

This is to certify that the

dissertation entitled

All Things Coherence

presented by

Zachary Andrew Constan

has been accepted towards fulfillment
of the requirements for

Ph D degree in Physics


Major professor

Date 29 April 2002

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
 TO AVOID FINES return on or before date due.
 MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
JUL 05 2008 05 05 08	SEP 29 2008 04 21 09	
NOV 21 2006 08 13 07		

ALL THINGS COHERENCE

By

Zachary Andrew Constan

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Physics

2002

ABSTRACT

ALL THINGS COHERENCE

By

Zachary Andrew Constan

Interaural coherence is a crucial element of binaural hearing. However, the nature of real-world environments is to reduce coherence through reverberation (especially in rooms) and extraneous noise. It becomes necessary to understand coherence discrimination and how somewhat-coherent sounds influence the use of auditory cues.

Because coherence is closely related to interaural time differences (ITDs), coherence discrimination may diminish for frequencies above the 1500 Hz “limit” where sinusoidal ITD cues lose effectiveness. Listeners’ sensitivity to coherence in 100-Hz wide bands above and below that limit was tested. Some listeners’ thresholds dropped at high frequencies, while other listeners were consistent on both sides of the limit. These results were attributed to variable envelope coherence sensitivity. Listeners showed poorer coherence discrimination as the stimulus duration decreased below 100 μ s and when the stimulus level lowered from 64 to 34 dB SPL.

While it is possible for human listeners to lateralize high-frequency noise on the basis of ITDs in the envelope, the ability to do so makes great demands on the interaural coherence of the noise. Chapter III explores listeners' ability to lateralize a broadband, high-passed, coherent-noise signal in the presence of a broadband incoherent masker. Results showed that as the high-pass cutoff frequency increased through a critical region from 1 to 4 kHz, the required interaural coherence increased rapidly, especially for small

changes in lateral position. This can be predicted by a neural model of lateralization based on the centroids of bandwise cross-correlation functions of model peripheral inputs that have been rectified and low-pass filtered [e.g., Bernstein and Trahiotis, *J. Acoust. Soc. Am.* **100**, 1754–1763 (1996)].

Coherent noise produces a compact lateralized image, and incoherent noise (independent signals to left and right ears) produces a diffuse image that fills the head. Nevertheless, listening experiments showed that coherent and incoherent noise images seem to be equally affected by small interaural level differences (ILD). The ILD threshold for lateralizing incoherent noise is less than 0.5 dB greater than that for coherent noise. In this sense, the human binaural system appears to behave like an ideal level meter, insensitive to the waveform and envelope fine structures that determine coherence. The small discrepancy (less than 0.5 dB) can be understood from a standard model of loudness perception—incorporating critical band filtering, half-wave rectification, amplitude compression (0.6 power law), and temporal integration (300 ms).

A final goal was to discover if listeners were sensitive to the incoherence produced by head dispersion. The dispersion was modeled using Kuhn's (*JASA* **62**, 157-167, 1977) derivation for pressure on a spherical surface due to a plane wave, which led to synthetic head-related transfer functions. In a headphone experiment, listeners attempted to distinguish these artificial head-related dispersions from perfectly-coherent stimuli of constant ITD. The results showed that listeners could discriminate the two stimuli, but only on the basis of lateral position. Another experiment, designed to eliminate the lateral cue, found listeners unable to consistently identify the head-dispersed sound.

ACKNOWLEDGMENTS

The following tome would never have been possible without the help of innumerable people, but I'll try to thank them all anyway. Tim McCaskey, Scott Lawton and fellow graduate student Xinya (Peter) Zhang provided a lot of data in return for subjecting me to their experiments. Mary Jo Hidecker taught me all I needed to know about analysis of variance. Dr. Brad Rakerd provided countless hours of listening, advising, and reviewing...not to mention an awful lot of useful C code. Thanks to many, many listeners who sat in a dark, quiet, warm room and tried not to fall asleep. Most importantly, I'd like to give much of the credit for this dissertation to my advisor: Dr. William M. Hartmann. He put up with my shenanigans for six years, instructed me in the ways of psychoacoustics, listened, suggested, waited, and mentored. I have nothing but admiration for Dr. Hartmann, which is why I'm so glad he has shaped my academic career. I aspire to be the professor he always has been.

I'd also like to thank those who looked after me in my graduate career. First, the late-night homework support group: Wayne and Marguerite Tonjes, Andrew Schnepf, Brian Sharpee, Viktoria Greanya, Jim Armstrong, Jack Ryan. Without these folks, I'd never have made it past the first semester. Phil Klimkewicz, Ryan Humphrey, Bob Slater, Martin Eltzroth and Jason Greanya also kept me sane with many, many diversions. My family (Louis, Karen and Megan) is the rock I stand on to see possibilities for the future...thanks for letting me see such big dreams. Finally, to my wife Diane, who provided all the basic love, support, wisdom, knowledge and maintenance necessary to finish graduate school: take a bow, darling!

TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
Chapter I. An introduction to coherence, and its importance in human auditory localization	
Binaural time cues.....	1
Real-world sources and noise.....	2
Interaural cross-correlation.....	3
Coherence – Mathematical.....	5
Coherence – Perceptual.....	6
The foundation for correlation-based models of the auditory system.....	8
The significance of coherence.....	10
References.....	12
Figures.....	13
Chapter II. A study of decorrelation sensitivity as a function of frequency band, duration, intracranial position, and stimulus intensity	
Introduction	23
Experiment 1: Coherence discrimination at low and high frequencies.....	25
Experimental Procedure.....	25
Stimulus Generation.....	27
Sources of Error.....	30
Presentation and Listeners.....	34
Results.....	35
Discussion.....	37
Experiment 2: Broadband coherence discrimination at short durations.....	40
Introduction.....	40
Part A: Variable Duration.....	42
Part B: Variable Duration with randomized ILD.....	43
Part C: Variable Duration with zero, constant, and random ILD.....	46
Experiment 3: Level-dependent coherence discrimination.....	48
Introduction.....	48
Results.....	49
Conclusions.....	50
Further study.....	52
References.....	54
Figures.....	56
Chapter III. ITD discrimination of several stimuli	
Introduction	86
Experiment 1: ITD discrimination as a function of interaural coherence.....	87
Method-Interference experiment.....	87
No-interference experiment.....	89

Absolute Threshold experiment.....	89
Listeners.....	89
Results.....	90
Discussion-Absolute Threshold.....	91
Non-interference.....	92
Interference.....	92
Experiment 2: ITD discrimination as a function of high-pass cutoff.....	93
Introduction-the salience of high-frequency ITD cues.....	93
High-frequency ITD cues in the room environment.....	94
Method.....	95
Results and Discussion.....	97
Experiment 3: ITD discrimination of high-pass-filtered targets and maskers.....	101
Introduction.....	101
Method.....	103
Results.....	103
Discussion-Analysis with respect to known models.....	105
The Jeffress model.....	105
The auditory nerve model.....	106
The equalization-cancellation model.....	107
The position-variable model.....	107
The coherence-based model.....	108
Model calculation results.....	111
Model predictions.....	113
Conclusions.....	115
References.....	117
Figures.....	119

Chapter IV. Intensity-based lateralization using four stimuli of varying coherences

Introduction.....	159
Models	160
Methods.....	162
Experiment 1: Broadband signal.....	165
Results.....	165
Discussion.....	166
Experiment 2: Low-passed signal.....	169
Introduction and method.....	169
Results and discussion.....	170
Experiment 3: Low-passed signal with randomized levels.....	171
Introduction and method.....	171
Results and discussion.....	173
Summary/Discussion.....	174
The inclusion of DLI as a test of the level-meter model.....	174
Experimental inconsistency with the level-meter model.....	176
The evolution from level-meter to loudness-meter.....	177

Conclusions.....	178
Appendix A: “The strange case of Listener C”.....	179
Appendix B: Noise-type discrimination project.....	180
Introduction.....	180
Experiment 1.....	180
Experiment 2.....	182
Experiment 3.....	183
Conclusions.....	185
References.....	187
Figures.....	189

Chapter V. Discrimination of decorrelation due to frequency-dependent phase shifts, as calculated by a model of sound pressure on a spherical head

Introduction	205
The current study.....	208
Experiment 1: Discrimination at two incident angles.....	209
Method.....	209
Results.....	213
Discussion.....	214
Experiment 2: Randomized angle	216
Introduction.....	216
Method.....	216
Preliminary Experiment: “Equivalence of Position”.....	217
Experimental task, Results and Discussion.....	223
Experiment 3: “Wrong” phase shifts.....	226
Introduction.....	226
Method.....	227
Results and Discussion.....	228
Head-shift coherence, and delayed coherence discrimination.....	229
Introduction.....	229
Coherence calculations.....	230
Analysis of peak cross-correlation lag values.....	232
Discussion-lateral position.....	232
Discussion-correlation.....	234
Experiment 4: Off-axis coherence discrimination.....	235
Introduction and Method.....	235
Results and Discussion.....	236
The nature of incoherence.....	237
Summary and conclusion.....	238
Appendix A.....	240
References.....	241
Figures.....	243

GENERAL APPENDICES.....	281
Signal Detection Theory.....	282
Levitt Staircase Method.....	284
References.....	287
Figures.....	288

LIST OF TABLES

Table II.1. Interval composition.....	28
Table II.2. Bandwise coherence thresholds.....	36
Table II.3. Level-dependent coherence thresholds.....	50
Table III.1. Coherence thresholds by ITD and cutoff.....	98
Table III.2. Calibration settings.....	102
Table III.3. Results comparison.....	104
Table IV.1. Broadband ILD thresholds according to stimulus type.....	167
Table IV.2. Broadband pairwise comparisons.....	168
Table IV.3. Narrowband ILD thresholds according to stimulus type.....	170
Table IV.4. Narrowband pairwise comparisons.....	171
Table IV.5. Random-level ILD thresholds according to stimulus type.....	173
Table IV.6. Random-level pairwise comparisons.....	174
Table IV.7. Informal noise-type discrimination.....	181
Table IV.8. Formal noise-type discrimination.....	183
Table IV.9. Noise-switched discrimination.....	184
Table IV.10. Summary of noise-type experiments.....	185
Table V.1. Ideal lateralization results.....	218
Table V.2. Experiment 2 lateralization data.....	221
Table V.3. Experiment 2 discrimination data.....	225
Table V.4. Wrong-head lateralization data.....	228
Table V.5. Wrong-head discrimination data.....	228

Table V.6. Stimulus cross-correlations.....	230
Table V.7. CC peak magnitudes and lags.....	231
Table V.8. ITD limits by stimulus.....	232
Table V.9. Lag/ITD comparison.....	233

LIST OF FIGURES

Figure I.1. Schematic of a head in the presence of a sound source.....	14
Figure I.2. Four signal waveforms.....	16
Figure I.3. Representation of coherent and incoherent sources.....	18
Figure I.4. Representations of intracranial images.....	20
Figure I.5. Depiction of Jeffress' coincidence-counter model.....	22
Figure II.1. Bandlimits for narrowband stimuli used in this experiment.....	57
Figure II.2. Description of the experimental task.....	59
Figure II.3. Schematic of the experimental setup in the form of a patch diagram.....	61
Figure II.4. Estimation of coherence in the case of Programmable Attenuator (PA) inaccuracy	63
Figure II.5. Actual measurement of error in the Programmable Attenuator (PA).....	65
Figure II.6. Results of Experiment 1	67
Figure II.7. Calculated normalized correlations in a Masking Level Difference experiment	69
Figure II.8. The concept of the envelope.....	71
Figure II.9. Results of Experiment 2: variable duration.....	73
Figure II.10. Results of Experiment 2: variable duration with randomized ILD.....	75
Figure II.11. Results of Experiment 2: variable duration with zero ILD.....	77
Figure II.12. Results of Experiment 2: variable duration with constant ILD.....	79
Figure II.13. Results of Experiment 2: variable duration with random ILD.....	81
Figure II.14. Graphic representation of the “fuzzy ball copier analogy.”.....	83
Figure II.15. Results of Experiment 3: level-dependence.....	85

Figure III.1. Depiction of a human head and a sound source in a room.....	120
Figure III.2. Schematic of the experimental setup in the form of a patch diagram.....	122
Figure III.3. Description of the experimental task.....	124
Figure III.4. Results of Experiment 1: ITD discrimination as a function of coherence for listener B.....	126
Figure III.5. Results of Experiment 1: ITD discrimination as a function of coherence for listener L.....	128
Figure III.6. Results of Experiment 1: ITD discrimination as a function of coherence for listener S.....	130
Figure III.7. Results of Experiment 1: ITD discrimination as a function of coherence for listener T.....	132
Figure III.8. Results of Experiment 1: ITD discrimination as a function of coherence for listener W.....	134
Figure III.9. Results of Experiment 1: ITD discrimination as a function of coherence for listener Z.....	136
Figure III.10. Results of Experiment 1: ITD discrimination as a function of coherence averaged over all six listeners.....	138
Figure III.11. Spectral diagram of experimental stimulus for Experiment 2.....	140
Figure III.12. Results of Experiment 2: ITD discrimination as a function of high-pass cutoff.....	142
Figure III.13. Results of Experiment 3: ITD discrimination of high-pass-filtered targets and maskers for listener W.....	144
Figure III.14. Results of Experiment 3: ITD discrimination of high-pass-filtered targets and maskers for listener X.....	146
Figure III.15. Results of Experiment 3: ITD discrimination of high-pass-filtered targets and maskers for listener Z.....	148
Figure III.16. Results of Experiment 3: ITD discrimination of high-pass-filtered targets and maskers averaged over three listeners.....	150
Figure III.17. Simulating the actions of the auditory periphery in the model.....	152

Figure-III.18. Binaural processing in the model.....	154
Figure III.19. Model calculation results.....	156
Figure III.20. Model predictions.....	158
Figure IV.1. Depiction of a two-interval ILD trial in Experiment 1.....	190
Figure IV.2. Depiction of a two-interval DLI trial in Experiment 1.....	192
Figure IV.3. Results of Experiment 1: Broadband signal.....	194
Figure IV.4. Results of Experiment 2: Low-pass filtered signal.....	196
Figure IV.5. Depiction of a two-interval ILD trial in Experiment 3.....	198
Figure IV.6. Results of Experiment 2: Low-pass filtered signal with randomized standard level.....	200
Figure IV.7. ILD discrimination according to signal detection theory (TSD).....	202
Figure IV.8. Loudness-meter model predictions compared to results from Exps. 2 and 3	204
Figure V.1. Depiction of a human head and an incoming sound wave.....	244
Figure V.2. Head-related dispersion according to Kuhn's spherical head model.....	246
Figure V.3. Results of Experiment 1: Discrimination at 30 degrees.....	248
Figure V.4. Results of Experiment 1: Discrimination at 30 degrees.....	250
Figure V.5. Results of Experiment 1: Discrimination at 30 degrees.....	252
Figure V.6. Results of Experiment 1: Discrimination at 30 degrees.....	254
Figure V.7. Results of Experiment 1: Discrimination at 45 degrees.....	256
Figure V.8. Results of Experiment 1: Discrimination at 45 degrees.....	258
Figure V.9. Results of Experiment 1: Discrimination at 45 degrees.....	260
Figure V.10. Results of Experiment 1: Discrimination at 45 degrees.....	262
Figure V.11. Head-related dispersion according to Kuhn's spherical head model, and the inverse.....	264

Figure V.12. Cross-correlation of broadband model head-dispersed noise at a 45-degree angle.....	266
Figure V.13. Cross-correlation of narrowband model head-dispersed noise at a 45-degree angle.....	268
Figure V.14. Cross-correlation of narrow-narrowband model head-dispersed noise at a 45-degree angle.....	270
Figure V.15. Cross-correlation of broadband coherent noise at 520 μ s ITD.....	272
Figure V.16. Cross-correlation of narrowband coherent noise at 520 μ s ITD.....	274
Figure V.17. Cross-correlation of narrow-narrowband coherent noise at 520 μ s ITD...	276
Figure V.18. Cross-correlation of broadband “wrong-head” noise at a 45-degree angle	278
Figure V.19. Results of Experiment 4: Off-axis coherence discrimination.....	280
Figure A.1. Graphical representation of signal detection theory (TSD).....	289

Chapter I: An introduction to coherence, and its importance in auditory localization

Binaural time cues

When performing sound localization in the real world, humans rely on several aspects of the sound in order to accurately identify its origin. The fact that we have two ears greatly improves our chances, because the comparison between sounds incident on both of them (termed “binaural hearing”) is a powerful tool. Given an external sound source at some azimuthal angle to the head’s forward direction (Figure I.1), there is a difference in path lengths to the ears. Assuming that the speed of sound is constant, it is clear that the sound will be incident on one ear before the other. This factor is called Interaural Time Difference (ITD). It provides one of the most critical cues for localization. The ITD is approximately equal to

$$\Delta t = \frac{3a}{c} \sin \theta$$

where a is the radius of the head (about 8.75 cm), and c is the speed of sound in air (343 m/s). This solution is predicted by a diffraction formula in the limit of low frequencies, and establishes a relationship where a particular time delay corresponds directly to an azimuthal angle. For purposes of scale, the maximum ITD expected (given a signal located on the extreme right or left) is about 800 μ s. The auditory system can use the detected ITD to determine the direction of (localize) the sound source, at least in the horizontal plane.

Real-world sources and noise

Accurate sound localization is impaired by several factors, not the least of which is the presence of competing sounds. Listeners often find it necessary to localize a single source of interest while trying to filter out the “clutter” of many extraneous sources or reflections. Yost (1997) gives a review of the related “cocktail party problem,” which simply asks how humans are able to identify and understand one conversation in a room where dozens of people are speaking at once. This environment involves a target signal (human voice) that must be comprehended in the presence of a masking noise (crowd).

Imagine that the target source is off to the listener’s right, and that its signal therefore leads in the right ear and lags in the left by some delay (top two panels of Figure I.2). The signals at left and right ears are identical aside from the delay, because they come from a single source. The listener can easily identify similar features in both. By measuring the time between corresponding peaks, the listener can establish an ITD and determine the direction of the target. The listener also has no trouble distinguishing characteristics of the target such as its nature as speech.

The bottom panels of Figure I.2 show the same target signal after an independent masking noise has been added to each ear. This noise may be constructed of multiple signals with different ITDs, as in the case of a talking crowd. The masker “muddies” the target, making it less distinguishable and localizable. The left and right signals are no longer identical, and common features are harder to detect. In the case of Figure I.2, the masker is of low amplitude with respect to the target, and so the target is still somewhat

recognizable. As the noise increases, there comes a point at which the target is “drowned out” altogether.

Interaural cross-correlation

The degree of similarity between left- and right-ear signals is referred to as interaural “correlation,” and measured quantitatively by the cross-correlation (CC) function. Plainly stated, the CC tests how close the relationship of two signals is across a range of interaural delays.

The mathematical definition of CC is described by the following equation, where x_L and x_R are simply the waveforms incident on left and right ears:

$$\gamma(\tau) = \frac{\int x_L(t)x_R(t+\tau)dt}{\sqrt{\int x_L^2(t_1)dt_1 \int x_R^2(t_2)dt_2}} \quad (1)$$

The symbol γ represents a measure of cross-correlation, which is a function of interaural delay time τ . The numerator is a time integral of the signal in the left ear multiplied by the signal in the right. The denominator is the square root of the energy in the left ear multiplied by the energy in the right.

To better understand the form of the CC function, let the reader concentrate on the numerator of the above equation, in essence calculating an unnormalized cross-correlation:

$$\gamma(\tau) = \int x_L(t)x_R(t+\tau)dt$$

By Fourier-transforming the signals, x_L and x_R are broken down into sums of cosines:

$$= \int dt \sum_n C_{nL} \cos(\omega_n t + \varphi_{nL}) \sum_m C_{mR} \cos[\omega_m(t+\tau) + \varphi_{mL} + \Delta\varphi_m]$$

Using the trigonometric identity for $\cos(A+B)$:

$$= \int dt \sum_n C_{nL} \cos(\omega_n t + \varphi_{nL}) \sum_m C_{mR} [\cos(\omega_m t + \varphi_{mL}) \cos(\omega_m \tau + \Delta\varphi_m) - \sin(\omega_m t + \varphi_{mL}) \sin(\omega_m \tau + \Delta\varphi_m)]$$

And distributing the multiplier:

$$= \sum_m C_{mR} \sum_n C_{nL} \int dt [\cos(\omega_n t + \varphi_{nL}) \cos(\omega_m t + \varphi_{mL}) \cos(\omega_m \tau + \Delta\varphi_m) - \cos(\omega_n t + \varphi_{nL}) \sin(\omega_m t + \varphi_{mL}) \sin(\omega_m \tau + \Delta\varphi_m)]$$

Performing the time integral within the limits $0 < t < T$, we can use orthogonality of cosines to sines and to cosines of different frequencies to collapse the sums into one. By returning the normalizing denominator, the equation is made complete:

$$\gamma(\tau) = \frac{\sum_n C_{nL} C_{nR} \cos(\omega_n \tau + \Delta\varphi_n)}{\sqrt{\sum_{n1} C_{n1L}^2 \sum_{n2} C_{n2R}^2}}$$

The value $\Delta\varphi_n$ represents the difference between right and left phases. Thus, the normalized CC function is simply constructed of cosines featuring various frequencies, phase differences, and amplitudes.

To improve understanding of the measure of cross-correlation, the following analysis involves a signal that is exactly the same in both ears, with all frequencies having a time delay of zero. Because $x_L(t) = x_R(t + 0)$ the numerator of equation 1 becomes a power integral at $\tau = 0$, and the two powers in the denominator become equal, yielding the power squared. Thus, the cross-correlation goes to one at $\tau = 0$, then falls off as a sum of cosines, indicating that the two signals are perfectly correlated given no delay (also known as an “autocorrelation” function). However, if the signals from the two ears are totally unrelated, having no time delays in common (uncorrelated), the integral in the

numerator goes to zero, and the correlation is zero everywhere. This is how we assign numerical values to correlation.

Coherence – Mathematical

The CC function, while an interesting and useful representation of the relationship between two signals, is not exactly concise. A simpler way to characterize the similarity of left and right channels is the coherence measurement, which is typically established by the peak value of the CC function. A mathematically rigorous definition begins with the CC equation. Let left and right signals contain a target A with some ITD τ_0 (as the top half of Figure I.2), as well as two masking noises, B and C respectively, that are mutually independent and uncorrelated with A (as added to the bottom half of Figure I.2). To simplify the derivation, let B and C also be equal in intensity. Thus,

$$\gamma(\tau) = \frac{\int x_L(t)x_R(t+\tau)dt}{\sqrt{\int x_L^2(t_1)dt_1 \int x_R^2(t_2)dt_2}} = \frac{\int [A(t)+B(t)][A(t-\tau_0+\tau)+C(t+\tau)]dt}{\sqrt{\int [A(t_1)+B(t_1)]^2 dt_1 \int [A(t_2)+C(t_2)]^2 dt_2}}$$

The peak value will occur when τ is equal to τ_0 . The time integral of independent functions goes to zero, so we can simplify the equation:

$$\text{Coherence} = \gamma(\tau_0) = \frac{\int [A^2(t)]dt}{\sqrt{\int [A^2(t_1)+B^2(t_1)]dt_1 \int [A^2(t_2)+C^2(t_2)]dt_2}}$$

This equation deals with the powers in A , B and C integrated over time. Because the two masking noises have equal intensities,

$$\int B^2(t)dt = \int C^2(t)dt$$

Let the time integral of X^2 (the total power in X) be symbolized by P_X :

$$\text{Coherence}(C) = \frac{P_A}{\sqrt{[P_A + P_B][P_A + P_B]}} = \frac{P_A}{P_A + P_B}$$

Therefore, coherence is simply the ratio of the correlated power in both ears to the total power. If the masker is absent ($P_B = P_C = 0$, resulting in the conditions on the top half of Figure I.2) then the coherence equals one, representing 100% correlation between the ears. This is a perfectly coherent sound. If the target is absent, then the coherence equals zero—a completely incoherent sound. The range between these limits represents various combinations of target and masker, and provides a straightforward and quantitative way to describe the relationship of binaural signals.

Coherence – Perceptual

Certainly, the ability to identify an ITD is dependent on the signal coherence, but how does coherence relate to ITD? To understand how the binaural system might use coherence, one must account for the frequency analysis done by the auditory periphery. The incoming signal is transformed from the time domain to the frequency domain, and thus, the measurement of coherence is important as a function of frequency bands.

For an illustration of coherence in terms of frequency and ITD, see Figure I.3. The two plots graph the frequencies (y-axis) that make up the signal by their interaural delay (x-axis). In the upper figure, the line labeled “coherent source” contains a range of frequencies with a common delay (in this case, $-667 \mu\text{s}$). The small breaks in the line indicate that the 1800-Hz-wide source is separated into several frequency bands in the auditory system, which then measures coherence within those bands. Frequency analysis into bands occurs at the auditory periphery (Pickles, 1982), and thus, central processes

like coherence perception must evaluate the signal within those bands. The common delay indicates that these frequencies issued from one direction, probably from the same source. Because they share one delay, the listener will have no trouble identifying the ITD and localizing the source. A CC calculation of this band in left and right ears peaks at one, indicating perfect coherence. The “incoherent source” in the lower figure, however, has randomly-distributed delays across frequency. Each open circle represents the delay for a particular frequency. The dotted curves on either side identify delays of half a period for each frequency, beyond which phase shifts indicate an image on the opposite side. Because different frequencies have different delays, the peak CC over any critical band is likely close to zero, and the sound has no identifiable direction. These frequencies do not produce a localizable image because they seem to come from many angles. The extent to which ITDs are consistent over a band of frequencies is called “straightness,” a factor which may have some significance in auditory localization (Buell and Trahiotis, 1997).

Coherence is intimately connected to one’s ability to localize a sound source. Consequently, it influences one’s perception of the sound as well. In the case of a sound produced by headphones, the binaural system places an internalized image of that sound along an axis in the listener’s head, much like the τ axis of a CC function, or the time delay axis of Figure I.3. If the source arrives at the left ear first, the internal image is pulled toward the left ear. Detection along this intracranial axis is called lateralization. The image location is in no small way determined by the delay at which the CC function peaks.

In order to better understand these images, and how they correspond to the two sources in Figure I.3, refer to Figure I.4. In the case of a coherent source, since all the frequencies are located at one time delay, they produce a single, compact image at one position along the lateralization axis. The clearly defined ball on the left represents this image. For an incoherent band, the time delays are spread out randomly, and thus appear at various lateral locations. They produce a fuzzy, head-filling image as depicted on the right. The middle image is a compromise between the two, the internal representation of coherence between zero and one.

The foundation for correlation-based models of the auditory system

For most real-world stimuli, the sound source is composed of a band of frequencies as depicted in Figure I.3. As mentioned above, once the signal is incident on the ears, the first act of the auditory system is to analyze it into component frequencies (Pickles, 1982). Any model of the auditory system must take that fact into account from the beginning. That approach is fundamental to correlation/coherence-based models that stem from the work of Jeffress (1948), who suggested a possible method of binaural interaction illustrated in Figure I.5 (from Hartmann, 1999). The bottom axis represents internal lag, while the vertical axes delineate channels where the frequency-analyzed auditory signal travels from the two ears. Each cluster of five lines represents a channel that receives only one characteristic frequency (CF) from the auditory input. The black dots where the vertical arrows originate symbolize actual neurons, which serve as “coincidence detectors”. If they receive simultaneous signals (within a time window)

from both sides, they will fire, signifying a coincidence of neural spikes in the left and right channels.

The purpose of the coincidence detectors is simple: take, for example, the detectors on the far left of the time axis. If they fire, the binaural system recognizes that the signal reached the right ear long before the left, allowing it time to travel a longer path to that neuron and still coincide with the signal from the left ear. Therefore, the source must be located towards the extreme right. The various neurons in each CF channel represent many different combinations of signal path lengths, and thus, ITDs. In this manner, Jeffress' construct uses a neural "place" mechanism to infer external ITDs from the activity of detectors at known internal delay (Stern and Trahiotis, 1995). Those detectors with higher firing rates simply indicate more coincidence for their characteristic time delay.

Recent physiological studies have revealed the existence of cells in the superior olivary complex and inferior colliculus that might play just such a part in binaural processing. Jeffress' model of coincidence counters may have a factual basis in ITD-sensitive cells found in the medial superior olive (Yin and Chan, 1990). The cells are particularly sensitive to a specific interaural delay, and investigators have identified a distribution of ITD-sensitive cells that includes all physically plausible (and some implausible) free-field delays. This evidence lends credence to the concept of coincidence detectors with a characteristic interaural delay.

Going back to Figure I.3, one can reinterpret the graph as a simpler view of Jeffress' coincidence matrix. Once again, it is a plot of frequency versus time delay. The points at each frequency that make up the coherent and incoherent bands could represent

the highest firing rate among coincidence detectors tuned to that particular CF. The plot now represents neural activity across internal delays for a selection of frequencies available to the auditory system. The interpretation of coherent and incoherent bands is unchanged: while the coincidence peaks have a common delay over CF, the band is coherent, and its ITD readily established. Interestingly, the pattern of coincidence response among detectors for a single CF can approximate the short-term CC function of the input signal at that frequency.

The proposed matrix of coincidence counters separated by internal delays forms the cornerstone of many auditory models. It is the primary binaural analyzer in the auditory-nerve-based model of Colburn (1977), the position-variable model of Stern and Colburn (1978), and the modern correlation-based model of Bernstein and Trahiotis (1996) to name a few. As such, the concepts outlined above are of great importance to the research reported hereafter, which deals with the effect of coherence on the use of auditory cues and as a cue itself.

The significance of coherence

What motivates this study of coherence? There are many answers, which tend to center around coherence's critical role in localization. When the listener is presented with incoherent sound, he or she cannot establish an interaural time delay because there are no comparable features in left- and right-ear signals. The reduction of coherence in real-world situations can be a daunting obstacle to communication. For instance, referring back to Figure I.1, note that l and l' are only two possible paths that the sound

can take to the ears. In a room, reflections of the walls can arrive shortly after the direct sound, providing a different directional cue and reducing coherence, which serves to confuse the binaural system. Also note that the head is not transparent to sound, and for the listener in the figure, the left ear will actually receive the signal that has dispersed around the head. This also lowers coherence, and may make localization more difficult.

