INFORMATION TO USERS

7923739

KOENIG, ARTHUR HOWARD
DIOTIC AND DICHOTIC PERCEPTION OF SINGLE ECHOES.

CITY UNIVERSITY OF NEW YORK, PH.D., 1979

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs ______
2. Colored illustrations ______
3. Photographs with dark background ✓
4. Illustrations are poor copy ______
5. Print shows through as there is text on both sides of page ______
6. Indistinct, broken or small print on several pages ✓ throughout ______
7. Tightly bound copy with print lost in spine ______
8. Computer printout pages with indistinct print ______
9. Page(s) ______ lacking when material received, and not available from school or author ______
10. Page(s) ______ seem to be missing in numbering only as text follows ______
11. Poor carbon copy ______
12. Not original copy, several pages with blurred type ______
13. Appendix pages are poor copy ______
14. Original copy with light type ______
15. Curling and wrinkled pages ______
16. Other ______

DIOTIC AND DICHOTIC PERCEPTION OF SINGLE ECHOES

by

ARTHUR H. KOENIG

A dissertation submitted to the Graduate Faculty in Speech and Hearing Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1979

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Science in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

March 27, 1979
Date

Chairman of Examining Committee

March 27, 1979
Date

Executive Officer

Anna K. Nabelek

Juergen Tonndorf

David Berkley

Irving Hochberg

Supervisory Committee

The City University of New York

ACKNOWLEDGEMENT

This dissertation represents the culmination of many months of effort. During this period copious doses of guidance, support and encouragement were provided by many individuals.

I would like to express my appreciation to my thesis advisor, Professor Harry Levitt, for the unfailing guidance which he provided and to my faculty committee, Professors Juergen Tonndorf and Irving Hochberg, for their guidance and suggestions. A special note of thanks must go to Dr. David A. Berkley, an outside member of my committee who encouraged a continuous exchange of ideas throughout all phases of the dissertation.

Several other individuals of Bell Laboratories willingly gave of their time in the form of useful discussions including: Drs. James Landwehr, Saul Sternberg, Allen Gersho, Jont Allen, Joseph L. Hall, Jr., and Linda Kaufman as well as Messrs. Ronald Knoll and Charles Lichtenwalner.

The Acoustics Research Department of Bell Laboratories provided the signal processing facility as well as a listening facility in which to run subjects. I would also like to

add a note of appreciation for the cooperation and support provided by the Environmental Health and Safety Department of Bell Laboratories.

Finally, thanks must go to my family and friends for their unstinting support. In particular, a note of very special thanks must go to my dear friend, Jane Schimel.

In Memory Of My Wife
Rosemary Joan Koenig

TABLE OF CONTENTS

List of Tables

List of Illustrations

GLOSSARY

ϕ = power spectrum

$\phi_{\emptyset}$ = average spectral power density

f = frequency in Hertz

τ = delay time in msec between the direct and delayed signal

m = measure of the modulation depth

g = measure of the level of the reflected signal in dB (2Ø log v)

v = amplitude measure of the reflected signal

Δf = frequency spacing in Hertz

$\hat{\mu}$ = estimate of the mean of the cumulative normal distribution

$\hat{\sigma}$ = estimate of sigma (one standard deviation) of the cumulative normal distribution

CHAPTER I

INTRODUCTION

Investigators have long recognized that when two or more sound sources radiate identical or nearly identical signals, the listener focuses on and hears an image located at the nearer source. Indeed, the literature review will show that this phenomenon was recognized over 100 years ago and has been referred to under several different names including: (a) the precedence effect, (b) the Haas effect, (c) the law of the first wavefront, as well as (d) the threshold of extinction effect. The studies which examined this effect, however, were mainly concerned with the issue of localization, although Wallach, Newman and Rosenzweig (1949) did recognize that two distinct processes are responsible for the formation of an auditory percept in a reverberant environment: (1) some form of weighting given to the earliest arriving information, and (2) some form of echo suppression. Undoubtedly both processes play a substantial role in determining the localization of a sound source.

To date, studies dealing with echo suppression have focused on: (a) reinforcement mechanisms related to improved speech intelligibility (e.g. Lochner and Berger, 1964), (b) forward-backward masking paradigms (e.g. Bekesy, 1971); (c) energy integration within critical bands, as well as other possible mechanisms. It has however been recognized that these studies do not adequately explain the relative insensitivity of the human listener to the serious distortion created by the superposition of a direct and reflected signal from the same source.

Several investigators dating back to Huygens (1693, cited in Bilsen, 1968) have observed that when a sound with a delayed version of itself reaches the ear of a listener, a pitch-like sensation is perceived. It would be misleading to suggest that ordinary listening environments have impulse responses which in any way resemble this highly simplified single-reflection case. Indeed, most hard-walled rooms provide multiple reflections which can persist for hundreds of milliseconds. Nevertheless, it is certainly reasonable to use such simplistic models in attempting to understand the underlying mechanisms involved. Atal, Schroeder and Kuttruff (1962) used such a series of simplistic models to examine the perceptibility of one or more reflections. More recently, Bilsen (1968, 1970) and colleagues have published several articles which examine the perceptibility of similar stimuli. Although these studies directly address the question of the perceptibility of single-reflections, the listening modes used were all monaural or diotic.

In a letter to the editor, W. Koenig (1950) (not related to the current author) noted that binaural (presumably dichotic listening) produced a noticeable and substantial degree of echo suppression relative to monaural listening. As reported by Koenig listening with two ears served to reduce the amount of echo and distortion which was perceived. Based on this observation along with a plethora of folklore regarding the advantages of binaural listening in reverberant environments, a preliminary study by Koenig, Allen, Curtis and Berkley (1975) was carried out. The results showed that, at least in the case of a single delay of 2 msec, substantial echo suppression was found in the dichotic listening mode relative to the diotic listening mode. To date, with the exception of a recently published doctoral dissertation, Zurek (1976), no additional data regarding the issue of echo suppression and dichotic listening has been presented. While Zurek's experiments confirm that dichotic listening can serve to reduce perceived echo, the conclusions are somewhat limited due to the use of independent sources for each ear.

CHAPTER II

Review of the Literature -- Binaural Hearing

Sound Localization

In an excellent review of the evolution of the vertebrate auditory system, Masterton and Diamond (1973) strongly support the argument that the auditory system developed selectively so as to maximize the ability of localization. This ability serves both to alert and to permit the appropriate visual orientation of the organism. Decisions such as whether to flee or approach are directly dependent on the auditory localization process. Bear in mind that while the localization experiments reviewed below demonstrate the sensitivity of the binaural auditory system they are all based on some form of binaural fusion where a single image is perceived. Indeed, the concept of localization is based on the localized percept of a single fused image.

Studies of auditory localization have generally been carried out using two distinct listening paradigms: (1) subjects are asked to localize an external source in space; and (2) subjects are asked to indicate the in-head location of an acoustic stimulus presented over headphones. Experiments employing the former paradigm are referred to as localization studies while experiments employing the latter paradigm are referred to as lateralization studies. Although localization experiments are more realistic, lateralization experiments afford the experimenter substantially more control over stimulus parameters. Jeffress and Taylor (1961) have demonstrated that subjects' ability to identify the direction of a source presented over headphones to be equivalent to results obtained with external source(s).

Physical Measurements

Lord Raleigh (1907) an early investigator in the area of binaural localization proposed the concept of a duplex theory whereby low frequency signals are localized on the basis of interaural time differences while high-frequency signals are localized on the basis of interaural intensity differences. This theory provided the impetus for much of the work which was to follow.

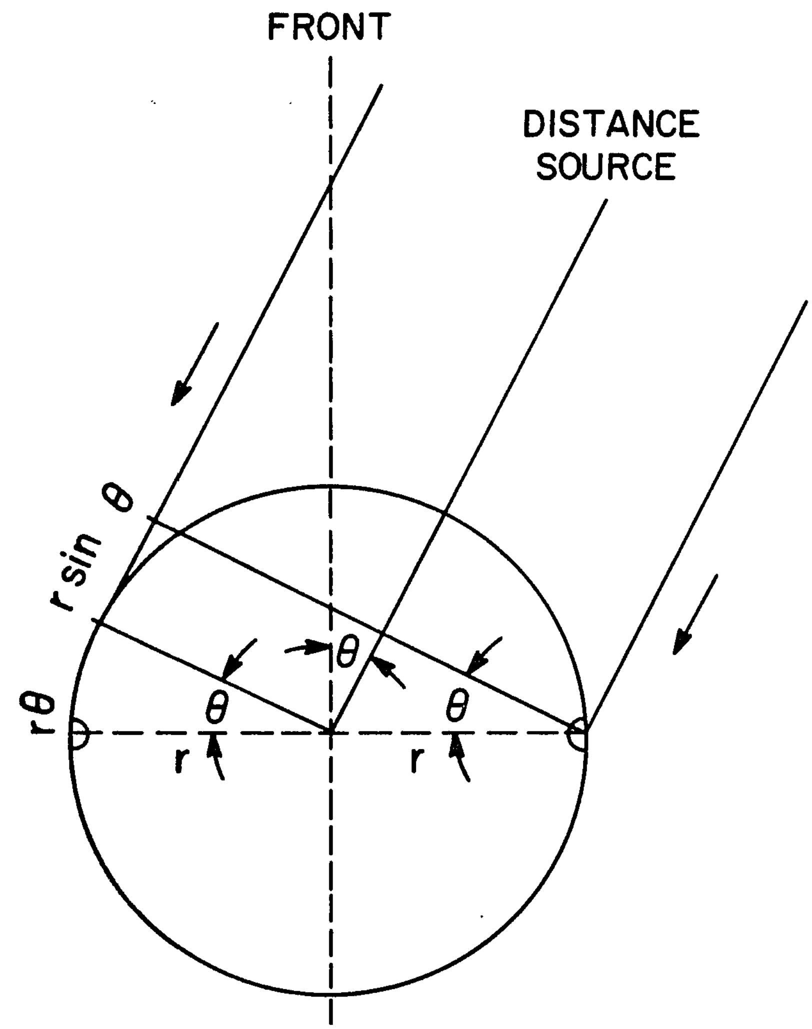
The physics of the problem becomes apparent when time and intensity differences which must exist when a sinusoidal source is displaced to one side of the head are considered. Low frequencies which have large wavelengths relative to the width of the head produce negligible intensity differences at the two ears since the sound wave will diffract or bend around the head. No shadows are produced in this instance.

On the other hand, high frequencies which have small wavelengths relative to the width of the head, produce substantial level differences between the two ears due to the shadow cast by the head. Recall that a 500 Hz sinusoid (low frequency) has a wavelength of 0.68 m while a 6000 Hz sinusoid (high frequency), on the other hand, has a wavelength of 0.0567 m.

Using a physical model, Woodworth and Schlosberg (1954) were able to examine the relationship of head width, wavelength and time differences. If the head is treated as a uniform sphere with two point receivers located on opposite sides as shown in Figure 2.1, the geometry becomes obvious. When the source lies in the median plane, the path or distance to both ears (from the point source) must be the same. If the source is moved away from the median plane, the sound wave must travel a longer distance to one ear than to the other ear. This distance is frequently labeled the binaural distance difference. If we operationally define a median source as representing 0° incidence, any offset in degrees from this plane can be represented by the direction angle or azimuth, θ. That is, θ is defined as the angle created by the source and the median plane of the listener. The difference in the length of the sound path is given by the formula

$$d = r(\theta+\sin \theta) \qquad (2.1.1)$$

Figure 2.1 Binaural distance difference when the source of sound is more than a few feet away so that the waves pass in practically parallel lines to the two ears. The construction lines show the geometry of the formula, $D = r(\theta + \sin\theta)$. (From Woodworth and Schlosberg, 1954).


FRONT
DISTANCE SOURCE
r sin θ
rθ
θ
θ
θ
r
r

where r is the radius of the sphere (head) and θ is the angle of incidence in radians. If we assume the radius of the head to be approximately air to be 344 m/sec. the time difference ($\triangle$t) required to reach the two ears is given by equation

$$\triangle t = 254(\theta + \sin\theta) \tag{2}$$

where $\triangle$t is in milliseconds (msec). If the source is straight ahead ($\theta = 0$), there will be no time difference between the two ears. If the source is located at 90°, however, where $\theta = \pi/2$, the time difference reaches a maximum of 653 µsec. This formula is an approximation in that it neglects reflections from the head as well as the interaction of these reflections with the wave front going directly to the two ears. Furthermore, this formulation assumes that a phase delay of more than one cycle is ambiguous, since we can not determine whether the right ear, for example, is leading or following by one cycle. Pure tones with periods of less than approximately 1 msec therefore become unreliable indicators of location (Green, 1976).

Fedderson, Sandel, Teas, and Jeffress (1957) have confirmed Woodworth and Schlosberg's calculations using probe microphones on live human heads. A click source was delivered by a small loudspeaker mounted on a 7 foot (2.13 m) boom which pivoted directly above the subjects head. Measurements were made at each 10° position around the subject as well as at $\pm 5^{\circ}$ and $\pm 175^{\circ}$ from the median plane. The results shown in Figure 2.2 for interaural time differences (ITD's) agree quite well with Woodworth's formula. Since the head can act as a reflector, we can expect interaction effects (reflected wave with direct wave) for continuous signals at certain frequencies. Obviously this is not the case when sharp transients are used.

Interaural intensity differences were also measured (Fedderson et. al. 1957) at the two ears as a function of source position. As shown in Figure 2.3, these results clearly demonstrate the lack of interaural intensity differences for low frequencies with large wavelengths and significant interaural intensity differences ($\sim$ 20dB) at high frequencies where the wavelengths are small relative to the size of the head. In addition, this figure demonstrates the existence of front-back asymmetry in intensity difference.

Localization of Sound Sources

Stevens and Newman (1936) provide data regarding the localization of actual sound sources. This study was conducted on the roof of a building with the subject positioned in such a way as to minimize reflections from nearby surfaces. This was accomplished by seating the subject on a swivel chair approximately 9 feet (2.74 m) above the roof. A speaker was mounted on a 12 foot (3.66 m) boom which was capable of rotating 360° in the horizontal plane. The task

Figure 2.2 Interaural time differences as a function of the aximuth position of the source. (From Fedderson, Sandel, Teas and Jeffress, 1957).

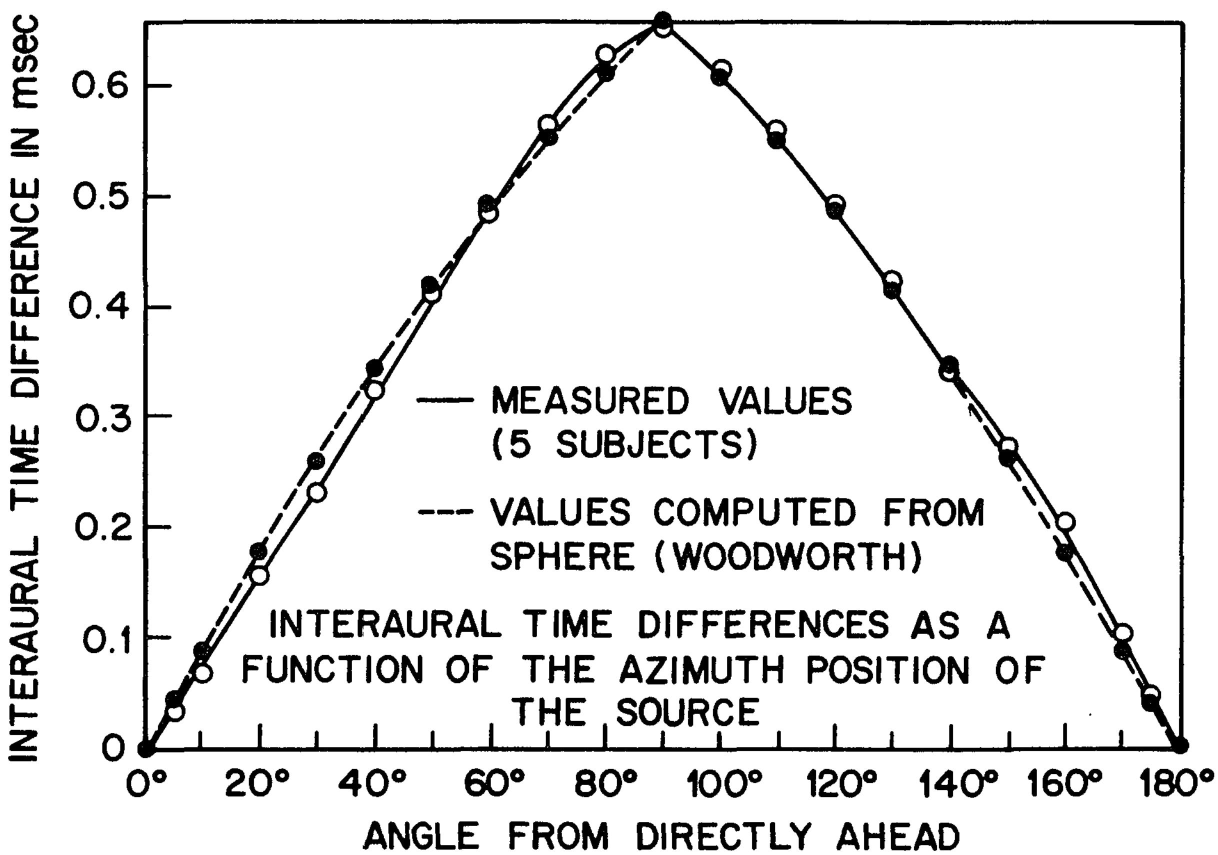
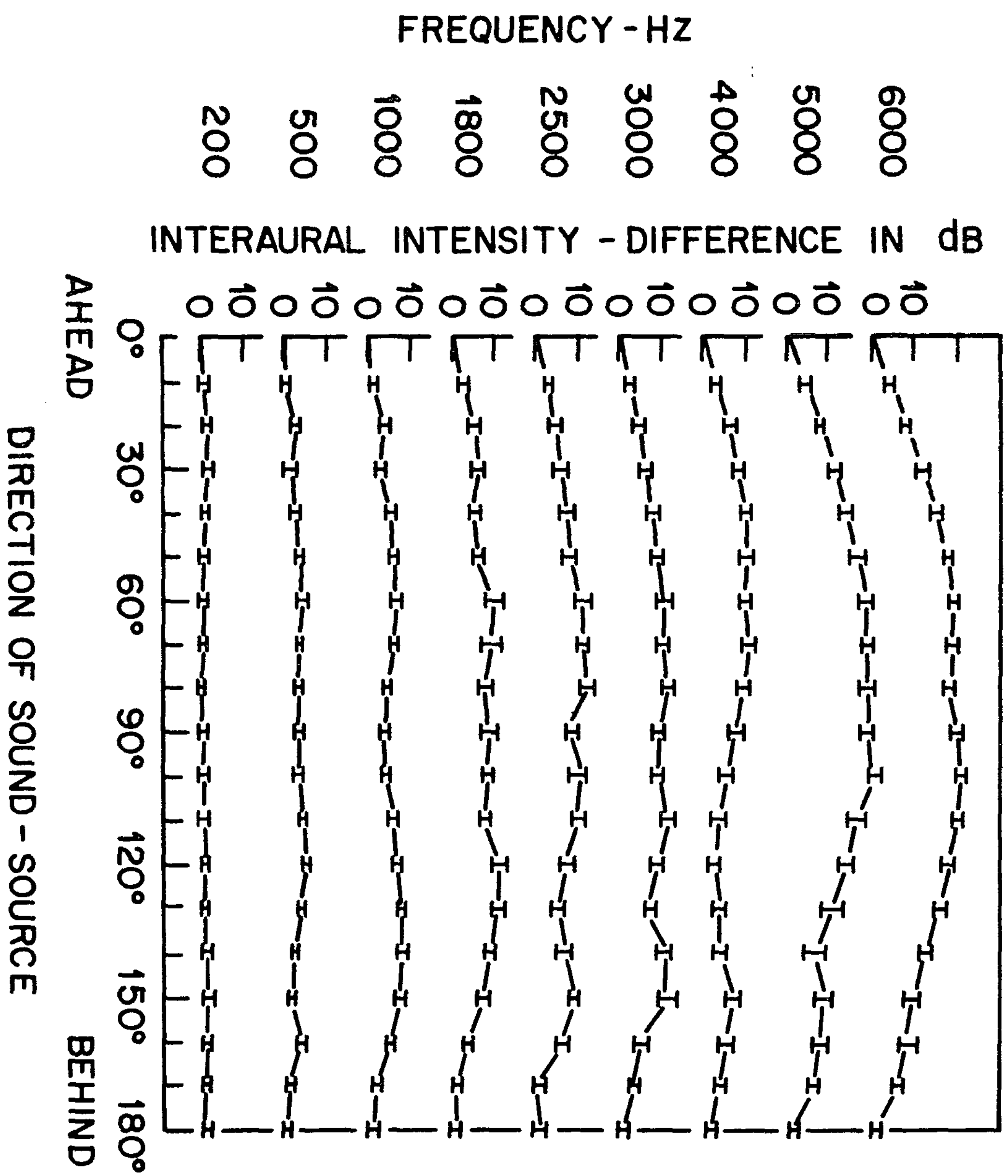

INTERAURAL TIME DIFFERENCE IN msec
0.6
0.5
0.4
0.3
0.2
0.1
0
— MEASURED VALUES (5 SUBJECTS)
--- VALUES COMPUTED FROM SPHERE (WOODWORTH)
INTERAURAL TIME DIFFERENCES AS A FUNCTION OF THE AZIMUTH POSITION OF THE SOURCE
0° 20° 40° 60° 80° 100° 120° 140° 160° 180°
ANGLE FROM DIRECTLY AHEAD

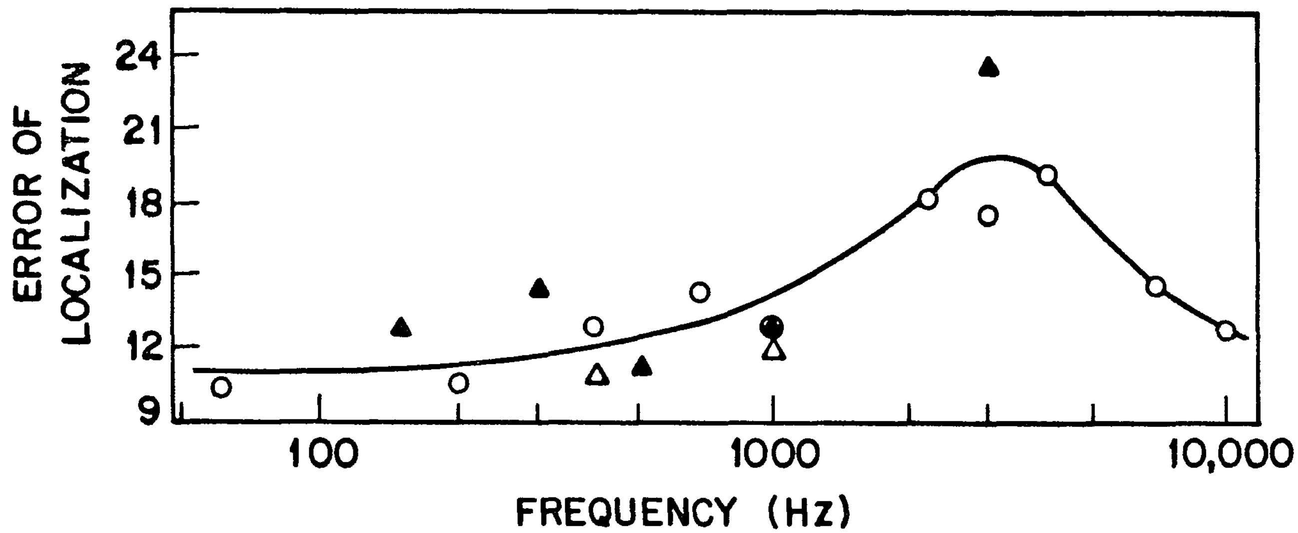
Figure 2.3 Interaural intensity-differences as a function of frequency and direction (five subjects). (From Fedderson, Sandel, Teas and Jeffress, 1957).


FREQUENCY - HZ
200
500
1000
1800
2500
3000
4000
5000
6000
INTERAURAL INTENSITY - DIFFERENCE IN dB
0
10
AHEAD
BEHIND
0°
30°
60°
90°
120°
150°
180°
DIRECTION OF SOUND - SOURCE

of the listener was to report the direction of the source in a horizontal field to the nearest 15°. Three different sound sources were used: (1) tones with dampened onset-offset properties; (2) clicks; and (3) a hiss-like sound. The results shown in Figure 2.4 indicate that errors of localization are relatively constant at low frequency, increase at 3 kHz and continue decreasing after 4 kHz up to the maximum frequency tested. The authors also reported that errors in localization were smallest near the median plane and increased as the tone was moved away from the median plane to one side. Many front-back confusions were also noted. These data were taken as support of the Duplex theory of sound localization, which maintains that two sets of cues are operative. Temporal cues are used for low frequency localization and intensive cues are used for high frequency localization. For middle-range frequencies neither mechanism is efficient thereby producing a maximum number of errors.

A follow-up study was undertaken by Sandel, Teas, Fedderson and Jeffress (1955) in an anechoic chamber. Subjects were asked to indicate the location of tone sources from three different positions: 0° and plus or minus 40° from the median plane. They were to adjust a second sound source (noise) to the position where the tone was localized. All tones had 50 msec rise and decay times to minimize transient effects, and were presented at a 35 dB sound pressure

Figure 2.4 Dependence of localization on frequency. The ordinate represents the average errors in degrees made by both subjects. (From Stevens and Newman, 1936)



ERROR OF LOCALIZATION
24
21
18
15
12
9
100
1000
10,000
FREQUENCY (HZ)

level (SPL). No attempt was made to control loudness. The results obtained from five subjects indicated the existence of large biases to underestimate the deviation of the source from 1500 Hz to 3000 Hz; however, when the data were replotted with the bias effects removed, localization was found to be most uncertain at 1500 Hz. That is, the maximum absolute error occurred in the frequency range of 1500 Hz to 3000 Hz, while the maximum relative error occurred at 1500 Hz. Although there are some differences between these results and those reported by Stevens and Newman, both experiments lend strong support to the Duplex theory of localization. In summary, cues for interaural time and intensity differences that are relevant to the localization of pure tones are strongly related to what we would predict from the physical nature of both cues and the human listener. Intensity cues are more important for high-frequency localization where wavelengths are small relative to head size, and phase or time differences are important for low-frequency tones below 1500 Hz where almost no intensity difference between the two ears exists but where significant phase differences can be measured.

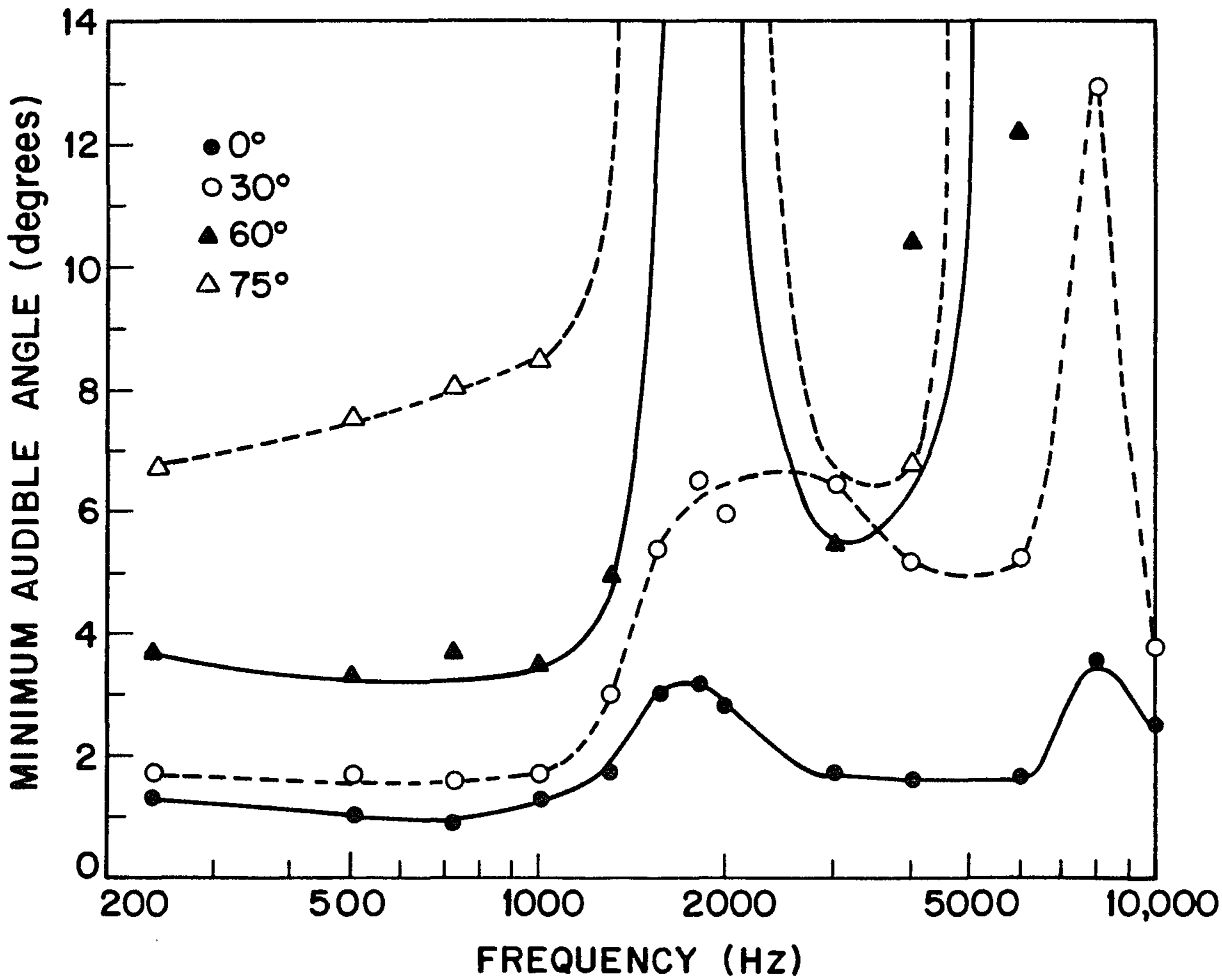
Mills (1958, 1972) determined the minimum change in a sound source that a subject could readily detect as a function of the initial position of the source. The measurement was labeled as the minimum audible angle (MAA) since it represented the minimum angle that a subject could detect.

MAA judgements serve to characterize the relative precision, or resolution, of auditory localization. Measurements were carried out in an anechoic chamber with the subject's head restrained to prevent uncontrollable head movement artifacts. Successive tone pulses were presented from two slightly different azimuths. Subjects had to indicate whether the second stimulus came from the right or left of the first. The smallest changes which subjects could detect 75% of the time were considered to represent the MAA for the specific frequency being investigated. At 0° azimuth (the median plane) the MAA was small (approximately 1° resolution for low and high frequencies and large for frequencies between 1500 and 2000 Hz. A decrease in sensitivity was noted again at 9 kHz. For all frequencies, the MAA increased as the reference azimuth changed from 0° to 75°. For tones between 1500 and 2000 Hz from sources at azimuths of more than 45°, the MAA was indeterminately large. These results are summarized in Figure 2.5. In summary, these data serve to confirm the lack of sensitivity for frequencies between 1500 and 3000 Hz and thus support the Duplex Theory. The marked decrement in sensitivity for non-median plane judgements also is noteworthy.

Lateralization of Signals

Klumpp and Eady (1956) measured interaural time differences (ITD) thresholds for stimuli presented over headphones. The stimuli used included: (1) noise - both

Figure 2.5 The smallest change in the angle of a source that is just detectable as a function of frequency. The initial angles of the source is indicated in the key. (From Mills, 1958).



14
12
10
8
6
4
2
0
MINIMUM AUDIBLE ANGLE (degrees)
0°
30°
60°
75°
200
500
1000
2000
5000
10,000
FREQUENCY (Hz)

broad-band as well as 5 different filter band-pass configurations with frequency content varying from only low-frequency (410-440 Hz) to only high-frequency (3056-3344 Hz); (2) clicks - a single click as well as a 2-second burst containing 30 clicks; and (3) tones ranging in frequency from 90 Hz to 3200 Hz. All stimuli were presented for a 2-second period with 300 msec rise and decay times. The results showed that broad band noise stimuli with low-frequency content have ITD's in a range of 9 to 19 µsec with the 150-1700 Hz band-passed noise exhibiting the most sensitive ITD threshold of 9 µsec. The single click ITD was 28 µsec while the repeated click ITD was substantially lower at 11 µsec. The ITD's for tones ranged from 11 to 75 µsec with the 1000 Hz tone exhibiting the lowest ITD for tones. In summary, Klump and Eady data show that (a) ITD's are lower (more sensitive) for stimuli with low frequency content vs higher frequency content, (b) band-pass configurations containing low frequency noise stimuli exhibited the lowest ITD, (c) ITD's are lower for multiple click vs single click presentations.

An interpretation of these results suggested by Moore (1977) and supported by Tobias and Zerlin (1959) is that since the noise stimulus varies continuously with time, it provides continuous transient information many times during the presentation of the stimulus; the pure tone for a particular frequency, on the other hand, provides only information regarding ongoing phase differences. The single click only provides one-time transient information. Furthermore, Moore ascribes the excellent low frequency tone performance of 11 μsec (only 2 μsec less sensitive than noise) to the long duration (1.4 sec) tone pulses.

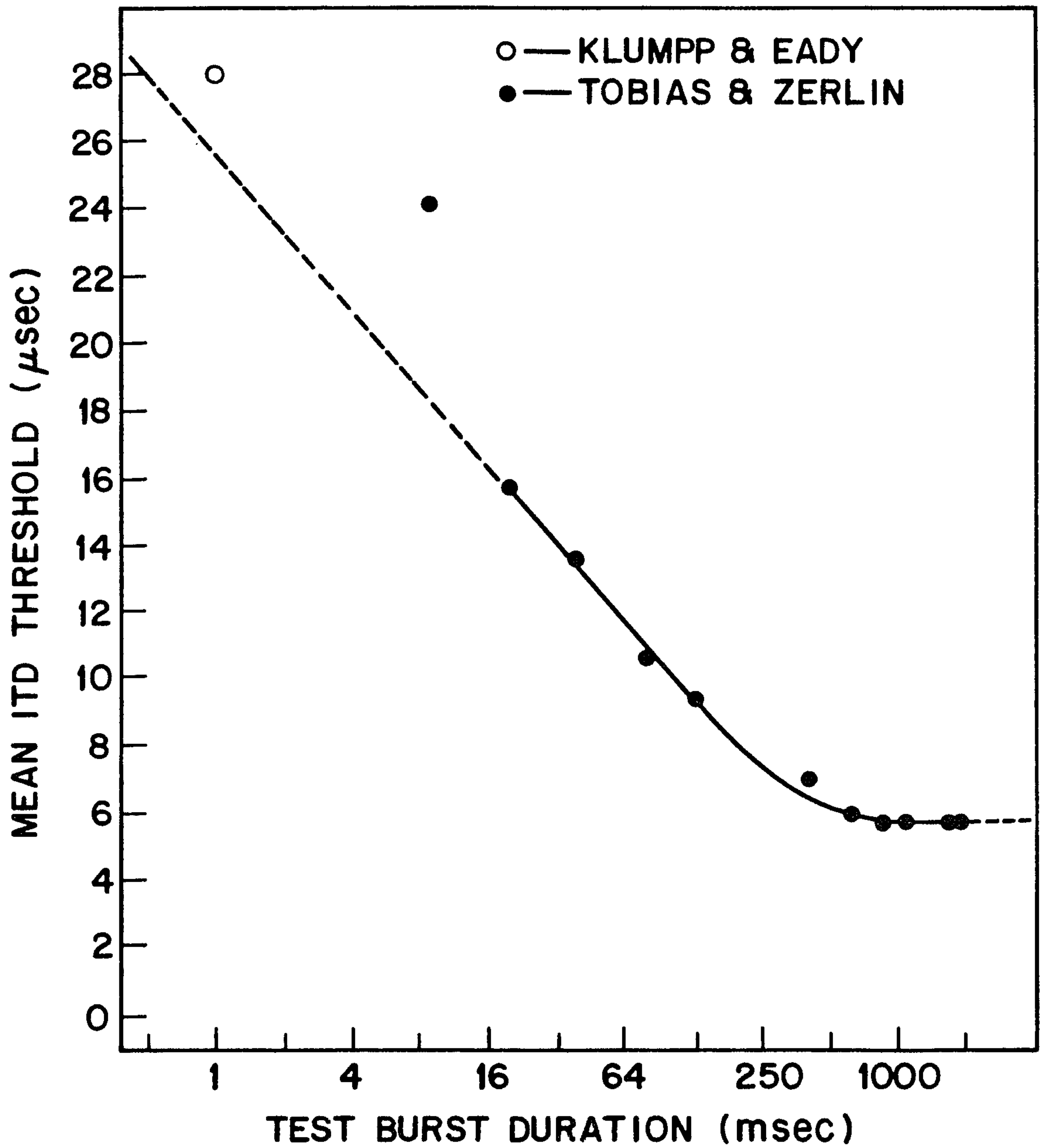
Zwislocki and Feldman (1956) report the results of a somewhat similar set of measurements. Pure tone pulses 1 sec in duration were used with 50 msec rise and decay times. Subjects were presented with a pair of tone pulses over headphones such that the first of the pair was always in the median plane while the second pulse was phase-shifted. The subjects had to indicate whether the second pulse in the pair appeared to the right or left of the first one. Both stimulus frequency and sound pressure levels were varied. Their results indicate that phase discrimination presented in degrees decreases at the low and high sensation levels; at a comfortable listening level, the just-noticeable difference (JND) reached a minimum of 2° for 250 Hz. Furthermore, the JND values also decreased as frequency increased. The highest frequency at which dichotic phase differences could be reliable determined was 1250 Hz.

Tobias and Zerlin (1959) directly addressed the issue of ongoing ITD discussed above. The study was directed at determining the relationship between stimulus duration and ITD for white noise stimuli. Each trial consisted of two binaurally presented noise bursts, the first of which was

1.4 sec in duration and appeared simultaneously at both ears. This stimulus essentially served as a reference stimulus and was designed to identify the intracranial center for the subject. The second burst, presented after a 100 msec interburst interval, contained a randomly placed ITD ranging from 0 to 40 µsec, and was randomly placed to the right or the left of the reference stimulus. The subjects' task was to indicate whether the second burst occurred to the right or left relative to the first reference burst. The duration of the second burst was varied from 0.01 sec to 1.94 sec.

Their results clearly demonstrated that when mean ITD thresholds are plotted as a function of noise burst duration (log scale) (see Figure 2.6), there is a linear decrease in the thresholds as duration increases. At a noise burst duration of 700 msec the ITD threshold curve reached an asymptote with a maximum sensitivity level of 6 µsec. The authors concluded that an ongoing disparity cue increases judgemental precision when judging movement away from a median plane and speculated that this effect is due to a central trace pattern mechanism which can store the interaural disparity information and scan it for a period not exceeding 700 msec. Whatever the mechanism, the ability of the binaural auditory system to detect ITDs in the order of 6-10 µsec must be considered remarkable, particularly when one considers synaptic latencies to be in the order of

Figure 2.6 ITD threshold as a function of the stimulus duration. Note that the curve reaches asymptote at about Ø.7 sec (Tobias and Zerlin, 1959). The point at 1 msec is from a report by Klumpp and Eady (1958).



KLUMPP & EADY
TOBIAS & ZERLIN
MEAN ITD THRESHOLD (μsec)
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
1
4
16
64
250
1000
TEST BURST DURATION (msec)

100 μsec.

Tobias and Schubert (1959), attempted to determine the trading relationship which exists for interaural transient disparity (ITD) and ongoing disparity. Gaussian noise was gated to one earphone slightly ahead of the other earphone. This difference, the ITD was varied from 0 to 400 μsec. Interaural ongoing disparity was achieved by passing the noise waveform through a delay line so that the fine structure of the noise arrived at the ears with a temporal disparity. This disparity ranged from -35 to 170 μsec and was always in the opposite direction of the interaural transient disparity. In addition, noise burst durations were varied from 10 msec to 1 sec. The presentation paradigm and the task required of the subject was identical to the task used by Tobias and Zerlin (1959): the subject was presented with a pair of binaural noise bursts and had to indicate the relative locations of the second pair using the first pair as a reference. The results showed that transient disparity becomes less important as duration increases. For 10 msec noise bursts, a 400 μsec transient disparity in one direction requires a 94 μsec ongoing disparity in the opposite direction to achieve a centered image. When the duration is increased to 300 msec, approximately zero ongoing disparity is necessary to achieve a centered image. The results for intermediate noise durations indicated that the ongoing disparity necessary to balance the transient offset appears to increase proportionately with the value of the transient disparity. The authors define an effective onset duration as ranging between 2 and 4 msec. Apparently, ongoing disparities are extremely potent cues in that they can offset much larger transient disparities. Any interpretation of these must take into consideration the fact that the stimuli configurations are artifical and therefore do not resemble real-life experiences. It is difficult to imagine any situation where conflicting cues of this type occur naturally. Nevertheless, these studies certainly help to characterize the incredible signal processing capabilities of the binaural auditory system.

Franssen (1961) demonstrated that transient cues can totally dominate localization judgments in the case of short duration tone bursts. Franssen used sinusoidal signals where one of two loudspeakers received an initial and final 40 msec tone burst. Both of these components (initial and final) started abruptly (transient) and decayed over a 40 msec period. The second speaker received the same tone, however, with an initial 40 msec rise time, a sustained portion of 3960 msec followed by a 40 msec decay period. Despite the fact that the first speaker only produced tonal transients, subjects identified all sound as one continuous tone emanating from the first loudspeaker which in fact had only produced the transients. This is a remarkable finding when one considers that the first loudspeaker produced sound for less than 1% of the listening period.

Interpretation of Franssen's data should consider an experiment reported by Witkin, Wapner and Leventhal (1952). These investigators demonstrated that subjects perceive speech sounds as emanating from a visible loudspeaker when in fact the speech sounds are being produced by a concealed speaker within approximately 20^{o} of the visible speaker. Nonauditory cues can obviously play a substantial role in auditory localization.

Yost, Wightman and Green (1971) measured the importance of the frequency content of brief transient pulses in the determination of interaural temporal acuity. Subjects were asked to discriminate between a centered image produced by presenting identically filtered clicks which arrived simultaneously at both ears versus a displaced image produced by slightly delaying the click to the left ear. The ability to lateralize deteriorated as clicks were high-pass filtered beyond 1500 Hz while low-pass filtering (removal of the high-frequency energy content) had little or no effect on the lateralization task. Yost et. al. also found that masking the transient with low-pass noise severly disrupted lateralization acuity of high-pass clicks while high-passed noise had no effect on the lateralization of low-pass transients. The authors interpret their results as suggesting that the discrimination of lateral position depends mainly on the low-frequency content of the click. These results agree well with the results reported on pure-tone localization where it has been shown that for frequencies above 1500 Hz, sensitivity to phase differences markedly decreases.

