APPLICATION OF MICROPHONE ARRAYS IN HANDS-FREE TELEPHONY

by

Takafumi Chujo

Memorandum No. UCB/ERL M86/70

2 September 1986

APPLICATION OF MICROPHONE ARRAYS IN HANDS-FREE TELEPHONY

by

Takafumi Chujo

Memorandum No. UCB/ERL M86/70

2 September 1986

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

TITLE PAGE

APPLICATION OF MICROPHONE ARRAYS IN HANDS-FREE TELEPHONY

by

Takafumi Chujo

Memorandum No. UCB/ERL M86/70

2 September 1986

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

APPLICATION OF MICROPHONE ARRAYS IN HANDS-FREE TELEPHONY

Takafumi Chujo

Fujitsu Laboratories Ltd.
1015 Kamikodanaka Nakahara-ku
Kawasaki, 211, Japan

ABSTRACT

This paper discusses the properties of microphone arrays as a voice input device for hands-free telephony. The goal is to reduce the effects of reverberation of speech sounds picked up in a room. Performance criteria for this application have not previously been clearly established. The main objective of this paper is to derive a performance criterion in terms of the intelligibility of speech, and to evaluate the microphone array performance according to it.

The signal-to-echo energy ratio determined from experiments is introduced as a criterion. Based on geometrical acoustics, this measure has been evaluated from the primary acoustic parameters of a room, such as room size and surface reflection coefficients. The spatial response of one-dimensional microphone array, especially the directivity index, has been analyzed. The degree of dereverberation is estimated in terms of the signal-to-echo energy ratio. By combining the predicted signal-to-echo energy ratio with the improvement, it is possible to predict the intelligibility of speech received by the microphone array. The results of computer simulations made for microphone array signal processing showed the signal-to-echo energy ratio agrees with the predicted value. The signal-to-echo energy ratio deteriorates for large rooms in which the reverberation time exceeds 1 second. A one-dimensional microphone array which has the directivity index of 10 dB would be feasible at least in terms of the intelligibility for all but the largest rooms.

This work was done when the author was at Electronics Research Laboratory, University of California, Berkeley, as a visiting industrial fellow from August 1985 to August 1986 and joined the acoustic echo canceller project.

ACKNOWLEDGEMENTS

I would like to thank to Professors David A. Hodges and David G. Messerschimitt for the many useful discussions and suggestions.

I wish to thank Wen-Bin Hsu for his kind support throughout the project.

I am also grateful to M. Sekigawa, Y. Mochida, Dr. K. Murano and Dr. T. Tsuda of Fujitsu Laboratories Ltd., and S. Hinoshita, T. Kudo and Dr. M. Yamasawa of Fujitsu America Inc. for their encouragement and support during the stay at University of California, Berkeley. Partial support was also provided by State of California Microelectronics Innovation and Computer Research Opportunities Program.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 1

CHAPTER 2 SYSTEM DESCRIPTION 3

2.1 System Configuration 3

2.2 Design Problems 3

CHAPTER 3 ROOM ACOUSTICS 5

3.1 Sound Reflection Model Based on Geometrical Acoustics 5

3.2 Computer Simulation Using the Image Method 8

CHAPTER 4 SUBJECTIVE EFFECTS OF SOUND REFLECTIONS 10

4.1 Signal-to-Echo energy ratio 10

4.1 Predicted Signal-to-Echo Energy Ratio 11

CHAPTER 5 MICROPHONE ARRAYS 13

5.1 Configuration 13

5.2 Spatial Response and Directivity Factor 13

5.3 Elimination of Near-End-Talker Echos 15

5.4 Computer Simulation 17

CHAPTER 6 CONCLUSION 18

REFERENCES 19

CHAPTER 1

Introduction

Among various communication media. such as mail. electronic mail and facsimile. the telephone has been most comfortable medium because real time person-to-person conversation is the most natural form of human communication. Even if in the upcoming information age. it will remain the dominant medium.

The demand for hands-free telephony has arisen in teleconferencing systems. Another demand is expected to come up in the forthcoming ISDN (Integrated Services Digital Network) age. The key feature of the ISDN is integration of a various communication media. such as voice. data and image and to allow users to use these media simultaneously. Integration of the phone with the data/image terminal means that people may talk with each other over the phone. each referring to a personal data base. for example their schedule. memo. mails and document files. According to this concept. it is necessary to use a keyboard and a pointing device while in conversation. In these situations. however. a conventional handset seems to be not an ideal human-machine interface since it forces the person to keep holding it. Thus hands-free voice communication may become more important in the ISDN environment.

However. many problems arise in the design of a hands-free telephone [1.1]. First of all. an acoustic feedback loop between a speaker and a microphone causes echoes (far-end-echoes). Second. it also causes singing if the feedback loop gain is high. Third. speech sound picked up by the microphone includes not only the direct speech sound but also the reflected speech sound from walls and this causes annoyance (near-end-echoes).

Although hands-free telephones were used as early as in 1950's [1.2]. [1.3]. they have not been commonly used. The main reason seems to be that surrounding environments are very complicated. There are many different size of rooms and many kinds of interior in the rooms because people work and live there. So communication equipments must handle not only phenomena in electric media but also human environments.

Thanks to advanced VLSI technologies, communication equipments may be within reach to handle such a complicated acoustic environment. For teleconferencing systems, an acoustic echo canceller has been implemented to eliminate the far-end-talker echos and increase stability margin against singing [1.4]. These approaches require a fairly long echo processing window because sound velocity in air is definitely longer than one in electric media. At present, current digital signal processors can handle reflections with acoustic delay of several hundreds milliseconds. However, this approach is not effective for the near-end-talker echos.

In this paper much emphasis has been put on elimination on near-end-talker echoes. For this end, microphone arrays have been chosen as the most promising technique. The basic idea is to use the directivity property of arrays, steer it to the direction of the near-end-talker and thus reduce near-end-talker sound reflections as well as far-end-talker sounds. The current approaches involve a single microphone and signal processing in one-dimensional space, such as time-domain and frequency-domain, after receiving sound by a microphone. Taking into account multi-talker and multi-noise environment, a one-dimensional approach may require sophisticated signal processing to separate the near-end-talker from the other talkers and the noise, such as recognition of the sound sources. Spatial filtering in three-dimensional space by using microphone arrays seems to be helpful to reduce the complexity of the signal processing.

CHAPTER 2

System Description

In this chapter the configuration of hands-free telephones will be described and then the design problems will be discussed.

2.1. System Configuration

Figure 2.1.1 shows the configuration of a hands-free digital telephone. The received digital signal is converted to an analog signal by a D/A converter and then fed to a power amplifier to drive the speaker. The gain of the power amplifier is adjusted so that the level of the far-end-talker speech is comfortable for the near-end-talker.

The near-end-talker speech is picked up by a microphone and then amplified to a certain level (Ps0) specified by an A/D converter. Because the distance between the microphone and the near-end-talker may be dynamically changed, the gain of the amplifier should be automatically adjusted so that the output level become constant. The output signal of the amplifier is converted to a digital signal by the A/D converter and then transmitted to the other party.

2.2. Design Problems

The far-end-talker speech sound wave emitted by the speaker may be picked up by the microphone and thus make an acoustic transmission path between them. This signal is transmitted to the other party as well as the near-end-talker speech. Because acoustic propagation delay of this loop is fairly large, this is very annoying for the far-end-talker. And also the returned speech may be picked up again by the microphone at the far-end. If the loop gain is high enough to satisfy the oscillation condition, this causes singing and prevents from the normal voice conversation. To overcome these problems, echo cancellation techniques have been developed. The echo canceller generates echo replica and subtracts it from the input signal.

Another problem is near-end-talker echoes. The microphone at the near-end picks up not only the direct sound from the near-end-talker but also sound reflections from the walls, the ceiling, and the floor in the room. This combination of the direct sound and the reflected sounds is very annoying for the far-end-listener. Aiming at dereverberation of the received sound, multi-microphone approaches have been proposed [2.2.1], [2.2.2]. In these papers, however, no performance criterion is clearly established and thus it would be desirable to specify how much echo power should be eliminated. Another fact we should take into account is that there may be noise source, such as a printer, and another talkers in the room. Because each acoustic path between each noise source and the microphone is different from each other, so those approaches mentioned above seem to be not effective for multi-talker and multi-noise environments.

We would like to introduce microphone arrays to eliminate the near-end-talker echoes and derive a performance criterion in terms of speech intelligibility to figure out the feasibility of microphone arrays.

CHAPTER 3

Room Acoustics

In this chapter the room acoustics will be reviewed to obtain a good insight of sound reflections. The basic idea of using a microphone array is to receive the direct sound from the near-end-talker and eliminate the sound reflections coming from the other directions. To evaluate this performance. it would be useful to model the sound reflections and then evaluate its characteristics. such as direct-to-reverberant acoustic energy ratio. temporal structure of sound reflections and relationship between these characteristics and acoustical parameters of the room.

3.1. Sound reflection model based on geometrical acoustics

To simplify the problems. geometrical acoustics are used to model sound reflections in a room. assuming that the room size is much larger than the wave length of the speech sound waves.

Consider a rectangular room whose dimensions are L_x ,L_y and L_z and model a talker in the room as a omnidirectional point sound source located at (x_s ,y_s ,z_s) and a receiver as a ideal omnidirectional microphone located at (x_r ,y_r ,z_r) The sound pressure at the receiver is composed of the direct sound pressure propagating from the source and the reflected sound pressure from the wall surfaces. As shown in Fig. 3.1.1. the basic idea of the image method is that each reflected sound wave is supposed to be propagating directly from the corresponding image source which is located at

$$(-x_r+2qx_r+2nL_x\,.\,-y_r+2jy_r+2mL_y\,.\,-z_r+2kz_r+2lL_z) \qquad (3.1.1)$$

where $(q\,.j\,,k)$ is an integer vector.

$$q\,.j\,,k = 0 \;\; or \;\; 1 \qquad (3.1.2)$$

and $(n\,,m\,,l)$ is also an integer vector.

$$-\infty < n, m, l < \infty \tag{3.1.3}$$

Furthermore, each image source emits as a same pressure wave as the original source and all of the sources are excited simultaneously. Thus, the sound pressure is expressed as

$$h(t) = \sum_{q,j,k} \sum_{n,m,l} g_{qjknml}\,\delta\left(t - \frac{d_{qjknml}}{c}\right) \tag{3.1.4}$$

where g_{qjknml} is the sound pressure at the receiver location emitted by the original source and the image sources, c is the sound velocity and d_{qjknml} is the distance between the image source and the receiver

$d_{qjknml} =$

$$\sqrt{(x_s - x_r + 2qx_r - 2nL_x)^2 + (y_s - y_r + 2jy_r - 2mL_y)^2 + (z_s - z_r + 2kz_r - 2lL_z)^2} \tag{3.1.5}$$

Each sound wave propagating away from the source suffers spherical divergence by a factor of

$$\frac{1}{4\pi d_{qjknml}}. \tag{3.1.6}$$

Furthermore each sound wave originating from the image source is attenuated by a factor of each reflection coefficient passed through the wall

$$\beta_{x1}^{|n-q|}\beta_{x2}^{|n|}\beta_{y1}^{|m-j|}\beta_{y2}^{|m|}\beta_{z1}^{|l-k|}\beta_{z2}^{|l|} \tag{3.1.7}$$

where the β 's are the reflection coefficients of the six walls, with the subscript 1 referring to walls adjacent to the coordinate origin and the subscript 2 to the opposing wall. [3.1.1]

Apply (3.1.4) and (3.1.5) to (3.1.2) , the complete impulse response of the room is

$$h(t) = \sum_{q,j,k} \sum_{n,l,m} \beta_{x1}^{|n-q|}\beta_{x2}^{|n|}\beta_{y1}^{|m-j|}\beta_{y2}^{|m|}\beta_{z1}^{|l-k|}\beta_{z2}^{|l|} \frac{\delta\left(t - \frac{d_{qjknml}}{c}\right)}{4\pi d_{qjknml}}. \tag{3.1.8}$$

In the following, however, we use the average value to obtain the temporal structure of the sound reflections rather than the exact value. Consider the average number of the reflections that arrive in the center of the room in the time interval from time t to $t+dt$. All those reflections are propagating away from the image sources which are located in a spherical shell with radius ct and thickness cdt. Since the volume of the shell is $4\pi c^3t^2$, it contains $4\pi c^3t^2/ V$ image sources on the average. Thus the temporal density of the reflections arriving at time t is

$$dN = \frac{4\pi c^3t^2}{V}dt \tag{3.1.9}$$

where V is the volume of the room. These reflections suffer from spherical divergence and also attenuation due to the absorption by the walls as mentioned above. The sound intensity is given by [3.1.2].

$$\begin{aligned} h^2(t) &= (\frac{4\pi c^3t^2}{V})\frac{A_0}{4\pi(ct)^2}\exp[-mc + n\ \ln(1-\bar{\alpha})]t \\ &= \frac{cA_0}{V}\exp[-mc + n\ \ln(1-\bar{\alpha})]t \end{aligned} \tag{3.1.10}$$

where m is the absorption coefficients of the medium (air) and n is the average total number of reflection .

$$n = \frac{c}{2}(\frac{1}{L_x} + \frac{1}{L_y} + \frac{1}{L_z}) = \frac{cS}{4V} \tag{3.1.11}$$

and $\bar{\alpha}$ is the average absorption coefficient

$$\bar{\alpha} = \frac{\sum_{i=1}^{6}\alpha_i S_i}{\sum_{i=1}^{6} S_i} \tag{3.1.12}$$

in which S_i and α_i is the area and the absorption coefficient. respectively. of each wall of the room. Assuming that $mc \ll - nln(1-\bar{\alpha})$. we get

$$h^2(t) = \frac{A_0}{4\pi d_0^2}\delta(t - \frac{d_0}{c}) + \frac{cA_0}{V}\exp[n\ \ln(1-\bar{\alpha})]\, t \tag{3.1.13}$$

where the first term is the direct sound intensity and d_0 is the distance between the source and the receiver.