However, in the field of sound reproduction, the introduction of incoherence is actually desirable. Since the listener wants to experience a realistic reproduction of the concert or theater environment, sound engineers must ensure that the illusion is not ruined by easily-localized sound at a speaker. By lowering the coherence of the signal (with multiple speakers and added reverberations), they produce a sound that seems to come from many directions, enveloping the listener. Finally, a key factor inspiring the study of coherence in the binaural system is that such study provides the experimenters with insight on how this part of the auditory system works.

References

- Bernstein, L.R. and Trahiotis, C. "The normalized correlation: Accounting for binaural detection across center frequency," *J. Acoust. Soc. Am.* **100**, 3774-3784 (1996).
- Buell, T.N and Trahiotis, C. "Recent Experiments Concerning the Relative Potency and Interaction of Interaural Cues," *Binaural and Spatial Hearing in Real and Virtual Environments*, ed. Gilkey and Anderson, Lawrence Erlbaum Associates, Mahwah, NJ, p. 139-149 (1997).
- Colburn, H.S. "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," *J. Acoust. Soc. Am.* **61**, 525-533 (1977).
- Hartmann, W.M. "How we localize sound," *Phys. Today*, November 1999.
- Kuhn, G.F. "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.* **62**, 157-167 (1977).
- Jeffress, L.A. "A place theory of sound localization," *J. Comp. Phys. Psych.* **41**, 35-39 (1948).
- Pickles, J.O. *An Introduction to the Physiology of Hearing*. Academic Press, New York, 1982. p. 24-70
- Stern, R.M. Jr. and Colburn, H.S. "Theory of Binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position," *J. Acoust. Soc. Am.* **64**, 127-140 (1978).
- Stern, R.M. and Trahiotis, C. "Models of Binaural Interaction," *Hearing*, ed. B. Moore, Academic Press, New York, p. 347-386 (1995).
- Yin, T.C.T. and Chan, J.C.K. "Interaural Time Sensitivity in Medial Superior Olive of Cat," *J. Neurophys.* **64**, p. 465-488 (1990).
- Yost, W.A. "The Cocktail Party Problem: Forty Years Later," *Binaural and Spatial Hearing in Real and Virtual Environments*, ed. Gilkey and Anderson, Lawrence Erlbaum Associates, Mahwah, NJ, p. 329-347 (1997).

Figure I.1. Schematic of a head in the presence of a sound source. a is the radius of the head, θ is the azimuthal angle to the source as measured from the forward direction, l and l' are the distances from the source to the left and right ears, respectively.

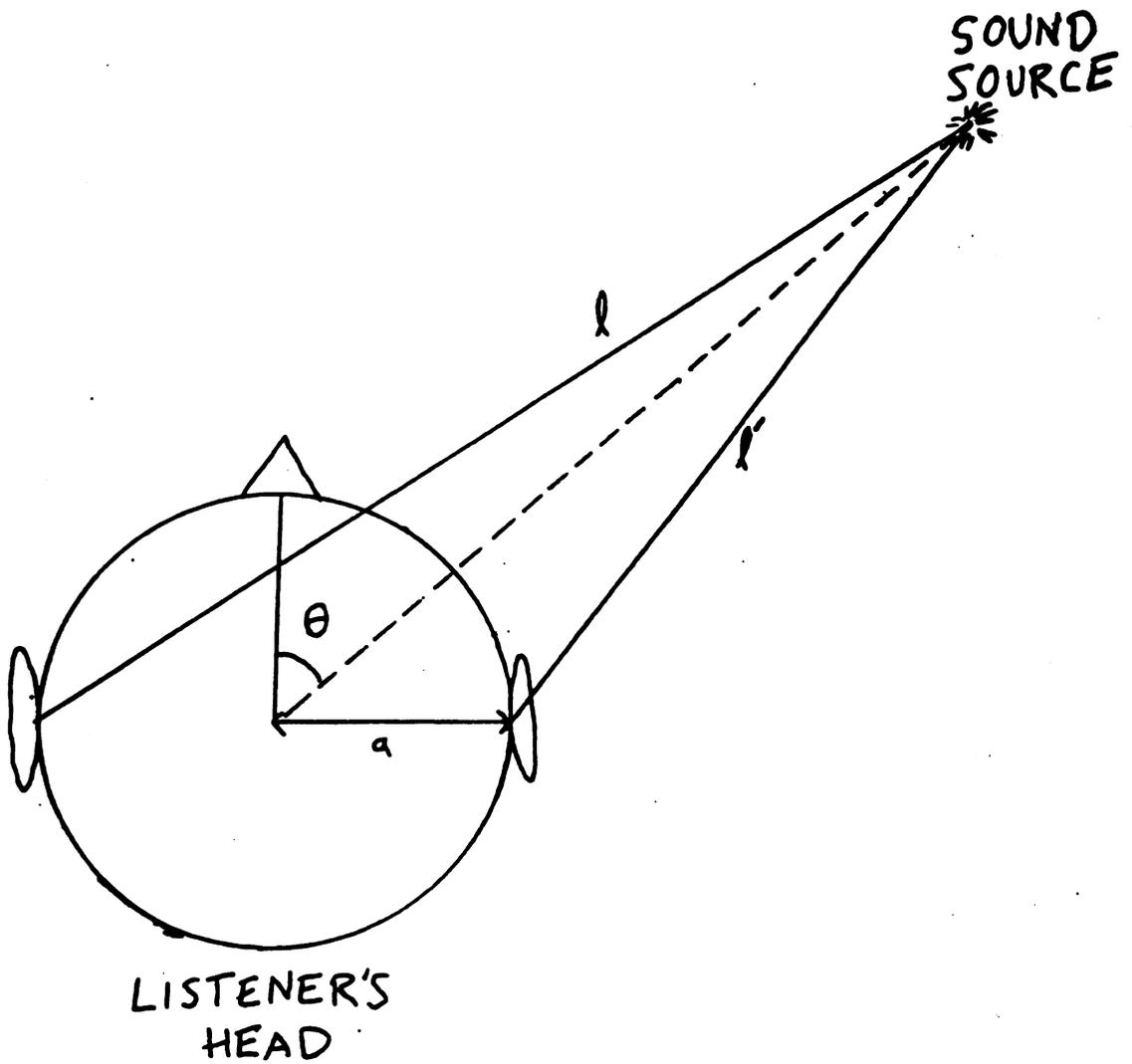


Figure I.1

Figure I.2. Four signal waveforms. The top two plots represent a 100-percent coherent signal in left and right ears ($x_L(t)$ and $x_R(t)$, respectively) with an interaural time delay imposed so that the right signal leads the left. The bottom two plots include the same coherent signal as well as a small amount of independent noise added to each ear, reducing the coherence to 90 percent. The arrows in top and bottom left-ear waveforms show the delay, where the left signal waveforms correspond to the earliest time in the right signal.

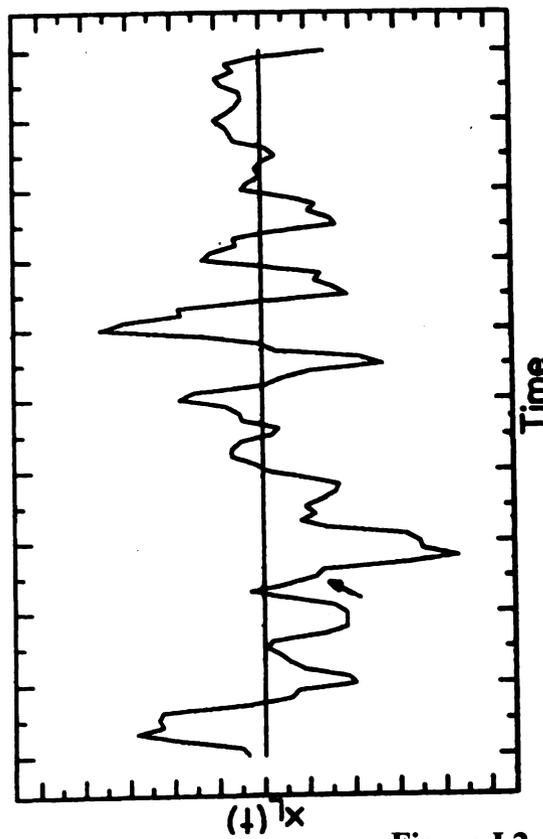
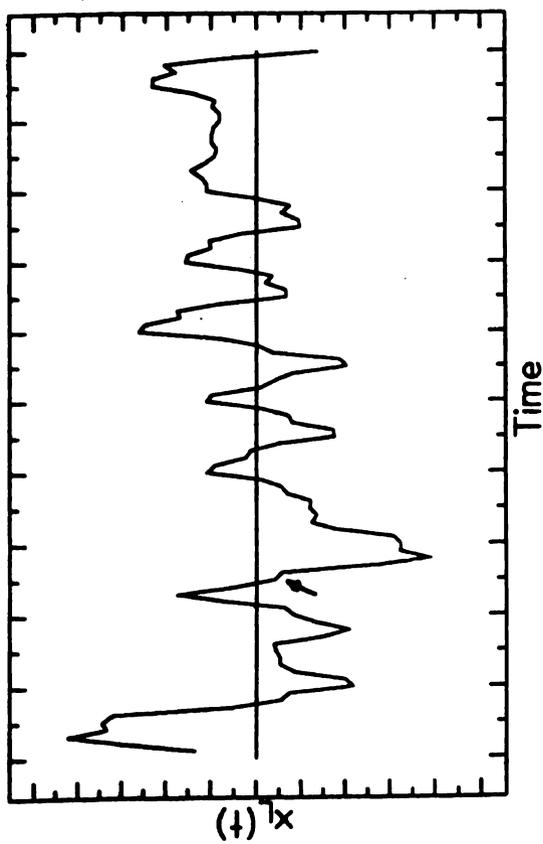
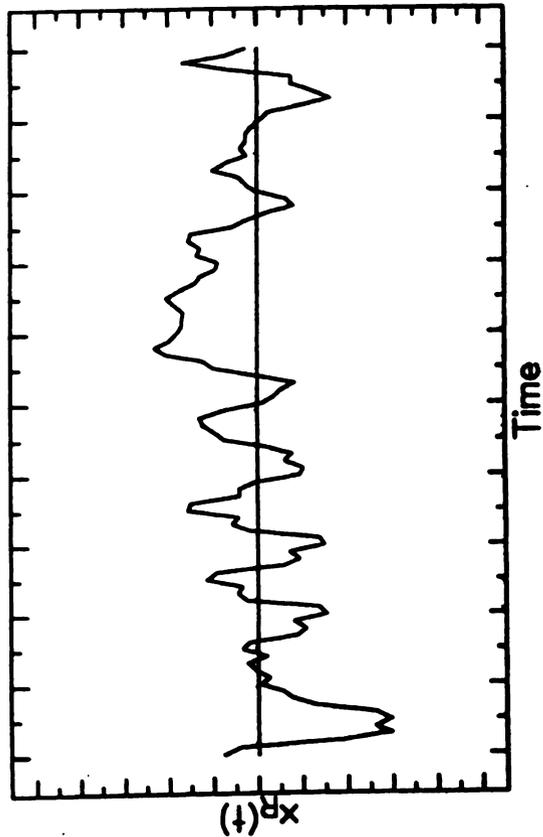
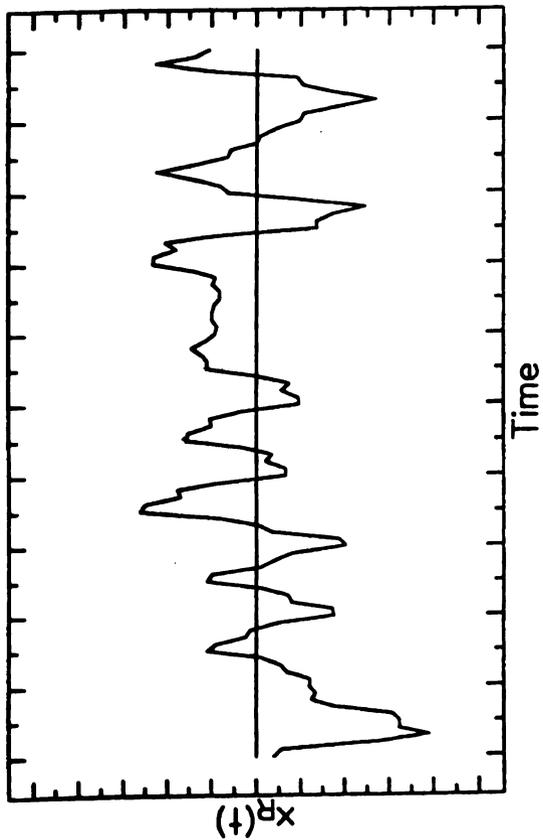


Figure I.2

Figure I.3. Representation of coherent and incoherent sources. In both plots, the y-axes are frequency scales, while the x-axes plot the interaural time differences (ITD) of a signal according to frequency. The upper plot contains a coherent source including frequencies from 200 to 2000 Hz, characterized by a constant ITD within frequency bands (indicated by breaks in the line). The lower plot shows an incoherent source composed of frequencies with random ITD in the same bandwidth. The dotted lines constrain the range of measurable ITDs because they mark delays of one-half period.

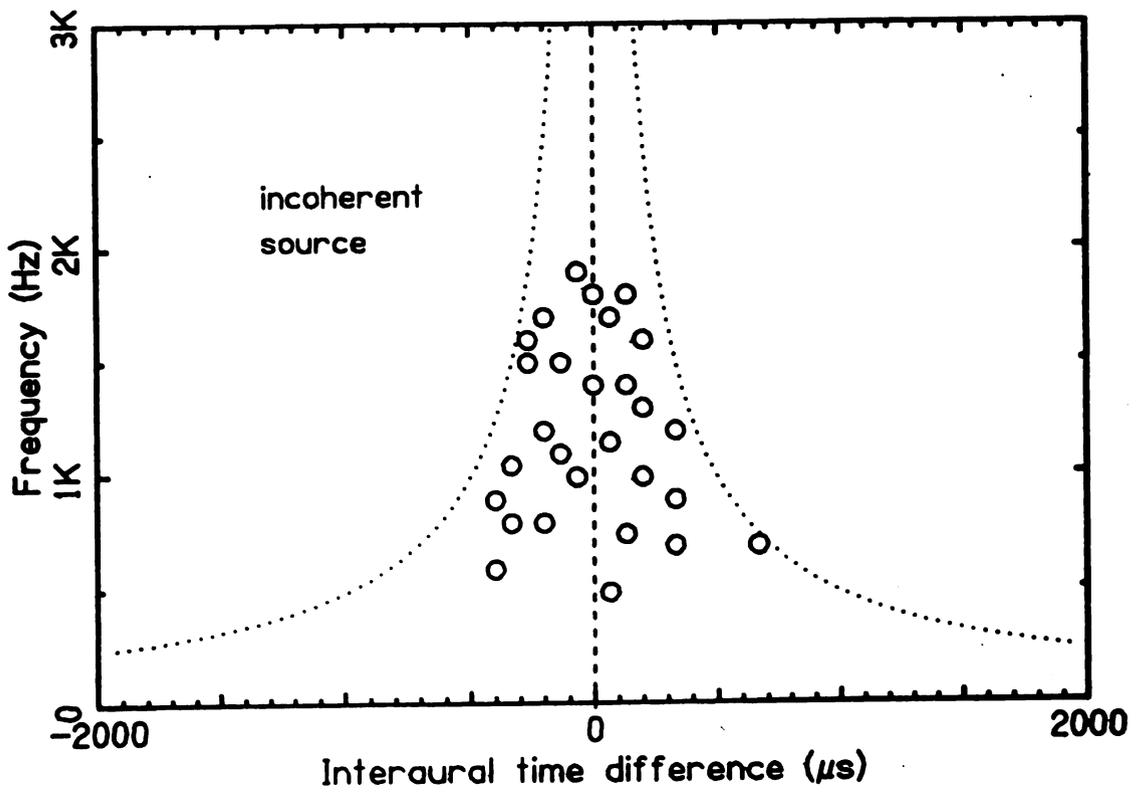
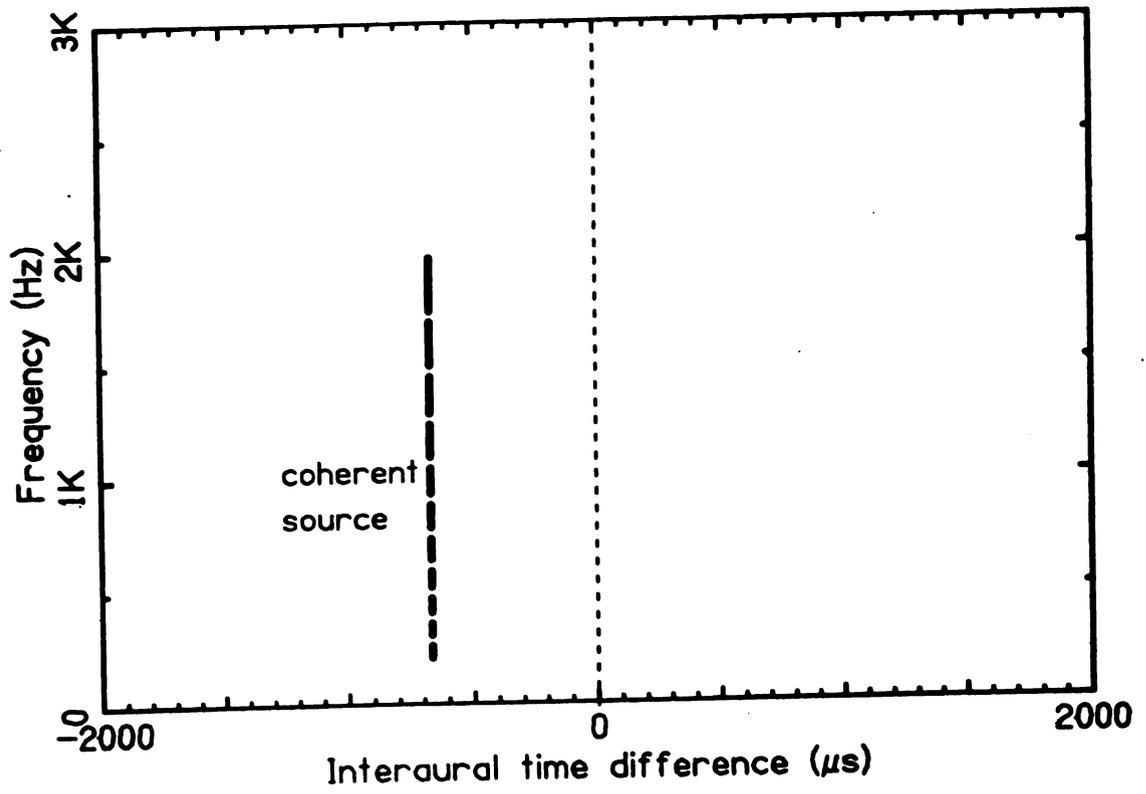


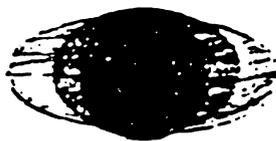
Figure I.3

Figure I.4. Representations of intracranial images. A coherent image (leftmost) is compact and easy to localize. An incoherent image (rightmost) is broad and indistinct, and has no definite intracranial location. A slightly-incoherent image (center) is broader than a coherent image, but more defined than an incoherent one.

**Coherent
Image**



**Slightly-incoherent
Image**



**Incoherent
Image**

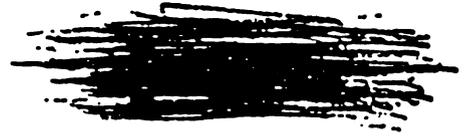


Figure I.4

Figure I.5. Depiction of Jeffress' coincidence-counter model. Each horizontal band of five lines represents a characteristic frequency, and the x-axis spans interaural lags of ± 3 ms. The dots at varying lags are coincidence counters, and their outputs are the vertical arrows that presumably transfer information to the binaural processor. (From Hartmann, 1999)

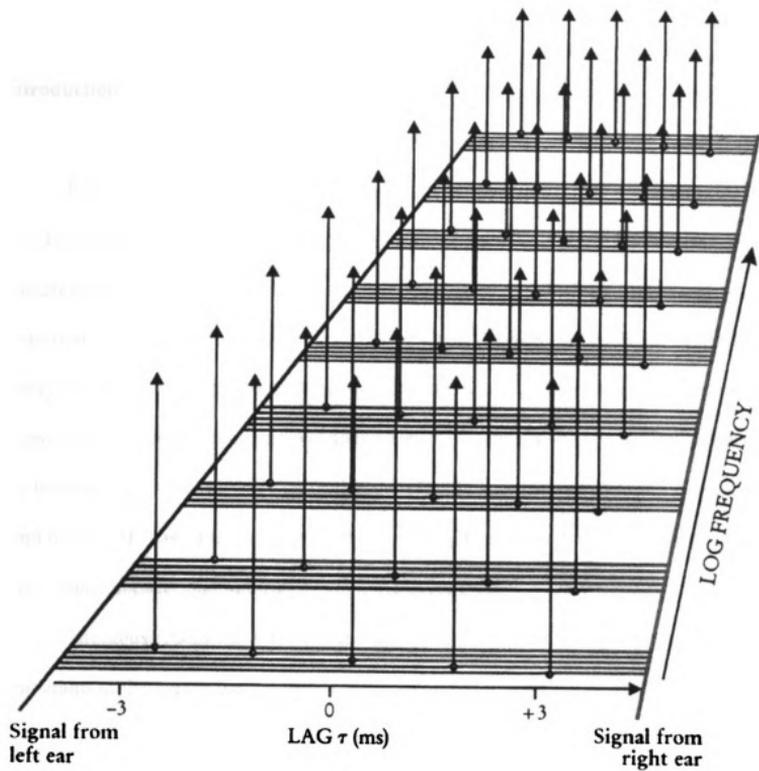


Figure I.5

Chapter II: A study of coherence sensitivity as a function of frequency band, duration, intracranial position, and stimulus intensity

Introduction

It is well-known that the binaural system can use interaural time difference (ITD) cues in a signal waveform for frequencies up to about 1500 Hz, but not above. This limitation may be due to neural firing rates not keeping up with high frequencies, so the spike trains lose “phase lock” with the incoming signal (Pickles, 1982). In his “duplex” theory, Rayleigh (1907) attributed sinusoid localization to ITD cues for all frequencies below 1500 Hz and to interaural level difference (ILD) cues above. His generalization of the localization process has held up well in decades of experiments to follow. The upper limit of 1500 Hz has appeared for many binaural effects, such as the lateralization of sine tones, binaural beats, and Masking Level Differences (MLD).

The study reported in this chapter is one of many currently exploring the limitations of the duplex theory. While the theory holds true for sinusoids, Klumpp and Eady (1956) have shown that listeners are sensitive to ITDs of high-frequency complex waveforms. Bernstein and Trahiotis (1995) discovered that for narrowband (100-Hz wide) noise presented at a center frequency of 4 kHz, threshold ITD sensitivity for listeners was around 250 μ s. While this threshold is better than no ITD sensitivity at all, it represents a large percentage of the total internal delay that the auditory system can detect, and is a very poor resolution compared to previously reported thresholds of 10 μ s for low-frequency tones (Durlach and Colburn, 1978). Since coherence discrimination is

inherently dependent on ITD distribution across frequencies (Trahiotis *et al.*, 2001), one may conclude that it should be significantly impaired for frequencies above 1500 Hz. As a test of this hypothesis, the study reported here examined listeners' coherence discrimination for 100-Hz-wide noise bands that were either entirely above or below this limit, as in Figure II.1. If the above assumption was correct, listeners should perform well in the lower bands, and poorly in the higher ones.

In order to actually measure coherence discrimination, an experiment was devised to test whether a subject could distinguish between a perfectly coherent stimulus and a slightly incoherent one. This kind of test finds a “just noticeable difference (jnd)” from a reference coherence of one. To generate this slightly-incoherent stimulus, a small amount of independent noise was added to the coherent noise. In terms of intracranial perception, the coherent noise produces a focused image of small auditory source width (ASW), while the addition of independent noise effectively blurs the edges to make the image a “fuzzy ball” with increased ASW (Figure I.3). This increase in ASW could relate to the standard deviation of probability density along the ITD axis (see General Appendix B on Signal Detection Theory). Listeners were then asked to identify which of the two stimuli was the slightly incoherent one to test their ability to discriminate coherence. This task may be compared to the well-known Masking Level Difference (MLD) experiments (Durlach and Colburn, 1978) that require listeners to detect a tonal signal in broadband noise. The name of these experiments comes from the fact that when the signal and masker have different interaural phases (for instance, the signal is shifted 180 degrees ($S\pi$) while the masker is diotic (No)), listeners can detect the signal at much lower levels than if the two had identical phase relationships. The addition of an out-of-phase signal

corresponds to a small drop in coherence, and some models indicate that NoS π MLDs are directly related to coherence discrimination (Jain *et. al.*, 1991). This interpretation of MLD suggests that it is easy for listeners to distinguish a very small change from a coherence of one (Robinson and Jeffress, 1963), and results show that the task becomes more difficult as the signal frequency increases to 1500 Hz due to the declining usefulness of ITD cues. The present results are compared to those of several MLD researchers at many points in this chapter, and the research takes hints from their discoveries.

Experiment 1: Coherence discrimination at low and high frequencies

Experimental procedure

As described in the introduction, the initial goal was to test coherence discrimination of 100-Hz bands both above and below the 1500-Hz cutoff for fine-structure ITD cues. The center frequencies of the narrow bands tested were 200, 500, 1000, 2000 and 4000 Hz—three below and two above the cutoff. See Figure II.1 for a frequency-scale representation of those bands. The experimental program also presented runs with a broadband stimulus from 20-6000 Hz, which included all the narrow bands within its limits. Listeners were expected to produce thresholds for the broadband stimulus that were at least as good as for the band that gave their best result, since the broadband stimulus contained the same auditory information in that band as well as all other bands. This is, of course, assuming that more coherence data allows the binaural processor to make finer distinctions—however, results from Gabriel and Colburn (1981)

actually indicate that coherence discrimination degrades as bandwidth increases. This result is non-intuitive, and the present study is intended to provide more information on the subject.

Subjects attempted to identify the less coherent interval in a two-interval, forced choice trial by pressing the appropriate button on a two-button response box. The intervals were 500 ms in duration, at 64 dB Sound Pressure Level (SPL). See Figure II.2 for a schematic of two possible trials. The coherent and slightly incoherent intervals could occur in two orders: if the slightly incoherent interval (represented by the fuzzy ball) were presented first, the left response button would be correct, and if it were second, then the reverse was true. The “fuzziness” of the signal was determined by the amount of independent noise introduced to reduce coherence, which was in turn controlled by a 3-down, 1-up staircase (Levitt, 1971) (see General Appendix A). If the listener answered correctly three times in a row, the program lowered the level of the independent noise, making the two intervals more alike and the task more difficult. However, if the listener answered incorrectly, the program intensified the independent noise, making the slightly incoherent interval less coherent and easier to distinguish. Each time the listener changed direction on the staircase, the “turnaround” intensity was recorded, and the run went on for 14 turnarounds. The first four turnaround intensities were discarded, and the other ten were averaged to establish the 79% correct level (defined as the threshold) at that condition. Listeners performed training runs on each task until their performance was deemed consistent, and then data were recorded for the next four runs. For each condition, the mean and (n-1) standard deviation of the values determined by those four runs formed the results.

Stimulus generation

The setup (in the form of a patch diagram) designed to produce the stimuli can be found in Figure II.3. Beginning at the left, noises A and B are generated from two channels of the digital noise generator. A and B have identical levels and band limits, but are completely independent. Because they are constructed in the frequency domain before transformation into time, both noises have infinitely sharp bandlimits. They enter separate low-pass filters (cutoff frequency = 8.2 kHz, rolloff = -115 dB/octave) to eliminate components above half the sample rate (these spurious components are a consequence of the digital-to-analog conversion), after which A is split into two branches. The top branch feeds directly into an exponential voltage-controlled amplifier (EVCA) that is ramped on and off by a trapezoidal envelope generator, and then sent to the listener's left ear. The lower branch of noise A is connected to a Programmable Attenuator (PA), as is B. The PAs lower the level of the input signals, scaling by factors designated as "h" and "g" so that noise A comes out as hA , and B as gB . Then the two are mixed to get signal $hA+gB$, which feeds through another EVCA into the listener's right ear.

The coherent and slightly incoherent intervals are generated as shown below in Table II.1. To achieve a coherent stimulus, one would set $h = 1$ and $g = 0$, and therefore the signals to both ears are identical. When adding some independent noise (B) to lower the coherence from one, the settings change to $h < 1$ and $g > 0$.

Table II.1 - Interval composition			
First Interval (100% Coherent)		Second Interval (<100% Coherent)	
Left Ear	Right Ear	Left Ear	Right Ear
A	1.0 A + 0.0 B	A	<i>h</i> A + <i>g</i> B

In order to prevent any interaural level differences (ILDs), the mixture of A and B must be properly weighted to match the level of A in the left ear. It is imperative that the ILD be the same (approximately zero) for both coherent and slightly-incoherent stimuli, or the listeners may be able to use it as a cue. Specifically, if the mixture of A and B was **always** quieter than A alone, the listener could learn to do the task by identifying which **interval** provided the lower level to the right ear. To properly test coherence **discrimination**, one must eliminate all other cues.

