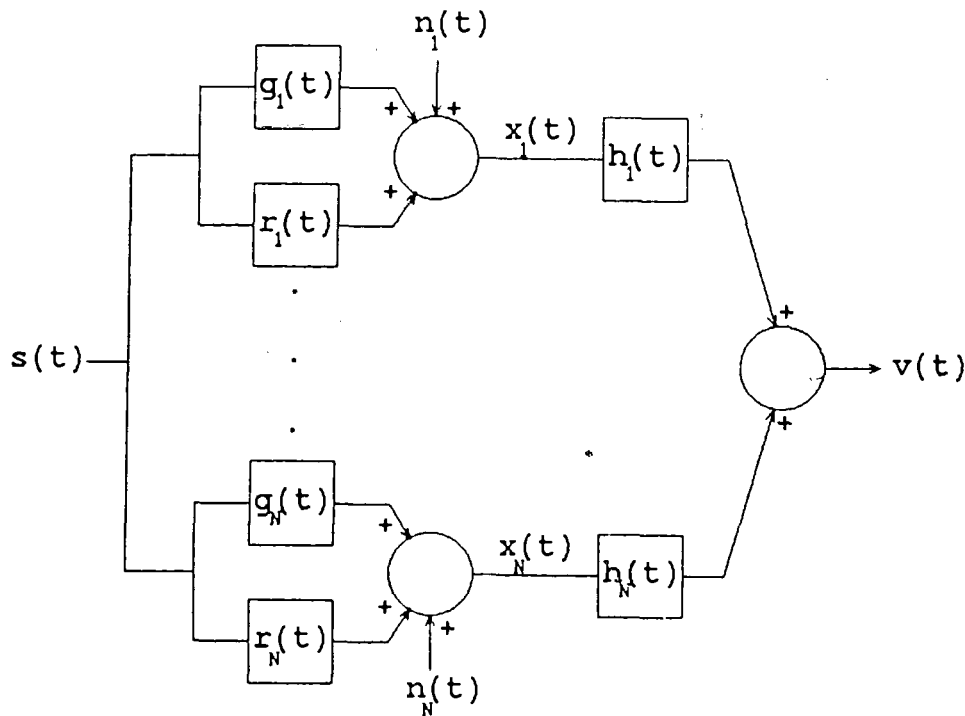
Figure 5.1 Model for Non-Adaptive Beamformer

5.1.1. A Distortionless Array

We begin by defining the output signal, $v(t)$

$$v(t) = s_{out}(t) + n_{out}(t) \tag{5.2}$$

where the output speech signal

$$s_{out}(t) = \sum_{i=1}^{N} g_i(t) * s(t) * h_i(t) \tag{5.3}$$

$$= \sum_{i=1}^{N} \frac{s(t-\tau_i)}{c\tau_i} * h_i(t) \tag{5.4}$$

from the definition of $g_i(t)$. The output noise signal, $n_{out}(t)$ is given by

$$n_{out}(t) = \sum_{i=1}^{N} \left[h_i(t)*r_i(t)*s(t) + h_i(t)*n_i(t) \right] \tag{5.5}$$

Consider the output speech power

$$E\{s_{out}^2(t)\} = E\left\{ \sum_{i=1}^{N}\sum_{j=1}^{N} \left[\frac{s(t-\tau_i)}{c\tau_i} * h_i(t) \right]\left[\frac{s(t-\tau_j)}{c\tau_j} *h_j(t) \right] \right\} \tag{5.6}$$

In order not to introduce any distortion of the speech, the $h_i(t)$ are constrained to be a simple function consisting of a scale factor and a delay:

$$h_i(t) = \alpha_i c\delta(t-\zeta_i) \tag{5.7}$$

where α_i is the variable weight, c is the speed of sound and ζ_i is the amount of delay. With this definition of $h_i(t)$

$$E\{s_{out}^2(t)\} = E\left\{ \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\alpha_i\alpha_j}{\tau_i\tau_j} s(t-\tau_i-\zeta_i) s(t-\tau_j-\zeta_j) \right\} \tag{5.8}$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\alpha_i\alpha_j}{\tau_i\tau_j} R_s(\tau_i+\zeta_i-\tau_j-\zeta_j) \tag{5.9}$$

From basic autocorrelation theory, for all τ

$$R_s(0) \geq R_s(\tau) \tag{5.10}$$

So, to maximize the output speech power, we want

$$\tau_i + \zeta_i - \tau_j - \zeta_j \equiv 0 \tag{5.11}$$

for all i and j. The solution is to let

$$\zeta_i = T - \tau_i \quad \text{where } T \geq \max_i(\tau_i) \tag{5.12}$$

Very simply, we are delaying each signal to combine the speech coherently at the output. The output speech signal will then be

$$s_{out}(t) = \sum_{i=1}^{N} \frac{\alpha_i}{\tau_i} s(t-T) \tag{5.13}$$

and the output speech signal power will be

$$E\{s_{out}^2(t)\} = \left| \sum_{i=1}^{N} \frac{\alpha_i}{\tau_i} \right|^2 E\{s^2(t-T)\} \tag{5.14}$$

$$= \left| \sum_{i=1}^{N} \frac{\alpha_i}{\tau_i} \right|^2 \sigma_s^2 \tag{5.15}$$

where $\sigma_s^2 =$ the input speech signal power.

Now, the output noise signal, $n_{out}(t)$ is given by

$$n_{out}(t) = \sum_{i=1}^{N} [h_i(t) * r_i(t) * s(t) + h_i(t) * n_i(t)] \tag{5.16}$$

$$= \sum_{i=1}^{N} [\alpha_i cr_i(t) * s(t-\zeta_i) + \alpha_i cn_i(t-\zeta_i)] \tag{5.17}$$

so that the output noise power is

$$E\{n_{out}^2(t)\} = \sum_{i=1}^{N} \sum_{j=1}^{N} [\alpha_i cr_i(t) * s(t-\zeta_i) + \alpha_i cn_i(t-\zeta_i)][\alpha_j cr_j(t) * s(t-\zeta_j) + \alpha_j cn_j(t-\zeta_j)] \tag{5.18}$$

but we assume that the $r_i(t)$ and $n_i(t)$ are uncorrelated so that

$$E\{n_{out}^2(t)\} = E\{\sum_{i=1}^{N}[\alpha_i cr_i(t) * s(t-\zeta_i)]^2\} + E\{\sum_{i=1}^{N}[\alpha_i cn_i^2(t-\zeta_i)]^2\} \qquad (5.19)$$

$$= \sum_{i=1}^{N}\alpha_i^2\left[c^2 E\{[r_i(t) * s(t-\zeta_i)]^2\}\right] + \sum_{i=1}^{N}\alpha_i^2\left[c^2 E\{n_i^2(t-\zeta_i)\}\right] \qquad (5.20)$$

$$= \sum_{i=1}^{N}\alpha_i^2(\beta_i + \eta_i) \qquad (5.21)$$

where we have defined β_i and η_i by the large brackets above.

The output signal-to-noise ratio, ρ_{out} will be

$$\rho_{out} = \frac{E\{s_{out}^2(t)\}}{E\{n_{out}^2(t)\}} \qquad (5.22)$$

$$= \frac{\left|\sum_{i=1}^{N}\dfrac{\alpha_i}{\tau_i}\right|^2 \sigma_s^2}{\sum_{i=1}^{N}\alpha_i^2(\beta_i + \eta_i)} \qquad (5.23)$$

Now, from Schwartz' inequality for complex numbers,[44]

$$\left|\sum_{i=1}^{N}\frac{\alpha_i}{\tau_i}\right|^2 \leq \left(\sum_{i=1}^{N}\frac{|1/\tau_i|^2}{\beta_i + \eta_i}\right)\left(\sum_{i=1}^{N}|\alpha_i|^2(\beta_i + \eta_i)\right) \qquad (5.24)$$

so that

$$\rho_{out} \leq \frac{\sum_{i=1}^{N}|1/\tau_i|^2\sigma_s^2}{\beta_i + \eta_i} \qquad (5.25)$$

The equal sign holds for

$$\alpha_i = \frac{K(\dfrac{1}{\tau_i})}{\beta_i + \eta_i} \qquad (5.26)$$

where K is a constant.

The results are self-evident. For maximal output signal-to-noise ratio, we delay the input signals so that the speech adds coherently. Those signals having higher input speech-to-noise ratios are given a larger weight than those with lower. The SNR is a function of $(\frac{1}{\tau_i})^2$; for maximal gain we want the microphones as close to the speaker as possible. Unfortunately, this will cause the correlation of the noise inputs to increase. This is the basic trade-off in the design of the array, and will be further discussed in section 5.4.

When the β_i and the η_i are assumed to be constants for all i, equation (5.26) becomes

$$\alpha_i = \frac{K}{\tau_i} \qquad (5.27)$$

For fair evaluation of the array's performance, we want the speech energy in the array to equal that from the microphone nearest the source. From equation (5.13) this implies that

$$\sum_{i=1}^{N} \frac{\alpha_i}{\tau_i} = \frac{1}{\tau_{min}} \qquad (5.28)$$

where τ_{min} is the delay from the speech source to the nearest microphone. When the definition of (5.27) is used in (5.28),

$$\sum_{i=1}^{N} \frac{K}{\tau_i^2} = \frac{1}{\tau_{min}} \qquad (5.29)$$

which we can solve for K, resulting in an expression for the individual α_i:

$$\alpha_i = \frac{1}{\tau_i \tau_{min} \sum_{j=1}^{N} \frac{1}{\tau_j^2}} \qquad (5.30)$$

We have chosen a very simple form for the 'filters' $h_i(t)$. An obvious question is whether we can do any better by increasing the complexity of the filters.

5.1.2. Alternate Filters for $h_i(t)$

In a recent paper, Zelinski[45] proposed an array that use the solution of equation (4.13) (the optimal Wiener filters) for the $h_i(t)$:

$$H_{i_{opt}}(\omega) = \frac{S_s(\omega)}{S_{x_i}(\omega)} \qquad (5.31)$$

To find the optimal Wiener $h_i(t)$, an estimate of the power spectra of the speech, $S_s(\omega)$ is required. Consider taking the cross-correlations between two inputs $x_i(t)$ and $x_j(t)$, where the two have been delayed so that the speech in each is coherent. Then the cross correlation is given by

$$R_{x_ix_j}(\tau) = E\left\{ x_i(t)x_j(t+\tau) \right\} \qquad (5.32)$$

$$= E\left\{ \left[\frac{s(t)}{c\tau_i} + r_i(t)*s(t) + n_i(t-\zeta_i) \right] \left[\frac{s(t+\tau)}{c\tau_j} + r_j(t)*s(t+\tau) + n_j(t+\tau-\zeta_j) \right] \right\} \qquad (5.33)$$

This would be an effective estimate if the noise were uncorrelated. Recall from chapter 3 that the coherence is significant at low frequencies so that low frequency noise, as well as the speech will be enhanced by the filter. Zelinski solved this problem by using a priori estimates of the low frequency energy. He could do this since he was dealing in a quasi-stationary environment (an office), but we aren't afforded the same luxury, since (from chapter 3) the noise power spectral density changes significantly under different vehicle conditions.

A more fundamental question is whether we would want to use Wiener filtering when our objective is to provide speech for computer based recognition. While the Wiener filter does give optimal performance *in the least square sense,* the frequency distortion it introduces would probably lessen its usefulness for speech recognition. Fortunately, the noise and typical speech spectra[27] aren't dissimilar, so we don't expect the

difference between a Wiener filter and our simple array to be significant. (Note that if the noise spectra and speech spectra are identical, the optimal Wiener filter is our simple array).

We will now analyze the gains of the array when the simple form of (5.6) is used for the $h_i(t)$.

5.2. EFFECT OF ARRAY ON NOISE

We will begin by deriving the theoretical noise gain of the array in section 5.2.1. Experimental results will be presented in section 5.2.2.

5.2.1. Theoretical

Consider figure 5.2.
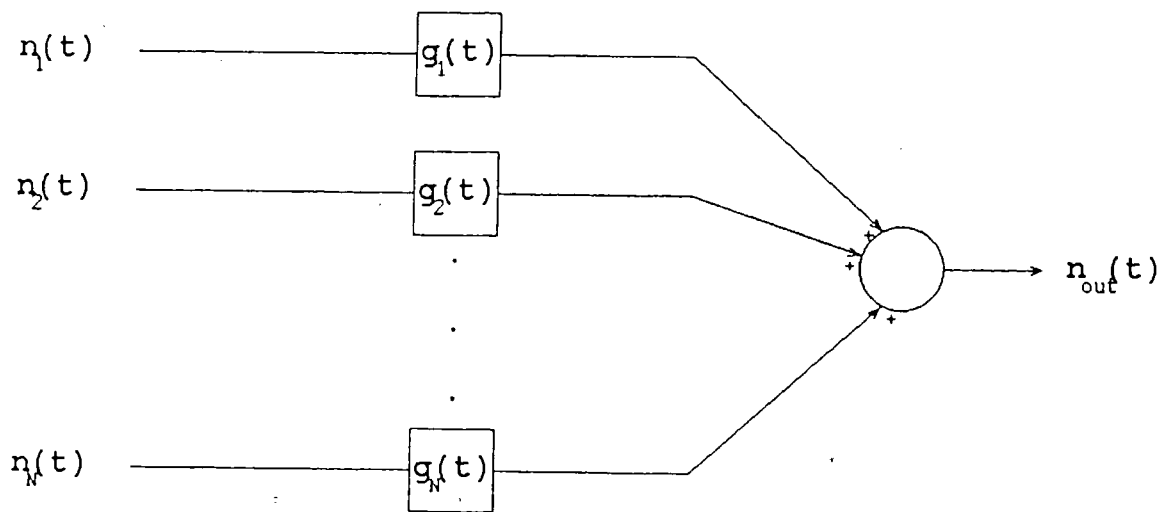
Figure 5.2 Effect of Array on Noise

The output noise, as with equation (5.5) is

$$n_{out}(t) = \sum_{i=1}^{N} \alpha_i n_i(t-\zeta_i)$$

(5.34)

The autocorrelation function of the output noise is

$$R_{n_{out}}(\tau) = E\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j n_i(t-\zeta_i)n_j(t-\zeta_j+\tau)\right\} \tag{5.35}$$

$$= \sum_{i=1}^{N}\alpha_i^2 R_{n_i}(\tau) + \sum\sum_{i\neq j}\alpha_i\alpha_j R_{n_i n_j}(\tau+\zeta_i-\zeta_j) \tag{5.36}$$

Taking the Fourier Transform to find the power spectral density,

$$S_{n_{out}}(\omega) = \sum_{i=1}^{N}\alpha_i^2 S_{n_i}(\omega) + \sum\sum_{i\neq j}\alpha_i\alpha_j S_{n_i n_j}(\omega)e^{j\omega(\zeta_i-\zeta_j)} \tag{5.37}$$