Henning (1974a) was able to demonstrate that amplitude-modulated waveforms containing only high-frequency energy could be readily lateralized. In fact, the precision of lateralization with these waveforms based solely on differences in interaural time, is as good or better than lateralization of low-frequency pure tones. Lateralization is apparently dependent on the time-delay of the envelope rather than the time delay of the fine structure within the envelope. Similar results have been reported by Leakey, Sayers, and Cherry (1958) in a study specifically directed at the binaural fusion of low and high frequency sounds. A complete description of their experiment and results is presented in a later section dealing with binaural fusion.

Localization in Reverberant Sound Fields

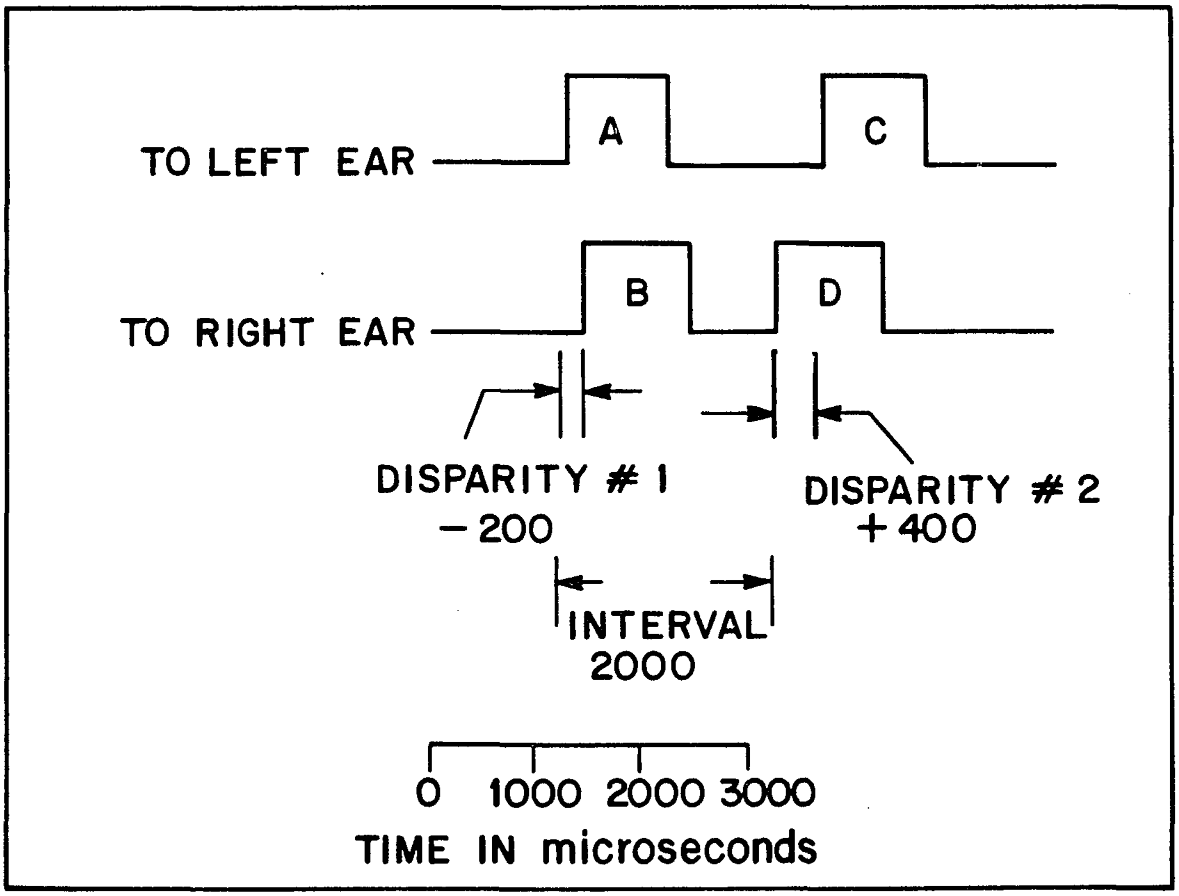
The normal listening environment rarely provides us with a reflection-free listening. Even though the sound may emanate from a single point source, reflections from hard surfaces throughout the environment are ubiquitous. Nevertheless, despite the existence of these reflection which can be thought of as spurious sound sources, we have no difficulty in locating the originating source (eg. a talker, loudspeaker). Indeed, accurate localization occurs in reverberant environments where the total energy in the reflected sound may be greater than that of the sound reaching our ears from a direct path (Moore, 1977).

During the 1930's Fay (1936) and Hall (1936) presented two papers before the 14th meeting of the Acoustical Society. Their reports focused on group address systems which would maintain the illusion that reproduced sound emanates from the speaker's mouth by arranging the reproducing system so that the original sound reached the listener's ear before the reproduced sound. Although their scheme obviously was taking advantage of the precedence effect phenomenon, neither individual followed-up their presentation with a publication.

Wallach, Newman and Rosenzweig (1949) examined the ability of subjects to correctly localize sound sources in the presence of echoes presented both in free-space as well as over headphones. The free-space study demonstrated that when two similar sounds (clicks and gated sinusoids) are presented in close sequence (approximately 1 msec), the sounds will fuse and be heard as a single source with the location primarily determined by the location of the first occurring or prior sound. This phenomenon was logically labeled the precedence effect. An analytic study using one msec clicks presented over matched earphones was carried out to confirm the free-space experiment.

Two clicks are heard as a single fused click when presented in succession as long as the interval between the clicks does not exceed approximately 2 msec. In their experiment Wallach, et. al. took advantage of this fact by fixing the temporal onset disparity of a pair of pulses to one ear while the temporal onset disparity of a second similar pair of pulses presented to the other ear was varied. The stimulus sequence used in their study is shown in Figure 2.7 where two pairs of pulses are used (one to each ear) for each trial. The first pair of pulses, labeled A and B, arrive at the ears with a small time disparity (left ear leading) as shown by the arrows, while the second pair of pulses, labeled C and D, in effect simulate a single echo and show an opposite interaural disparity where the signal now arrives first to the right ear. If the initial and delayed clicks arrive earlier at the left earphone, subjects will report the fused image to be located to the left. It is easy to see, however, as shown in Figure 2.7, that by fixing the delay between clicks presented to the left ear, Δt_1 and varying the delay between the clicks in the second ear, Δt_2, the relative importance of the two conflicting sources can be determined. In this case, the first pair points left, while the second pair points right. All that need be done is to determine the extent to which the right ear has to lead so as to obtain a centered image. These results showed that subjects gave the initial pulse substantially more weight in determining their response. In one

Figure 2.7 Voltages delivered to the right and left ears. Diagram shows the four separate clicks and three time-intervals (Disparity #1, Interval and Disparity #2) which could be manipulated by the subject. (From Wallach, Newman and Rosenzweig, 1949).


TO LEFT EAR
A
C
TO RIGHT EAR
B
D
DISPARITY # 1
- 200
DISPARITY # 2
+ 400
INTERVAL
2000
0
1000
2000
3000
TIME IN microseconds

instance, for example, centering of a 50 μsec lead in one direction for the first pair of clicks required a 500 μsec lag in the second pair of clicks. Subjects were asked to indicate the location of the fused image in terms of a six sector arc circumscribing the ear to ear distance. The ratio of the two delays can be used as an indicator of the temporal importance of both cues. Wallach et al. found the value of this ratio to vary between 6 and 10, so that an initial disparity of say, 30 μsec, in the leading pair required an opposite disparity of between 180 and 200 μsec in the second pair to center the image. The results of their study are shown in summary form in Figure 2.8.

Wallach et al. offer the following summary of their results:

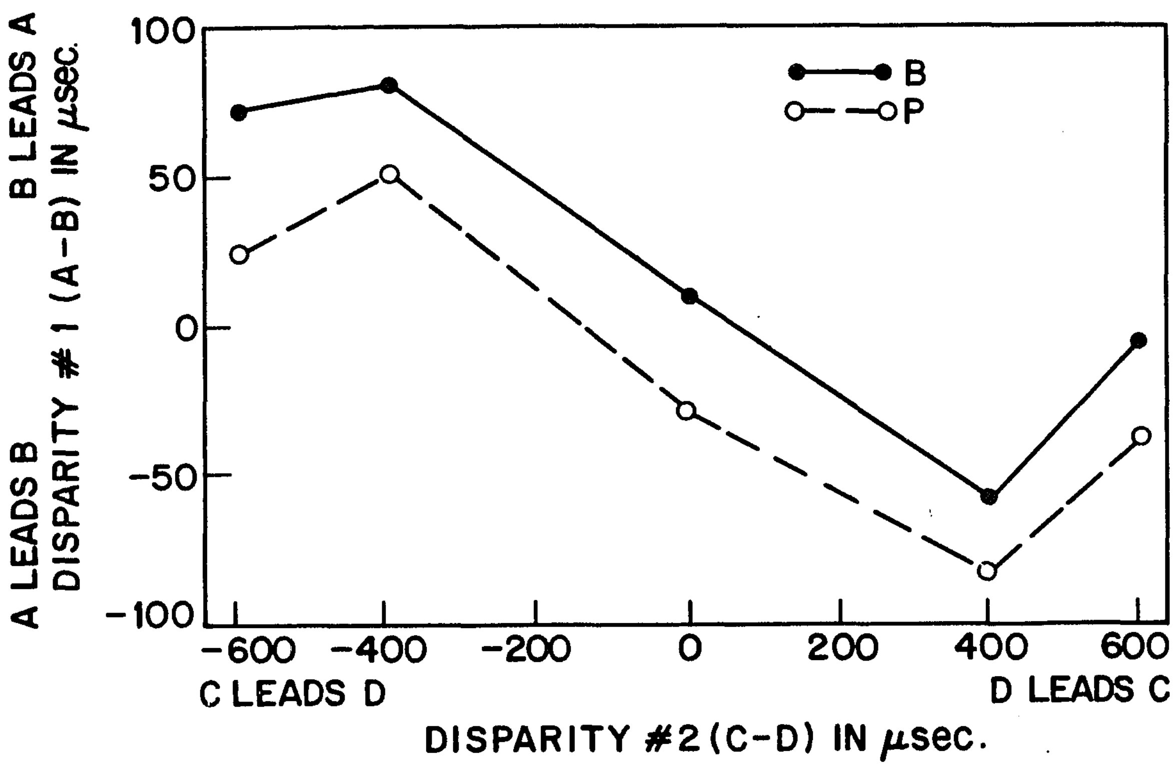
1. Two brief sounds that reach an observer ear in close succession are heard as a single click if the interval between them is in the order of 2 msec or less; this interval is not the same for all sounds since the limit can be extended out to 40 or 50 msec for complex stimuli such as speech and music.

2. If two brief sounds are heard as a single fused sound, the location of the fused image is determined largely by the localization of the initial sound; they termed this a precedence effect in sound localization versus centralization.

3. The effect is only shown for sounds of a discontinuous or transient nature.

4. The second sound or echo can be shown to have a smaller but measurable effect; if the location of the second sound departs more and more from the location of the first, it will pull the total sound with it up to a maximal movement of about 7 degrees and then becomes progressively less effective.

Figure 2.8 Disparity of First Pair (ordinate) that just balances disparity of Second Pair (Abscissa). Results are plotted separately for subjects B and P. (From Wallach, Newman and Rosenzweig, 1949).


B LEADS A
A LEADS B
DISPARITY #1 (A-B) IN μsec.
100
50
0
-50
-100
B
P
-600
-400
-200
0
200
400
600
C LEADS D
D LEADS C
DISPARITY #2 (C-D) IN μsec.

5. If the interval between the arrival of the two sounds is less than one msec the precedence effect does not operate; rather some compromise or average location of the image results.

6. If the second sound is made sufficiently intense, say 15 dB above the first sound it overrides the precedence effect.

7. The precedence effect is favored for sounds which are qualitatively similar; that is if the two stimuli are sufficiently different in character so as too prohibit binaural fusion, the precedence effect will not occur. The advantage of a system which weighs more heavily the initial arriving pair is obvious for environments where reflections are apt to be abundant (Wallach et al, 1949, p. 171).

Almost simultaneous with the publication of the Wallach et al. study, Haas (1949) submitted a doctoral dissertation to the University of Gottingen entitled, "The Influence of a Single Echo on the Audibility of Speech." This review of Dr. Haas's work is based on a translation by K. P. R. Ehrenberg (1972). The object of the Haas study was to determine the influence of echoes on the audibility of speech in binaural hearing. The experiments concerned with small delays (i.e., delays not exceeding 50 ms) were initially conducted on a roof where Haas hoped to be able to minimize reflection problems.

It should be noted that although Haas was interested in examining the same binaural echo suppression phenomenon as Wallach et. al., he used a significantly different approach. The suppressive effect of the first arriving waveform was determined by asking subjects to reduce the level of the first arriving signal, until both loudspeakers sounded equally loud.

Haas used two loudspeakers which were placed 3 m from the observer at an angle of 45^{o}. Subjects were instructed to adjust the intensity of the primary loudspeaker with an attenuator until both the primary and delayed loudspeaker appeared to be equal in level. The first experiment used continuous spoken text as source stimuli with delay times varying from 1 msec to 40 msec. According to Haas the equal loudness judgement did not produce a direct-in-front fused perception, but rather a sensation of two sound sources emanating from two different directions.

The results obtained for 15 subjects are shown in Figure 2.9. The necessary reduction in intensity of the primary loudspeaker to achieve equal loudness is plotted as a function of delay time. Equal loudness judgements in the 5 to 30 msec range require a 10 dB reduction in level of the primary speaker. Beyond 30 msec the intensity reduction required for equal loudness drops slightly and the points are more scattered. This result is often referred to as the "Haas effect."

Haas repeated the above experiment using two different noise sources: (a) band-passed Gaussian white noise with a frequency range of 200-400 Hz (b) band-passed Gaussian white noise with a frequency range of 3200 to 6400 Hz. The results of these experiments are shown in Figure 2.10 and

Figure 2.9 Echo suppression effect as a function of delay time for speech. (From Haas, 1972).

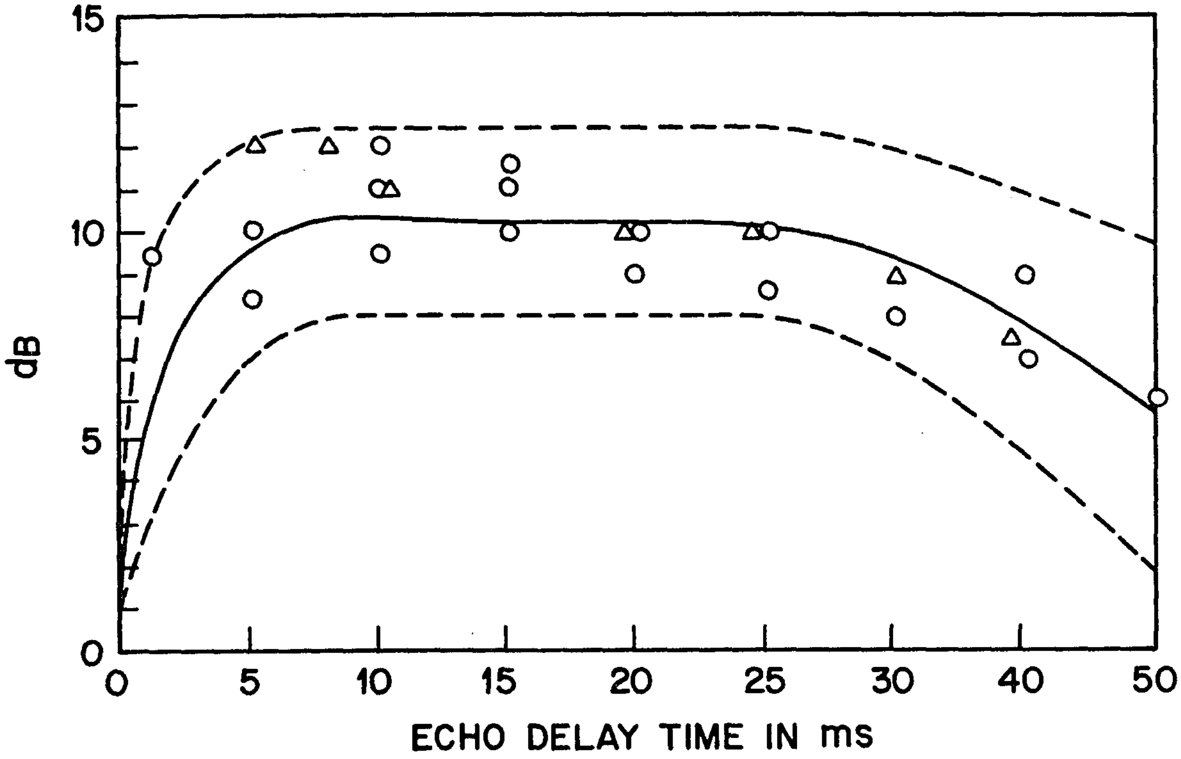
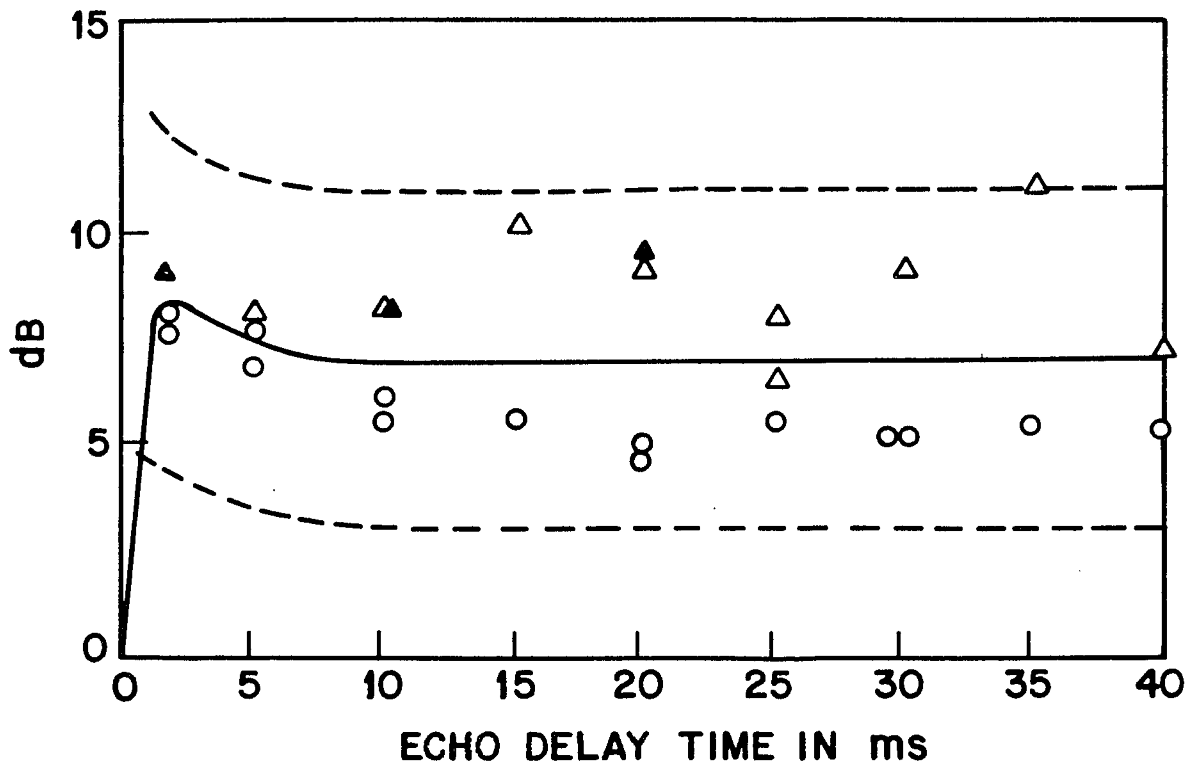

15
10
5
0
dB
0
5
10
15
20
25
30
40
50
ECHO DELAY TIME IN ms

Figure 2.10 Echo suppression effect as a function of delay time for noise with frequencies from 3200 - 6400 Hz. (From Haas, 1972).

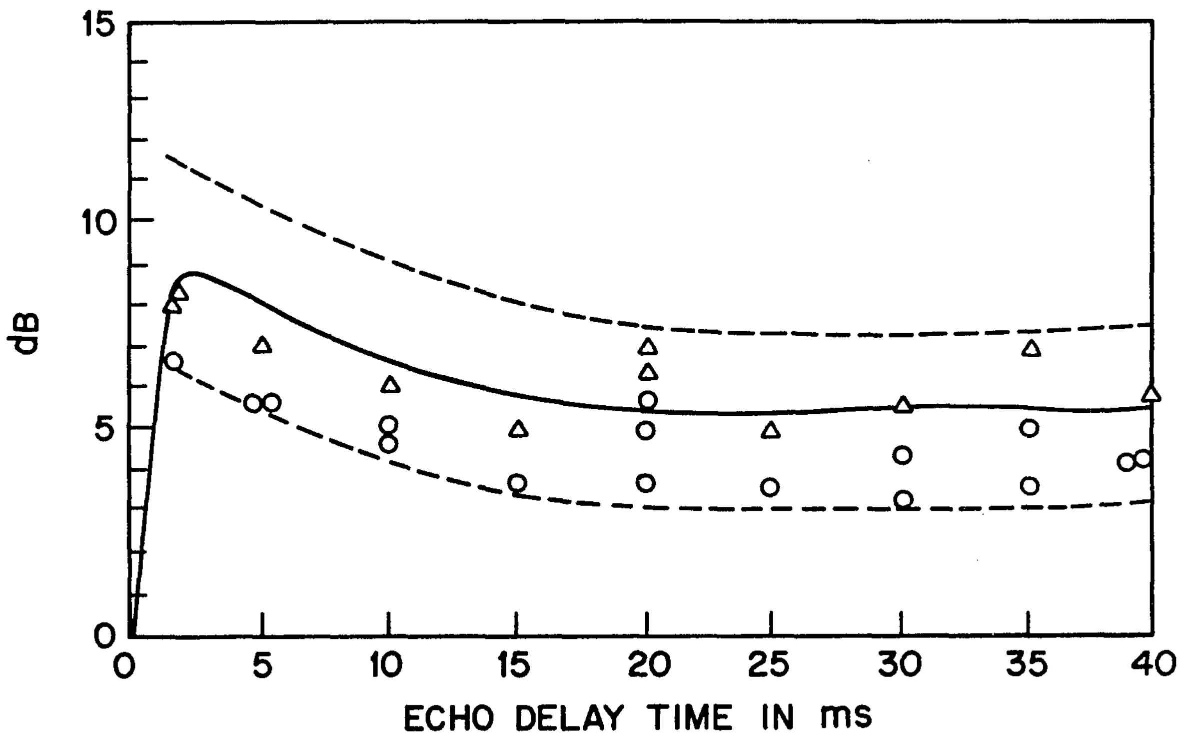

15
10
5
0
dB
0
5
10
15
20
25
30
35
40
ECHO DELAY TIME IN ms

2.11. In both those instances where Gaussian noise stimuli were used, less attenuation of the primary signal was required for equal loudness than was required for equal loudness with speech stimuli. This effect was most pronounced with high-frequency noise stimuli.

Harris, Flanagan and Watson (1963) proposed a neural gate model to explain the precedence effect. A neural gate is speculated to exist prior to the level of binaural interaction (the cochlear nucleus is specifically identified as a good candidate) which would close, approximately one msec after receiving the first neural response thereby permitting no further timing information to be sent to higher brain centers. This gate would then remain closed for approximately 2 msec. They further speculate that self-inhibition and inhibition of lateral nerves by the nerves firing on the initial stimulus as the underlying mechanism for this effect.

Guttman (1962, 1965) used a three-click paradigm to study binaural interactions. Two clicks were temporally fixed in opposite ears (ie. over headphones) while a third click always appeared in the reference ear and could be moved about in time relative to the fixed click in that ear. Using this technique, Guttman was able to examine the interfering properties of the third roving click on the fixed image produced by the temporally fixed (binaural) clicks. Consider the case where a single click is presented

Figure 2.11 Echo suppression effect as a function of delay time for noise with frequencies 200 and 400 Hz. (From Haas, 1972).


15
10
5
0
dB
0
5
10
15
20
25
30
35
40
ECHO DELAY TIME IN ms

to each ear with zero interaural time delay. In the absence of the third click, a lateralized image is perceived in the center of the head. The addition of a second click to the reference ear 2 msec prior to the fixed clicks, produces a shift of the lateralized image toward this ear. The image location is now totally dominated by the newly added click. Removal of the fixed (later) click in this ear had no effect on the location of the lateralized image. That is, the image was still lateralized at the ear receiving earlier stimulation. The loudness and timbre of the lateralized image, however, was altered.

Harris et. al (1963) interpret this observation as supporting their speculation that even when the timing information is blocked, other information relating to timbre and loudness is not blocked. That is, the lateralized image remains stationary despite the removal of the later monaural click due to an interruption of the timing pathway,while timbre and loudness are free to change since these pathways are not blocked.

Masking-Level Differences (MLD's)

Thus far, we have concentrated on those aspects of the binaural auditory system which relate to localization of sounds in space. Several studies (see Green and Henning, 1969 and Green and Yost, 1975 for an excellent review) have demonstrated that the binaural system also functions to permit the selection and detection of specific acoustic sources out of a background of interfering and unwanted sounds. This phenomenon was originally and simultaneously identified by Licklider (1948) and Hirsh (1948). The paradigm typically used for demonstrating this effect in the laboratory has been frequently described (eg. Jeffress, 1972) as follows: A low frequency interrupted tone along with a relatively narrow-band noise is initially applied to one ear through a headset; the subject is asked to adjust the interrupted tone until it is barely audible against the noise background. The identical noise waveform is then applied to the headphone over the opposite ear whereupon a dramatic improvement in the detectability of the tone is observed; that is, with this listening configuration subjects can now attenuate the signal (which still appears at only one ear) by approximately 9 dB without losing the tone. However, if the tone is now added to the second ear, there is a marked drop in the detection threshold which is approximately equal to that of the initial condition where only one ear received both the signal and masker. More dramatic effects are found when the relative phase of the noise waveform at the two ears is reversed while the tone is present in phase at both ears. Finally, the largest improvement using wide-band maskers has been obtained when the signal (an interrupted tone) is phase reversed at one ear, with the noise waveform remaining in phase at both ears. The latter condition produces a 15 dB increase (improvement) in the sensitivity threshold (Green and Yost, 1975). A masking level difference (MLD) is typically specified in terms of the signal level necessary in dB for detectability in the experimental condition relative to the signal level for the S_oN_o condition where both the tone and noise appear in phase at both ears (Levitt and Voroba, 1976).

Conventional notation used in masking-level difference experiments is shown in Table 2.1. The subscript m refers to the monaural listening mode, the subscript o denotes the diotic listening mode where the same input is presented to both ears and u refers to an uncorrelated masker at the two ears. The various interaural signal and masker combination are compared in Table 2.2 (Green, 1976). Manipulation of the signal frequency, phase and duration have been shown to effect the obtained MLD.

As the signal frequency is raised from 250 Hz to 2000 Hz, the MLD has been shown to decrease from a maximum value of 15 dB to a uniform level of approximately 3 dB (Hirsh, 1948). Masking level differences are presumed to depend upon the processing of phase or timing information with the lack of improvement in MLD values above 1500 Hz being linked to the insensitivity of the auditory system to temporal cues from signals above 2 kHz (Levitt and Veroba, 1976). Some controversy exists over MLD's obtained for signals below 250 Hz. While Hirsh (1948) reported an MLD value of approximately 5 dB for 100 Hz, Durlach (1963) found an MLD value of

S_o:	Signal presented to both ears with no interaural differences (lateralized image in center of head).
N_o:	Noise presented to both ears with no interaural differences (lateralized image in center of head).
S_m:	Signal presented to only one ear, no signal presented at the other ear (sound image at one ear).
N_m:	Noise presented to only one ear, no noise presented at the other ear (sound image at one ear).
S_π:	Signal presented to both ears but the waveform inverted at one of the ears (sound image at both ears or on a thin line from center toward both ears - distinctly different from S_o). For a sinusoid this amounts to 180° phase delay between the two ears.
N_π:	Noise presented to both ears but inverted at one of the ears.
N_u:	Noise uncorrelated at the two ears (sound image in center of head but very diffuse compared with N_o).

TABLE 2.1: Conventional Notation Used in Masking-Level Difference Paradigms. (From Green, 1976).

Interaural Condition	Relative Improvement (in dB, $\pm$.5 dB)
S_oN_o, S_mN_m, S_mN_u, $S_\pi N_\pi$	0
$S_\pi N_u$	3
S_oN_u	4
S_mN_π	6
S_mN_o	9
S_oN_π	13
$S_\pi N_o$	15

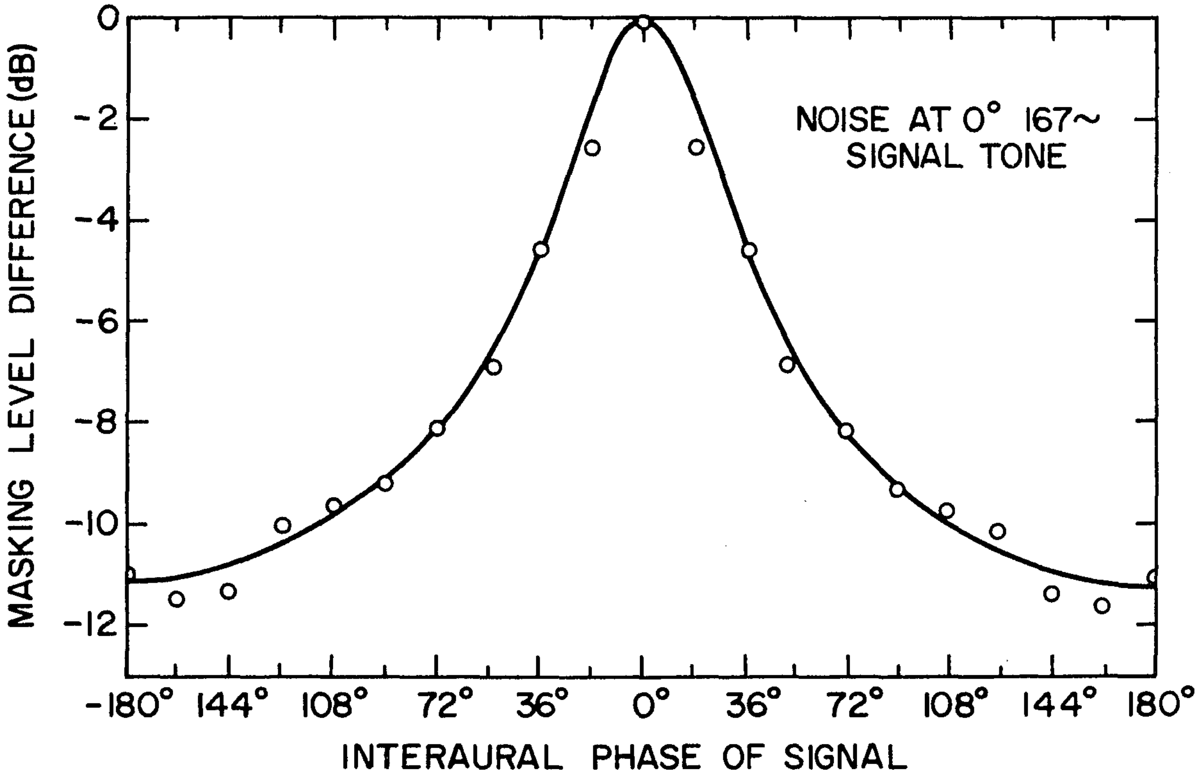
TABLE 2.2: Typical Results Reported For Various Stimulus-Masker Configurations Used in MLD Experiments. (From Green, 1976).

15 dB for this frequency. More recently, Dolan (1968) has reported MLD's for signals in the 150 Hz to 300 Hz frequency range to be a function of the intensity of the masking noise surrounding the tone.

The second signal parameter which has been investigated is that of interaural phase. Jeffress, Blodgett and Deatherage (1962) have shown that as the interaural phase difference is changed from 0° (i.e. S_oN_o) to 180°, (i.e. $S_{\pi}N_o$) the obtained MLD's show a gradual increase as the interaural phase relationship approaches $\pm 180^{\circ}$ with the maximum MLD obtained at 180° (see Figure 2.12).

Masking level differences are also dependent on several parameters of the masking noise including overall level, and interaural correlation or time delay. Overall noise levels have been shown to affect the size of obtained MLD's. For the $S_{\pi}N_o$ - S_oN_o condition using a 200 Hz signal, MLD values have been shown to range from about 4 dB at 9.1 dB spectrum level to about 14 dB at a more intense spectrum level of 59.1 dB (Hirsh, 1948). This effect has been ascribed to the fact that internal auditory noise is only partially uncorrelated (Green and Henning, 1969). Apparently, the uncorrelated portion of the internal noise serves to disrupt the correlation in the effective stimulus at the two ears. An intense external noise source will obviously serve to override or mask the effect of the uncorrelated internal noise. It has also been shown that sizable binaural effects occur

Figure 2.12 Size of the masking level difference as a function of interaural phase of the 167 Hz signal. The noise is in phase at the two ears. At 0^{o}, the condition is S_oN_o while at ± 180 the condition is $S_{\pi}N_o$. (From Jeffress, Blodgett and Deatherage, 1962).


NOISE AT 0° 167~
SIGNAL TONE
MASKING LEVEL DIFFERENCE (dB)
0
-2
-4
-6
-8
-10
-12
-180° 144° 108° 72° 36° 0° 36° 72° 108° 144° 180°
INTERAURAL PHASE OF SIGNAL

even when the noise in the nonsignal ear (i.e. S_mN_o) differs by ± 10 dB from the noise in the signal ear. Intensity differences greater than this amount serve to decrease the binaural advantage (Green and Yost, 1976).

The degree of correlation between the noise signals present at the two ears also has been shown to be an important variable. When two independent but identical random noise generators are used, one for each ear, an MLD value of 3 dB is obtained (Robinson and Jeffress, 1963). This case, of course, represents an uncorrelated (interaural) noise condition. The introduction of a time delay between the noise waveforms delivered to each ear also serves to decorrelate the noise waveforms. Using a 500 Hz signal and starting with a S_oN_o reference condition, Jeffress, Blodgett and Deatherage (1957) have shown that when the narrow-band noise is delayed by some value τ, a cyclic MLD pattern is observed in which maxima of detectability occur at half-periods of the signal and minima occur at full periods with the difference between maxima and minima decreasing as delay increases. The obtained MLD's oscillate in a pattern almost identical to the correlation coefficient over a small band of frequencies near the frequency of the signal (Green, 1976). Of course, at long delays where the interaural correlation approaches zero, an MLD of 3 dB is obtained.

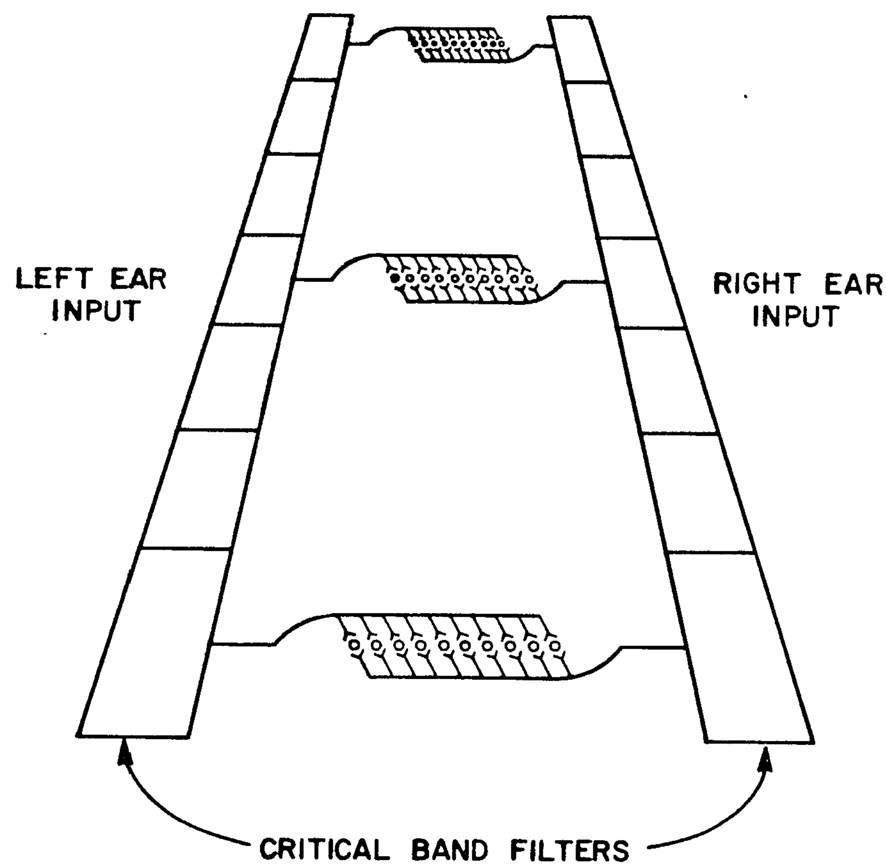
Although several models, Webster (1951), Jeffress (1972) Durlach (1963, 1972), Osman (1971), Hafter (1969) and

Colburn (1973) have been proposed to explain the MLD phenomenon, only a combination of the first two models referred to as the Webster-Jeffress hypothesis as well as Durlach's Equalization and Cancellation model have been extensively examined. The Webster-Jeffress hypothesis assumes that the increased detectability observed in binaural experiments arises out of the more fundamental localization mechanism, where the noise and signal are lateralized in different places (Green, 1976). Cues are obtained from the different temporal properties of the signal and noise. Webster uses a vector diagram where the noise waveform in the N_o case is represented by a fixed length vector since the interaural phase is zero. A centered image will be perceived in this instance due to the absence of a time difference in the arrival of the waveforms at the two ears. When a signal now is added to one ear, a phase change between the signal plus noise vector in one ear and the noise vector in the other is produced which in turn serves as a detection cue. High frequencies will produce small phase changes relative to low frequency signals which produce larger time or phase differences (Green and Henning, 1969). This is consistent with MLD results where low frequencies produce large MLD's and high frequencies produce small MLD's. Actual calculations require that a narrow-band of noise frequencies surrounding the signal frequency be treated as a slowly changing sinusoidal process.

Jeffress' extension of the Webster hypothesis is somewhat more physiologically based. Each signal is initially filtered by a series of critical band filters. The left and right critical bands for any given frequency region send off axons toward the opposite ear. These axons in turn send off collaterals which can converge on cells from the collaterals coming off the opposite ear. Essentially, then, a sort of coincidence detector exists. For example, in-phase stimuli will cause neurons in the center section of the coincidence detector to fire, while a signal delayed, say to the left ear, will cause its neurons to fire later, and coincidence will move off center to the right. A diagram of the vector model along with the coincidence model is shown in Figure 2.13.

Durlach's initial interest was to devise a model which could predict the binaural results obtained in empirical experiments. The Equalization and Cancellation (E-C) model assumes the waveforms at the two ears to be filtered (via critical band filters) and amplified or attenuated (via a scaling factor). Once the gain (scaling factor) for each channel is adjusted, the two channels are combined, where either subtraction (left minus right) or addition (left plus right) is permitted. The cancellation device shown in Figure 2.14, will perform the function which best serves to cancel the masker and thus permit detection of the signal. The ability to switch to a monaural mode of operation when

Figure 2.13 (a) An illustration of Jeffress' hypothesis concerning the physiological mechanism responsible for the interaural time cue. The signal is first filtered by critical bands and initiates neural impulses that travel on the fibers extending toward the opposite ear. Each fiber sends off collaterals to a cell body, indicated by a small circle. If the signal is delayed to the left ear, the finite velocity of neural transmission will cause the impulses generated in each ear to arrive simultaneously at a cell body located on the left side of the diagram. (From Green, 1976); (b) Vector diagram illustrating how adding a signal S added at angle ϕ to noise component produces a phase change Θ. The resulting interaural phase cue, Θ, or its temporal equivalent, is the basis of detection in this lateralization model. (From Green, 1976).