As mentioned above. the room impulse response is composed of the direct impulse and the reverberant impulses. Assuming that the direct impulse is independent of the reverberant impulses. the total power of the room impulse may be expressed by sum of the power of the direct impulse and the power of the reverberant impulses

$$P = P_d + P_r. \tag{3.1.14}$$

We define the direct-to-reverberant acoustic energy ratio as

$$D/R = \frac{P_d}{P_r} \tag{3.1.15}$$

From the temporal distribution function of sound reflections, P_d and P_r is

$$P_d = \frac{A_0}{4\pi d_0^2}, \tag{3.1.16}$$

$$P_r = \int_{t=0}^{\infty} \frac{cA_0}{V} \exp[n\ \ln(1-\bar{\alpha})]t\,dt, \tag{3.1.17}$$

respectively. We obtain the direct-to-reverberant energy ratio

$$D/R = -\frac{S\ln(1-\bar{\alpha})}{16\pi d_d^2}. \tag{3.1.18}$$

The critical distance d_c of a room is defined as a distance from a omnidirectional point sound source at which the direct and reverberant energy densities are equal. From (3.1.5).

$$1 = -\frac{S\ln(1-\bar{\alpha})}{16\pi d_0^2} \tag{3.1.19}$$

we get

$$d_c = \sqrt{-S\ln(1-\bar{\alpha})/16\pi}. \tag{3.1.20}$$

The reverberation time T_{60} is defined as the time interval in which the reverberation level drops down 60 dB [3.1.1]. From (3.1.10),

$$-60 = 10\log(\exp[n\ \ln(1-\bar{\alpha})]t\ ) \tag{3.1.21}$$

and thus we get

$$T_{60} = -\frac{24V\ln 10}{cS\ln(1-\bar{\alpha})}. \tag{3.1.22}$$

3.2. Computer Simulation Using the Image Method

Based on geometrical acoustics, a computer simulation program which calculates the impulse response of a room has been developed [3.1.1] This program exactly calculates the direct sound impulse response and the reverberant impulse responses from the image source, according to (3.1.8) . Using this program, the analytical results described in the previous section will be compared with those values which are calculated from simulated impulse responses in terms of the direct-to-reverberant energy ratio and the reverberation time.

In the computer simulation a rectangular room is also assumed. The room size is 10×15×12.5 ft. The receiver was at coordinates (5,7,5), taking a corner of the room as the origin and the sound source, the talker, assumed to be moving along a straight line. The impulse response was calculated at five sound source positions on the line, which were at (A): (5.5,6.75,5.5) , (B): (6,6,5.5) , (C): (7,4.5,5.5), (D): (8,3,5.5) , and (E): (9,1.5,5.5) . The distance between each source location and the receiver was 0.75, 1.5, 3.2, 5.0, and 6.8 ft, respectively. Each wall was assumed to have a uniform reflection coefficients of 0.9.

The impulse response for each source location has been computed. Figure 3.2.1 shows the computer-simulated impulse response for the source location (E). Figure 3.2.2 shows the direct-to-reverberant ratio for the source location (A) through (E). Figure 3.2.3 also shows the power decay curve of each impulse response for the source location (A) through (E). The power decay curve was calculated according to Schroeder Integration method [3.2.1], given by

$$p(t) = \int_t^{\infty} h^2(\tau) d\tau \qquad (3.2.1)$$

Table 3.2.1 shows reverberation time of the room calculated from the decay curve. As seen in Fig. 3.2.2 and Table (3.2.1), the results calculated from the computer-simulated impulse responses agree with those derived theoretically.

CHAPTER 4

Subjective Effects of Sound Reflections

The purpose of this chapter is to derive the performance criterion for the dereverberation in terms of the speech intelligibility and then predict it from the primary acoustic parameters, such as the size of the room and the average absorption coefficients of the walls. The subjective effects of sound reflections have been investigated to predict level of the speech intelligibility prior to construction of a room, an auditorium and a concert hall. In the following, one of these criterions will be reviewed to apply it to the microphone array design.

4.1. Signal-to-Echo Energy Ratio

The intelligibility of speech in a room is a function of speech level and masking noise level as shown in Fig. 4.1.1 [4.1.1]-[4.1.3].

Classical experiments carried out by Hass [3.1.2] indicate that a reflection with a very short time delay is not perceived at all. The intelligibility of speech in a room seems to be increased by reflections arriving within a short period after the direct sound and decreased by later reflections. This subjective effect suggests that human ears integrate some of the early reflections with the direct sound. Lochner and Burger carried out many subjective experiments and obtained the integration curve.

Based on the results, they defined a useful energy and a detrimental energy for the speech intelligibility in a room [4.1.1]-[4.1.3]. The useful energy is

$$E_E = \int_0^{95_{ms}} a(t)h^2(t)dt \qquad (4.1.1)$$

where $a(t)$ is the weighting function shown in Fig. 4.1.2 and the $h(t)$ is the impulse response of the room. The reflections within 40 ms are fully integrated and from 40 ms to 95 ms partially integrated. The detrimental energy is

$$E_{NE} = \int_{95_{ms}}^{\infty} h^2(t)dt. \tag{4.1.2}$$

All reflections arrive after 95 ms are considered as noise. The signal-to-echo energy ratio is defined as

$$S/E = 10\log\frac{E_E}{E_{NE}}. \tag{4.1.3}$$

This is regraded as some sort of "signal-to-noise-ratio". If we measure an impulse response in a room and then compute the signal-to-echo energy ratio according to (4.1.3) , the intelligibility of speech will be predicted from Fig. 4.1.1.

4.2. Predicted signal-to-echo energy ratio

In this section we will introduce the temporal distribution function of the sound reflections derived in the previous chapter to calculate the signal-to-echo energy without measuring the impulse response in the room of interest. The computer simulation results also will be shown and compared the signal-to-echo energy ratio calculated from the simulated impulse response with the predicted one.

As mentioned in the previous chapter, the temporal distribution of the sound reflections can be predicted from the acoustic conditions of the room. Apply (3.1.13) to (4.1.1) and (4.1.2) ,

$$E_E = \frac{A_0}{4\pi d_0^2} + \int_0^{95_{ms}} E_0 e^{-\gamma t}\,dt \tag{4.2.1}$$

where E_0 is cA_0/V and γ is $-n\ln(1-\overline{\alpha})$.

$$E_{NE} = \int_{95_{ms}}^{\infty} E_0 e^{-\gamma t}\,dt. \tag{4.2.2}$$

Then the weighting function $a(t)$ shown in Fig. 4.1.2 is approximated as follows.

$$a(t) = \begin{cases} 1 & 0 < t < t_1 \\ \dfrac{(t_2 - t)}{(t_2 - t_1)} & t_1 \leqslant t \leqslant t_2 \\ 0 & t_2 < t \end{cases} \tag{4.2.3}$$

where t_1 is 35_{ms} and t_2 is 95_{ms}. The signal-to-echo energy ratio is

$$S/E = \frac{D + 1 - \dfrac{e^{-\gamma t_1} - e^{-\gamma t_2}}{\gamma(t_2 - t_1)}}{e^{-\gamma t_2}} \qquad (4.2.4)$$

where D is the direct-to-reverberant energy ratio at the receiver location.

The signal-to-echo energy ratios have been calculated for rooms: 10×15×12.5 ft (A), 20×30×12.5 ft (B) and 20×30×25 ft (C). The average reflection coefficients were assumed to be 0.9. The results are shown in Fig. 4.2.1.

For the room (A), the signal-to-echo energy ratio have been calculated from the computer-simulated impulse responses which were used for the calculation of the direct-to-reverberant energy ratios shown in Fig. 3.2.2. The results are shown in Fig. 4.2.2 and compared with the predicted signal-to-echo energy ratios. The result shows excellent agreement with the theoretical calculation. Thus, the temporal distribution function can be used to predict the signal-to- echo energy ratio.

In Fig. 4.2.2, the further the receiver is placed away from the source, as lower the signal-to-echo energy is. This minimum signal-to-echo energy in the room could be calculated as the direct-to-reverberant energy ratio of 0 in (4.2.4). The result is shown in Fig. 4.2.3. It shows that the degradation of the signal-to-echo energy ratio and thus the annoyance of echoes would occur in a relatively large room whose reverberation time exceeds 1 second.

CHAPTER 5

Microphone Arrays

In this chapter we will analyze the properties of microphone arrays, especially on the directivity factor and predict the degree of improvement of the signal-to-echo energy ratio.

5.1. Configuration

Figure 5.1.1 shows the configuration of the microphone array. It consists of M microphone elements, variable delays, a summation circuit and a beam steering control circuit. The output signals of the microphone elements are fed to variable delay circuits which are controlled so as to steer the beam to a certain direction by the beam steering circuit. The delayed signals are then summed to conform the array output signal.

Figure 5.1.2 shows an example of an office where the hands-free telephone using microphone arrays is installed. The beam of the microphone array is steered so as to eliminate the sound reflections originating from the near-end-talker as well as the far-end-talker speech sounds coming from the speaker. For the beam steering, a simple algorithm has been proposed by J.L.Flanagan [5.1.1].

5.2. Spatial Response and Directivity Factor

In this section the spatial response of microphone arrays will be analyzed and the traditional directivity index will be derived.

Figure 5.2.1 shows geometry of a one-dimensional microphone array which consists of 2N+1 equally spaced microphone elements. All microphone elements have the same omnidirectional sensitivity which is equal to unity.

Consider a plane wave arriving from the direction (θ, ϕ) as given by the coordinate system in Fig. 5.2.2. The time-domain response of the unweighted one-dimensional array is,

$$h(t,\theta,\phi) = \sum_{n=-N}^{N} \delta(t + nT) \qquad (5.2.1)$$

where T is the inter-element delay.

$$T = \frac{d}{c}(\cos\phi - \cos\phi') \qquad (5.2.2)$$

in which d is the spacing between microphone elements and ϕ' is the beamforming direction.

The corresponding amplitude-frequency response is

$$H(j\omega,\theta,\phi) = \sum_{n=-N}^{N} e^{j\omega nT}. \qquad (5.2.3)$$

The power-frequency response is

$$|H(j\omega,\theta,\phi)|^2 = 2N + 1 + \sum_{k,l=-N,k\neq l}^{N} \cos\omega(k-l)T \qquad (5.2.4)$$

and then the power response is

$$|f(\theta,\phi)|^2 = \int_{\omega_l}^{\omega_u} |H(j\omega,\phi)|^2 d\omega = (2N+1)(\omega_u - \omega_l) +$$

$$\sum_{k,l=-N,k\neq l}^{N} \frac{1}{(k-l)T}(sin\,\omega_u(k-l)T - \sin\omega_l(k-l)T). \qquad (5.2.5)$$

where $f_l(=\omega_l/2\pi)$ is the lowest useful frequency and $f_u(=\omega_u/2\pi)$ is the upper useful frequency.

The ability to reject random-incidence sound by directional microphones is measured by the directivity factor Q. Because the one-dimensional microphone array has the symmetrical directivity pattern in terms of ϕ. Q is given by [5.2.1]

$$Q = \frac{2}{\int_0^{\pi} |f(\theta,\phi)|^2 \sin\theta\, d\theta}. \qquad (5.2.6)$$

For the parameters of microphone arrays. such as the microphone element spacing d and the number of microphone elements $2N+1$, a design technique has been developed by J.L. Flanagan [5.2.2] and the microphone spacing d and the number of microphone elements are

$$d = \frac{1}{\frac{f_u}{c}|\cos\phi - \cos\phi'|_{max}} \qquad (5.2.7)$$

$$N = \left[\frac{1}{2}(\frac{f_u}{f_l} - 1)\right]_{nhi} \qquad (5.2.8)$$

Fig. 5.2.3 through 5.2.5 shows the spatial response $|H(\omega,\theta,\phi)|$ of the microphone array at 500 Hz, 1 kHz, 3 kHz, respectively.