Consider re-writing the second term

$$\sum\sum_{i\neq j}^{N}\alpha_i\alpha_j S_{n_i n_j}(\omega)e^{j\omega(\zeta_i-\zeta_j)} = \sum\sum_{i<j}^{N}\alpha_i\alpha_j S_{n_i n_j}(\omega)e^{j\omega(\zeta_i-\zeta_j)} + S_{n_j n_i}(\omega)e^{-j\omega(\zeta_i-\zeta_j)} \tag{5.38}$$

$$= \sum\sum_{i<j}^{N}\alpha_i\alpha_j\left[S_{n_i n_j}(\omega)e^{j\omega(\zeta_i-\zeta_j)} + \left[S_{n_i n_j}(\omega)e^{j\omega(\zeta_i-\zeta_j)}\right]^*\right] \tag{5.39}$$

$$= 2\sum\sum_{i<j}^{N}\alpha_i\alpha_j \operatorname{Re}(S_{n_i n_j}(\omega)e^{j\omega(\zeta_i-\zeta_j)}) \tag{5.40}$$

where we have used the identities that $S_{xy}^*(\omega) = S_{yx}(\omega)$ and $z+z^* = 2\operatorname{Re}(z)$ for any complex z. Now from equation 3.10,

$$\gamma_{ij}(\omega) = \frac{S_{n_i n_j}(\omega)}{[S_{n_i}(\omega)S_{n_j}(\omega)]^{\frac{1}{2}}} \tag{5.41}$$

We next make the assumption that the noise power spectral densitity in each of the inputs is the same, that is $S_{n_i}(\omega) = S_n(\omega)$ for all i. Then

$$\gamma_{ij}(\omega) = \frac{S_{n_i n_j}(\omega)}{S_n(\omega)} \tag{5.42}$$

and combining equations (5.33), (5.36) and (5.40), we can write the output noise power spectral density as

$$S_{n_{out}}(\omega) = \left[\sum_{i=1}^{N} \alpha_i^2 + 2\sum\sum_{i<j}^{N} \alpha_i \alpha_j \text{Re}\left[\gamma_{ij}(\omega)e^{j\omega(\zeta_i - \zeta_j)} \right] \right] S_n(\omega) \qquad (5.43)$$

Consider the noise output from a single mic with weighting $= 1$. Then the output noise spectral density of the single microphone is $S_n(\omega)$, and the gain or the noise reduction capability of the array over a single microphone is

$$\text{Gain}(\omega) = \cfrac{1}{\sum_{i=1}^{N} \alpha_i^2 + 2\sum\sum_{i<j}^{N} \alpha_i \alpha_j \text{Re}\left[\gamma_{ij}(\omega)e^{j\omega(\zeta_i - \zeta_j)} \right]} \qquad (5.44)$$

It is instructive to investigate the limits of equation (5.44). Consider the case where we have a circular array, so that the ζ_i are equal and therefore the α_i are all $= \frac{1}{N}$. For the worst scenario, consider the noise in the microphones to be completely correlated so that $\gamma_{ij}(\omega)$ is always 1. Then we intuitively expect no gains from the array, and the equation does indeed predict this since

$$2\sum\sum_{i<j}^{N} \frac{1}{N^2} = \frac{1}{N^2}[N(N+1) - 2N] = 1 - \frac{1}{N} \qquad (5.45)$$

so that

$$\text{Gain} \geq \cfrac{1}{\frac{1}{N} + 1 - \frac{1}{N}} = 1 \qquad (5.46)$$

For the worst case, we expect no improvement (but no degradation).

At the other limit, where the noise inputs are mutually uncorrelated, so that $\gamma_{ij}(\omega) = 0$ for all $i \neq j$, the gain is N. (see equation 5.44)

5.2.2. Noise Cancelling Performance of the Array in an Automobile

When combined with the results of chapter 3 and appendices 1 and 2, equation (5.44) allows us to predict the expected array gain for any configuration of microphones in a diffuse noise field.

Consider the semi-circular array of figure 5.3. The distance between source and microphones, and therefore the α_i and ζ_i are all equal. For this geometry, the optimal array is simply the average of the inputs at the microphones. Figure 5.4 gives the gains predicted by equation (5.44) for arrays of 5 cardioid and 5 omnidirectional microphones in a diffuse noise field. Note the array gains are relative to a single microphone of the same type as are in the array (recall from chapter 2 that an omnidirectional microphone receives 4.8 dB more noise). The major difference between the curves is the gain predicted at low frequency for cardioid microphones, which is due to a coherence value of less than one (from the misalignment of the microphone axes).