LEFT EAR
INPUT
RIGHT EAR
INPUT
CRITICAL BAND FILTERS

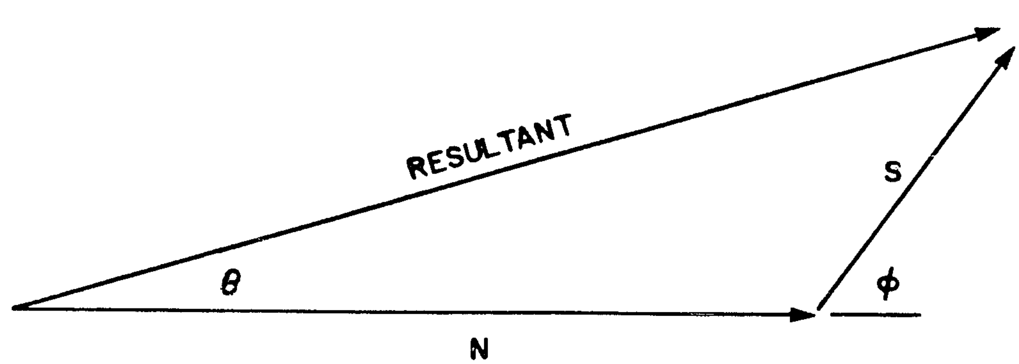
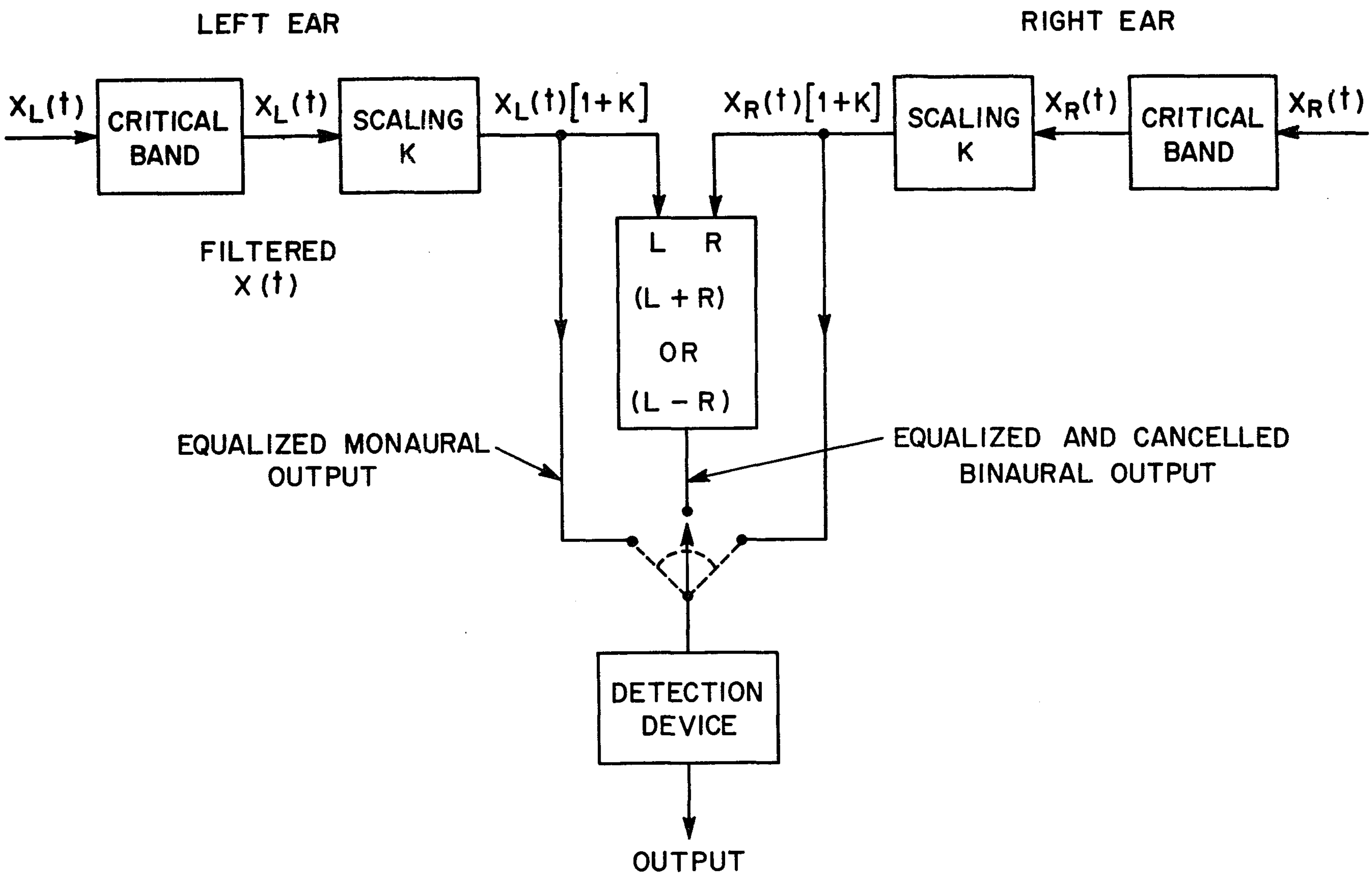

RESULTANT
S
θ
φ
N

Figure 2.14 Durlach's Equalization and Cancellation model. The operators of this model are indicated by the boxes. The detection device will select either of the binaural interactions to maximize signal detectability. (From Green, 1976)


LEFT EAR
RIGHT EAR
$X_L(t)$
CRITICAL BAND
$X_L(t)$
SCALING K
$X_L(t)[1+K]$
$X_R(t)[1+K]$
SCALING K
$X_R(t)$
CRITICAL BAND
$X_R(t)$
FILTERED X (t)
L R
(L + R)
OR
(L − R)
EQUALIZED MONAURAL OUTPUT
EQUALIZED AND CANCELLED BINAURAL OUTPUT
DETECTION DEVICE
OUTPUT

either ear alone produces a better signal-to-noise ratio than is provided by the cancellation device. Although not depicted in the model shown in Figure 2.14, the E-C model also included two additional constants: (1) an internal noise constant to account for the fact that infinite release from masking was not feasable as well as (2) a filter constant to account for interaural timing errors.

Binaural Fusion

Localization and lateralization experiments have demonstrated that subjects can often readily fuse binaurally presented stimuli to produce a single image. Although most of these studies apply similar stimuli to each ear differing only in time of arrival or level, investigators have found that dissimilar or unlike signals applied separately to each ear also are readily fused. Broadbent (1955) found that when subjects listened to filtered speech over earphones with one ear receiving only high-frequency components while the other ear received only low-frequency components, 14 out of 18 subjects readily heard a single fused image. These subjects were then exposed to additional dissimilar binaural stimuli including: (1) a metronome signal - high-passed to one ear, low-passed to the other; (2) a speech recording high-passed to one ear, low-passed to the other with a delay of 250 msec; (3) a pure tone of 3000 Hz to one ear and a pure tone of 500 Hz to the other ear. The results showed that 11 out of the 14 subjects heard (1) as a fused single image, while all subject reported hearing (2) and (3) as separate sounds at each ear. Broadbent identified the temporal relation between the arrival of the sounds as the factor most likely responsible for determining whether fusion does or does not occur.

In a follow-up study, Broadbent and Ladefoged (1957) used synthesized formants with the first formant presented to one ear and the second formant to the other ear. The laryngeal driving source was also varied so that in some instances the same larynx generator (125 Hz) was used for both formants while in other cases two independent but identical laryngeal sources were used. For synthetic speech with a single laryngeal generator exciting both formants, most subjects reported hearing one voice in one place. The results for independent laryngeal generators showed that the most common response was for subjects to hear more than one voice in more than one place. Sustained formants one to each ear were also used as stimuli. The results showed that sustained formants could be heard as coming from a single source as long as the envelope frequency was the same. That is, if one ear received a 125 Hz fundamental and a 1250 Hz formant while the other ear received a fundamental of 125 Hz with a formant of 375 Hz, a single fused image was perceived. However, when the fundamental was raised to 150 Hz with the same 375 Hz formant, no subjects were able to fuse the sources.

Leakey, Sayers and Cherry (1958) also examined the relationship of spectral content and binaural fusion. One ear received a 4 kHz tone while the other ear received a 4.1 or 4.2 kHz tone. Both tonal stimuli, presented separately to each ear over headphones were modulated by the same signal, either a low-frequency (200 Hz) tone or a low frequency (500 low-pass) band of noise. In all cases, subjects were able to lateralize a fused image. The authors concluded that for binaural signals with energy exclusively in the spectral region above 1500 Hz, the binaural fusion mechanism is operated by the low-frequency envelope of these signals, while for binaural signals with energy content below 1500 Hz, the binaural fusion mechanism is controlled by the microstructure of the signals. The latter conclusion is based on the fact that the auditory system is insensitive to interaural time differences for high-frequency tonal stimuli (e.g. Klumpp and Eady, 1956). No measureable change in apparent location is found in such instances. Furthermore, David, Guttman and van Bergeijk (1959) have shown that the apparent location of high frequency clicks and noise bursts can be altered by balancing interaural delays against interaural intensity to produce a centered image. In summary, timing information provided by (a) the low frequency envelope of a waveform as well as individual high frequency complex waveforms (i.e. transients) can serve to produce a binaurally fused image which can be localized (Henning, 1975).

Scharf (1974) measured the ability of subjects to localize a pair of tones from two loudspeakers placed 3 m from the listener's head, one 45^{o} to the right, the other 45^{o} to the left of the median plane. The tones covered a range of 500 to 5000 Hz and were adjusted to be equally loud. All tones were ramped with a 1 or 2 msec rise and fall time and were presented for a duration of either 6 or 500 msec. Stimuli were repeatedly presented every 2 sec until the subject responded. Onset time differences between the two loudspeakers ΔT, was varied from 0 to 500 msec with one set of measurements taken with the same frequencies applied to both speakers; a second set of measurements were taken with a frequency separation, ΔF between the tone in the left speaker and the tone in the right speaker. The frequency separation varied from 0 to 4200 Hz around a geometric center of 2000 Hz. All presentations were made at loudness levels of 45 and 65 phons. Judgements were scored as left or right when they were more than 15^{o} to the left or right of the median plane, whether in-front or in-back of the listener.

Localization judgements were more accurate at the 65 phon level relative to the 45 phon level, as well as for short duration (6 msec) tones versus long duration (500 msec) tones. When the same frequency was simultaneously applied to both speakers, a single fused sound was heard until the temporal separation became greater than 20 to 60 msec. When different frequencies were applied to each speaker, a single fused sound was heard only with small temporal and frequency differences. Despite the existence of more than one distinct pitch in these cases, the auditory image, nevertheless, appeared to emanate from a single source. For a ΔT of 0 sec with different frequencies applied to both speakers, a single image was reported 90% of the time even when the two tones differed by as much as 900 Hz. The introduction of an interaural time difference caused the tones to separate at a narrower ΔF. In these instances, fusion was level dependent. At the lower presentation level of 45 phons, fused images were reported when values of ΔF as large as 600 Hz and values of ΔT as large as 40 msec; for the higher presentation level of 65 phons, fusion persisted until ΔF reach 500 Hz and ΔT reached 30 msec.

Other Phenomena Related to the Perception of Single Echoes

Studies reviewed in the preceding section on the precedence effect clearly demonstrate that appropriate binaural information in the form of the earliest arriving information serves us well in permitting accurate localization in environments with conflicting cues. A related issue which has recently received a great deal of attention concerns overall perception of signals which have been corrupted by an echo. It will be recalled that although the precedence effect facilitates localization by directing one's attention at the direct or undelayed sound source, this does not occur to the perceptual total exclusion of the delayed source. The abrupt termination of the delayed source produces a noticeable change in the perception, although localization does not change. Physical measurements in the form of room transfer functions indicate that the presence of single (or multiple) echoes serves to produce irregular frequency responses. A simple example of this problem is that of a direct sound interacting with a single reflection. The effect of combining a signal with a delayed version of itself is to produce a highly irregular rippled spectrum where the number of ripples are determined by the delay and their depth by its level. The mechanism involved is straight-forward. Depending on the delay, certain frequencies are reinforced while others are diminished. In the case where the level of the direct and delayed signal are equivalent, the reinforced frequencies reach a maximum peak while the diminished frequencies are totally cancelled. Despite the existence of these rather large frequency irregularities (ie. nulls of 40 dB are not uncommon) in our listening environment, we are rarely cognizant of them.

In an early study addressed at this issue, Atal, Schroeder and Kuttruff (1962) determined detection thresholds for well-defined impulse responses with single and multiple reflections. They wished to determine the listener's ability to distinguish between white (flat spectrum) Gaussian noise and filtered noises with irregular power spectra. The filter impulse responses used along with their corresponding power spectra are shown in Figure 2.15. Power spectra with significant deviations from a flat spectrum are often referred to as colored. It should be noted that these filters have a periodic frequency response with maxima which are separated by f = 1/time delay (in sec). All noise stimuli were presented diotically at a sensation level of 60 dB. Threshold measurements (50% correct) for the perception of coloration were obtained using a paired comparison paradigm where a white noise spectrum was always paired with a colored noise spectrum.

The results of this study are shown in Figure 2.16 for filters 1-4 (shown in Figure 2.15) where the threshold value of the delayed signal (in dB) is plotted as a function of delay. The threshold values obtained for all of these filters are similar. Although no statistics are presented in this brief article, the authors claim the threshold curves for all filters to be the same. For delays exceeding 77 msec, thresholds were not obtainable even when the level of the delayed signal (g) equaled that of the direct signal. They invoked two separate but related models to explain these results. The first model depicts the auditory mechanism as a short-time power spectrum analyzer. That is, if the level differences between the maxima and minima of the short-time power spectrum exceed some fixed threshold,

Figure 2.15 Filter impulse responses and associated power spectra. (From Atal, Schroeder and Kuttruff, 1962).

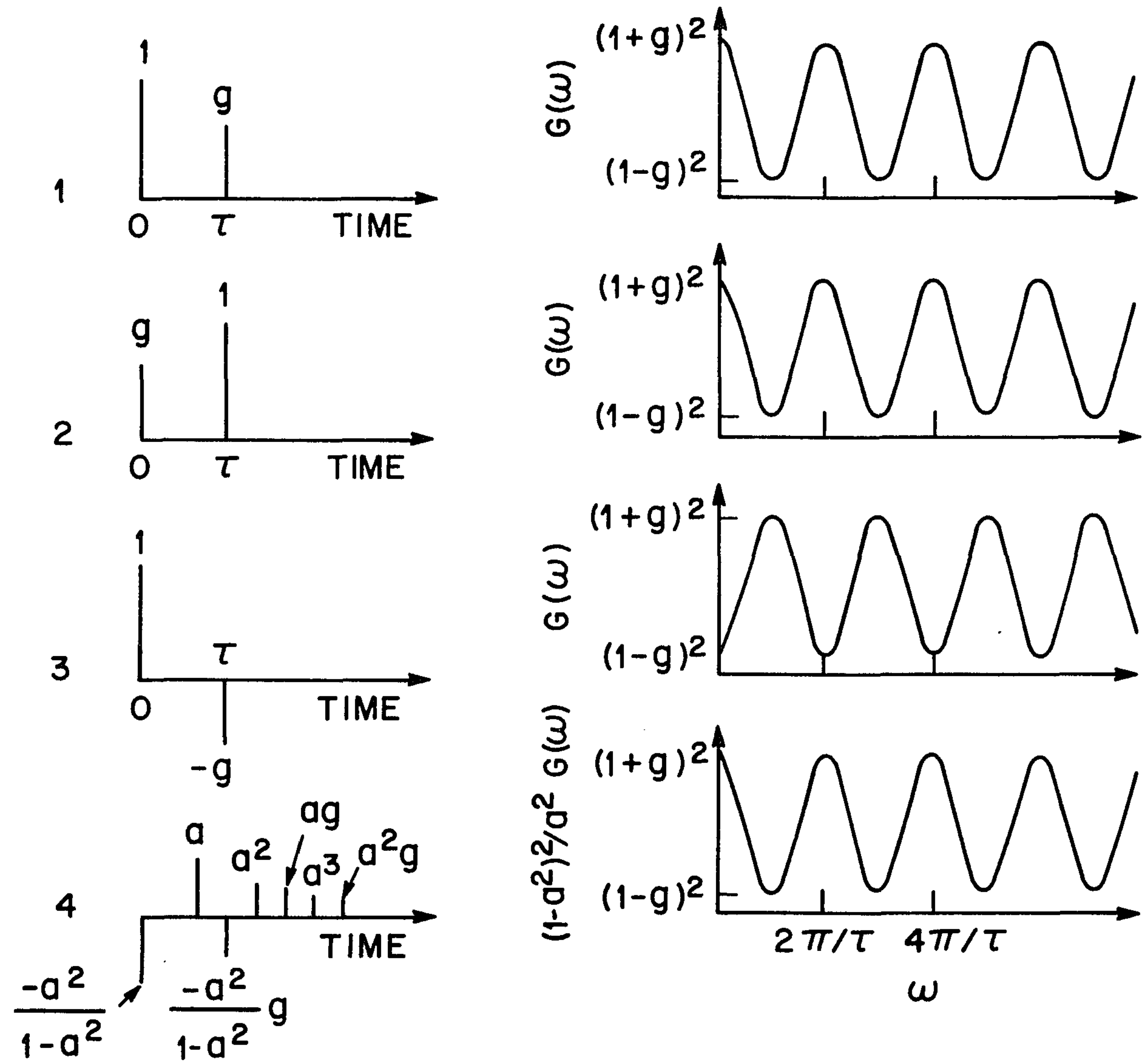
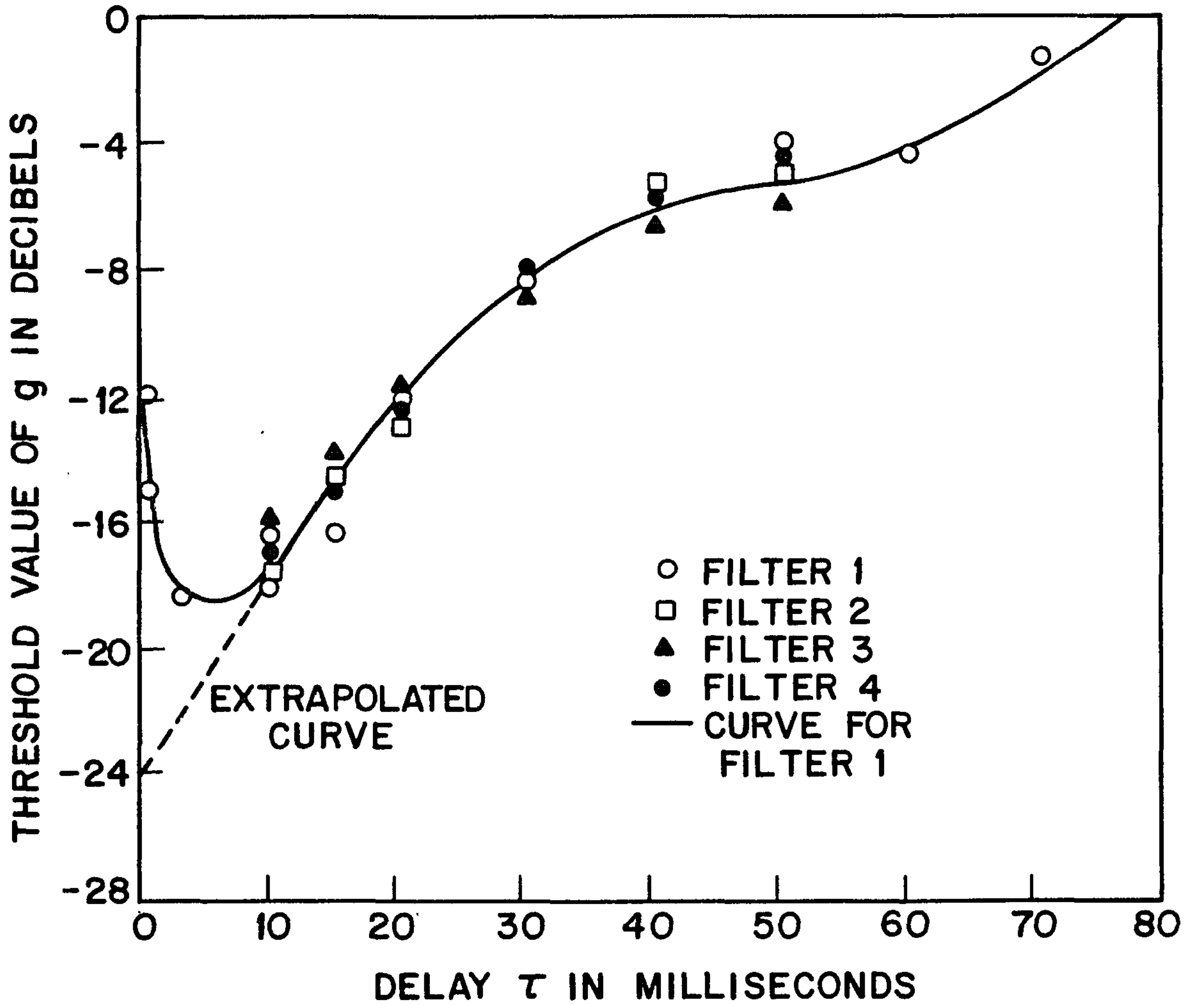

1
g
1
0
τ
TIME
(1+g)²
G(ω)
(1-g)²
g
1
2
0
τ
TIME
(1+g)²
G(ω)
(1-g)²
1
3
τ
0
TIME
-g
(1+g)²
G(ω)
(1-g)²
a
ag
a²
a³
a²g
4
TIME
-a²/1-a²
-a²/1-a² g
(1-a²)²/a² G(ω)
(1+g)²
(1-g)²
2π/τ
4π/τ
ω

Figure 2.16 Threshold levels obtained with noise bursts passed through simplistic filters (See Figure 2.15). (From Atal, Schroeder, and Kuttruff, 1962).



THRESHOLD VALUE OF g IN DECIBELS
0
-4
-8
-12
-16
-20
-24
-28
0
10
20
30
40
50
60
70
80
DELAY τ IN MILLISECONDS
EXTRAPOLATED CURVE
FILTER 1
FILTER 2
FILTER 3
FILTER 4
CURVE FOR FILTER 1

coloration is perceived. This model assumes all judgements to be based on place or frequency domain information.

A second model based on time-domain analysis is also presented by Atal et. al. This model is similar to a pitch-extractor autocorrelation model presented by Licklider (1956). Licklider's Triplex Theory of Pitch is based upon a three-dimensional processor: (1) an x-dimension which represents the power spectrum distribution; (2) a z- dimension which reflects the periodicity of the stimulus waveform; and (3) a y-dimension which reflects the location, or distribution of locations of the sound source (s). The latter position is determined by the times of arrival of neural volleys at the opposite sides of this region. The y-dimension analysis is achieved by a coincidence analysis which serves to correlate (via cross-correlation) the signals from the two ears so as to mediate sound localization. The z-dimension which serves as an envelope-periodicity detector also depends on coincident analysis. An autocorrelation process, however, is assumed for this dimension which can be based on periodicities present only in the envelope and not necessarily in the actual waveform.

Monaural Pitch Perception

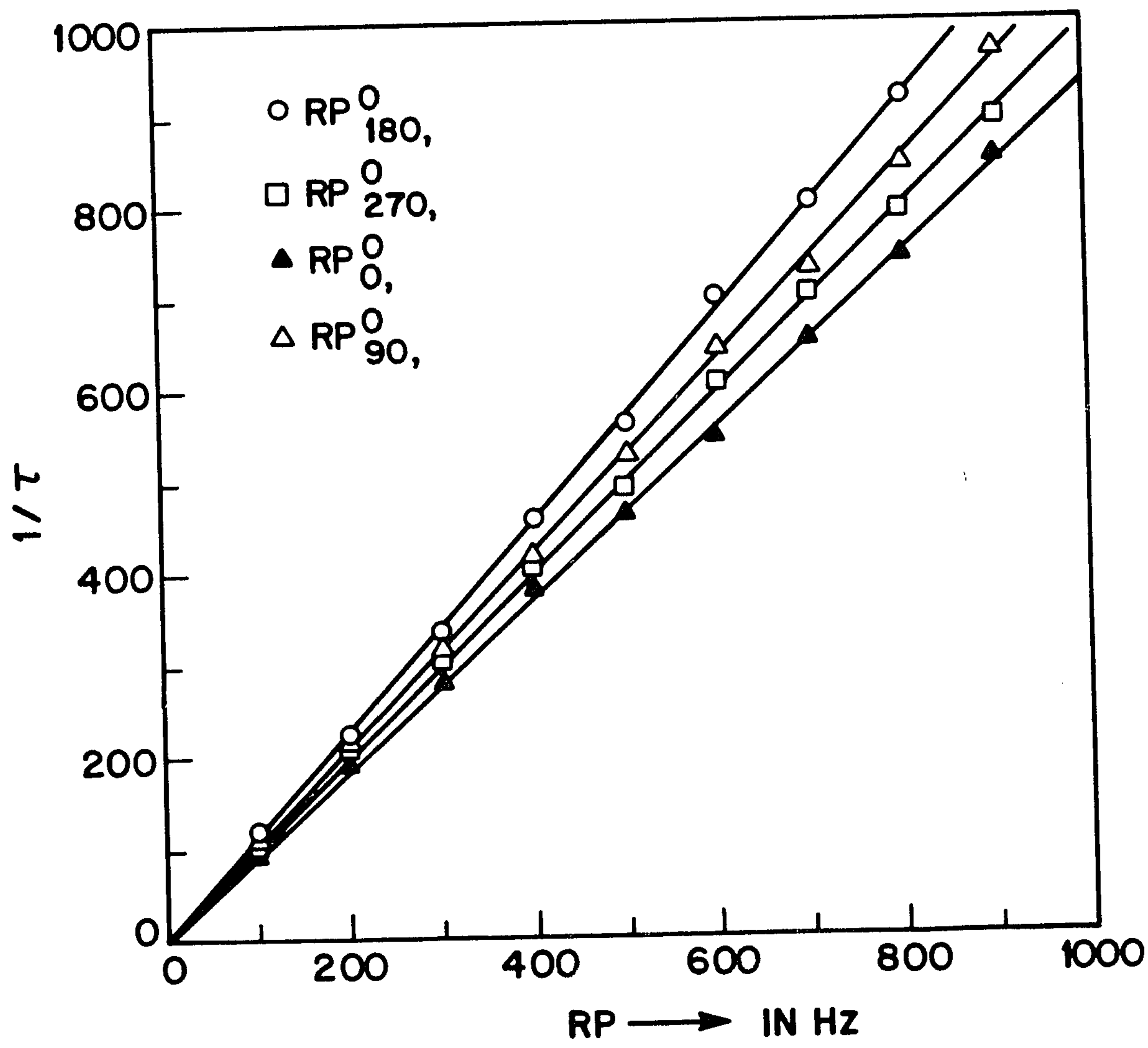
In a series of studies using stimuli similar (comb-filter spectra) to those used by Atal et al., Bilsen (1966, 1968) examined repetition pitch (RP) thresholds. When presented with a signal consisting of a sound and a repetition of that sound following a time delay, subjects experience a well-defined strong pitch sensation whose frequency corresponds to the reciprocal of the delay. This pitch-like perception is perceived along with the timbre or coloration sensation investigated by Atal et al. The differences ascribed to these percepts may be more apparent then real, with threshold differences resulting from the psychometric paradigm used. Bilsen's and colleagues' effort was specifically addressed at some basic questions regarding the functioning of the auditory system rather than the specific issue of the perception of reflections. The repetition pitch (RP) being examined has properties which are identical to periodicity pitch possessed by complex periodic sounds.

In the first published report, Bilsen (1966) measured the perceived pitch for two different types of stimuli: (1) noise, and (2) pulse pairs. In both cases, the parameters varied were: (a) the time separation or delay between the signal and its replication, and (b) phase shift of the replication. The task of the subject was to match the pitch of a pure tone (which the subject could control) to the perceived repetition pitch. Both the matching tone and the RP signal were presented over separate loudspeakers. In order to explain the results, the following notation (adopted from Bilsen) must be explained: for the symbol $RP^{\phi 1}_{\phi 2}$, ϕ_1 designates the phase shift of the undelayed sound and ϕ_2 the phase shift of the delayed sound. The results of the first experiment showed that the RP: corresponds to the reciprocal of $1/\tau$ where τ represents delay time in msec. That is, the Gaussian noise signal added to itself with a delay of 2 msec, produced a pitch-like sensation of 500 Hz while a delay of 4 msec produced a pitch-like sensation of 250 Hz. However, when the delayed signal was phase shifted by 90^{o} (using a Hilbert transformer) the perceived pitch (RP) changed so that the RP^{ϕ}_{90} pitch now was equivalent to $1.08/\tau$. In this case, the signal now produced a pitch of 270 Hz.

Changes in RP were also found for the other phase shifted conditions. RP^{ϕ}_{270} produced a pitch equivalent to $0.94/\tau$ while the 180^{o} phase shifted condition produced two different pitch matchings (despite the fact that both matches are to the same signal) of $RP^{\phi}_{180} = 0.87/\tau$ and $1.14/\tau$.

The results reported for the second experiment using pulse pairs (a pulse followed by a second pulse with fixed delay) and with similar phase manipulation of the delayed pulse produced virtually identical data with the exception of the RP^{ϕ}_{180} case. For these stimuli neither of the two subjects made different matching as was the case for noise stimuli. The relationship for all matching for this condition was 0.88 (eg., $RP^{\phi}_{180} = 0.88/\tau$. These results are summarized in Figure 2.17.

Figure 2.17 RP as a function of $1/\tau$ for pairs of impulses. (From Bilsen, 1966).



1000
800
600
400
200
0
1/τ
○ RP 0 180,
□ RP 0 270,
▲ RP 0 0,
△ RP 0 90,
0
200
400
600
800
1000
RP ⟶ IN Hz

In a follow-up study, Bilsen and Ritsma (1969/1970) attempted to explain the results reported above with a model which includes both frequency and time domain processing. Repetition Pitch is shown to correspond well with the reciprocal value of the time interval between two prominent positive peaks in the temporal fine structure of the displacement waveform, evoked by the signal, at a dominant frequency region on the basilar membrane. Furthermore, for each pitch value, this dominant region appears to be situated around a frequency of about four times the value of the pitch.

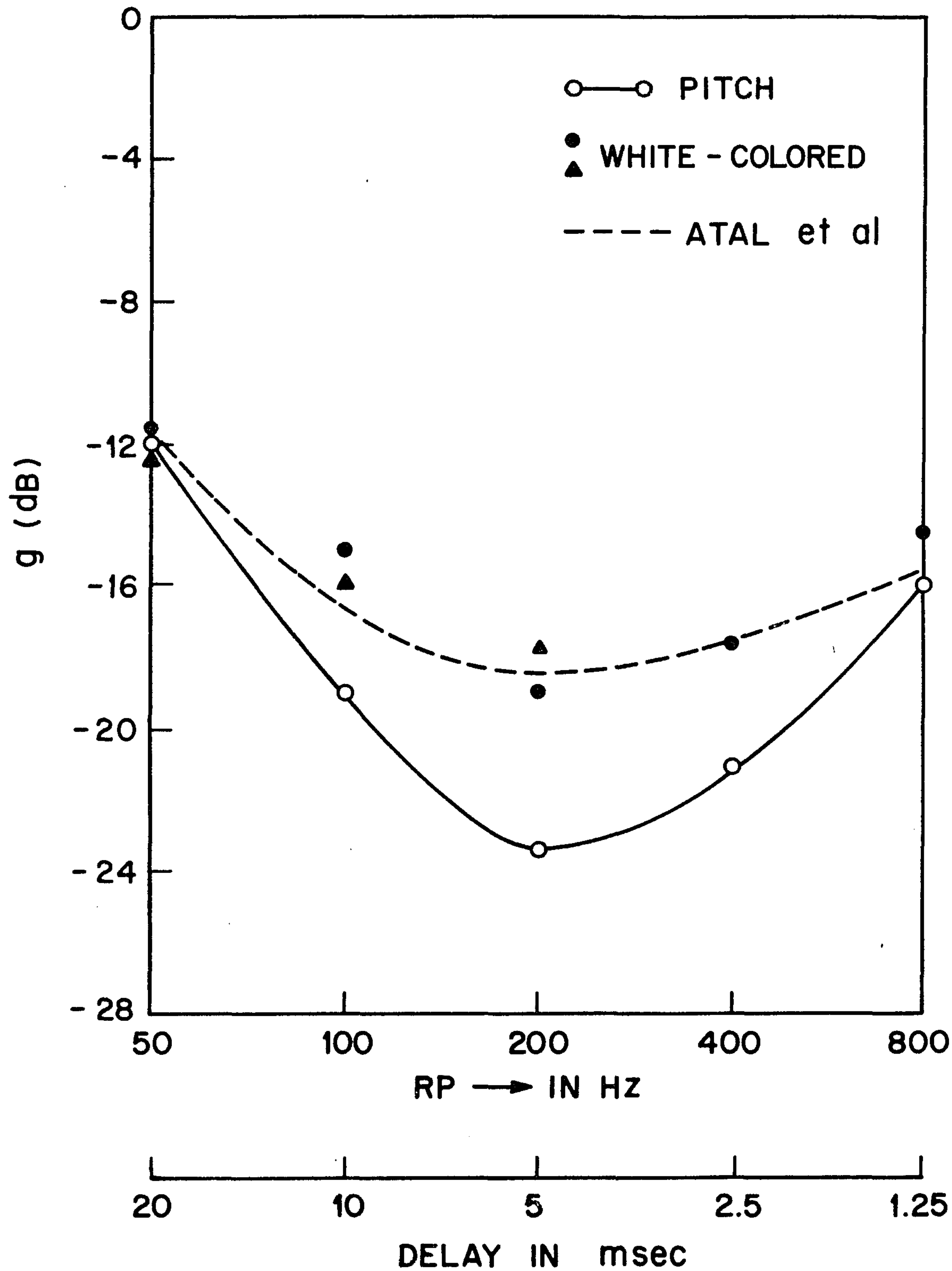
Bilsen and Ritsma (1970) examined some of the parameters which could markedly affect the perception of Repetition Pitch. The results showed that: (1) digital noise with its repetition produces optimal pitch perception when compared to pitch perception with individual pairs presented repetitively. (2) For low sensation levels below about 30 dB, Repetition Pitch is perceptible if the absolute level of the repetition is just above the hearing threshold. From 30-70 dB the level of the delayed signal necessary to produce a pitch sensation was independent of sensation level. (3) Perceptibility of RP is a function of delay time. As shown in Figure 2.18, an optimal RP occurs at 200 Hz corresponding to a 5 msec delay time. Although the exact experimental procedure is not reported, this figure shows the level of the repetition (g) as a function of delay time required to obtain a pitch sensation.

Bilsen and Ritsma (1970) also measured the threshold for perceived coloration as a function of delay time. Subjects were presented with two successive stimuli, a digital Gaussian noise, followed its attenuated repetition or echo. After each presentation of the pair, subjects had to indicate whether the second stimulus was colored or white with respect to the first. The average results are plotted in Figure 2.18 as closed circles. Also plotted in this figure, using a dotted line, are the results reported by Atal et. al. for the threshold of coloration as a function of the relative level of the repetition (g). Besides noting the close correspondence between the two sets of data, it should be noted that the pitch sensation also plotted in this figure fades away beyond 800 Hz (a delay of 1.25 msec) as well as below 50 Hz (a delay of 20 msec) leaving all remaining judgements to be based on timbre perception alone.

Dichotic Pitch Perception Studies

Huggins (1954, cited in Licklider, 1956) and Cramer and Huggins (1958) demonstrated that a faint pitch-like sensation could be created using a single Gaussian noise source and presenting to one ear an unaltered version of the noise while the other ear received the same noise source with a narrow frequency region phase transformed. An all-pass filter served to produce a sharp phase change or shift in a

Figure 2.18 The threshold of perceptibility of pitch (shown as open circles) and the threshold of perception of coloration in g (shown as closed circles) as a function of RP. The closed triangles represent control measurements using uncorrelated masking noise. (From Bilsen and Ritsma, 1970).


PITCH
WHITE - COLORED
ATAL et al
0
-4
-8
-12
-16
-20
-24
-28
g (dB)
50
100
200
400
800
RP → IN Hz
20
10
5
2.5
1.25
DELAY IN msec

small frequency region while passing all audible frequencies without amplitude change. This pitch-like sensation was shown to correspond to the frequency of the all-pass filtered noise. The later study, Cramer and Huggins (1958) showed that the effect was most pronounced for low-frequency, narrow banded phase shifts. Moore (1977) has suggested that the phase shifting of the narrow noise band produces a difference in localization (lateralization in this case) so that this narrow-band of noise is perceived separately from the rest of the noise in subjective space.

Fourcin (1970) showed that a pitch-like sensation could be perceived when two independent sources were simultaneously fed to both ears (through headphones) with a delay placed in the leg of one source feeding one ear. In this configuration, the output of one noise source arrives simultaneously at both ears and produces a sound image lateralized in the center of the head. The output of the second source reaches one ear later (eg. following a fixed delay) than the other ear and produces a second sound image lateralized to one side. Some subjects reported a faint pitch whose position was not well defined but which was centrally located. Frequency matching to pure tones showed the pitch to be related to the delay used. No pitch sensation was obtained when a single source was used with a delay at one ear.

More recently, Bilsen and Goldstein (1974) have shown that a single source can be used to create a pitch-like sensation. A continuous wide band noise is presented to one ear while the same noise delayed in time is presented to the other ear. For a delay shorter than approximately 2 msec, Bilsen and Goldstein report that the noises fuse and a single noise image is perceived whose position depends on the delay. When the delay is increased beyond 2 msec, the noise image remains at one side of the head but becomes more diffuse. Their observation was that in addition to this increase in diffuseness, a faint but distinct pitch could be perceived in the middle of the head corresponding in frequency to $1/\tau$. Furthermore, this dichotic pitch is quite similar to the monaural repetition pitch reviewed above having both subjective pitch and timbre properties. Whereas the monaural repetition pitch was reported for delays which varied from $\tau \geq 1$ msec to $\tau \leq 10$ msec, dichotic repetition pitch exists for $10 > \tau \geq 3$ msec. When a second independent but statistically similar noise source was applied 180° out of phase to both ears, no change in image location or pitch was observed. However, when the second noise source was added in phase to both ears, it produced a change of image location within the head along with a pitch change. Bilsen and Goldstein propose that the centrally represented spectral patterns which have been created through binaural addition can account for these results.

A binaural power spectrum analyzer model fashioned after Durlach's (1963, 1972) equalization and cancellation model of binaural signal detection is presented to account for the binaural addition. As shown in Figure 2.19, binaural output is formed as the result of monaural band-pass filtering (critical band filters), noisy binaural addition, time averaging of binaural power and ensemble averaging of the binaural parameters. The modulation depth in the power spectra deteriorates both for large and small values of interaural delay. For large values of interaural delay the deterioration is due to deteriorating resolution of the monaural filter at low frequencies (recall large delay produces low pitch frequencies). For small delays, the deterioration is due to random delays or jitter in the binaural addition. A similar model is presented for monaural repetition pitch and is reproduced in Figure 2.20. The authors point out that since the monaural spectrum does not suffer any loss in modulation due to random internal delays, monaural repetition pitch can exist for higher pitch frequencies (small delays) than can dichotic repetition pitch.

Comparison of Single Reflection Thresholds For Diotic and Dichotic Listening Modes

As already described experimenters have examined monaural and diotic pitch perception thresholds, single reflection thresholds in the diotic listening mode as well

Figure 2.19 Binaural power spectrum analyzer after Durlach (1963). Shown for one characteristic frequency. (From Bilsen and Goldstein, 1974).

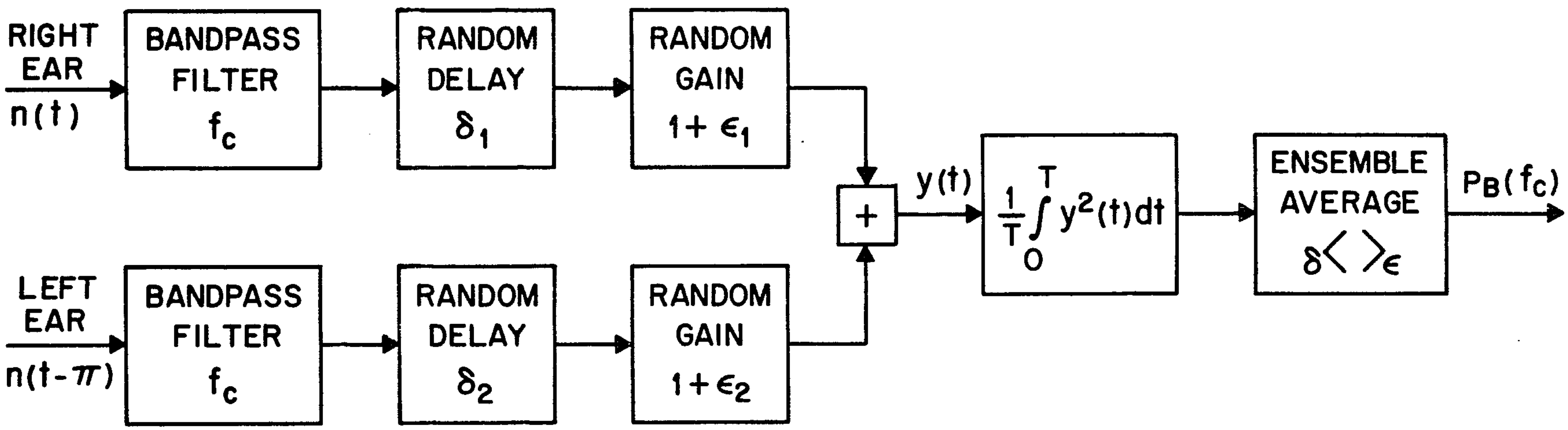
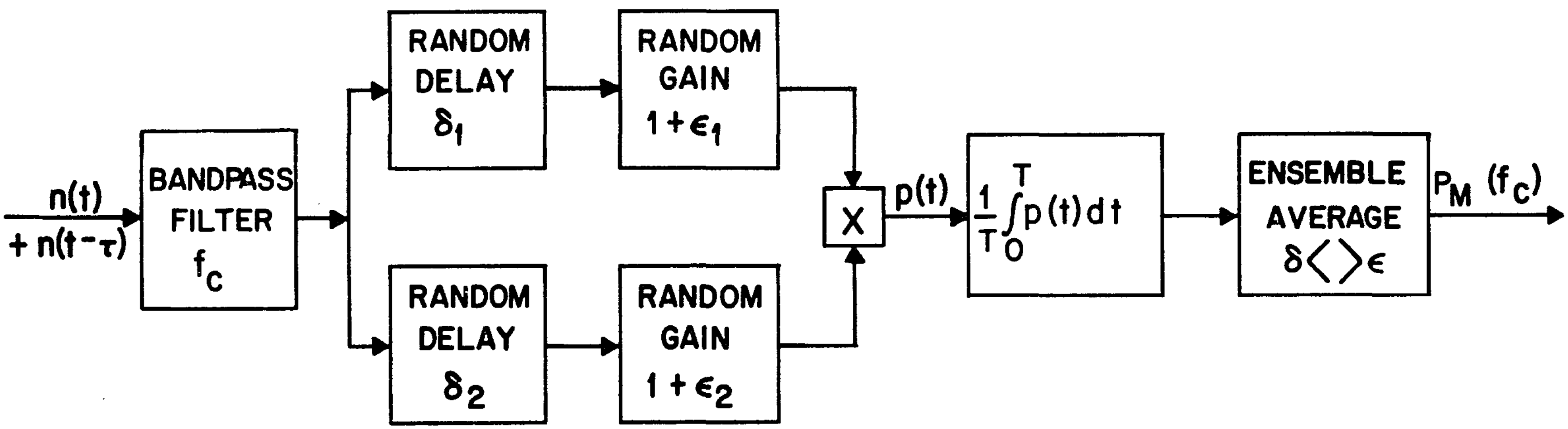

RIGHT EAR
n(t)
BANDPASS FILTER f_c
RANDOM DELAY δ_1
RANDOM GAIN $1+\epsilon_1$
LEFT EAR
$n(t-\tau)$
BANDPASS FILTER f_c
RANDOM DELAY δ_2
RANDOM GAIN $1+\epsilon_2$
+
y(t)
$\frac{1}{T}\int_0^T y^2(t)dt$
ENSEMBLE AVERAGE
$\delta\langle\ \rangle_\epsilon$
$P_B(f_c)$

Figure 2.20 Monaural analog of binaural power spectrum analyzer. (From Bilsen and Goldstein, 1974).


n(t)
+ n(t-τ)
BANDPASS FILTER
f_c
RANDOM DELAY
δ_1
RANDOM GAIN
$1+\epsilon_1$
RANDOM DELAY
δ_2
RANDOM GAIN
$1+\epsilon_2$
X
p(t)
$\frac{1}{T}\int_0^T p(t)\,dt$
ENSEMBLE AVERAGE
$\delta\langle\ \rangle\epsilon$
$P_M(f_c)$

as dichotic pitch perception thresholds. None of these studies were directed however at the role of binaural (specifically dichotic) processing as a mechanism of echo suppression.

As mentioned in the Introductory Chapter, A. Koenig et al. (1975) reported that listeners were less sensitive to the presence of single (2 msec) reflections for the dichotic vs diotic listening mode. Dichotic stimuli were created by reversing the phase of the reflection in the impulse response (a more complete explanation of this procedure is presented in the Methodology Chapter of this dissertation).