In the case of a coherent interval, the settings are simple: no attenuation on A and **maximum** attenuation on B (as shown in Table II.1). However, when the interval is **meant** to have a coherence slightly less than one, the settings require some calculation. The **powers** in each ear are as follows, with *A* and *B* representing the two independent **noise** waveforms, and the horizontal bars above them indicate the average value over **time**:

$$P_L = \overline{A^2}$$

$$P_R = \overline{(hA + gB)^2} = \overline{(hA)^2} + \overline{(gB)^2}$$

There is no cross term of *A* and *B* in the right ear power because they are **completely** independent waveforms, and so when multiplied and averaged, this term does **not** **contribute** to the power. By equating the left and right powers, and noting that the

power in noise A equals that of noise B , one can then solve for a relationship between factors h and g :

$$P_L = P_R \Rightarrow \overline{A^2} = \overline{(hA)^2} + \overline{(gB)^2} = (h^2 + g^2) \overline{A^2}$$

$$h^2 + g^2 = 1 \rightarrow h = \sqrt{1 - g^2}$$
(1)

Now, for any scale factor g chosen to impose on the independent noise (thus, determining how much incoherence will be introduced), the computer can choose an h that will ensure a balance of power between the two ears. With the knowledge of this relationship, one can make use of the equation for cross-correlation $\gamma(\tau)$, and insert signals for the slightly incoherent interval. This calculation will establish the coherence for a particular combination of h and g :

$$\gamma(\tau) = \frac{\int x_L(t) x_R(t + \tau) dt}{\sqrt{\int dt_1 x_L^2(t_1) \int dt_2 x_R^2(t_2)}}$$

$$x_L = A$$

$$x_R = hA + gB$$

$$\text{Coherence} = \gamma(\tau = 0) = \frac{\overline{(A)(hA + gB)}}{\sqrt{\overline{A^2} \overline{(hA + gB)^2}}}$$

$$= \frac{\overline{hA^2}}{\sqrt{\overline{A^2} (h^2 \overline{A^2} + g^2 \overline{B^2})}} = \frac{\overline{hA^2}}{\sqrt{\overline{A^4} (h^2 + g^2)}} = \frac{\overline{hA^2}}{\overline{A^2}} = h$$
(2)

From the above calculation, it is apparent that the coherence value for this interval is, in fact, simply scale factor h for noise A . Note that the result depends on several

assumptions: that A is independent of B, A and B are of the same level, and $h^2 + g^2 = 1$.

The experimenters needed to ensure that the setup does not conflict with these assumptions, and that the results can be trusted.

Sources of Error

One possible source of error is the digital noise generator. It is fairly simple to generate two independent noises, but to guarantee that they have equal levels is something else altogether. There is an inherent randomness to noise that defies prediction, and that is no better illustrated than by Experiment 1's first noise-generating algorithm. The original design used Rayleigh-distributed noise as a stimulus, which is defined as a type of "random-amplitude, random-phase" noise. This noise type is commonly found in everyday stimuli. The sample period was 50 μ s per channel, buffer length was 16384 points for a frequency spacing of 1.22 Hz and a top frequency of 10 kHz. The phases were selected randomly after the program's number generator received a random seed between 0 and 9999. In the broadband case, when the digital signal is composed of many thousand components, the Rayleigh-distributed amplitudes are not a problem. Thanks to the law of large numbers, the overall power tends to be the same from trial to trial. However, the experiment called for well-defined 100-Hz bands, which included only 82 component frequencies (a result of the 1.22-Hz spacing). With fewer random numbers, the power tended to vary every time the noise was generated, fluctuating by as much as one decibel between intervals—obviously in violation of the assumption that the A and B powers were identical. This fluctuation influenced the decision to change the stimulus to an equal-amplitude, random-phase (EARP) noise.

Since the EARP amplitudes were constant between trials, the power also remained constant. It is also important to note that EARP noise is perceptually indistinguishable from Rayleigh noise, so the experiment did not lose any relevancy with respect to real-world stimuli.

The Programmable Attenuators were another element of variability in the experimental design. These components were crucial to setting the scale factors for noises A and B, and the coherence calculation depended on the relationship between them holding true (namely, that $h^2 + g^2 = 1$). The task was then to establish how sensitive the coherence is to fluctuations in level, and the accuracy of the PAs with which parameters h and g were determined. First, knowing that the PAs are only precise to 0.1 dB, the effects that an uncertainty of that magnitude would have on coherence were calculated. Using the relationship between level and amplitudes:

$$0.1 \text{ dB} = 20 \log \left(\frac{h_{\text{actual}}}{h_{\text{program}}} \right)$$

Let the program-set coherence be 0.90, which is one of the typical coherences involved in this experiment. Thus, $h_{\text{program}} = 0.9$:

$$\frac{0.1}{20} = \log \left(\frac{h_{\text{actual}}}{0.9} \right)$$

$$h_{\text{actual}} = 0.9 * 10^{0.1/20} = 0.910422$$

The calculation cannot blindly stop here, since h is no longer the coherence as was previously established. Now, it must include the denominator from the original calculation:

$$coh = \frac{h_{actual}}{\sqrt{h_{actual}^2 + g^2}}$$

$$g = \sqrt{1 - h_{program}^2} = 0.43589$$

$$coh = \frac{0.910422}{\sqrt{0.828867 + 0.19}} = 0.901953$$

According to this corrected calculation, there is a change of less than 0.2 percent coherence from the original estimation. This is a reasonable error compared with the range of coherences found in the results. To illustrate the negligible nature of these deviations, Figure II.4 graphs the effect of a ± 0.1 dB error on expected coherence. The x-axis represents the coherences as calculated by the computer, while the y-axis uses the recalculation above to account for “actual” coherence. The range of coherences from 0.7 to 1 includes all values of h used in the experiments. Note that the dotted and dashed lines never deviate from perfect accuracy by more than 1% coherence. Also, most runs found listeners staying in the higher coherences, where the error is smallest. Thus, Experiment 1 could tolerate attenuation errors of this magnitude.

Since the parameter g was also dependent on a PA, these calculations were repeated with a corrected version, and found that an error in noise B has the same effect on coherence as it did for noise A, but with the signs reversed. Thus, if a +0.1 dB error

on noise A caused the coherence to increase by 0.001953, a -0.1 dB error on noise B would do exactly the same.

All the above calculations were performed under the assumption that the PAs could have an error on the order of their precision. To be thorough, the PAs were tested for accuracy by comparing a sine wave input of known amplitude with measured output voltage. A program cycled the PAs from 0 to 80 dB of attenuation in 0.1 dB steps, while the disparity between the calculated (actual) and expected attenuations were recorded by a computer-controlled Keithley multimeter and saved in a file at each step. Both PAs were tested in this manner, with the input amplitude at 3 or 9 volts. One example of the results can be found in Figure II.5, a plot of attenuation error vs. expected attenuation from 0 to 60 dB. All four cases (two PAs, two input voltages) showed errors of much less than 0.1 dB up to an attenuation of about 70 dB. However, since the attenuation for noise B during the sensitive mixing process tended to be less than 20 dB, one can assume that the apparatus is quite satisfactory for experimental purposes.

A final test attempted to ensure that noises A and B were treated equally throughout their signal paths, save for the computer-controlled variation in the PAs. Having altered the noise generation technique to guarantee that A and B had equal levels upon creation, it was necessary to check that they were still equal upon mixing (refer to Figure II.3). There were many opportunities for bias along their signal path, from the DAC outputs, through the low-pass filter and PAs, then across long cables to one of three mixer inputs. Measurement of the DAC outputs with a voltmeter confirmed that they were virtually identical. In order to account for the rest of the signal path, the experimental program added two steps to the calibration procedure that alternated which

PA was passing the signal, but substituted sine waves for noise. This change was made for ease of measurement, since the RMS voltage reading on the multimeter would have been erratic and uncertain for narrowband noise. The voltage at the mixer output provided comparison data for each combination of channel and mixer input (two channels times three inputs gave us six results). The results guided the selection of inputs for A and B which offered the lowest voltage difference. The level difference between them was less than 0.1 dB in this configuration – once again, certainly low enough for this experiment.

Presentation and Listeners

Experiments were conducted in an Acoustic Systems double-walled soundroom. Stimuli were presented through Sennheiser HD-480 II headphones at a standard level of 64 dB. In all, six listeners were involved in the initial coherence discrimination experiments, designated by the letters M, N, P, V, W and Z. All were males, between the ages of 18 and 26, except for listener W, who was 60 at the time of these trials. P and M were left-handed. All had musical experience. Listeners W and Z had extensive prior listening experience, while the rest had never before participated in this type of hearing study. The only hearing problems reported were possible high-frequency loss for W, consistent with that for middle-age males, and childhood ear surgery for Z (adenoid removal). Audiograms for V, W and Z exhibited normal hearing over the range of 125-8000 Hz, except for W, who did lose some sensitivity above 3000 Hz.

Results

Figure II.6 plots the results for all six listeners who participated. The x-axis represents the center frequencies of each 100-Hz-wide band (including the broadband case at far right), while the y-axis plots the coherence threshold necessary to discriminate from a perfectly-coherent image in these bands. The difference of these thresholds from one is thus the coherence jnd. First, note that all six listeners did well for bands centered at low frequencies, based on how close their thresholds are to one. The higher the value, the more similar the coherent and slightly incoherent intervals were, and the more difficult the task. Most of our listeners achieved thresholds around 0.99 for the lowest-frequency bands, indicating a jnd of only one percent coherence. This is excellent discrimination, and it matches closely (to two significant figures) with results from Gabriel and Colburn's (1981) two listeners for a 115-Hz wide band centered at 500 Hz. Note that one of their listeners achieved an incredible threshold of 0.997, compared to 0.989 for the other. The second feature of interest is the broadband (BB) thresholds, shown on the right side of the graphs. The BB results are in line with those of Pollack and Trittipoe (1959), who performed a similar discrimination experiment with "wideband" noise and found that listeners could distinguish between coherences of one and 0.98 in about 75 percent of trials. Each listener's broadband threshold is nearly equal to his or her best threshold among 100-Hz bands. However, some listeners actually performed better in their best band. See Table II.2 for comparisons, which also includes listeners' thresholds for the 500-Hz-centered band when that did not happen to be their best band.

Table II.2 - Bandwise coherence thresholds						
Listener	M	N	P	V	W	Z
Broadband	0.962 (0.022)	0.981 (0.006)	0.977 (0.006)	0.985 (0.005)	0.980 (0.005)	0.988 (0.003)
Best band	0.975 (0.010)	0.986 (0.005)	0.991 (0.003)	0.983 (0.007)	0.993 (0.003)	0.993 (0.002)
500-Hz center band	Same	0.982 (0.006)	Same	Same	0.988 (0.003)	Same

Coherence jnd thresholds (n-1 standard deviation) for best 100-Hz band and broadband stimuli

The slight degradation of thresholds according to bandwidth may indicate some of the effect discovered by Gabriel and Colburn (1981) mentioned earlier. Broadband thresholds in the two experiments appear to be similar; Gabriel and Colburn found correlation jnds of two to three percent for a 10-kHz broadband noise, while the six listeners in the present study showed jnds of about one to four percent in the BB case. Concentrating solely on the bands centered at 500 Hz, matching the condition tested by Gabriel and Colburn, most of the listeners in Experiment 1 showed improved thresholds. However, a t-test ($t(5) = 6.168, p = 0.056$) does not show significant improvement for the narrower band. These results do not indicate declining coherence discrimination with increasing bandwidth as was suggested in the previous study.

Finally, in the two highest bands where fine-structure ITD discrimination is poor, most subjects still performed extremely well. Listeners M and V both show a significant reduction in discrimination relative to the lower-frequency bands, but others held relatively steady across all center frequencies. Relatively consistent performance above and below the 1500-Hz limit seems contrary to the described concept of the binaural system, although Bernstein and Trahiotis (1995) found very similar results in their tests of

the MLD. Their subjects could be split into two distinct groups, much like those reported in Experiment 1: a few that struggled with the high-frequency stimuli, and the rest that showed little frequency dependence in their thresholds.

Discussion

One can explain the ability to discriminate high-frequency narrowband coherence with the existence of envelope ITD cues. Bernstein and Trahiotis (1996) predicted that the binaural system's use of the normalized correlation (which includes the dc value of the envelope) in MLD tasks would allow discrimination at high frequencies. The first figure in their paper (Figure II.7) graphs the calculated normalized correlations of narrowband stimuli for several center frequencies of $S\pi$ and signal-to-noise ratios, after processing the stimuli to extract the high-frequency envelope. By comparison, Experiment 1 typically used very little incoherent noise (signal), so listeners were mostly operating within the leftmost region of this graph. In this region, all center frequencies have nearly the same normalized correlation, and thus are equally difficult to distinguish from one. This implies that the use of an envelope cue allows listeners to establish the coherence of MLD stimuli at very high frequencies, and that coherence is almost invariant with respect to center frequency for low signal-to-noise ratios. Granting that listeners perform the MLD task by coherence discrimination, their analysis leads to the prediction that listeners should have relatively constant thresholds for narrowband noise at center frequencies above and below the 1500-Hz "limit". This was indeed the case for some, but not all, of the listeners in Experiment 1.

The envelope is illustrated in Figure II.8, which plots a sample 100-Hz-wide stimulus centered at 2000 Hz. While the waveform itself is oscillating rapidly, the envelope (joining the peaks of the waveform) varies slowly. The auditory neurons may not be able to keep up with the high frequencies involved in the fine structure, but could still follow the fluctuations of the envelope. One's binaural system would find it much easier to identify an interaural phase difference (and thus, an ITD) from the shift in this envelope's peak than to identify a shift in a particular waveform peak.

To quantitatively illustrate the frequencies involved in envelope fluctuation, the following mathematical development of the envelope $E(t)$ begins with the absolute value of the sum of all contributing components:

$$E(t) = |\tilde{x}(t)| = \left| \sum_{n=1}^N C_n e^{i(\omega_n t + \phi_n)} \right|$$

$$E^2(t) = \left| \sum_{n=1}^N C_n e^{i(\omega_n t + \phi_n)} \right|^2 = \sum_{n=1}^N C_n e^{i(\omega_n t + \phi_n)} \sum_{m=1}^N C_m e^{-i(\omega_m t + \phi_m)}$$

$$\rightarrow E^2(t) = \sum_{n=1}^N C_n^2 + 2 \sum_{n=1}^N \sum_{m=1}^{n-1} C_n C_m \cos[(\omega_n - \omega_m)t + (\phi_n - \phi_m)] = P_E(t)$$

As seen on the last line, the envelope power is governed by the “difference frequencies” that exist between waveform components ($\omega_n - \omega_m$). The envelope power is therefore “translation-invariant,” meaning that one could shift all the frequencies by the same amount, and the envelope would be unchanged. This leads to an important result: while the difference frequencies include all possible combinations of two components, the highest frequency present in the envelope can only be the largest distance between any

two parts of the frequency spectrum. In the case of our stimulus, a 100-Hz wide band, the center frequency has no relevance for the envelope function. The highest frequency of oscillation in the envelope is thus 100 Hz, regardless of whether the band lies above or below the 1500 Hz limit. The implication is that while listeners cannot follow fine-structure ITDs at high frequency, the slower-varying envelopes can still provide a usable ITD cue. This cue would account for the frequency-independent coherence thresholds of some listeners.

With the above knowledge, one can revisit Figure II.6 and better understand the patterns of data. For low-frequency bands, the listeners did equally well (to an extent, neglecting the contribution of listener M). However, their coherence thresholds diverged rapidly as the center frequency increased. Specifically, listeners Z, W, P and N made efficient use of envelope cues. A logarithmic curve fit to the average results of those four listeners resulted in an R^2 of 0.839, indicating that there was some dependence on center frequency. Z and P were nearly unaffected by the loss of fine-structure cues. Listeners V and M, on the other hand, did not glean as much information from envelope cues, and their thresholds suffered. Bernstein and Trahiotis (1998) discovered similar disparity among listeners with respect to MLD at center frequencies of 500 and 4000 Hz. Their listeners' results also suggest differing abilities to identify envelope correlation cues.

The inter-subject variability seems to indicate the presence of two mechanisms for coherence discrimination. All listeners were able to make use of the traditionally-understood fine structure cues to measure interaural coherence. However, once those cues failed, some listeners found that they could perform the task just as well with envelope information. Others did not benefit as much from the envelope.

The potential degradation of coherence discrimination at high frequencies may also provide a clue to the dependence of jnds on bandwidth, which is a discrepancy between the results of Experiment 1 and those of Gabriel and Colburn. Jain *et. al.* (1991) noted that some listeners show decreased decorrelation sensitivity at greater bandwidths, and suggested that may be a result of the gradual inclusion of higher frequencies where sensitivity is poorer. Rather than providing useful additional information to the binaural processor, perhaps the high-frequency coherence cues reduce the processor's overall performance. This hypothesis necessarily suggests that listeners who show poor coherence sensitivity at high frequencies (in this study, listeners V and M) should also demonstrate degraded sensitivity as a function of bandwidth. A check of Table II.2 shows that listener M certainly follows this pattern, as his broadband threshold falls well below his best band threshold. However, listener V's broadband result is actually better than his best band. To confuse matters more, listener P, whose narrowband thresholds are most consistent across frequency, shows a significant decline for the broadband case. Thus, Experiment 1 cannot provide definitive support for the hypothesis of Jain *et. al.*

Experiment 2: Broadband coherence discrimination at short durations

Introduction

Imagine a model that explains the coherence jnd as a function of resolution of the number of neural spikes. Assume that discrimination is dependent on the accumulation of enough data from Jeffress' coincidence counters (Stern and Trahiotis, 1995), as defined in Chapter I. While the accumulation of signal (number of spikes) increases as

number N , the standard deviation (variability in spike count) only increases as the square root of N . Increasing the amount of information gathered leads to a more favorable signal-to-noise ratio, improving discrimination. This concept of information-gathering parallels a model of Durlach *et al.* (1986) used to predict MLD and coherence-discrimination results. Their model is also based on the ratio of useful signal to interfering noise.

One method of augmenting the available data is to increase stimulus bandwidth, which would stimulate nerves of different characteristic frequencies and cause output from additional coincidence counters, supplementing the overall coherence measurement. The increase in information available to the system should lower the jnd. Of course, results from Experiment 1 and previous studies show that this may not be the case.

While additional bandwidth can provide more data, another dimension of information-gathering is over time. By limiting the time window for data accumulation, one could theoretically decrease the available information, thus raising the coherence jnd. It is well-known that the binaural system can make use of data accumulated over a duration called the “integration time.” Assume an integration time of 100 ms, a common value found in the literature (Durlach and Colburn, 1978). The implication is that the binaural system only “stores” information gathered over the last 100 ms, and performs analyses ignoring data that arrived before that period. It stands to reason that the binaural system should perform equally well whether provided with signal durations equal to or longer than the integration time. On the other hand, if the signal is shortened below 100 ms, the binaural system has less information available, and coherence discrimination should suffer. Blodgett, Jeffress and Taylor (1958) found results that support the above

“information-gathering” model in the course of an MLD experiment, estimating that a direct proportion existed between the logarithm of stimulus duration and detection threshold (in dB) until the threshold reached its asymptote.

Part A: Variable Duration

For the duration experiment, the task was exactly the same as the broadband (20-6000 Hz) stimulus of Experiment 1, but the duration was set at one of ten values: 500, 300, 170, 100, 80, 60, 40, 20, 10, and 5 ms. The signal level was 64 dB SPL. Experiment 2 provided the listeners with a broadband stimulus to ensure that their binaural systems had access to a large amount of information in the frequency domain, even at very short durations. The choice of broadband also avoided the problem of “spectral splatter” for short stimuli. In order to provide very accurate signal durations for *this* time-sensitive experiment, the trapezoidal envelope was removed from the setup in *Figure II.3*. The stimulus was then “windowed” by a rectangular envelope, with *effectively* instantaneous onsets and offsets controlled by a precise timing generator *independent* of the computer, so that expected and actual duration were identical. The “*hard*” onsets were audible at the beginning of the stimulus, but experimenters deemed *them* unobtrusive enough to not distract the listeners. Listeners included W and Z from *the* previous experiment, as well as two new subjects: S and T. Both were very *experienced* in hearing studies and had normal hearing.

The results of the variable-duration experiment are plotted in *Figure II.9*. On the *x-axis* are the various stimulus durations (the label for 80 ms is omitted, but the results *are still* present), and the y-axis once again represents the coherence threshold at the 79%

discrimination level. The four participating listeners are separated into two graphs for clarity. An analysis of variance (ANOVA) test measured the variation in thresholds for all listeners, and gave a quantitative assessment of any significant differences across conditions. The alpha level for significance was set at 0.05, meaning that differences between two means are significant with a 95% probability. Essentially, one in twenty of these measurements may show significance when there is none. The mean-difference calculations were performed under a least-significant-difference (LSD) procedure. The ANOVA showed that there is a significant effect on thresholds due to duration ($F(9,27) = 9.505, p < 0.001$). The upper graph of listeners S and T shows many of the features expected from the discussion above. For durations of 100 ms and longer, their thresholds are consistent. This implies that the integration time was indeed near 100 ms, and additional information beyond that time did not make the task any easier. At shorter durations, they required more incoherence to make an accurate judgement, showing a strong duration effect ($F(9,9) = 21.036, p < 0.001$). The graph for listeners W and Z shows a different pattern. Their performance was not only consistent above 100 ms, but also below, and almost as good at 5 ms as at 500 ms. These two plots show very little downturn at short periods, a smaller though still significant duration effect ($F(9,9) = 8.446, p = 0.02$). Once again, the results find two distinct groups of listeners.

Part B: Variable Duration with randomized ILD

Because W and Z performed well even at the lowest durations in part A, where S and T struggled, the experimenters theorized they may be using alternate cues to do the task. In this case, it was possible that the constant position of the image inside the

subject's head might have simplified their decision-making process. Refer back to Figure II.2 for a look at the two intervals. These images were presented with no ITD or ILD, placing them in the center of the head. The coherent image is a tightly-defined point along the internal delay axis, while the slightly incoherent image is somewhat broader due to the addition of independent noise. Instead of judging which image had a greater ASW, the listener could simply listen for the presence of off-center noise, which would indicate that the signal was not totally coherent. Thus, the task becomes one of position detection, rather than coherence discrimination. This concept is termed the "position-cue hypothesis."

As a check of this conjecture, Experiment 2 was changed in order to eliminate the possibility of a position cue. The stimulus modification included the addition of a random ILD to each trial. ILD has the same effect on image lateralization as ITD. By randomizing lateral position each trial, the listeners would not be able to learn a particular location on which to focus their attention. The computer picked an ILD from integers within the range of ± 4 dB, and all values had an equal probability. The easiest method of introducing an ILD was obviously the programmable attenuators that were used to control the levels of noises A and B. In order to ensure that the overall coherence was not altered by this perturbation, it was necessary to perform a recalculation of h and g . It was a two-equation, two-variable problem, given that the power in the right ear was a certain percentage of that in the left (as determined by the ILD), and we wanted the coherence unchanged. Thus, the calculation found a new h_{ild} and g_{ild} (that met criteria) in terms of the previous h and g . Let x be the amplitude ratio of right ear over left. Whereas

equation 1 equated left and right powers by setting h and g appropriately, this derivation uses **them** to affect an ILD:

$$ILD = 20 \log x$$

$$h_{ild}^2 + g_{ild}^2 = x^2$$

Because the coherence set by h_{ild} and g_{ild} should be no different from coherence h set by the **non**-ILD task (refer to the calculation of equation 2),

$$\frac{h_{ild}}{\sqrt{h_{ild}^2 + g_{ild}^2}} = h \rightarrow \frac{h_{ild}}{\sqrt{x^2}} = h \rightarrow h_{ild} = hx$$

$$h^2 x^2 + g_{ild}^2 = x^2 \rightarrow g_{ild}^2 = x^2(1 - h^2) = x^2 g^2 \rightarrow g_{ild} = gx$$

As the equations show, attenuating or boosting both A and B noises equally in the **right channel** does not change the coherence, or in simpler terms, coherence is level-**invariant**. This is an expected result. Part B of Experiment 2 used durations of 500, 100, 40, 10 and 5 ms as a representative set of the original stimuli. Aside from these changes, the **experiment** was identical to the duration tests as described in part A.

The addition of a random ILD is not without precedent; Bernstein and Trahiotis (1997) **introduced** just such a “roving” binaural cue in their tests of correlation sensitivity. They **found** that random ILD did not significantly degrade their listener’s performance from **tests** performed without the ILD. This experiment was performed with the

expectation that any changes in threshold from part A would be due to the loss of a **position** cue, and not a result of random ILD.

The results of the “random ILD duration experiment” may be found in Figure II.10, **and** should be analyzed with respect to Figure II.9. Note that the thresholds of **listeners** T and S are relatively unchanged, and the new ILD factor has not affected their **performance**. They show a significant effect over duration ($F(4,4) = 20.364, p = 0.006$). **However**, for durations less than 100 ms, W and Z now show a distinct degradation **coherence** discrimination, a behavior that agrees better with their fellow listeners ($F(4,4) = 17.65, p = 0.008$). These data fit the predictions of the information-gathering model, **because** jnds increase at the shorter durations where coincidence data are more limited. **Overall**, the ANOVA shows a strong duration effect ($F(4,12) = 34.948, p < 0.001$). The **change** in W’s and Z’s results also suggests that they were indeed using a position cue to **boost** their short-duration performance.

Part C: Variable Duration with zero, constant, and random ILD

Analysis of part B’s results could lead to a different conclusion. What if W and Z did **worse** simply because their coherence discrimination was best when the image was **centered** in their heads? Is performance dependent on intracranial location? These **questions** shed doubt on the position-cue hypothesis, and the experiment was redesigned **again** to answer them. Part C re-evaluated durations 40, 10 and 5 ms, as they were the **stimuli** that showed some change between the original duration experiment (part A) and the **random** ILD revision (part B). Listeners were run at each duration under three **conditions**: a constant ILD of 0 dB (which was identical to the experiment of part A), a

constant ILD of +4 dB, and a random integer ILD within the range of ± 4 dB (reproducing the experiment of part B). If the theory of listeners using the position cue over coherence discrimination were correct, one would expect the constant ILD of +4 to produce similar results to the constant ILD of 0. But if discrimination truly were not as accurate outside the center of the head, one would expect it to resemble the random ILD results.

The answers to this debate lie in Figures II.11, II.12 and II.13. The first is a plot of constant ILD = 0 results, and shows an interesting feature. Whereas the listeners for the duration experiment of part A were originally split into two distinct groups, this retrieval exhibits more consistent thresholds for all subjects down to 5 ms. These results don't quite show a significant duration effect ($F(2,6) = 5.076, p = 0.051$). In comparison to the results in Figure II.9, it is apparent that S and T have improved their short-duration performance. This change could indicate that they had discovered the use of the position cue, but judgement should be reserved until all the facts are analyzed. Figure II.12 contains data for the ILD = +4 condition, which are not significantly different from those for ILD = 0 according to statistical t-tests ($t(3) = 2.197, p = 0.159$). However, the same tests confirm that the random ILD results in Figure II.13 are appreciably lower than their ILD = +4 counterparts ($t(3) = -7.442, p = 0.018$). These results serve to disprove the hypothesis that ILD randomization reduced discrimination performance because the images were off-center, and lend credence to the position-cue hypothesis. It seems all listeners were using the position cue by the beginning of part C.

The duration-insensitivity found in Figure II.11 raises a question: how do these results relate to MLD duration dependence, which Blodgett, Jeffress and Taylor (1958) found to be very significant? In the same range of duration, from 40 to 5 ms, they

measured a 10-dB increase in the required signal-to-noise ratio. The four listeners of Experiment 2 averaged coherence thresholds of 0.955 at 40 ms, which lowered to 0.927 at 5 ms. The change was small enough that the ANOVA showed no significance. However, referring back to the small S/N region of Figure II.7, one might discover that very small changes in coherence near one correspond to large S/N ratio changes for an MLD task. The coherence reduction stated above appears to span a change of 3-4 dB. However, if one uses the 40- and 5-ms results in Figure II.9 (before listeners S and T used the position cue), the average coherences vary from 0.943 to 0.890, respectively. That change corresponds to a S/N shift of 6-7 dB on Figure II.7, closer to the sensitivity of Blodgett, Jeffress and Taylor's subjects.

Experiment 3: Level-dependent coherence discrimination

Introduction

The final iteration of the coherence-discrimination experiments altered the one value that had been constant through all previous variations: stimulus level. The binaural system is notably better at analysis when provided with a high level (Durlach and Colburn, 1978), and the experimenters wanted to determine how coherence jnds might change as the level decreased from our standard of 64 dB. A brief study by Gabriel and Colburn (1981) indicated that decreasing spectral level led to larger coherence jnds for their two listeners. When measuring with a 115-Hz wide stimulus centered at 500 Hz, they found coherence jnds for two levels: 0.008 at 75 dB, increasing to 0.013 at 44 dB.

One approach to the question of level has been dubbed the “fuzzy ball copier analogy” by the experimenters. For an illustration, see Figure II.14, which shows the familiar coherent and slightly incoherent images. Each subsequent copy of the pair, as they descend on the page, was made by lowering the darkness setting. This is analogous to lowering the overall level of the intervals. Note how the fuzzy edges of the slightly incoherent image fade away towards the bottom, leaving a more compact image that resembles that on the left. This demonstrates why it may be more difficult to distinguish between the two at low levels, as the edges fall below the threshold of hearing. Listeners would essentially lose the “position cue” that was postulated in Experiment 2.

The darkness of the images could represent regions where the number of neural spikes is above the detectable minimum. As intensities decrease, the spike count is reduced until it ceases to register in the coincidence detector, which harkens back to the information-gathering model. A lower intensity stimulates less neural activity, or the spike trains do not synchronize as well with the input signal (two elements of Colburn’s (1978) auditory-nerve based model), providing fewer/noisier data to the coincidence counters. The information-gathering model finds this situation no different than if the input was limited in time or frequency, because both have lower signal-to-noise ratios. Thus the model predicts that listeners will establish larger coherence jnds for lower intensities.

Methods

To test the copier analogy, Experiment 3 ran the discrimination task at levels of 64, 44 and 34 dB. The signal was broadband (20 Hz – 6 kHz, as the BB condition of

Experiment 1), and 100 ms in length. Listeners S, T, W and Z each performed four runs at the **three** levels.