Fig. 5.2.6 shows the power-angle response $|f(\theta,\phi)|^2$ of the microphone arrays which have 3 to 23 microphone elements. And also Fig. 5.2.7 shows the directivity indices. As shown in the figure the directivity index of 10 dB was obtained with 19 elements.

5.3. Elimination of Near-End-Talker Echoes

Figure 5.3.1 shows the acoustic transmission path from the talker to the microphone. If we assume that the talker is not moving in the room and also there is no movement in the room (that might change the acoustic transfer function between the source to the receiver), the transfer function is time-invarient.

The output impulse response is expressed by convolution of the impulse responses of the acoustic path and the microphone

$$h(t) = \int_{\Omega} \int_{t=-\infty}^{\infty} h_a(\tau,\theta,\phi)h_m(\tau - t,\theta,\phi)dtd\,\Omega \qquad (5.3.1)$$

where $h_a(t,\theta,\phi)$ is the impulse response arriving from the direction (θ,ϕ) and $h_m(t,\theta,\phi)$ is the impulse response of the directional microphone for the direction (θ,ϕ) and $d\,\Omega$ is the small area located at the coordinate (θ,ϕ) on the unit sphere surface whose center is at the microphone position.

We assume that the frequency characteristics of the microphone is all-pass, i.e.

$$h_m(t,\theta,\phi) = f(\theta,\phi)\delta(t) \qquad (5.3.2)$$

where $f(\theta,\phi)$ is the spatial response of the directional microphone and it is steered to the wave arrival direction (θ_0,ϕ_0) of the direct sound and the sensitivity is normalized as follows

$$f(\theta,\phi) = \frac{f'(\theta,\phi)}{f'(\theta_0,\phi_0)}. \tag{5.3.3}$$

Furthermore, we assume that each echo path (θ,ϕ) is independent, and thus the power of the impulse response is

$$h_a^2 = g_0^2\,\delta(t-t_0,\theta-\theta_0,\phi-\phi_0) + \sum_{i=1}^{\infty} g_i^2\delta(t-t_i,\theta-\theta_i,\phi-\phi_i) \tag{5.3.4}$$

where the first term is the direct sound energy and the second term is all of sound reflections coming from the image sources. And also we assume that the room is uniform, and thus the sound reflection energy is expressed as

$$\sum_{i=1}^{\infty} g_i^2\delta(t-t_i,\theta-\theta_i,\phi-\phi_i) = \frac{E_0}{4\pi}e^{-\gamma t} \tag{5.3.5}$$

Thus the impulse response is expressed by as follows.

$$h^2(t) = g_0^2\,\delta(t-t_0) + \frac{E_0}{Q}e^{-\gamma t} \tag{5.3.6}$$

where Q is the directivity factor.

$$Q = \frac{4\pi}{\int_\Omega f^2(\theta,\phi)d\,\Omega}. \tag{5.3.7}$$

Now we evaluate the impulse response based on the subjective criterion described in Chapter 4. From (4.1.1) and (4.1.2), the useful energy is

$$\begin{aligned} E_E &= g_0^2 + \int_0^{t_2} a(t)\frac{E_0}{Q}e^{-\gamma t}dt \\ &= g_0^2 + \frac{E_0}{aQ}\left\{1 - \frac{e^{-\gamma t_1} - e^{-\gamma t_2}}{\gamma(t_2-t_1)}\right\}. \end{aligned} \tag{5.3.8}$$

The detrimental energy is

$$E_{NE} = \int_{t_2}^{\infty}\frac{E_0}{Q}e^{-\gamma t}dt = \frac{E_0 e^{-\gamma t_2}}{aQ}. \tag{5.3.9}$$

Then the signal-to-echo energy is

$$S/E = \frac{DQ + 1 - \dfrac{e^{-\gamma t_1} - e^{-\gamma t_2}}{\gamma(t_2-t_1)}}{e^{-\gamma t_2}} \tag{5.3.10}$$

The signal-to-echo energy ratios have been calculated for the room conditions of Fig. 4.2.1.

and for the directivity factor of 10. As shown in Fig. 5.3.2 and 5.3.3, the improvement by the directivity factor depends on the distance between the source and the microphone. If the distance is short, the improvement as same as the directivity factor is obtained.

5.4. Computer Simulation

The one-dimensional microphone array shown in Fig. 5.1.1 has been simulated on the computer. The impulse response received by each microphone element was computed by using the image method program for the room condition of Fig. 4.2.1. The source position was (8, 3, 5.5) (C). Each impulse response was over-sampled at 64 kHz and then each steering delay was adjusted with the resolution of the sampling period. After low-pass filtering, each impulse response sample was decimated by a factor of 8.

The impulse response received by omnidirectional microphone which is one of microphone array elements is shown in Fig. 5.4.1 and the microphone array output is also shown in Fig. 5.4.2. The signal-to-echo ratio was calculated from the microphone array output for the number of microphone elements, 5, 7, 9, and 11. The results are shown in Figure 5.4.3. Owing to the random arrival times of the reflections , the signal-to-echo energy ratio of the output of the microphone arrays agrees with the predicted one from its directivity index.

CHAPTER 6

Conclusions

The properties of microphone arrays as a voice input device for hands-free telephony has been discussed.

The signal-to-echo energy ratio was introduced to the performance criterion for dereverberation of the near-end-talker echoes. Based on geometrical acoustics. the relationship between the signal-to- echo energy ratio and the primary acoustic parameters of the room was illustrated. This makes it possible to predict the signal-to-echo energy ratio from room size and reflection coefficients of the walls. the ceiling and the floor. The calculation made for various room conditions shows that the intelligibility of speech would degrade in a relatively large room in which the reverberation time is longer than one second.

The properties of uniform. unweighted. one-dimensional microphone arrays. specificly the directivity factor. was analyzed and also the relation between the improvement of the signal-to-echo energy ratio and the directivity factor was described. The computer simulation results of microphone arrays show the relation derived here can be used to predict the signal-to-echo energy ratio at the output of microphone arrays. In terms of the intelligibility of speech. one-dimensional microphone array is feasible to eliminate the near-end-talker echos in a large room.

REFERENCES

CHAPTER 1

[1.1] D.A.Berkley and O.M.M.Mitchell. "Seeking the Ideal in "Hands-Free" Telephony." Bell Labs. Record. pp.318-325. Nov.1974.

[1.2] W.F.Clemency. F.F.Romanow. and A.F.Rose. "The Bell System Speakerphone." A.I.E.E. Trans.. Pt.1. 76. p.148. 1957.

[1.3] J.W.Emiling. "General Aspects of Hands-Free Telephony. " A.I.E.E. Trans.. Pt.1. p.201. 1957.

[1.4] O.Horna. "Cancellation of Acoustic Feedback." COMSAT Technical Review. Vol.12. No.2. pp.320-333. Fall 1982.

CHAPTER 2

[2.2.1] J.L.Flanagan and R.C.Lummis. "Signal Processing to Reduce Multipath distortion in Small Rooms." J. Acoust. Soc. Am.. Vol.47. No.6(Part 1). pp.1475-1481. 1970.

[2.2.2] J.B.Allen. D.A.Berkley and J.Blauert. "Multimicrophone Signal Processing Technique to Remove Room Reverberation from Speech Signals." J. Acoust. Soc. Am.. Vol.62. No.4. pp.912-915. Oct. 1977.

CHAPTER 3

[3.1.1] J.B.Allen and D.A.Berkeley. "Image Method for Efficiently Simulating Small-Room Acoustics." J. Acoust. Soc. Am.. Vol.65. No.4. pp.943-950. Apr. 1979.

[3.1.2] H.Kuttruff. "Room Acoustics-2nd ed.." Applied Science Publishers Ltd. p.74. 1979.

[3.2.1] M.R.Schroeder. "New Method of Measuring Reverberation Time." J. Acoust. Soc. Am., Vol.37, pp.409-412, 1965.

CHAPTER 4

[4.1.1] J.P.A.Lochner and J.F.Burger. "Optimum Reverberation Time for Speech Rooms Based on Hearing Characteristics." Acustica, Vol.10, pp.394-399, 1960.

[4.1.2] J.P.A.Lochner and J.F.Berger. "The Intelligibility of Speech under Reverberant Conditions." Acustica, Vol.11, pp.195-200, 1961.

[4.1.3] J.P.A.Lochner and J.F.Berger. "The Influence of Reflections on Auditorium Acoustics." J. Sound Vib., Vol.1, No.4, pp.426-454, 1964.

CHAPTER 5

[5.1.1] J.L.Flanagan, J.D.Johnston, R.Zahn, and G.W.Elko. "Computer-Steered Microphone Arrays for Sound Transduction in Large Rooms." J. Acoust. Soc. Am., Vol.78, No.5, pp.1508-1518, Nov. 1985

[5.2.1] L.L.Beranek. "Acoustic Measurements." John Willey & Sons, Inc., p.648, 1949.

[5.2.2] J.L.Flanagan. "Beamwidth and Useable Bandwidth of Delay-Steered Microphone Arrays." AT&T Tech. J., Vol.64, No.4, pp.983-995, Apr. 1985.

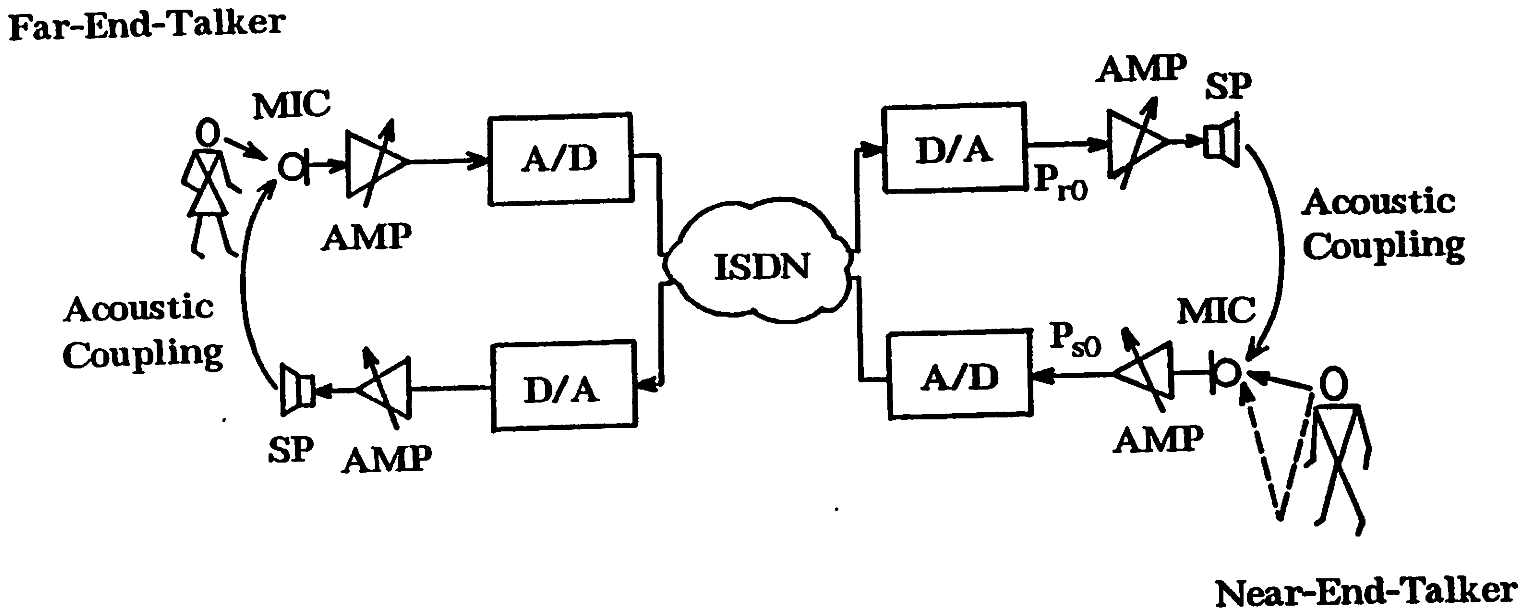


Fig. 2.1.1. Configuration of hands-free digital telephone

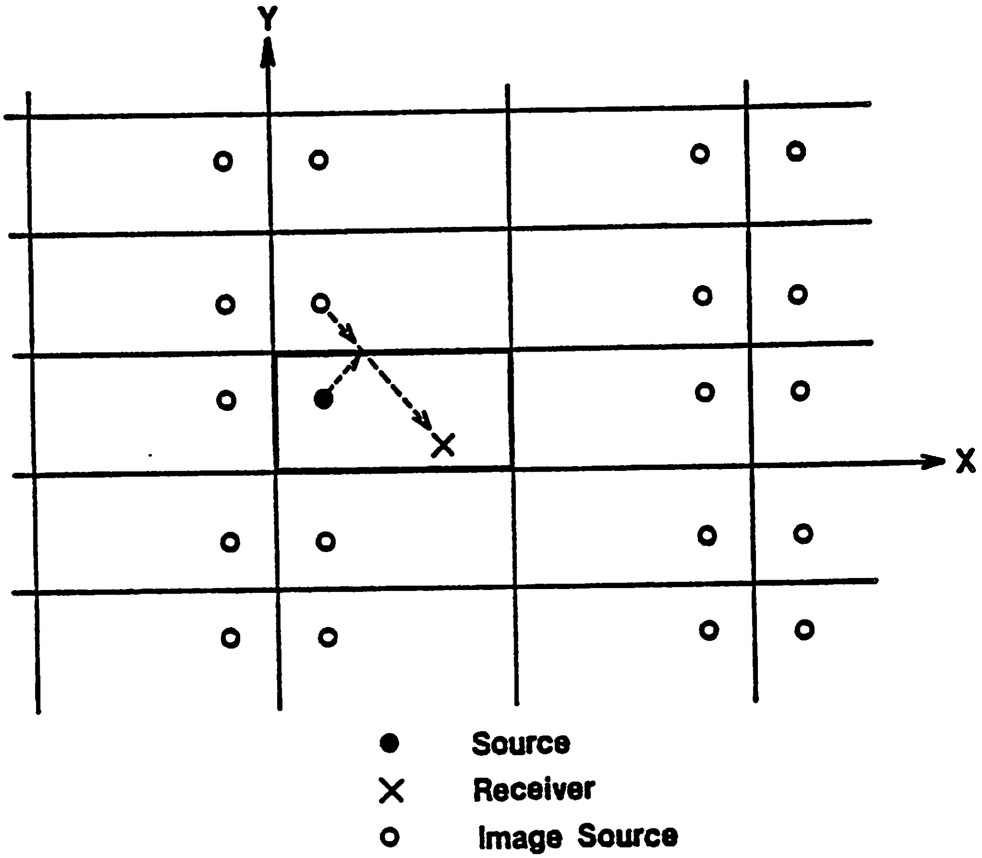


Fig. 3.1.1. Image sound sources for a rectangular room. The solid box represents the original room.