Figures 5.5 to 5.10 give the experimental performance of the circular array of 5 directional microphones under various conditions. The array performs well under all but the quietest conditions. In figure 5.5, where the idling engine is the only source of noise, it is no surprise that the array doesn't give great results, since the coherence is expected to be much higher (and therefore the gain much lower). This explains the gains that other researchers[4] achieved through ANC when the engine was idling. Increasing the noise level by turning up the fan (figures 5.6 and 5.7), increasing the speed (5.8 through 5.10), and opening the window (5.9) all reduce the coherence and increase the gain of the array.
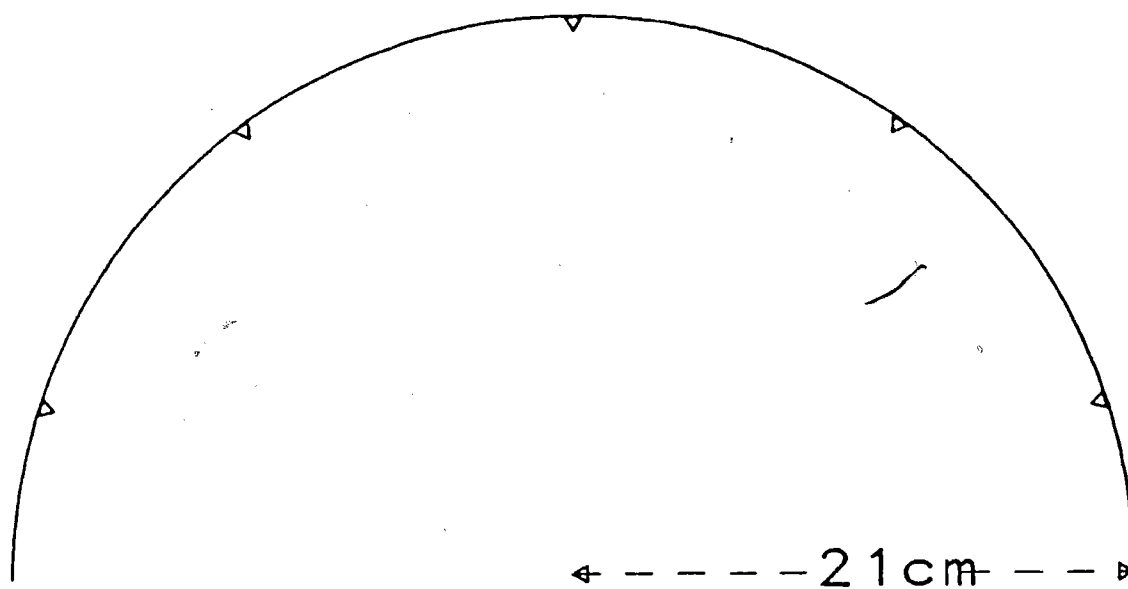
Figure 5.3 Semi-Circular Array
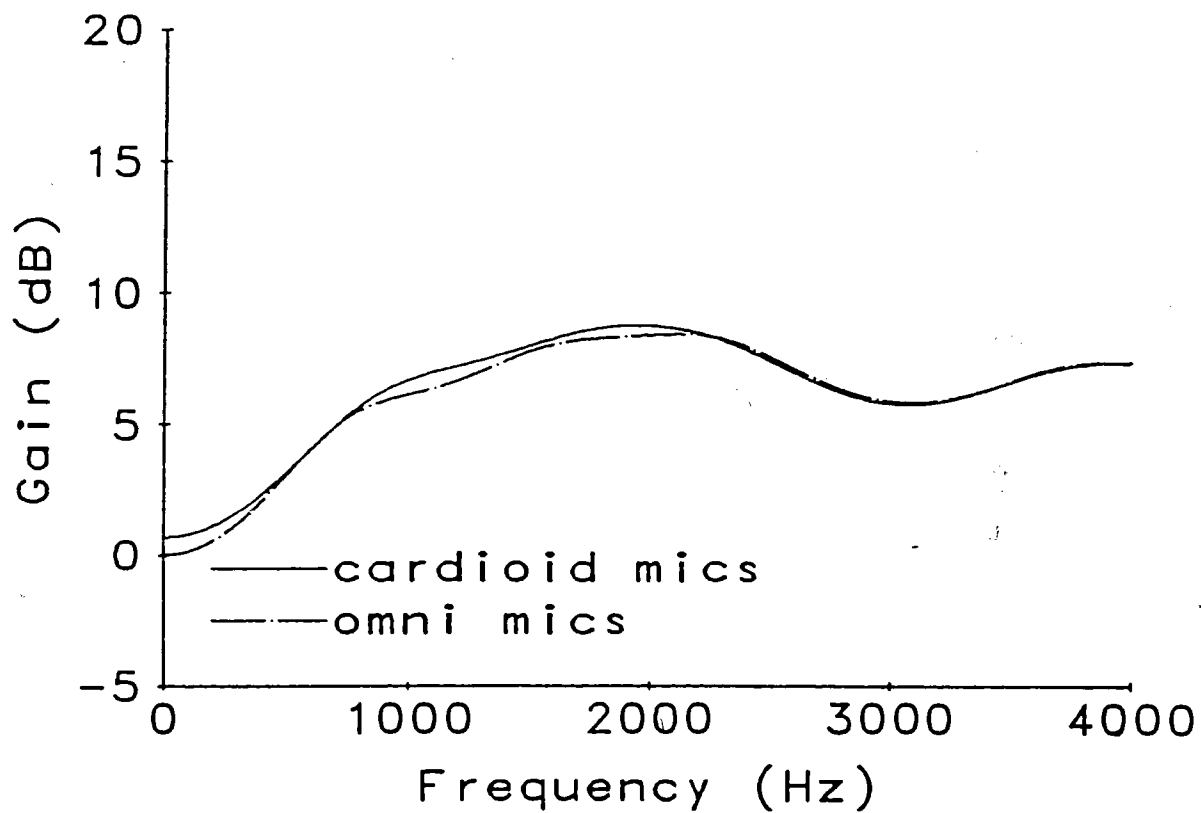
Figure 5.4 Theoretical Array Gains for Semi-Circular Array of
5 Microphones in Diffuse Noise Field
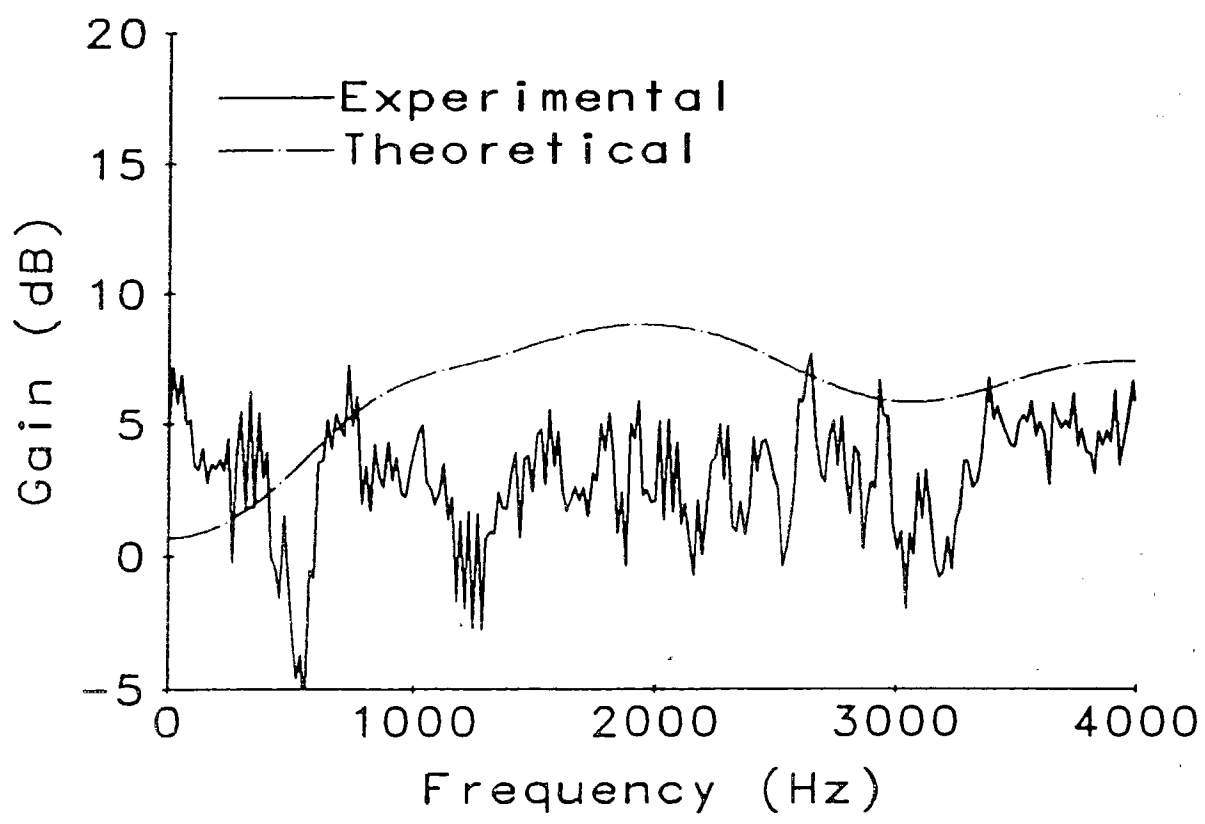
Figure 5.5 Array Gain With Engine Idling and Fan Off
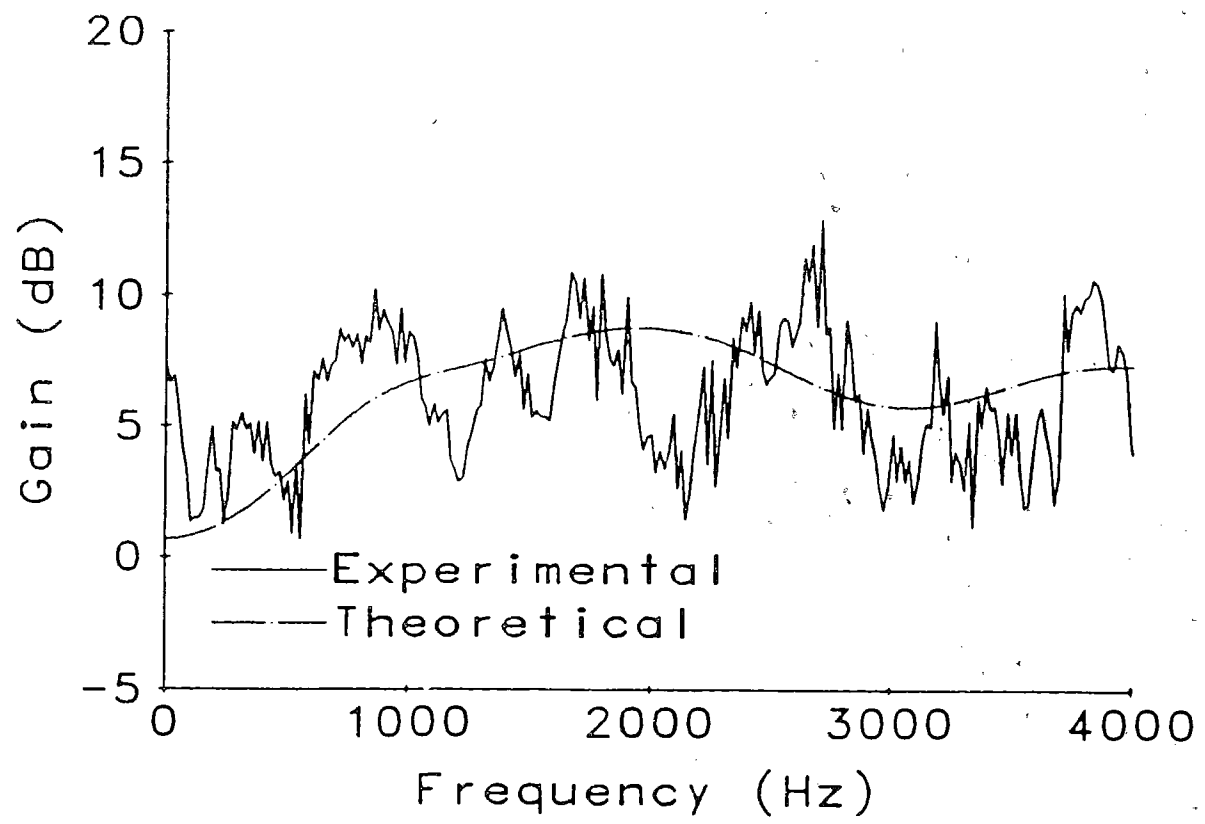
Figure 5.6 Array Gain With Engine Idling and Fan Low
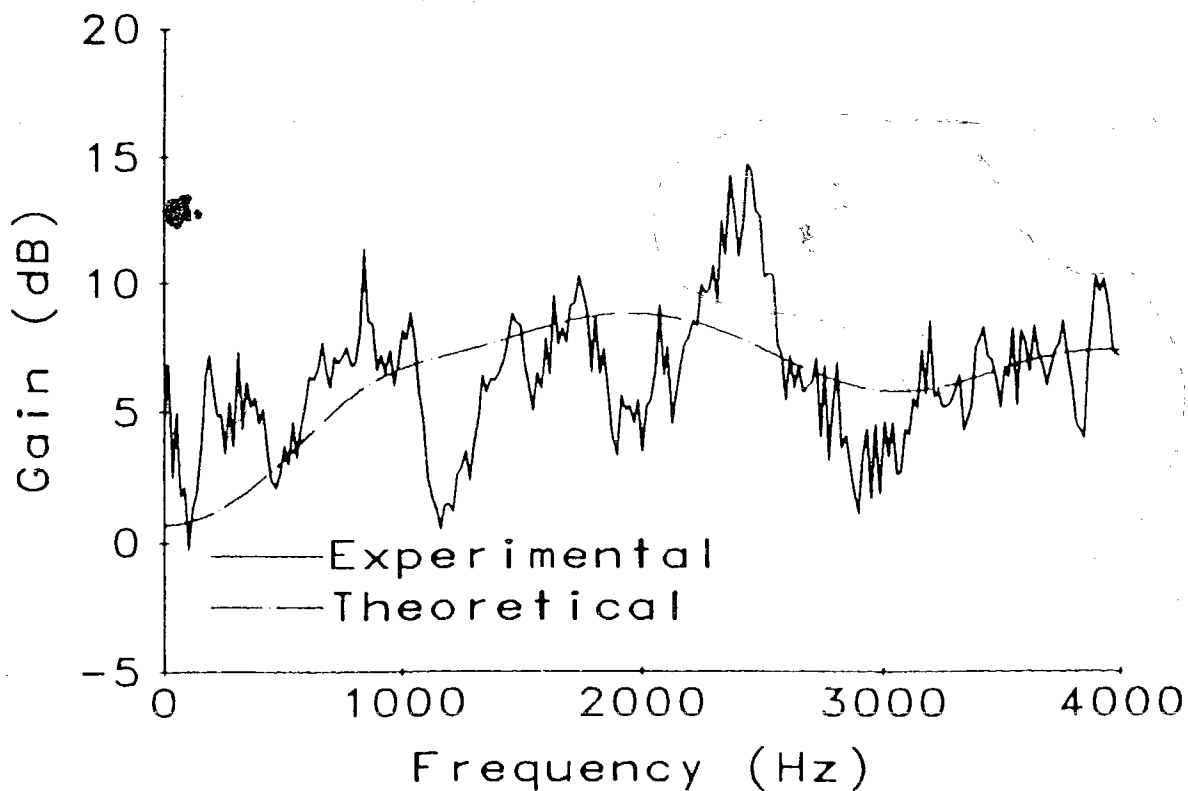
Figure 5.7 Array Gain With Engine Idling and Fan High
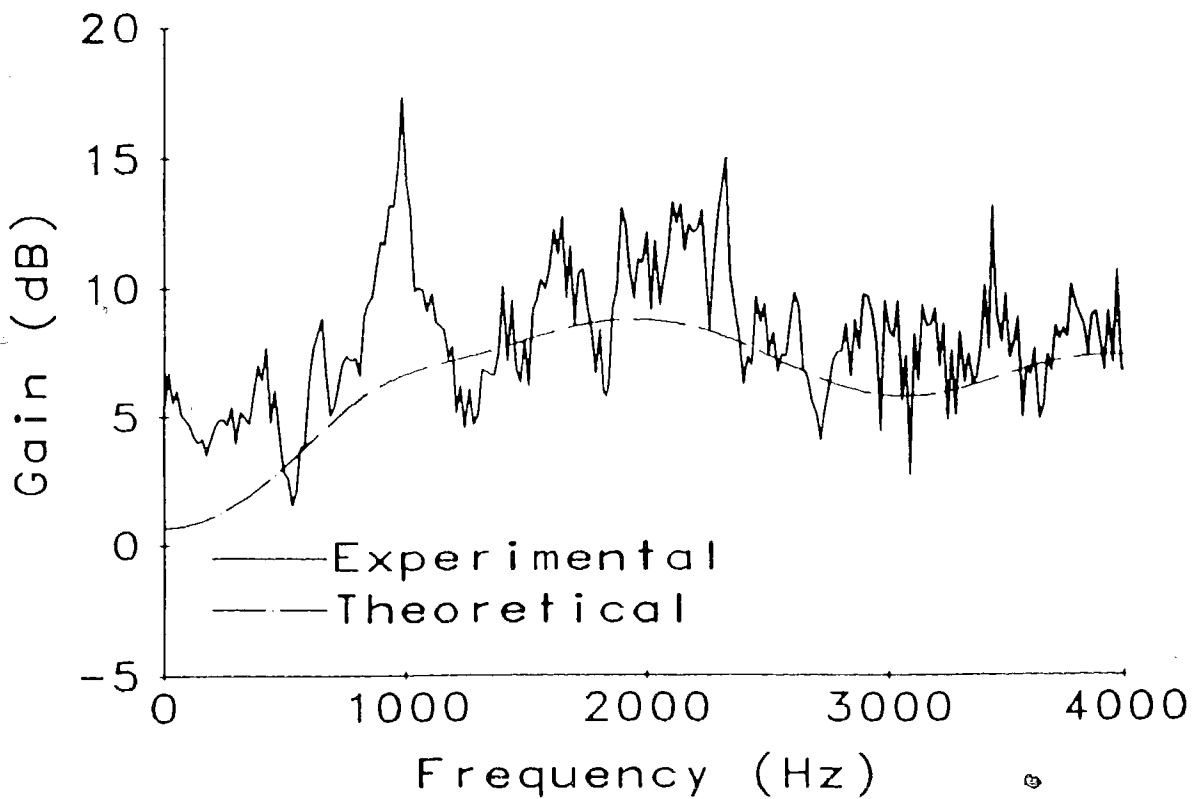
Figure 5.8 Array Gain at 50km/h With Fan Off and Windows Closed
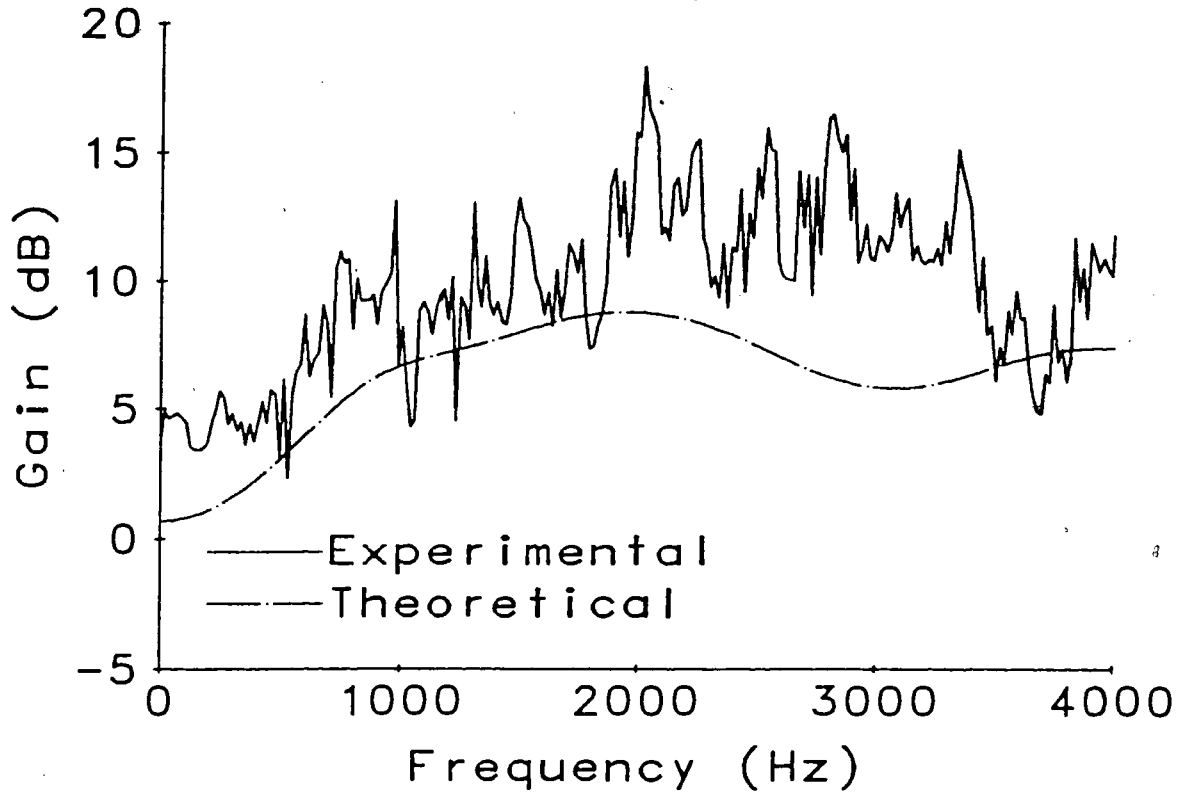
Figure 5.9 Array Gain at 50km/h With Fan Off and Windows Open
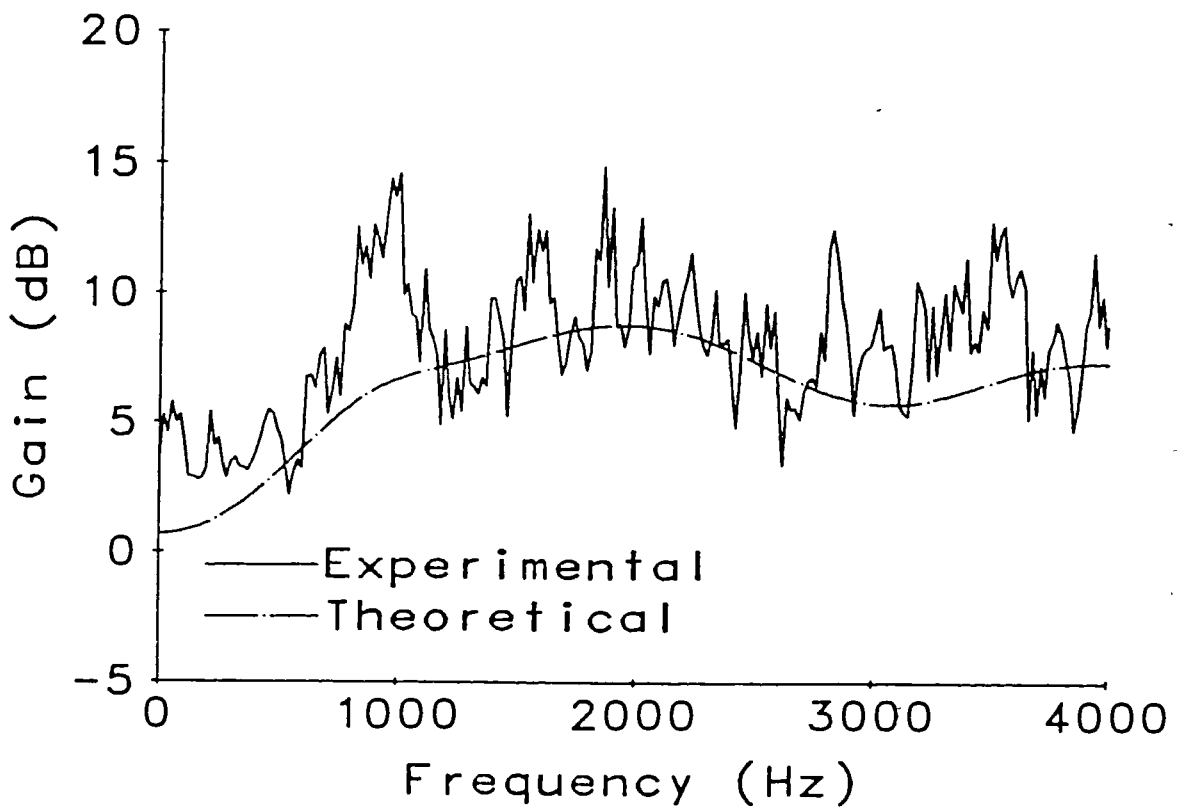
Figure 5.10 Array Gain at 100km/h With Fan Off and.Windows Closed

5.3. REVERBERATION REDUCTION PROERTIES OF THE ARRAY

If the reverberant speech does indeed become uncorrelated after passing through the $r_i(t)$, we expect the reverberant energy to reduce the same way the noise energy did. (Recall from chapter 2 that a reduction in reverberation results in a flatter frequency response.)

To investigate the potential reverberation reduction, we used the image model of chapter 2 to compute impulse responses for a single directional microphone and a semi-circular array. The cement tank described in chapter 2 was the first environment modelled. When the Fourier Transforms of the modelled responses were calculated, the spectral flatness (recall that this the variance of the frequency response calculated over the pass-band from 100 to 4000 Hz) was 29.1 for the array, and 35.1 for the single microphone, indicating a reduction in reverberation. Figures 5.11 and 5.12 show the modelled impulse responses, after convolution with the isolated impulses. The reverberation reduction effect of the array can be seen.