Results

Results from Experiment 3 appear in Figure II.15. The three levels appear on the **x-axis**, while the y-axis is still the coherence threshold for the listener to distinguish **correctly** 79% of trials. At the experiment-standard level of 64 dB, listeners performed **equivalently** to their previous results. As the level decreased, however, their performance **clearly** degraded. The results averaged over listener are found in Table II.3.

Table II.3 - Level-dependent coherence thresholds			
Level (dB)	64	44	34
Average Threshold	0.972 (0.006)	0.936 (0.016)	0.873 (0.043)

Average coherence jnd threshold (n-1 standard deviation) for three stimulus levels

The pattern of thresholds follows Gabriel and Colburn's findings, although the **listeners** of Experiment 3 exhibited a much larger coherence jnd (0.064) at 44 dB than **reported** in the cited work (0.013). Once again, their listeners appear to be much more **sensitive** to changes in coherence. An ANOVA reveals a significant effect due to level ($F(2,6) = 26.45, p = 0.001$). It would seem that coherence discrimination is indeed **sensitive** to level, and the "copier" analogy may be a fitting one. The model's predictions based **on** nerve-activity threshold were borne out by the results.

Conclusions

disc

four

coll

and

bro

AN

exc

sho

det

dir

II.F

lev

Exp

wh

lev

pos

inc

In Experiment 1, some listeners demonstrated that envelope coherence discrimination can be just as good as that for fine-structure. However, not all listeners found envelope cues as detectable as fine-structure cues in this context, leading to higher coherence jnds for narrow bands above 1500 Hz. Another comparison, between narrow- and broadband discrimination, found that some listeners did have larger jnds for a broadband stimulus. However, the effect was extremely variable among subjects, and ANOVA results showed no significant difference.

Experiment 2 showed that while some listeners exhibit increasing jnds for extremely short stimuli, others can discriminate coherence consistently well even at very short durations. This result is likely due to an ability to perform the task using position detection cues. Once those cues are eliminated, coherence jnds increase with decreasing duration for all listeners. The position of intracranial images (shifted off-center by an ILD) apparently has little or no effect on coherence discrimination.

Experiment 3 results revealed that coherence jnds also increase as the stimulus level decreases, possibly due to the loss of the same position detection cues used in Experiment 2 (as illustrated by the copier analogy). The information-gathering model, which establishes expected results based on the limitation of data in frequency, time and level, was able to accurately predict the patterns shown by our listeners' thresholds. A possible exception was the case of bandwidth dependence, where results were inconclusive.

Further Study

This line of experiments could lead to many other studies of coherence. Possible future directions are listed below:

- A. Some readers have suggested that randomizing ILD by trial is not quite enough to eliminate the position cue. While the listeners would not be able to focus on one off-center location, they would still be able to compare positions within a trial rather than overall “width.” The experimental program could instead randomize by interval.
- B. All the coherence discrimination tasks described essentially tested coherence jnds with a reference of one. Another interesting problem is to start with a completely incoherent noise and increase the coherence until listeners could tell the difference, corresponding to a reference-zero coherence jnd. Blodgett, Jeffress and Taylor (1958) report a very small MLD when the masking noise is uncorrelated. This could have to do with decreased coherence sensitivity in the region near zero correlation.
- C. Many experimenters are studying the effect of critical bands (frequency ranges that the binaural system uses for analysis) on other tasks, and it has implications for the information-gathering model. It may prove interesting to rerun Experiment 1 with a wide range of stimulus bandwidths.

- D. The reported results could relate to the study of coherence in rooms, and the use of ITD cues in rooms. This direction is pursued in Chapter III.
- E. Using the current experimental task, one might replace the slightly-incoherent interval with noise modified to simulate dispersion around the head from a certain incident angle, and test how well our listeners discriminate between the two. While the compact image of a coherent noise may be easily identifiable, the listeners will no doubt be used to experiencing phase shifts due to dispersion. This very problem is addressed in Chapter V.

References

- Bernstein, L.R. and Trahiotis, C. "Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms," J. Acoust. Soc. Am. **98**, 155-163 (1995).
- Bernstein, L.R. and Trahiotis, C. "The normalized correlation: Accounting for binaural detection across center frequency," J. Acoust. Soc. Am. **100**, 3774-3784 (1996).
- Bernstein, L.R. and Trahiotis, C. "The effects of randomizing values of interaural disparities on binaural detection and on discrimination of interaural correlation," J. Acoust. Soc. Am. **102**, 1113-1120 (1997).
- Bernstein, L.R. and Trahiotis, C. "Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise," J. Acoust. Soc. Am. **103**, 2069-2078 (1998).
- Blodgett, H.C., Jeffress, L.A. and Taylor, R.W. "Relation of Masked Threshold to Signal-Duration for Various Interaural Phase-Combinations," Am. J. Psych. **71**, 283-290 (1958).
- Colburn, H.S. and Durlach, N.I. "Models of Binaural Interaction," Handbook of Perception vol. 4. ed. E. Carterette, pp. 498-503 (1978).
- Durlach, N.I. and Colburn, H.S. "Binaural Phenomena," Handbook of Perception vol. 4. ed. E. Carterette, pp. 365-466 (1978).
- Durlach, N.I., Gabriel, K.J., Colburn, H.S. and Trahiotis, C. "Interaural correlation discrimination: II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548-1557 (1986).
- Gabriel, K.J. and Colburn, H.S. "Interaural correlation discrimination: I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394-1401 (1981).
- Hartmann, W.M. *Signals, Sound and Sensation*. AIP press, Woodbury, New York, 1997. p. 256
- Jain, M., Gallagher, D., Koehnke, J. and Colburn, H.S. "Fringed correlation discrimination and binaural detection," J. Acoust. Soc. Am. **90**, 1918-1926 (1991).
- Klumpp, R.G. and Eady, H.R. "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859-860 (1956).
- Levitt, H. "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467-477 (1971).

Pickles, J.O. *An Introduction to the Physiology of Hearing*. Academic Press, New York, 1982. pp. 173-177

Pollack, I. and Trittipoe, W.J. "Binaural Listening and Interaural Noise Cross Correlation," *J. Acoust. Soc. Am.* **31**, 1250-1252 (1959).

Rayleigh, Third Baron "On our perception of sound direction," *Philos. Mag.* **13**, 214-232 (1907).

Robinson, D.E. and Jeffress, L.A. "Effect of varying the Interaural Noise Correlation on the Detectability of Tonal Signals," *J. Acoust. Soc. Am.* **35**, 1947-1952 (1963).

Stern, R.M. and Trahiotis, C. "Models of Binaural Interaction," Hearing ed. B. Moore, pp. 347-386 (1995).

Trahiotis, C., Bernstein, L.R. and Akeroyd, M.A. "Manipulating the 'straightness' and 'curvature' of patterns of interaural cross correlation affects listeners' sensitivity to changes in interaural delay," *J. Acoust. Soc. Am.* **109**, 321-330 (2001).

Figure II.1. Bandlimits for narrowband stimuli used in this experiment. The vertical axis shows frequency in Hz. The 1500-Hz “cutoff” for fine-structure ITD cues is marked by a dotted line along the center axis. The 100-Hz-wide bands (delineated by solid horizontal lines) fall above and below that cutoff at center frequencies 200, 500, 1000, 2000 and 4000 Hz.

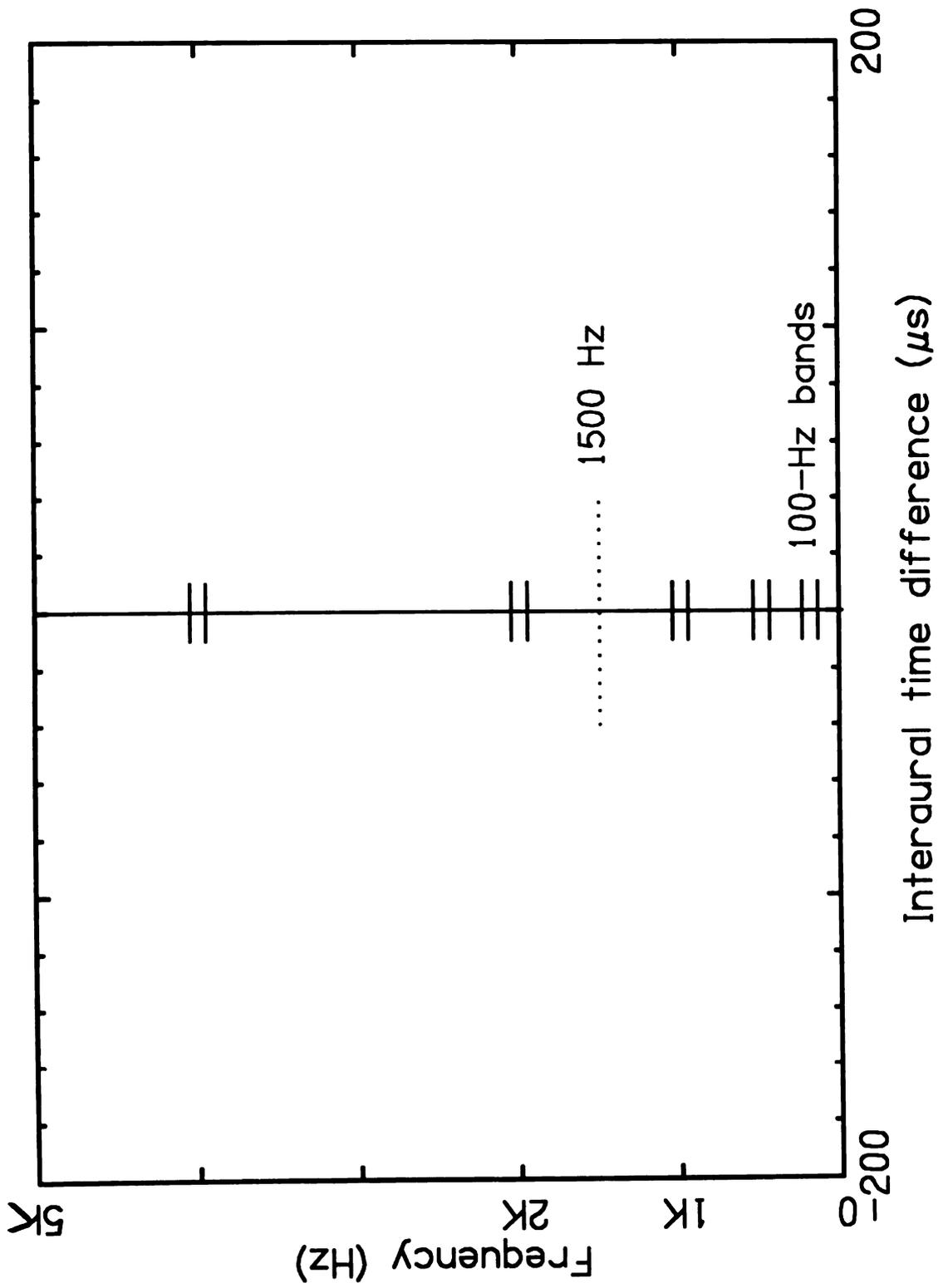
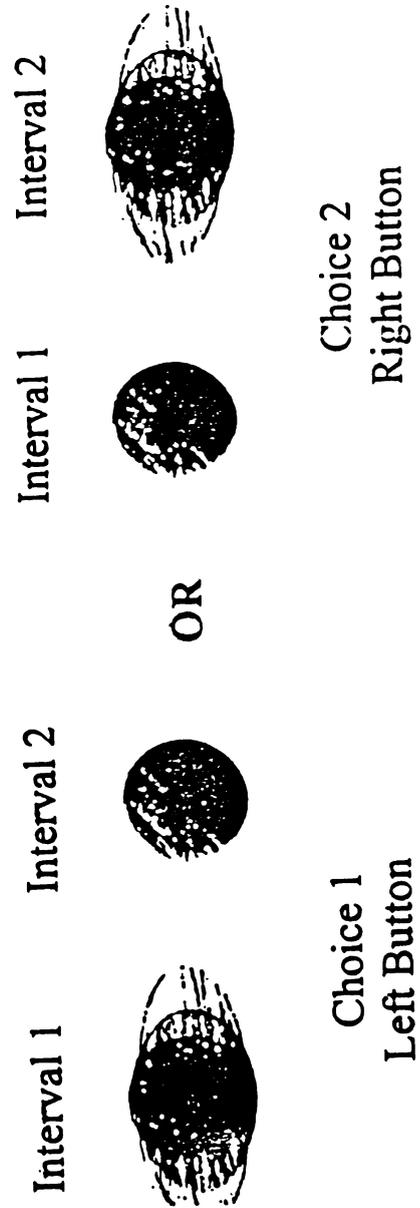


Figure II.1

Figure II.2. Description of the experimental task. Each trial, listeners are asked to choose the least coherent of two intervals. The intervals are ordered randomly so the coherent interval may occur first or second, as shown in two possible trials: the left two images or the right two images. The correct response is listed below each trial.

Experimental Task

Each trial is two-interval, forced-choice



Incoherence controlled by a 3-down, 1-up staircase

Figure II.2

Figure II.3. Schematic of the experimental setup in the form of a patch diagram. Begin at the left, where noises A and B are generated. LPF stands for low-pass filter, PA is the Programmable Attenuator, “hA” and “gB” are the attenuated noises, Amp stands for Amplifier and Trap. Env. for the trapezoidal envelope that ramps signals on and off. The signals feed into left and right ears as shown.

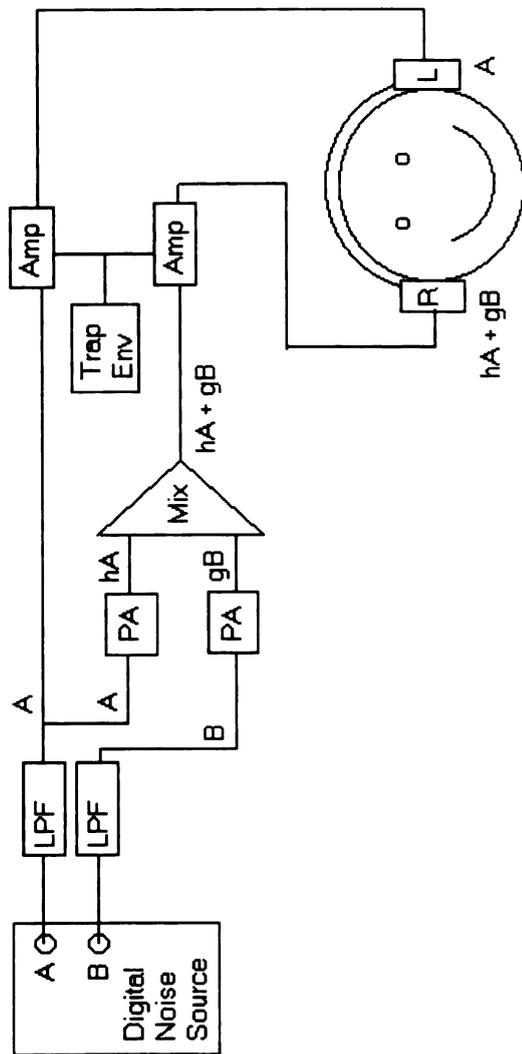


Figure II.3

Figure II.4. Estimation of coherence in the case of Programmable Attenuator (PA) inaccuracy. The x-axis is a range of coherences calculated given the PA settings, and the y-axis is the actual coherence based on assumed error. The solid-line plot is the diagonal, where expected and actual coherence agree. The dotted and dashed lines represent the actual coherences if the PA has a ± 0.1 dB error.

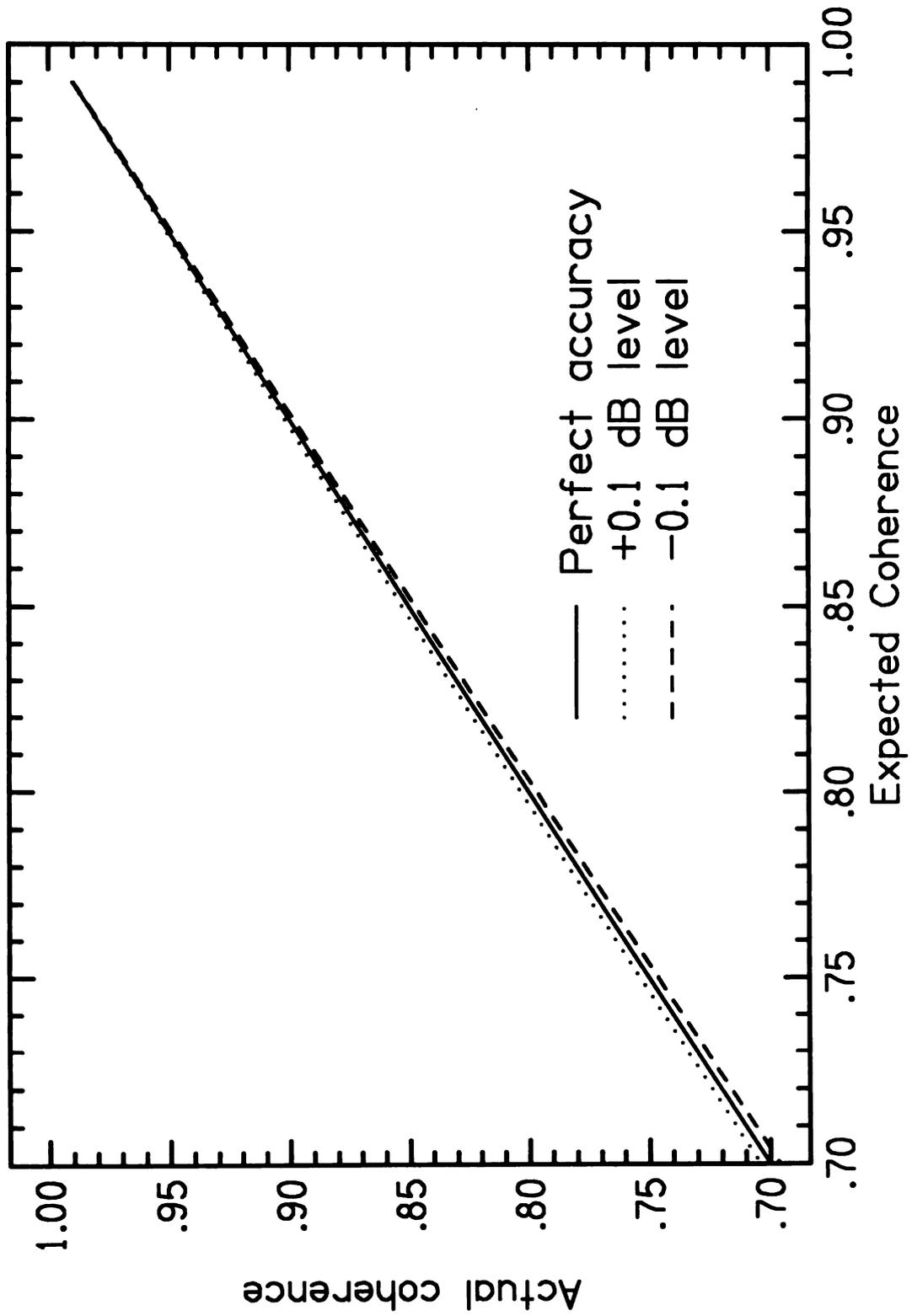


Figure II.4

Figure II.5. Actual measurement of error in the Programmable Attenuator (PA). The x-axis is the range of attenuations set on the PA, while the y-axis shows the measured deviations from those settings.

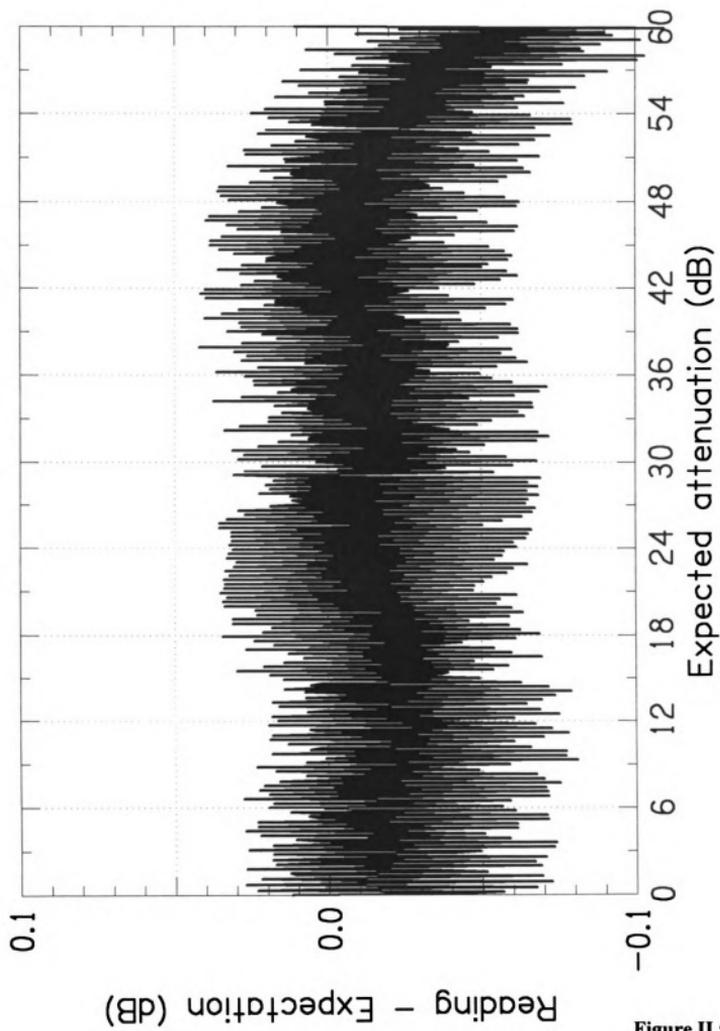


Figure II.5

Figure II.6. Results of Experiment 1. Thresholds for six listeners are broken up into two plots for ease of viewing. Both plots list the band limitations along the x-axis (five center frequencies for the 100-Hz bands and the broadband case listed at the extreme right) and the coherence thresholds along the y-axis are just noticeably different from perfect coherence. The alternate y-axis on the right side of each plot lists the relative level of the independent noise (with respect to the coherent target noise) necessary to achieve the coherences marked. Results for listeners Z, V, and W are plotted in the upper graph; P, M and N are in the lower. Error bars on data points represent two standard deviations.

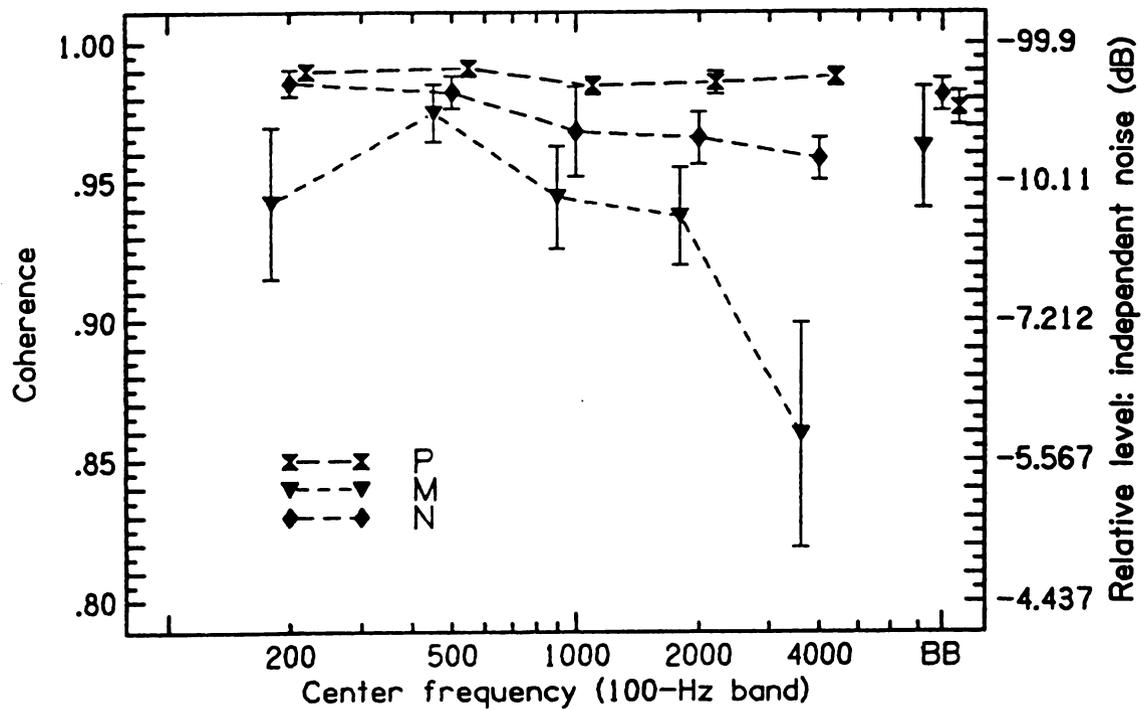
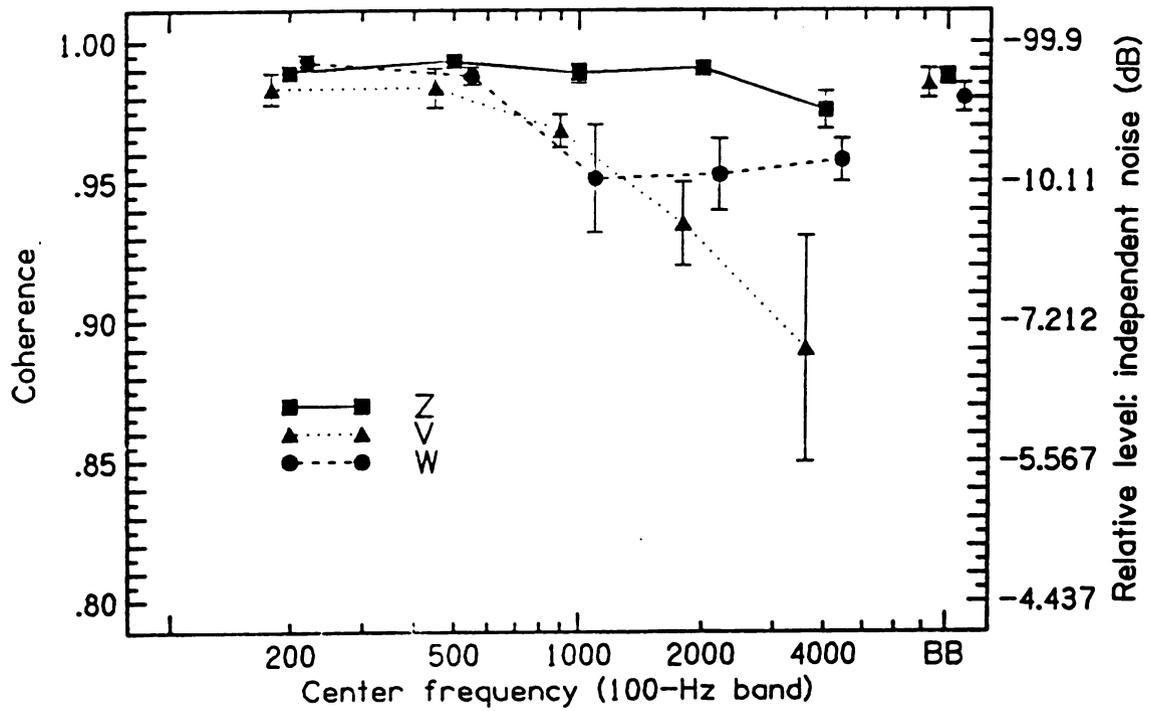


Figure II.6



Figure II.7. Calculated normalized correlations in a Masking Level Difference experiment. The stimuli include No (100-Hz wide band of diotic noise) and $S\pi$ (antiphasic tone centered on the noise band), and their relative levels are established by the signal-to-noise ratio along the x-axis. The y-axis gives the calculated normalized correlation subsequent to half-wave square-law rectification and low-pass filtering. The various plots show correlations dependent on a variety of tonal signal frequencies. (Bernstein and Trahiotis, 1996)

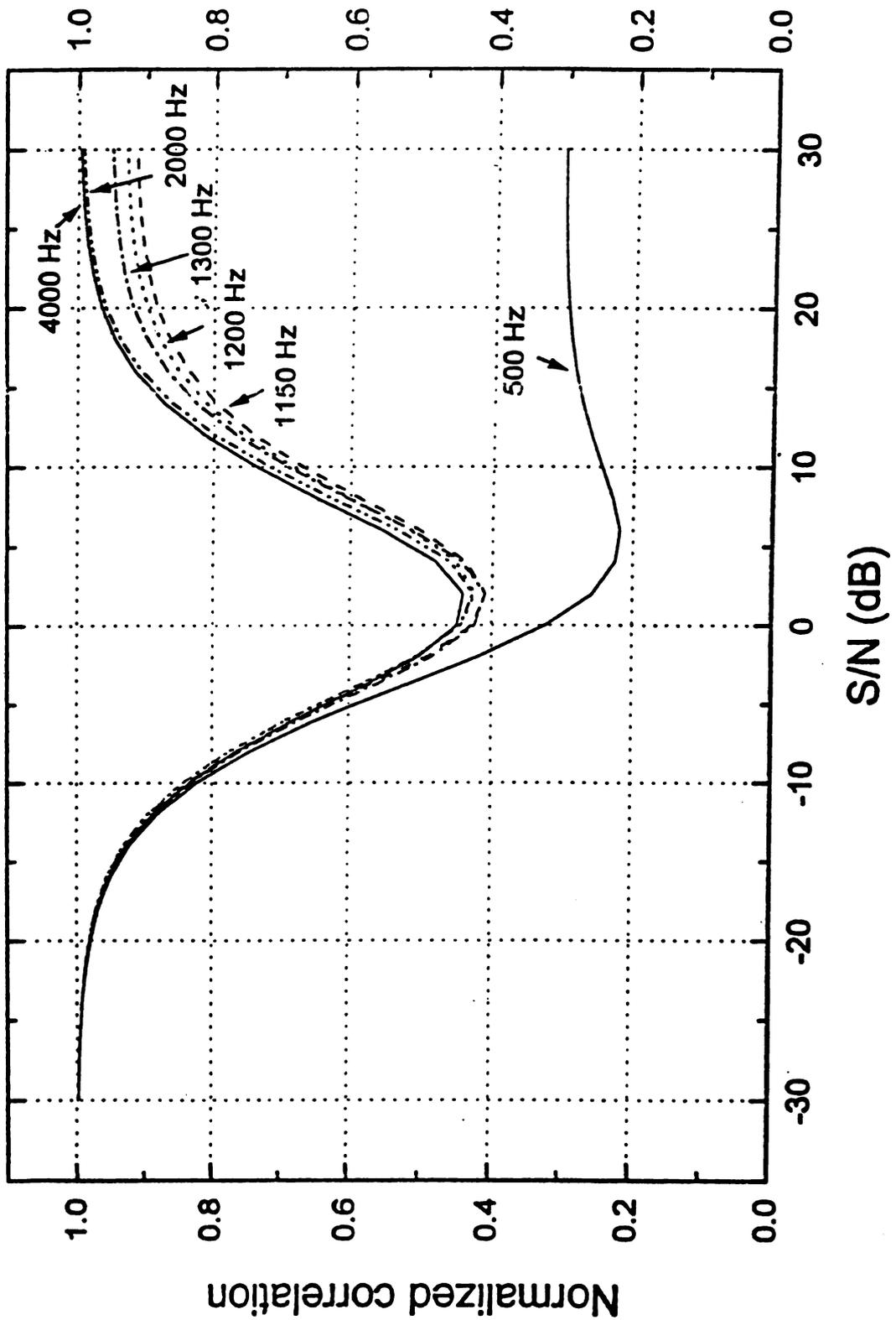


Figure II.7

Figure II.8. The concept of the envelope. A 100-Hz wide band of noise centered on 2 kHz produced this signal, shown as a function of time. The signal waveform is rapidly-varying. By contrast, the signal envelope (marked by dashed lines along the waveform peaks) is slowly-varying, with a most common frequency of 50 Hz.