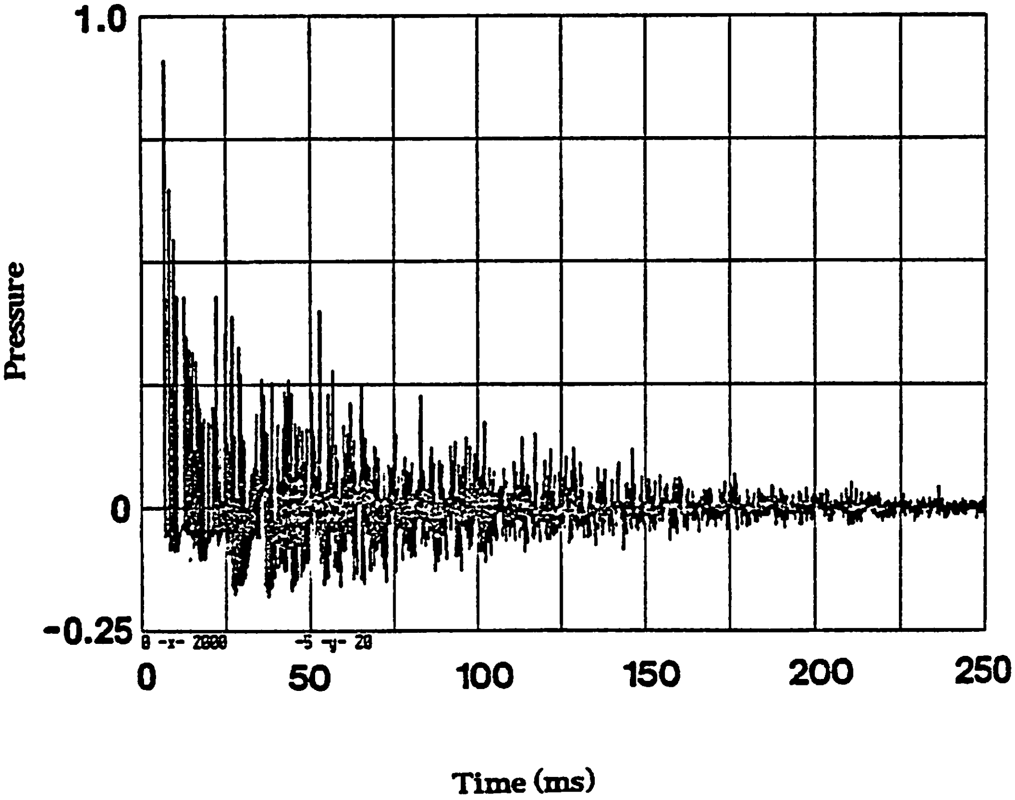


Fig. 3.2.1. Computer-simulated impulse response for a room: 10×15×12.5 ft. Wall reflection coefficients were all 0.9. Source was at (9, 1.5, 5.5). Receiver was at (5, 7, 5.5). Sampling rate was 8 kHz.

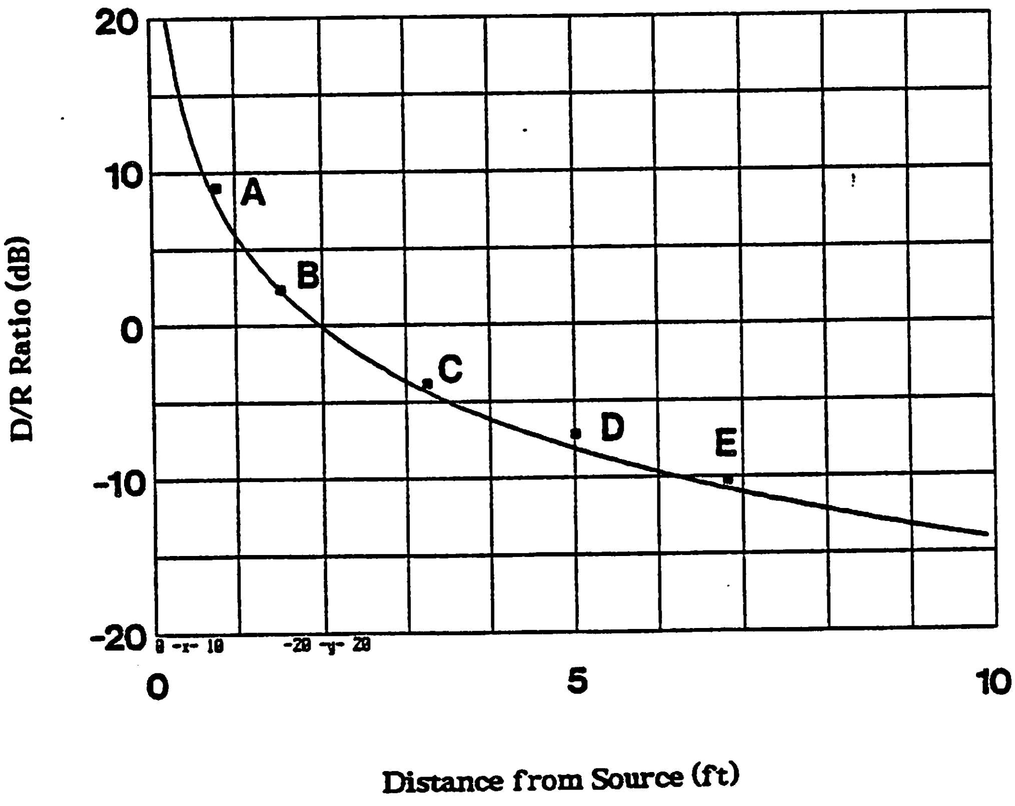


Fig. 3.2.2. Direct-to-reverberant energy ratios in a room: 10X15X12.5 ft. Wall reflection coefficients were all 0.9. Receiver was at (5, 7, 5.5). Source was at (5.5, 6.75, 5.5) (A), (6, 6, 5.5) (B), (7, 4.5, 5.5) (C), (8, 3, 5.5) (D), and (9, 1.5, 5.5) (E). Sampling rate was 8 kHz.

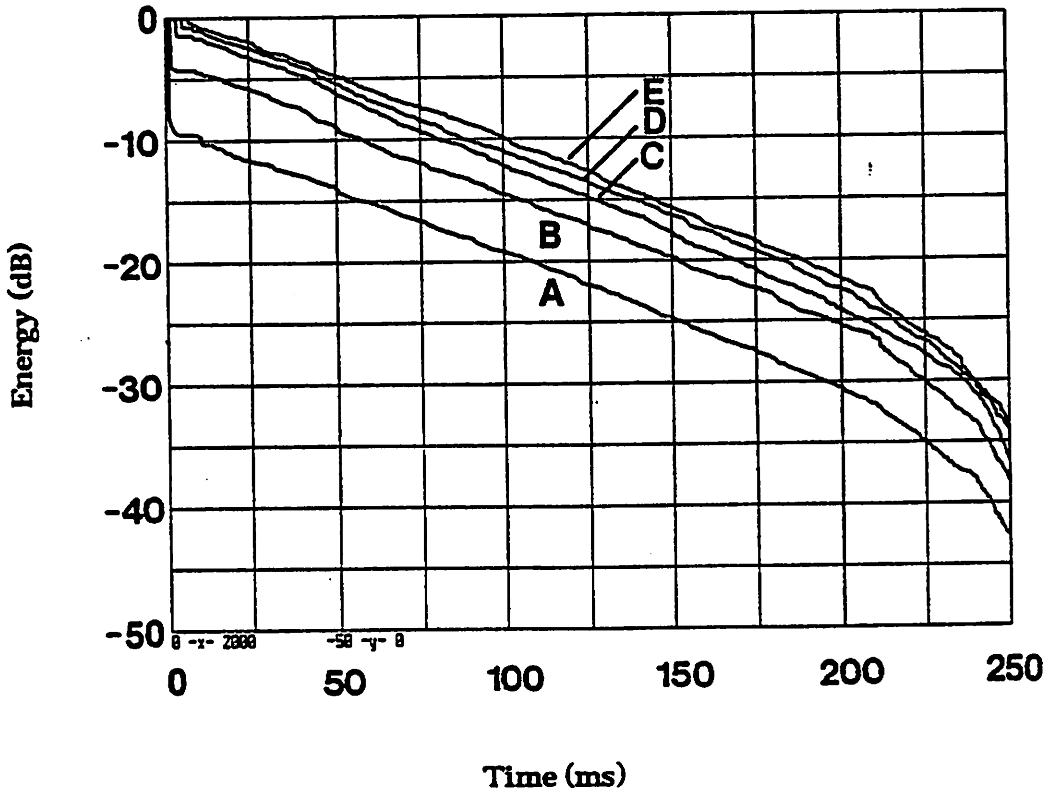


Fig. 3.2.3. Energy decay curves in a room: 10×15×12.5 ft. Wall reflection coefficients were all 0.9. Receiver was at (5. 7. 5.5). Source was at (5.5. 6.75. 5.5) (A). (6. 6. 5.5) (B). (7. 4.5. 5.5) (C). (8. 3. 5.5) (D). and (9. 1.5. 5.5) (E). Sampling rate was 8 kHz.

Table 3.2.1. Reverberation time calculated from the energy decay curves of Fig. 3.2.3.