More recently, Zurek (1976) in a doctoral dissertation, studied the perceptability of single echoes for diotic and dichotic listening modes. Low-passed (5 kHz), gated noise stimuli (10 msec rise and decay envelope), 250 msec in duration were used as stimuli. The reverberated noise bursts were created by summing the initial noise source with a delayed version of itself. The delays used ranged from 1 to 40 msec. In the diotic listening mode, the same signal was presented to both ears over headphones; a dichotic listening mode was created by adding an additional delay of 500 μsec to one ear beyond the delay in the opposite ear. The additional delay in one ear served to alter the spectrum of the signal received so that signals of different spectra were received at each ear. A two-alternative forced-choice adaptive psychophysical method was used with a 70.7% correct performance level defined as threshold. The results are shown in Figure 2.21. For short delays (ie. $\Delta t <$ ~7 msec), diotic thresholds were lower than dichotic thresholds. The diotic-dichotic thresholds, for all subjects, however, exhibited a crossover effect with increasing delay. Diotic thresholds decreased (ie. became less sensitive) while dichotic thresholds increased (ie. became more sensitive) as delay increased.

A follow-up experiment using independent sources (one to each ear) was also reported by Zurek. The diotic stimuli were similar to the diotic stimuli reported in the earlier experiment in that they possessed the same spectrum; however, due to the independence of the noise sources, a zero interaural correlation existed in the follow-up experiment. Dichotic stimuli were created by inverting the phase (ie. 180° phase shift) of the delayed signal to one ear. Diotic and dichotic listening mode thresholds were obtained using the same psychophysical procedure described in the first experiment. As shown in Figure 2.22, the results demonstrate substantially lower thresholds for the diotic vs dichotic listening mode. Although the diotic thresholds, once again, increase with increasing delay, the dichotic thresholds remain relatively constant as delay is increased. The cross-over effect seen for the 40 msec delay case is now almost entirely the result of the increasing <u>diotic</u> threshold. Zurek ascribes the diotic-dichotic threshold

Figure 2.21 Detection thresholds for a signal that was a delayed version of the direct signal. The signal was presented with an interaural delay of either 0 or 500 μsec. (From Zurek, 1976).

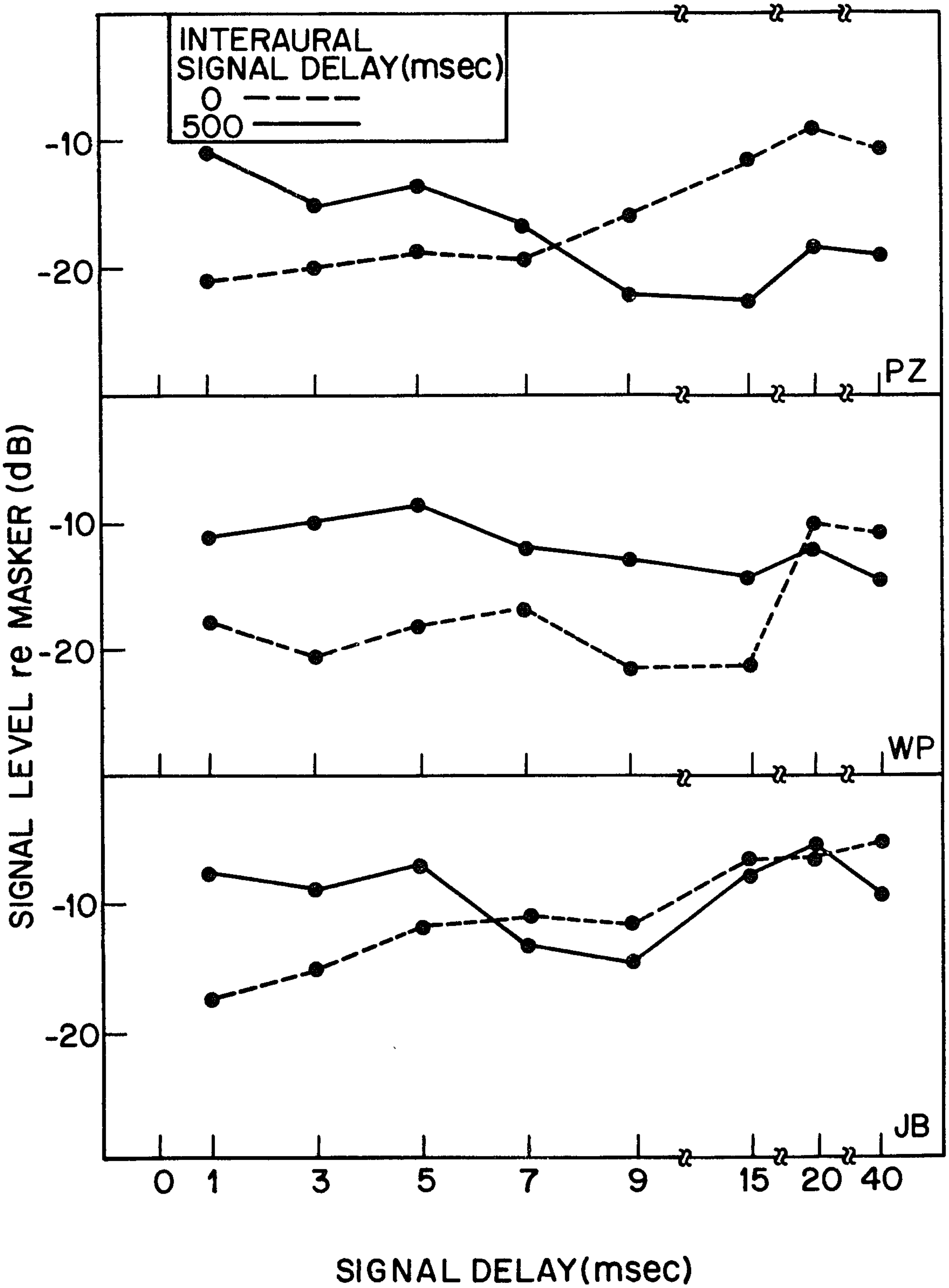
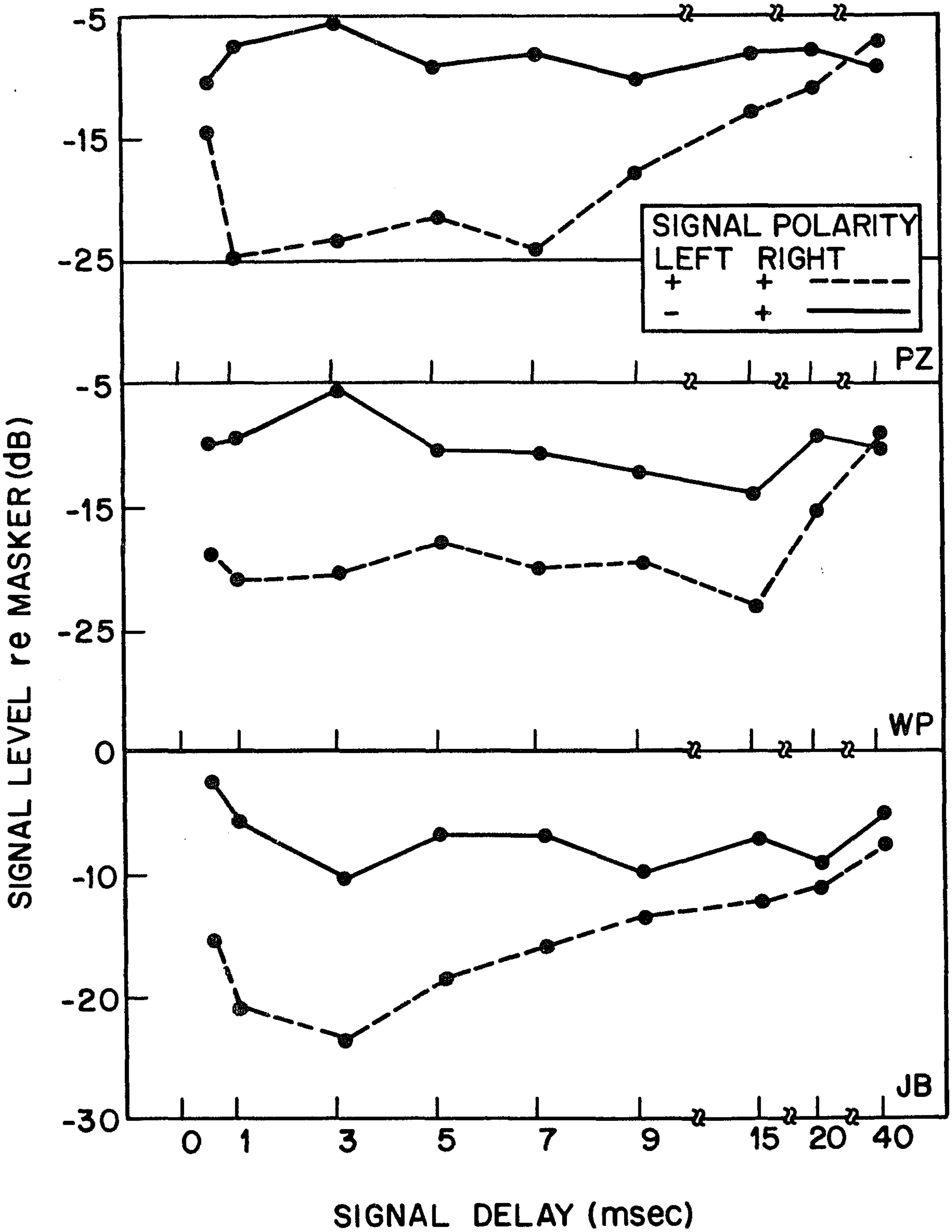

INTERAURAL SIGNAL DELAY (msec)
0
500
-10
-20
PZ
WP
JB
SIGNAL LEVEL re MASKER (dB)
0 1 3 5 7 9 15 20 40
SIGNAL DELAY (msec)

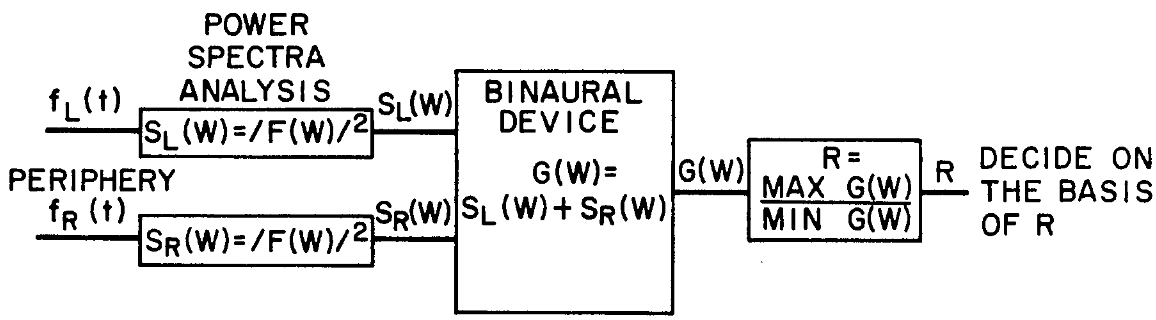
Figure 2.22 Thresholds for the detection of an echo which is either added to each earphone channel with the same polarity as the undelayed signal (+,+ dashed line) or subtracted in one channel (+,-, solid line). (From Zurek, 1976).


SIGNAL LEVEL re MASKER (dB)
SIGNAL DELAY (msec)
SIGNAL POLARITY
LEFT RIGHT
+ +
- +
PZ
WP
JB
-5
-15
-25
0
-10
-20
-30
0 1 3 5 7 9 15 20 40

difference to spectral flattening produced by some form of binaural summation of the power spectra. Perfect power summation in the dichotic case, however, would always result in a flat spectrum.

In a third experiment, Zurek attempted to examine the validity of the simple binaural summation model. Zurek's model shows individual power spectral analysis (See Figure 2.23) of the incoming time signal f_L and f_R being performed at each ear. The monaural power spectra are then fed into a Binaural Device which produces a summed power spectra $G(\omega) = S_{L(\omega)} + S_{R(\omega)}$. The decision as to whether coloration is perceived is based on the peak to valley ratio, $R = \frac{MAX(\omega)}{MIN(\omega)}$ or the modulation depth of the binaural spectra. By manipulating the overall level present at one ear, the validity of this model could be explicitly tested. That is, with one ear completely attenuated, no spectral flattening due to binaural summation can exist. Systematically increasing the level in the attenuated ear until the overall level was equivalent to the overall level in the unattenuated ear (assuming equal amplitude of the positive going and phase inverted negative going delay) however, should the growth function of this effect. Zurek found that while his model explicitly called for almost no spectral flattening when one ear was attenuated by 10 dB, the data in fact, showed residual spectral flattening until an attenuation level of approximately 25 dB was reached. That is, despite a

Figure 2.23 Binaural power summation model. (From Zurek, 1976).


POWER
SPECTRA
ANALYSIS
$f_L(t)$
$S_L(W)=/F(W)/^2$
$S_L(W)$
PERIPHERY
$f_R(t)$
$S_R(W)=/F(W)/^2$
$S_R(W)$
BINAURAL
DEVICE
$G(W)=$
$S_L(W)+S_R(W)$
$G(W)$
$R=\frac{\text{MAX } G(W)}{\text{MIN } G(W)}$
R
DECIDE ON
THE BASIS
OF R

substantial reduction (20-30 dB) of the signal in one ear, precluding any spectral flattening by a simple binaural addition, the data indicated that the reduced signal still served to reduce the perception of coloration. Zurek concluded that these results showed this model to be inadequate.

The Proposed Study

The first experiment reported in the present dissertation examines the detectability of single echoes for both diotic and dichotic listening modes. The dichotic stimulus is created by presenting a single reflection added to a direct sound to one ear, while the other ear receives the same direct sound with the reflection reversed in polarity. Use of this dichotic configuration serves to produce the largest obtainable differences in interaural spectral modulations. That is, a spectral peak in one ear coincides with a spectral null in the other ear. This is the result of the phase-reversed reflection which serves to shift the spectrum present in one ear by one-half cycle relative to the other ear. The diotic stimulus consists of a direct sound followed by a single reflection (positive polarity) presented to both ears. The second experiment reported in this dissertation examines listening mode differences, for the perception of single reflections, at above absolute threshold levels. That is, using the same stimulus configuration, described above, the necessary increment in the level of the reflected signal required to obtain a just-noticeable difference, is measured. The second experiment therefore determines whether effects measured in the absolute threshold experiment, persist at levels above threshold.

CHAPTER III

METHODOLOGY

Stimulus Generation and Processing

A Data General S/200 digital computer capable of acoustic signal processing was used for the generation of all stimuli used in this study. The acoustics signal processing facility is supported by the Acoustics Research Department of Bell Laboratories at Murray Hill, New Jersey.

Computer Processing of Stored Noise Waveforms

The output of a Gen Rad (Model #1382) random noise generator was low-pass filtered at 6 kHz (48 dB/octave roll-off) and fed into one channel of an eight-channel 15-bit Phoenix Data (Model DAS 6000) analog-to-digital (A/D) converter with simultaneous sample and hold. The sampling rate was fixed at 15 kHz and the input level was adjusted to obtain full-scale input without overload. After storing the file on disk, it was edited to obtain 25 (256-word) contiguous blocks, each block containing 17.06 msec of the noise waveform (17.06 msec = $\frac{256\text{wds/blk}}{15\text{wds/msec}}$). The total duration of the noise waveform stored in the computer was thus 25 x 17.06 = 426.6 msec. The frequency spectrum of the stored noise waveform is shown in Figure 3.1. The spectrum is flat within ± 2 dB from 50 Hz to 5 kHz followed by a rapid roll-off of 48 dB/octave beyond 6 kHz. The spectrum was obtained with a Rockland Spectrum analyzer with the bandwidth set at 10 kHz. The noise file was played back cyclically into the analyzer so as to obtain a spectrum average of 128 passes.

Digital Filtering Process

A general purpose filtering program (Allen, 1979) was used in the processing of the stored noise waveform. The following functional description of this program outlines the essential processing steps:

a. An impulse response (IR) is created by entering the appropriate amplitude and delay values;

b. The IR is high-passed filtered at 150 Hz with approximately 6 dB/octave roll-off;

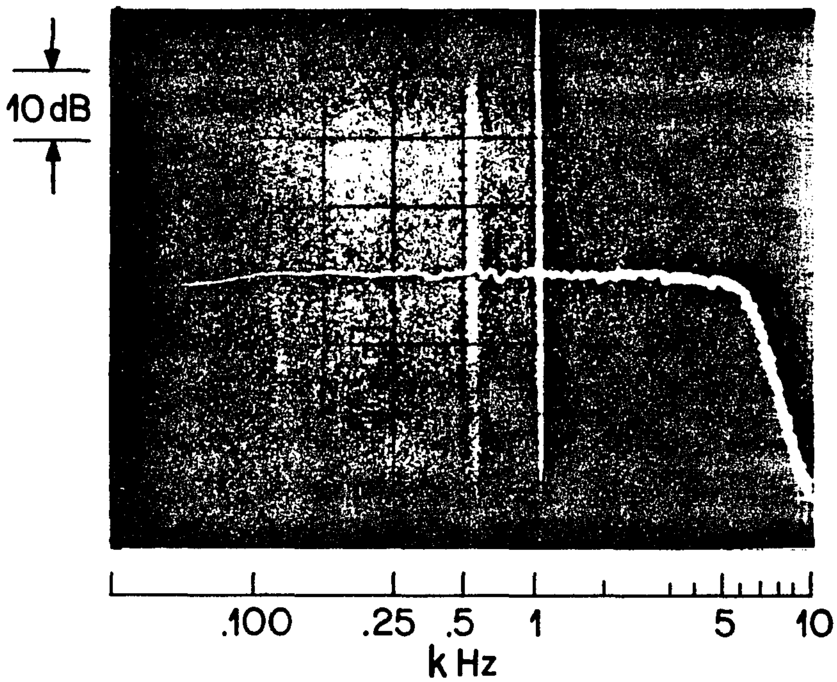
c. The high-passed IR is intensity equalized by adjusting the overall RMS level to a given arbitrary level which was the same for all impulse responses.

d. the noise waveform is convolved with the high-passed, equalized IR.

e. The processed noise is then truncated to a 374 msec duration with a linear 20 msec rise and decay.

The processing scheme described above was designed to deal with complex impulse response structures and is more sophisticated than is necessary for this study. The processed noise waveforms used in this study consisted of a noise waveform added to itself with a predetermined delay. This type of processing could have been obtained with the

Figure 3.1 Power spectrum of stored noise waveform. Abscissa - log frequency in Hz. Ordinate - relative amplitude in dB (1Ø dB per div).


10 dB
.100
.25
.5
1
5
10
k Hz

simple technique shown schematically in Figure 3.2. The resulting time and frequency domain information is also shown in this figure.

The addition of a white Gaussian noise waveform to itself delayed for a specified time period, τ, (i.e. a reflection) produces a signal whose spectral power density is a cosinusoidal function of frequency with maxima at n/τ $(n = 0,1,2,3,\ldots)$ and minima (valleys) at $(n+1/2)/\tau$ (Bilsen, ten Kate, Brunner, and Raatgever, 1975). The modulation depth of the cosinusoidal power spectrum is varied by changing the intensity of the delayed noise relative to the undelayed noise. This power spectrum has been mathematically derived (Bilsen et al, 1975) as:

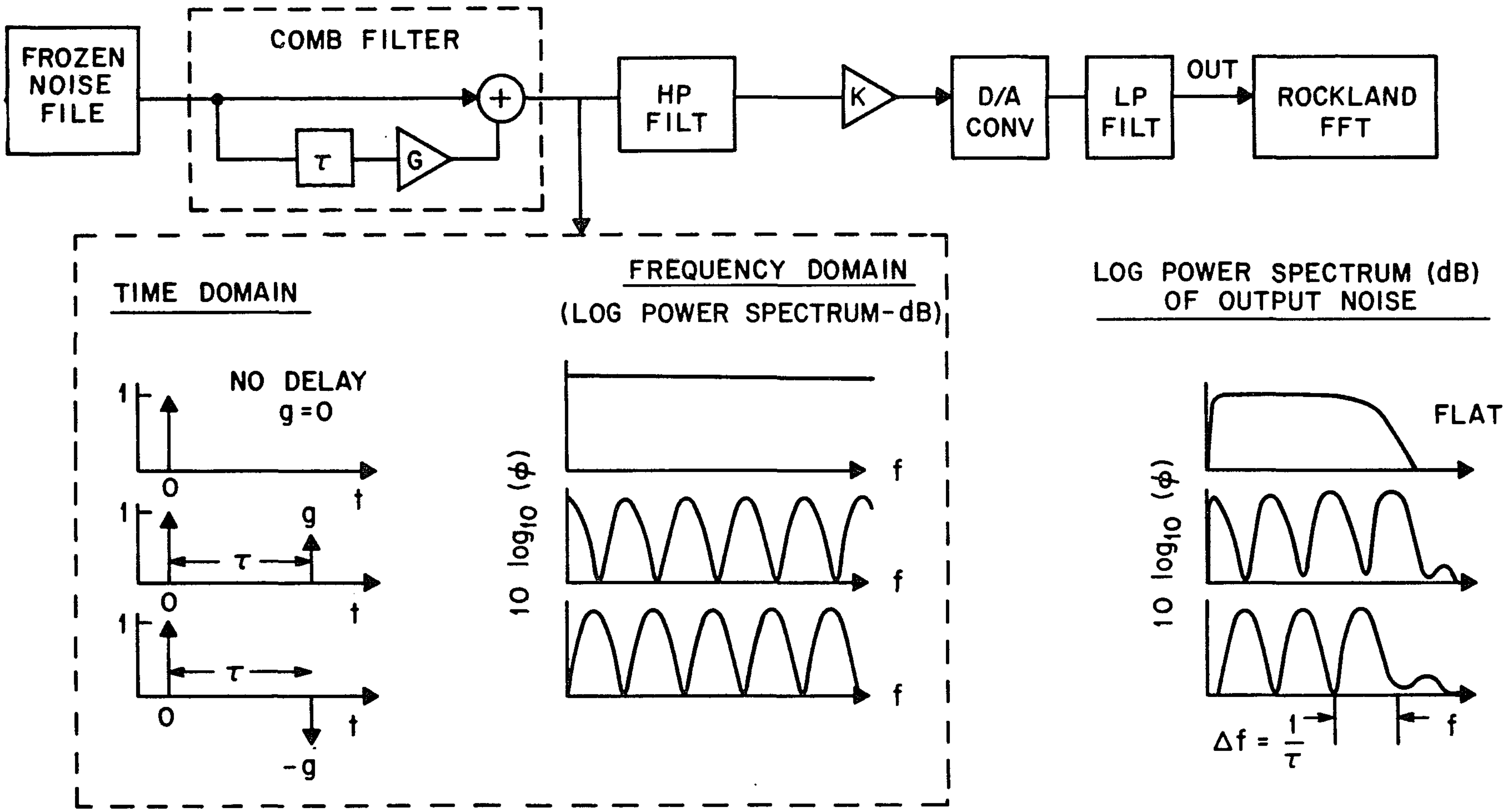
$$\phi(f,\tau,m) = \phi_0(1+m\mathrm{Cos}2\pi f\tau) \qquad (3.3.1)$$

where τ is the delay between the undelayed and delayed noise, m is a measure of the modulation depth of the spectrum; ϕ_0 is the average spectral power density. The relationship of the modulation depth and the level of the delayed signal is expressed by the equation:

$$m = \frac{2g}{(1+g^2)} \qquad (3.3.2)$$

where g represents the amplitude of the delayed noise relative to the undelayed noise. As shown in Figure 3.2, if the level (g) of the delayed noise is set to zero, a flat spectrum is obtained. When $1 \geq g > 0$ a modulated or rippled

Figure 3.2 Schematized functional diagram of digital processing technique.



COMB FILTER OUTPUT

τ; 2-16 msec

g; 0 → ±1

K; OVERALL SCALING FACTOR
FOR RMS EQUALIZATION

frequency spectrum is produced. This method of shaping the frequency spectrum is known as comb filtering. The frequency spacing of the teeth in these comb filters is given by:

$$\Delta f = 1/\tau \qquad (3.3.3)$$

where Δf represents the frequency spacing in Hz and τ represents the delay of the second impulse (see Figure 3.2) in seconds. When the delayed impulse is phase reversed the maxima (peaks) and minima (valleys) in the amplitude spectrum are shifted in frequency by one-half cycle. We see therefore, that positive delayed impulse (where $1 \geq g > 0$) and negative delayed impulses (where $-1 \leq g < 0$) produce complementary spectra with minima and maxima in opposite phase (See Figure 3.2). When the delayed signal is equal in absolute level to the undelayed signal ($g = \pm 1$), the modulation depth is 100% (ie. the spectrum density at a peak is equal to twice the unprocessed spectrum power density, and the spectrum density at a valley is zero). The overall scaling factor designated K in Figure 3.2 provides for RMS equalization. Corliss and Winzer (1965) have shown that RMS equalization serves as a good approximation of loudness equalization for broad band noise bursts.

For the experiments reported in this dissertation, 20 log (g) represents the level difference in dB between undelayed and delayed noise.

Merging of Monaural Processed Noise Waveforms to Create Diotic and Dichotic Stimuli

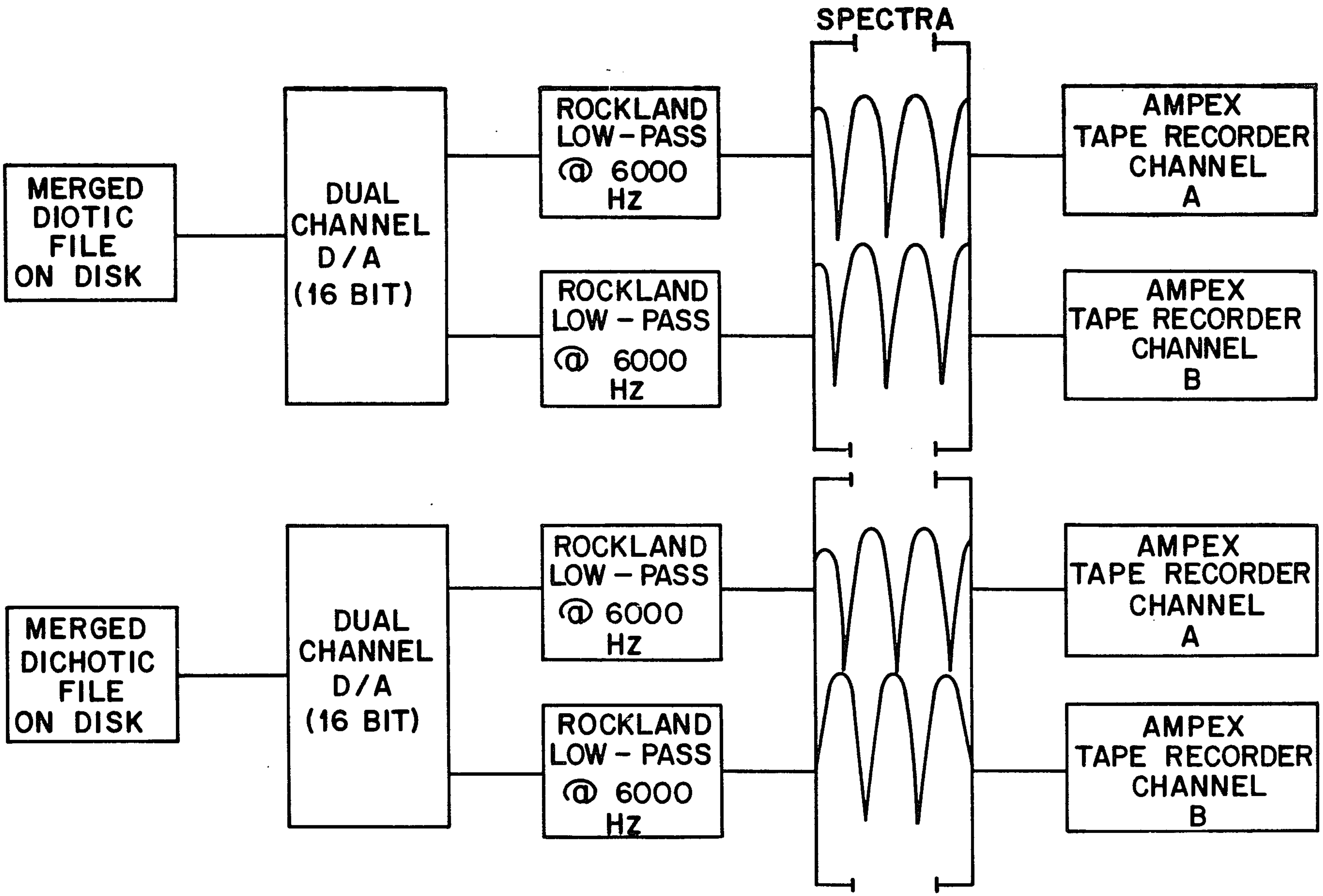
The processed noise waveforms are merged to create two-channel (diotic and dichotic) waveforms using an interleave program. Interleaved files are required so as to obtain simultaneous 2-channel output from the digital-to-analog (D/A) device used.

Creation of Analog Magnetic Tapes

The final analog magnetic tapes were created by outputting the stored interleaved files through two channels of a 4-channel Phoenix Data 16-bit D/A converter with simultaneous loading. The digital output from each channel was low-passed through two separate banks of a Rockland filter (Model 1042) each set at a cutoff frequency of 6000 Hz with a roll-off of 48 dB/octave. (The Rockland filter was always used in the Butterworth mode.) The low-passed analog outputs were then fed into a two-channel Ampex tape recorder (Model AG-350) and recorded at a speed of 7-1/2 inches per second.

A schematized representation of the resulting merged files is shown in Figure 3.3. The upper portion of this figure shows the generation of diotic stimuli where the output of both D/A channels is identical. Dichotic stimuli, shown in the lower portion of this figure are generated by simultaneously outputting complementary spectra from each D/A channel. For both listening modes, however, each D/A

Figure 3.3 Schematized representation of merged binaural stimuli files.


SPECTRA
MERGED DIOTIC FILE ON DISK
DUAL CHANNEL D/A (16 BIT)
ROCKLAND LOW-PASS @ 6000 Hz
ROCKLAND LOW - PASS @ 6000 Hz
AMPEX TAPE RECORDER CHANNEL A
AMPEX TAPE RECORDER CHANNEL B
MERGED DICHOTIC FILE ON DISK
DUAL CHANNEL D/A (16 BIT)
ROCKLAND LOW - PASS @ 6000 Hz
ROCKLAND LOW - PASS @ 6000 Hz
AMPEX TAPE RECORDER CHANNEL A
AMPEX TAPE RECORDER CHANNEL B

output was recorded on a separate channel of an Ampex two-channel tape recorder. The bracketed section shown in the upper and lower portions of this figure depicting the stimulus spectra are displayed for instructive purposes only and were not part of the recording system.

Subjects

Five female high school students whose ages ranged from 14 to 16 years were initially recruited. Following an initial four week training period, one subject had to be eliminated from the study since she could not perform the task with any degree of consistency. Each subject participated in the study on a voluntary basis with the agreement that their efforts would be remunerated at the rate of $3.00 per hour. Testing covered a four month period.

Automatic audiometric tests were administered to all subjects using a Grason-Stadler (Model No. 1703) recording audiometer where thresholds are obtained in the form of fixed frequency Bekesy tracings. The results showed that in all cases hearing threshold levels (500-8000 Hz) were within 10 dB of audiometric zero. The recorded Bekesy tracings for each subject are reproduced in Appendix A.

Psychophysical Paradigm - Stimulus Presentation

An ABX paradigm (Rosenblith and Stevens, 1953; Pierce and Gilbert, 1958; MacMillan, Kaplan and Creelman, 1977) was used throughout this study where subjects receive a triad of stimuli such that the first two successive items, A and B are necessarily different in some respect. While the third stimulus X is identical to either the first or second stimulus. The task requires that the subject attempt to distinguish between the initial two stimuli (A and B) and then to determine whether the final stimulus X was more like the first or the second stimulus. The paradigm adopted was a forced-choice task in that subjects were required to guess when they could not make a distinction between the first two stimuli.

The advantage of the ABX paradigm is that subjects need not concern themselves with the nature of the discrimination being examined (Harris, 1952). That is, subjects merely have to perceive a difference between A and B. This paradigm is particularly useful in studies involving difficult discriminations. Although some investigators (eg. Atal et al., 1962) have required subjects to indicate which stimulus, in an A-B paired comparison test, is more colored or reverberant, a preliminary pilot test conducted by this author showed that naive subjects have a great deal of difficulty with this distinction.

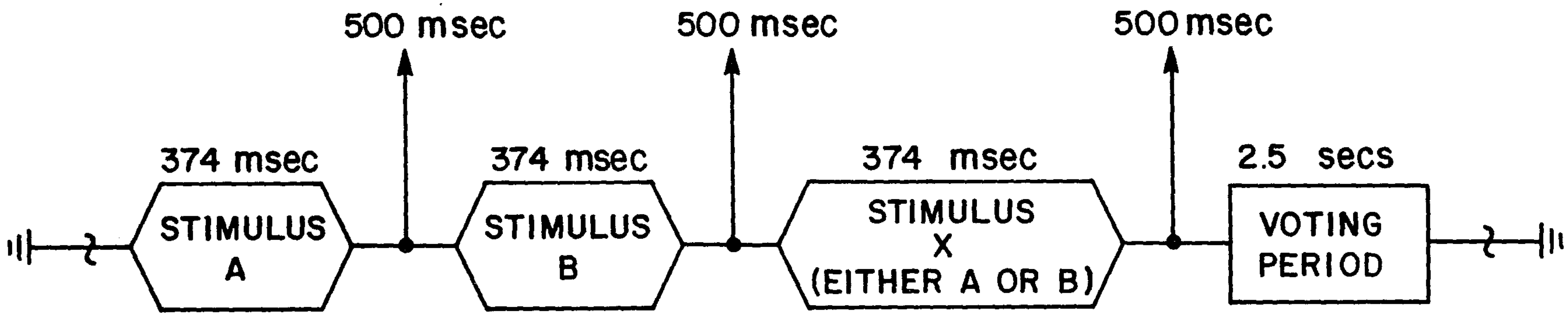
One noteworthy disadvantage of the ABX task is that it places some degree of cognitive loading on the subject, in that the judgement requires the subject to recall the first two stimuli after receiving the third so as to make the appropriate match. An incorrect response can therefore occur either as the result of not being able to distinguish the first two stimuli or as the result of not being able to recall the order in which the stimuli were presented.

A diagram showing the sequence of events in each trial is shown in Figure 3.4. As shown in this figure, each trial consisted of: presentation of Stimulus A, an inter-stimulus period of 500 msec, presentation of Stimulus B followed by a second inter-stimulus period of 500 msec, presentation of Stimulus X (the stimulus to be matched to either stimulus A or B), followed by a voting period of 3.0 sec. During this period, the subject was asked to indicate whether the final stimulus was identical to the first or second stimulus of the trial. The actual response was recorded on an answer sheet by having the subject circle either 1 or 2. Each trial lasted for approximately 5 seconds. Each experimental session consisted of 4 units of 52 trials with a short break of between 5-10 minutes between units. The duration of each unit was approximately 5 minutes and the total duration of an experimental session including breaks ranged from 35-45 minutes. The breaks served to permit the changing of analog tapes, calibration of output levels, as well as to give the subjects a rest.

Experimental Conditions

This study consisted of two separate experiments:

Figure 3.4 Stimulus trial sequence.



STIMULUS TRIAL SEQUENCE

Experiment I - Determination of detection thresholds for the presence of a single reflection; and

Experiment II - Determination of difference limens for a change in level of the reflection.

In Experiment I the subject was asked to discriminate between a flat Gaussian noise source with no reflection present and a filtered Gaussian noise source produced by the presence of a single reflection. For this experiment then, threshold represents the smallest amplitude of the delayed signal which can be reliably detected by the subject. Experiment II on the other hand, required the subject to discriminate between some fixed amplitude value of the delayed signal (referred to as the reference or standard) and some incremented level of the delayed signal.

As shown in Figures 3.5 and 3.6, the same parameters were used in both experiments. These were:

1. Listening Modes - two listening modes were examined:

    a. Diotic Listening Mode - Identical signals applied to both ears.

    b. Dichotic Listening Mode - Sign of the reflected signal was reversed in one ear.

2. Delay times - four values were examined: 2,4,8 and 16 msec. These delay times were selected on the basis of previous diotic experiments (ie. Atal et al., 1962;

Figure 3.5 Experimental conditions used in threshold experiment.

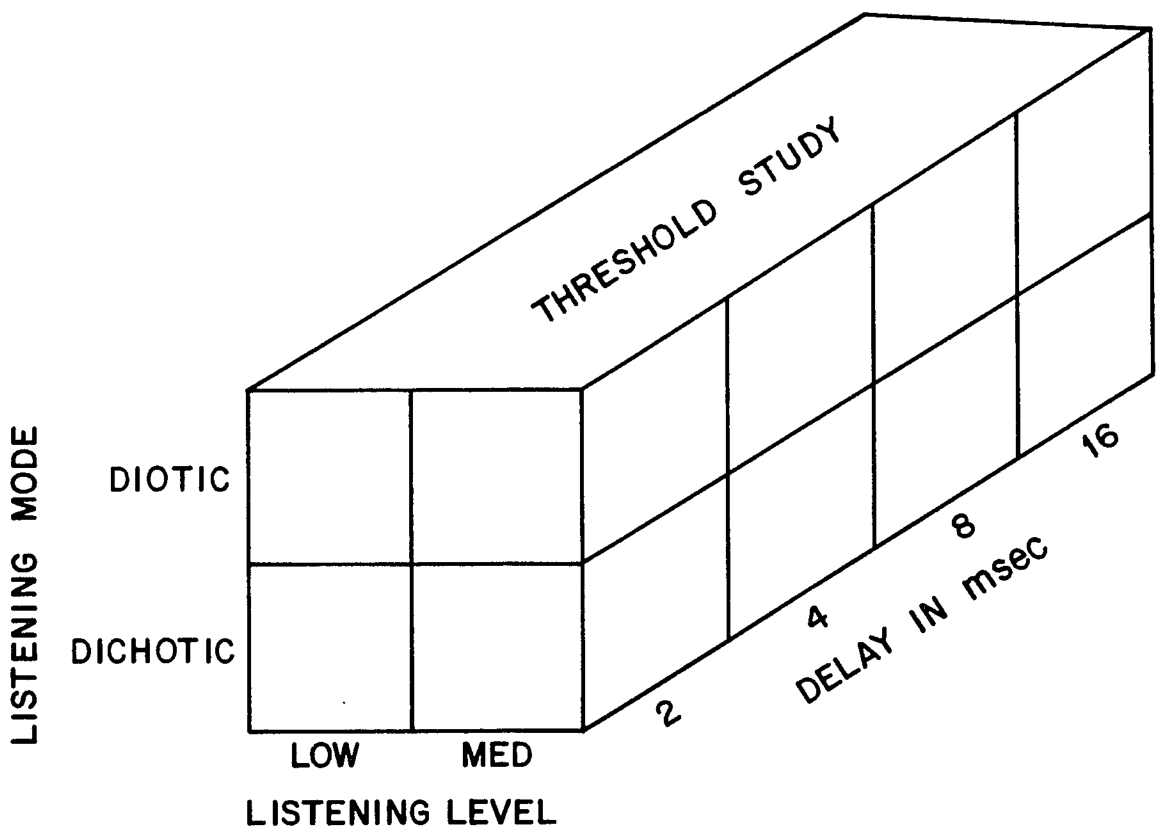
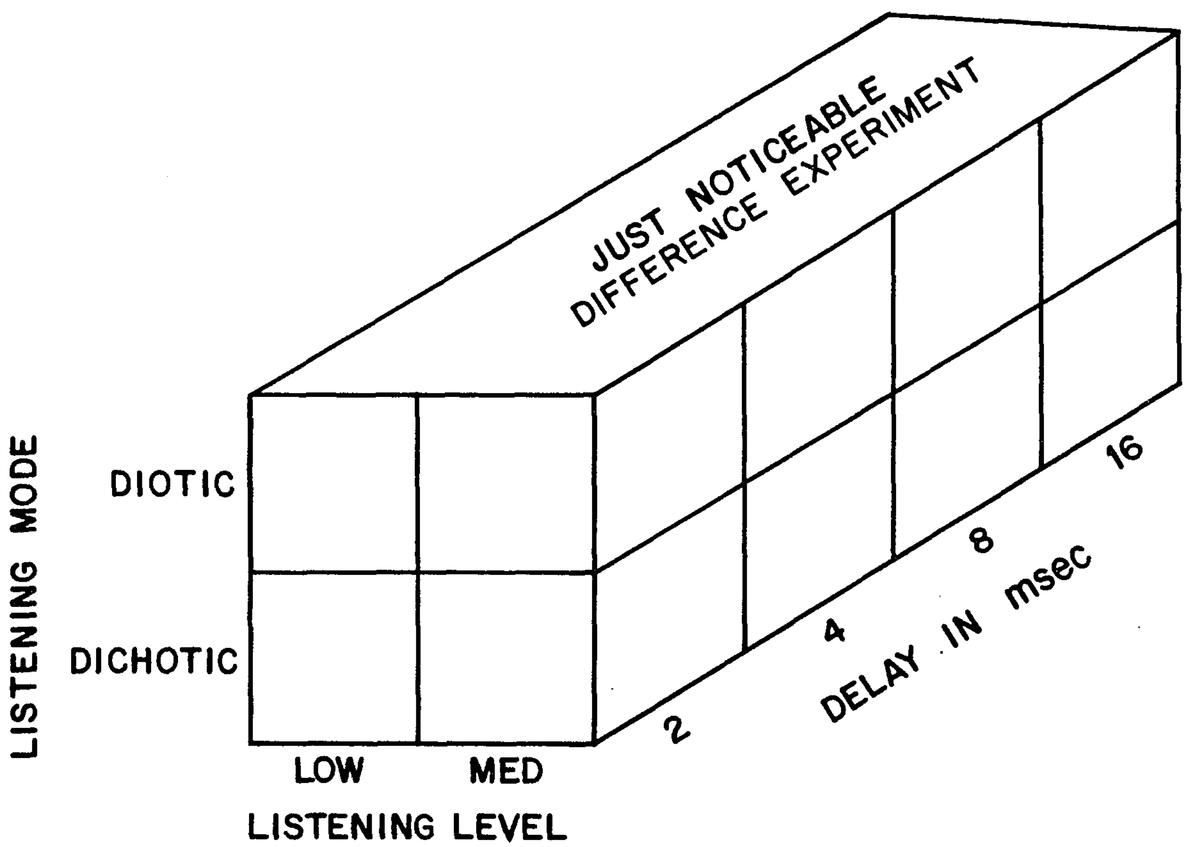

THRESHOLD STUDY
LISTENING MODE
DIOTIC
DICHOTIC
LOW
MED
LISTENING LEVEL
2
4
8
16
DELAY IN msec

Figure 3.6 Experimental conditions used in JND experiment.


JUST NOTICEABLE
DIFFERENCE EXPERIMENT
LISTENING MODE
DIOTIC
DICHOTIC
LOW
MED
LISTENING LEVEL
2
4
8
16
DELAY IN msec

Zurek, 1976) which showed that single reflections are most perceptible for this range.