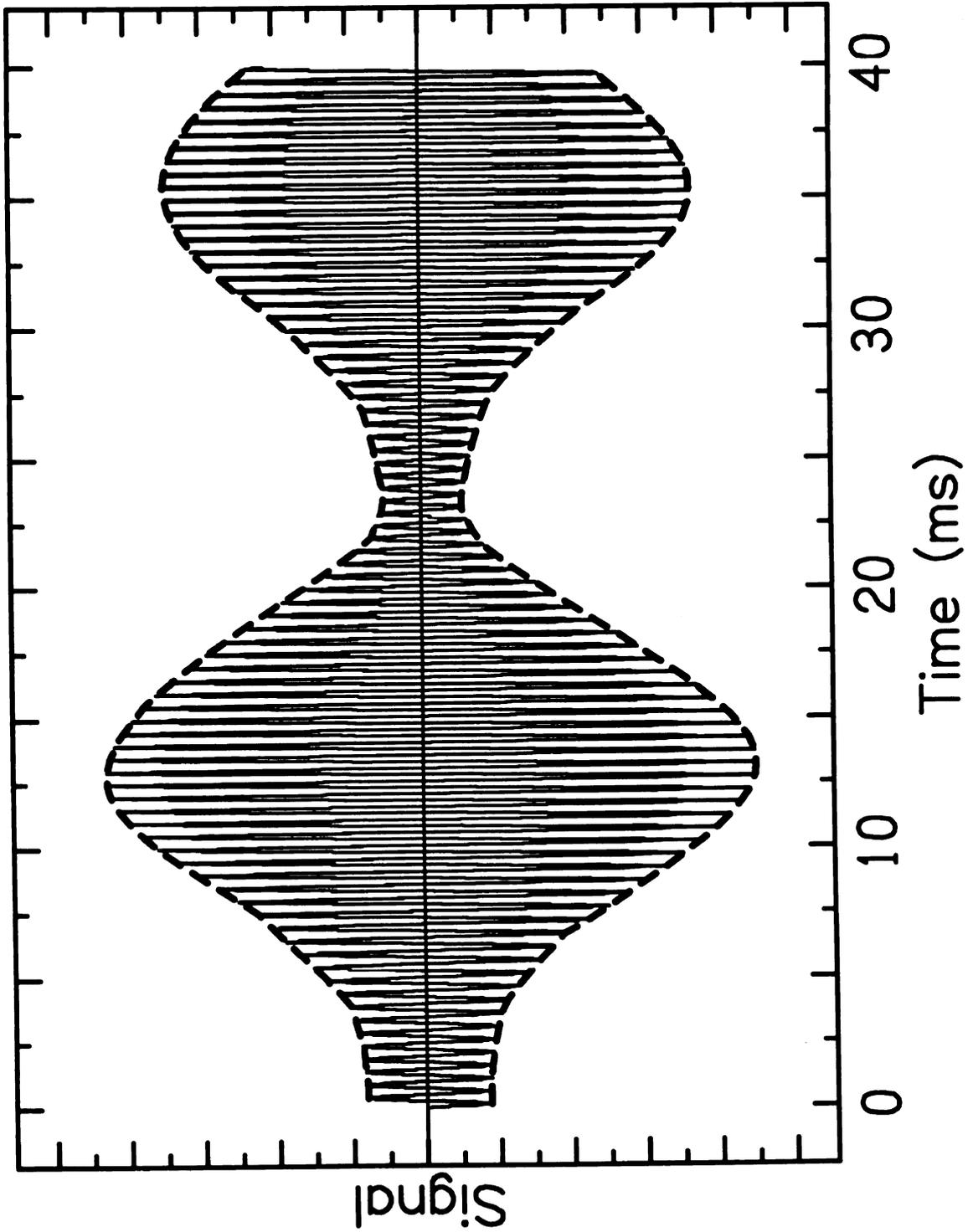


Figure II.8

Figure II.9. Results of Experiment 2: variable duration. Coherence thresholds for four listeners are divided into two plots for purposes of comparison. Both plots have a logarithmic scale of signal duration (nine values between 5 to 500 ms) on the x-axis, and the coherence necessary to discriminate from perfect coherence along the y-axis. The alternate y-axis on the right side of each plot lists the relative level of the independent noise (with respect to the coherent target noise) necessary to achieve the coherences marked. Results for listeners S and T are found in the upper graph, while those for listeners W and Z are in the lower. Error bars on data points represent two standard deviations.

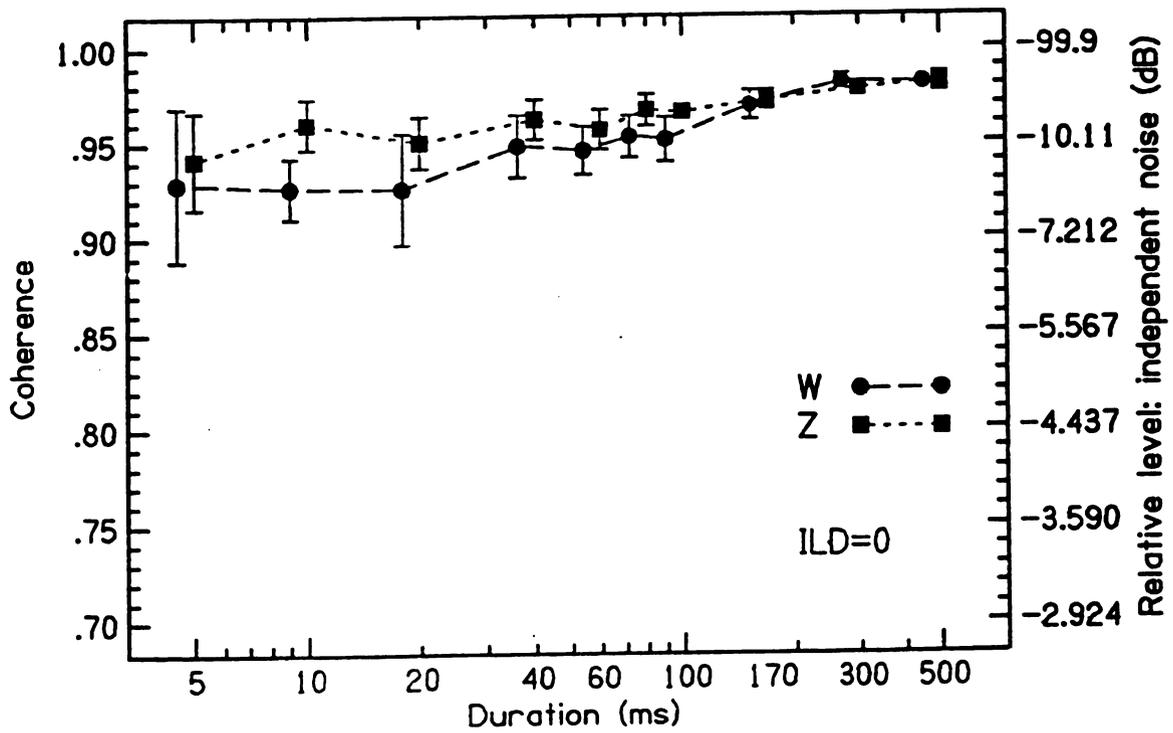
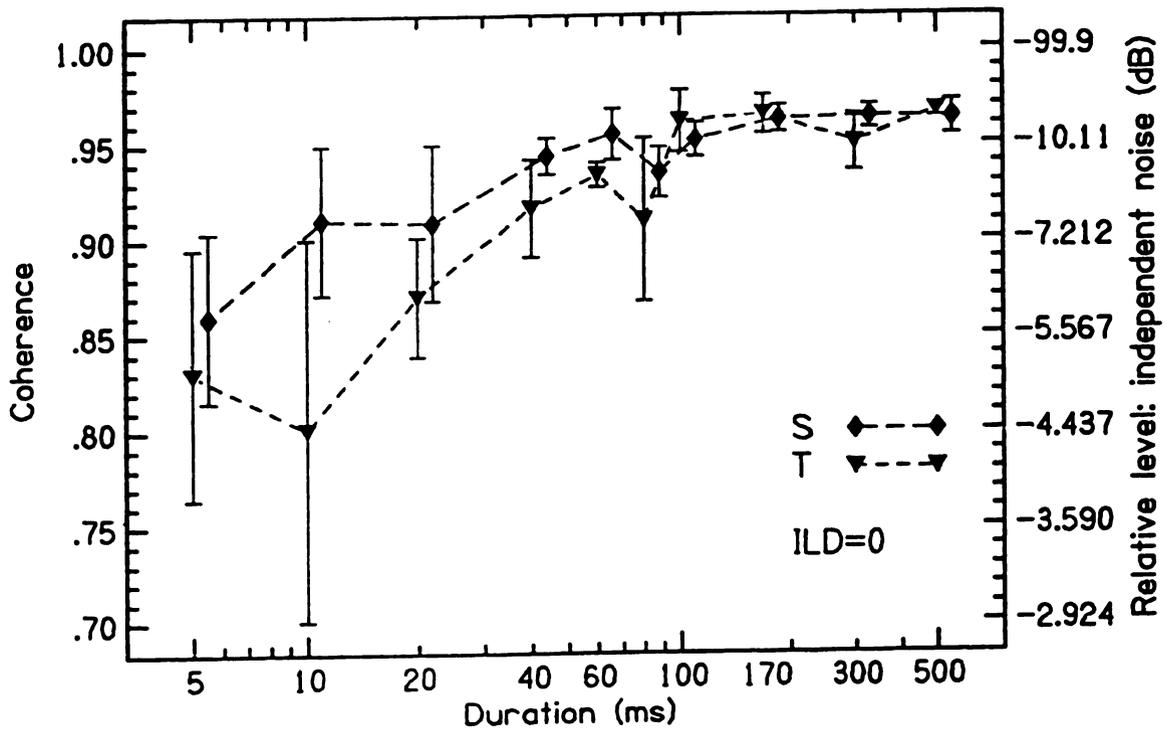


Figure II.9

Figure II.10. Results of Experiment 2: variable duration with randomized ILD. Coherence thresholds for four listeners are divided into two plots for purposes of comparison. Both plots have a logarithmic scale of signal duration (five values between 5 to 500 ms) on the x-axis, and the coherence necessary to discriminate from perfect coherence along the y-axis. The alternate y-axis on the right side of each plot lists the relative level of the independent noise (with respect to the coherent target noise) necessary to achieve the coherences marked. Results for listeners S and T are found in the upper graph, while those for listeners W and Z are in the lower. Error bars on data points represent two standard deviations.

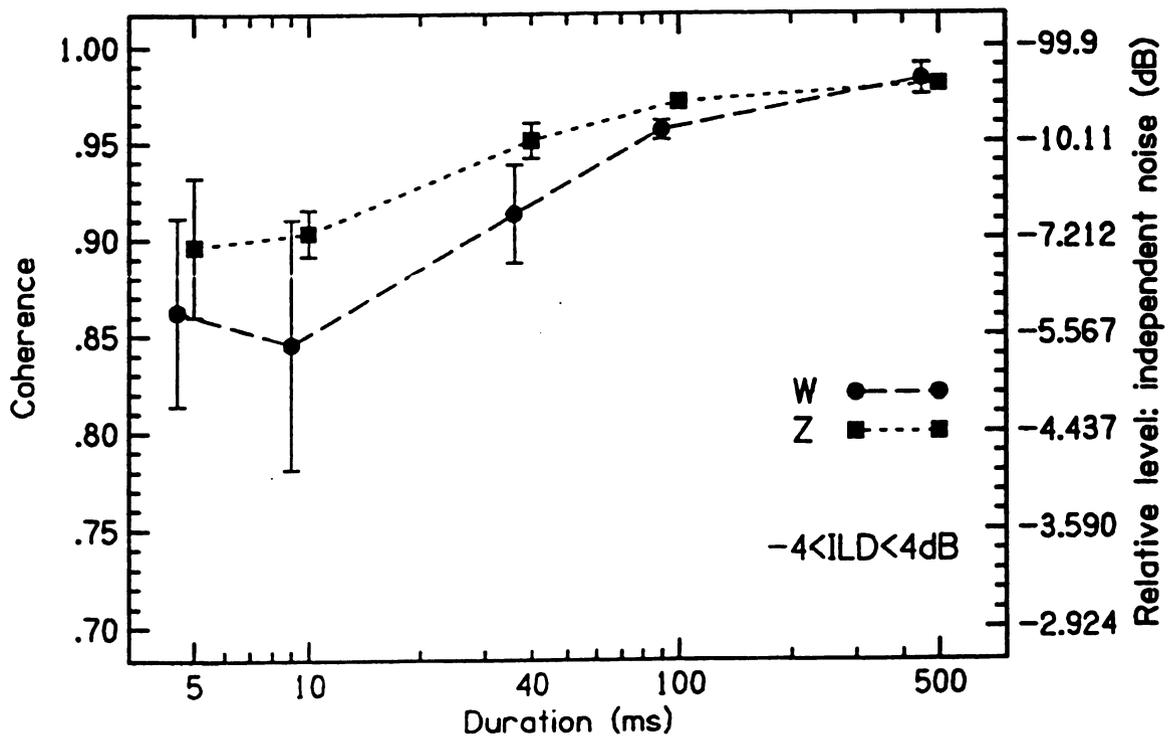
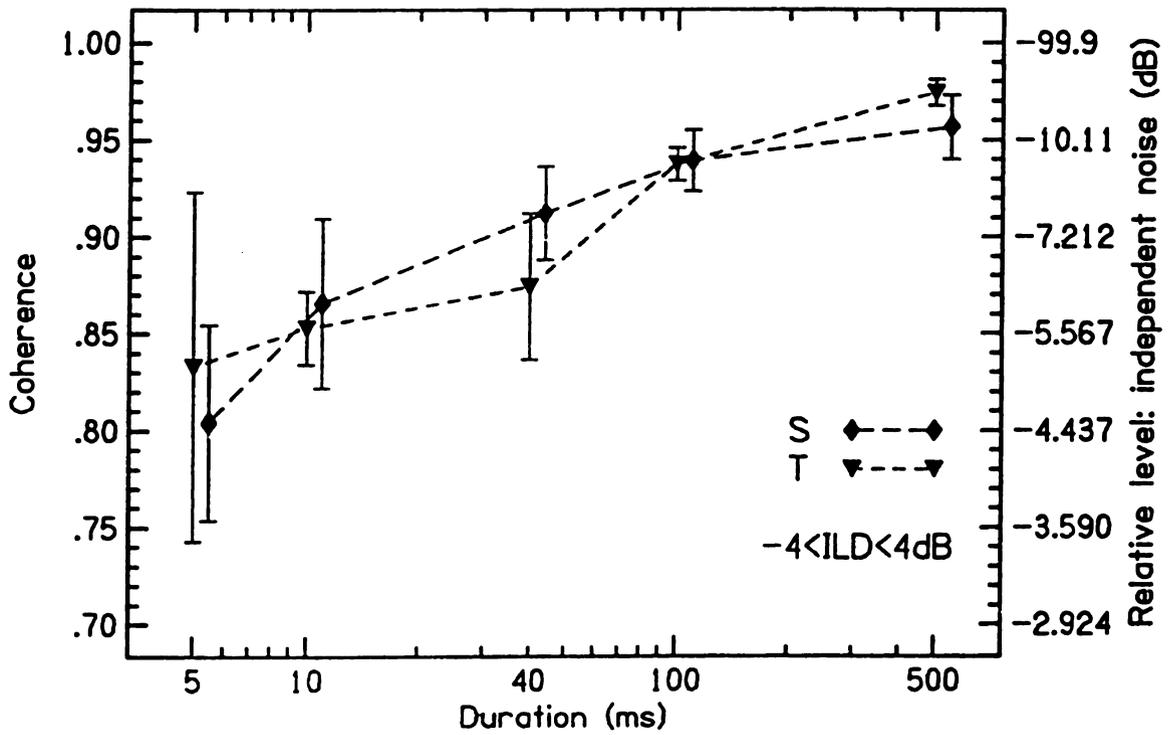


Figure II.10

Figure II.11. Results of Experiment 2: variable duration with zero ILD. This graph shows coherence discrimination thresholds for four listeners versus a logarithmic scale of signal duration (four values between 5 to 40 ms) on the x-axis. The alternate y-axis on the right side of each plot lists the relative level of the independent noise (with respect to the coherent target noise) necessary to achieve the coherences marked. Error bars on data points represent two standard deviations.

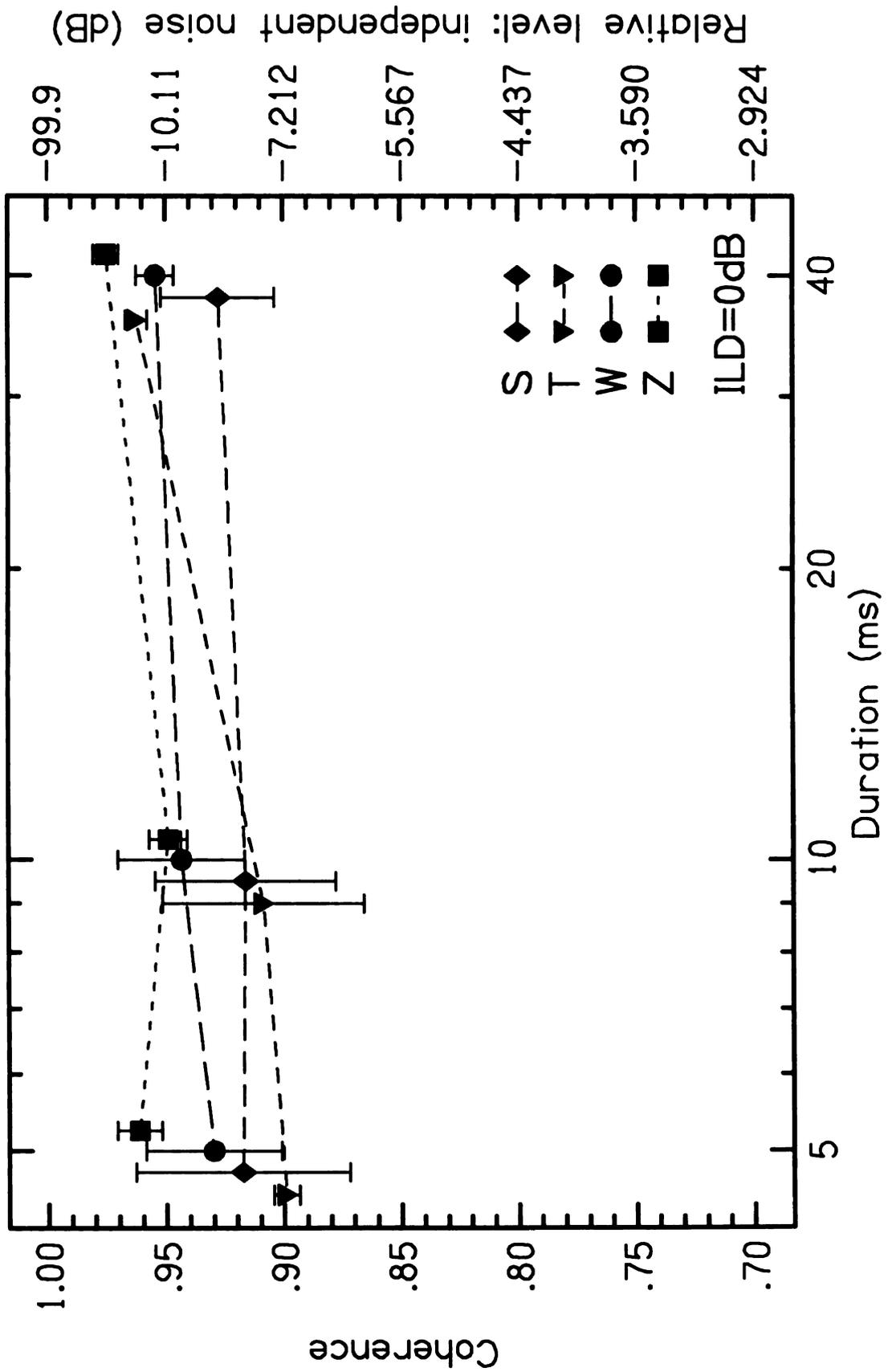


Figure II.11

Figure II.12. Results of Experiment 2: variable duration with constant ILD. This graph shows coherence discrimination thresholds for four listeners versus a logarithmic scale of signal duration (four values between 5 to 40 ms) on the x-axis. There was a constant 4-dB ILD during the experiment. The alternate y-axis on the right side of each plot lists the relative level of the independent noise (with respect to the coherent target noise) necessary to achieve the coherences marked. Error bars on data points represent two standard deviations.

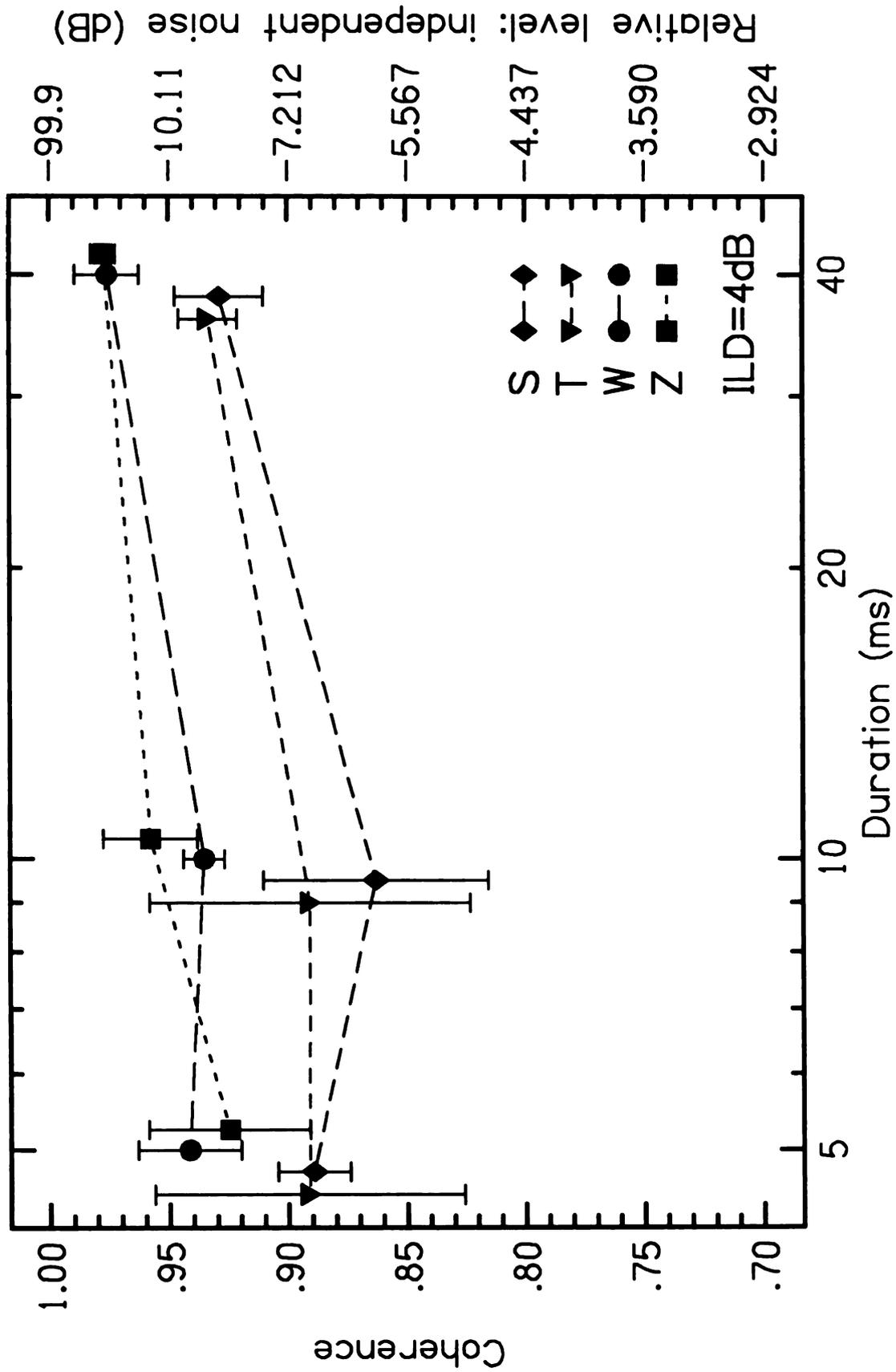


Figure II.12

Figure II.13. Results of Experiment 2: variable duration with random ILD. This graph shows coherence discrimination thresholds for four listeners versus a logarithmic scale of signal duration (four values between 5 to 40 ms) on the x-axis. There was a random ILD during the experiment, which was selected from integer values between ± 4 dB. The alternate y-axis on the right side of each plot lists the relative level of the independent noise (with respect to the coherent target noise) necessary to achieve the coherences marked. Error bars on data points represent two standard deviations.

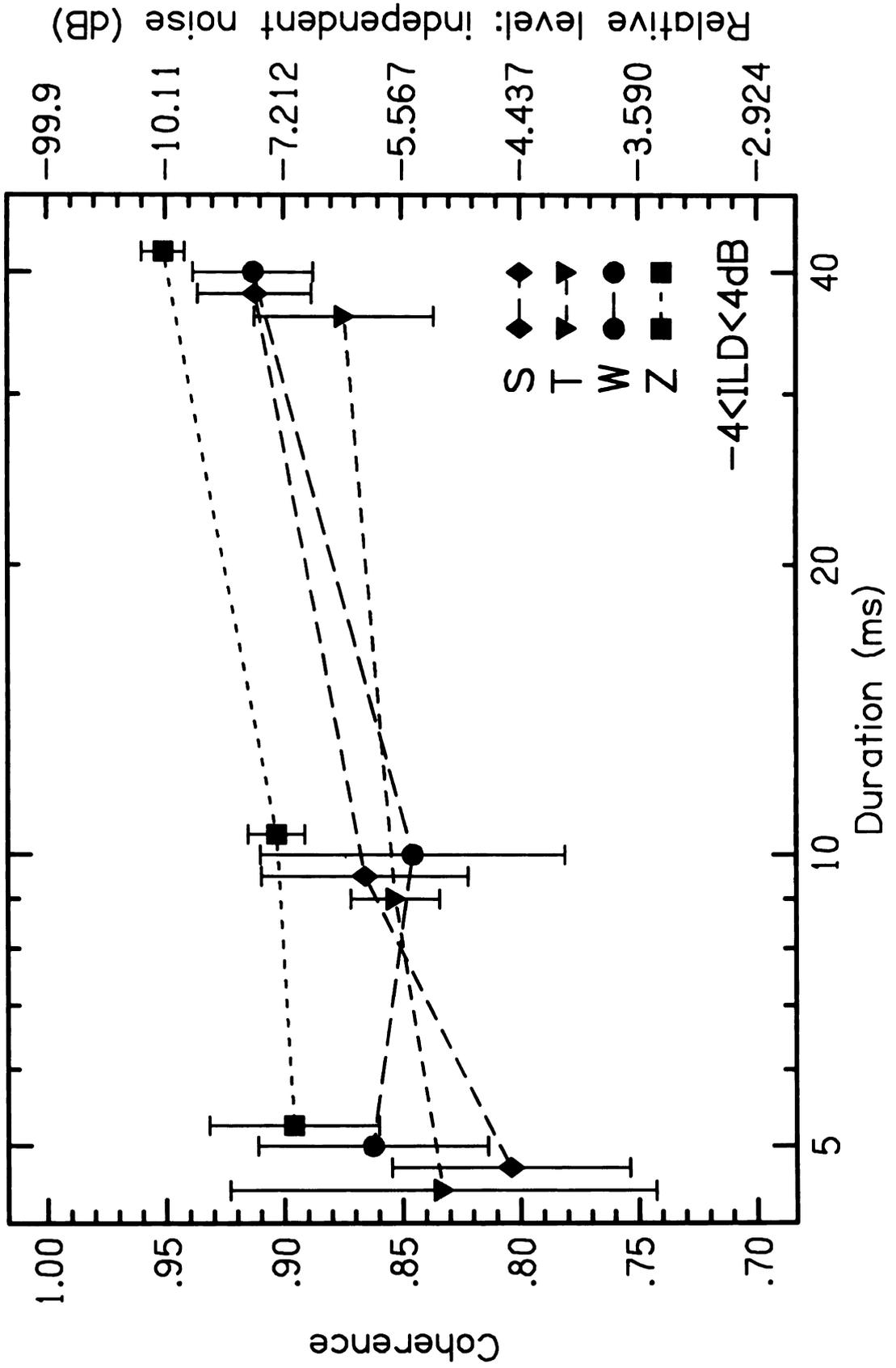


Figure II.13

Figure II.14. Graphic representation of the “fuzzy ball copier analogy.” The left column represents coherent intracranial images, while the right column represents slightly-incoherent intracranial images. The five rows model a decrease in stimulus level from top to bottom, which was simulated by lowering the darkness setting on a photocopier.