Source position	Reverberation time (ms)	
	Simulated	Model
A (5.5. 6.75. 5.5)	545	466
B (6.0. 6.00. 5.5)	552	466
C (7.0. 4.50. 5.5)	511	466
D (8.0. 3.00. 5.5)	542	466
E (9.0. 1.50. 5.5)	547	466

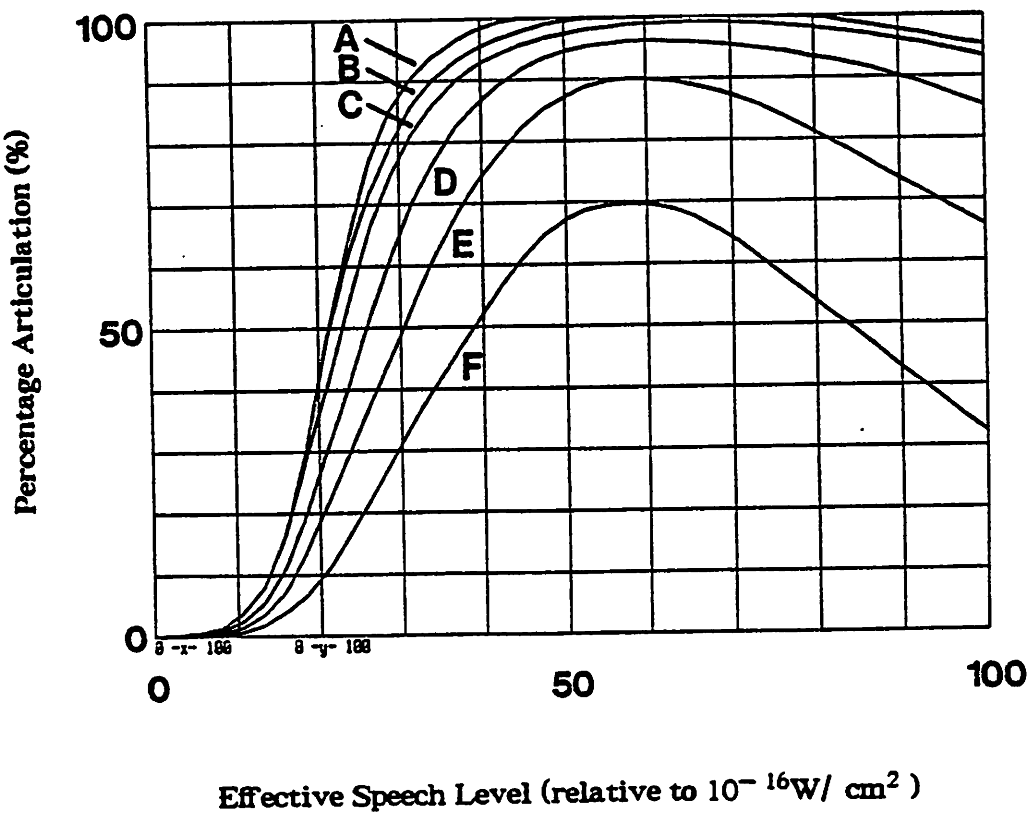
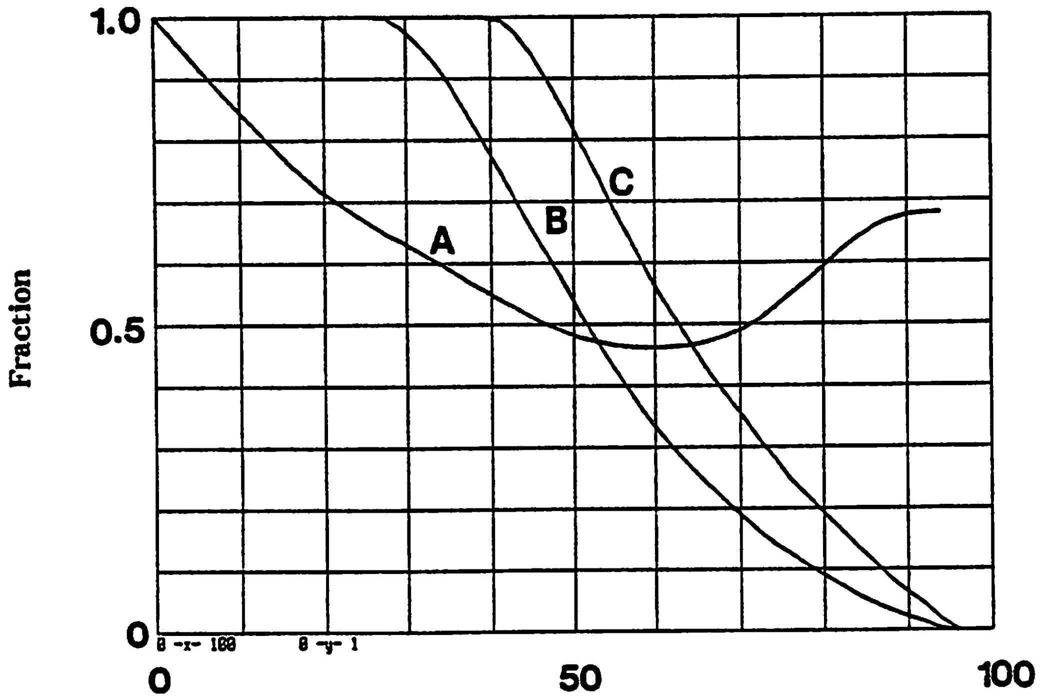


Fig. 4.1.1. Curves of percentage articulation vs effective speech level for signal-to-noise ratio of ∞ (A). 10 dB (B). 5 dB (C). 0 dB (D). -5 dB (E) and -10 dB (F). derived by Lochner and Burger [4.1.1]-[4.1.3].



Time (ms)

Fig. 4.1.2. Fraction of echo integrated with primary sounds for + 5 dB (A). 0 dB (B). -5 dB (C) echoes. determined by Lochner and Burger [4.1.1]-[4.1.3].

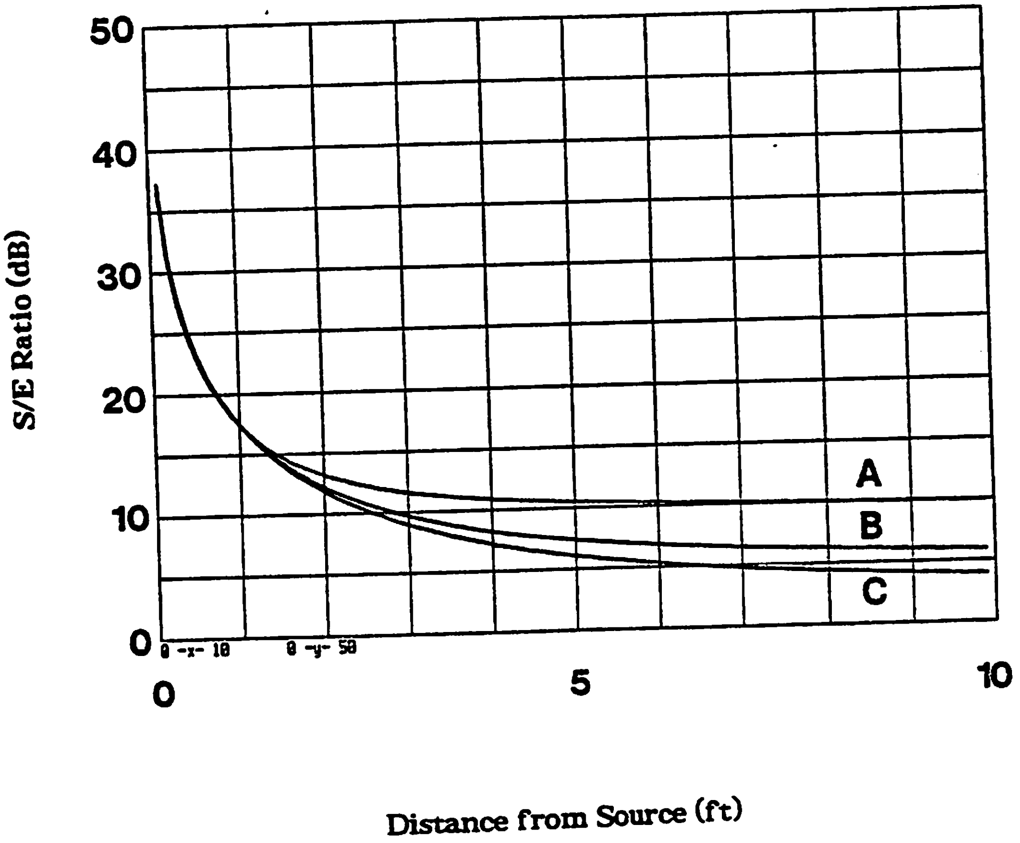


Fig. 4.2.1. Predicted signal-to-echo energy ratio in rooms: 10×15×12.5 ft (A). 20×30×12.5 ft (B). and 20×30×25 ft (C). Wall reflection coefficients were all 0.9.

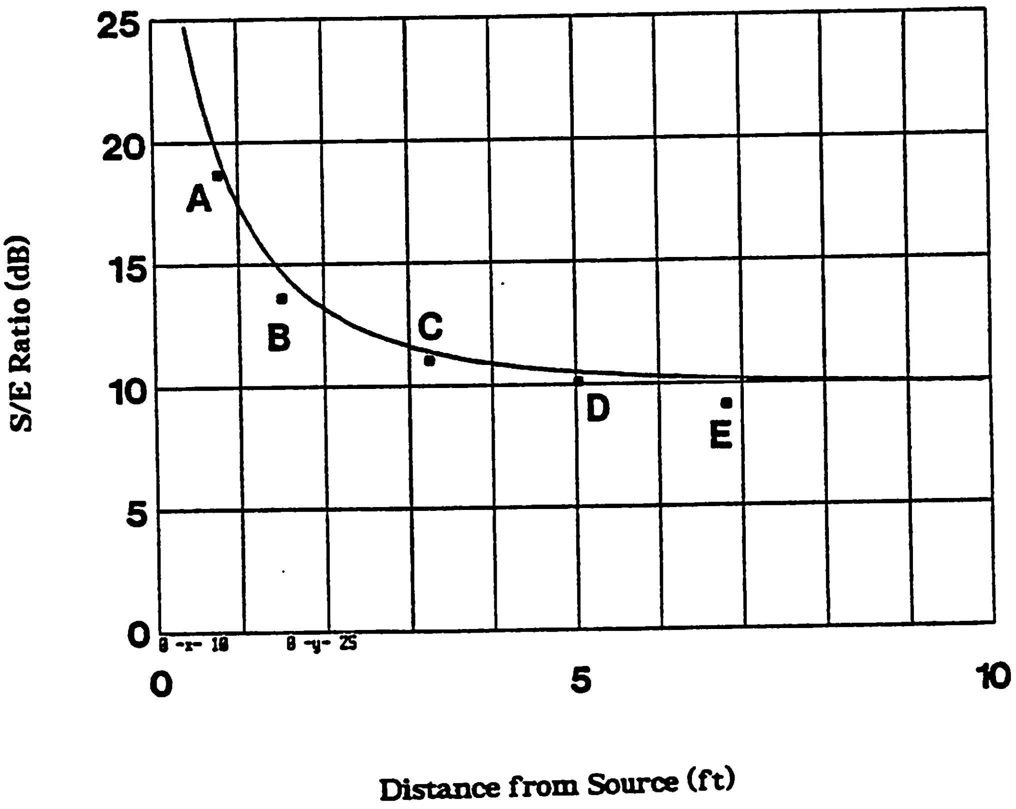


Fig. 4.2.2. Signal-to-echo energy ratios calculated from computer-simulated impulse responses. Room size was 10×15×12.5 ft. Wall reflection coefficients were all 0.9. Receiver was at (5, 7, 5.5). Source was at (5.5, 6.75, 5.5) (A), (6, 6, 5.5) (B), (7, 4.5, 5.5) (C), (8, 3, 5.5) (D), and (9, 1.5, 5.5) (E). Sampling rate was 8 kHz.

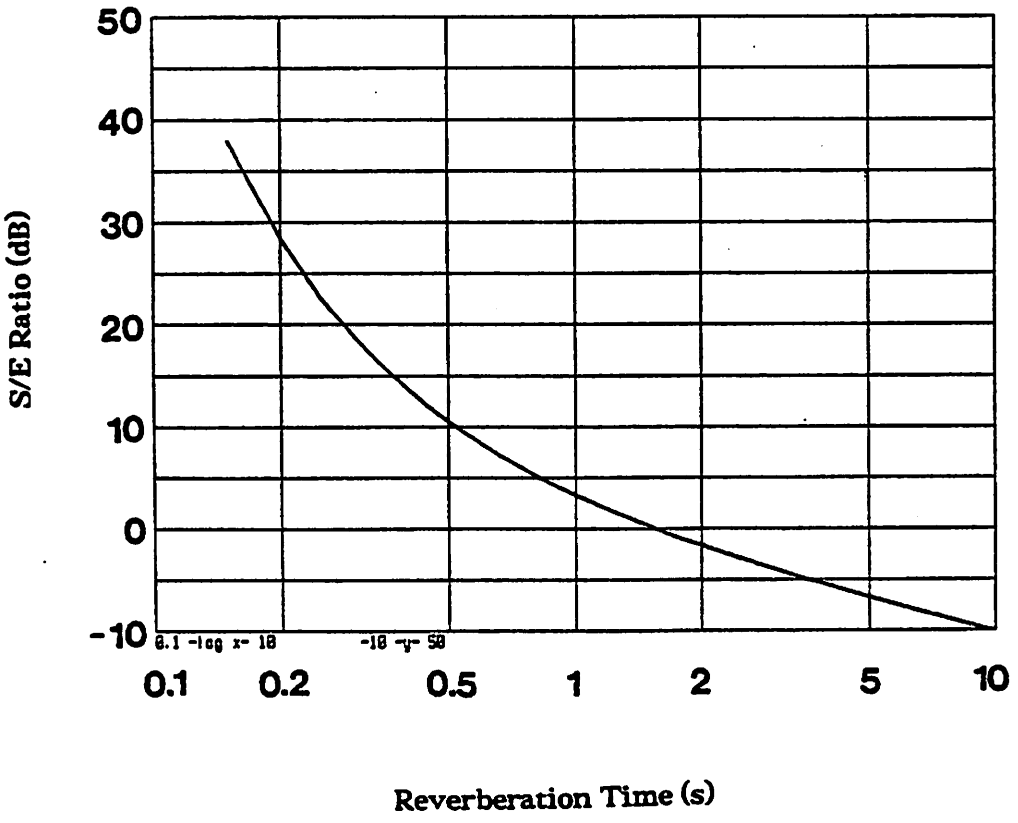


Fig. 4.2.3. Predicted signal-to-echo energy ratios. calculated from (4.2.4) with direct-to-reverberant energy ratio D equal to 0.