With the impulse source located at the focus of the array, the impulse response of a single microphone and the array were simultaneously recorded. The responses are shown in figures 5.13 and 5.14. Comparing these figures with the modelled responses (figures 5.11 and 5.12) the image model's ability to predict array performance is seen.

The responses in figures 5.13 and 5.14 have been normalized to have the same maximum amplitude after filtering. Due to losses (to be discussed in the next section), the peak amplitude of the impulse output by the array actually was 86% of that for the single microphone. This 0.66 dB loss is due to the fact that the impulse was not placed exactly equadistant from each of the microphones in the array.

Note the single microphone receives considerably more reverberant energy than the array. For this reason, the array gains cannot be calculated from the SNR values of speech embedded in noise. While the array reduces the noise content, it also reduces the reverberant speech energy. Since there is no way of separating the clean speech from the reverberant speech, SNR values calculated by this method introduce a bias against the array. (However, as explained in chapter 4, SNR isn't an accurate indicator of signal

enhancing performance when the noise is predominately lowpass.)

The results of the model for the array and single microphones in an automobile are shown in figures 5.15 and 5.16; figures 5.17 and 5.18 show the experimental responses. The modelled array had a spectral flatness of 5.6, improving on the single microphone's 24.7. The effects of reverberation in the automobile interior are severely reduced by the array.

In the vehicle, the array output an initial impulse at a level 92% that of the single microphone, representing a loss of 0.38dB.
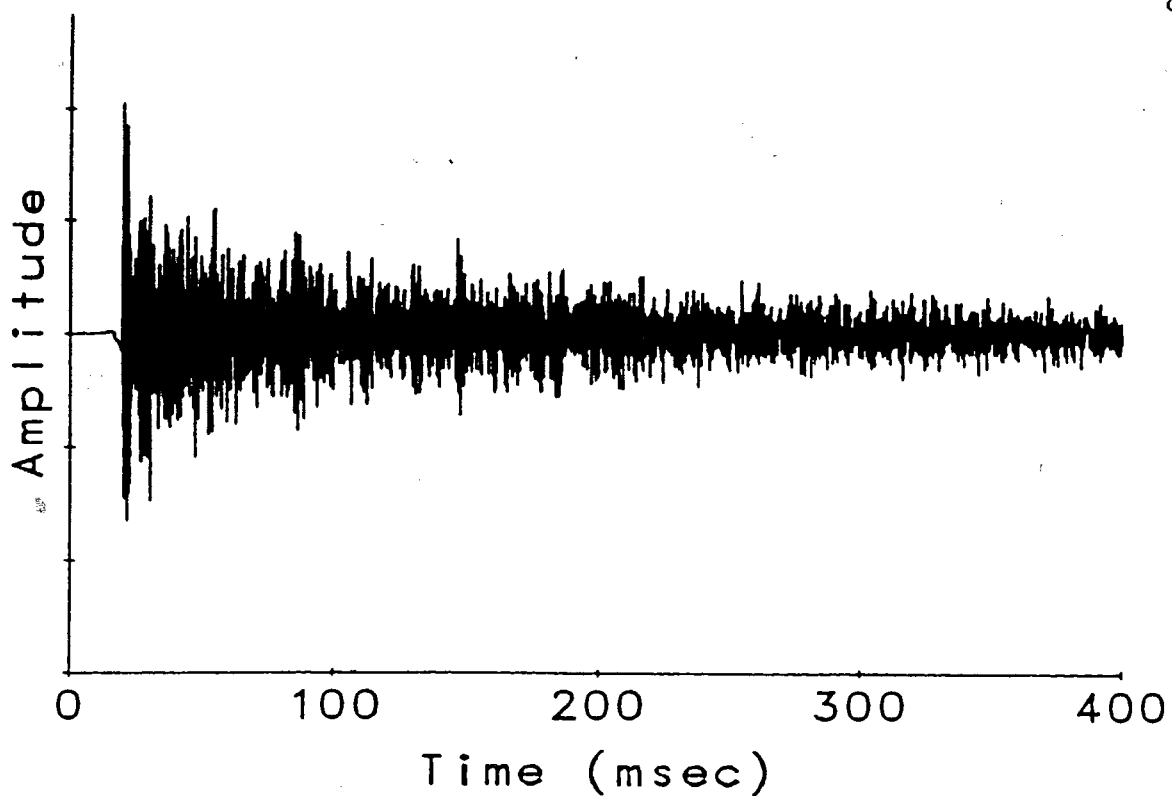
Figure 5.11 Modelled Impulse Response for Single Directional
Microphone in Cement Tank
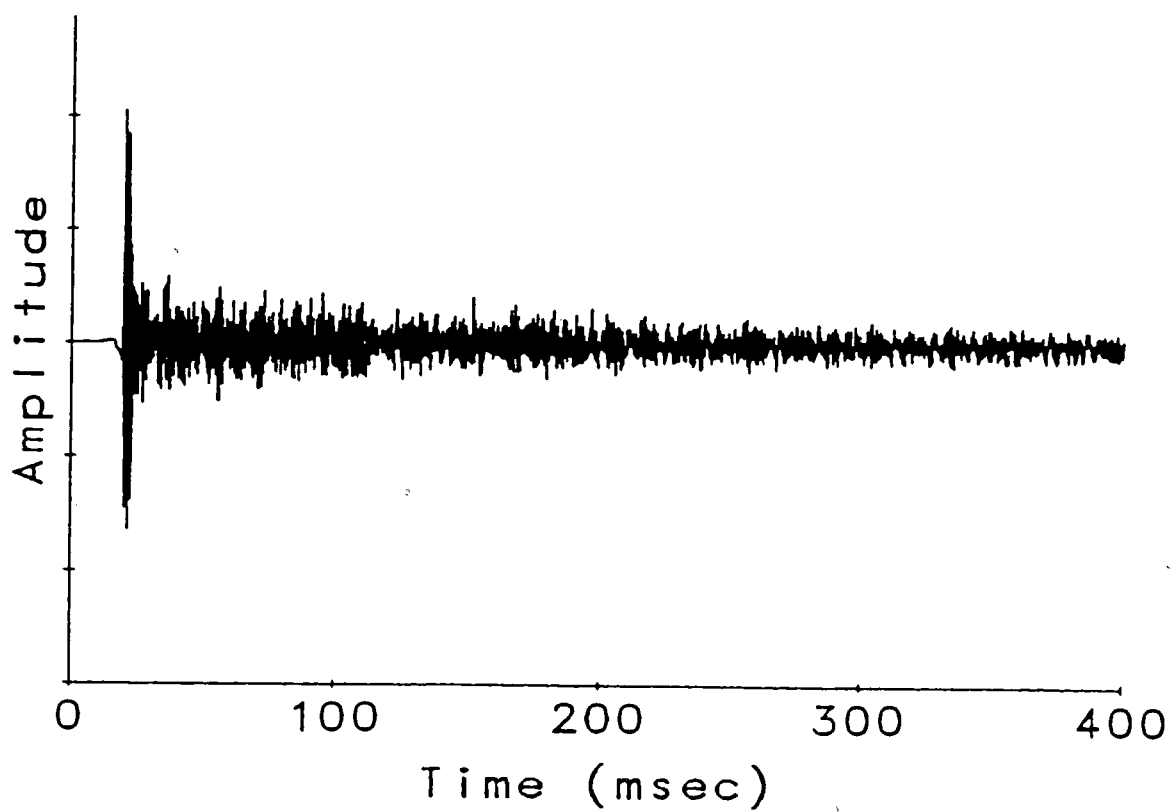
Figure 5.12 Modelled Impulse Response for Array of Directional Microphones in Cement Tank
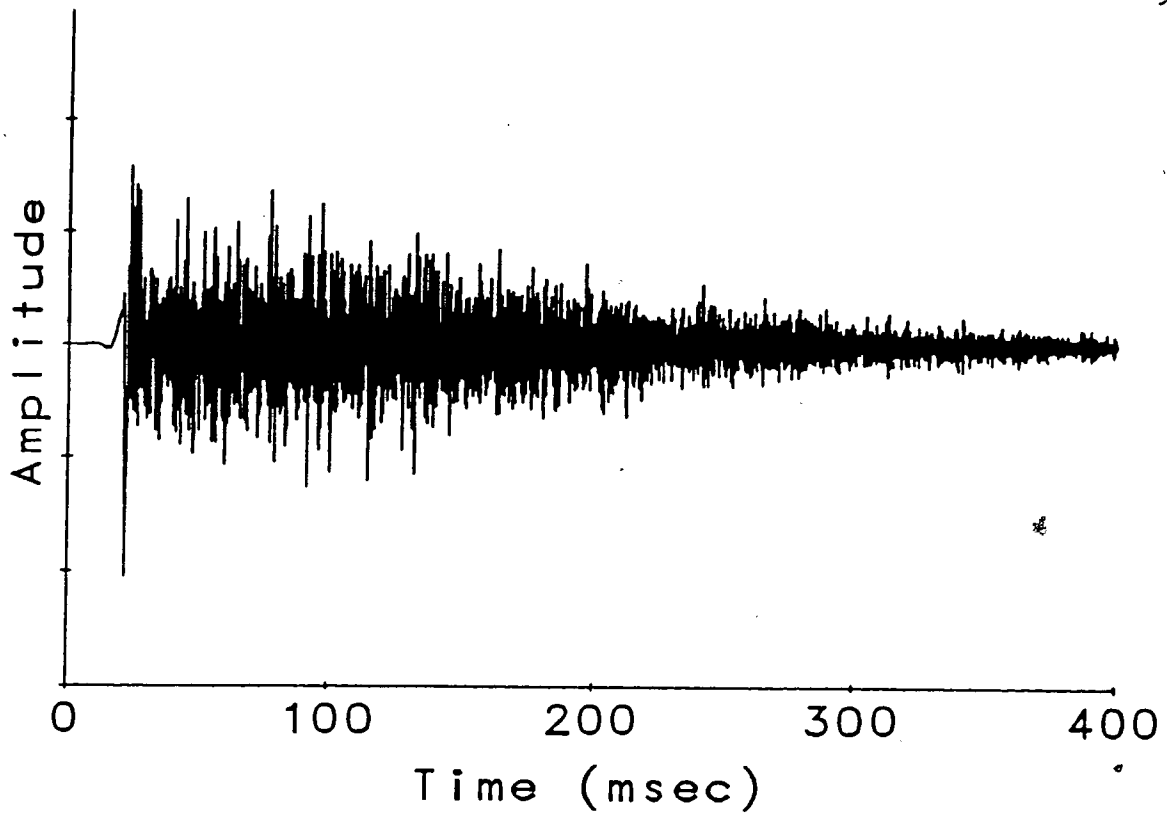
Figure 5.13 Actual Impulse Response for Single Directional
Microphone in Cement Tank
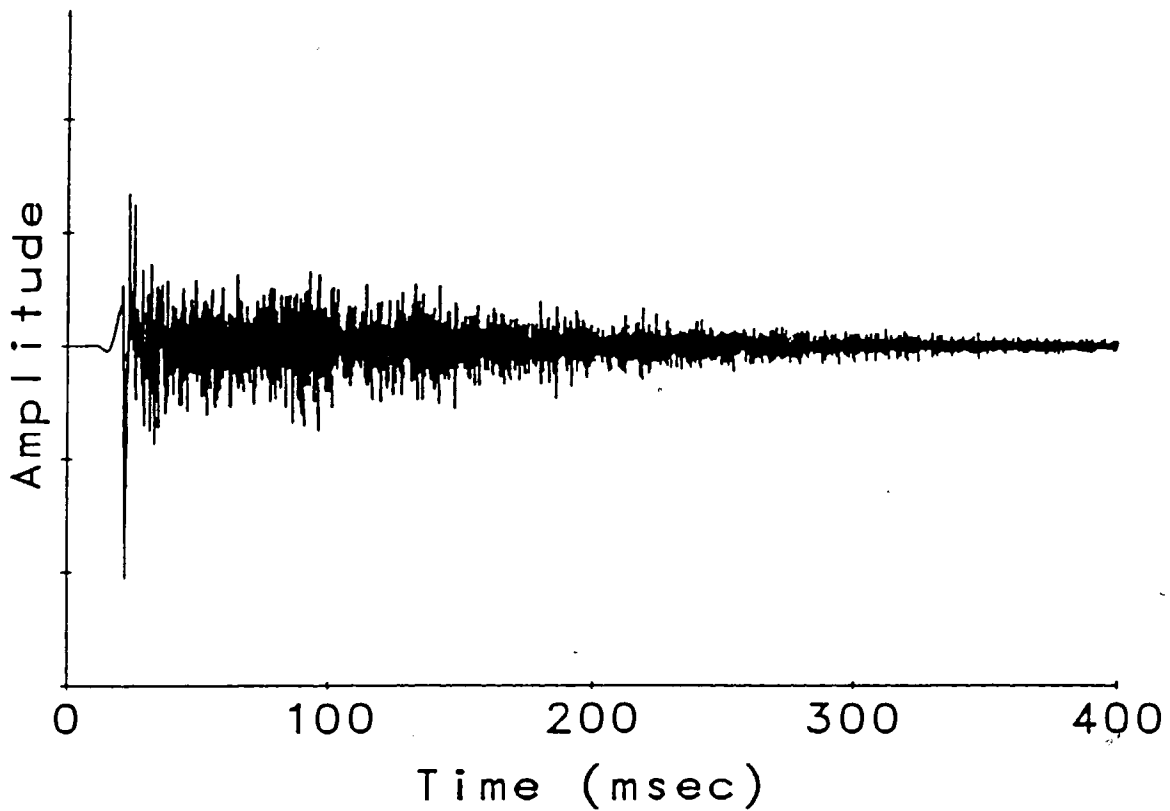
Figure 5.14 Actual Impulse Response for Array of Directional
Microphones in Cement Tank
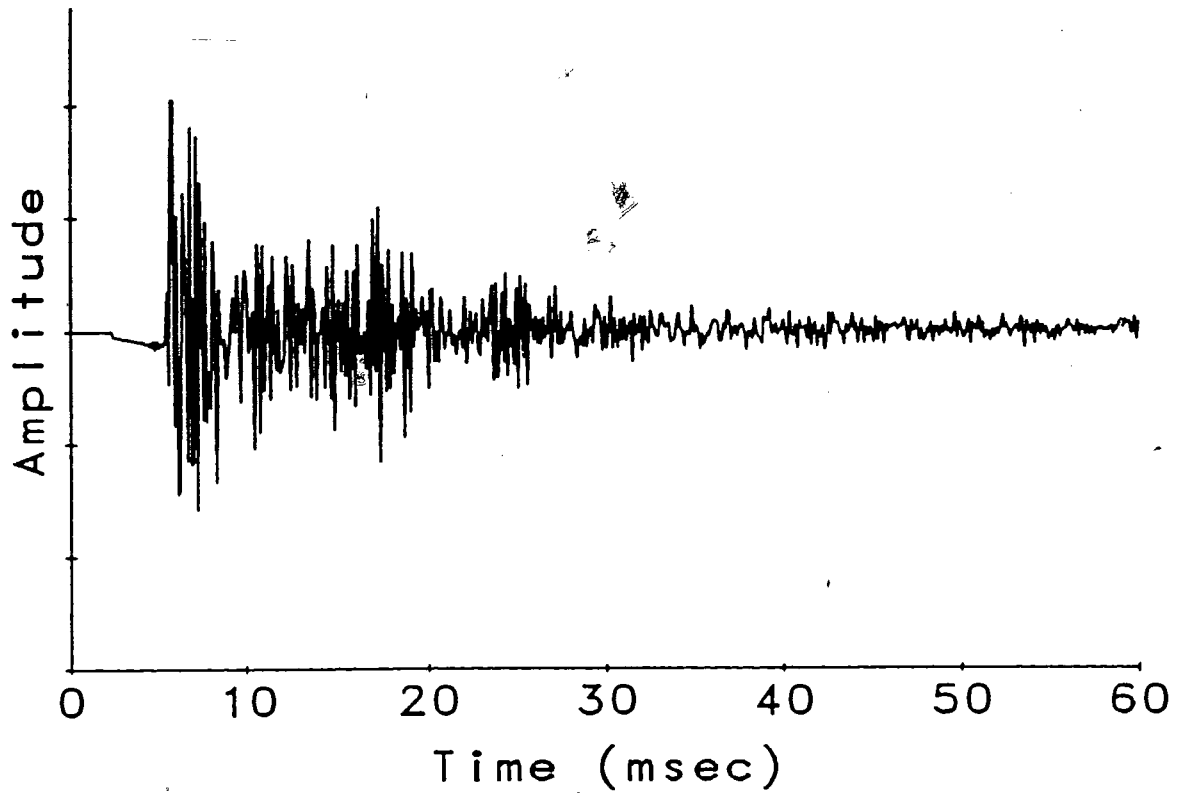
Figure 5.15 Modelled Impulse Response for Single Directional Microphone in Automobile
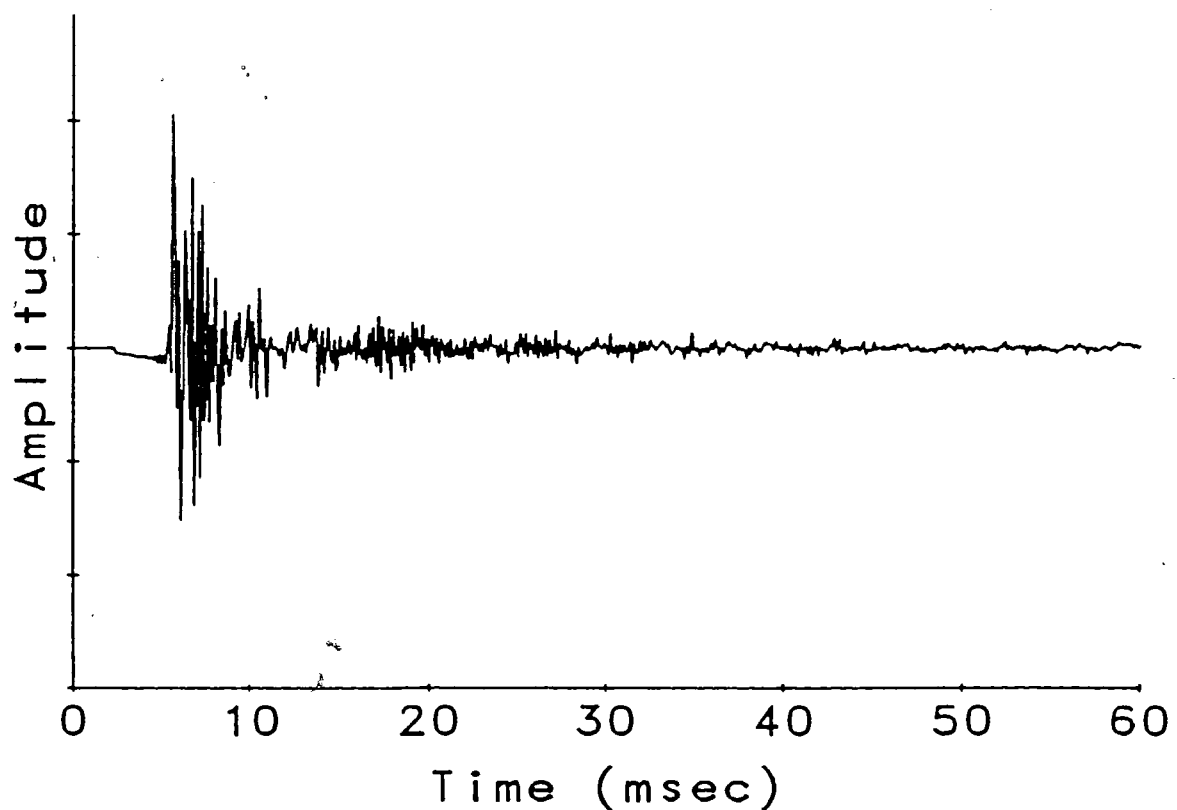
Figure 5.16 Modelled Impulse Response for Array of Directional Microphones in Automobile
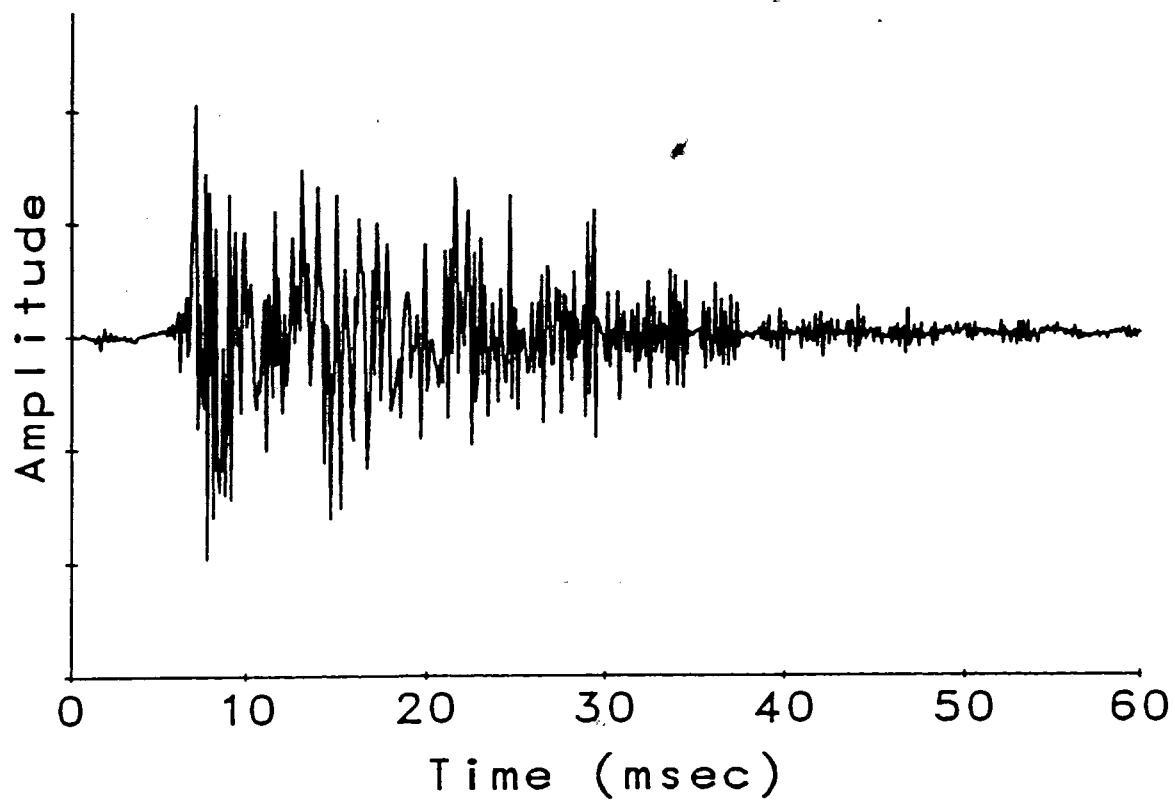
Figure 5.17 Actual Impulse Response for Single Directional Microphone in Automobile
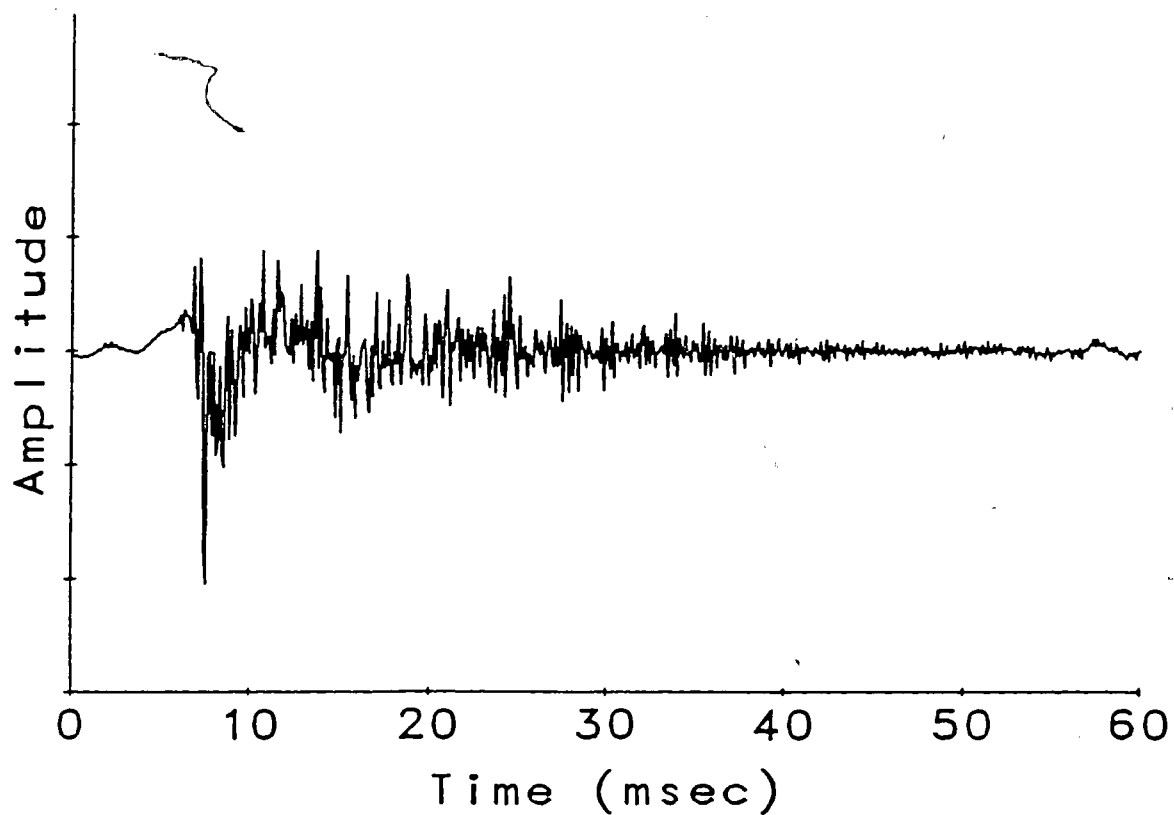
Figure 5.18 Actual Impulse Response for Array of Directional Microphones in Automobile