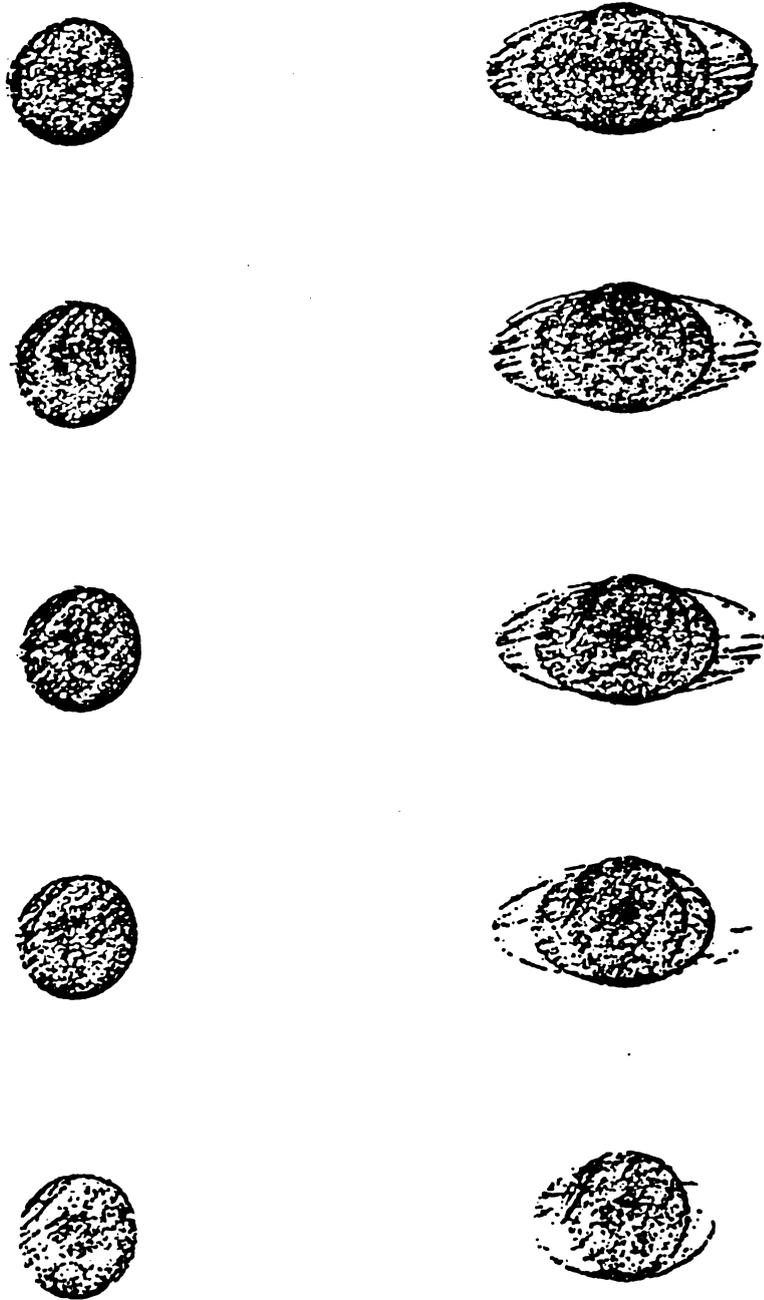


Figure II.14

Figure II.15. Results of Experiment 3: level-dependence. This graph shows coherence discrimination thresholds for four listeners versus a scale of signal level (values of 34, 44, and 64 dB SPL) on the x-axis. The alternate y-axis on the right side of each plot lists the relative level of the independent noise (with respect to the coherent target noise) necessary to achieve the coherences marked. Error bars on data points represent two standard deviations.

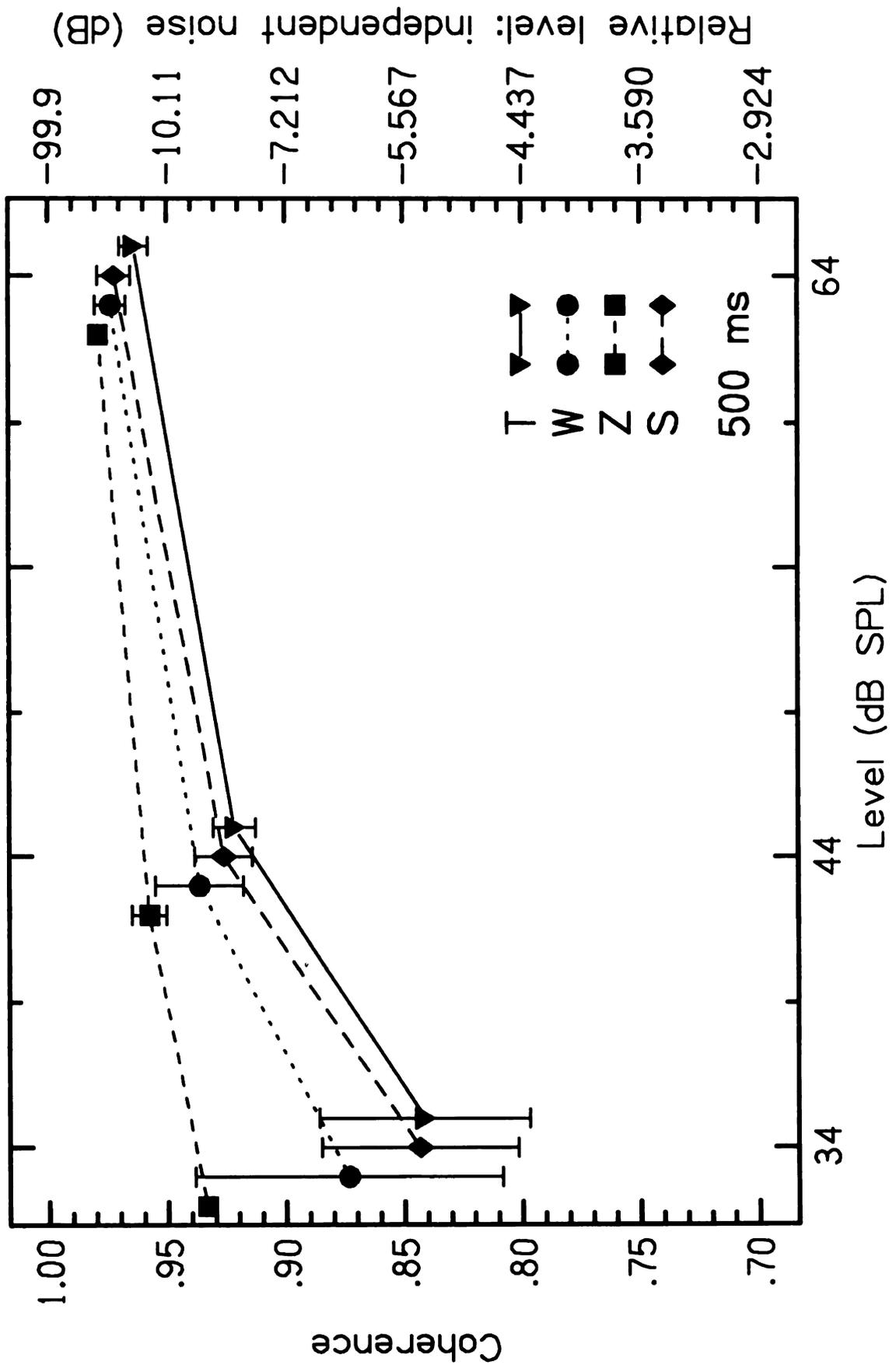


Figure II.15

Chapter III: ITD discrimination of several stimuli

Introduction

Coherence is important for a listener in a room with a sound source (see Figure III.1). The direct sound (solid line) will reach the listener unobstructed, producing a very coherent signal that also has an interaural time difference (ITD) because it reaches his left ear first. However, the listener also receives sound that has reflected off walls and other obstructions. These sounds follow trajectories that are delayed and will have various ITDs due to their multiple reflections. The combination of reflected sounds may be very random and incoherent, and interfere with the listener's ability to hear the direct sound. In a highly reverberant environment, it becomes difficult to identify the direction of the source as the direct sound is "drowned out."

As a test of listeners' ability to detect ITD, experimenters often measure their ITD resolution or just-noticeable-difference (jnd) (Durlach and Colburn, 1978). The jnd is defined as the smallest range of signal image movement along the internal delay axis (the "lateralization" axis) that a listener can reliably identify. Clearly, an image that moves 800 μs along the axis is easier to lateralize than one that moves 40 μs . This has much to do with the variability of lateral judgments. For an image that moves 40 μs , there is a higher probability of misjudging the direction of motion. See the discussion of Signal Detection Theory in the General Appendix B for more details. With the introduction of a masking noise, the task becomes even more difficult. Jeffress *et al.* (1962) asked listeners to center a correlated target noise in their heads by varying its ITD. As the

interfering masker intensified (decreasing the interaural correlation), their judgments became more erratic to the point where they selected random ITD values. Trahiotis *et al.* (2001) tested ITD discrimination for stimuli of varying “straightness” (an adjustment they made that happened to lower coherence), and concluded that ITD jnds were poorest for lower cross-correlations. The study reported in this chapter took this research a step further, with an experiment designed to establish the interaural coherence necessary to lateralize various ITDs.

Experiment 1: ITD discrimination as a function of interaural coherence (COH)

Method - Interference experiment

See Figure III.2 for a diagram of the experimental setup. The target signal was a coherent target noise (CTN) in both ears. This analog noise (N2) was generated by a Zener diode, and was white, thermal, and Gaussian. The signal bandwidth ranged from 0 to 10 kHz, where it was cut off by a Brickwall filter with -115 dB/octave rolloff. This target was delayed $x \mu\text{s}$ in the left ear, where x was selected each run from the set of 20, 40, 100, 200, and 400. The signal to the right ear was delayed either 0 or $2x \mu\text{s}$, producing an ITD of $\pm x \mu\text{s}$ that served to offset the target image to either the left or right inside the listener's head. Both delays were introduced using a BSS PCM delay line, model TCS-803, which had three taps and a resolution of $10 \mu\text{s}$.

Two other mutually-independent noise sources (N1 and N3), one for each ear, provided a 50-dB incoherent masker when mixed with the target. Both interfering masker noises were low-pass filtered, cutting off at 10 kHz with a rolloff of -48 dB per

octave. The mixed signals were sent through matched exponential voltage-controlled amplifiers (EVCAs), which were in turn ramped on and off by a trapezoidal envelope generator (Trap. Env.).

Experimental runs presented listeners with trials composed of two 500-ms intervals. The first interval would have a positive or negative ITD imposed on the CTN, offsetting the coherent image to the left or right. After a 500-ms pause, the second interval would have the ITD reversed, so the signal image would seem to move from left to right or vice versa (see Figure III.3). The listeners were faced with a two-alternative forced-choice (2AFC) task to identify the masked coherent image's direction of "motion" by pushing a button. The run did not continue until the listener had responded. There were no onset cues available for lateralization, because both left and right channels were ramped on simultaneously by matched voltage-controlled amplifiers with a 30-ms rise time. Subjects performed runs in an Acoustic Systems soundroom, listening through Sennheiser HD-480 headphones.

The intensity of the CTN was determined by a 3-down 1-up staircase as defined by Levitt (1971), so that if the listener answered correctly three times in a row, the target became softer, but just one incorrect response caused it to become louder (see General Appendix A). The target level was 60 dB at the beginning of a run, and the step size started at 4 dB, changing to 2 dB after the first turnaround. The higher initial step size allowed listeners to approach their threshold region quickly, and the subsequent step size increased precision in that region of interest. Listeners did three runs lasting two to three minutes at each of the five ITDs.

No-interference experiment

To discover how much the interfering masker noise degraded the resolution of lateral movement across different ITDs, the same experiment was performed with no masker. Once again, each listener performed three runs at the same five delays.

Absolute Threshold experiment

Finally, to see how the results from the non-interference experiment compared to absolute thresholds (the level necessary to detect the signal's presence), listeners participated in a detection experiment. The listeners were given a 2AFC task, where the intervals were defined by two lights (yellow and orange). The CTN was only presented concurrently with only one of the lights, and the listener had to identify which. All other parameters were essentially the same, except that each ITD was tested once with the left-leading delay, and then with the right-leading delay, to insure no preferential treatment of either ear. Thus, listeners performed two runs at five delays, each lasting from three to four minutes. One would expect all delays to yield similar thresholds, because there is no reason to believe that detection would depend on ITD; nevertheless, all were tested in order to be certain and averaged each delay over left- and right-leading conditions.

Listeners

The listening group was comprised of six subjects, each identified by one letter.

- B, male age 45, common middle-age high-frequency hearing loss, very experienced in hearing studies

- L, female age 19, normal hearing, previous experience in localization studies
- S, male age 17, normal hearing, previous experience in localization studies
- T, male age 18, normal hearing, previous experience in localization studies
- W, male age 58, common middle-age high-frequency hearing loss, very experienced in hearing studies
- Z, male age 24, normal hearing, previous experience in localization studies

Results

Figures III.4-9 show each listener's individual results, while Figure III.10 contains results averaged over all listeners. The individual results are intended to provide the reader with a sense of inter-subject variability. The staircase-method thresholds for the three experiments are plotted with respect to ITD. The dashed line indicates the level of the interfering masker noise on the runs in which it was present. Triangles, circles, and diamonds represent the interference, no-interference, and absolute threshold experiments, respectively. The right axis lists the coherence between left and right noises as a function of the correlated target level with respect to the uncorrelated masker level, and is only relevant for the interference results. Because the coherence scale is not linear, the coherence value corresponding to each data point can be found next to that point. Coherence is calculated according to the ratio of coherent (identical) noise power in both

ears to total incident power. Coherence is related to the threshold levels using the following derivation:

$$Coherence = \frac{P_{coh}}{P_{tot}} = \frac{P_T}{P_M + P_T} = \frac{P_T/P_M}{1 + P_T/P_M}$$

$$level \ L(dB) = 10 \log(P/P_0) \rightarrow P = P_0 10^{(L/10)}$$

$$Coherence = \frac{10^{(L_T - L_M)/10}}{1 + 10^{(L_T - L_M)/10}}$$

The subscripts T and M refer to the target and masker, respectively.

Discussion

Absolute Threshold

The plots bear out the expectation that all absolute thresholds averaged over left- and right-leading delays should be similar. There was some inter-subject variability, however. Listeners W and B's thresholds are somewhat higher than those of the other subjects, which is not surprising because they were older. However, all listeners exhibited thresholds 10-20 dB above the threshold of hearing (0 dB). We can explain this apparent insensitivity by noting that the CTN had a 10-kHz bandwidth. The threshold of hearing is based on measurements of a sine tone. Thus, the CTN had a much lower spectral density at each frequency, and the listeners had less signal power available at their most sensitive frequencies (around 4 kHz, according to equal loudness contours in Hartmann, 1997). Assuming a 2-kHz-wide band of highest sensitivity, the actual level of this band would be 7 dB below the level of the 10-kHz-wide CTN. For the younger listeners, whose thresholds fell near 10 dB, this translates to a more reasonable absolute

threshold of 3 dB. Once satisfied that these detection thresholds met expectations, one can use them as a basis of comparison for the discrimination data garnered from no-interference runs.

No-interference

Note that the no-interference curves do not parallel the absolute threshold curves and have significantly higher thresholds. All listeners follow the same trend of higher thresholds at the lower delay times, with a monotonic and gradual shift as the ITD decreases. At 400 μ s ITD, the lateralization (non-interference) thresholds are still 10 dB higher than those necessary to detect the CTN (absolute thresholds). In order to see if the coherence thresholds vary significantly as a function of ITD shift, analysis of variance (ANOVA) was employed. The test shows a significant difference overall ($F(4,20) = 30.343, p < 0.001$), and so Experiment 1 finds that the magnitude of ITD shift does affect discrimination, requiring higher coherences at smaller shifts. This result is a natural consequence of Signal Detection Theory (see General Appendix B), which states that uncertainty in each ITD measurement will limit listeners' ability to correctly identify the direction of shift. It is quite likely that lower stimulus levels introduce more ITD uncertainty.

Interference

The interference-case data show a very similar behavior to the non-interference thresholds with respect to ITD magnitude, although the thresholds are consistently higher than those for the no-interference case. The higher thresholds show that the 50-dB

masker degrades the ability to distinguish direction. As a check whether the two curves are truly parallel, each was averaged over the five ITDs. By subtracting the difference between the two means from each interference threshold, the interference curve was shifted down to the non-interference level. The resulting curves may be essentially identical. The ANOVA test on the modified results ($F(4,20) = 0.652$, $p = 0.632$) concludes that there is no difference, and thus, the average listener may have a constant difference between interference and non-interference thresholds. This parallelism between curves indicates that the external and internal interfering noises behave similarly. The internal, neural noise that represents the randomness of binaural processing produces the same ITD-dependent effect at a lower threshold. This result is an integral assumption of many binaural-interaction models (Colburn and Durlach, 1978).

Experiment 2: ITD discrimination as a function of high-pass cutoff (COB)

Introduction-The salience of high-frequency ITD cues

It has been previously noted by numerous psychoacousticians that the interaural time difference (ITD) information provided by high-frequency tones is useless to the auditory system (Durlach and Colburn, 1978). This is true simply because the waveform is cycling so rapidly, and peaks are so close together in time, that a given delay at high frequencies could cause many full cycles of phase shift (e.g. θ , $\theta+2\pi$, $\theta+4\pi\dots$).

Physiologists have also noted that auditory nerve spikes lose phase-locking with high-frequency waveforms (Pickles, 1982), which could render the auditory system unable to identify an interaural delay. ITD information is too uncertain at high frequencies to be an

accurate tool for lateralization. Previous papers have estimated that fine-structure ITD cues lost their salience above 1500 Hz (Durlach and Colburn, 1978), which is in agreement with Rayleigh's (1907) duplex theory.

As reported by McFadden and Pasanen (1978) and Henning (1980), it is possible for human listeners to lateralize high-passed noise on the basis of interaural time differences in the envelope. High-frequency narrowband signals featuring rapidly-varying waveforms can still have slowly-varying envelopes, which are just larger constructs composed of the waveform peaks. See Figure II.9 in Chapter II for an example. Since envelope structures may have a much longer period, the auditory system can extract meaningful ITD cues even when fine-structure cues are unusable. However, the ability to do so makes great demands on the interaural coherence of the noise, as found by Hartmann, Constan and Rakerd (1998).

High-frequency ITD cues in the room environment

Experiment 1 verified that smaller shifts in ITD required more coherence to lateralize in the broadband regime. Experiment 2 involves expansion of the "head in a room" model in Figure III.1. An interesting aspect of the reflected sound is that all surfaces have an absorption coefficient, which tends to increase with frequency. Thus, the reflections are low-pass filtered somewhat, and sound in the higher frequency bands will therefore be more coherent than lower-frequency components. To further test the limits of the auditory system, and more accurately represent the room conditions described above, a new experiment featured a high-pass-filtered coherent stimulus, effectively eliminating fine-structure ITD cues.

The expanded concept of the room environment inspired a new study much like Experiment 1. The new stimulus consisted of a broadband incoherent noise, representing the reflections and any ambient interference, and a high-pass filtered CTN with cutoff frequency f_c , representing the direct sound. The coherent portion also had an applied ITD. Using this revised stimulus, the goal of the experiment was to determine the coherence threshold required to lateralize a high-pass-filtered target noise in the presence of a broadband masker.

Method

Refer to the diagram of the stimulus (Figure III.11), composed of two parts for each ear, left and right. The stimulus was a coherent noise (A) that was 60 dB SPL in the broadband case. Its spectrum ranged from the high-pass cutoff to 10 kHz. Spectral density was held constant throughout the experiment, so the total signal power varied with cutoff frequency f_c . Using a Stanford high-pass filter with -115 dB/octave rolloff, Experiment 2 set f_c to seven values: 0, 500, 1000, 1500, 2000, 3000, and 4000 Hz. One should recognize that for $f_c < 1500$, the listener still had access to fine-structure ITD cues in the target. However, when the cutoff was 1500 Hz or greater, only the envelope structure provided ITD cues. This noise was digitally generated, and was white, thermal, and gaussian. A delay was imposed on the left or right ear to produce an ITD. Two thermal noise sources (B and C), one in each ear, provided the incoherent masker. These noises were independent of noise A and each other. Their spectra ranged from 20 Hz to 10 kHz. Through the use of programmable attenuators, the experimental software was able to set their levels between 40 and 80 dB.

At first inspection, it would seem that the coherence of the high-pass-filtered target in a broadband masker would be lower than for a broadband target, since there is less power in the high-passed signal. According to a conventional measurement, the overall coherence is indeed lower...however, the binaural processor analyzes the broadband input in a different manner. According to critical band theory (Moore, 1995) the auditory periphery separates stimuli into discrete bands that are processed separately. Thus, only masker power in the same bands as the target noise provides any interference. Coherences calculated within each band better represents the information available to the auditory system, and tend to be higher than broadband values. This approach will be reviewed in more detail later in the chapter. Experiment 2 extends the masking noise below the target's high-pass cutoff simply to better simulate the expanded room environment model.

During the run, stimulus coherence was controlled by again varying the relative levels of target and masker according to a 3-down, 1-up Levitt staircase function. Unlike Experiment 1, the masker level was the staircase variable (due to setup requirements), though this makes no difference in the coherence calculation. The change is significant because in Experiment 1, the CTN level tended to be lower than or equivalent to the masker, and so when its level moved on the staircase the overall level did not change much (a desirable result). In Experiment 2, the masker levels are low enough with respect to the CTN that choosing them as the staircase variable is the best way to maintain consistency in overall level. If the listener answered correctly on three successive trials, the program raised the incoherent masking noise level by 2 dB, lowering the overall stimulus coherence and making the task more difficult. If listener

answered incorrectly, the program lowered the masking noise level by 2 dB, increasing coherence and making the task easier. Runs continued for 14 turnarounds with the coherence value recorded each time. The thresholds were an average of the last 10 coherences.

Between runs, the experimenter had control of two parameters. They were the same five ITDs used in Experiment 1 (20, 40, 100, 200 and 400 μ s), in addition to seven high-pass cutoffs (0, 0.5, 1.0, 1.5, 2.0, 3.0 and 4.0 kHz). The experimenter randomly varied these parameters so no two successive runs had identical stimuli. All listeners did three runs at each combination of ITD and high-pass cutoff, except for those combinations at which they failed on three runs (defined as requiring a coherence $> .99$ to do the task). Thus, their input totaled a maximum of 105 runs. Four experienced listeners participated in this experiment, including the previously-mentioned T, W and Z. A new listener was added for this study:

- M, female age 21, normal hearing, no previous listening experience

Results and Discussion

Average coherence thresholds for all four listeners can be found in Table III.1 below. They are grouped by cutoff values for a particular ITD.

Table III.1 - Coherence thresholds by ITD and cutoff					
ITD	HP cutoff	M	T	W	Z
20 us	0	0.772129	0.414501	0.830445	0.328759
	500	0.802909	0.528751	0.827178	0.344179
	1000	0.888184	0.57428		0.403371
40 us	0	0.359935	0.236075	0.454078	0.15098
	500	0.403371	0.257464	0.42572	0.189905
	1000	0.655821	0.375998	0.90717	0.212073
	1500		0.957242		0.839951
	2000				0.901187
100 us	0	0.223845	0.109549	0.20076	0.063327
	500	0.253087	0.121295	0.261891	0.10299
	1000	0.618584	0.197091	0.517263	0.10299
	1500	0.833662	0.755519	0.823862	0.681322
	2000	0.953309	0.823862		0.72911
	3000		0.910976		0.823862
200 us	0	0.189905	0.109549	0.20076	0.054413
	500	0.13411	0.066114	0.15098	0.05561
	1000	0.27089	0.15098	0.42572	0.09283
	1500	0.738109	0.540208	0.763925	0.4201
	2000	0.78013	0.613137	0.738109	0.634743
	3000	0.94676	0.86589	0.914641	0.629389
	4000		0.966488		0.79552
400 us	0	0.148052	0.06755	0.13411	0.042758
	500	0.28946	0.063327	0.176138	0.066114
	1000	0.585499	0.261891	0.540208	0.333861
	1500	0.557312	0.488489	0.528751	0.414501
	2000	0.645352	0.482737	0.607661	0.197091
	3000	0.772129	0.751241	0.681322	0.505756
	4000	0.86589	0.84902	0.876229	0.700977

Coherence thresholds for various ITDs and cutoffs in Experiment 2

Note that the shorter-ITD data do not extend to all seven cutoff values, and some cells are blank. This is because at several conditions, listeners could not complete the task regardless of the interaural coherence. After three failures, that combination was

dropped. Some listeners were successful where others failed, but all ran into limitations at some point. The combination of short ITDs and high cutoff values proved too difficult overall.

To facilitate a clear understanding of the relationship between coherence, frequency distribution, and ITD lateralization, Figure III.12 contains data averaged over all listeners and plotted on a graph of coherence versus high-pass cutoff. The x-axis is the range of high-pass cutoffs, from 0 to 4 kHz. The y-axis is the interaural coherence threshold necessary to lateralize with 79% accuracy, as targeted by the staircase. The five plots are for the various ITD values. While taking the average threshold may eliminate some of the subtler effects within each listener's results, those results all shared a common behavior much like that shown in Figure III.12. Also note that the graph only includes points where data were obtained for all listeners. If one or more listeners failed at a particular condition, the experimenters did not average over the other listeners' results to establish a threshold.

Some general trends in the data are readily apparent. Runs with the longest ITD value tended to require the least coherence for adequate performance. While there are exceptions to that rule, the 400 μ s plot is clearly well below that of the 20 μ s condition. Attending to the finer details reveals that the ITD dependence of thresholds is nowhere near as smooth as in Experiment 1. Observe all points with 500- or 1000-Hz cutoffs: the threshold does not increase monotonically with ITD. The reversal of 200- and 400- μ s points at a 500-Hz cutoff is almost negligible. However, their positions at a 1000-Hz cutoff, where every listener required a lower coherence to discriminate 200- μ s ITDs, cannot be ignored. It is counter-intuitive that listeners might lateralize a smaller ITD

more effectively, a fact in opposition with nearly all other data on Figure III.12. A possible explanation arises if listeners tended to concentrate on the lowest available frequencies (i.e. those near the cutoff) when discriminating ITD. A time shift between $\pm 400 \mu\text{s}$ at 1000 Hz corresponds to a phase shift of nearly π , and interaural phase differences of almost half a period lead to confusion and reversal of the image along the internal delay axis. Thus, one might expect listeners to have more difficulty at that condition.

Runs with the lowest high-pass cutoffs tended to require lower coherence thresholds. Note the shortest ITD, 20 μs : once the high-pass cutoff exceeded 500 Hz, listeners went above the 99% coherence threshold quickly, and so this plot drops off the top of the graph (indicating failure). Longer ITDs do the same at higher cutoffs: listeners fail at 40 μs ITD above 1 kHz, 100 μs above 1.5 kHz, 200 μs above 3 kHz, and listeners could still lateralize a 400 μs ITD for a cutoff of 4 kHz. Also note that all listeners exhibited maximum slope in the vicinity of 1kHz, a region near the frequency limit for fine-structure ITD cues.

It would seem that envelope ITDs proved quite useful, since listeners could still perform the task for cutoffs of 1500 Hz and above. However, the required interaural coherence grew rapidly as the high-pass cutoff frequency increased through the critical region from 1 to 4 kHz, especially for the smaller ITD shifts. Thus, it appears that envelope cues are very sensitive to coherence, indicating that they are more vulnerable to masking and reverberated sounds. Fine-structure ITD cues, on the other hand, were still effective at very low coherences. This has implications for the localization of broadband sounds in rooms where the interaural coherence tends to increase with increasing

frequency, but possibly not rapidly enough to allow the envelope timing information to contribute usefully. The ability to lateralize could depend very much on the room characteristics.

Experiment 3: ITD discrimination of high-pass-filtered targets and maskers (COD)

Introduction

In Experiment 2, the CTN was high-pass filtered, while the masker was always broadband (20-10000 Hz). The calibration procedure required the experimenter to keep the target's spectral density constant, which led to different total powers depending on the cutoff (see the top half of Table III.2). This procedure was justified by citing critical band theory, which states that masker power outside the target band should not contribute to the incoherence. However, the auditory filters measured by several experimenters do not feature the ideal bandlimits that our program was able to achieve. The cutoffs are much more gradual (Moore, 1995), especially for frequencies below the band. This inaccuracy leads to an effect called the "upward spread of masking." Hartmann (1997) shows how a lower-frequency component can excite fibers of much higher CF, effectively masking signals of higher frequency. Trahiotis *et al.* (2001) found that interference by adjacent bands reduced the binaural release from masking in an MLD experiment centered at 500 Hz. Their listeners' thresholds decreased as they widened the presented band, and continued to worsen as the bandwidth exceeded the critical bandwidth of 100 Hz.