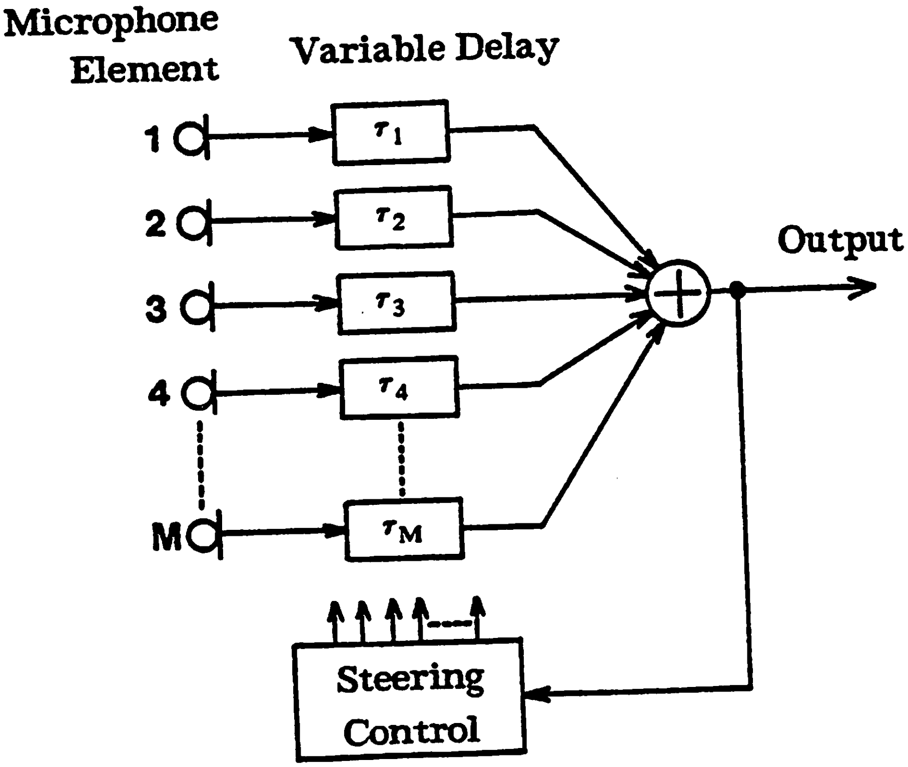


Fig. 5.1.1. Configuration of microphone array.

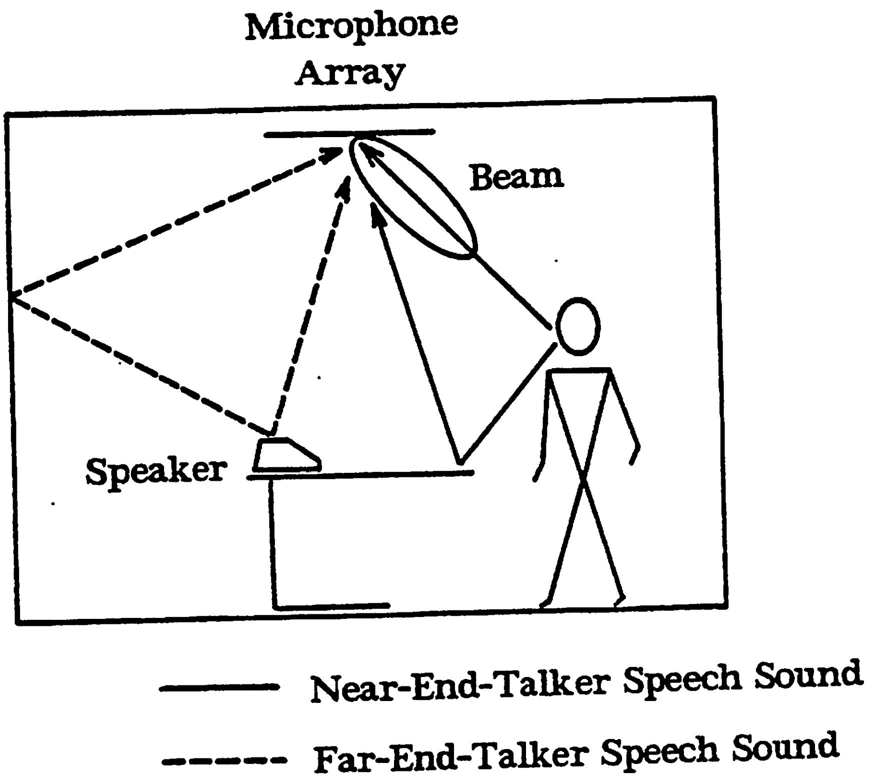

Microphone
Array
Beam
Speaker
—— Near-End-Talker Speech Sound
----- Far-End-Talker Speech Sound

Fig. 5.1.2. Microphone array in an office.

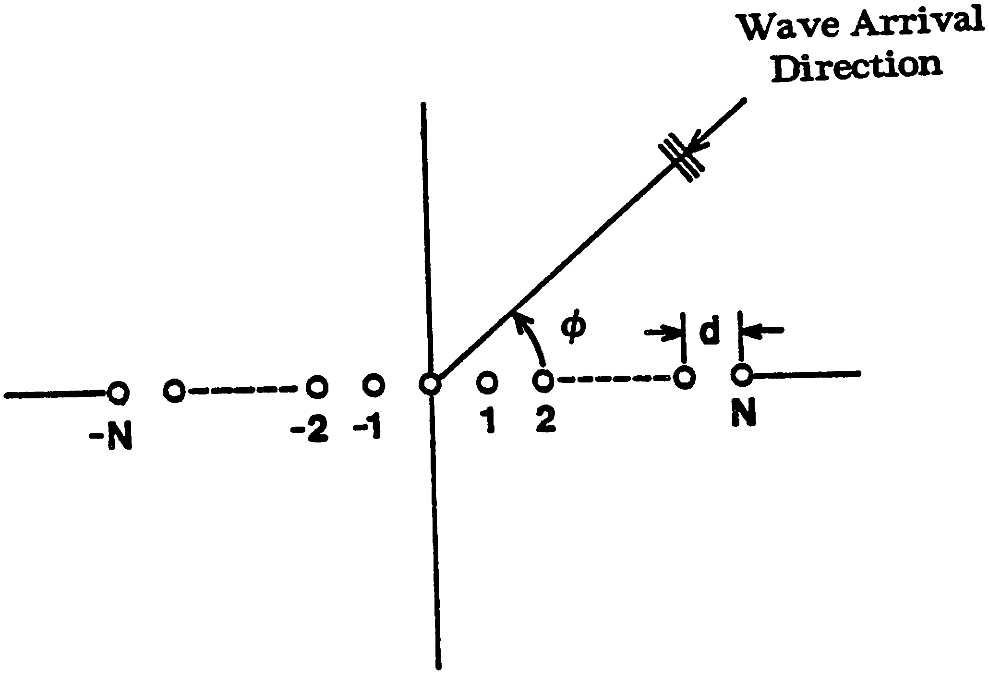


Fig. 5.2.1. Geometry for one-dimensional microphone array receiving a plane wave from the direction (θ ,ϕ) .

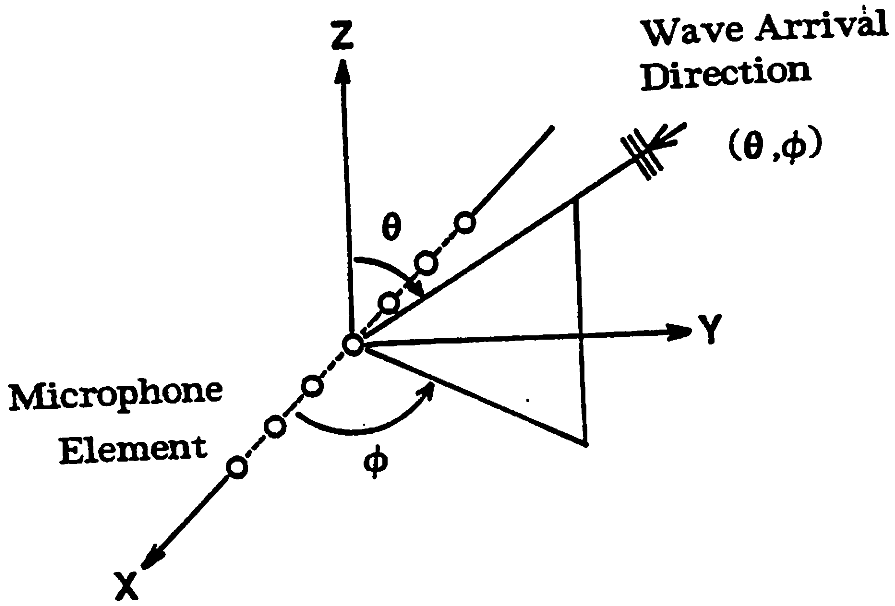


Fig. 5.2.2. Coordinate system for plane wave arriving from the (θ ,ϕ) direction.

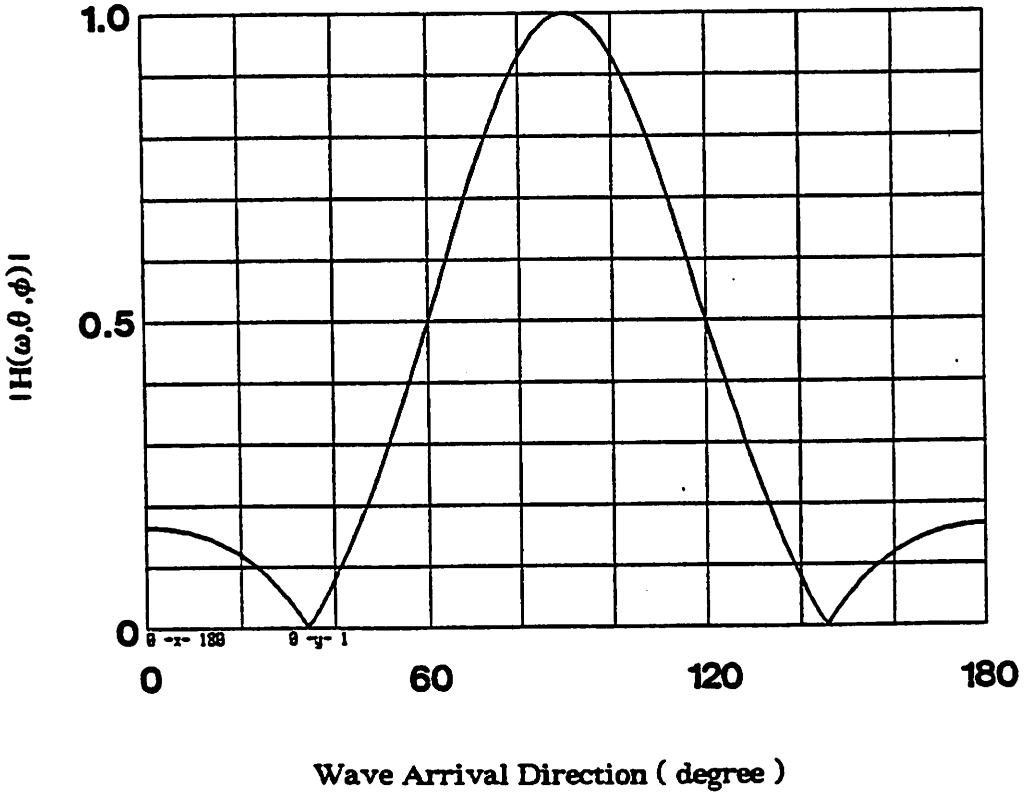


Fig. 5.2.3. Spatial response $|H(\omega,\theta,\phi)|$ vs wave arrival direction ϕ for steering direction $\phi' = 90$ *degree* for frequency $f = 500Hz$. Microphone spacing d was 0.22 ft. Response was calculated from (5.2.4) .