5.4. PRACTICAL ISSUES

The semi-circular array of figure 5.3 was chosen for simplicity. Because all the delays were equal, the "signal processing" for the array became a simple average, which could be implemented in analog form. The semi-circular array also has the advantage of pointing the directional microphones in different directions, so that the coherence between the received noise signals is reduced. Furthermore, it provides maximal separation between the microphones using the smallest area. The major design trade-off in the microphone constellation of the array lies in desiring the microphones to be close to the speaker (so that the SNR at the inputs is high) but wanting the microphones separated by as much distance as possible (to keep the coherence low and the gain of the array high). The number of microphones possible, and their maximal separation is dictated by the size of the vehicle interior.

It is not practical to mount an array of microphones in a semi-circle in an automobile (for this study, the microphones were mounted from a helmet that a passenger wore). A more feasible constellation of microphones is a linear array, with, say 15 cm separation between nearest neighbours. With this practicality comes a number of problems. First of all, the delays are no longer the same, so the array is no longer the simple average of the inputs.

Each of the τ_i of figure 5.1 must be estimated, and will have an associated estimation error of Δ_i. Consider equation (5.44), for the case of $\gamma_{ij}(\omega) = 1$ for all i and j (as for speech). Then the degradation of performance due to the Δ_i is given by

$$\text{Degradation}(\omega) = \frac{1}{\sum_{i=1}^{N} \alpha_i^2 + 2\sum\sum_{i<j} \alpha_i\alpha_j\cos(\omega(\Delta_i-\Delta_j))} \tag{5.47}$$

Since the τ_i are no longer equal, digital sampling and delaying is required. However, the τ_i will not, in general, be integral multiples of the sampling period. Therefore, the optimal filter $h_i(t)$ are no longer simple delays, but rather interpolators.

The best way to determine these optimum interpolating filters is not obvious. If one assumes that the speaker remains in the same position at all times, physical measurements of the distances between his mouth and the microphones could be made, and the optimum filters designed. Installation of such a system would be extremely difficult for the consumer or salesperson.

A more practical solution is to "train" the array by having the user say a sentence when the car is silent. While cross-correlation methods[46] could be used to estimate the delays between the the inputs, and these estimates could be used to design the interpolaters, the broad-band and quasi-periodic characteristics of speech make exact determination of the delays difficult. Figure 5.19 is an example of the cross-correlation between two microphones in an automobile recording the author saying the phonetically balanced sentence, "Cats and dogs each hate the other." Note the delay appears to be approximately one sample, but determining the value to sub-sample accuracy isn't a simple task.[46]

For an alternative approach, consider figure 5.20, where speech makes up the two inputs of a classic ANC filter. To minimize the output power, the adaptive filter $\hat{H}$ will become the optimum interpolator. (Note that the interpolater will not have a flat frequency response; it will introduce some shaping, so the training sentences must be representative of all speech).

When this technique was applied to the data of figure 5.19, the filter converged to the weights displayed in figure 5.21. Note that the filter places maximal weight for sample 1, corresponding to the delay of 1 sample from figure 5.19. Only a very short filter is required for each of the microphone inputs of the array.