Table III.2 – Calibration settings						
	Target cutoff	Target level	Target spec lvl	Masker cutoff	Masker cal lvl	Masker cal spec lvl
EXP 2	0	60.0	20.0	0	60.0	20.0
	500	59.8	20.0	0	60.0	20.0
	1000	59.5	20.0	0	60.0	20.0
	1500	59.3	20.0	0	60.0	20.0
	2000	59.0	20.0	0	60.0	20.0
	3000	58.5	20.0	0	60.0	20.0
	4000	57.8	20.0	0	60.0	20.0
EXP 3	0	60.0	20.0	0	60.0	20.0
	500	60.0	20.2	500	60.0	20.2
	1000	60.0	20.5	1000	60.0	20.5
	1500	60.0	20.7	1500	60.0	20.7
	2000	60.0	21.0	2000	60.0	21.0
	3000	60.0	21.5	3000	60.0	21.5
	4000	60.0	22.2	4000	60.0	22.2

Comparison of overall and spectral levels set by calibration routines in Experiments 2 and 3 as a function of target and masker cutoff frequencies

The third and final experiment presented listeners with a different stimulus--one that limited both target and masker to the same frequency range, so their respective levels were identical on initial calibration (compare the bottom half of Table III.2 with the top). Experiment 3 is useful as a check on our hypothesis that the masker power below the target bands in Experiment 2 did affect the lateralization threshold. If this hypothesis is true, then the inclusion of upward spread of masking (from the low-frequency portion of the masker present in Experiment 2) indicates we should see lower coherence thresholds (i.e. higher masker-level thresholds) for Experiment 3 at higher cutoffs. If not, then the threshold spectral level for a particular high-pass cutoff should be the same in both experiments.

Method

Experiment 3 reran the 200- μ s ITD condition of Experiment 2, with two significant differences. First, the bandlimits have changed as noted above, so that both target and masker are filtered identically. Second, this filtering is done when both noises are generated digitally, to ensure accuracy. The program arranges a frequency spectrum in the Tucker-Davis Array Processor, assigns random phases to each frequency, and constructs the noise with an inverse FFT. The experimenter's only responsibility is to ensure that each run starts at the same level: 60 dB SPL for both noises, or 63 dB for the combination. During the run, the program uses the same level staircase (2-dB steps) as Experiment 2, but it calculates the relative intensity of the incoherent noise and scales the digital waveform appropriately, so the Programmable Attenuators are unnecessary in this particular task.

Three listeners participated, including the aforementioned W and Z. A new listener was also included, and he also did runs of the 200- μ s Experiment 2 condition in order to provide comparison data:

- X, male age 26, normal hearing, previous experience in localization studies

Results

Table III.3 includes masker-level thresholds for all three listeners in both experiments, arranged according to high-pass cutoff of the masking noise. The calculated spectrum levels corresponding to each threshold are listed as well. When comparing the results of Experiments 2 and 3, one might note how they are similar when the masker

cutoff is low in Experiment 3, and deviate from each other as that cutoff increases. These data were converted to coherence values and plotted in Figures III.13-15, one for each listener, then averaged over listeners in Figure III.16. The x-axis charts high-pass cutoff, while the y-axis measures the threshold coherence at that condition. Thus, when target and masker are at equal spectrum level, data points fall along the “0.5” line. The figures clearly illustrate a strong divergence that begins near 1500 Hz (or earlier, as in the case of listener X), a region that has some significance for fine-structure and envelope ITD cues as mentioned earlier. The difference is rather clear and dramatic for each listener.

Table III.3 – Results comparison								
	Target cutoff	Masker cutoff	Masker threshold W	Masker spec lvl W	Masker threshold X	Masker spec lvl X	Masker threshold Z	Masker spec lvl Z
EXP 2	0	0	66.0	26.0	66.7	26.7	72.4	32.4
	500	0	67.5	27.5	64.8	24.8	72.3	32.3
	1000	0	61.3	21.3	59.9	19.9	69.9	29.9
	1500	0	54.9	14.9	56.9	16.9	61.4	21.4
	2000	0	55.5	15.5	56.8	16.8	57.6	17.6
	3000	0	49.7	9.7	53.0	13.0	57.7	17.7
	4000	0			51.7	11.7	54.1	14.1
EXP 3	0	0	64.8	24.8	68.3	28.3	73.8	33.8
	500	500	65.5	25.7	66.0	26.2	73.3	33.5
	1000	1000	60.9	21.4	63.3	23.8	69.7	30.2
	1500	1500	59.1	19.8	59.7	20.4	62.3	23.0
	2000	2000	59.4	20.4	59.7	20.7	62.1	23.1
	3000	3000	58.4	19.9	59.2	20.7	62.1	23.6
	4000	4000	57.3	19.5	57.6	19.8	61.4	23.6

Comparison of threshold overall and spectral levels in Experiments 2 and 3

In Experiment 3, listeners were relatively unaffected by increases in cutoff frequency above the 1500-Hz limit. This would seem to indicate that once fine-structure ITD cues have been eliminated, the usefulness of envelope cues is relatively unaffected by the loss of further high-frequency bands. However, the additional low-frequency masker in Experiment 2 seems to degrade thresholds at higher cutoffs. As the target

cutoff increases, the loss of bandwidth is compounded by the remaining masker power in frequencies removed from the target. The dependence of threshold on cutoff in Experiment 2 supports the hypothesis that listeners experience upward spread of masking in this experiment.

Discussion

Analysis with respect to known models

Experiments 2 and 3 tested the ability of listeners to follow the movement of a high-pass target coherent noise in an incoherent masker. It is clear that the task becomes more difficult as the signal-to-noise ratio decreases, but how the auditory system's decision-making process is affected by increasing the high-pass cutoff is not as obvious. Experiment 2 revealed upward spread of masking, and established a general behavior for thresholds at various ITDs and cutoffs. How do these findings fit into known models of the auditory system?

The Jeffress model

These experiments are intimately connected to Jeffress' cross-correlation model of binaural perception. The results clearly show that increasingly difficult lateralization tasks (e.g. shorter ITDs) require a higher amount of interaural coherence to maintain the same level of performance. The coherence is represented in the Jeffress model by coincidence counters, and a higher coherence threshold means that the signal from one of the coincidence detectors must be stronger to establish a reliable indication of ITD, or

else the broad incoherent noise (which produces random coincidence data) will effectively mask the coincidence signal. Physiologists have noted a high concentration of comparatory cells sensitive to low-frequency sound in the medial superior olive (Yin and Chan, 1990), which would support the notion that ITD is primarily a low-frequency cue. When lateralizing high-pass signals with increasing cutoff frequencies, one does still have access to envelope ITDs, but has lost the waveform cue. To produce an equally strong coincidence signal for ITD lateralization, a high-frequency signal should require increased interaural coherence.

The auditory nerve model

The auditory-nerve-based model of Colburn (1977) presents an effective way to predict how the addition of incoherent noise should affect ITD discrimination. It assumes that neural firing rates in the auditory nerve are distinguished by a Poisson process, depending on the characteristic frequency (CF) of the particular fiber and temporal features of the presented stimulus. Colburn's approach thus incorporates an internal, neural noise in addition to the external masking noise. Fibers of the same CF from left and right ears feed into a set of coincidence counter neurons distinguished by different delay lines, which operates much like the matrix of counters described by Jeffress (Colburn and Durlach, 1978). The model assumes that coincidence is a function of simultaneous firings of fibers from left and right, and thus, the counter output is another Poisson process that depends on the interaural lag and a time window. This output, when measured across several counters featuring different lags for one CF, forms a crude representation of the interaural cross-correlation for the neural signal at that frequency.

The auditory system can identify ITD by locating the lag with the most coincidences, which corresponds to the cross-correlation peak. Incoherent noise adds a random element to the Poisson firing of left and right ears, necessarily reducing the coincidence rate and lowering the peak. The flatter cross-correlation function introduces uncertainty to the ITD decision variable.

The equalization-cancellation model

The equalization-cancellation (E-C) model of Durlach (1972) suggests that the auditory system's ability to detect a signal depends on noise suppression through various attenuations, time-shifts, and subtractions. It also includes the concept of neural "jitter," which provides an internal noise that prevents the auditory system from achieving perfect performance. The jitter in the time domain is assumed to be independent of frequency. At higher frequencies, the constant time jitter translates into a large variability in phase, and thus the internal noise is much more effective. In this manner, the jitter introduces more incoherence to the system. Note that the E-C model is typically applied to MLD tasks rather than lateralization, but its definition of neural jitter would suggest that ITD discrimination should weaken as the high-pass cutoff increases.

The position-variable model

An updated version of the Jeffress model, the Colburn-Stern position variable model (Stern and Colburn, 1978), predicts the location of a lateralized image by computing the center of mass of all the coincidence-counter outputs. Because of its relation to the Jeffress model, the position variable model relates to this chapter's results

in much the same way: a higher value of coherence will produce a larger response from the coincidence counters. This sharper peak leads to a better-defined centroid, and improves lateralization. The use of the centroid as a decision variable is more robust than simply including the cross-correlation peak. However, this model has no explicit elements to include the contribution of the envelope, which leads us to the next model.

The coherence-based model

Bernstein and Trahiotis (1996) have developed a model in which the auditory system detects target signals by their normalized cross-correlation, after the auditory stimulus has been low-pass filtered. The filter has no effect on low-frequency signals, but as the frequency increases, it serves to extract the envelope of the waveform while eliminating the fine structure that conveys no useful ITD information. This model accounts for envelope-based ITD cues, predicting that high-pass filtering will still allow the listener to complete the task.

This section analyzes the results of Experiments 2 and 3 using the model developed by Bernstein and Trahiotis and their colleagues (Bernstein and Trahiotis, 1996; Trahiotis *et al.*, 2001), as well as Colburn, Stern, Kohlrausch and all their colleagues (Colburn, 1977; Stern and Colburn, 1978; Kohlrausch *et al.*, 1997). This model has become a standard for narrowband binaural hearing, and has been applied to Masking Level Difference tasks that involved detection of a tone in narrow-band noise (van der Heijden and Trahiotis, 1997), as well as narrow-band lateralization tasks (Trahiotis *et al.*, 2001). However, it could be paired with a wide variety of phenomena, providing a “unified” model of processing. The following steps employ this model in a broadband

situation, separating the stimulus into critical bands before processing so each band was similar to the prior narrow-band stimuli used in the model. This separation into narrow bands agrees with many disparate models of binaural interaction (Colburn and Durlach, 1978), which share the common element of bandpass filters at the auditory input. By making calculations of normalized correlation and establishing decision criteria, the model calculation attempted to predict the 200 μ s data of Experiments 2 and 3.

Figures III.17 and III.18 are a road map of the model calculation process. It began with the generation of stimulus files according to specifications in Experiment 3, with high-pass-filtered coherent targets and incoherent maskers. The discussion of calculations involving Experiment 2 stimuli will follow later in this chapter. The relative levels of targets and maskers were determined by the coherence value alpha. The process involved separation of the stimuli into 15 odd-numbered critical bands from number 1 to 31, including center frequencies from 124 to 7021 Hz. To ensure that the model could be applied generally to both Experiments 2 and 3, allowing for the inclusion of upward spread of masking, critical band filters were designed to fit the profile of an actual auditory filter. A formula for the “roex(p)” auditory filter shape $H(f)$ is found in Moore (1995).

$$|H(f)|^2 = (1 + pg)e^{-pg}$$

The roex (rounded exponential) function falls off exponentially with deviation from center frequency $g=|(f-f_c)/f_c|$, and uses parameter p to define the width of a critical band (limiting the area under the curve to approximate the bandpass filters that describe critical bands). Note that at higher CFs, this filter includes a much wider frequency band due to the normalization according to CF. Filters near the target noise’s

high-pass cutoff would include masker power below the cutoff frequency, and the increased incoherence due to additional masker power would only rise as the high-pass cutoff increased. Thus, the coherences for the higher cutoffs in Experiment 2 may have been significantly less than we assumed. By including filters with a gradual rolloff, the model has a somewhat accurate representation of the auditory system, and should be able to simulate the case of Experiment 2 successfully. The filters are also applicable to the Experiment 3 stimulus because the shape should have no effect when the target and masker have identical bandlimits.

After rectifying each band's left and right channels according to a half-wave square-law, the model process included a stage of low-pass filtering according to the formula used by Bernstein and Trahiotis (1996) (see bottom of Figure III.17).

$$G(f) = \sqrt{\frac{1}{1 + \left(\frac{f}{f_0}\right)^{n/2}}}$$

$$f_0 = \frac{f_c}{\sqrt{2^{1/n} - 1}}$$

Bernstein and Trahiotis selected parameters $n = 4$ and corner frequency = 425 Hz, which best fit their data. The same parameters were adopted for this calculation. At their suggestion, a subsequent low-pass filter was introduced as a further development of the model. This additional filter was second-order Butterworth with a cutoff frequency of 150 Hz. It was applied only to bands with center frequencies falling above 1500 Hz. All of the processes thus far applied to the stimulus are meant to simulate the actions of the auditory periphery on a real-world source.

Following the filter stage, the model proceeded to cross-correlate left and right signals over internal lag τ from -4 ms to +4 ms without normalization (top of Figure III.18). Note that the cross-correlation function extends well beyond 4 ms in either direction. To eliminate the need to calculate out to infinite delays, the model applied Colburn's $p(\tau)$ function (center of Figure III.18), which gives more weight to lower ITDs—this ensured that extreme ITDs provided little contribution. The function may represent the density of neurons available for detection along the lag axis, and is described mathematically below.

$$\begin{aligned}
 p(\tau) &= C && \text{for } |\tau| \leq 0.15 \\
 p(\tau) &= C \exp\left(-\frac{|\tau| - 0.15}{0.6}\right) && \text{for } 0.15 < |\tau| \leq 0.2.2 \\
 p(\tau) &= C(0.33) \exp\left(-\frac{|\tau| - 2.2}{2.3}\right) && \text{for } |\tau| > 2.2
 \end{aligned}$$

The τ values in the above equations are measured in milliseconds. For the purpose of this calculation, the constant C was set to one. From the weighted cross-correlation function for one band, shown at the bottom of Figure III.18 along with the original function for comparison, the model process calculated the centroid to establish a corresponding point along the internal lag axis (indicated by an “x”), as predicted by Stern and Colburn (1978). The process has led to a lateral position for one band of the stimulus at one delay and one coherence.

Model calculation results

The sample output shown in Figure III.19 is a summary of twenty centroid calculations in each critical band using the described process. The input stimuli had a 200 μ s ITD, and the coherence level γ for this example was 0.5. The vertical axis

is the lag of the centroid, with 0 μs being the center of the head and values above shifted off to the right. The horizontal axis includes all fifteen bands tested in this calculation. Model results show that the centroids for an ITD of 200 μs are shifted well towards 0 ITD...none of them are delayed more than 60 μs . This is due to the weighting by $p(\tau)$. The distribution shows the well-known strength of lateralization at low frequencies due to fine-structure ITD—the centroids for bands 1 through 6 are shifted well away from the center, compared with centroids for higher bands. On the other hand, the decay of lateral position is non-monotonic with frequency, and there appears to be a slight peak at band 12 that may be the result of envelope ITD cues.

The most important features of the model results are as follows:

- The mean of the 20 centroid calculations is greater than zero in each band. This stimulus should be clearly lateralized to the right because all bands are in agreement.
- Going even further, almost *none* of the points fall below zero. The standard deviation of the points, while large in some cases, never extends to the left side of the listener's head. This result suggests no room for confusion.
- Centroids approach zero lag for the higher bands, where coherence is strongest (in the room-environment model). Also, the majority of coincidence counters are found near zero lag, which indicates that the auditory system works with smaller, more subtle lateral positions (Trahiotis and Stern, 1989)
- For decreasing coherences, centroids in all bands tend to shift towards the center of the head, which should degrade lateralization. At $\alpha = 0.4$, some points start to appear below zero. At $\alpha = 0.1$, those bands which are the strongest

lateralized in Figure III.19 (1-6) are still generally above zero, while higher bands show no preference to either side.

Model predictions

The experimenters have attempted to establish a detection criterion that would analyze the centroid distributions of Figure III.19 and match the data. The selected criterion made predictions according to the fraction of centroid values greater than zero in each band. If the stimulus contained a band where more than 71% of points are above zero, then the criterion considers it lateralized to the right. As is evident, this criterion would tend to predict equal performance regardless of high-pass cutoff frequency due to the vast majority of centroids being above zero in all bands at low coherences. Figure III.20 includes the replotting of 200- μ s Experiment 3 data (represented by pentagons), as well as attempts at fitting the model to those results. The solid triangles are the result of the model as stated: because the calculated centroids were well-lateralized in all bands when stimulus coherence was 0.2, one would expect listeners to perform adequately at that level over a range of cutoff frequencies. The model predictions are somewhat close to the actual data for cutoffs of 1 kHz and below, but they clearly overestimate the ability of listeners to use ITD cues in the “envelope region” above 1500 Hz.

Raab and Goldberg (1975) note that several models have been reconciled with real-world results by the inclusion of internal neural noise, for which the above coherence-based model has not yet accounted. The concept of neural noise, a representation of randomness in binaural processing, can be of some significance to sensitive discrimination tasks. Indeed, many estimations of neural noise find it several

times larger than the noise applied externally (Raab and Goldberg, 1975). To ensure completeness of the model, one must consider the effect of neural noise that is additive with the interfering noise, and thus reduces the overall coherence from its expected value.

To better fit the model to results, it is clear from Figure III.20 that the actual coherences must be much lower. For instance, the average threshold at 2000-Hz cutoff was 0.5 coherence, while the model predicted only 0.2. At the expected coherence of 0.5, the coherent power must equal the incoherent power:

$$\gamma = \frac{P_c}{P_c + P_I} = 0.5$$

$$\therefore P_c = P_I$$

One can reconcile the model and results by the addition of estimated neural noise that has three times the power of the interfering noise to establish an *actual* coherence of 0.2:

$$\gamma = \frac{P_c}{P_c + P_I + 3P_I} = \frac{P_c}{P_c + P_c + 3P_c} = \frac{P_c}{5P_c} = 0.2$$

In support of the above derivation, model calculations at a γ of 0.5 (as in Figure III.19) that include the estimated neural noise result in centroid distributions closely matching those of a $\gamma=0.2$ calculation without neural noise. Continuing with the addition of neural noise at constant power, one can equate all expected coherences on Figure III.20 with actual coherences that will be used for a new set of predictions. Open circles mark those predictions on the graph (the “N-noise model”). Note how well the model matches actual performance in regions below 1500 Hz (where listeners still had access to fine-structure ITD cues) and above (where they were limited to envelope cues). The

inclusion of neural noise appears to lend more realism to the coherence-based model as implemented here.

Unfortunately, the model is much less successful at predicting the results of Experiment 2. The implementation of $\text{roex}(p)$ filters that would accept masker power below the target noise cutoff should lead to lower centroids in bands near the cutoff, and thus, require more coherence to lateralize those bands. This is clearly the case when comparing the results of Experiments 2 and 3 in Figure III.16. However, model predictions based on calculations with broadband maskers and high-pass filtered targets are exactly the same as the open circles in Figure III.20. There is no increase in required coherence. The experimenters continue to refine the model and look for new interpretations that might account for this discrepancy. One possible alteration is to include a frequency-dependent version of Colburn's $p(\tau)$ function, or $p(\tau, f)$. Also, switching to $\text{roex}(p, r)$ filters (Hartmann, 1997) would incorporate much more masker power outside the target band, which may improve predictions for Experiment 2.

Conclusions

Experiment 1 presented listeners with three varied tasks. The first test verified that absolute threshold was unaffected by interaural delay. The second revealed that listeners required increased stimulus levels to adequately lateralize smaller ITD shifts. The third found that listeners needed increased coherences (i.e. S/N ratios) to adequately lateralize targets exhibiting smaller ITD shifts. In runs with and without a masking noise, listeners produced parallel thresholds as a function of the shift. This suggests an intrinsic

noise that interferes with the binaural system's ability to discriminate in a manner similar to the masker.

The key conclusion of Experiment 2 is that the envelope ITD cues are indeed useful for a high-passed signal. The coherence graph clearly shows that larger ITDs were still effectively lateralized out to increasingly higher cutoff frequencies above 1500 Hz. However, the results clearly show that listeners required high interaural coherence in those cases, suggesting that envelope cues are susceptible to masking and reverberated noise (similar to the conclusion reached in Chapter II). In contrast, waveform cues seem very effective at low coherences, as evidenced by listener performance for low cutoffs. This would seem to indicate that envelope ITDs may play a less significant role than fine-structure ITDs in the lateralization process.

Experiment 3 was motivated by the suspicion that Experiment 2 had included some upward spread of masking. Eliminating the low-frequency masker power (an integral part of the described room environment) below the target's high-pass cutoff yielded significantly lower coherence thresholds for cutoffs of 1500 Hz and above. These results strongly supported the hypothesis that masker outside of the target band was still reducing the coherence.

The coherence-based model was not very effective at forecasting the results of Experiment 3. The inclusion of estimated neural noise improved predictions dramatically, to fit thresholds that varied according to the effectiveness of both fine-structure and envelope ITD cues. Modifications to approximate the stimuli of Experiment 2 did not change the model's predicted thresholds.

References

- Bernstein, L.R. and Trahiotis, C. "The normalized correlation: Accounting for binaural detection across center frequency," *J. Acoust. Soc. Am.* **100**, 3774-3784 (1996).
- Colburn, H.S. "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," *J. Acoust. Soc. Am.* **61**, 525-533 (1977).
- Colburn, H.S. and Durlach, N.I. "Binaural Phenomena," *Handbook of Perception vol. 4* ed. E. Carterette, pp. 467-518 (1978).
- Durlach, N.I. "Binaural Signal Detection: Equalization and Cancellation Theory," *Foundations of Modern Auditory Theory* ed. J. Tobias, pp. 371-461 (1972).
- Durlach, N.I. and Colburn, H.S. "Binaural Phenomena," *Handbook of Perception vol. 4* ed. E. Carterette, pp. 365-466 (1978).
- Hartmann, W.M. *Signals, Sound and Sensation*. American Institute of Physics, Woodbury, NY, 1997. p. 60, 248.
- Hartmann, W.M., Constan, Z.A., Rakerd, B. "Binaural coherence and the localization of sound in rooms," *J. Acoust. Soc. Am.* **103**, 3081 (1998).
- Henning, B. "Some observations on the lateralization of complex waveforms," *J. Acoust. Soc. Am.* **68**, 446-454 (1980).
- Jeffress, L.A., Blodgett, H.C. and Deatherage, B.H. "Effect of interaural correlation on the precision of centering a noise," *J. Acoust. Soc. Am.* **34**, 1122-1123 (1962).
- Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S. and Oxenham, A.J. "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," *Acustica* **83**, 659-669 (1997).
- Levitt, H. "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467-477 (1971).
- McFadden, D. and Pasanen, E.G. "Binaural detection at high frequencies with time-delayed waveforms," *J. Acoust. Soc. Am.* **63**, 1120-1131 (1978).
- Moore, Brian C.J. *Hearing*. Academic Press, New York, 1995. pp. 161-205.
- Pickles, J.O. *An Introduction to the Physiology of Hearing*. Academic Press, New York, 1982. p. 82
- Raab, D.H. and Goldberg, I.A. "Auditory intensity discrimination with bursts of reproducible noise," *J. Acoust. Soc. Am.* **57**, 437-447 (1975).

Rayleigh, Lord (3rd Baron Rayleigh) "On our perception of sound direction," *Philos. Mag.* **13**, 214-232 (1907).

Stern, R.M. Jr. and Colburn, H.S. "Theory of Binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position," *J. Acoust. Soc. Am.* **64**, 127-140 (1978).

Trahiotis, C., Bernstein, L.R., Akeroyd, M.A. "Manipulating the 'straightness' and 'curvature' of patterns of interaural cross correlation affects listeners' sensitivity to changes in interaural delay," *J. Acoust. Soc. Am.* **109**, 321-330 (2001).

Trahiotis, C. and Stern, R.M. "Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase," *J. Acoust. Soc. Am.* **86**, 1285-1293 (1989).

Trahiotis, C. and Stern, R.M. "Across-frequency interaction in lateralization of complex binaural stimuli," *J. Acoust. Soc. Am.* **96**, 3804-3806 (1994).

van der Heijden, M. and Trahiotis, C. "A new way to account for binaural detection as a function of interaural noise correlation," *J. Acoust. Soc. Am.* **101**, 1019-1022 (1997).

Yin, T.C.T. and Chan, J.C.K. "Interaural Time Sensitivity in Medial Superior Olive of Cat," *J. Neurophys.* **64**, p. 465-488 (1990).

Figure III.1. Depiction of a human head and a sound source in a room. The source is represented by a dot in the upper left corner, while the head is at bottom right. A solid line from the source pointing to the head represents the path of a direct sound that is unimpeded by obstacles. The dashed lines indicate a few possible paths for reflected sounds arriving at the head.

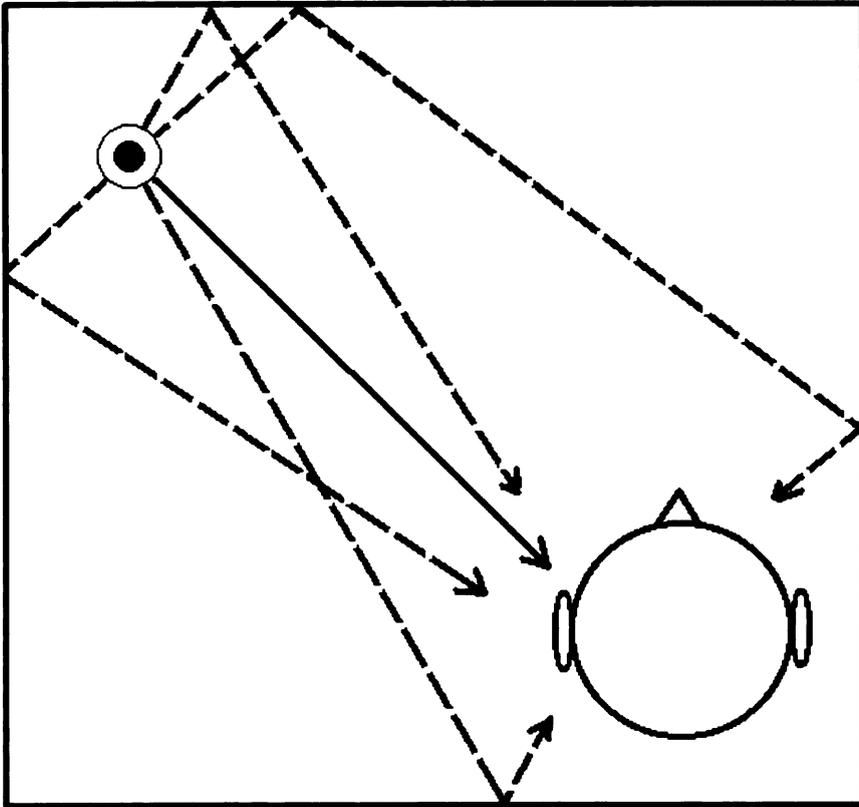


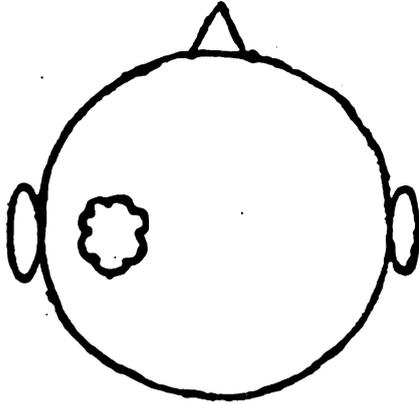
Figure III.1

Figure III.2. Schematic of the experimental setup in the form of a patch diagram. Begin at the left, where noises N1, N2 and N3 are generated. LPF stands for low-pass filter, Delay x/2x represents the delay line and the relative settings in μs , AS is the analog switch, EVCA is the exponential voltage-controlled amplifier and Trap. Env. stands for the trapezoidal envelope that ramps signals on and off.

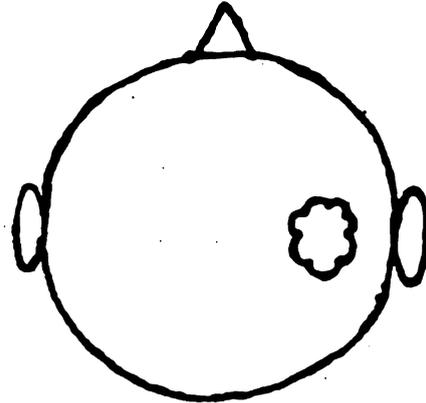
Figure III.3. Description of the experimental task. Images portray listener's head with an intracranial image produced by the stimulus (depicted by a small "cloud"). Each trial, the first interval is equally likely to have a positive or negative ITD imposed so the intracranial image appears offset to the left or right side of the head. The second interval features an opposite ITD of the same magnitude so the image moves from left to right (as in the top two images) or vice versa (as in the bottom two images).

Two-interval forced choice task

First interval



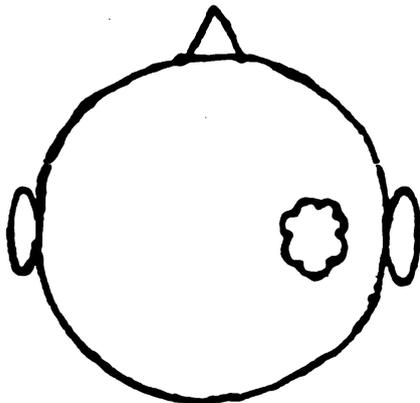
Second Interval



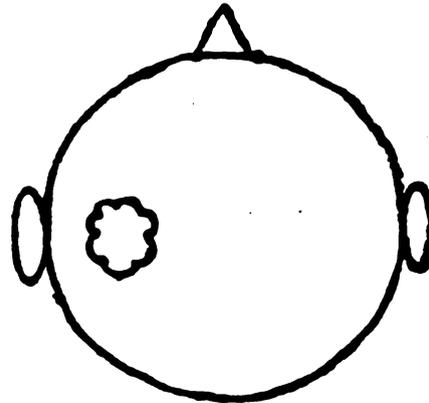
Left to Right Trial

OR

First Interval



Second Interval



Right to Left Trial

Figure III.3

Figure III.4. Results of Experiment 1: ITD discrimination as a function of coherence for listener B. The x-axis shows all tested ITDs from 20 to 400 μs , and the y-axis is the threshold target level necessary to perform the task (with 79% accuracy) at each. The three curves plot results for interference, non-interference, and absolute threshold conditions. For the interference condition, also note the dashed line that represents the level of the masking noise and the alternate y-axis indicating the corresponding coherences to target levels on the right side of the graph. The actual coherence values corresponding to interference data points are listed next to those points for ease of evaluation. Error bars on data points represent two standard deviations.