3. Overall level of presentation - 2 levels were examined:

   a. Medium level - 50 dB sensation level

   b. Low level - 30 dB sensation level

Amplitude levels of the reflected signal tested for each delay time are shown in Table 3.1 and 3.2. Amplitude values are expressed in volts (the first impulse is set at one volt) as well as in dB representing the ratio of the reflected and initial impulse (ie. 20 log g). Overall scaling factors used for RMS equalization in both experiments are also shown in these tables. The scaling values shown in Tables 3.1 and 3.2 can be closely approximated with the formula:

$$\sqrt{\frac{1}{1+g^2}} \qquad (3.8.1)$$

Scale values shown in Tables 3.1 and 3.2 are based on noise file calculations which take the high-pass filtering into account.

Diotic and dichotic stimuli were generated as follows: twelve amplitude levels of the reflected signal were paired with a reference condition (either flat Gaussian noise for the threshold experiment or g= -9.18 dB of reflected signal for the difference limen experiment). A third stimulus

'g' (in Volts)	'g' (in dB)	EQUALIZATION SCALING VALUES (2 msec+) DICHOTIC*	DIOTIC**
Ø.ØØ	ØØ	1.ØØØ	1.ØØØ
Ø.Ø5	-26.Ø2	Ø.9989	Ø.9982
Ø.Ø75	-22.44	Ø.9965	Ø.9974
Ø.1ØØ	-2Ø.ØØ	Ø.9942	Ø.9954
Ø.125	-18.Ø6	Ø.9913	Ø.9928
Ø.15Ø	-16.47	Ø.9878	Ø.9896
Ø.18Ø	-14.89	Ø.9829	Ø.985Ø
Ø.2ØØ	-13.97	Ø.9729	Ø.9815
Ø.225	-12.95	Ø.9741	Ø.9766
Ø.25Ø	-12.Ø4	Ø.9685	Ø.9713
Ø.3ØØ	-1Ø.45	Ø.956Ø	Ø.9592
Ø.35Ø	-9.18	Ø.9419	Ø.9454
Ø.375	-8.519	Ø.9343	Ø.938Ø
Ø.4ØØ	-7.958	Ø.9263	Ø.93Ø2
Ø.45Ø	-6.935	Ø.9Ø96	Ø.9138
Ø.5ØØ	-6.Ø2Ø	Ø.892Ø	Ø.8964
Ø.55Ø	-5.192	Ø.8738	Ø.8783
Ø.6ØØ	-4.436	Ø.855Ø	Ø.8596
Ø.65Ø	-3.741	Ø.8359	Ø.84Ø6
Ø.7ØØ	-3.Ø9	Ø.8167	Ø.8214
Ø.75Ø	-2.498	Ø.7977	Ø.8Ø22
Ø.8ØØ	-1.938	Ø.7784	Ø.783Ø

+ Equalization scaling values for 4, 8 and 16 msec cases deferred by no more than Ø.Ø5.

* Equalization scaling value for the 'phase-shifted delay' noise file.

** Also, used as equalization scaling value for 'zero phase-delay' (positive going) noise file.

TABLE 3.1: Amplitude Values of Reflected Signals Along With Equalization Scaling Values for the Threshold Experiment.

'g' (in Volts)	'g' (in dB)	EQUALIZATION SCALING VALUES (2 msec+) DICHOTIC*	DIOTIC**
Ø.35Ø	-9.18	Ø.9419	Ø.9454
Ø.375	-8.519	Ø.9343	Ø.938Ø
Ø.4ØØ	-7.958	Ø.9263	Ø.93Ø2
Ø.45Ø	-6.935	Ø.9Ø96	Ø.9138
Ø.5ØØ	-6.Ø2Ø	Ø.892Ø	Ø.8964
Ø.55Ø	-5.192	Ø.8738	Ø.8783
Ø.6ØØ	-4.436	Ø.855Ø	Ø.8596
Ø.65Ø	-3.741	Ø.8359	Ø.84Ø6
Ø.7ØØ	-3.Ø9	Ø.8167	Ø.8214
Ø.75Ø	-2.498	Ø.7977	Ø.8Ø22
Ø.8ØØ	-1.938	Ø.7784	Ø.783Ø
Ø.85Ø	-1.411	Ø.7595	Ø.7641
Ø.9ØØ	-Ø.9151	Ø.74Ø9	Ø.7454
Ø.95Ø	-Ø.445	Ø.7226	Ø.727Ø
1.ØØØ	Ø	Ø.7Ø4Ø	Ø.7Ø9Ø

+ Equalization scaling values for 4, 8 and 16 msec cases deferred by no more than Ø.Ø5.

* Equalization scaling value for the 'phase-shifted delay' noise file.

** Also, used as equalization scaling value for 'zero phase-delay' (positive going) noise file.

TABLE 3.2: Amplitude Values of Reflected Signals Along With Equalization Scaling Values for the JND Experiment.

(either A or B) was added to each of these 12 pairs to complete the triad. So as to avoid order artifacts, the triads for each level were balanced so that subjects received all combinations of the triad where the initial two elements were necessarily constrained to be different:

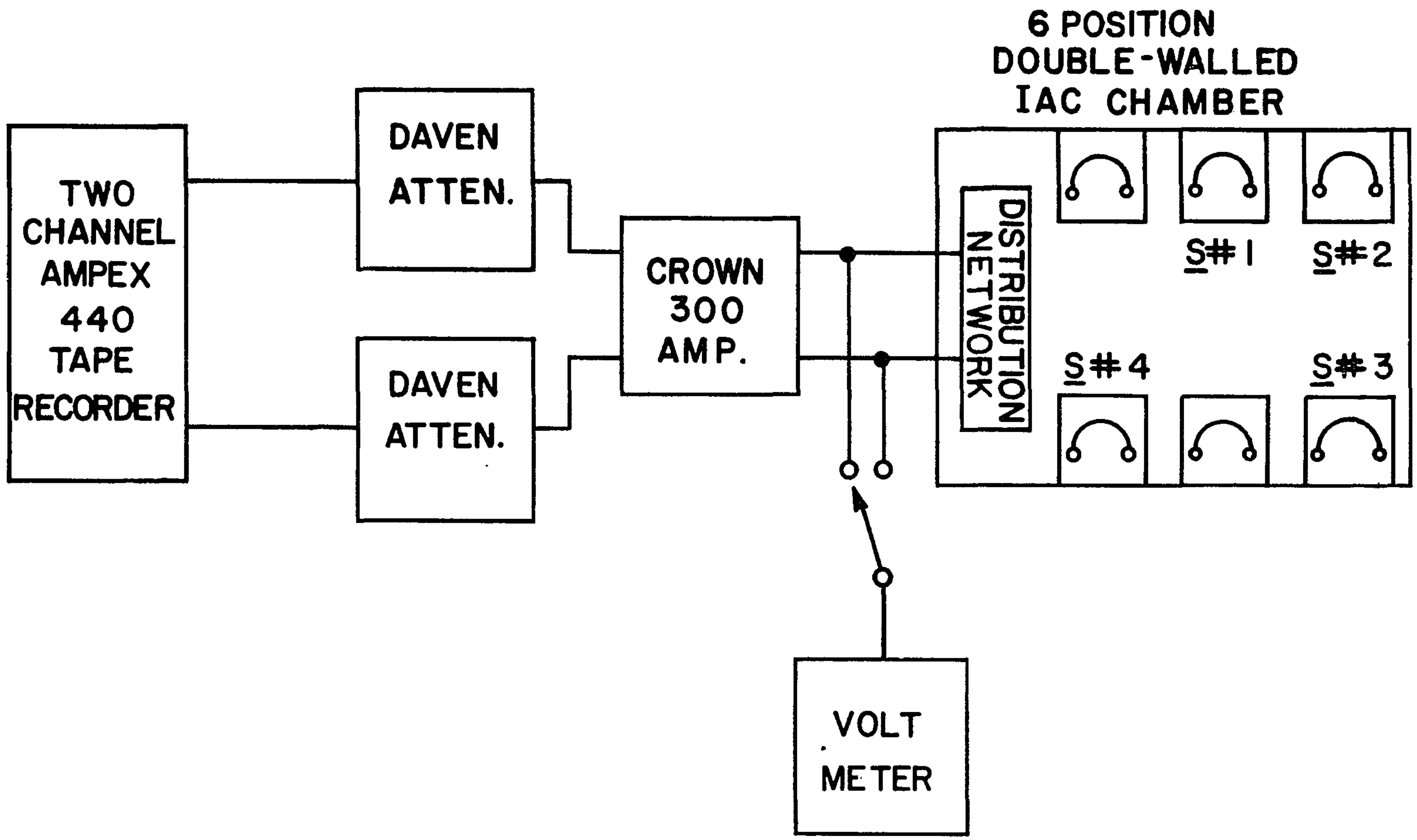
ABA -- ABB -- BAA -- BAB

Each of the 12 amplitude levels was randomly presented four times producing a set of 48 triads plus four AAA triads which served as a check for response bias. For these AAA triads, responses were expected not to significantly differ from a 50% correct score in the absence of a response bias. Finally these 52 triadic stimuli were randomized prior to analog recording.

Stimulus Presentation Apparatus

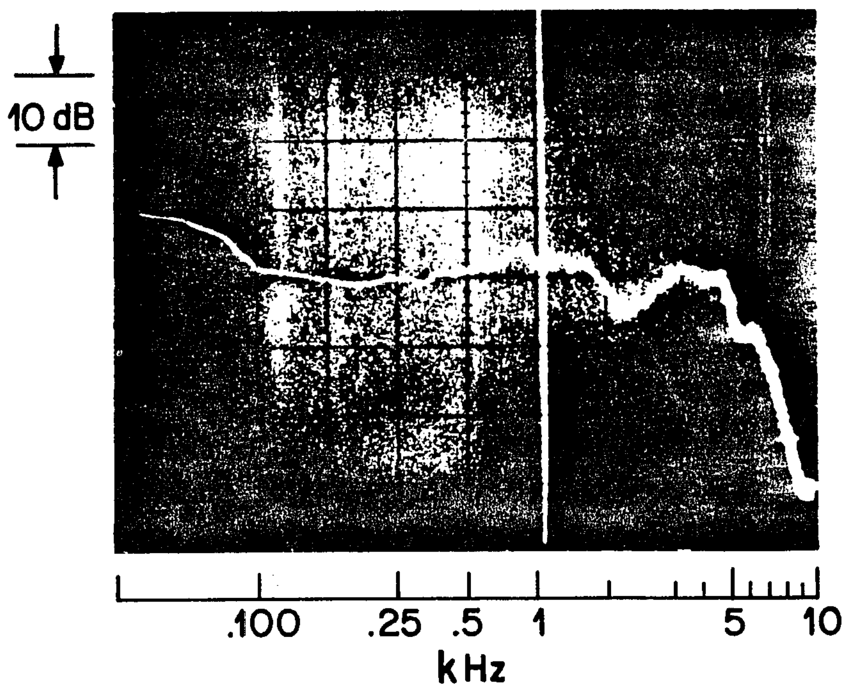
An Ampex 440 tape recorder was used for playback. Both channels were adjusted to produce identical output levels through a calibration tone recorded at the start of each unit on a tape. The output of each playback channel was fed into a separate Daven attenuator whose output in turn was fed into separate channels of a Crown (Model 300) stereo amplifier. The Crown outputs in turn were passed into an IAC double-walled acoustical chamber and were simultaneously made available at four headphones via a distribution network. The listening set-up is shown in block diagram form in Figure 3.7.

Figure 3.7 Block diagram of listening set-up.


TWO CHANNEL AMPEX 440 TAPE RECORDER
DAVEN ATTEN.
DAVEN ATTEN.
CROWN 300 AMP.
VOLT METER
DISTRIBUTION NETWORK
6 POSITION DOUBLE-WALLED IAC CHAMBER
S#1
S#2
S#3
S#4

Stimuli were presented over Pioneer SE-700 headsets with each headset containing two speakers, one for each ear, composed of piezoelectric high-polymer diaphragms. The frequency response of the overall system from the stored noise waveform (low-passed white Gaussian noise) to the headphones is shown in Figure 3.8. These results were obtained with a flat-plate coupler and a one-inch B&K Type 4144 microphone. The ambient noise level present in the IAC accoustical chamber was 35 dBA.

Figure 3.8 Overall system frequency response as measured on a flat-plate coupler. Abscissa - log frequency in Hz. Ordinate - relative amplitude in dB (10 dB/div).


10 dB
.100
.25
.5
1
5
10
kHz

CHAPTER IV

RESULTS

A Psychometric Model

The cumulative normal distribution has been assumed as a model for representing threshold and difference limen data, with the mean (μ) of this distribution defined as threshold and the standard deviation (σ) determining the differential sensitivity. The probability P(x) of a correct response at level x for this study is given by:

$$P(x) = 0.5 \left| 1 + \frac{1}{\sigma\sqrt{(2\pi)}} \right| \int_{-\infty}^{x} \exp \left| -\frac{1}{2\sigma} (x-\mu)^2 \right| dx \quad (4.1.1)$$

where μ and σ represent the mean and variance of the distribution, respectively. Inspection of this equation will show that these two parameters completely determine the function.

Numerical Method for Obtaining Estimates of μ and σ

A method similar to Probit Analysis (Finney, 1964) was used to obtain estimates of μ and σ. The method is based on a parametric formulation which assumes the psychometric function to be cumulative normal (described above). Although Finney makes use of the maximum likelihood approach to obtain the parameter estimates, the actual calculation uses an iterative method of approximation (using tables) that converges on the maximum likelihood solution. The maximum likelihood estimates obtained in this study were obtained numerically using a computerized optimization routine. A sample output of the program is provided in Appendix B. It should be noted that the cumulative normal function shown in equation 4.1.1 is adjusted so as to cover the range of proportion correct from 0.5 to 1.0 rather than 0 to 1.0. This adjustment was necessary since the ABX paradigm produces a chance response rate of 50 percent.

Calculated Estimates of μ and σ for the Threshold Study

Estimates of the parameter, μ, the mean of the cumulative normal distribution for all threshold study conditions are shown in Table 4.1. The mean threshold estimates, $\hat{\mu}$ represent the level of g (the amplitude of the delayed signal relative to the direct signal in dB) required to reach a 75% correct criterion. Recall that subjects were required to select either A or B in the ABX task; the inability to distinguish between stimulus A and stimulus B should therefore produce a 50% correct response rate. For this task then, the 75% correct criterion was defined as threshold. In two instances, the obtained threshold estimate was greater than zero. That is, for these two conditions, a 75% correct level was not achieved. These estimates are shown in parenthesis in Table 4.1.

CONDITION	SUB. NO.	DICHOTIC $(\hat{\mu})$	DIOTIC $(\hat{\mu})$	CONDITION	DICHOTIC $(\hat{\mu})$	DIOTIC $(\hat{\mu})$
2 msec MED	1.	-10.15	-15.81	8 msec MED	-13.01	-15.35
	2.	-3.05	-17.63		-9.98	-8.84
	3.	-7.07	-18.56		-7.52	-13.50
	4.	-6.25	-14.73		-5.43	-8.80
2 msec LOW	1.	-7.59	-14.96	8 msec LOW	-14.97	-14.04
	2.	(235.8)+	-15.22		-7.29	-11.45
	3.	-10.82	-17.91		-7.12	-14.41
	4.	-5.07	-14.11		-5.79	-12.13
4 msec MED	1.	-12.27	-15.09	16 msec MED	-14.68	-15.06
	2.	-5.42	-10.27		-11.25	-7.34
	3.	-6.84	-19.96		-10.29	-6.90
	4.	-4.34	-16.38		-11.80	-7.33
4 msec LOW	1.	-10.00	-16.40	16 msec LOW	-12.25	-14.69
	2.	(7.09)+	-7.94		-7.65	-5.31
	3.	-6.93	-16.89		-9.48	-6.94
	4.	-4.18	-15.03		-5.71	-5.91

+ Parentheses indicate that the obtained threshold estimate was greater than 0.

TABLE 4.1: THRESHOLD ESTIMATES OF μ (in dB)

Obtained threshold estimates where the amplitude of the delayed signal is greater than the direct signal are difficult to interpret since the depth of the spectral null (minima) will in fact begin to decrease as the delayed signal is increased in level above that of the direct signal. To facilitate statistical testing, two separate strategies were adopted to deal with the two cases where the threshold estimate exceeded 0 dB (ie. $g > 0$ dB): (1) a ceiling value of 0 dB was inserted as the threshold estimate; and (2) the two cases were treated as missing observations in the analysis.

Using a mixed model (ie. fixed and random effects), the threshold estimates of μ were entered into an Analysis of Variance (ANOVA) program, (BIOMED, UCLA, 1971). All main effects except subjects were treated as fixed effects. Subjects were treated as a random effect. For the ANOVA results reported in Table 4.2 the ceiling value of 0 dB was used for the two cases described above. This analysis shows that (a) the main effect, listening mode, is statistically significant at the $P < 0.01$ level; and (b) the interaction effect, delay by listening mode, is statistically significant at the $P < 0.01$ level. Using the highest order interaction (ABCD) as the error term, subjects are shown to be significantly different at the $P < 0.01$ level.

The results of the second ANOVA test for threshold estimates ($\hat{\mu}$) where the 2 unobtained thresholds were treated

	SOURCE	ERROR TERM	SIG. LEV.	F	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARE
1	MEAN	D		71.74	6994.81	1	6994.81
2	A Overall Level	AD		4.37	16.72	1	16.72
3	B Delay	BD		0.66	22.32	3	7.44
4	C Listening Mode	CD	$P < 0.01$	64.95	403.40	1	403.40
5	D Subjects	ABCD	$P < 0.01$	52.12	292.48	3	97.49
6	AB	ABD		3.44	16.62	3	5.54
7	AC	ACD		1.40	4.14	1	4.14
8	BC	BCD	$P < 0.01$	10.64	333.26	3	111.08
9	AD				11.48	3	3.82
10	BD				101.24	9	11.24
11	CD				18.63	3	6.21
12	ABC	ABCD		0.72	4.05	3	1.35
13	ABD				14.46	9	1.60
14	ACD				93.88	9	10.43
15	BCD				93.88	9	10.43
16	ABCD				16.83	9	1.87

TABLE 4.2: Analysis of Variance Results for Threshold Estimates* of μ.

* Threshold Estimate Exceeding 0 dB Were Fixed At 0.dB

as missing observations are shown in Table 4.3.

The results are basically the same as those shown in Table 4.2. The main effect, listening mode, is statistically significant at the $P < 0.01$ level; and the interaction effect, delay by listening mode is statistically significant at the $P < 0.01$ level. Again, when the highest order interaction term (ABCD) is used as an error term, subjects are shown to be significantly different at the $P < 0.01$ level. In this instance, the degrees of freedom were reduced by 2 due to the two missing observations.

Obtained estimates across all conditions for μ are shown graphically in Figures 4.1 - 4.4 for each of the four subjects. Although there are apparent threshold differences across subjects, an obvious diotic-dichotic relationship (for both listening levels) exists:

1. Diotic listening mode thresholds are lower (more sensitive) than are dichotic listening mode thresholds for the 2 and 4 msec delay conditions;

2. A reversal in sensitivity occurs at 16 msec - in all instances, at least for one of the listening levels, the dichotic listening mode exhibits a lower threshold; inspection of the curves indicates that the reversal is due to both a marked decrease in diotic sensitivity as well as a slight increase in dichotic sensitivity.

Figure 4.5 shows the relationships described above for the

	SOURCE	ERROR TERM	SIG. LEV.	F	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARE
1	MEAN	D		86.57	7201.85	1	7201.85
2	A Overall Level	AD		8.73	8.18	1	8.18
3	B Delay	BD		1.08	31.86	3	10.62
4	C Listening Mode	CD	$P < 0.01$	44.15	355.55	1	355.55
5	D Subjects	ABCD	$P < 0.01$	35.64	249.55	3	83.18
6	AB	ABD		3.70	15.66	3	5.22
7	AC	ACD		0.36	.65	1	.65
8	BC	BCD	$P < 0.01$	10.90	293.44	3	97.81
9	AD				2.81	3	.93
10	BD				88.02	9	9.78
11	CD				24.15	3	8.05
12	ABC	ABCD		1.46	10.23	3	3.41
13	ABD				12.67	9	1.40
14	ACD				5.32	3	1.77
15	BCD				80.72	9	8.96
16	ABCD				21.00	9	2.33

TABLE 4.3: Analysis of Variance Results for Threshold Estimates* of μ.

* Threshold Values Exceeding 0 dB Were Treated as Missing Observations

Figure 4.1 Threshold estimates of μ for Subject #1.

THRESHOLD DATA - SUBJECT #1

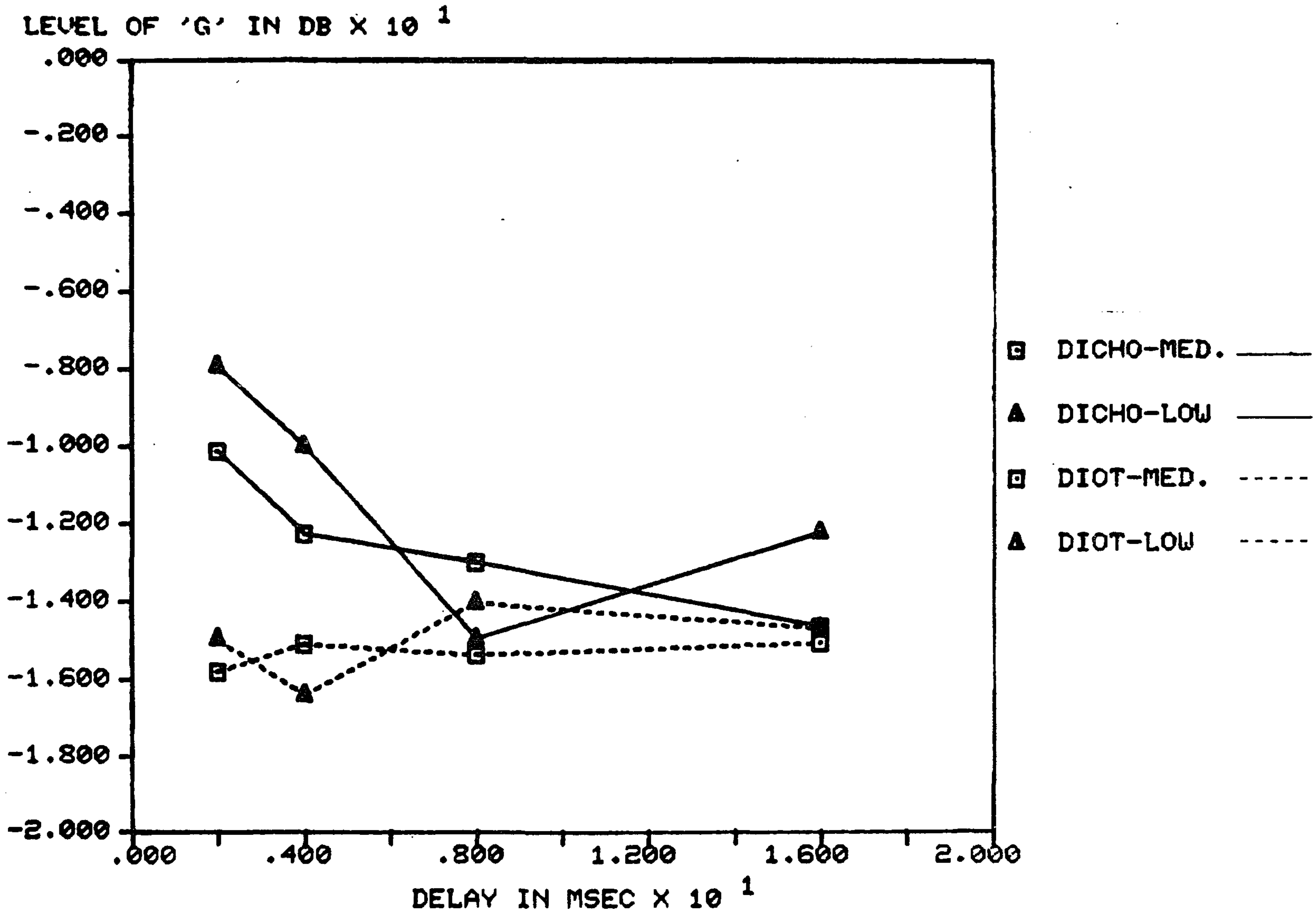


Figure 4.2 Threshold estimates of μ for Subject #2.

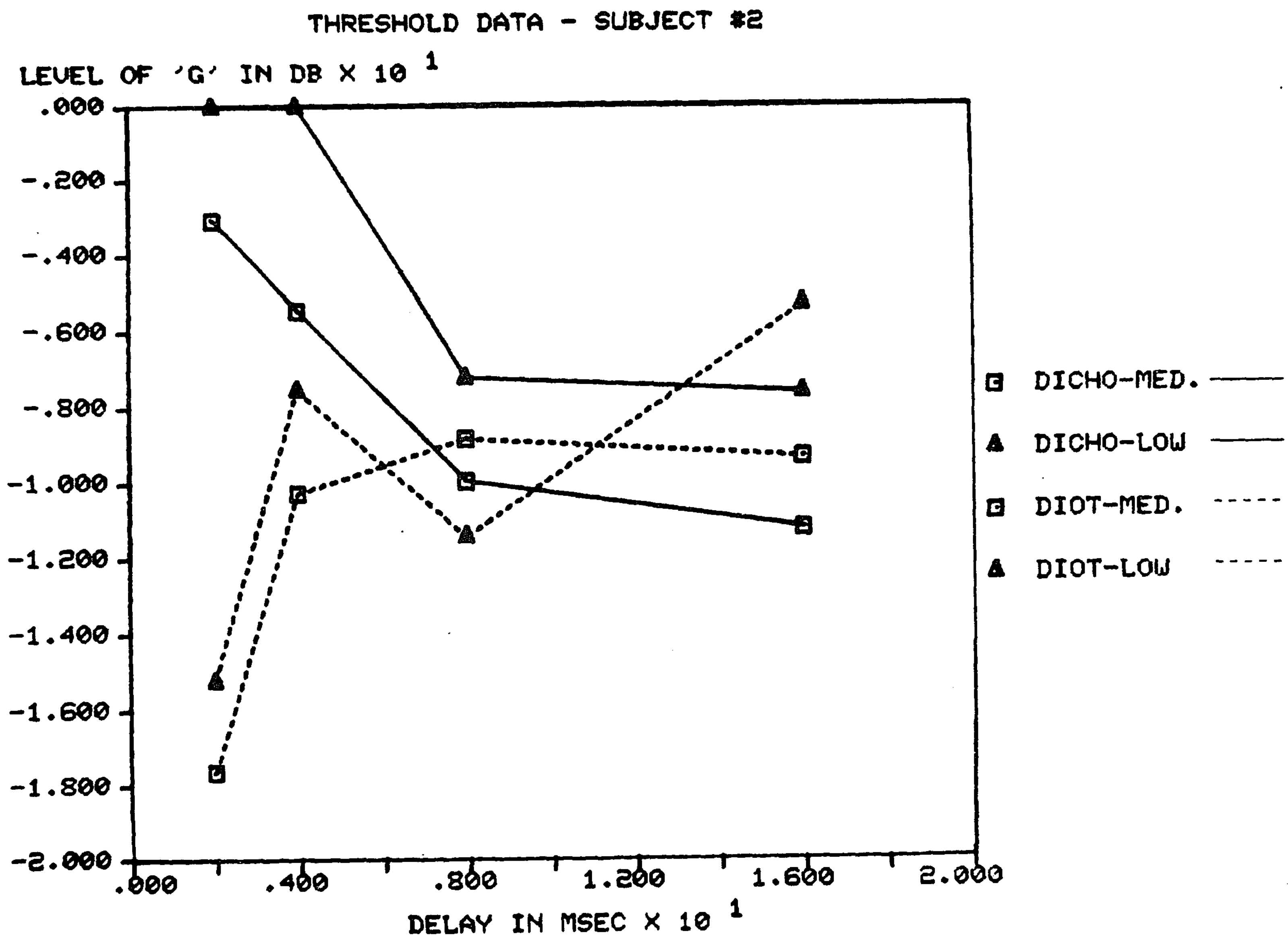
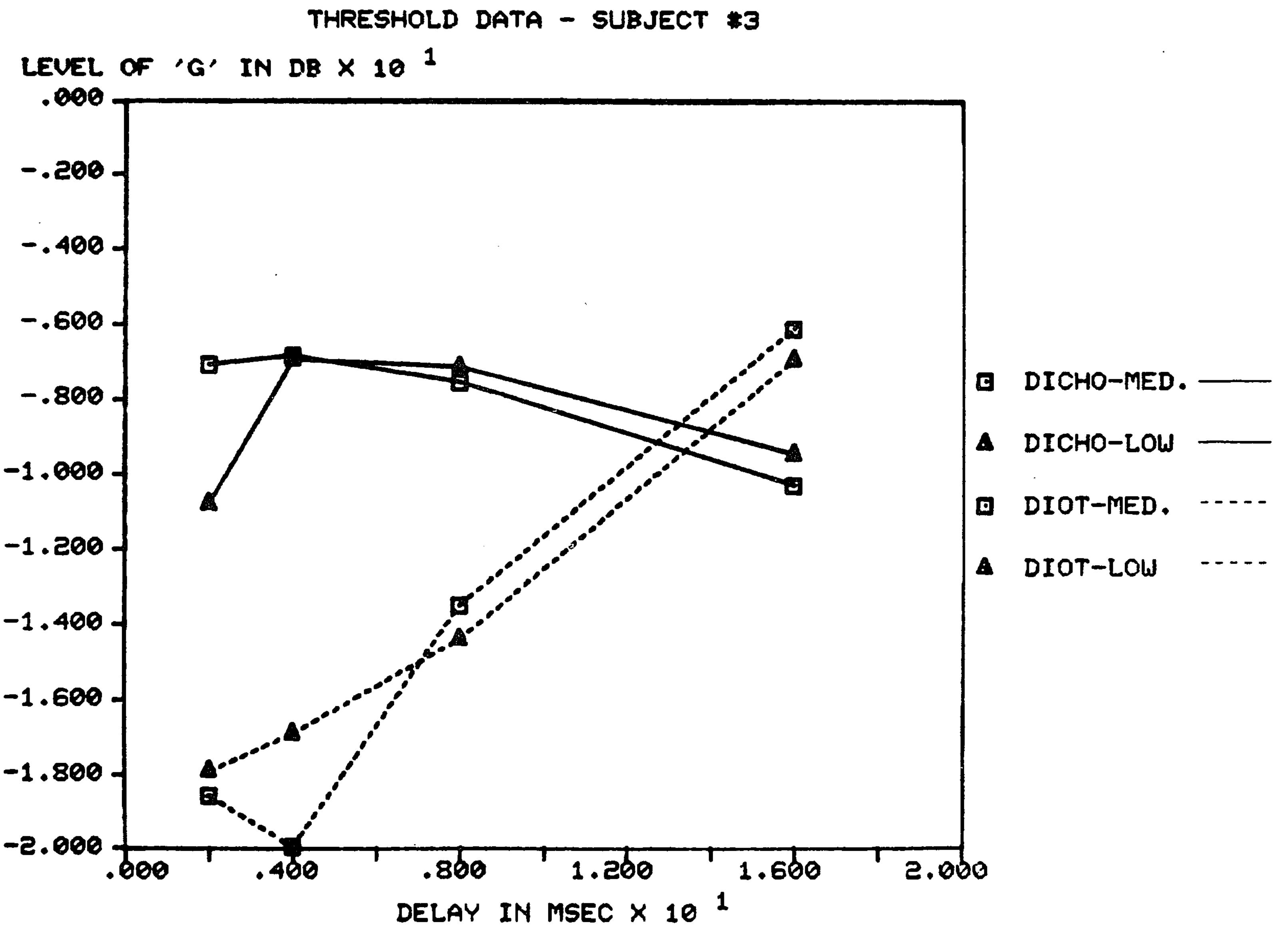

THRESHOLD DATA - SUBJECT #2
LEVEL OF 'G' IN DB X 10 1
.000
-.200
-.400
-.600
-.800
-1.000
-1.200
-1.400
-1.600
-1.800
-2.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

Figure 4.3 Threshold estimates of μ for Subject #3.



THRESHOLD DATA - SUBJECT #3
LEVEL OF 'G' IN DB X 10 1
.000
-.200
-.400
-.600
-.800
-1.000
-1.200
-1.400
-1.600
-1.800
-2.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

Figure 4.4 Threshold estimates of μ for Subject #4.

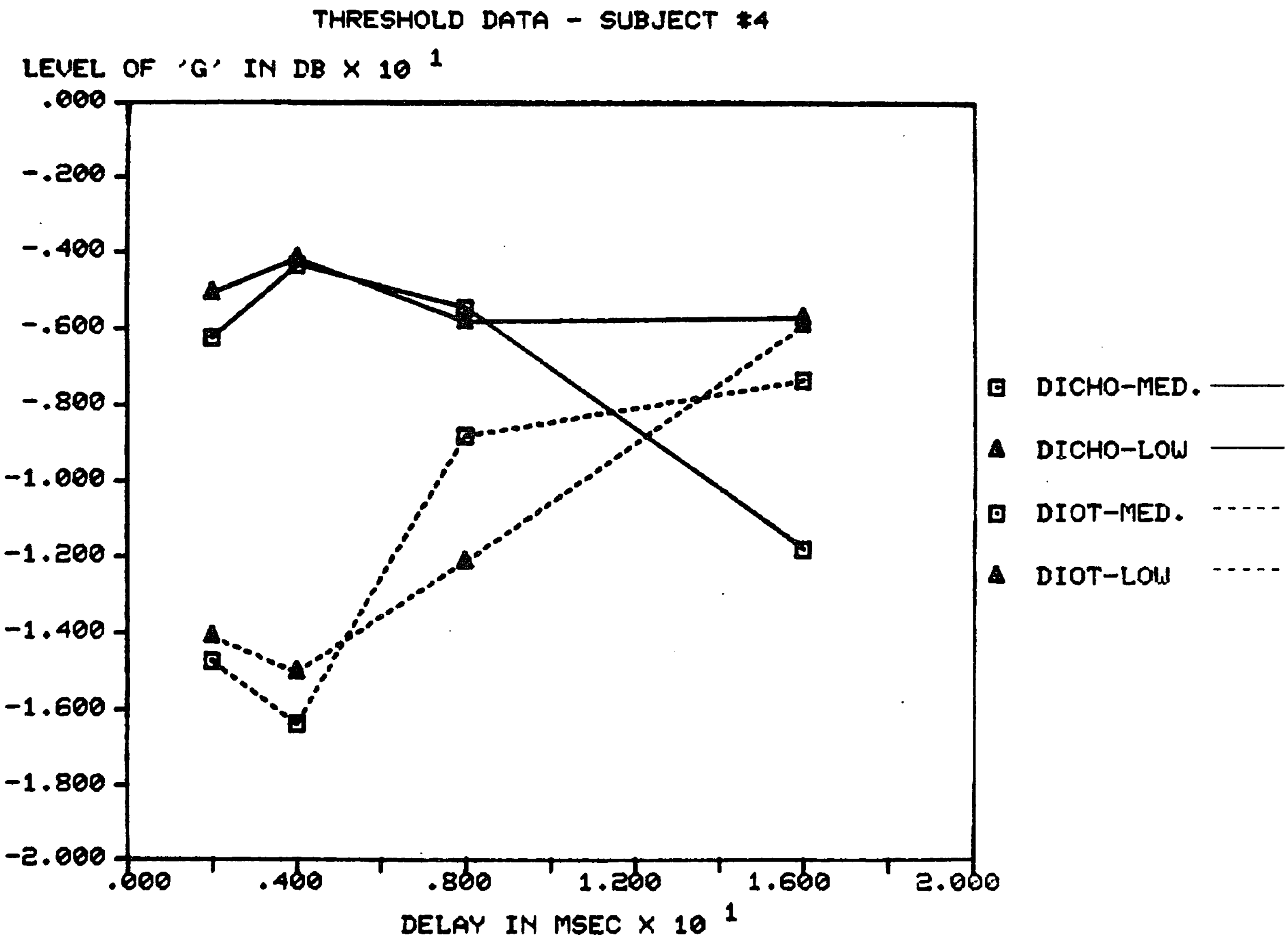
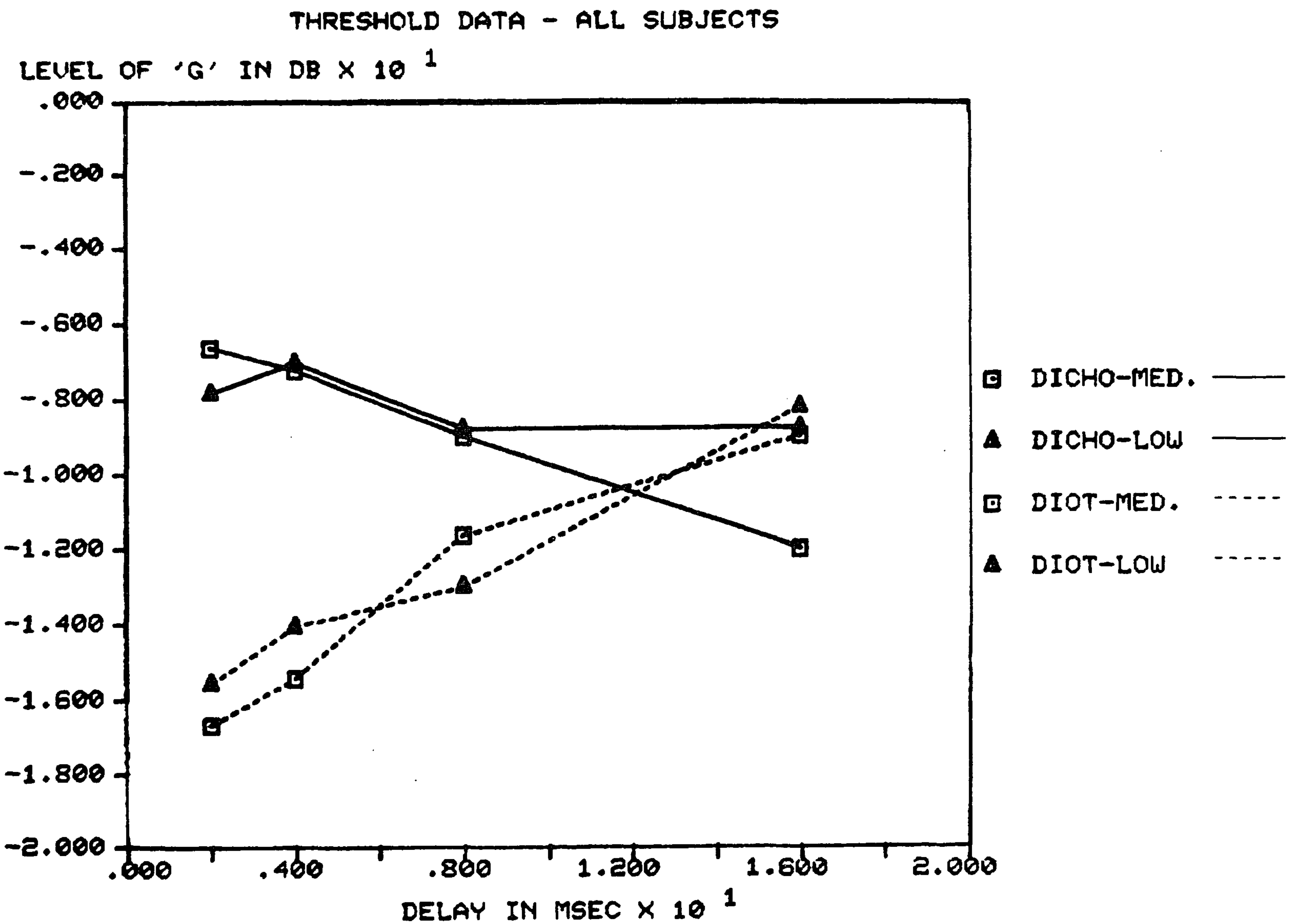

THRESHOLD DATA - SUBJECT #4
LEVEL OF 'G' IN DB X 10 1
.000
-.200
-.400
-.600
-.800
-1.000
-1.200
-1.400
-1.600
-1.800
-2.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

Figure 4.5 Threshold estimates of μ averaged across Subjects.


THRESHOLD DATA - ALL SUBJECTS
LEVEL OF 'G' IN DB X 10 1
.000
-.200
-.400
-.600
-.800
-1.000
-1.200
-1.400
-1.600
-1.800
-2.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

data pooled across subjects.

Estimates of σ (the standard deviation of the distribution) for individual subjects is given in Table 4.4. Estimates of σ associated with the 2 unobtainable thresholds (described above and shown in parenthesis) were treated as missing observations. The ANOVA results shown in Table 4.5 for the σ estimates were obtained with the same mixed model described above and show that all main effects are not statistically significant. The delay by listening mode interaction is significant at the $P < 0.01$ level. The estimated variance of the cumulative normal response curves for each condition pooled across subjects are shown in Figure 4.6.

Calculated Estimates of μ and σ for the JND (Just-noticeable Difference) Experiment

The parameter estimates reported in this section were obtained using the same procedures described above. The reference stimulus for the JND experiment was a colored noise burst consisting of a noise signal summed with itself following a fixed delay. The amplitude of the reflected component of this reference signal was fixed at $g = -9.18$ dB. The JND estimates reported then represent the level of g required for subjects to achieve a 75% correct response rate when attempting to distinguish between this reference signal and a comparison signal with an incremented value of

CONDITION	SUB NO.	DICHOTIC ($\hat{\sigma}$)	DIOTIC ($\hat{\sigma}$)	CONDITION	DICHOTIC ($\hat{\sigma}$)	DIOTIC ($\hat{\sigma}$)
2 msec MED	1.	3.76	3.70	8 msec MED	2.65	4.18
	2.	11.62	5.74		8.42	5.01
	3.	3.33	3.24		2.72	5.72
	4.	4.40	2.95		2.13	6.98
2 msec LOW	1.	11.08	3.94	8 msec LOW	3.94	1.77
	2.	(15.64)*	4.10		1.56	1.61
	3.	5.40	5.18		0.23	4.16
	4.	0.74	2.46		2.95	3.26
4 msec MED	1.	2.81	3.05	16 msec MED	1.96	5.11
	2.	6.61	2.87		3.33	8.62
	3.	1.87	3.95		6.79	16.85
	4.	6.84	13.72		7.24	8.60
4 msec LOW	1.	2.09	6.43	16 msec LOW	1.47	3.88
	2.	(35.77)*	6.96		3.46	2.71
	3.	0.21	2.89		4.54	6.19
	4.	3.07	4.52		6.23	5.89

*Conditions for which the obtained estimates of μ were greater than zero.