Further work remains to study the effect of such an approach on speech. It is the author's feeling that the best performance will result from a non-adaptive system, with the interpolators determined and fixed first of all, then the speech recognition system trained with the interpolators in place. In this manner, any distortion introduced by the interpolators is consistent in both training and recognizing.
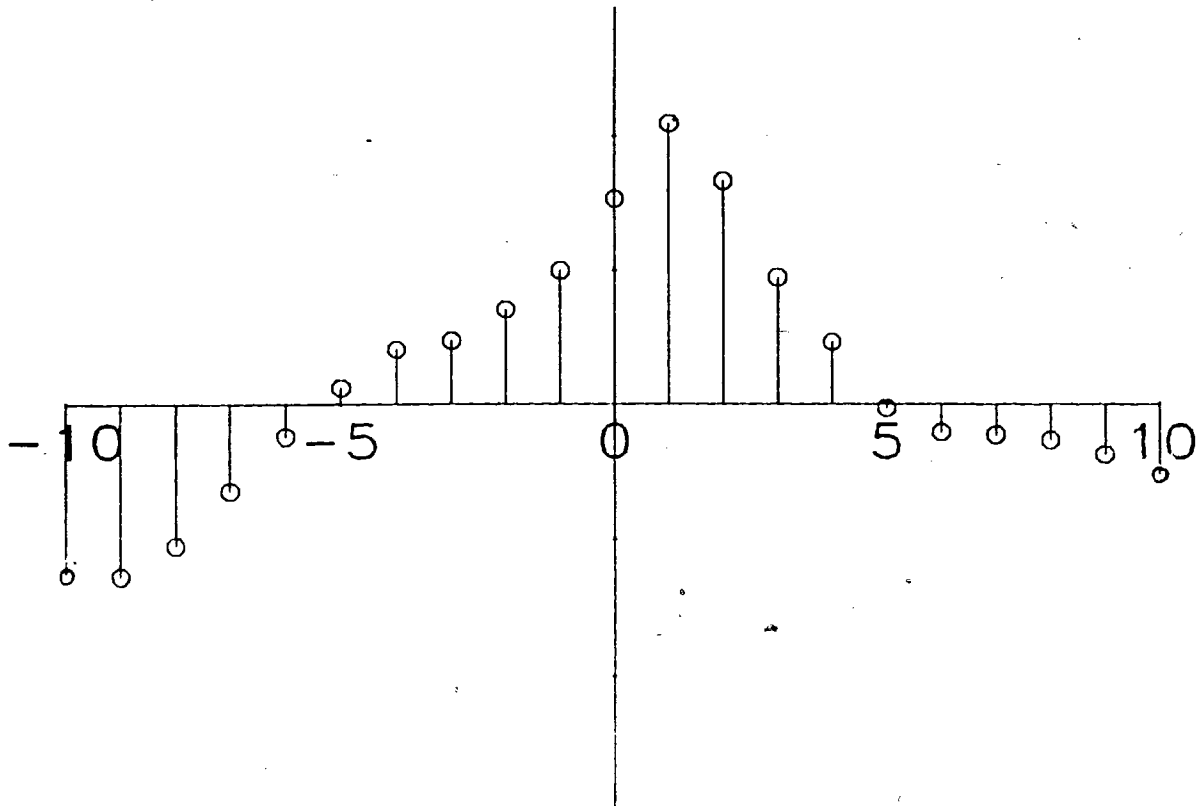
Figure 5.19 Cross-Correlation Between Two Microphones With Speech Recorded in Quiet Automobile

5.5. CONCLUSIONS

The delay-equalized array performs as predicted, with noise gains approximately equal to the number of elements. When coupled with high-pass filtering, the array produced discernibly cleaner speech, which should in turn produce better speech recognition results.[3] While some work remains for a practical implementation, the array holds significant potential for enhancing speech without introducing distortion. The array has very low complexity, requiring only a microphone, sampler and a very short interpolating filter for each element. If the interpolators are determined and fixed, no adaptation is required.
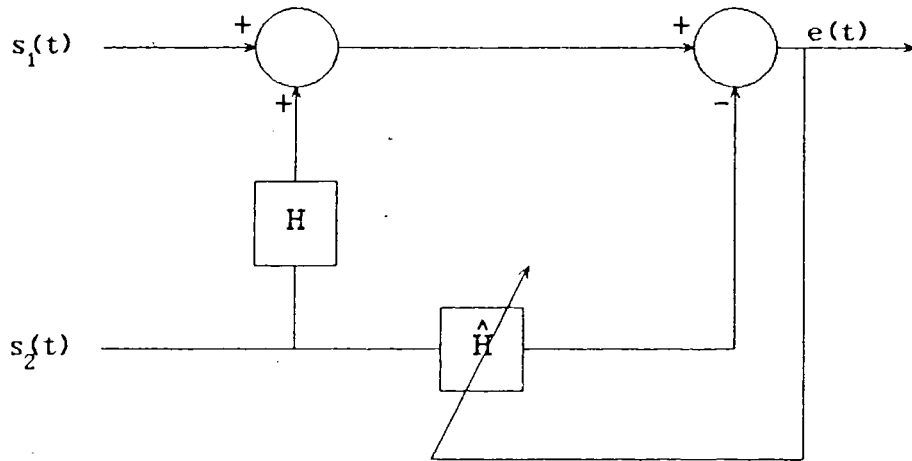
Figure 5.20 Using ANC Techniques to Calculate Optimal Interpolator
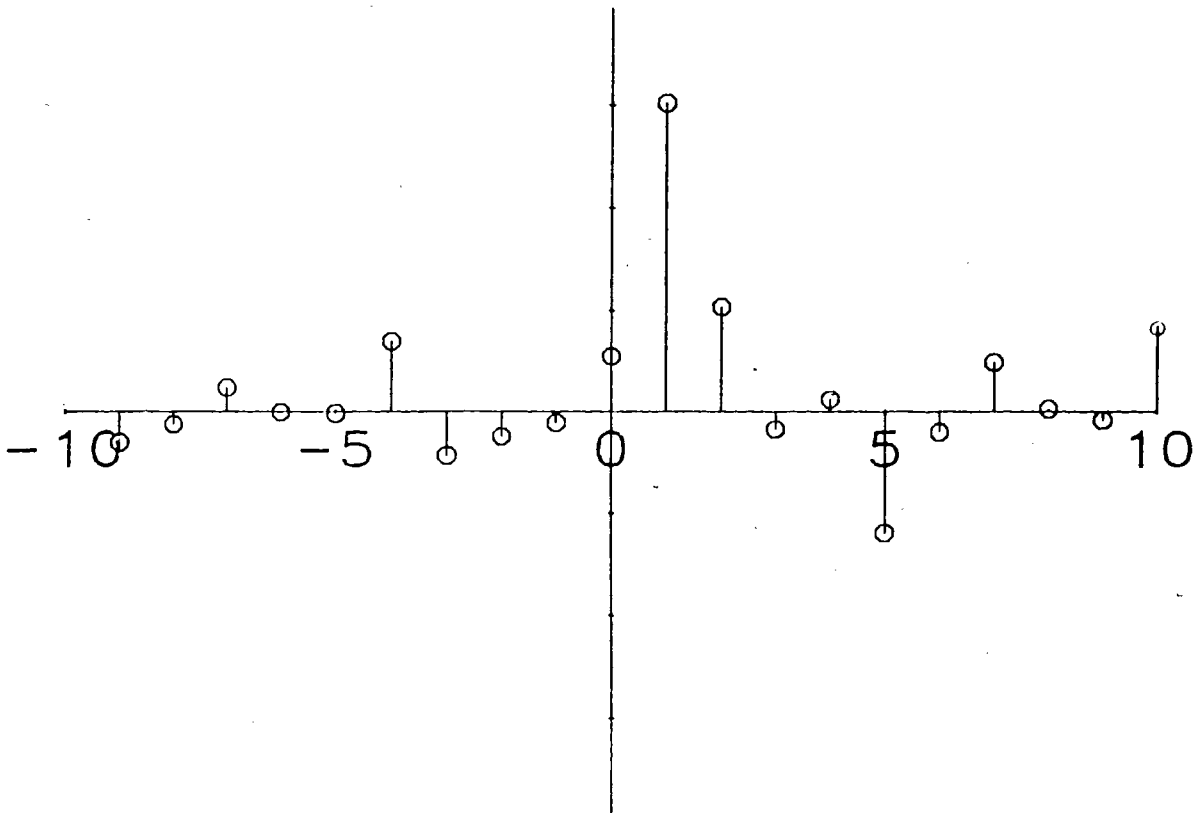
Figure 5.21 ANC-Derived Interpolator from figure 5.19

6. CONCLUSIONS

The problem of speech enhancement for mobile telephony is important, because current systems are unsafe. A voice-activated phone which responds to verbal commands will provide the ultimate in safety but is currently unfeasible due to the inability of speech recognizers to cope with the effects of the acoustical environment on the speech. The reverberation and noise in an automobile call for low-distortion, high-performance speech enhancement algorithms. As a side benefit, these techniques will also provide better voice quality for transmission than current "hands-free" systems.

Other studies involving ANC for speech enhancement in a vehicle have been attempted, but their results were inconclusive. Rather than attempting numerous speech enhancement techniques, we chose to concentrate on the principles governing the acoustical field within the automobile interior. By considering the acoustical environment, we have been able to gain understanding into why certain methods will or won't be effective in a vehicle.

The reverberant nature of the vehicle interior affects both the speech and noise, distorting the speech and causing the noise field to become more diffuse. The exact effects of reverberation are very sensitive to changes in the positions of the speaker, microphones and the room structure. We have successfully developed an image model for reverberation, not to exactly simulate the acoustical field, but rather to allow us to study a related environment. The model not only allowed us to gain intuition regarding the nature of reverberation, but also enabled us to make predictions regarding the performance of arrays in reverberant enclosures. A very simple method of introducing directionality into the microphones of the model has also been developed.

A careful study of the properties of the noise field in an automobile has been undertaken. The noise field is predominately low-pass, with much of the energy susceptible to high-pass filtering, below the frequency range of speech. The state of the vehicle can vastly change the noise power spectral density, with a dynamic range of some 30 dB between the case of a stationary car and one travelling at highway speed.

The coherence of the noise field was studied, and seen to be low, approaching that for a diffuse field. We also developed a general equation for the coherence between two cardioid microphones in a diffuse noise field.

We have been able to show why Adaptive Noise Cancellation is ineffective in a vehicular environment. The two requirements for ANC, coherence between the two noise inputs, and speech isolation (speech present at only one of the inputs) cannot be simultaneously satisfied in an automobile. The SNR gains reported by others are deceptive, representing the low-frequency cancellation since the noise spectral density is low-pass in nature. A high-pass filter would achieve the same cancellation without adding noise at higher frequencies.

A new approach to speech enhancement for the vehicular environment, based on a new model for the acoustical noise field, has been presented. Subsequent analysis of the array's theoretical noise reduction performance shows that it requires low coherence to be effective. The experimental performance is consistent with theoretical predictions for the array in a diffuse noise field.

Since reverberation is highly dependent on position, the reverberant tails received by the various microphones of the array become uncorrelated. Therefore, the array is able to reduce the reverberant energy in the same manner as it reduces the noise energy. The reverberation reducing properties of the array was successfully predicted by the image model.

The delay-equalized array holds great potential for speech enhancement for mobile telephony, but a couple of implementation details remain to be studied. The problems left for future study are how best to estimate the delays between the speaker and microphones, and then how to use these estimates to derive the optimal interpolators. A related question is whether or not the interpolators should be allowed to adapt. It is clear that these interpolators need only be very short filters so that the complexity of a practical array is expected to be quite low.

When these implementation details are resolved, the array will be provide robust, low-distortion enhancement at a very low complexity. When coupled with high-pass

filtering, the array will provide considerable improvement in speech recognition performance.

APPENDIX 1: DIFFUSE FIELD COHERENCE
FOR DIRECTIONAL MICROPHONES

In this appendix, the coherence between the signals received by two cardioid microphones in a diffuse noise field is derived. From equation 3.35, the coherence between two directional microphones with directional patterns $m_x(\theta,\phi)$ and $m_y(\theta,\phi)$ is

$$\gamma_{xy}(\omega) = \frac{E\left\{m_x(\theta,\phi)m_y(\theta,\phi)e^{-jkd\cos\theta}\right\}}{\left[E\{m_x(\theta,\phi)^2\}E\{m_y(\theta,\phi)^2\}\right]^{\frac{1}{2}}} \equiv \frac{N(\omega)}{D(\omega)} \tag{A1.1}$$

In general, for a cardioid pattern with axis along $\vec{r}_1 = (x_1,y_1,z_1)$, the response to a wave arriving from direction $\vec{r}_0 = (x_0,y_0,z_0)$, is given by equation 2.1.

$$\text{Gain}(\psi) = \frac{1}{2}(1 + \cos\psi) \tag{A1.2}$$

where ψ is the angle between $\vec{r}_0$ and $\vec{r}_1$. (see figure A1.1)
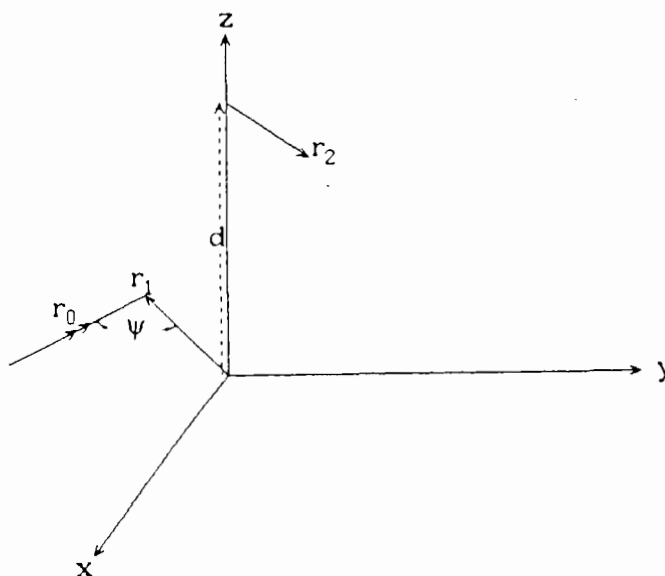
Figure A1.1 Geometry for Deriving Coherence Between Directional Mics

But from linear algebra,[22] the cosine of the angle between two vectors $\vec{x}_0$ and $\vec{x}_1$ in space is given by

$$\cos\psi = \frac{\vec{x_0}\cdot\vec{x_1}}{|\vec{x_0}||\vec{x_1}|} \tag{A1.3}$$

where the dot (·) denotes the vector inner product. If $\vec{r_0}$ is a unit vector, its representation in spherical co-ordinates is given by $\vec{r_0} = (\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)$ then the microphone gain for microphone with axes along $\vec{r_1}$ to a plane wave arriving from direction r_0 is

$$\text{Gain}(\theta,\phi) = \tfrac{1}{2}(1 + x_1\sin\theta\cos\phi + y_1\sin\theta\sin\phi + z_1\cos\theta) \tag{A1.4}$$

The numerator of equation A1.1 is then

$$N(\omega) = \frac{1}{4}E\left\{ \left[1 + x_1\sin\theta\cos\phi + y_1\sin\theta\sin\phi + z_1\cos\theta \right] \right.$$
$$\left. \left[1 + x_2\sin\theta\cos\phi + y_2\sin\theta\sin\phi + z_2\cos\theta \right] e^{-jkd\cos\theta} \right\} \tag{A1.5}$$