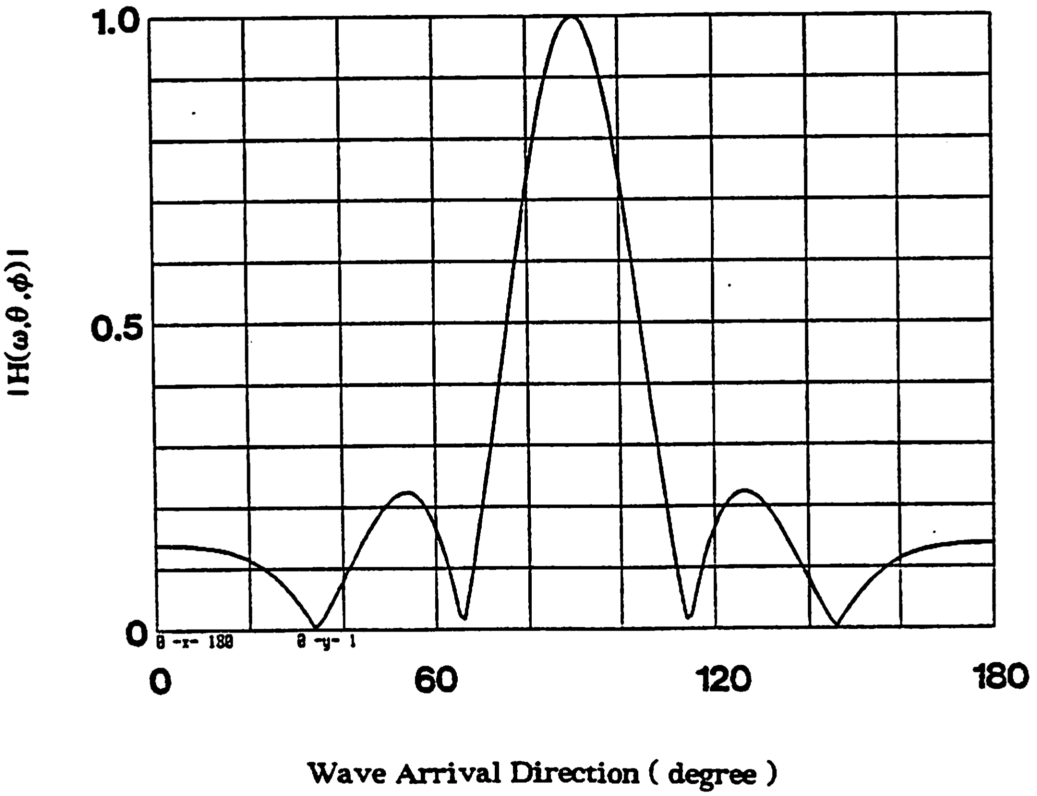


Fig. 5.2.4. Spatial response $|H(\omega,\theta,\phi)|$ vs wave arrival direction ϕ for steering direction $\phi' = 90$ *degree* for frequency $f = 1kHz$. Microphone spacing d was 0.22 ft. Response was calculated from (5.2.4) .

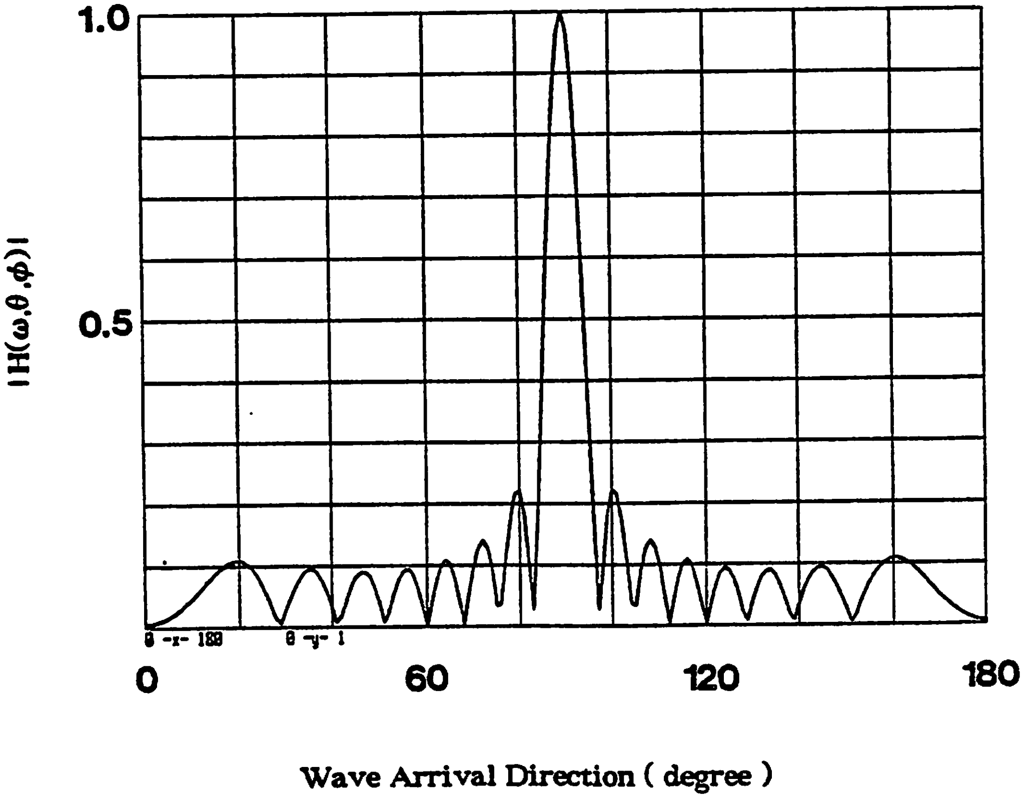


Fig. 5.2.5. Spatial response $|H(\omega,\theta,\phi)|$ vs wave arrival direction ϕ for steering direction $\phi' = 90$ *degree* for frequency $f = 3kHz$. Microphone spacing d was 0.22 ft. Response was calculated from (5.2.4) .

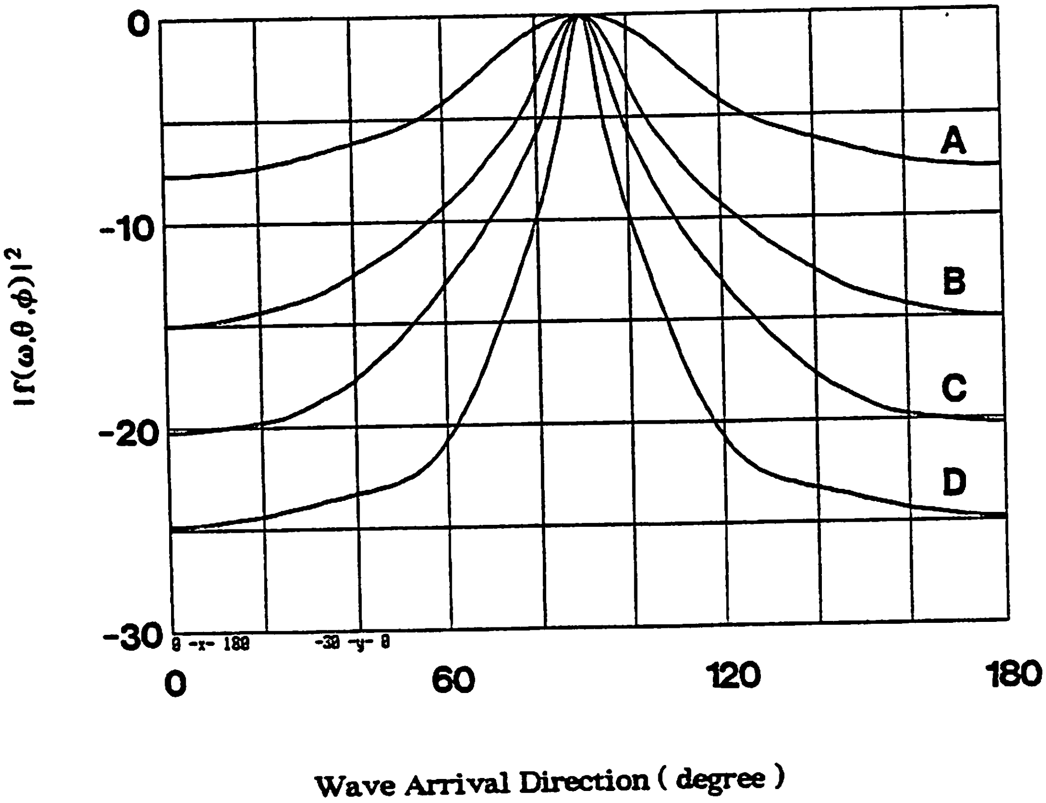


Fig. 5.2.6. Spatial response $|f(\theta,\phi)|^2$ vs wave arrival direction ϕ for steering direction $\phi' = 90$ *degree*. Microphone spacing *d* was 0.22 ft. Number of microphone elements was 3 (A), 7 (B), 11 (C), and 21 (D). Response was calculated from (5.2.5) .

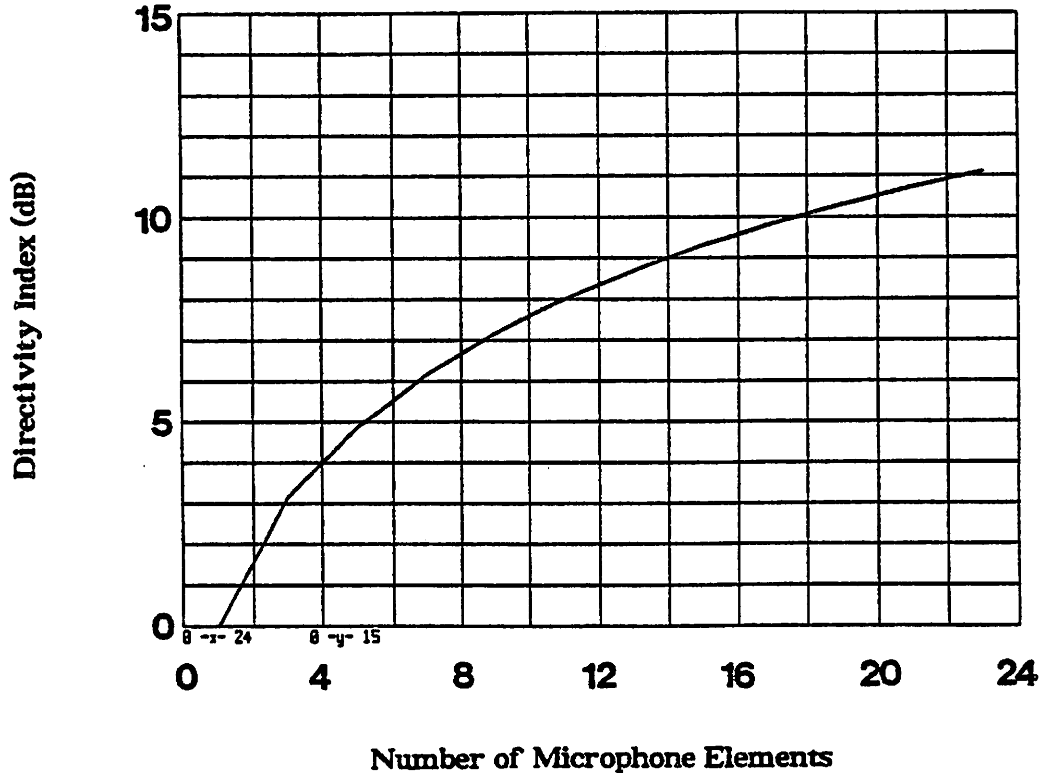


Fig. 5.2.7 Directivity index for steering direction $\phi' = 90$ *degree* . Microphone spacing *d* was 0.22 ft. The index was calculated from (5.2.6) .

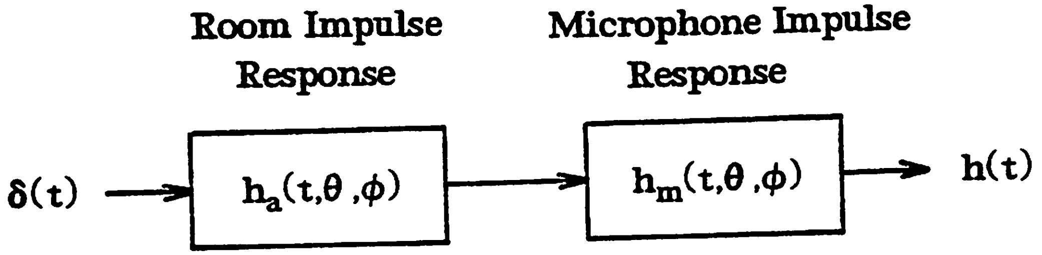


Fig. 5.3.1 Acoustic transmission path from talker to microphone.