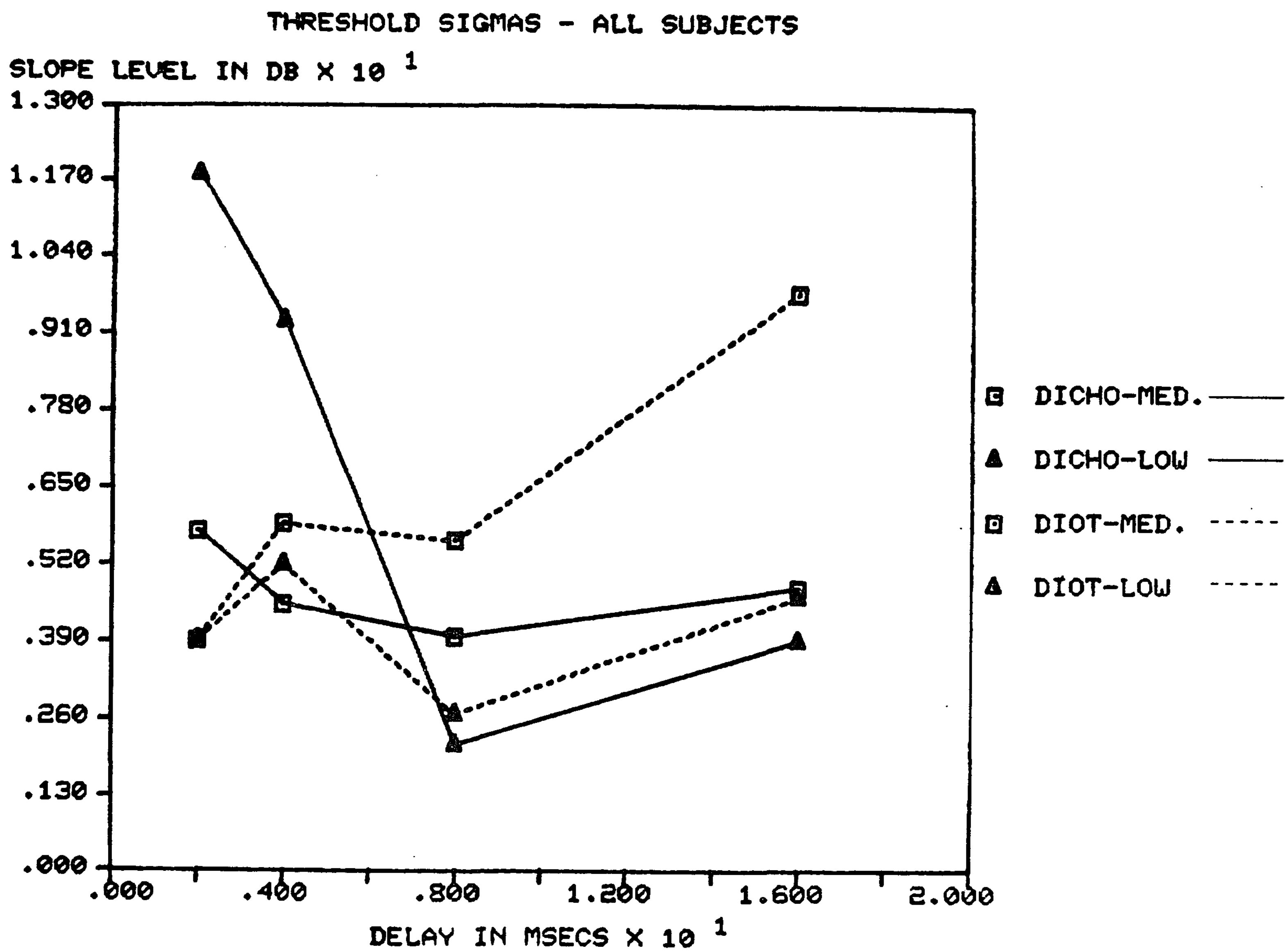
<u>TABLE</u> <u>4.4</u>: THRESHOLD ESTIMATES OF σ (in dB).

	SOURCE	ERROR TERM	SIG. LEV.	F	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARE
1	MEAN	D		262.11	1382.81	1	1382.81
2	A Overall Level	AD		3.59	49.08	1	49.08
3	B Delay	BD		0.84	41.56	3	13.85
4	C Listening Mode	CD		1.92	19.12	1	19.12
5	D Subjects				15.82	3	5.27
6	AB	ABD		1.01	19.94	3	6.64
7	AC	ACD		0.49	2.49	1	2.49
8	BC	BCD	$P < 0.01$	7.20	53.98	3	17.99
9	AD				40.94	3	13.64
10	BD				147.85	9	16.42
11	CD				29.85	3	9.95
12	ABC	ABCD		1.18	20.33	3	6.77
13	ABD				58.99	9	6.55
14	ACD				15.16	3	5.05
15	BCD				22.47	9	2.49
16	ABCD				51.68	9	5.74

TABLE 4.5: Analysis of Variance Results for Estimates* of σ
(Threshold Study)

* Conditions For Which $\hat{\mu}$ Exceeded 0 dB Were Treated As Missing Observations.

Figure 4.6 Threshold estimates of σ averaged across Subjects.


THRESHOLD SIGMAS - ALL SUBJECTS
SLOPE LEVEL IN DB X 10 1
1.300
1.170
1.040
.910
.780
.650
.520
.390
.260
.130
.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSECS X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

g. Estimate of μ for the JND experiment are shown in Table 4.6. As might be expected since the reference condition is closer to a ceiling g value of 0 dB, we find a larger number of instances where a 75% correct response rate is not obtained. Thirteen such occurrences appear in Table 4.6. Since no subject reached threshold for the 16 msec diotic-low listening mode, and only one subject reached threshold for the medium listening level, the 16 msec data was not included in the ANOVA analysis described in the next paragraph. Five unachieveable JND's remained following removal of the 16 msec delay condition. Four of these 5 unachieveable JND's however, were associated with the same subject. Since estimates of these 5 missing observations are not readily determined, a decision was made to drop the subject with 4 unobtained JND's. The final unachieveable observation was treated in a similar fashion as the threshold data reported earlier.

Two separate ANOVA's were calculated, again, using the mixed model described in the threshold experiment section. A value of 0 dB was entered as the JND estimated for the first ANOVA analysis shown in Table 4.7 while the remaining unobtained JND estimate was treated as a missing observation in the second ANOVA analysis shown in Table 4.8. Both analyses (Tables 4.7 and 4.8) show the overall listening level by delay interaction and the delay by listening mode interaction to be significant at the $P < 0.10$ and $P < 0.05$

CONDITION	SUB NO.	DICHOTIC $(\hat{\mu})$	DIOTIC $(\hat{\mu})$	CONDITION	DICHOTIC $(\hat{\mu})$	DIOTIC $(\hat{\mu})$
2 msec MED	1.	-3.59	-5.65	8 msec MED	-7.28	-5.69
	2.	-1.04	-3.23		-5.19	-1.24
	3.	-2.16	-6.10		-3.82	-1.46
	4.	-0.43	-3.62		-2.91	(2.61)$^{+}$
2 msec LOW	1.	-2.56	-5.48	8 msec LOW	-5.94	-3.80
	2.	-0.38	-1.87		-3.63	-0.97
	3.	-1.43	-3.59		-2.41	-0.73
	4.	-1.03	-1.86		-2.79	(30.00)$^{+}$
4 msec MED	1.	-5.37	-6.16	16 msec MED	-6.90	-3.16
	2.	(0.40)$^{+}$	-2.12		-1.73	(12.00)$^{+}$
	3.	-0.74	-4.85		4.86	(7.37)$^{+}$
	4.	-0.63	(4.75)		-4.90	(7.37)$^{+}$
4 msec LOW	1.	-7.61	-5.84	16 msec LOW	-7.81	(1.83)$^{+}$
	2.	-0.56	-0.61		-1.40	(8.45)$^{+}$
	3.	-0.72	-7.53		-3.01	(1.03)$^{+}$
	4.	(3.28)$^{+}$	-2.71		(0.50)$^{+}$	(10.55)$^{+}$

$^{+}$ Parentheses indicate that the obtained JND value was greater than 0.

TABLE 4.6: ESTIMATES OF μ (in dB) FOR THE JND EXPERIMENT.

	SOURCE	ERROR TERM	SIG. LEV.	F	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARE
1	MEAN	D		9.72	409.05	1	409.05
2	A Overall Level	AD	$P < 0.05$	20.92	2.79	1	2.79
3	B Delay	BD		0.15	1.41	2	.70
4	C Listening Mode	CD		0.88	4.33	1	4.33
5	D Subjects	ABCD	$P < 0.01$	34.74	84.15	2	42.07
6	AB	ABD	$P < 0.10$	5.19	6.10	2	3.05
7	AC	ACD		0.36	.12	1	.12
8	BC	BCD	$P < 0.05$	7.70	43.27	2	21.63
9	AD				.26	2	.13
10	BD				18.40	4	4.60
11	CD				9.76	2	4.88
12	ABC	ABCD		0.23	.57	2	.28
13	ABD				2.34	4	.58
14	ACD				.69	2	.34
15	BCD				11.23	4	2.80
16	ABCD				4.85	4	1.21

TABLE 4.7: Analysis of Variance Results for Estimates* of μ for the JND Experiment.

* (1) 16 msec condition not included in the analysis;
(2) Subject 4 not included;
(3) Estimate of exceeding 0 dB was fixed at 0 dB.

	SOURCE	ERROR TERM	SIG. LEV.	F	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARE
1	MEAN	D		9.99	412.42	1	412.42
2	A Overall Level	AD		15.31	3.08	1	3.08
2	B Delay	BD		0.18	1.57	2	.78
4	C Listening Mode	CD		.78	3.99	1	3.99
5	D Subjects	ABCD	$P < 0.01$	36.19	82.53	2	41.26
6	AB	ABD	$P < 0.10$	2.96	5.53	2	2.76
7	AC	ACD		.23	.73	1	.73
8	BC	BCD	$P < 0.05$	7.53	42.62	2	21.31
9	AD				.40	2	.20
10	BD				17.38	4	4.30
11	CD				10.18	2	5.09
12	ABC	ABCD		0.21	.48	2	.24
13	ABD				2.74	4	.69
14	ACD				.64	2	.32
15	BCD				11.31	4	2.82
16	ABCD				4.58	4	1.14

TABLE 4.8: Analysis of Variance Results for Estimates* of μ for the JND Experiment.

* (1) 16 msec condition not included in the analysis;
(2) Subject 4 not included;
(3) Estimate of μ exceeding 0 dB was treated as a missing observation.

levels respectively. Subjects (the random effect in this mixed model) were shown to be significant at the $P < 0.01$ level with the highest order interaction term being used as the error term. The main effect, overall level was significant at the $P < 0.05$ level for the analyses where a fixed value of 0 dB was inserted (Table 4.7).

Although not reported here, an additional ANOVA was performed using all subjects and conditions. The ceiling value of 0 was used for all 13 unobtained JND estimates. These results showed the interaction effects overall listening level by delay and delay by listening mode to be significant at the $P < 0.01$ level. All other effects with the exception of subjects were shown to be statistically non-significant.

The estimates of μ (JND) are plotted in Figures 4.7 - 4.10 for each subject. Comparison of these data with the individual μ estimates for the threshold experiment shows that the JND $\hat{\mu}$ data are not as orderly and therefore more difficult to interpret. Nevertheless, all subjects exhibit larger $\hat{\mu}$ values at 2 msec for dichotic versus diotic estimates, while the reverse is true at 16 msec. These observations are clearly demonstrated in the pooled data (across subjects) shown in Figure 4.11.

Individual subject $\hat{\sigma}$ are shown in Table 4.9. For reasons described above, all 16 msec σ estimates were not

Figure 4.7 Estimates of μ for the JND experiment for Subject #1.

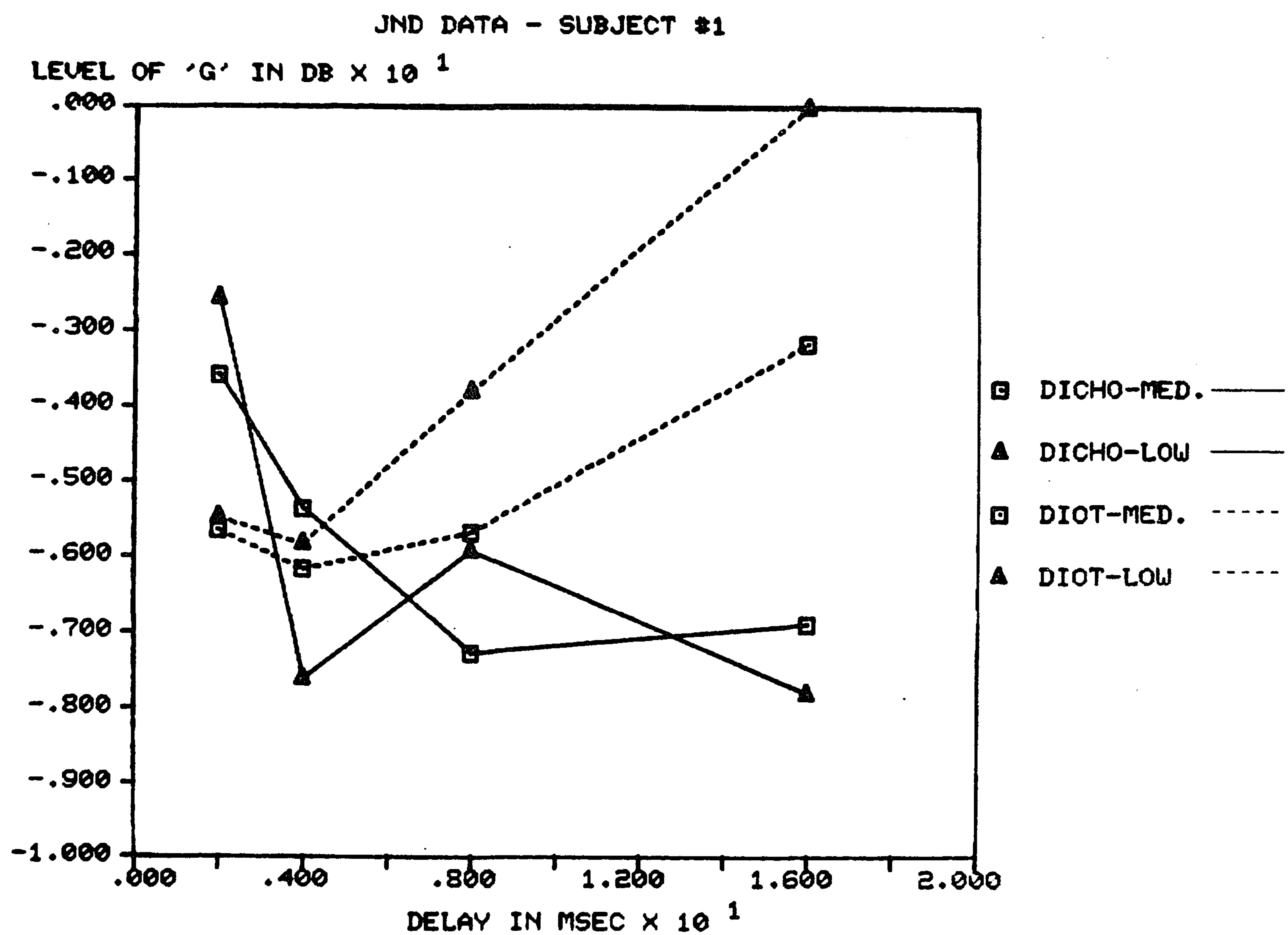

JND DATA - SUBJECT #1
LEVEL OF 'G' IN DB X 10 1
.000
-.100
-.200
-.300
-.400
-.500
-.600
-.700
-.800
-.900
-1.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

Figure 4.8 Estimates of μ for the JND experiment for Subject #2.

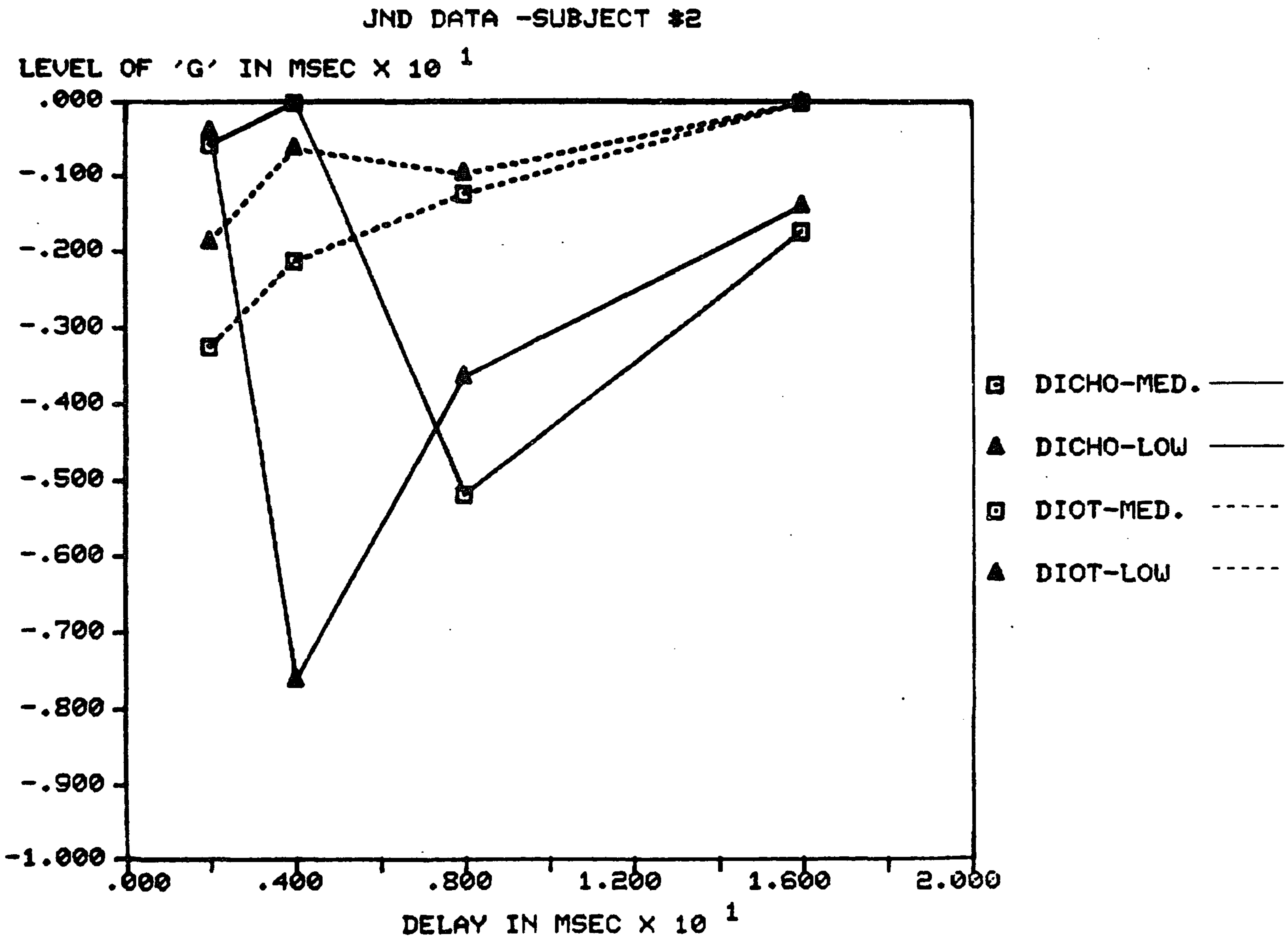

JND DATA -SUBJECT #2
LEVEL OF 'G' IN MSEC X 10 1
.000
-.100
-.200
-.300
-.400
-.500
-.600
-.700
-.800
-.900
-1.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

Figure 4.9 Estimates of μ for the JND experiment for Subject #3.

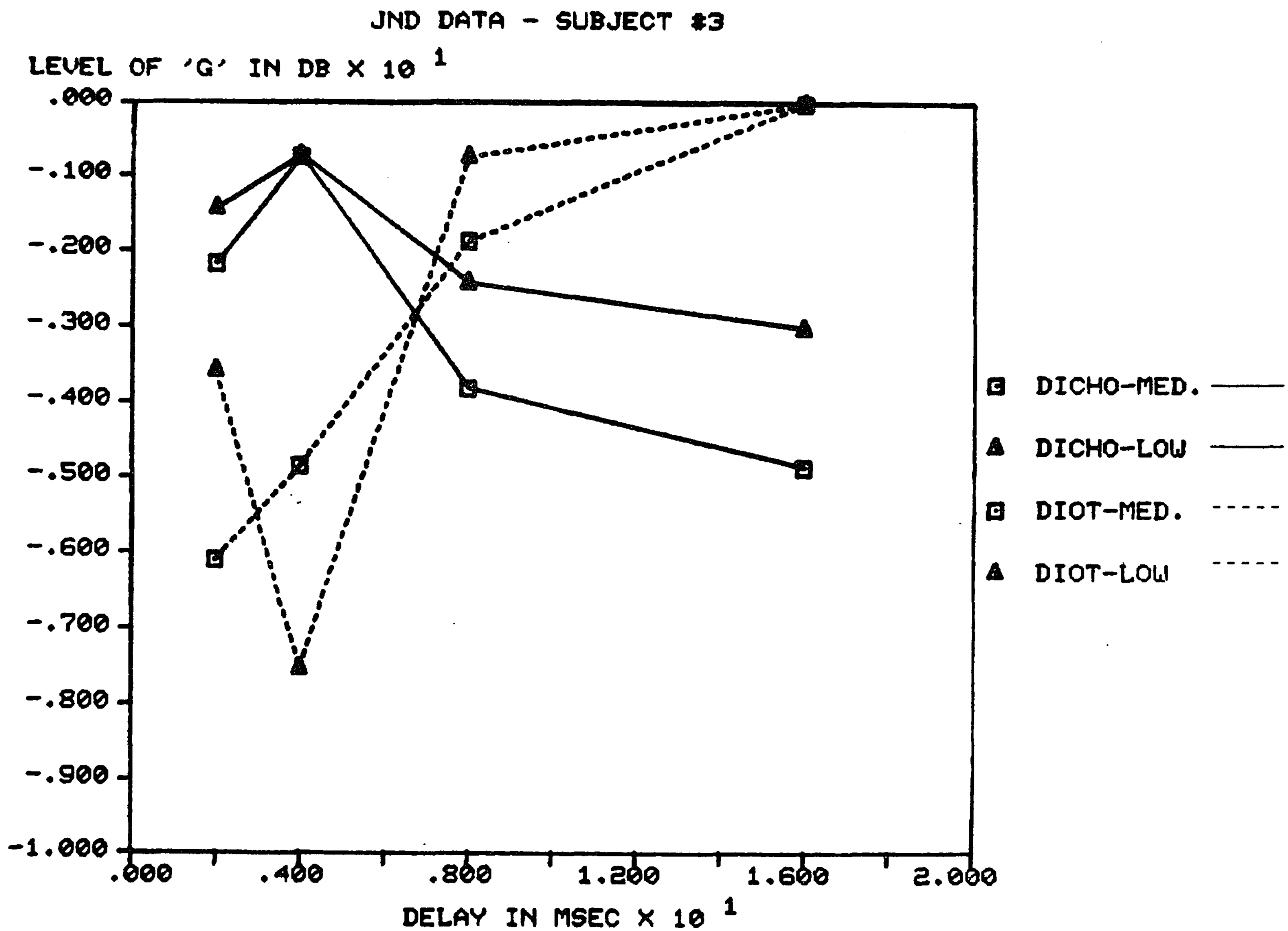

JND DATA - SUBJECT #3
LEVEL OF 'G' IN DB X 10 1
.000
-.100
-.200
-.300
-.400
-.500
-.600
-.700
-.800
-.900
-1.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

Figure 4.1Ø Estimates of μ for the JND experiment for Subject #4.

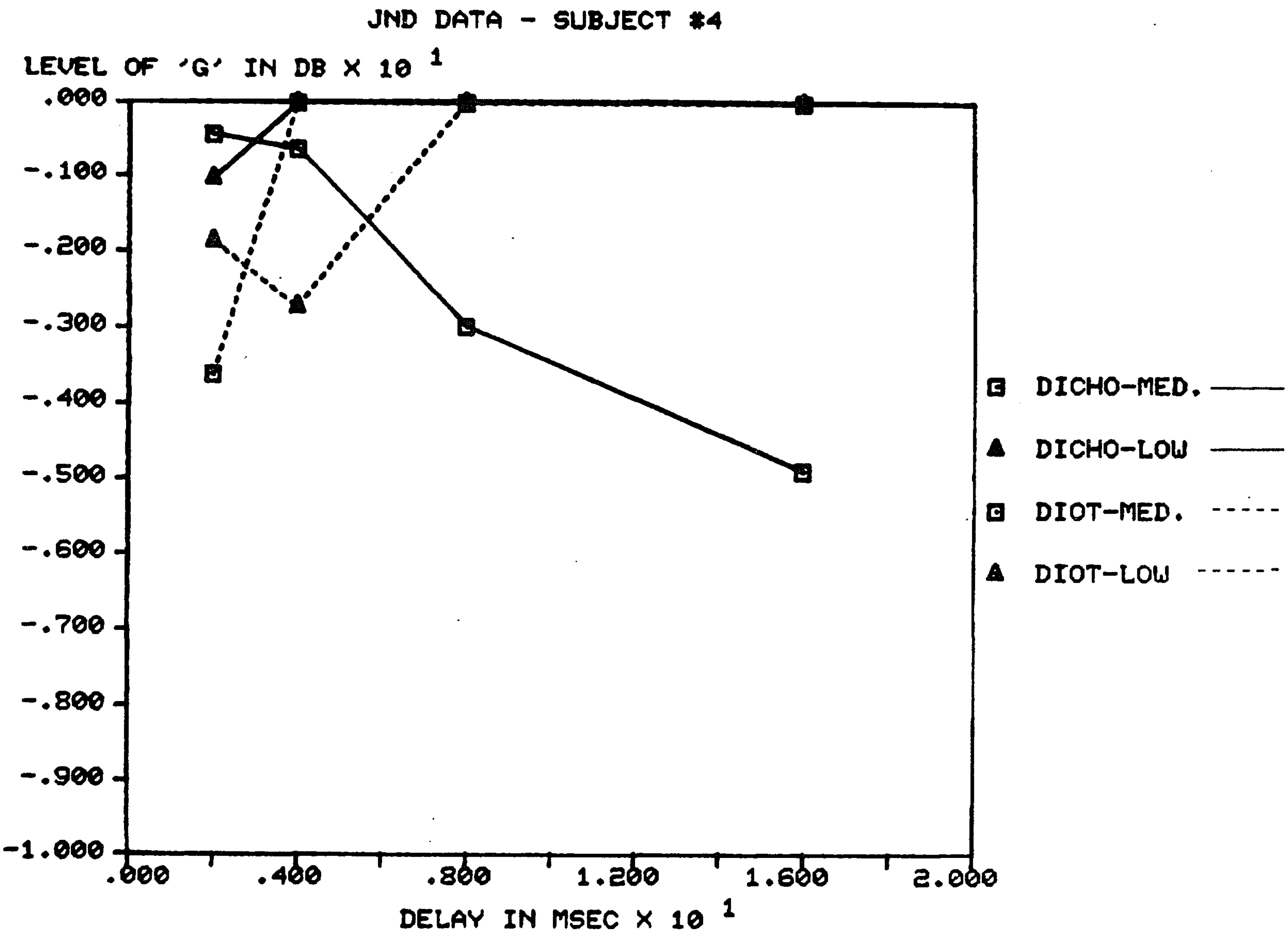
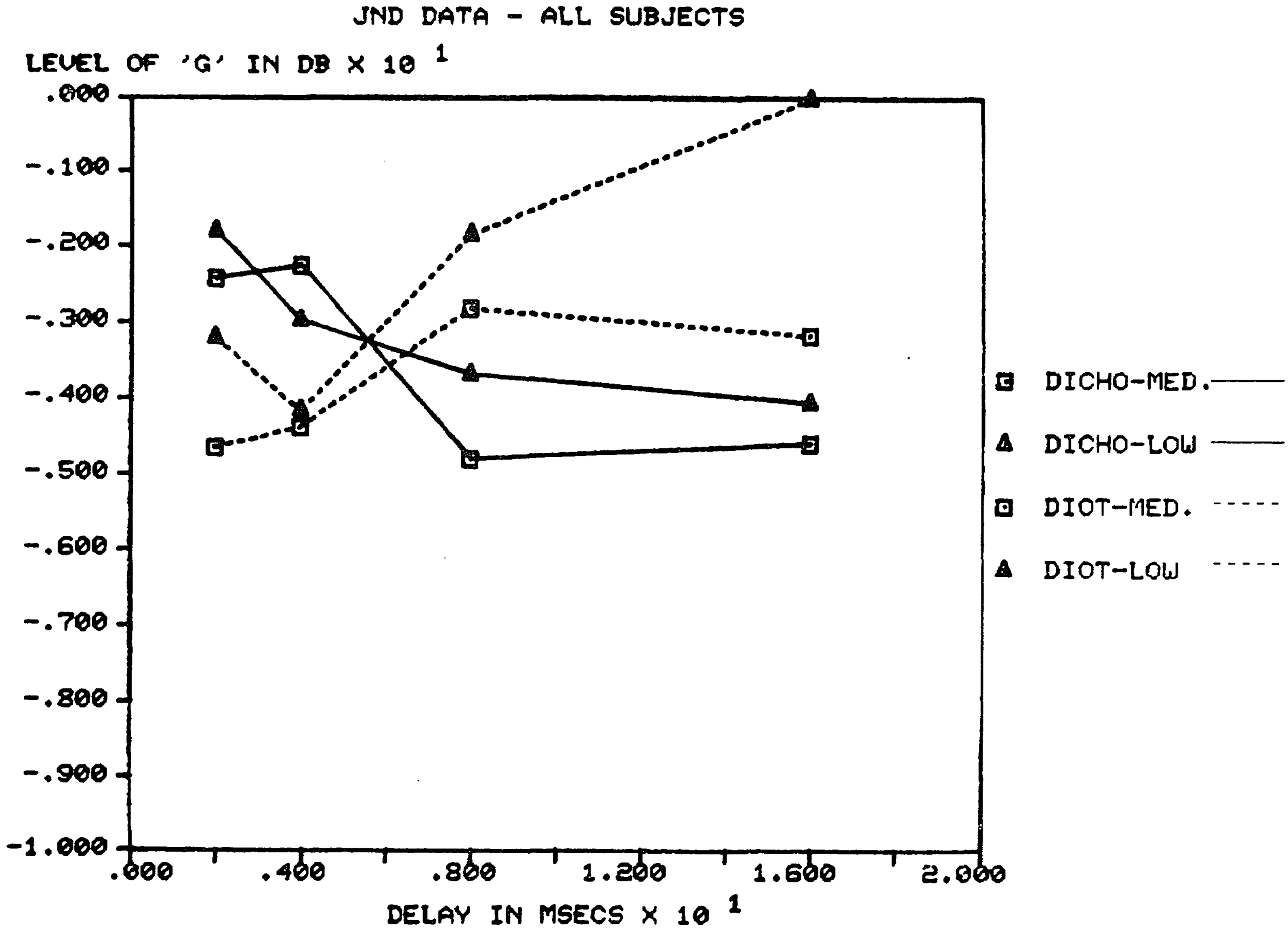

JND DATA - SUBJECT #4
LEVEL OF 'G' IN DB X 10 1
.000
-.100
-.200
-.300
-.400
-.500
-.600
-.700
-.800
-.900
-1.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

Figure 4.11 Estimates of μ for the JND experiment averaged across Subjects.


JND DATA - ALL SUBJECTS
LEVEL OF 'G' IN DB X 10 1
.000
-.100
-.200
-.300
-.400
-.500
-.600
-.700
-.800
-.900
-1.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSECS X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

CONDITION	SUB NO.	DICHOTIC ($\hat{\sigma}$)	DIOTIC ($\hat{\sigma}$)	CONDITION	DICHOTIC ($\hat{\sigma}$)	DIOTIC ($\hat{\sigma}$)
2 msec MED	1.	2.32	2.07	8 msec MED	1.77	3.53
	2.	0.01	1.33		4.65	4.31
	3.	1.81	2.35		3.05	1.44
	4.	1.70	2.74		5.41	(0.57)*
2 msec LOW	1.	3.29	3.00	8 msec LOW	1.90	2.77
	2.	1.66	7.19		3.66	3.36
	3.	3.79	1.25		1.17	3.53
	4.	0.97	4.87		1.37	(2.00)*
4 msec MED	1.	1.83	5.62	16 msec MED	0.07	2.25
	2.	(4.27)*	0.09		3.83	(16.55)*
	3.	4.22	5.58		2.49	(0.55)*
	4.	2.47	9.60		2.95	(9.96)*
4 msec LOW	1.	5.73	2.65	16 msec LOW	3.83	(2.14)*
	2.	1.67	7.81		2.94	(5.62)*
	3.	1.29	6.14		3.18	(2.09)*
	4.	(10.73)*	4.18		(6.24)*	(8.92)*

* Conditions for which the obtained estimates of μ were greater than 0.

<u>TABLE 4.9</u>: ESTIMATES OF σ (in DB) FOR THE JND EXPERIMENT.

included in the analysis. The five cases for which thresholds were not obtained were treated in a similar fashion to the estimates of μ. That is, all data from the subject with 4 unobtained JND's was dropped. Since it is not clear how to set a ceiling value for these data, the ANOVA analyses shown in Table 4.10 only was performed with the one remaining unobtained JND value being treated as a missing observation. When these data were submitted to an ANOVA analysis, main and interaction effects are shown not to be statistically significant. Estimates of σ pooled across all subjects are plotted in Figure 4.12.

Goodness of Fit - χ^2 Analysis

Several methods are available both parametric (eg. Chi Square) and nonparametric (eg. Kolmogoroff-Smirnoff) for testing the validity of a model (Bury, 1975). All of these tests are directed at the question of whether the observed distribution differs significantly from the assumed underlying distribution. Finney (1952) recommends that the Chi Square (χ^2) test be used for determining the heterogeneity of the discrepancies between the observed and expected values obtained with the maximum likelihood solution. The relationship of the observed and expected values obtained in this study was therefore examined with the χ^2 test. The Pearson chi-square statistic is defined as:

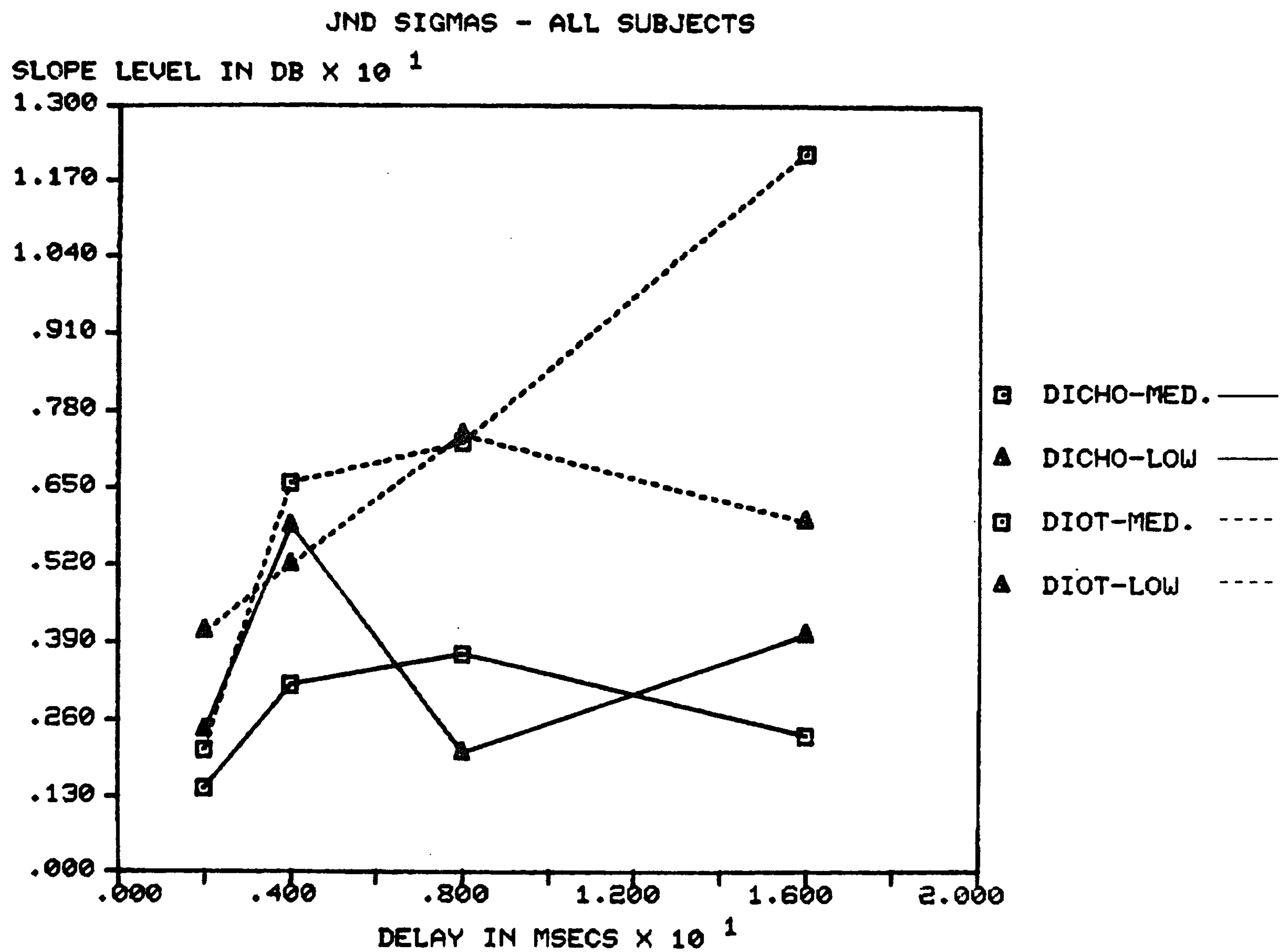
$$\chi^2 = \Sigma \frac{(\text{observable frequencies - expected frequencies})^2}{\text{expected frequencies}}$$

	SOURCE	ERROR TERM	SIG. LEV.	F	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARE
1	MEAN	D		508.96	327.18	1	327.18
2	A Overall Level	AD		0.59	3.04	1	3.04
3	B Delay	BD		1.84	11.59	2	5.79
4	C Listening Mode	CD		4.88	6.07	1	6.07
5	D Subjects	ABCD		0.11	1.28	2	.64
6	AB	ABD		3.16	9.63	2	4.81
7	AC	ACD		.17	1.59	1	1.59
8	BC	BCD		0.53	3.98	2	1.99
9	AD				10.28	2	5.14
10	BD				12.60	4	3.15
11	CD				2.48	2	1.24
12	ABC	ABCD		0.11	1.28	4	.64
13	ABD				6.09	4	1.52
14	ACD				18.67	2	9.33
15	BCD				15.02	4	3.75
16	ABCD				21.65	4	5.41

TABLE 4.10: Analysis of Variance Results for Estimates* of σ for the JND Experiment.

* (1) 16 msec not included in the analysis;
(2) Subject #4 not included;
(3) σ value for conditions where $\hat{\mu}$ exceeded 0 dB was treated as a missing observation.

Figure 4.12 Estimates of σ for the JND experiment averaged across Subjects.


JND SIGMAS - ALL SUBJECTS
SLOPE LEVEL IN DB X 10 1
1.300
1.170
1.040
.910
.780
.650
.520
.390
.260
.130
.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSECS X 10 1
DICHO-MED.
DICHO-LOW
DIOT-MED.
DIOT-LOW

Each set of data on a subject by condition basis was examined with this statistic and the results are reported for the Threshold and JND experiments in Tables 4.11 and 4.12, respectively.

Also shown in Table 4.11 and 4.12 are collapsed x^2 calculations with associated significance levels where the first 3 levels and last 3 levels of each distribution were collapsed into two categories. The degrees of freedom have been appropriately reduced for these cases. The rationale for collapsing the upper and lower tails (extremes) of the distribution is: when nP or nQ, the expected values, for any class (level in these experiments) is less than 5, the distribution may give misleading results (Finney, 1964, pg. 56). It is interesting to note, that in many cases (certainly not all cases), the collapsed probability levels change significantly obviously in favor of retaining the underlying model.