Averaging over all angles of arrival

$$= \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\frac{1}{4}\left[1 + x_1\sin\theta\cos\phi + y_1\sin\theta\sin\phi + z_1\cos\theta \right]$$
$$\left[1 + x_2\sin\theta\cos\phi + y_2\sin\theta\sin\phi + z_2\cos\theta \right] e^{-jkd\cos\theta}\sin\theta \, d\theta \, d\phi \tag{A1.6}$$

Performing the integration over ϕ first, and remembering that

$$\int_0^{2\pi}\cos\phi \, d\phi = \int_0^{2\pi}\sin\phi \, d\phi = \int_0^{2\pi}\sin\phi\cos\phi \, d\phi = 0 \tag{A1.7}$$

and that

$$\int_0^{2\pi}\cos^2\phi \, d\phi = \int_0^{2\pi}\sin^2\phi \, d\phi = \pi \tag{A1.8}$$

the surviving terms from equation A1.5 are

$$N(\omega)=\frac{1}{16}\int_0^\pi \left[2 + 2(z_1+z_2)\cos\theta + (x_1x_2+y_1y_2)\sin^2\theta + 2z_1z_2\cos^2\theta \right] e^{-jkd\cos\theta}\sin\theta\,d\theta \quad (A1.9)$$

From equations 3.25 - 3.31,

$$\int_0^\pi e^{-jkd\cos\theta}\sin\theta\,d\theta = 2\frac{\sin kd}{kd} \quad (A1.10)$$

Now, the remaining integrals of equation A1.9 are all evaulated by using the same substitution, namely

$$u = kd\cos\theta \text{ so that } du = -kd\sin\theta\,d\theta \quad (A1.11)$$

Under this substitution,

$$\int_0^\pi \cos\theta\sin\theta e^{-jkd\cos\theta}d\theta = \frac{-1}{(kd)^2}\int_{kd}^{-kd} ue^{-ju}du$$

$$= \frac{-1}{(kd)^2}\left[-jkde^{jkd} - jkde^{-jkd} + e^{-jkd} - e^{-jkd} \right]$$

$$= 2j\left[\frac{\cos kd}{kd} + \frac{\sin kd}{(kd)^2} \right] \quad (A1.12)$$

And

$$\int_0^\pi \cos^2\theta\sin\theta e^{-jkd\cos\theta}d\theta = \frac{-1}{kd}\int_{kd}^{-kd} \frac{u^2}{(kd)^2}e^{-ju}du$$

$$= \frac{1}{(kd)^3}\left[-j(kd)^2\left(e^{jkd}-e^{-jkd}\right) + +2kd\left(e^{jkd}+e^{-jkd}\right) + 2j\left(e^{jkd}-e^{-jkd}\right) \right]$$

$$= 2\frac{\sin kd}{kd} + 4\frac{\cos kd}{(kd)^2} - 4\frac{\sin kd}{(kd)^3} \quad (A1.13)$$

Now

$$\int_0^\pi \sin^3\theta e^{-jkd\cos\theta}d\theta = \int_0^\pi (1 - \cos^2\theta)\sin\theta e^{-jkd\cos\theta}d\theta$$

$$= \int_0^{\pi} \sin\theta \, e^{-jkd\cos\theta} \, d\theta - \int_0^{\pi} \cos^2\theta \, \sin\theta \, e^{-jkd\cos\theta} \, d\theta$$

$$= 4\frac{\sin kd}{(kd)^3} - 4\frac{\cos kd}{(kd)^2} \tag{A1.14}$$

from the results in equations A1.10 and A1.13.

To calculate $D(\omega)$, we assume without loss of generality that the axis of the microphone is along the z-axis. Then

$$E\{m_x^2(\theta,\phi)\} = E\{m_y^2(\theta,\phi)\}$$

$$= \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} \frac{1}{4}(1 + \cos\theta)^2 \sin\theta \, d\theta \, d\phi$$

$$= \frac{1}{3} \tag{A1.15}$$

from equation 3.8. Then, combining equations A1.1,A1.10,A1.12-A1.15, we have the result

$$\gamma_{xy}(\omega) = \frac{3}{4}\left[\frac{\sin kd}{kd} + \left(x_1 x_2 + y_1 y_2\right)\left(\frac{\sin(kd)}{(kd)^3} - \frac{\cos(kd)}{(kd)^2} \right) + z_1 z_2 \left(\frac{\sin(kd)}{kd} + \frac{2\cos(kd)}{(kd)^2} - \frac{2\sin(kd)}{(kd)^3} \right) \right.$$

$$\left. +j\left(z_1 + z_2\right)\left(\frac{\cos(kd)}{kd} - \frac{\sin(kd)}{(kd)^2} \right) \right] \tag{A1.16}$$

APPENDIX 2: MAPPING TO GET d ALONG THE Z-AXIS

Consider diagram A2.1, where we have plotted the difference between the position of two microphones and labelled it d. Microphone 1 is at position $\vec{p}_1 = (x_{p_1}, y_{p_1}, z_{p_1})$ and has its axis of pointing along $\vec{a}_1 = (x_{a_1}, y_{a_1}, z_{a_1})$, with similar definitions for the position and pointing direction of microphone 2.
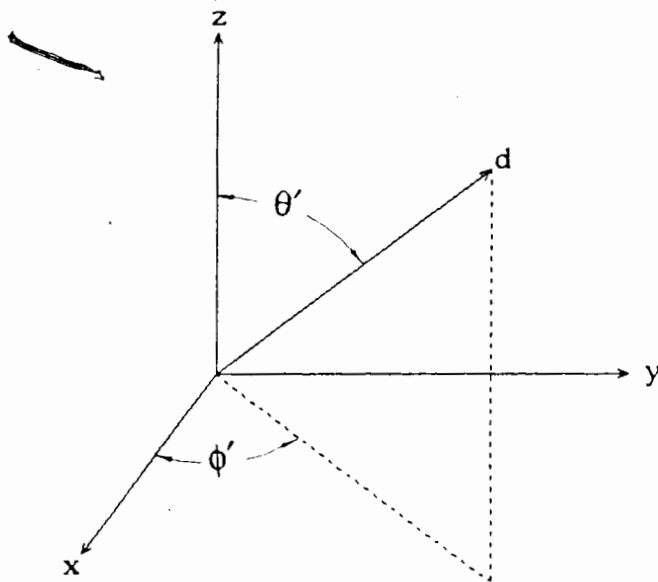
Figure A2.1 Arbitrary d Vector

In order for (A1.16) to be valid, $\vec{d} = \vec{p}_1 - \vec{p}_2$ must lie along the z axis. We begin by normalizing $\vec{d}$ so it is a unit vector, in direction (ϕ', θ') where

$$\phi' = \tan^{-1}\left(\frac{y_{p_1} - y_{p_2}}{x_{p_1} - x_{p_2}}\right) \tag{A2.1}$$

and

$$\theta' = \cos^{-1}\left(\frac{z_{p_1} - z_{p_2}}{|\vec{p}_1 - \vec{p}_2|}\right) \tag{A2.2}$$

Then, to transform $\vec{d}$ we must rotate through ϕ' about the z axis and then through θ' about the y axis. The transformation to rotate $\vec{d}$ through ϕ' around the z axis is[22]

$$T_1 = \begin{bmatrix} \cos\phi' & -\sin\phi' & 0 \\ \sin\phi' & \cos\phi' & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A2.3}$$

and the transformation to rotate through θ' around the y axis is

$$T_2 = \begin{bmatrix} \cos\theta' & 0 & \sin\theta' \\ 0 & 1 & 0 \\ -\sin\theta' & 0 & \cos\theta' \end{bmatrix} \tag{A2.4}$$

The transformations are applied one after the other so that the complete transformation is

$$T = T_2 T_1 = \begin{bmatrix} \cos\theta'\cos\phi' & -\cos\theta'\sin\phi' & \sin\theta' \\ \sin\phi' & \cos\phi' & 0 \\ -\sin\theta'\cos\phi' & \sin\theta'\sin\phi' & \cos\theta' \end{bmatrix} \tag{A2.5}$$

Thus to apply (A1.16), one first finds θ' and ϕ', then applies T to the vectors $\vec{a}_i$. The resulting pointing directions $\vec{a}_i'$ are normalized, and (A1.16) can be used.

APPENDIX 3: FORTRAN LISTING FOR REVERBERATION MODEL

```fortran
      program simsroom
c
c programmed by Martie Meryle Goulding
c copyright  Simon Fraser University
c
c program to call sroom (i.e. simulator of room to give impulse
c response)
c
c Adapted from 'Image method for efficiently simulating
c small room acoustics' by Allen and Berkley
c see subroutine sroom.f for further documentation
c
c r = array of vector radius to receivers
c rn = vector radius to receiver for single receiver
c ro = vector radius to source
c rL = vector of box dimensions
c ************************************************************
c         NOTE:
c all of the above are in sample periods
c to convert from physical distance to sample periods:
c  sample period = length*fs/c
c  where c = speed of sound   331 m/s
c       fs = sampling frequency
c ************************************************************
c
c beta = vector of six wall reflection coefs (0<= beta <= 1)
c beta(i,j) = coefficient for wall in direction i
c j = 1 .. x direction         i = 1 .. adjacent to origin
c    2 .. y                2 .. opposite to origin
c    3 .. z
c
      real r(3,10),rn(3),ro(3),rL(3),beta(2,3),ht(5000),out(5000)
      real pi,wt(10),fs,fc,b
      real fo(3,10),delay(10),micaxis(3),absmicaxis
      integer npts,i,j,point,mictype(10),nummics,range,offmax
      character*32 filnam
c
      pi = 3.14159265359
c
c ****************************************************************
c      READ IN PARAMETERS FOR SIMULATION
c
c
      print *,'Number of points to be computed:'
      read (5,101)npts
      print *,'Sampling frequency, Cutoff frequency (kHz):'
      read (5,102)fs,fc
```

```
c read in coefficients for walls
      print *,'Reflection coefficients (0<b<1) '
      print *,'Coefficients for walls:'
      read *,b
      do 20 j=1,3
        do 20 i=1,2
          beta(i,j)=b
 20   continue
      print *,'Coefficient for floor:'
      read *,beta(2,3)
      print *,'Coefficient for ceiling:'
      read *,beta(1,3)
      print *,'Output coefficients to :'
      read 103,filnam
      open(unit=1,file=filnam)
c  check Nyquist criterion satisfied
      if (fs .lt. 2*fc) then
        print *,'Error - aliasing will occur'
        goto 999
      endif
      print *,'Input room dimensions (meters):'
      read (5,100)rL(1),rL(2),rL(3)
      print *,'Source Position:'
      read (5,100)ro(1),ro(2),ro(3)
      print *,'Number of microphones in array (<=10):'
      read (5,101)nummics
      offmax = 0
c convert to sample periods
      do 25 j = 1,3
        rL(j) = rL(j)*fs/.331
        ro(j) = ro(j)*fs/.331
 25   continue
c read in weightings and directional characteristics for microphones
      do 60 i=1,nummics
        print *,'For microphone #',i
        print *,'Weighting:'
        read(5,104)wt(i)
 30     continue
        print *,'Type of mic response (0=omni,1=cardioid):'
        read (5,101)mictype(i)
        if ((mictype(i) .ne. 0) .and. (mictype(i) .ne. 1)) goto 30
        print *,'Receiver Position:'
        read (5,100)r(1,i),r(2,i),r(3,i)
 40     continue
c for directional mics, find pointing direction
        if (mictype(i) .ne. 0) then
          print *,'Point receiver at source? (0=yes,1=no)'
          read(5,101) point
          if (point .eq. 1) then
            print *,'Point mic towards focus:'
```

```
         read(5,100)fo(1,i),fo(2,i),fo(3,i)
         if ((fo(1,i).eq.r(1,i)).and.(fo(2,i).eq.r(2,i))
   +                  .and.(fo(3,i).eq.r(3,i))) then
          print *,'error: receiver cannot point at itself!!!!'
          goto 40
         else
           do 45 j = 1,3
           fo(j,i) = fo(j,i)*fs/.331
45         continue
         endif
c point at source
       else if (point .eq. 0) then
          do 50 j=1,3
          fo(j,i) = ro(j)
50        continue
       else
          goto 40
       endif
c omnimics use source as focus to delay to
       else
         do 55 j=1,3
         fo(j,i) = ro(j)
55       continue
       endif
       delay(i) = 0.
c calculate delay from mic to source in sample periods
       do 56 j = 1,3
        r(j,i) = r(j,i)*fs/.331
        delay(i) = delay(i) + (fo(j,i) - r(j,i))**2
56     continue
       delay(i) = sqrt(delay(i))
       offmax = nint(max(offmax,delay(i)))
60     continue
c calculate maximum range of possible response and zero output array
     range=npts+offmax+41
     do 66 i=1,range
     out(i) = 0.
66   continue
c ***************************************************************
c                MAIN LOOP
c
    do 90 i = 1,nummics
c copy mic vector into one dimensional array
      do 70 j = 1,3
      rn(j) = r(j,i)
70    continue
c if directional type microphone desired then
c calculate axis of mic and normalize
c
      if (mictype(i) .ne. 0) then
```

```fortran
          absmicaxis = 0.
          do 75 j = 1,3
            micaxis(j) = fo(j,i) - r(j,i)
            absmicaxis = absmicaxis + micaxis(j)**2
75        continue
          absmicaxis = sqrt(absmicaxis)
c normalize
          do 78 j = 1,3
            micaxis(j) = micaxis(j)/absmicaxis
78        continue
        endif
c
c call subroutine to calculate response for single mic
c
      call sroom(rn,ro,rL,beta,ht,range,fs,fc,mictype(i)
     +              ,micaxis,delay(i),offmax)
c
c weight response by mic weight factor and add into array output
c
      do 80 j = 1,range
        out(j) = ht(j)*wt(i) + out(j)
80      continue
90    continue
c
c normalize for number of mics and output results
c
      i = 0
c don't output zeroes at beginning
91    continue
      i = i+1
      if (out(i) .ne. 0) goto 92
      goto 91
92    continue
      do 95 j = i,i+npts
        write(1,*)out(j)/nummics
95    continue
c
c
100   format(3f7)
101   format(i5)
102   format(2f7)
103   format(a)
104   format(f7)
999   continue
      end
```

```
c
c ********************************************************
c   main subroutine sroom
c ********************************************************
c
        subroutine sroom(r,r0,rL,beta,ht,range,fs,fc,mictype,
     +                   micaxis,offset,offmax)
c
c 1988 01 19
c subroutine to calculate a room impulse response
c adapted from 'Image method for efficiently  simulating
c small-room acoustics' by Alllen and Berkley
c
c********************************************************
c update 1:
c this version introduces a sinc-type filter (after 'Simulating
c the response of multiple microphones to a single acoustic source
c in a reverberant room' by Peterson) to give the accurate time
c of arrival of the image impulses
c
c ********************************************************
c update 2:
c this version introduces directionality to simulate the direction
c of arrival of each pulse, and the direction-dependency of the
c mics, modeled by a given function
c
c ********************************************************
c update 3:
c now, the responses are delayed and added coherently so
c that the focus point is re-inforced
c
c ********************************************************
c update 4: (1988 07 11)
c program revamped to include directionality by simple dot product
c method (see notes for 1988 07 10)
c
c
c r= vector radius to receiver in sample periods = length/(c*t)
c r0=vector radius to source in sample periods
c rL = vector of box dimensions in sample periods
c beta = vector of six wall reflection coefs (0<=beta<=1)
c beta(i,j) = coefficient for wall in direction i
c j = 1 .. x     i = 1 .. adjacent to origin
c    2 .. y         2 .. opposite to origin
c    3 .. z
c ht = calculated impulse response array
c zero delay is in ht(1)
c mic is aligned along z axis by mapping with matrix A
c fo = vector of focal point
c micaxis = unit vector in direction of micaxis
```

```
c
c
      real ht(range),dis,fs,fc,gid,off,micaxis(3)
        real r(3),r0(3),rL(3),delp(8),beta(2,3),array(41)
      real x(8),y(8),z(8),micresponse,mic,offset
      integer offmax
        integer nr(3),nx,ny,nz,n1,n2,n3,range
      integer i,j,k,m,mm,mictype
        equivalence (nr(1),nx),(nr(2),ny),(nr(3),nz)
c
      do 5 i=1,range
      ht(i)=0.
5     continue
      dis=0.
c compute distance from mic to source and check for zero
      do 6 i=1,3
        dis=(r(i)-r0(i))**2+dis
6     continue
      if(dis.lt. 1e-2) then
      ht(1) = 1
      return
      endif
c calculated range of sum
      n1=range/(rL(1)*2)+1
      n2=range/(rL(2)*2)+1
      n3=range/(rL(3)*2)+1
      do 20 nx=-n1,n1
    do 20 ny=-n2,n2
      do 20 nz=-n3,n3
c get eight image locations for mode # nr
        call lthim(r,r0,rL,nr,delp,x,y,z)
        i0=0
        do 10 l=0,1
        do 10 j=0,1
        do 10 k=0,1
        i0=i0+1
c check for delay less than range
        if (delp(i0).le.range) then
        off=delp(i0)-float(ifix(delp(i0)))
c low-pass filter the signal to get accurate time arrival
          call banlim(off,fc,fs,array)
c put in loss factor once for each wall reflection
          gid=beta(1,1)**iabs(nx-l)
    +       *beta(2,1)**iabs(nx)
    +       *beta(1,2)**iabs(ny-j)
    +       *beta(2,2)**iabs(ny)
    +       *beta(1,3)**iabs(nz-k)
    +       *beta(2,3)**iabs(nz)
    +       /delp(i0)
c calculate response of mic to arrival  from (x,y,z)
```

```
           mic = 1
           if (mictype .ne. 0) then
              mic = micresponse(mictype,micaxis,x(i0),y(i0),z(i0))
           endif
c add computed impulses to appropriate bins if within range of points
c(incorporate offset from source to focus - note this only
c affects the time of output, not the magnitude)
c
           do 14 m=1,41
              mm=m+delp(i0)-offset+offmax
              if((mm.ge.1).and.(mm.le.range)) then
                ht(mm)=ht(mm)+array(m)*gid*mic
              endif
14         continue
         endif
10      continue
20    continue
      return
      end
```

```
c
c ***********************************************************
c  subroutine banlim
c ***********************************************************
c
      subroutine banlim(off,fc,fs,array)
c subroutine to lowpass filter impulse response
c filter is hanning windowed brick wall filter
c
c fc = cutoff frequency
c fs = sampling frequency
c off = offset of arrival of impulse from sampling instant
c array = filter response returned
c
      real array(1),pi,wc,var,t,off,fc,fs,con
      pi=3.14159265359
      wc=2.*pi*fc/fs
      con=2.*pi/40.
      do 10 j=1,41
        jj=j-21
        t=float(jj)-off
        var=0.5*(1+cos(con*t))
        if(abs(t).lt.1.e-5) go to 20
        var=var*sin(wc*t)/(pi*t)
        go to 30
20      continue
        var=var*wc/pi
30      continue
        array(j)=var
10    continue
      return
      end
```

```
c *******************************************************
c    subroutine lthim
c *******************************************************
c
      subroutine lthim(dr,dr0,rL,nr,delp,x,y,z)
c subroutine to compute eight images of a point in a box
c
c dr = vector radius to receiver in sample periods
c dr0 = vector radius to source in sample periods
c rL = vector of box dimensions in sample periods
c nr = vector of mean image number
c delp = vector of eight source to image distances
c      in sample periods
c rl = vector of positions of images
c x,y,z = normalized direction vectors for various images
c
      real r2l(3),rL(3),delp(8),dr0(3),dr(3),rp(3,8)
      real r1(3),delsq,x(8),y(8),z(8)
      integer nr(3)
c loop over all sign permutations and compute r+/r-
      i0=1
      do 10 l=-1,1,2
      do 10 j=-1,1,2
      do 10 k=-1,1,2
c nearest image is l=j=k=-1
       rp(1,i0)=dr(1)+l*dr0(1)
       rp(2,i0)=dr(2)+j*dr0(2)
       rp(3,i0)=dr(3)+k*dr0(3)
       i0=i0+1
10    continue
c add in mean radius to eight vectors to get total delay
      r2l(1)=2.*rL(1)*nr(1)
      r2l(2)=2.*rL(2)*nr(2)
      r2l(3)=2.*rL(3)*nr(3)
c calculate distances to images
      do 30 i=1,8
      delsq=0.
       do 20 j=1,3
        r1(j)= r2l(j)-rp(j,i)
        delsq=delsq+r1(j)**2
20    continue
      delp(i)=sqrt(delsq)
c calculate normalized (unit) vector of direction to image from mic
       x(i) = r1(1)/delp(i)
       y(i) = r1(2)/delp(i)
       z(i) = r1(3)/delp(i)
30    continue
      return
      end
```

```
c
c ****************************************************
c   function micresponse
c ****************************************************
c
      Function micresponse(mictype,micaxis,x,y,z)
c
c this function calculates the spatial response of a microphone
c inputs:
c      mictype = type of response desired
c      x,y,z = unit vector for source direction (from mic)
c      micaxis = unit vector for mic-axis (from mic)
c (for now, only the cardioid is available  i.e. mictype = 1)
c
c   cardioid of revolution response:
c          micresponse = 0.5*(1 + cos(theta))
c
c but from linear algebra
c  theta = acos (a*b)
c if a and b are unit vectors and * denotes the dot product
c
c
      real micresponse,x,y,z,micaxis(3)
c
      micresponse = 1.
      if (mictype .eq. 1) then
        micresponse=0.5*(1+ x*micaxis(1)+ y*micaxis(2)+ z*micaxis(3))
      else
        print *,'Error in function micresponse...'
        print *,'Non-existent response specified'
      endif
      return
      end
```