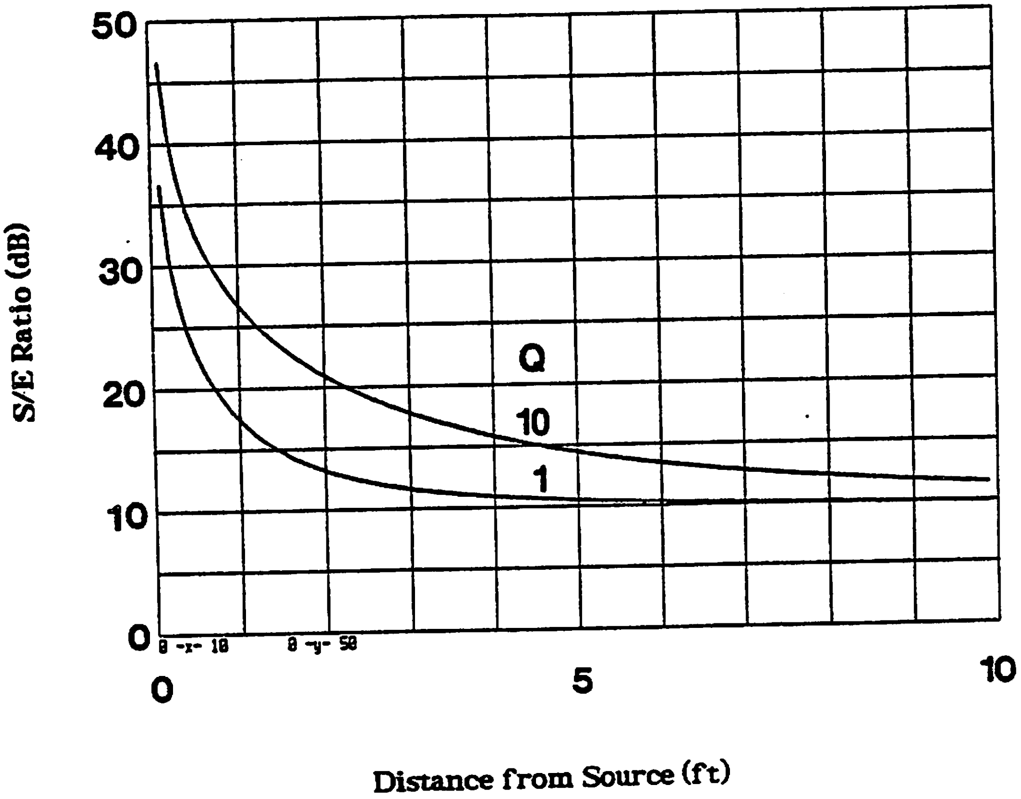


Fig. 5.3.2. Predicted signal-to-echo energy ratio at a directional microphone output with directivity index of 10 dB for a room: 10×15×12.5 ft . Wall reflection coefficients were all 0.9.

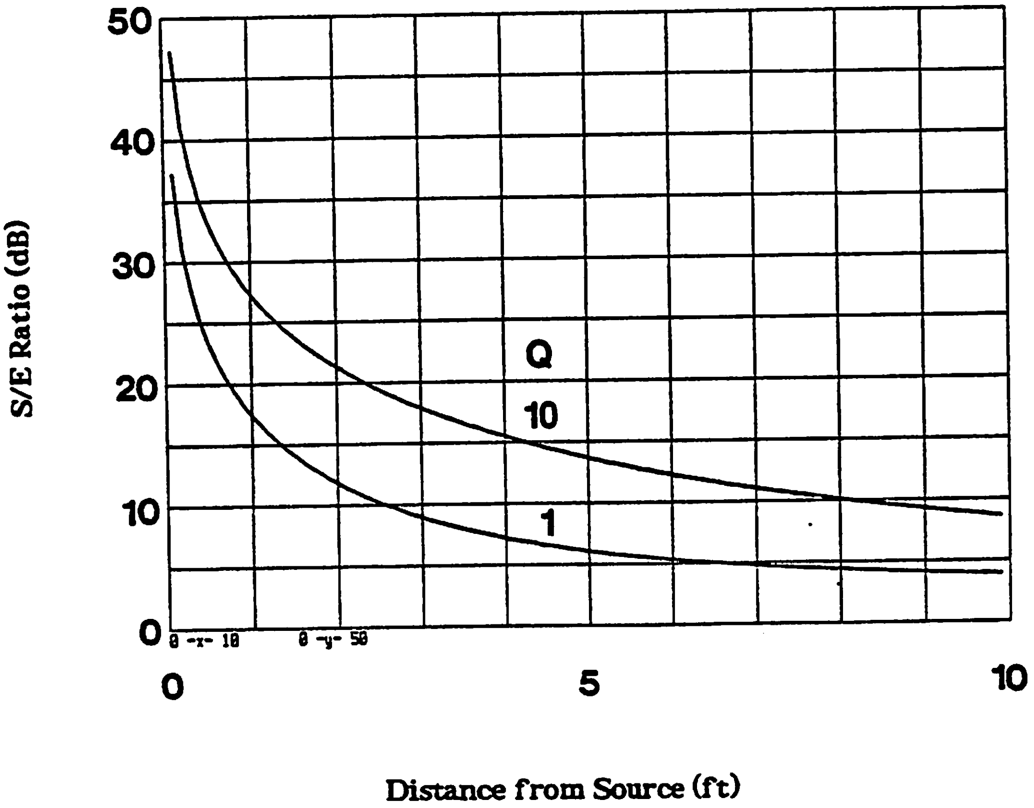


Fig. 5.3.3. Predicted signal-to-echo energy ratio at a directional microphone output with directivity index of 10 dB for a room: 20×30×25 ft . Wall reflection coefficients were all 0.9.

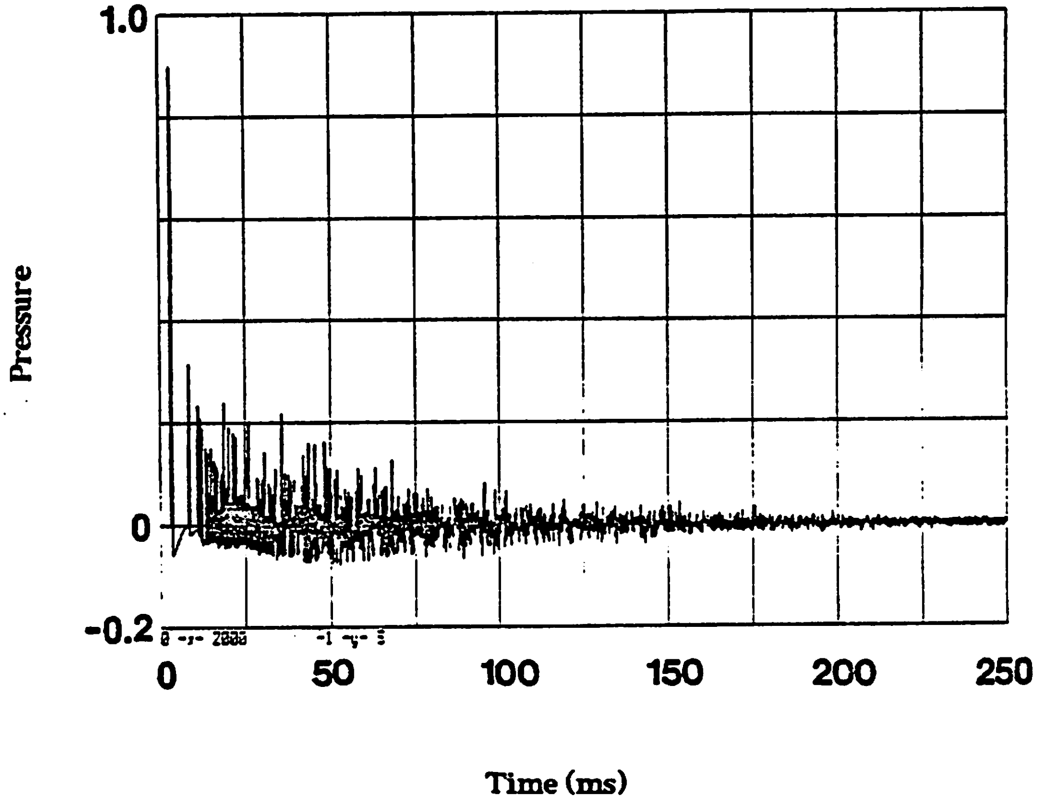


Fig. 5.4.1 Impulse response received by omnidirectional microphone in a room: 10×15×12.5 ft. Wall reflection coefficients were all 0.9. Receiver was at (5, 7, 5.5). Source was at (7, 4.5, 5.5). Sampling rate was 8 kHz.

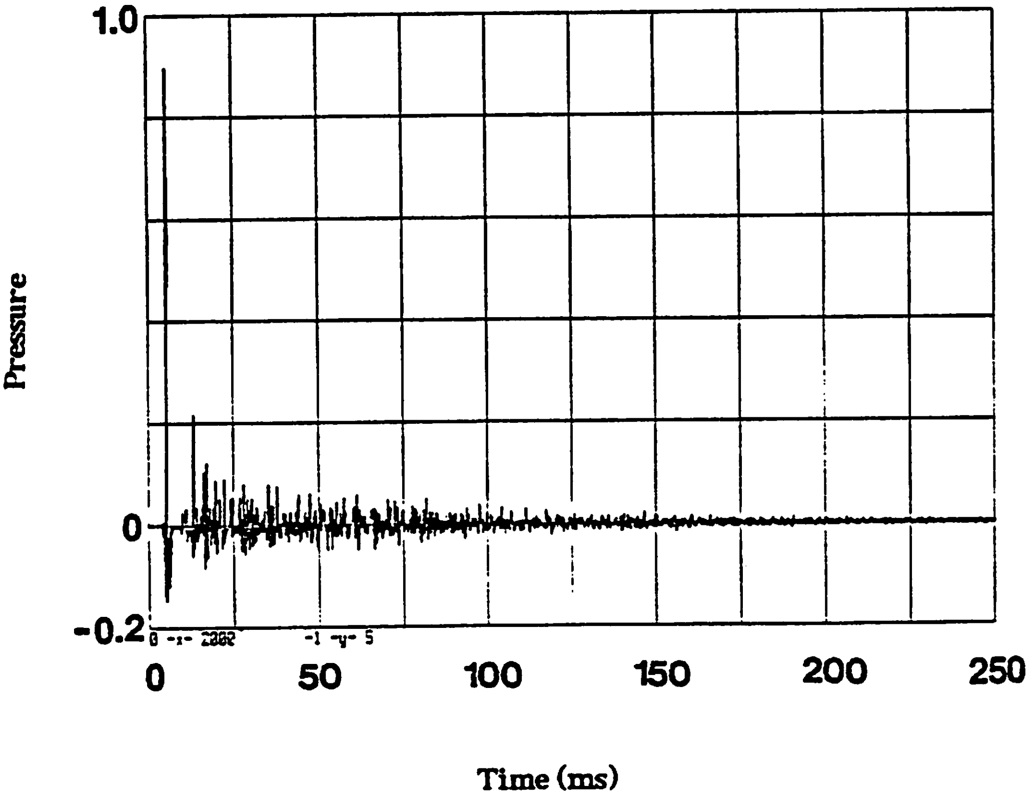


Fig. 5.4.2 Impulse response received by one-dimensional microphone array for the room condition of Fig. 5.4.1. Number of microphone elements was 11 and microphone spacing *d* was 0.22 ft.

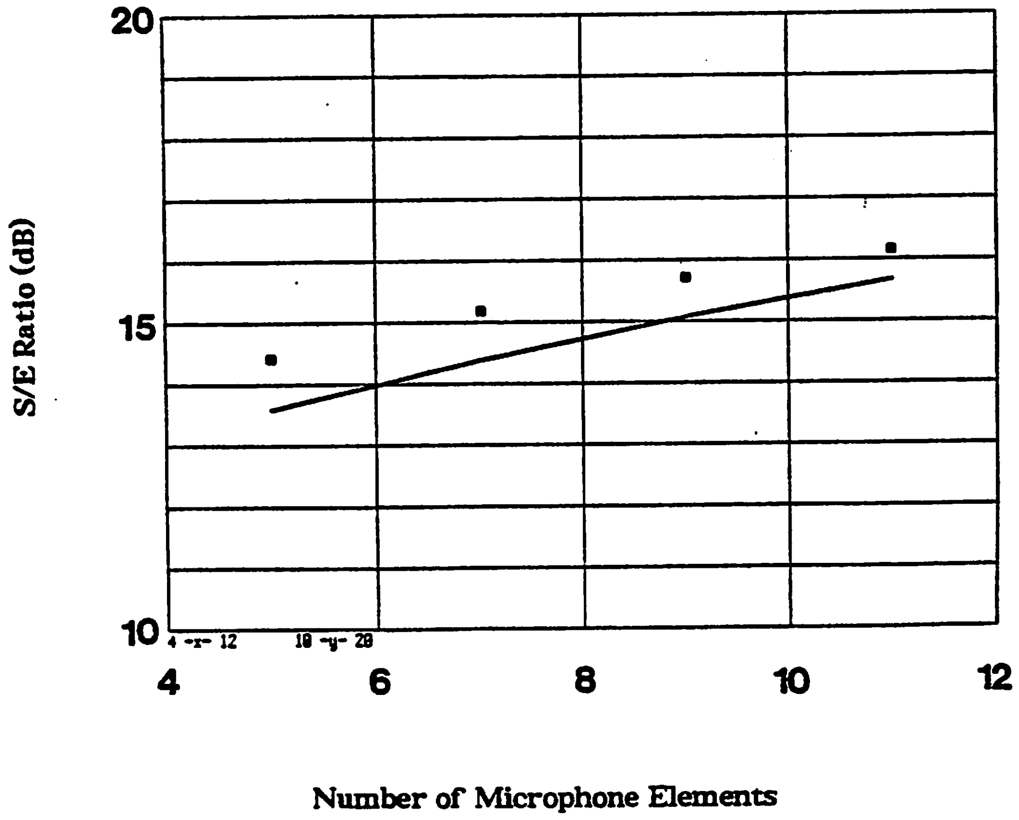


Fig. 5.4.3 Signal-to-echo energy ratio at one-dimensional microphone array output for the room condition of Fig. 5.4.1. Microphone spacing *d* was 0.22 ft.