Inspection of Tables 4.11 and 4.12 indicates that at least for some individual cases (seven cases reach the $P \leq 0.05$ significance level) the cumulative normal model may not fully account for the data. This however, does not suggest abandonment of the model since no obvious systematic departure of the observed data from the expected values is seen. Furthermore, due to the large number of tests being performed, a small number of these tests would be expected, by chance alone, to show a significant departure from the

COND.	NO.	DICHOTIC x^2	d_F	SIG. LEVEL	COLLAPSED x^2	d_F	SIG. LEVEL	DIOTIC x^2	d_F	SIG. LEVEL	COLLAPSED x^2	d_F	SIG. LEVEL
2 msec MED	1.	15.96	12	.25>P>.10	11.34	8	.25>P>.10	9.78	14	.90>P>.75	7.77	10	.75>P>.50
	2.	17.64	13	.25>P>.10	8.97	9	.50>P>.25	22.93	14	.10>P>.05	7.17	10	.75>P>.50
	3.	9.21	13	.90>P>.75	8.81	9	.50>P>.25	11.07	14	.75>P>.50	10.54	10	.50>P>.25
	4.	8.17	13	.90>P>.75	5.91	9	.75>P>.50	21.09	14	.10>P>.05	18.73	10	.05>P>.025
2 msec LOW	1.	15.79	11	.25>P>.10	5.90	7	.75>P>.50	16.36	10	.10>P>.05	13.62	6	.05>P>.025
	2.	11.57	11	.50>P>.25	8.49	7	.50>P>.25	5.39	10	.40>P>.75	2.27	6	.90>P>.75
	3.	13.59	10	.25>P>.10	5.79	6	.50>P>.25	2.23	10	.995>P>.99	1.99	6	.95>P>.90
	4.	6.39	11	.90>P>.75	4.38	7	.75>P>.50	2.46	10	.995>P>.99	2.30	6	.90>P>.75
4 msec MED	1.	9.87	11	.75>p>.50	7.01	7	.5>P>.25	3.62	12	.99>P>.975	3.09	8	.95>P>.90
	2.	12.35	11	.50>P>.25	8.08	7	.5>P>.25	5.66	12	.95>P>.90	5.66	8	.75>P>.50
	3.	8.37	11	.75>P>.50	7.48	7	.5>P>.25	5.03	12	.975>P>.95	4.34	8	.90>P>.75
	4.	9.23	11	.75>P>.50	6.87	7	.5>P>.25	12.61	12	.50>P>.25	8.38	8	.50>P>.25
4 msec LOW	1.	11.92	11	.50>P>.25	11.57	7	.125>P>.10	10.39	10	.50>P>.25	7.16	6	.50>P>.25
	2.	4.46	11	.975>P>.95	3.10	7	.90>P>.75	3.34	10	.975>P>.95	1.33	6	.975>P>.95
	3.	5.00	11	.95>P>.90	4.00	7	.90>P>.75	1.54	10	P=.995	1.45	6	.975>P>.95
	4.	0.975	11	P=.995	0.714	7	P=.995	8.33	10	.75>P>.50	3.97	6	.75>P>.50
8 msec MED	1.	4.36	12	.99>P>.975	3.97	8	.90>P>.75	11.19	16	.90>P>.75	10.75	12	.75>P>.50
	2.	27.77	12	.010>P>.005	10.61	8	.25>P>.10	11.28	16	.90>P>.75	8.64	12	.75>P>.50
	3.	17.72	12	.25>P>.10	10.39	8	.25>P>.10	16.55	16	.50>P>.25	13.56	12	.50>P>.25
	4.	9.31	12	.75>P>.50	5.66	8	.75>P>.50	9.73	16	.90>P>.75	7.78	12	.75>P>.50
8 msec LOW	1.	8.75	12	.75>P>.50	8.74	8	.50>P>.25	14.99	15	.50>P>.25	10.29	11	.75>p>.50
	2.	12.33	12	.50>p>.25	11.12	8	.25>P>.10	9.21	15	.90>P>.75	8.46	11	.75>P>.50
	3	10.95	12	.75>P>.5	10.15	8	.25>P>.10	15.97	15	.50>P>.25	14.79	11	.25>P>.10
	4.	7.63	12	.90>P>.75	3.66	8	.90>P>.75	14.36	15	.50>P>.25	12.97	11	.50>P>.25
16 msec MED	1.	_	10		_	6		6.93	10	.75>P>.5	4.58	6	.75>P>.5
	2.	9.19	10	.75>P>.50	8.10	6	.25>P>.10	5.65	10	.90>P>.75	1.36	6	.975>P>.95
	3.	8.17	10	.75>P>.50	8.06	6	.25>P>.10	2.94	10	.975>P>.950	1.32	6	.975>P>.95
	4.	6.35	10	.90>P>.75	3.80	6	.75>P>.50	7.05	10	.75>P>.5	6.57	6	.50>P>.25
16 msec LOW	1.	5.21	10	.90>P>.75	3.97	6	.75>P>.5	8.71	10	.75>P>.50	5.72	6	.50>P>.25
	2.	18.38	10	.05>P>.025	8.47	6	.25>P>.10	5.82	10	.90>P>.75	4.00	6	.75>P>.50
	3.	9.03	10	.75>P>.5	4.32	6	.75>P>.50	6.64	10	.90>P>.75	5.99	6	.50>P>.25
	4.	15.77	10	.25>P>.10	2.46	6	.90>P>.75	5.15	10	.90>P>.75	2.30	6	.90>P>.75

Table 4.11 - χ^2 values and significance levels for Threshold Data

COND.	NO.	DICHOTIC x^2	d_F	SIG. LEVEL	COLLAPSED x^2	d_F	SIG. LEVEL	DIOTIC x^2	d_F	SIG. LEVEL	COLLAPSED x^2	d_F	SIG. LEVEL
2 msec MED	1.	16.97	12	.25>P>.10	11.93	8	.75>P>.10	8.31	13	.90>P>.75	7.64	9	.75>P>.50
	2	9.0	12	.75>P>.50	5.12	8	.75>P>.50	17.04	13	.25>P>.10	15.83	9	.10>P>.05
	3.	27.88	12	.01>P>.005	13.88	8	.10>P>.05	11.87	13	.75>P>.5	10.92	9	.50>P>.25
	4.	5.05	12	.975>P>.95	3.43	8	.95>P>.90	8.00	13	.90>P>.75	6.83	9	.75>P>.50
2 msec LOW	1.	15.69	10	.25>P>.10	8.99	6	.25>P>.10	17.21	10	.10>P>.05	9.81	6	.25>P>.10
	2.	8.33	10	.75>P>.50	5.76	6	.50>P>.25	7.83	10	.75>P>.5	2.15	6	.95>P>.90
	3.	7.59	10	.75>P>.50	7.54	6	.50>P>.25	8.49	10	.75>P>.5	8.08	6	.25>P>.10
	4.	8.79	10	.75>P>.5	4.54	6	.75>P>.50	4.72	10	.95>P>.90	3.80	6	.75>P>.5
4 msec MED	1.	5.27	11	.95>P>.90	3.70	7	.90>P>.75	7.68	11	.75>P>.5	6.99	7	.5>P>.25
	2.	10.18	11	.75>P>.5	5.37	7	.75>P>.5	6.59	11	.9>P>.75	3.11	7	.90>P>.75
	3.	9.63	11	.75>P>.5	6.94	7	.50>P>.25	7.23	11	.9>P>.75	4.17	7	.90>P>.75
	4.	5.65	11	.90>P>.75	4.53	7	.75>P>.5	8.59	11	.75>P>.5	2.47	7	.95>P>.90
4 msec LOW	1.	7.31	10	.75>P>.5	5.33	6	.75>P>.5	11.96	10	.50>P>.25	6.01	6	.5>P>.25
	2.	7.31	10	.75>P>.5	5.33	6	.75>P>.5	7.83	10	.75>P>.5	6.51	6	.5>P>.25
	3.	10.17	10	.50>P>.25	5.72	6	.50>P>.25	5.81	10	.90>P>.70	3.85	6	.75>P>.5
	4.	9.59	10	.75>P>.50	6.42	6	.50>P>.25	5.07	10	.90>P>.70	4.18	6	.75>P>.5
8 msec MED	1.	32.77	13	.000>P>.001	31.13	9	P=.00	26.15	14	.05>P>.025	16.26	10	.10>P>.05
	2.	14.61	13	.50>P>.25	12.34	9	.25>P>.10	9.75	14	.90>P>.75	6.07	10	.90>P>.75
	3.	5.66	12	.95>P>.90	2.54	8	.975>P>.95	19.80	13	.25>P>.10	16.35	9	.10>P>.05
	4.	14.06	13	.50>P>.25	12.64	9	.25>P>.10	11.50	14	.75>P>.50	6.50	10	.90>P>.75
8 msec LOW	1.	6.96	12	.90>P>.75	5.05	8	.90>P>.75	7.84	13	.90>P>.75	5.07	9	.90>P>.75
	2.	17.4	12	.25>P>.10	9.49	8	.50>P>.25	16.91	13	.25>P>.10	7.03	9	.75>P>.50
	3.	11.50	12	.50>P>.25	10.97	8	.25>P>.10	17.30	13	.25>P>.10	14.51	9	.25>P>.10
	4.	6.87	12	.90>P>.75	6.19	8	.75>P>.5	10.0	13	.75>P>.5	7.75	9	.75>P>.5
16 msec MED	1.	2.50	10	.995>P>.99	0.167	6	P=.995	7.53	10	.75>P>.50	4.71	6	.75>P>.50
	2.	3.75	10	.975>P>.95	2.25	6	.90>P>.75	9.81	10	.5>P>.25	7.42	6	.50>P>.25
	3.	2.81	10	.99>P>.975	1.71	6	.95>P>.90	4.0	10	.95>P>.9	2.3	6	.75>P>.75
	4.	5.82	10	.90>P>.750	3.37	6	.90>P>.75	7.29	10	.75>P>.5	3.47	6	.75>P>.50
16 msec LOW	1.	16.79	10	.10>P>.05	5.85	6	.5>P>.25	3.72	10	.975>P>.95	3.02	6	.90>P>.75
	2.	7.82	10	.75>P>.5	5.28	6	.75>P>.5	6.89	10	.75>P>.5	4.44	6	.75>P>.5
	3.	12.44	10	.50>P>.25	7.81	6	.5>P>.25	11.89	10	.50>P>.25	9.13	6	.25>P>.1
	4.	16.07	10	.10>P>.05	7.90	6	.25>P>.10	4.35	10	.95>P>.90	2.09	6	.95>P>.9

Table 4.12 - χ^2 values and significance levels for JND Data.

assumed underlying cumulative normal distribution.

CHAPTER V

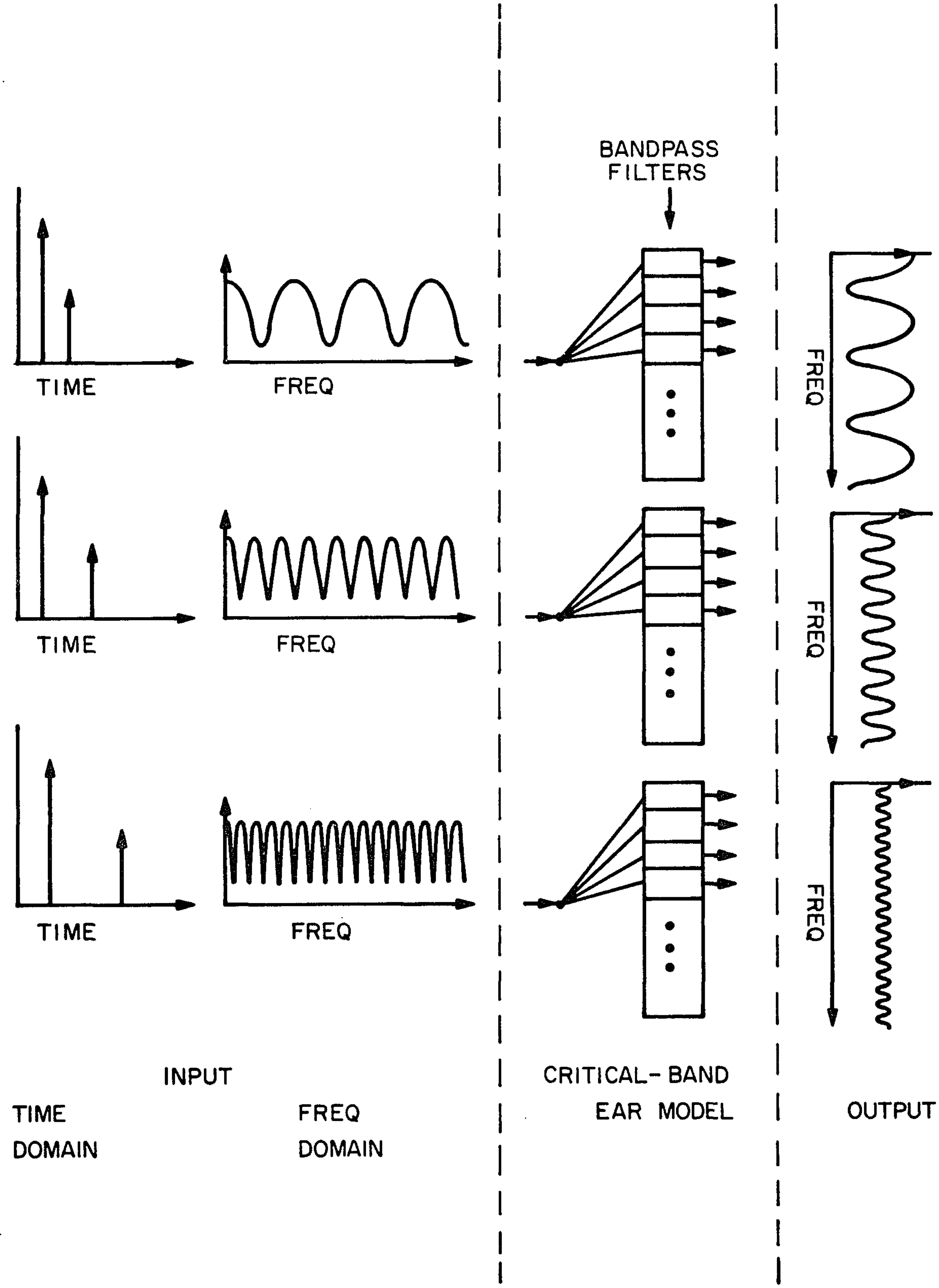
DISCUSSION

Threshold Data - Dichotic vs. Diotic Listening Mode

The pooled diotic threshold data (Figure 4.5) indicate that sensitivity for the delayed signal decreases with increasing delay for the diotic listening mode. Furthermore, this decreasing sensitivity is seen to be independent of overall listening level. The dichotic threshold results, also shown in Figure 4.5, demonstrate an increasing sensitivity to the delayed signal as a function of increasing delay. More importantly, a large difference in threshold sensitivity was shown for the diotic vs. dichotic listening modes. For the smaller delay times of 2, 4 and 8 msec, the pooled diotic thresholds are seen to be significantly lower (more sensitive) than are the pooled dichotic thresholds. While the Results Section showed that individual subject performance for the two listening modes exhibited significant differences, the relative difference in thresholds for the diotic vs. dichotic listening modes remained essentially unchanged at 2 and 4 msec for all subjects. The salient features of these data can be summarized as follows: (1) diotic thresholds increase with increasing delay, (2) dichotic thresholds decrease slightly with increasing delay, (3) significant threshold differences exist between diotic and dichotic listening modes, and (4) significant differences exist across subjects in terms of obtained thresholds as well as in terms of relative diotic vs. dichotic differences in listening mode.

The fact that threshold sensitivity decreases as a function of delay as observed in the diotic listening mode, has been ascribed to the deteriorating resolution of peripheral (critical-band) filters (Bilsen and Goldstein, 1974). Recall that the spectrum of a noise source added to itself with an intervening delay produces a comb-filter spectrum. As the delay is increased the frequency spacing between peaks decreases. By using the simple expression $1/\tau$, where τ represents the delay (in seconds) to define this spacing, it is readily shown that a 2 msec delay produces a comb-filter spectrum with spectral peaks occurring every 500 Hz, while a 16 msec delay produces a comb-filter spectrum with spectral peaks occurring every 62.5 Hz. The closely spaced peaks produced by the 16 msec delay are smoothed within each critical band; the more widely spaced peaks produced by the 2 msec delay extend across critical-band boundaries and do not undergo the same degree of smoothing. A block diagram of a simplified ear model designed to demonstrate this effect is shown in Figure 5.1. In each case, the input filter characteristics (for both the time and frequency

Figure 5.1 Simplified ear model demonstrating critical-band effect.


BANDPASS FILTERS
TIME
FREQ
TIME
FREQ
TIME
FREQ
FREQ
FREQ
FREQ
INPUT
TIME DOMAIN
FREQ DOMAIN
CRITICAL-BAND EAR MODEL
OUTPUT

domain) through which the noise source is processed are shown. The noise source then passes through the bank of critical-band band pass filters and the resulting spectral response is shown on the right hand column of the figure. In the frequency region of interest, the critical-band bandwidths are relatively constant. Smoothing resulting from critical-band filtering is evident at the spectral output of the critical-band ear model. The upper panel demonstrates faithful reproduction of the input signal in the frequency domain; the center panel, using input stimuli with additional delay shows a modest degree of smoothing; the lower panel with further delay demonstrates extreme smoothing. It is assumed that detection takes place when the depth of the spectral variations exceeds a threshold value. In summary, it seems reasonable to invoke the well-established critical-band model as an explanation for the decrease in sensitivity as a function of delay observed for the diotic listening mode.

In contrast, the dichotic thresholds (see Figure 4.5) decrease markedly in sensitivity for a delay of 2 msec. A reversal in slope relative to the diotic data is then seen for the longer delays, up to 8 msec, with no further change at 16 msec. The marked overall shift in sensitivity as well as the slope reversal as a function of increasing delay must both be attributed to some form(s) of binaural interaction(s). Introspective comments offered by the subjects as well as by other observers (including the author), experienced in psychoacoustic testing, indicated (a) that the diotic stimuli produce a fairly compact median plane image, with little or no diffusion; (b) that the dichotic stimuli produce a more diffuse image; and (c) that the image appears to become more diffuse both for increasing delays as well as for increasing levels of the delayed signal. These observations suggest that subjects' decisions are based on at least two distinct cues: (1) a pitch-like effect, presumably produced by the envelope of the modulated spectra, and (2) a diffusion cue produced by binaural interaction of the 180° phase shifted reflection present at one ear. The availability of an additional level cue resulting from binaural interaction will be discussed in the model section of this chapter. While these cues may well account for the decreasing sensitivity, a substantially different mechanism must be operative for the shorter delay times to account for the differences observed between the diotic and dichotic listening modes. Discussion of possible, hypothetical mechanisms to account for these results will be postponed to be presented in a subsequent section dealing with binaural models.

<u>JND (Difference Limen) Data - Diotic and Dichotic Listening Mode</u>.

The JND results pooled across subjects and plotted in Figure 4.11 indicate that a similar pattern exists as that observed for the threshold data: (1) diotic JND thresholds increase with increasing delay; (2) dichotic JND thresholds decrease with increasing delay through 8 msec, while the 16 msec JND thresholds do not show any further decrease; (3) dichotic JND thresholds are substantially greater than diotic JND thresholds with the relative difference diminishing at 4 msec; a reversal in sensitivity occurs for the 8 and 16 msec conditions; (4) significant differences are observed across subjects. Although the data appear similar, one should not assume that the underlying mechanisms leading to detection are the same. The cues available to the subject may be quite different. For example, absolute threshold detection may be dependent on whether a pitch-like sound is audible (eg. Bilsen, 1968), whereas in the JND experiment, where this cue is present all of the time, detection may be primarily based on a change in pitch level.

In studying JND's one is tempted to think of describing the data in terms of a Weber fraction. The discussion which follows is not based on the classical JND paradigm in which just-noticeable differences would have been determined for different levels of m (see Glossary) for a fixed delay time. Rather, the approach taken is based on the hypothesis that the same underlying mechanism and the same criteria for detection are used for all delays, so that a constant Weber fraction would be expected across all delay times. Weber's law* holds that the fraction $\frac{\Delta I}{I}$, (where I represents a reference level and where ΔI represents the necessary increment to achieve a perceptual change), is constant. As indicated in the above paragraph, when JND's are plotted in terms of g both listening modes show a non-zero slope and therefore assuming the same mechanism independent of τ, appear to violate this law. However, the following discussion demonstrates that this violation is not significant, except in the 16 msec diotic case. Recall from the Methodology Chapter that the coefficient m, in the equation representing the input power spectrum of comb-filtered noise (see equation 3.2.1), is a measure of the modulation depth of the spectrum. Furthermore, since modulation depth directly provides a measure of spectral information, it was felt that this measure would be more closely related to the cue upon which subjects' decisions are based. Indeed, it can be shown that the ratio $\frac{m'-m}{m}$ obtained by comparing the modulation depth at the JND reference level of 0.35v to the modulation depth at the 75% correct level (obtained JND estimate) is roughly equivalent to a Weber fraction. Modulation depth ratios were calculated for the diotic JND estimates and are shown in Table 5.1. The resulting ratios are

---

* More recently, Stevens (1957) has questioned the validity of this relationship and has further suggested that the relationship of sensory magnitude and stimulus intensity be more accurately described by a power function shown in the following expression:

$$\psi = k(I-Io)^n$$

When the exponent n is much less than 1, the logarithmic function proposed by Fechner and the Power function proposed by Stevens are very similar.

	Diotic	Dichotic
2 msec	3.44 dB	2.38 dB
4 msec	3.68 dB	2.72 dB
8 msec	2.61 dB	3.67 dB
16 msec	-	3.76 dB

TABLE 5.1: Modulation ratio $(\frac{m'-m}{m})$ results in dB $(10 \log_{10} \frac{(m'-m)}{(m)})$ for diotic and dichotic listening modes.

seen to be reasonably constant varying over a range of less than 1.07 dB for the 2, 4 and 8 msec delay conditions. The reason why most subjects were not able to reach the 75% correct criterion for the 16 msec diotic conditions is straight-forward: it is the result of the proximity of the JND reference value of 0.35v (g=-9.18 dB) and the measured absolute threshold estimate for this condition of 0.36v ($g\cong-9.00$ dB). Weber fractions obtained for stimuli near <u>absolute</u> threshold are known to be substantially larger than the constant fraction obtained for levels well above threshold. The maximum possible increase in modulation depth, provided by equal levels of the direct and delayed signals (g=0 dB) yields an $\frac{m'-m}{m}$ ratio of 2.19 dB. For three out of four subjects this range (2.19 dB) was inadequate to provide a measureable JND.

Modulation depth ratios $(\frac{m'-m}{m})$ obtained for the dichotic JND data are also shown in Table 5.1. These data are quite similar to the diotic values. The fact that an approximately constant and similar value is obtained suggests that modulation depth is a central variable for detection (JND) in both diotic and dichotic listening modes, and furthermore that, whatever the form of binaural interaction, both listening modes conform to the Weber fraction by yielding a relatively constant value. These results are also consistent with those reported by A. H. Koenig et al. (1975).

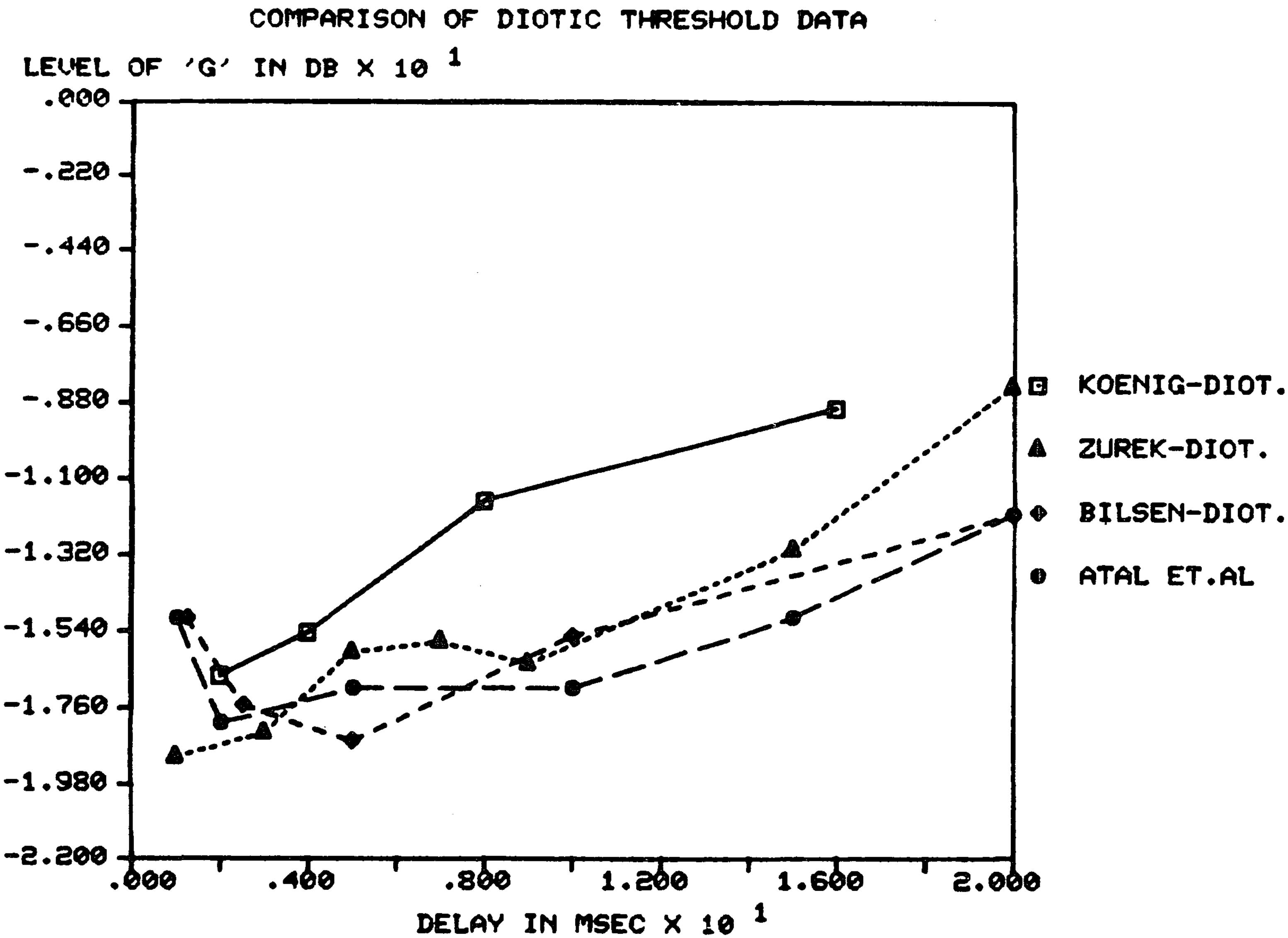
Comparison with Published Data

Diotic Threshold Data

Several investigators (Atal et al, 1962; Bilsen, 1968, and Bilsen and Ritsma, 1970; and Zurek, 1976) reported diotic threshold data for single echoes. These studies differed from the present study both in terms of overall listening levels and psychophysical testing procedures used. (A comparison of these test parameters used in the studies is shown in Table 5.2) Figure 5.2 shows a comparison of the earlier results with those obtained in the present study. To facilitate this comparison, the results for the medium and low listening levels obtained in the current study have been pooled. Inspection of Figure 5.2 shows: (1) a decrease in sensitivity as a function of increasing delay for all data with $\tau \geq 2$ msec; (2) for delays of 8 and 16 msec, the present threshold data are less sensitive by several dB.

The use of the ABX paradigm in the current study may have resulted in slightly elevated threshold measurements due to the inherent memory component. That is, subjects first had to compare stimulus A with stimulus B and retain this distinction so as to make the final match with stimulus X. The paired comparison paradigm used by Atal et. al., however, offered no control over the criterion level used by the subject (Green and Swets, 1966).

Figure 5.2 Comparison of diotic threshold data.


COMPARISON OF DIOTIC THRESHOLD DATA
LEVEL OF 'G' IN DB X 10 1
.000
-.220
-.440
-.660
-.880
-1.100
-1.320
-1.540
-1.760
-1.980
-2.200
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
KOENIG-DIOT.
ZUREK-DIOT.
BILSEN-DIOT.
ATAL ET.AL

In addition to the differing psychophysical procedures used, the previous studies shown here did not mention that any attempt was made to equate for loudness. The increase in intensity produced by the addition of the reflected signal, in the absence of some form of loudness adjustment, may well have served to facilitate discrimination and thereby to lower the threshold. However, since detection studies (eg. Moore, 1977, pg. 68), using white noise, have shown that intensity level differences in the range of Ø.5-1.Ø dB are required for intensity discrimination, the lack of an intensity adjustment should only contaminate thresholds where the intensity difference (between the reference and comparison stimulus) exceeds Ø.5 dB. It can be shown (See equation 3.8.1 in the Methodology Chapter) that a level of g = -9.11 dB must be reached before a Ø.5 dB difference in level is obtained. Inspection of Figure 5.2 shows that the diotic threshold levels of g reported by other investigators did not reach this level, and it is therefore unlikely that this contaminant is significant. The salient differences between the reported studies and the current study have been presented in an attempt to account for diotic threshold differences (particularly for the 8 and 16 msec delay conditions). While significant methodological differences exist (see Table 5.2) it is unclear whether the threshold differences can be ascribed to one or more of the methodological differences.

	Overall Listening Level	Experimental Paradigm	Level Correction	Stimulus Duration	Threshold Criteria (% correct)	Number of Subjects
Atal et. al. (1962)	60 dB SL	Paired Comparison (A-B)*	None	Not Specified	75	8
Bilsen (1968)	40 dB SL	Same - Different**	None	200 msec	75	3
Zurek (1976)	Not Specified	Two-alternative forced choices adaptive procedure	None	250 msec	70.7	3
Koenig (1979)	30 dB SL 50 dB SL	A-B-X Paradigm	Equal rms	374 msec	75	4

* Subjects were asked to indicate which of the two stimuli presented was more colored.

** Subjects were asked to indicate whether they perceived a difference between two stimuli.

TABLE 5.2: Comparison of Test Parameters Used in Diotic Threshold Experiments.

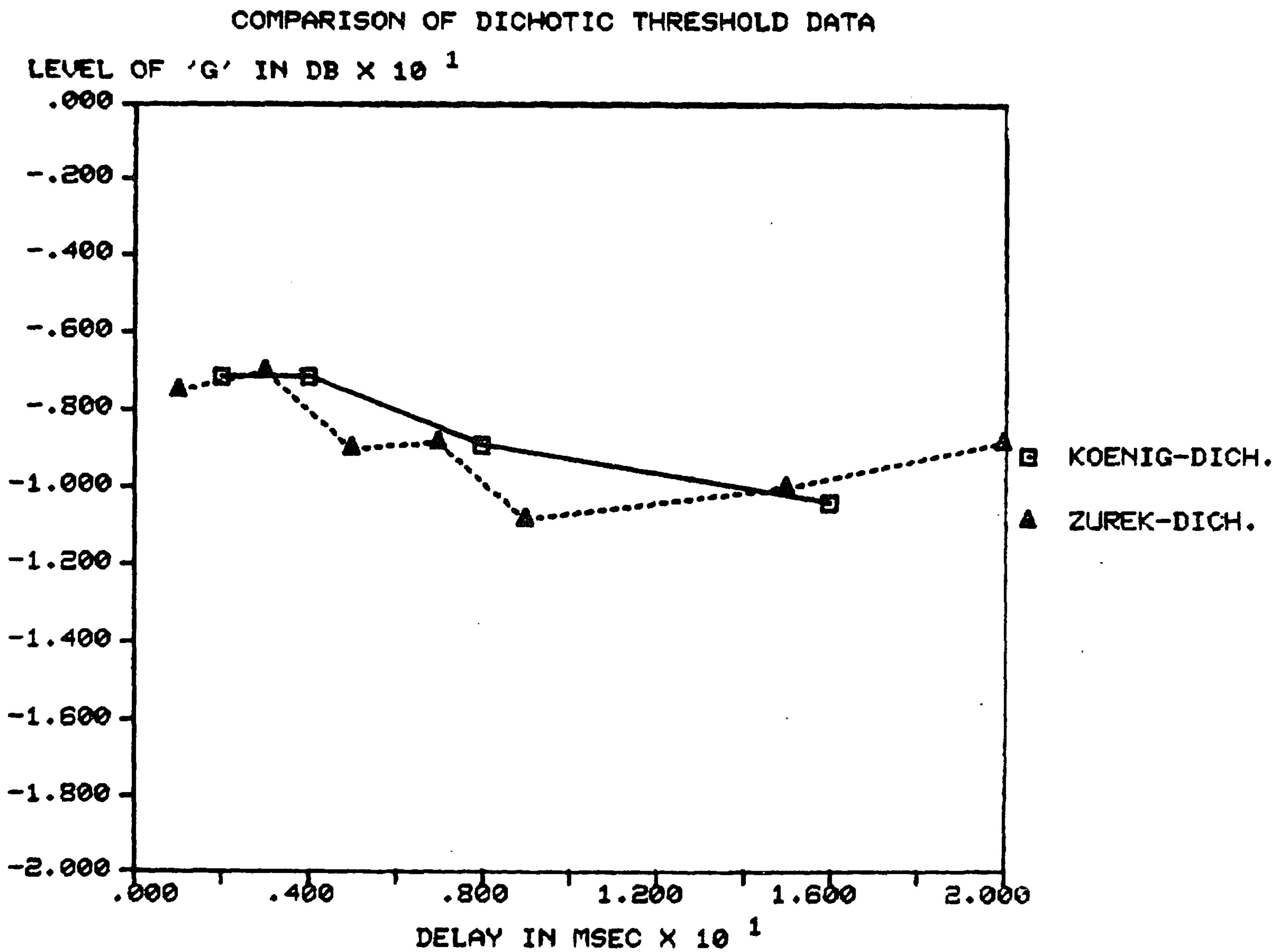
Dichotic Threshold Data

To date, Zurek (1976) has published the only dichotic single reflection data. These values along with the dichotic data reported here are shown in Figure 5.3. While two different dichotic experiments were reported by Zurek, only one of these experiments is directly relevant to the present dissertation. This experiment was similar to the dichotic threshold experiment reported here. The major difference between the two experiments is that Zurek used independent noise sources for both ears while the same noise source was used in the present experiment. Despite this major difference, the two sets of data (triangles and squares) exhibit marked similarity. Furthermore, for delays up to 8 msec both sets of data show the dichotic thresholds to be elevated relative to the diotic thresholds. The Zurek data demonstrate that the binaural interaction, presumably responsible for the observed diotic-dichotic differences, is not dependent on whether or not the noise sources are correlated at both ears. This, of course, assumes that the same underlying effect is responsible for both sets of data.

Exploration For A Model

A most striking feature of the data reported here is the difference between the diotic and dichotic data. In this section, four heuristic models are considered in an attempt to further understand this difference.

Figure 5.3 Comparison of dichotic threshold data.


COMPARISON OF DICHOTIC THRESHOLD DATA
LEVEL OF 'G' IN DB X 10 1
.000
-.200
-.400
-.600
-.800
-1.000
-1.200
-1.400
-1.600
-1.800
-2.000
.000
.400
.800
1.200
1.600
2.000
DELAY IN MSEC X 10 1
KOENIG-DICH.
ZUREK-DICH.

Simple Spectrum Addition

One of the simplest heuristic models which has been suggested (Zurek, 1976) is one which assumes an unweighted summation of the monaural power spectra. By analogy with the critical-band model, detection should occur when the resulting modulation depth exceeds a critical threshold. However, if each ear is presented with a comb-filtered noise spectrum such that the modulations appearing at each ear are complementary to each other, (ie. a maximum in the spectrum for any given frequency will find a minimum in the spectrum for that frequency at the other ear), simple summation of these spectra will produce a smooth spectrum. Thus, we are forced to reject this intuitively attractive model on the grounds that the formation of a smooth spectrum resulting from addition of complementary noise sources would never permit the subject to reach threshold.

Repetition Pitch Model

The second model is an adaptation of the Bilsen and Goldstein (1974) dichotic repetition pitch model (see Figure 2.19) discussed in the Review chapter. The adaptation of this model discussed below differs from the original in that it is not concerned specifically with the perception of pitch, but rather with the use of periodic information in the spectrum to account for the detectability of the echoes. Again, as in the first model, in this case, the criterion for detection is the periodic variation in spectral level. This model differs from the first model in that addition of signals rather then spectra form the dichotic result. The present model also includes the peripheral bandpass filters which were invoked previously to account for the decreasing sensitivity with increasing delay observed in the diotic listening mode as well as stochastic gain and delay parameters in each channel. It can be shown (Sondhi, 1979) that when complementary stimuli are applied to each input, this model also predicts perfect spectral summation in that a smooth spectrum is obtained, (i.e. the periodic frequency information contained in the comb-filter spectrum is lost). This spectrum results from the fact that the random amplitude and time jitter constants (see Figure 2.19) ($1-\epsilon_1$ and $1-\epsilon_2$, $1 - \delta$ and $1 - \delta$) are defined as statistically independent, zero-mean Gaussian random variables with equal variance. Performing an ensemble average as required by the model necessarily leads to a smooth binaural spectrum which is independent of τ.

The absence of frequency information in the spectrum as predicted by the previous models requires us to consider other cues which might be available. Two possible candidates are: (1) level cues and (2) monaural cues. These are considered in the following two models.

Waveform Summation Model

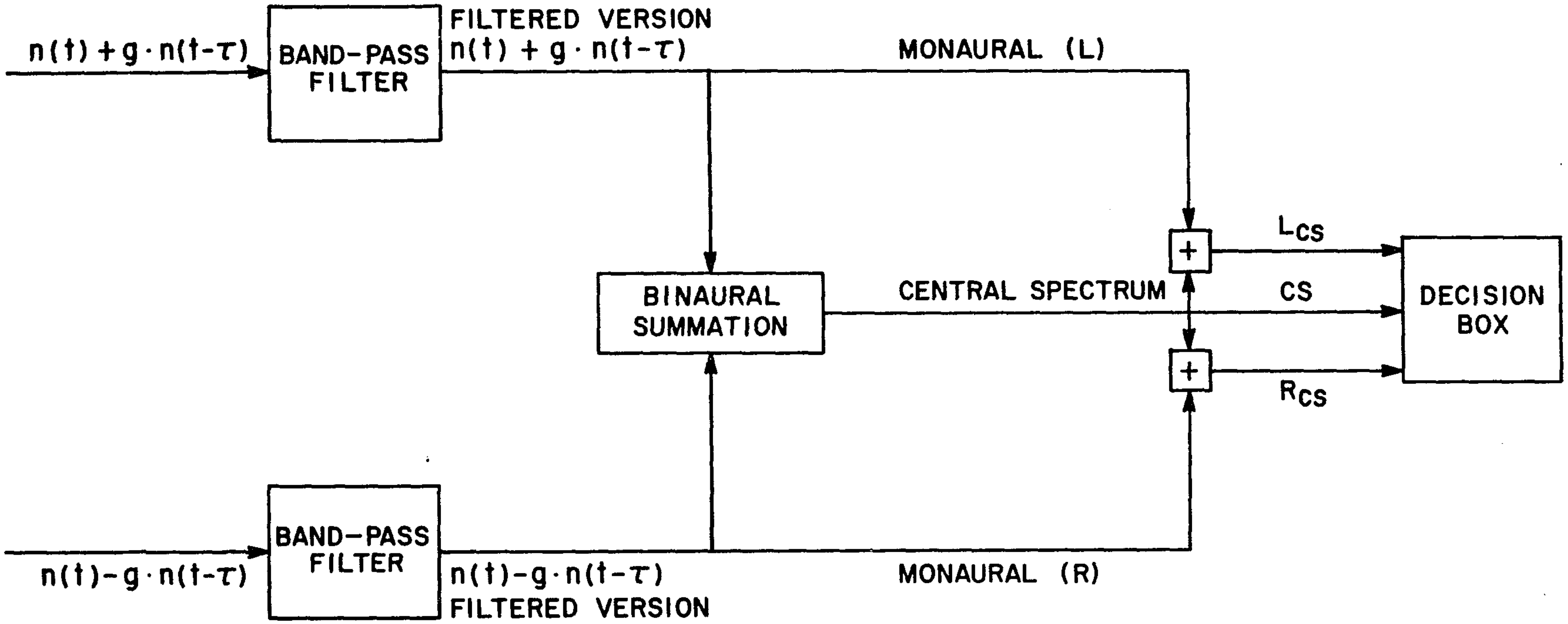
This model is formally the same as the repetition pitch model, but without the stochastic variables. Detection can occur on the basis of absolute loudness discrimination. Thus it represents the simple addition of input waveforms filtered by the critical-band filters. Addition of the dichotic waveforms after filtering by each critical-band (See Methodology Chapter - Figure 3.1) serves to cancel the delayed component of the input waveform. Since this is a linear process the direct component must now be multiplied by a coefficient of 2. That is, the summed spectrum (presumably, a central spectrum) will be flat with twice the original amplitude. However, all stimuli were equalized in intensity (i.e. all stimuli were normalized to have the same rms level). Thus, stimuli with a delayed component had to be scaled down (scaling factors are shown in Tables 3.1 and 3.2) due to the presence of the delayed signal. Summation of the scaled-down; direct components will therefore yield a level which is less than twice the level of the unscaled direct signal. However, since the delayed component was absent for the dichotic reference stimulus (Threshold experiment), the scaling factor was set to 1.Ø; that is, no level adjustment was required. The net result of the intensity adjustment is to produce at the output of this model, a signal level difference between the reference stimulus and the dichotic comparison stimulus. It becomes clear then that dichotic thresholds requiring large values of g (the level of the delayed signal) might well be based on level differences between the dichotic reference and comparison output signals levels. This effect will necessarily increase as the level of g increases. For the 2 msec dichotic threshold condition where the obtained value of g equals -7.14 dB, (low and medium 2 msec dichotic threshold values have been pooled.) a scaling factor of about 0.91* was used representing a decrease in overall level of 0.82 dB which may well be detectable. (A similar level cue also occurs in each of the previously cited models. However, the role of this level cue was not included in their formulation.)

<u>Monaural Detection with Binaural Masking</u>

Both models presented above neglect the striking fact that the individual monaural signals, although available to subjects, were not fully utilized as was generally assumed by other authors. The third model addresses this aspect of the experimental results. This model shown in Figure 5.4 is similar to the waveform summation model, but it is more akin to the traditional E-C model in that the monaural inputs are individually available and play a role in the formation of the perceived signal.

* This value is obtained by determining the value g in volts (ie. g in volts = antilog of the (dB value / 20) = 0.44 volts; this value of g is then inserted in equation 3.8.1.

Figure 5.4 Binaural interaction model with binaural masking of monaural signals.


n(t) + g · n(t-τ)
BAND-PASS FILTER
FILTERED VERSION
n(t) + g · n(t-τ)
MONAURAL (L)
BINAURAL SUMMATION
CENTRAL SPECTRUM
+
Lcs
CS
Rcs
DECISION BOX
n(t)-g · n(t-τ)
BAND-PASS FILTER
n(t)-g · n(t-τ)
FILTERED VERSION
MONAURAL (R)

The output of each monaural bandpass filter is fed into a binaural summation device. The output of the summation device is defined as the central spectrum. (Binaural summation presumably produces a fused image. The inability to fuse the monaural inputs, will eliminate the central spectrum component. In this instance, the perceived spectrum is based on either filtered monaural signal).

The central spectrum then forms one component of the perceived spectrum. In addition, the central spectrum is also summed with, or serves to mask, the individual monaural spectra, all of resulting spectra are used in the decision making process. It is assumed that when the monaural input levels are not equal, the more intense monaural signal is summed with the central spectrum. In the case of equally intense monaural inputs, the selection is based on a maximal information criterion. (Either input (or both) are used. If inputs are unequal some form of equalization would be appropriate). The perceived spectrum is thus based on the central spectrum and the summation of this central spectrum with <u>one</u> of the monaural spectra. Spectral variation alone, as in the first two models, provides the criterion for detection. Overall loudness effects are not considered.

If we assume equal monaural input levels, as used in both experiments reported in the present dissertation, the perceived spectrum is readily determined: (See Figure 5.4)

$$\text{Left Monaural(L)} = \left| n(t) + g \cdot n(t-\tau) \right|^{*} \tag{5.4.1}$$

$$\text{Right Monaural(R)} = \left| n(t) - g \cdot n(t-\tau) \right| \tag{5.4.2}$$

thus:

$$\text{Central Spectrum (CS)} = L + R = \left| 2n(t) \right| \tag{5.4.3}$$

and,

$$L_{CS} = L + CS = \left| 3n(t) + g \cdot n(t-\tau) \right| \tag{5.4.4}$$

so,

$$L_{CS} \propto \left| n(t) + (g/3) \cdot n(t-\tau) \right| \tag{5.4.5}$$

In this instance the perceived spectrum will not be a perfectly smooth spectrum, as predicted by simple binaural spectral addition, but rather will result in a modulated spectrum where the modulation depth has been reduced. Equation (5.4.5) shows that the level of g will be reduced by one-third or 9.54 dB (20 log .333) relative to that of the unmasked monaural signals (which are equivalent to the signals obtained in the diotic listening mode).

This model predicts a constant difference of 9.5 dB between diotic and dichotic thresholds. That is, according to this model, the dichotic thresholds should be parallel to the diotic thresholds but displaced (decreased sensitivity)

** The brackets indicate that these signals are filtered by the critical-band filters which has no substantive effect on the calculation. To simplify the calculation, g is assumed small and level correction is neglected. It can be shown to have no effect on detection.

by 9.5 dB. Examination of Figure 4.5 shows this not to be the case. While agreement with the pooled difference for the 2 and 4 msec delay conditions is reasonable (8.9 and 7.0 dB respectively) the difference between the two listening modes at 8 msec diminishes to 3.4 dB. Thus, at the longest delay times, there is an obvious departure of the observed values from these predicted. Although somewhat speculative, it seems reasonable to assume that for the longer times dichotic diffusion, as discussed previously, could account for the variation of dichotic thresholds with τ. Recall, however, that the underlying consideration for development of the masking model was to provide a mechanism which specifically recognizes the importance of the individual monaural inputs as they have been provided for in existing binaural models (e.g. Durlach, 1972). This is especially important since the major models existing in the literature have assumed the monaural inputs to be fully available for decision making.