LIST OF REFERENCES

References

1. Thomas Boozer, "For Safety's Sake," *Cellular Business*, pp. 44-48, Sept. 1984.

2. Stuart Crump, *Cellular Telephones: A Layman's Guide,* Tab Books Inc., Blue Ridge Summit, PA, 1985.

3. Vladimir Cuperman and Neil Fried, "Evaluation of Speech Recognition Equipment in Vehicular Environment," *Submitted to the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, Victoria, B.C., June 1989.

4. R.A. Goubran and H.M Hafez, "Background Acoustic Noise Reduction in Mobile Telephony," *Proc. 1986 IEEE Vehic. Tech. Conf.*, pp. 72-75, Dallas, 1986.

5. M. H. Savoji, "A Variable Length Lattice Filter for Adaptive Noise Cancellation," *Proc. 1986 ICASSP*, pp. 2935-2938, Tokyo, April 1986.

6. Jont B. Allen and David A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943-950, Apr. 1979.

7. W. Armbruster, R. Czarnach, and P. Vary, "Adaptive Noise Cancellation With Reference Input - Possible Applications and Theoretical Limits," *Proc. European Signal Process. Conf EUPISCO-86*, pp. 391-394, The Hague, 1986.

8. Neviano Dal Degan and Claudio Prati, "Acoustic Noise Analysis and Speech Enhancement Techniques for Mobile Radio Applications," *Signal Processing*, vol. 15, no. 1, pp. 43-56, July 1988.

9. Glenn D. White, *The Audio Dictionary,* University of Washington Press, Seattle, 1987.

10. J. M. Berman, "Behavior of Sound in a Bounded Space," *J. Acoust. Soc. Am.*, vol. 57, no. 6, pp. 1275-1291, June 1975.

11. Patrick M. Peterson, "Simulating the Response of Multiple Microphones to a Single Acoustic Source in a Reberberant Room," *J. Acoust. Soc. Am.*, vol. 80, no. 5, pp. 1527-1529, Nov. 1986.

12. A. Craggs, "A Finite Element Model for Acoustically Lined Small Rooms," *J. Sound Vib.*, vol. 108, no. 2, pp. 327-337, 1986.

13. Richard G. DeJong, "A Study of Vehicle Interior Noise Using Statistical Energy Analysis," *Surface Vehicle Noise and Vibration Conf. Proc. P-161*, pp. 1-6, Warrendale, PA., May 1985.

14. M.L. Gayford, *Electroacoustics,* American Elsevier Publishing Company, New York, N.Y., 1971.

15. Alan V. Oppenheim and Ronald W. Schafer, *Digital Signal Processing*, Prentice-Hall Signal Processing Series, Prentice-Hall, Englewood Cliffs, N.J., 1975.

16. William E. Boyce and Richard C. DiPrima, *Elementary Differential Equations and Boundary Value Problems,* Wiley & Sons, New York, 1977.

17. S. D. Dodd, *Assisted resonance, theory of source-receiver transmission characteristics*, M.Sc. Dissertation, Inst. Sound Vib., Southampton U., 1969.

18. Philip M. Morse and K. Uno Ingard, *Theoretical Acoustics*, Princeton University Press, Princeton, N.J., 1986.

19. S. Thomasson, "Reflection of Waves from a Point Source by an Impedance Boundary," *J. Acoust. Soc. Am.*, vol. 59, pp. 780-785, 1976.

20. E. Andresdottir and R. W. Schafer, "Application of Adaptive Noise Cancelling in a Noisy Reverberant Environment," *Proc. 1983 ICASSP*, pp. 57-60, Boston, Mass., April 1983.

21. Stephen T. Neely and Jont B. Allen, "Invertibility of a Room Impulse," *J. Acoust. Soc. Am.*, vol. 66, no. 1, pp. 165-169 , July 1979..

22. Gilbert Strang, *Linear Algebra and its Applications*, Academic Press, Orlando, Fla., 1980.

23. Allan D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications*, McGraw-Hill Series in Mechanical Engineering, McGraw-Hill, New York, N.Y., 1981.

24. P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 70-73, June 1970.

25. Arthur A. Giordano and Frank M. Hsu, *Least Square Estimation with Applications to Digital Signal Processing*, Wiley & Sons, New York, N.Y., 1985.

26. A.G. Piersol, "Use of Coherence and Phase Data Between Two Receivers in Evaluation of Noise Environments," *J. Sound and Vib*, vol. 56, no. 2, pp. 215 - 227, 1978.

27. Harvey F. Silverman, "Some Analysis of Microphone Arrays for Speech Data Acquisiton," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 12, pp. 1699-1711, Dec. 1987.

28. Bernard Widrow et al., "Adaptive Noise Cancelling: Principles and Applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692 - 1716, Dec. 1975.

29. William A. Harrison, Jae S. Lim, and Elliot Singer, "Adaptive Noice Cancellation in a Fighter Cockpit Environment," *Proc. 1984 ICASSP*, pp. 18A.4.1 - 18A4.4, San Diego, Cal., March 1984.

30. J.J. Rodriguez, "Adaptive Noise Reduction in Aircraft Communication Systems," *M.I.T. Lincoln Laboratory Technical Report 756*, Lexington, Mass., Jan. 1987.

31. G.A. Powell, P. Darlington, and P.D. Wheeler, "Practical Adaptive Noise Reduction in the Aircraft Cockpit Environment," *Proc. 1987 ICASSP*, pp. 173-176, Dallas, Tx., April 1987.

32. J. Dunlop, M.J. Al-Kindi, and L.E. Kirr, "Application of Adaptive Noise Canelling to Diver Voice Communications," *Proc. 1987 ICASSP*, pp. 1708-1711, Dallas, Tx., April 1987.

33. Bernard Widrow and Samuel D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.

34. Dennis Pulsipher et al., "Reduction of Nonstationary Acoustic Noise in Speech Using LMS Adaptive Noice Cancelling," *Proc. 1979 ICASSP*, pp. 204-207, Washington, D.C., April 1979.

35. Hen-Geul Yeh, "Adaptive Noise Cancellation For Speech With a TMS3020," *Proc. 1987 ICASSP*, pp. 1171-1174, Dallas, Tx., April 1987.

36. George S. Kang and Lawrence J. Fransen, "Experimentation with an Adaptive Noise-Cancellation Filter," *IEEE Trans. Circuits Sys.*, vol. CAS-34, no. 7, pp. 753-758, July 1987.

37. Bernard Widrow et al., "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," *Proc. IEEE*, vol. 64, no. 8, pp. 1151 - 1161, Aug. 1976.

38. Eweda Eweda and Odile Macchi, "Convergence of the RLS and LMS Adaptive Filters," *IEEE Trans. Circuits and Systems*, vol. CAS-34, no. 34, pp. 799 - 803, July 1987.

39. L.J. Griffiths, "A New Approach to Partially Adaptive Arrays," *Proc. 1987 ICASSP*, pp. 1999-2002, Dallas, Tx., April 1987.

40. P. Darlington, P.D. Wheeler, and G.A. Powell, "Adaptive Noise Reduction in Aircraft Communication Systems," *Proc. 1985 ICASSP*, pp. 716-719, Tampa, Fla., March 1985.

41. L.J. Griffiths and C.W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," *IEEE Trans. Ant. Prop.*, vol. AP-30, no. 27, pp. 27-34, 1982.

42. Patrick M. Peterson, "Using Linearly-Constrained Adaptive Beamforming to Reduce Interference in Hearing Aids From Competing Talkers in Reverberant Rooms," *Proc. 1987 ICASSP*, pp. 2364-2367, Dallas, Tx., April 1987.

43. A. H. Koenig, J. B. Allen, and D. A. Berkley, "Determination of Masking-Level Differences in a Reverberant Environment," *J. Acoust. Soc. Am.*, vol. 61, no. 5, pp. 1374-1376, May 1977.

44. William C. Y. Lee, *Mobile Communications Engineering*, p. 305, McGraw Hill, New York, N.Y., 1982.

45. Rainer Zelinski, "A Microphone Array with Adaptive Post-Filtering for Noise Reduction in Reverberant Rooms," *Proc. 1988 ICASSP*, pp. 2578-2581, New York, N.Y., April 1988.

46. Henry M. Dante, "Effect of Signal Bandwidth on the Accuracy of Adaptive Interpolation of Discrete-Time Signals," *Proc. 1987 ICASSP*, pp. 1292-1295, Dallas, Tx., April 1987.