Each of the four models presented above represents an attempt to understand one aspect of the differences between diotic and dichotic listening mode results. It is not possible to say, based on current data, that one or the other model is valid, partially valid, or invalid. However, certain conclusions can be drawn from the discussion:

1. Power spectrum or waveform addition is the basis of central spectrum formation in all the models presented.

No comb-filter variations are seen in the resulting central spectra and hence such variations in the spectrum cannot explain the existence of dichotic thresholds.

2. Differences in absolute spectral levels exist in all these models and are explicitly recognized as a detection cue in the third (waveform summation) model. If any mechanism of this sort is active such a central spectrum level is probably inherent in spectral flattening and could account for some aspects of the data.

3. Reduction of monaural input cues cannot be ignored since, otherwise, no differences would exist between diotic and dichotic results. This is an important experimental result but the model presented only suggests one possible direction of thought and doesn't even predict all the currently obtained data.

4. A further cue, diffusion, has been observed. Although its role is unclear it may account for the variation of dichotic threshold with τ.

Practical Implications of Spectral Smoothing

We have already demonstrated that when noise is added to itself with an intervening delay, a frequency spectrum with spectral variations characteristic of a comb-filter is produced. The comb-filtered noise has a distinct pitch-like quality and is readily distinguished from the original white noise signal. Speech signals processed in this way can also be shown to possess similar spectral characteristics. That is, the spectrum of a speech signal combined with its reflection will show the same spectral variations as found in the noise case. Furthermore, the spectrally distorted speech signal is often reported to possess a rain barrel-like quality and is referred to as colored. This perceptual effect is usually judged as unpleasant and undesirable. Preliminary experiments have confirmed that the perception of this barrel-like quality is directly dependent on the delay between the direct signal and the reflected signal as well as on the level of the reflected signal.

It has been long recognized that this interaction of a speech signal with a delayed version of itself (reflection) is frequently encountered in conference telephony when a single microphone is so situated as to pick up a strong reflection. In an attempt to reduce the perceived coloration, Flanagan and Lummis (1968) suggested a system consisting of two microphones be used in conference telephony. Essentially their approach was to band-pass filter the output of each microphone, select the output for each band with the greater power over some defined time period and then recombine the bands for single channel transmission. This attempt to emulate the presumed behavior of the binaural hearing system reduces the spectral irregularities in the recombined channel relative to the irregularities found in either channel alone. Reducing these irregularities then serves to reduce the perceived distortions i.e., the barrel-like sensations.

Recently Allen, Berkley and Blauert (1977) have presented a more elaborate multimicrophone technique for removing spectral distortions produced by reflections. The outputs of individual microphones are divided into frequency bands whose outputs are cophased (i.e. delay differences are compensated) and then the signals for each frequency are recombined to form a single monaural output also in an attempt to simulate two-ear hearing.

While both techniques serve to reduce (to some degree) the perceived coloration, the normal binaural listener apparently possesses a natural capability of reducing these colorations as is suggested by the present thesis.

CHAPTER VI

SUMMARY, CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

The perception of single echoes for two binaural listening modes, diotic and dichotic, was examined in this dissertation. As early as 1950, W. Koenig observed that two-channel listening serves to reduce or squelch perceived reverberation relative to that experienced in single-channel listening. Previous psychoacoustic studies had focused mainly on the ability to localize and lateralize sounds in the presence of reflections. The purpose of this disseration was to examine the perception of echoes using a simplified model of reverberation.

Two separate experiments have been reported: Experiment I is an absolute threshold experiment dealing with the perceptibility of a single echo while Experiment II is a just-noticeable-difference (JND) experiment which examines the perceptibility of an increase in amplitude of an echo above a fixed reference level.

An ABX psychophysical paradigm was used in both experiments. Four delay values, 2,4,8 and 16 msec at overall sensation levels of 30 and 50 dB SL were presented to 4 subjects. All conditions were presented in two listening modes: (1) diotic listening mode where both ears received identical signals and (2) dichotic listening mode where the polarity of the reflected signal was reversed in one ear relative to the other.

Estimates of the stimulus level for 75-percent correct identification criterion were obtained for each experimental condition. A maximum-likelihood estimation procedure assuming a cumulative normal response function was used. For the absolute threshold experiment, diotic listening-mode thresholds were found to be significantly different from dichotic listening mode thresholds. For delays of 2 and 4 msec, the diotic listening-mode thresholds were substantially elevated relative to the dichotic listening-mode thresholds. Inspection of the pooled data showed that diotic thresholds exhibited decreasing sensitivity with increasing delay while dichotic thresholds exhibited an opposite trend; that is, sensitivity increased with increasing delay.

Data obtained in the JND experiment exhibited a pattern similar to that obtained in the threshold experiment: diotic thresholds increased with increasing delay while dichotic thresholds decreased with increasing delay. While the effect of listening mode was not statistically significant for the JND experiment, the listening mode by delay interaction was significant at the $P < 0.05$ level.

The diotic threshold data are shown to be consistent with existing critical-band models of hearing. The dichotic threshold results lend support to a mechanism of binaural combination which serves to reduce spectral variation. Four separate but related models which incorporate some form of binaural interaction have been discussed in an attempt to account for the diotic-dichotic threshold differences. In addition, a number of factors were discussed which may also contribute to the effect including: image diffusion, monaural spectral cues, and level cues.

Investigators have typically demonstrated that the detectability of a signal is primarily determined by those cues, either monaural or binaural, which maximize the signal-to-noise ratio (e.g. Durlach, 1963). The data obtained in this study are unique in that the available monaural cues apparently are not used by the auditory system for detecting the echo. These data are viewed as suggestive of a binaural interaction mechanism resulting in some degree of binaural spectral flattening of the perceived signal. This plus the practical importance of spectral flattening (cited in the Discussion Chapter) are clear signposts for further investigation.

Implications For Future Research

While the data are insufficient to validate any of the models discussed, an underlying mechanism is suggested. The decreased sensitivity thresholds reported for the dichotic listening mode relative to the diotic listening mode indicate that spectral flattening (the reduction of spectral variations) as the result of some form of binaural processing is a likely mechanism. Further investigation is clearly needed to elucidate the details of this mechanism. In particular, additional attention must be given to the influence of concomitant auditory cues (ie. lateralization and/or diffusion cues) on this effect. Furthermore, consideration should be given to the effect of loudness equalization and the influence of individual monaural spectral cues. An additional question which must be addressed is the extent to which spectral flattening occurs when spectra are not precisely complementary.

REFERENCES

[1] Allen, J. B., Fastfit - An FFT Based Filtering Program, in IEEE Press Book on Digital Signal Processing, IEEE Press, New York 1979.

[2] Allen, J. B., D. A. Berkley and Blauert, J., "Multimicrophone Signal-Processing Technique to Remove Room Reverberation From Speech Signals," Journal of the Acoustical Society of America, 62, 912-915, (1977).

[3] Atal, B. S., Schroeder, M. R. and Kuttruff, H., "Perception of Coloration in Filtered Gaussian Noise - Short Time Spectral Analysis by the Ear," Fourth International Congress on Acoustics, Copenhagen, 21-28 August, 1962.

[4] Bilsen, F. A., "Repetition Pitch: Monaural Interaction of a Sound with the Repetition of the Same but Phase Shifted Sound," Acoustica, 17, 295-300, (1966).

[5] Bilsen, F. A., "On the Interaction of a Sound with its Repetitions," Doctoral Dissertation, Delft University of Technology, (1968).

[6] Bilsen, F. A., "Pitch of Noise Signals: Evidence for a Central Spectrum," Journal of the Acoustical Society of America, 61, 150-161, (1977).

[7] Bilsen, F. A. and Ritsma, R. J., "Repetition Pitch and Its Implication for Hearing Theory," Acoustica, 22, 63-73, (1969/1970).

[8] Bilsen, F. A. and Ritsma, R. J., "Some Parameters Influencing the Perceptibility of Pitch," Journal of the Acoustical Society of America, 47, 469-475, (1970).

[9] Bilsen, F. A. and Goldstein, J. L., "Pitch of Dichotically Delayed Noise and Its Possible Spectral Basis," Journal of the Acoustical Society of America. 55, 292-296, (1974).

[10] Bilsen, F. A., ten Kate, J. H., Buunen, T. J. F., and Raatgever, J., "Responses of Single Unites in the Cochlear Nucleus of the Cat to Cosine Noise," Journal of the Acoustical Society of America, 58, 858-866, (1975).

[11] Biomed - Analysis of Variance, Health Sciences Computing Facility, UCLA, (1971).

[12] Broadbent, D. E., "A Note on Binaural Fusion," Quarterly Journal of Experimental Psychology, 7, 46-47, (1955).

[13] Broadbent, D. E. and Ladefoged, P., "On the Fusion of Sounds Reaching Different Sense Organs," Journal of the Acoustical Society of America, 29, 708-710,

(1957).

[14] Bury, K. V. Statistical Models in Applied Science, John Wiley and Sons: New York, (1975).

[15] Cherry, C., "Two-Ears -- But One World," Chapter 6, Sensory Communication, Walter A. Rosenblith, Ed., The M. I. T. Press: Cambridge, Mass., 1961, pp. 529-536.

[16] Colburn, H. S., "Theory of Binaural Interaction Based on Auditory Nerve Data. I. General Strategy and Preliminary Results on Interaural Discrimination," Journal of the Acoustical Society of America, 54, 1458-1470, (1973).

[17] Corliss, E., and Winzer, G. E., "Studies of Methods of Estimating Loudness," Journal of the Acoustical Society of America, 38, 424-428, (1965).

[18] E. M. Cramer and W. H. Huggins, "Creation of Pitch through Binaural Interaction," Journal of the Acoustical Society of America, 30, 413-417, (1958).

[19] David, E. E., "Closing the Binaural Gap," Journal of the Acoustical Society of America, 34, 728(A), (1962).

[20] David, E. E., Jr., Guttman, N., and van Bergeijk, W. A., "Binaural Interaction of High Frequency Complex Stimuli," Journal of the Acoustical Society of America, 31, 774-782, (1959).

[21] David, E. E., Guttman, N., and van Bergeijk, W. A., "On the Mechanism of Binaural Fusion," Journal of the Acoustical Society of America, 30, 801-802, (1958).

[22] Dolan, T. R., "Effects of masker spectrum level on masking-level differences at low signal frequencies," Journal of the Acoustical Society of America, 44, 1507-1512, (1968).

[23] Durlach, N. I., "Equalization and Cancellation Theory of Binaural Masking-Level Differences," Journal of the Acoustical Society of America, 35, 1206-1218, (1963).

[24] Durlach, N. I., "Binaural Signal Detection: Equalization and Cancellation Theory," in Modern Foundations of Auditory Theory, J. V. Tobias, Ed., Academic Press: New York, vol. II, (1972).

[25] Fay, R. D., "A Method for Obtaining Natural Directional Effects in a Public Address System," Journal of the Acoustical Society of America, 7, 239, (1936).

[26] Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A., "Localization of High-Frequency Tones," Journal of the Acoustical Society of America, 29, 988-991, (1957).

[27] Finney, D. J., Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve, 2nd ed., Cambridge University Press: London, (1964).

[28] Flanagan, J. L., and Lummis, R. C., "Signal Processing to Reduce Multipath Distortions in Small Rooms," Journal of the Acoustical Society of America, 47, 1475-1481, (1970).

[29] Fourcin, A. J., "An Aspect of the Perception of Pitch, in Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki (Mouton, The Hague), pp. 355-359 (1962).

[30] Fourcin, A. J., Central Pitch and Auditory Lateralization," in Frequency Analysis and Periodicity Detection in Hearing, ed. by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 319-328, (1970).

[31] Franssen, N. V., "Eigenschaften des natuslechen Richteingshorens und ihre Anwendung auf die Stereophonic," Proceedings International Congress on Acoustics, 3rd Stuttgart, 2, 228-290, (1961).

[32] Gardner, M. B., "Historical Background of the Haas and/or Precedence Effect," Journal of the Acoustical Society of America, 43, 1243-1248, (1968).

[33] Gardner, M. B., "Image Fusion, Broadening, and Displacement in Sound Location," Journal of the Acoustical Society of America, 46, 339-349, (1969).

[34] Green, D. M., and Swets, J. A., Signal Detection Theory and Psychophysics, John Wiley and Sons: New

York (1966).

[35] Green, D. M., An Introduction to Hearing, Lawrence Erlbaum Associates: Hillsdale, New Jersey, (1976).

[36] Green, D. M., and Henning, G. B., "Audition," Annual Review of Psychology, 20, 105-128, (1969).

[37] Green, D. M. and Yost, W. A., "Binaural Analysis," Ira W. Keidel and N Neff (eds.), Handbook of Sensory Physiology, Springer-Verlag: Berlin, Hiedelburg and New York, (1975).

[38] Guttman, N., "Binaural Click Lateralizations with a Three-Click Paradigm," Journal of the Acoustical Society of America, 34, 728(A), (1962).

[39] Guttman, N., "Binaural Interactions of Three Clicks," Journal of the Acoustical Society of America, 37, 145-150, (1965).

[40] Haas, H. H. "The Influence of a Single Echo on the Audibility of Speech," Dissertation at the University of Goettingen. Translated by K. P. R. Ehrenberg, Department of Scientific and Industrial Research, Building Research Station, Library Communication No. 363, Garston, Watford, Herts, England.

[41] Haas, H. "The Influence of a Single Echo on the Audibility of Speech," Journal of the Audio Engineering

Society, 20, 146-159, (1972).

[42] Hall, W. M., "A Method for Maintaining in a Public Address System the Illusion that the Sound Comes From the Speaker's Mount," Journal of the Acoustical Society of America, 7, 239, (1936).

[43] Hafter, E. R., Bourbon, W. T., Blocker, A. S., Tucker, A., "A Direct Comparison Between Lateralization and Detection Under Conditions of Antiphasic Masking," Journal of the Acoustical Society of America, 46, 1452-1457, (1969).

[44] Harris, J. D., "Remarks on the Determination of a Differential Threshold by the So-Called ABX Technique," Journal of the Acoustical Society of America, 24, 417, (1952).

[45] Harris, G. G., Flanagan, J. L., and Watson, B. J., "Binaural Interaction of a Click with a Click Pair," Journal of the Acoustical Society of America, 35, 672-678, (1963).

[46] Henning, G. B., "The Detectability of Interaural Delay in High-Frequency Complex Waveforms," Journal of the Acoustical Society of America, 55, 84-90, (1974a).

[47] Henning, G. B., "Lateralization and the Binaural Masking-level Difference," Journal of the Acoustical Society of America, 55, 1259-1262, (1974b).

[48] Henning, G. B., "Auditory Localization," in Handbook of Psychobiology, M. S. Gazzinga and Blakemore, Colin, eds., Academic Press: New York, (1975).

[49] Hirsh, I. J., "The Influence of Interaural Phase on Interaural Summation and Inhibition," Journal of the Acoustical Society of America, 2Ø, 536-544, (1948).

[5Ø] Huggins, W. H., "Outline of an Experiment for Measuring Nonperipheral Pitch-Perception," Interdepartmental memorandum, The Johns Hopkins University (1954).

[51] Jeffress, L. A., "A Place Theory of Sound Localization" Journal of Comparative and Physiological Psychology, 41, 35-39 (1948).

[52] Jeffress, L. A., "Binaural Signal Detection: Vector Theory," in Modern Foundations of Auditory Theory, J. V. Tobias, Ed., Academic Press: New York, vol. II, (1972).

[53] Jeffress, L. A. and Taylor, R. W., "Lateralization vs Localization," Journal of the Acoustical Society of America, 33, 482-483, (1961).

[54] Jeffress, L. A., Blodgett, H. C. and Deatherage, B. H., "The Masking of Tones by White Noise as a Function of the Interaural Phase of Both Components - I. 5ØØ Cycles," Journal of the Acoustical Society of America 24, 523-527 (1957).

[55] Jeffress, L. A., Blodgett, H. C. and Deatherage, B. H., "Masking and Interaural Phase II - 167 CPS," Journal of the Acoustical Society of America, 34, 1124-1126, (1962).

[56] Klumpp, R. G., and H. R. Eady, "Some Measurements of Interaural Time Difference Thresholds," Journal of the Acoustical Society of America, 28, 859-860, (1956).

[57] Koenig, A. H., Berkley, D. A., Curtis, T. H., and Allen, J. B., "Magnitude of JND's for Diotic and Dichotic Perception of Spectrally Colored Noise," Journal of the Acoustical Society of America, 58, S55(A), (1975).

[58] Koenig, A. H., Allen, J. B., Berkley, D. A. and Curtis, T. H., "Determination of Masking-Level Differences in an Reverberant Environment," Journal of the Acoustical Society of America, 61, 1374-1376 (L) (1977).

[59] Koenig, W., "Subjective Effects in Binaural Hearing," Journal of the Acoustical Society of America, 22, 61-62, (L), (1950).

[60] Leakey, D. M., Sayers, B., McA. and Cherry, C., "Binaural Fusion of Low-and-High-Frequency Sounds," Journal of the Acoustical Society of America, 30, 222, (1958).

[61] Levitt, H., and Voroba, B., "Binaural Hearing," in Introductory Hearing Science: Physical and Psychological Concepts, S. E. Gerber, Ed., W. B. Saunders: Philadelphia, (1974).

[62] Licklider, J. C. R., "The Influence of Interaural Phase Relations Upon the Masking of Speech by White Noise," Journal of the Acoustical Society of America, 20, 150-159, (1948).

[63] Licklider, J. C. R., "Auditory Frequency Analysis," in Information Theory, ed. Colin Cherry, Academic Press: New York, 1956.

[64] Licklider, J. C. R., "Three Auditory Theories," in S. Koch (Ed.) Psychology: A Study of a Science. Vol. I., McGraw-Hill, New York (1959).

[65] Lochner, J. P. A. and Burger, J. F., "The Subjective Masking of Short Time Delayed Echoes By Their Primary Sounds and Their Contribution to the Intelligibility of Speech," Acoustica, 8, 1-10, (1958).

[66] Lochner, J. P. A. and Burger, J. F., "The Influence of Reflections on Auditorium Acoustics," Journal of Sound and Vibration, 4, 426-454, (1964).

[67] Macmillan, N. A., H. L. Kaplan and C. D. Creelman, "The Psychophysics of Categorical Perception," Psychological Review, 84, 452-471, (1977).

[68] Masterton, B. and Diamond, I. T., "Hearing: Central Neural Mechanisms" in Handbook of Perception; Biology of Perceptual Systems, Chapter 18, V3, Academic Press: New York, (1973).

[69] Mills, A. W., "On the Minimum Audible Angle," Journal of the Acoustical Society of America, 30, 237-246, (1958).

[70] Mills, A. W., "Lateralization of High-Frequency Tones," Journal of the Acoustical Society of America, 32, 132-134, (1960).

[71] Mills, A. W., "Auditory Localization," in Modern Foundations of Auditory Theory, J. V. Tobias, Ed., Academic Press: New York, Vo. II, (1972).

[72] Moore, B. C. J., Introduction to the Psychology of Hearing, University Park Press: Baltimore, Maryland, (1977).

[73] Osman, E., "A Correlation Model of Binaural Masking Level Differences," Journal of the Acoustical Society of America, 50, 1494-1511, (1971).

[74] Pierce, J. R. and Gilbert, E. N., "On AX and ABX Limens," Journal of the Acoustical Society of America, 30, 593-595, (1958).

[75] Plomp, R., Aspects of Tone Sensation, A Psychophysical

Study, Academic Press: New York, 1976.

[76] Raleigh, Lord, "On Our Perception of Sound Direction," Philosophical Magazine, 13 214-232, (1907).

[77] Robinson, D. E. and Jeffress, L. A., "Effect of Varying the Interaural Noise Correlation on the Detectability of Tonal Signals," Journal of the Acoustical Society of America, 35, 1947-1952, (1963).

[78] Rosenblith, W. A. and Stevens, K. N., "On the DL for Frequency," Journal of the Acoustical Society of America, 25, 980-985 (1953).

[79] Sandel, T. T., Teas, D. C., Fedderson, W. E. and Jeffress, L. A., "Localization of Sound from Single and Paired Sources," Journal of the Acoustical Society of America, 27, 842-852, (1955).

[80] Sayers, B. M. A. and Cherry, E. C., "Mechanisms of Binaural Fusion in the Hearing of Speech," Journal of the Acoustical Society of America, 29, 973-987, (1957).

[81] Scharf, B., "Localization of Unlike Tones From Two Loudspeakers," in Sensation and Measurement, Eds. Moskowitz, H. R., Scharf, B., and Stevens, J. C., D. Reidel Publishing Company: Dordrecht - Holland, (1974).

[82] Schroeder, M. R., "Natural Sounding Artificial Reverberation," Journal of the Audio Engineering Society, 10, 219-223, (1962).

[83] Sondhi, M. M., (1979) Personal Communication.

[84] Stevens, S. S., and Newman, E. B., "The Localization of Actual Sources of Sound," American Journal of Psychology, 48, 297-306, (1936).

[85] Tobias, J. V. and Schubert, E. R., "Effective Onset Duration of Auditory Stimuli," Journal of the Acoustical Society of America, 31, 1595-1605, (1959).

[86] Tobias, J. V.. and Zerlin, S., "Lateralization Threshold as a Function of Stimulus Duration," Journal of the Acoustical Society of America, 31, 1591-1594, (1959).

[87] Vermeulen, R., "Stereo-Reverberation," Journal of the Audio Engineering Society, 6, 124-130, (1958).

[88] von Bekesy, G., "Auditory Backward Inhibition in Concert Halls," Science, 17, 529-536, (12 February 1971).

[89] Wallach, H., Newman, E. B., and Rosenzweig, M. R., "The Precedence Effect in Sound Localization," American Journal of Psychology, 62, 315-336, (1949).

[90] Webster, F. A., "The Influence of Interaural Phase on Masked Thresholds," Journal of the Acoustical Society

of America, 23, 452-462, (1951).

[91] Witkin, H. A., Wapner, S., and Leventhal, T., "Sound Localization with Conflicting Visual and Auditory Cues," Journal of Experimental Psychology, 43, 58-67, (1952).

[92] Woodworth, R. S. and Schlosberg, H., Experimental Psychology, (Revised Edition), Holt, Rinehart and Winston, New York, (1954).

[93] Yost, W. A., Wightman, F. L. and D. M. Green, "Lateralization of Filtered Clicks," Journal of the Acoustical Society of America, 50, 1526-1531, (1971).

[94] Zurek, P., "An Investigation of the Binaural Perception of Echoed Sound," Doctoral Dissertation, Arizona State University, 1976.

[95] Zwislocki, J. and Feldman, R. S., "Just Noticeable Differences in Dichotic Phase," Journal of the Acoustical Society of America, 28, 860-864, (1956).

APPENDIX A

Date 2/3/78 By DHK Audiometer No.

Test

Retest

1703 Recording Audiometer
Grason-Stadler a GR company

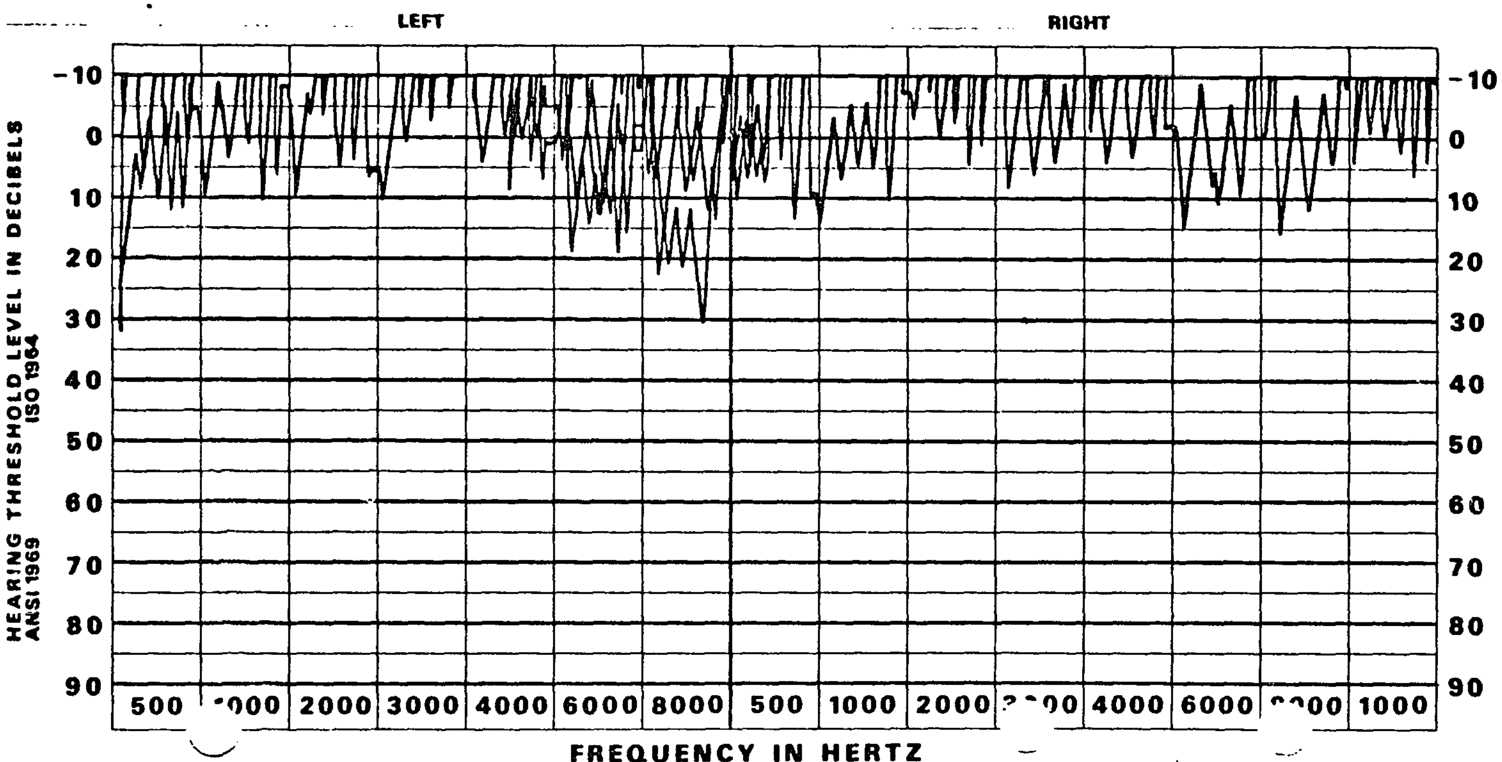


CHART No 1703 9102

REORDER FROM GRAPHIC CONTROLS CORP., BUFFALO N.Y. 14210

PRINTED IN U.S.A.

Subject No. 1

Date 4/3/78 By [illegible] Audiometer No.

Test

Retest

1703 Recording Audiometer
Grason-Stadler a GR company

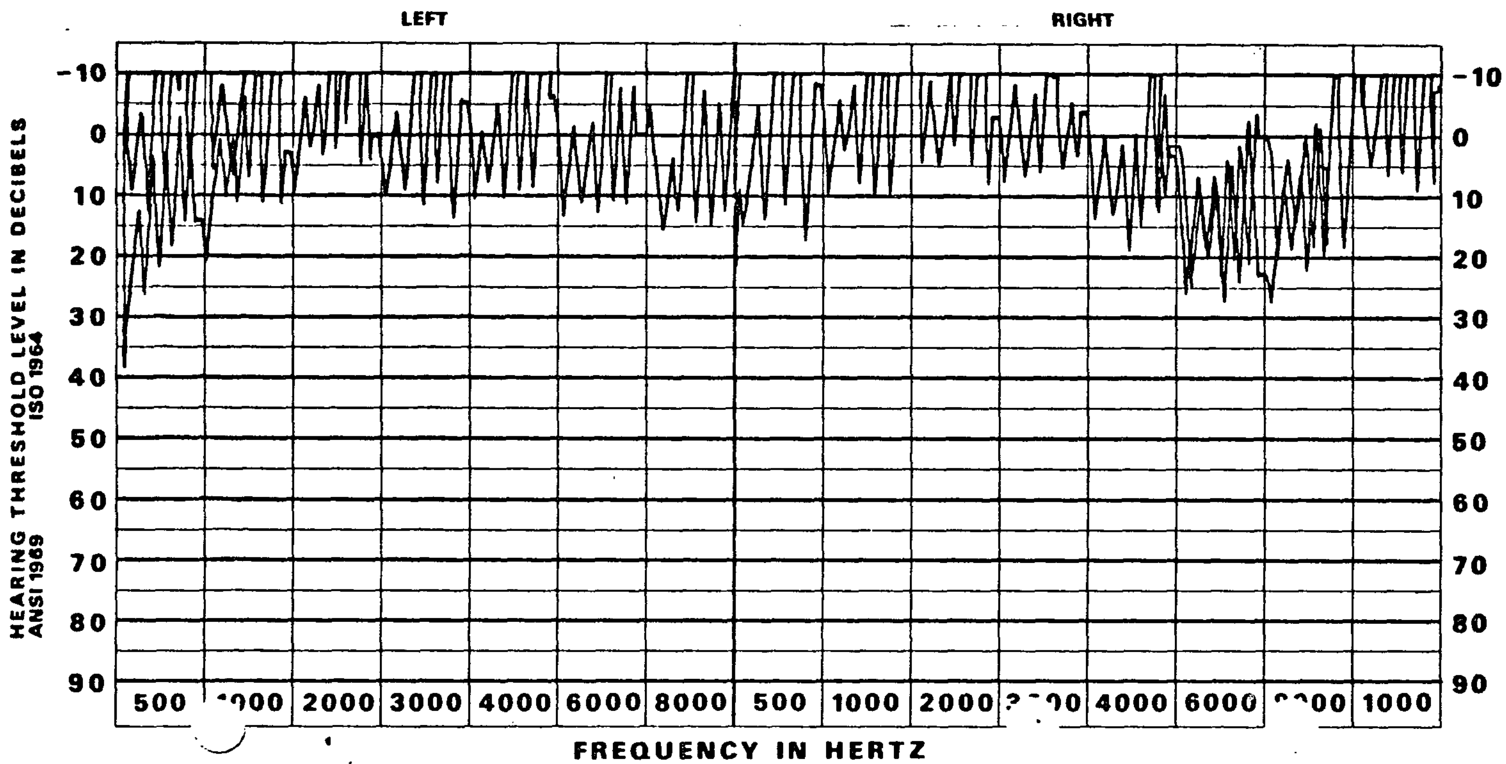
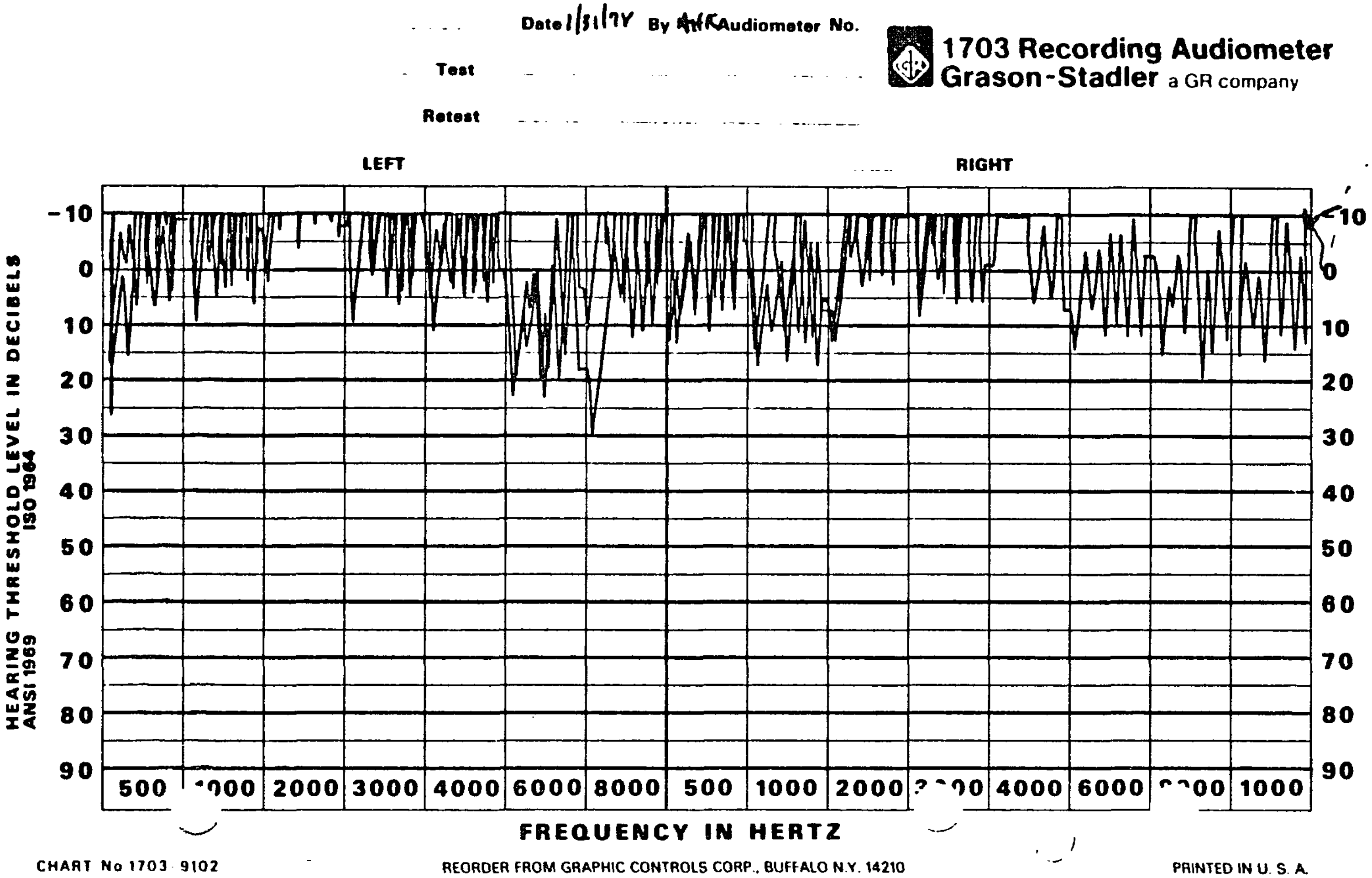


CHART No 1703 9102

REORDER FROM GRAPHIC CONTROLS CORP., BUFFALO N.Y. 14210

PRINTED IN U. S. A.

Subject No. 2


Date 1/31/74 By AHK Audiometer No.
Test
Retest
1703 Recording Audiometer
Grason-Stadler a GR company
LEFT
RIGHT
HEARING THRESHOLD LEVEL IN DECIBELS
ISO 1964
ANSI 1969
-10
0
10
20
30
40
50
60
70
80
90
500
1000
2000
3000
4000
6000
8000
500
1000
2000
4000
6000
1000
FREQUENCY IN HERTZ
CHART No 1703 9102
REORDER FROM GRAPHIC CONTROLS CORP., BUFFALO N.Y. 14210
PRINTED IN U. S. A.

Subject No. 3

Date 2/3/79 By A H Audiometer No.

Test

Retest

1703 Recording Audiometer
Grason-Stadler a GR company

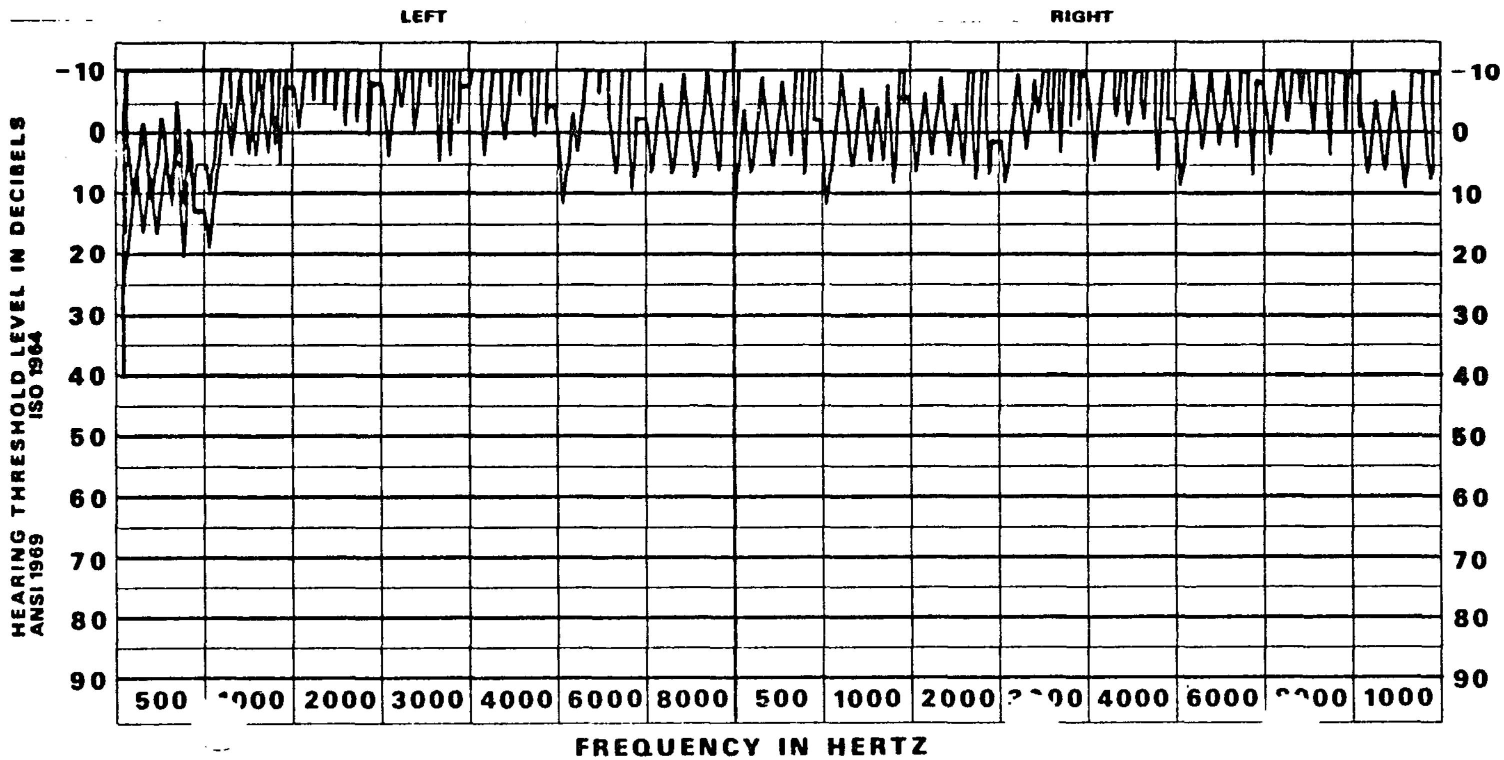
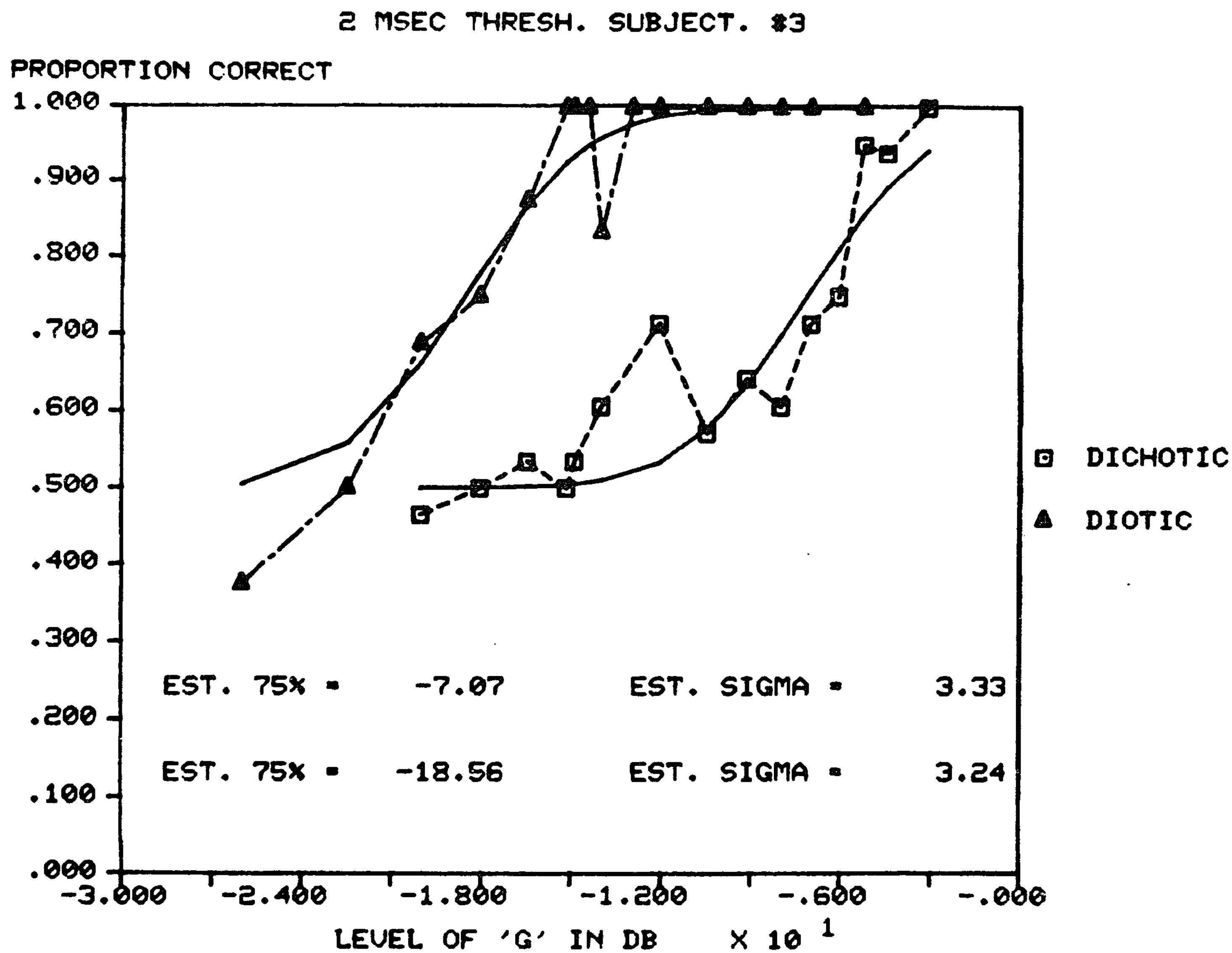


CHART No 1703 9102 REORDER FROM GRAPHIC CONTROLS CORP., BUFFALO N.Y. 14210 PRINTED IN U.S.A.

Subject No. 4

APPENDIX B



Sample Output of Optimization Routine