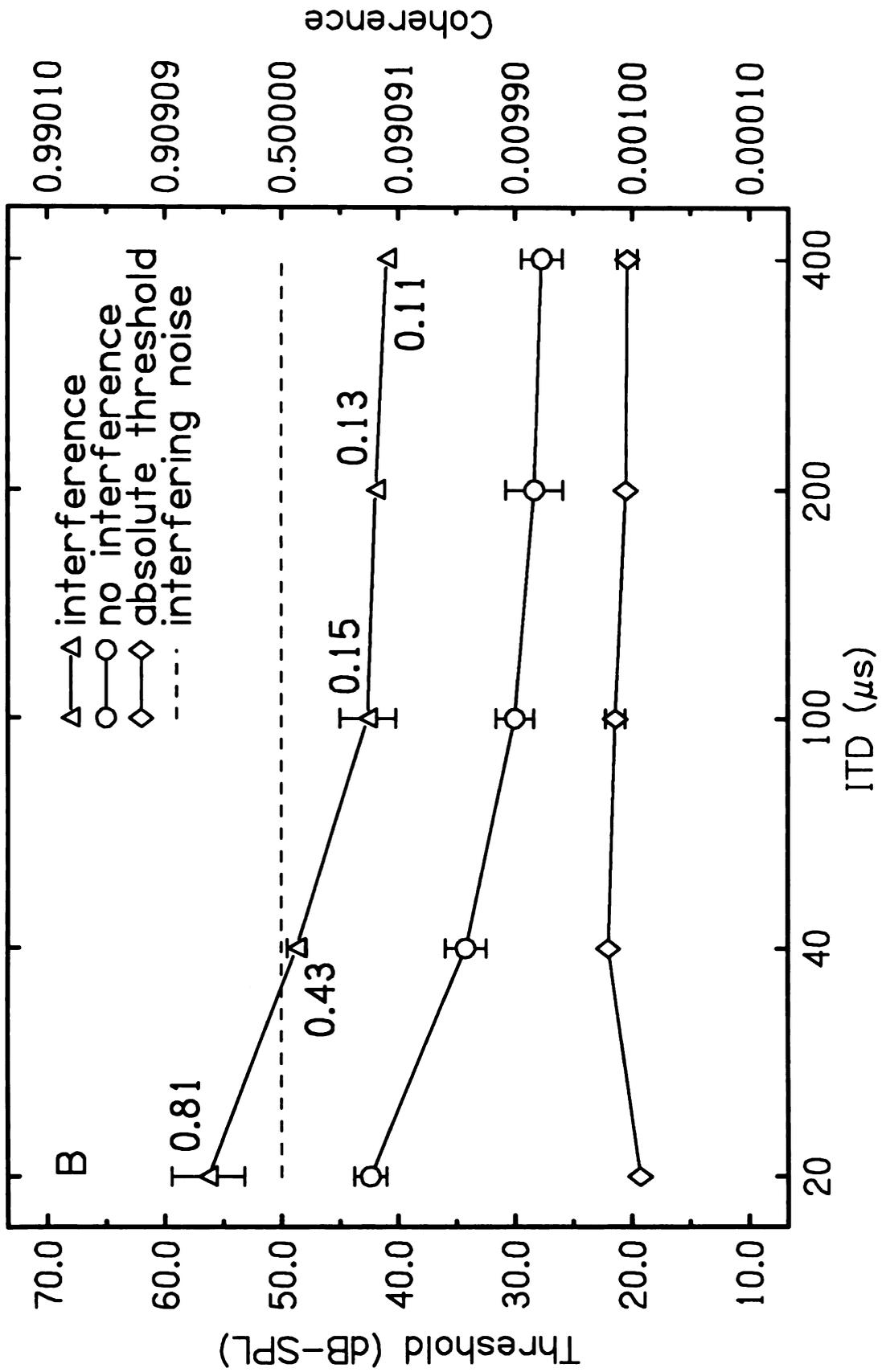


Figure III.4

Figure III.5. Results of Experiment 1: ITD discrimination as a function of coherence for listener L. The x-axis shows all tested ITDs from 20 to 400 μ s, and the y-axis is the threshold target level necessary to perform the task (with 79% accuracy) at each. The three curves plot results for interference, non-interference, and absolute threshold conditions. For the interference condition, also note the dashed line that represents the level of the masking noise and the alternate y-axis indicating the corresponding coherences to target levels on the right side of the graph. The actual coherence values corresponding to interference data points are listed next to those points for ease of evaluation. Error bars on data points represent two standard deviations.

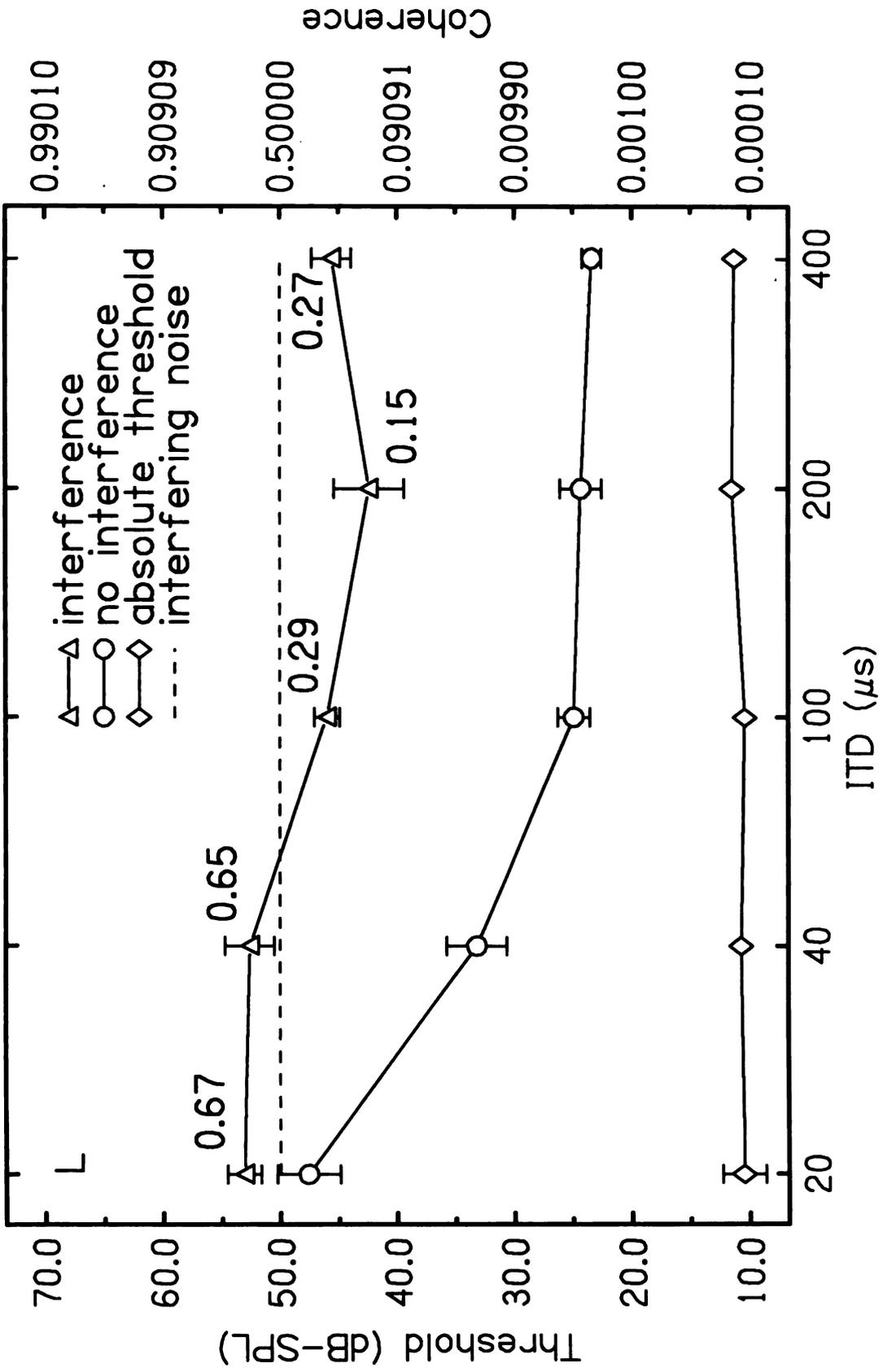


Figure III.5

Figure III.6. Results of Experiment 1: ITD discrimination as a function of coherence for listener S. The x-axis shows all tested ITDs from 20 to 400 μ s, and the y-axis is the threshold target level necessary to perform the task (with 79% accuracy) at each. The three curves plot results for interference, non-interference, and absolute threshold conditions. For the interference condition, also note the dashed line that represents the level of the masking noise and the alternate y-axis indicating the corresponding coherences to target levels on the right side of the graph. The actual coherence values corresponding to interference data points are listed next to those points for ease of evaluation. Error bars on data points represent two standard deviations.

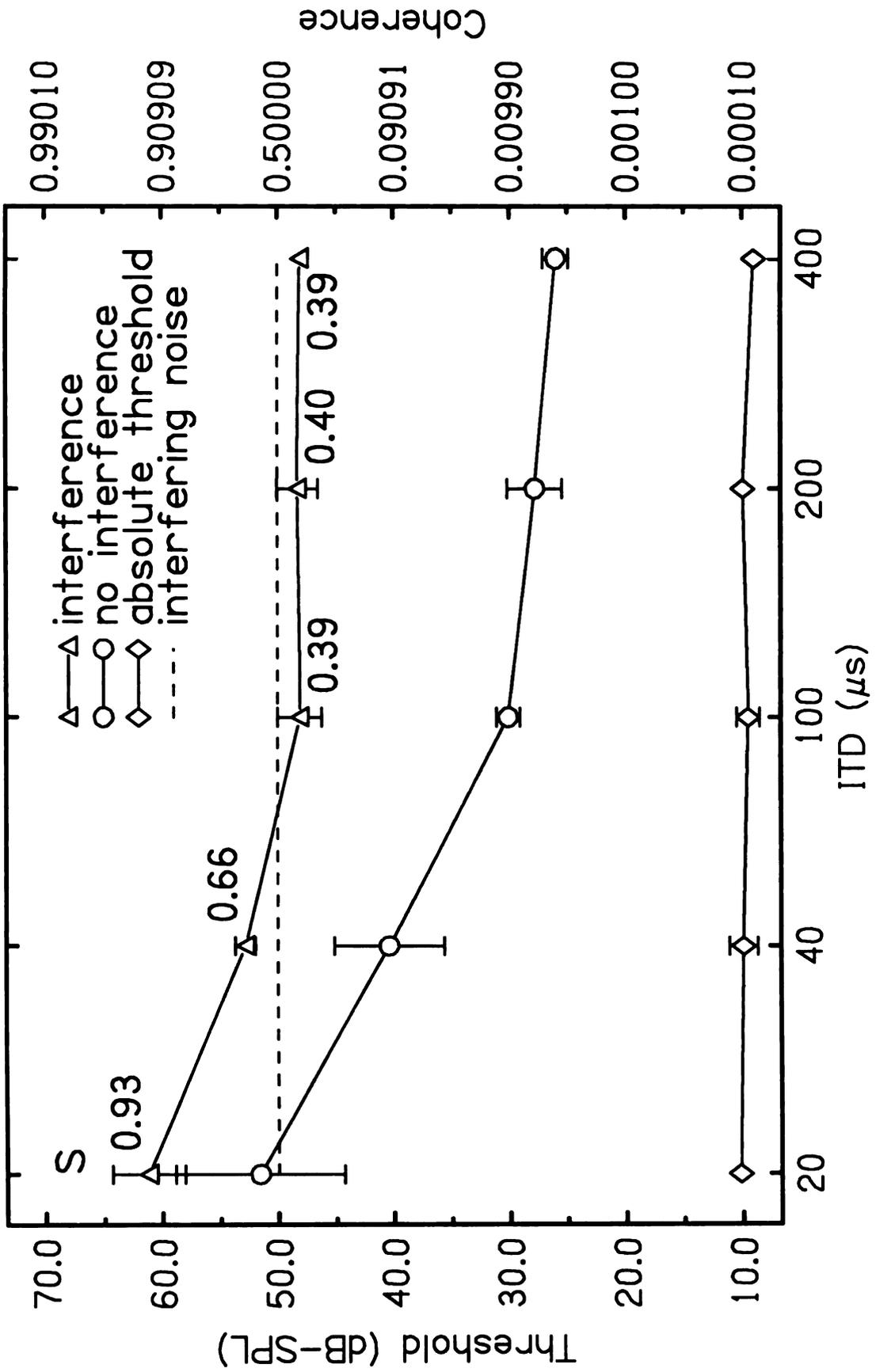


Figure III.6

Figure III.7. Results of Experiment 1: ITD discrimination as a function of coherence for listener T. The x-axis shows all tested ITDs from 20 to 400 μ s, and the y-axis is the threshold target level necessary to perform the task (with 79% accuracy) at each. The three curves plot results for interference, non-interference, and absolute threshold conditions. For the interference condition, also note the dashed line that represents the level of the masking noise and the alternate y-axis indicating the corresponding coherences to target levels on the right side of the graph. The actual coherence values corresponding to interference data points are listed next to those points for ease of evaluation. Error bars on data points represent two standard deviations.

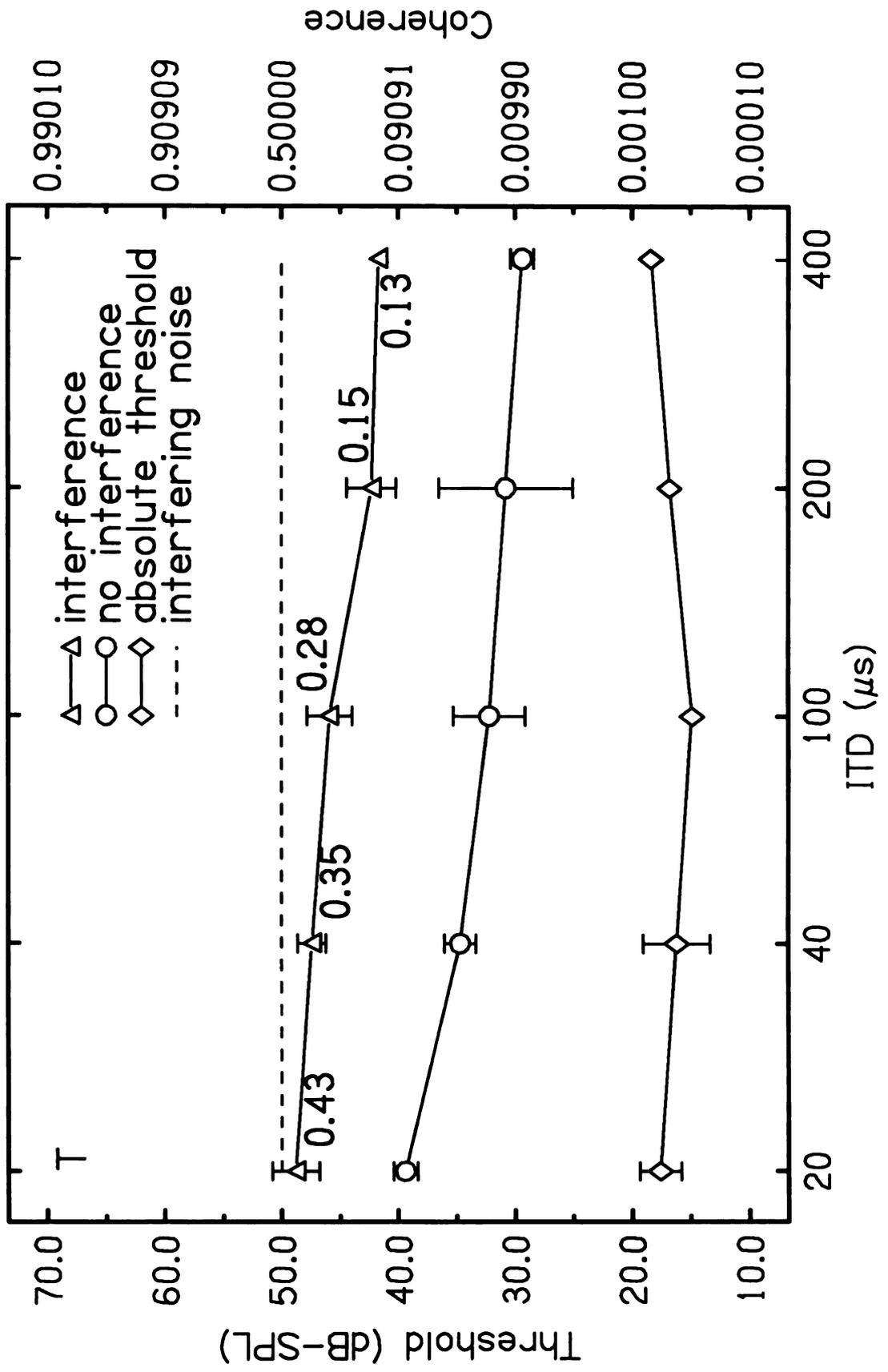


Figure III.7

Figure III.8. Results of Experiment 1: ITD discrimination as a function of coherence for listener W. The x-axis shows all tested ITDs from 20 to 400 μ s, and the y-axis is the threshold target level necessary to perform the task (with 79% accuracy) at each. The three curves plot results for interference, non-interference, and absolute threshold conditions. For the interference condition, also note the dashed line that represents the level of the masking noise and the alternate y-axis indicating the corresponding coherences to target levels on the right side of the graph. The actual coherence values corresponding to interference data points are listed next to those points for ease of evaluation. Error bars on data points represent two standard deviations.

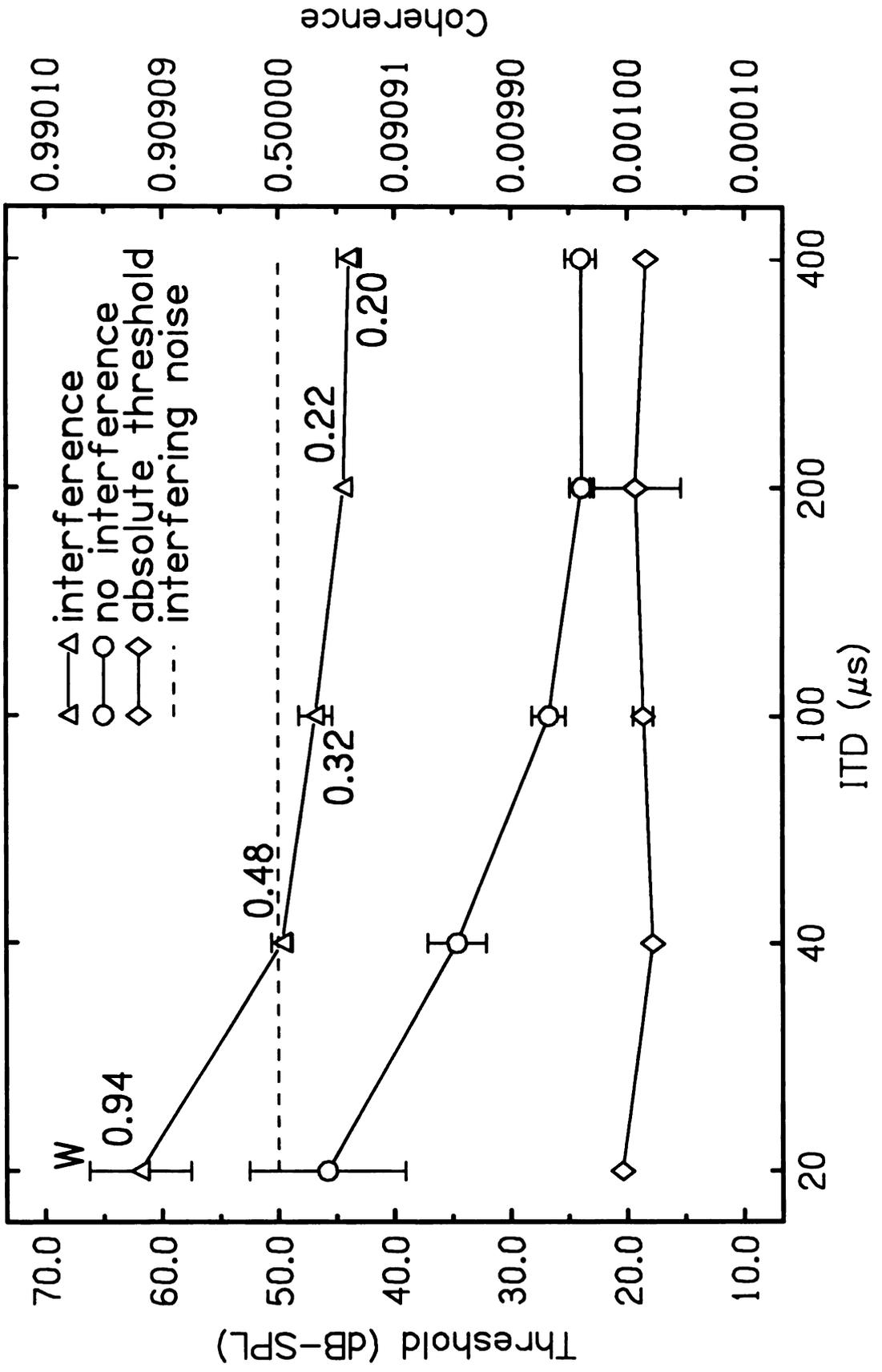


Figure III.8

Figure III.9. Results of Experiment 1: ITD discrimination as a function of coherence for listener Z. The x-axis shows all tested ITDs from 20 to 400 μ s, and the y-axis is the threshold target level necessary to perform the task (with 79% accuracy) at each. The three curves plot results for interference, non-interference, and absolute threshold conditions. For the interference condition, also note the dashed line that represents the level of the masking noise and the alternate y-axis indicating the corresponding coherences to target levels on the right side of the graph. The actual coherence values corresponding to interference data points are listed next to those points for ease of evaluation. Error bars on data points represent two standard deviations.

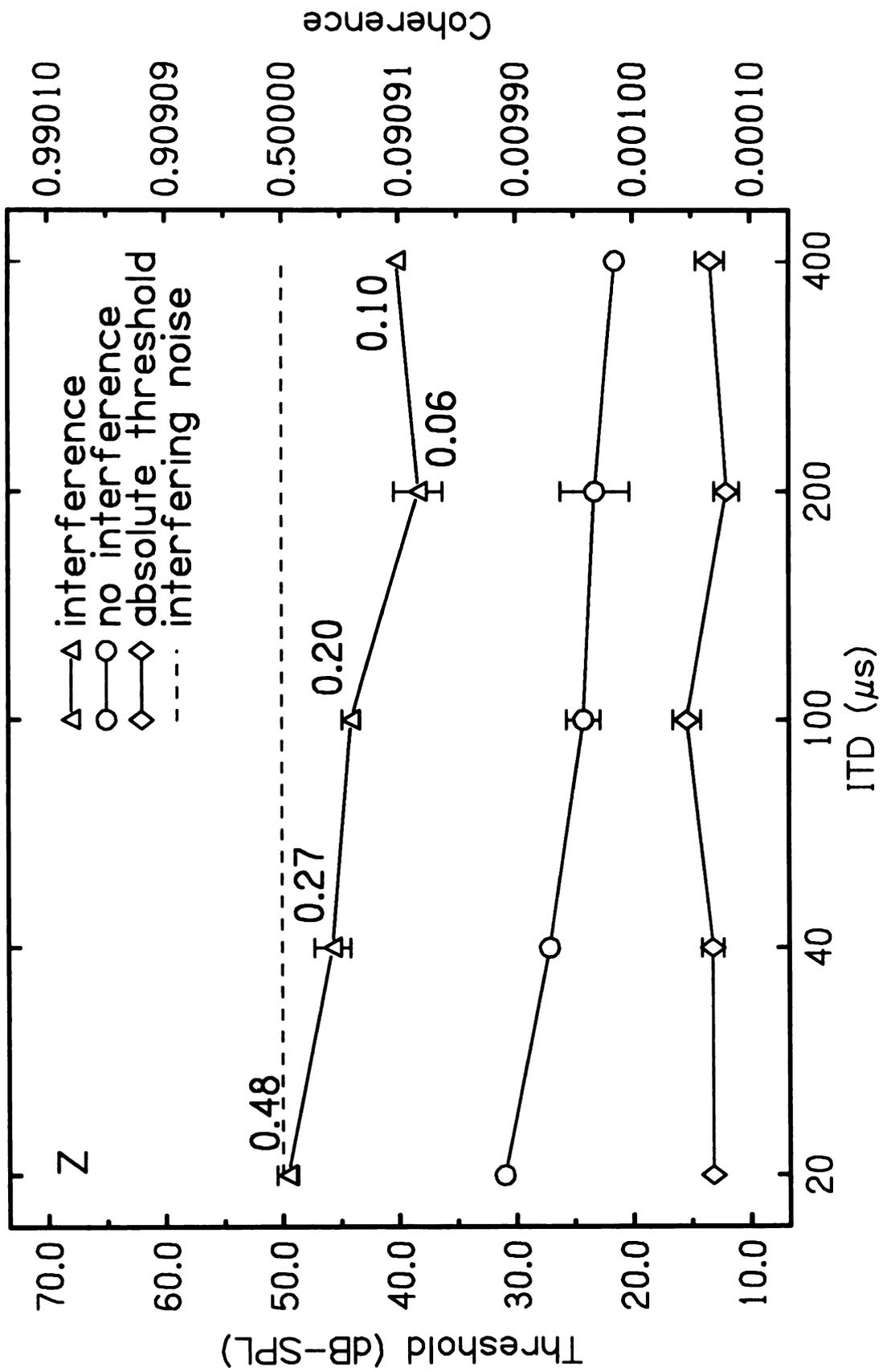


Figure III.9

Figure III.10. Results of Experiment 1: ITD discrimination as a function of coherence averaged over all six listeners. The x-axis shows all tested ITDs from 20 to 400 μ s, and the y-axis is the threshold target level necessary to perform the task (with 79% accuracy) at each. The three curves plot results for interference, non-interference, and absolute threshold conditions. For the interference condition, also note the dashed line that represents the level of the masking noise and the alternate y-axis indicating the corresponding coherences to target levels on the right side of the graph. The actual coherence values corresponding to interference data points are listed next to those points for ease of evaluation.

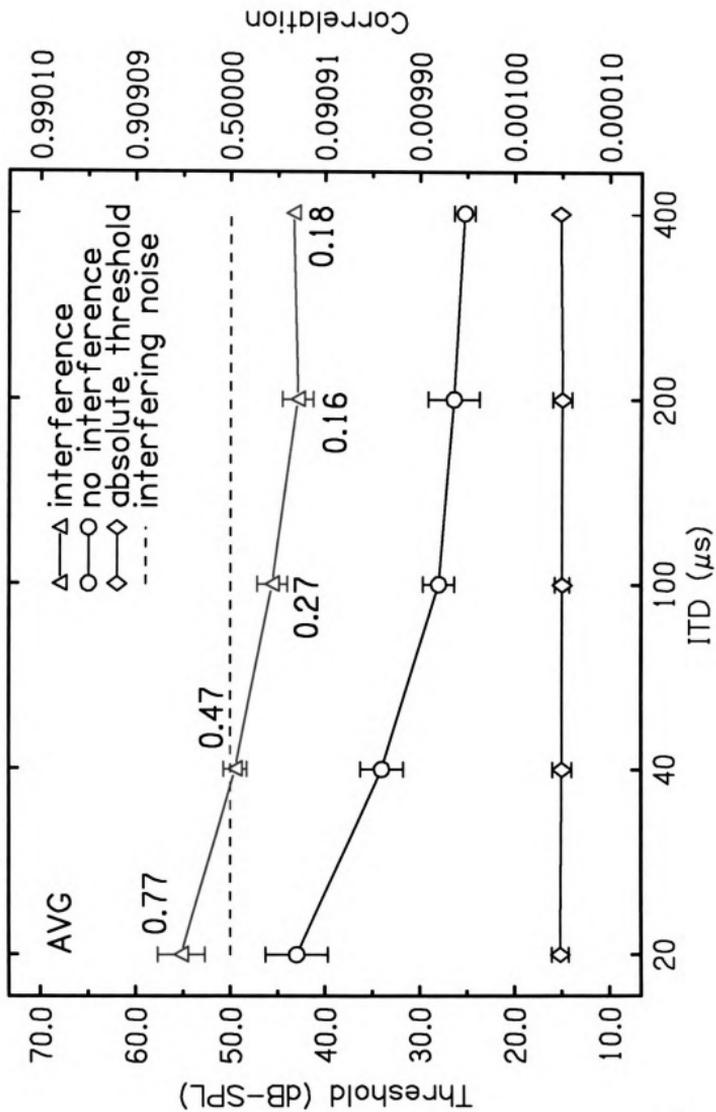


Figure III.10

Figure III.11. Spectral diagram of experimental stimulus for Experiment 2. The four plots show stimulus intensity versus frequency. The two columns represent left and right ears. The top row shows the characteristics of the coherent target noise (A), which extends from high-pass cutoff frequency f_c to 10 kHz. The dashed line extending to lower frequency indicates that the cutoff can be adjusted by the experimenter. The bottom row represents the incoherent masking noises (B and C) in each ear, which are broadband to 10 kHz. The dashed line extending to higher intensity indicates that the experimental program can vary the masking noise level during runs.

Experimental Stimulus

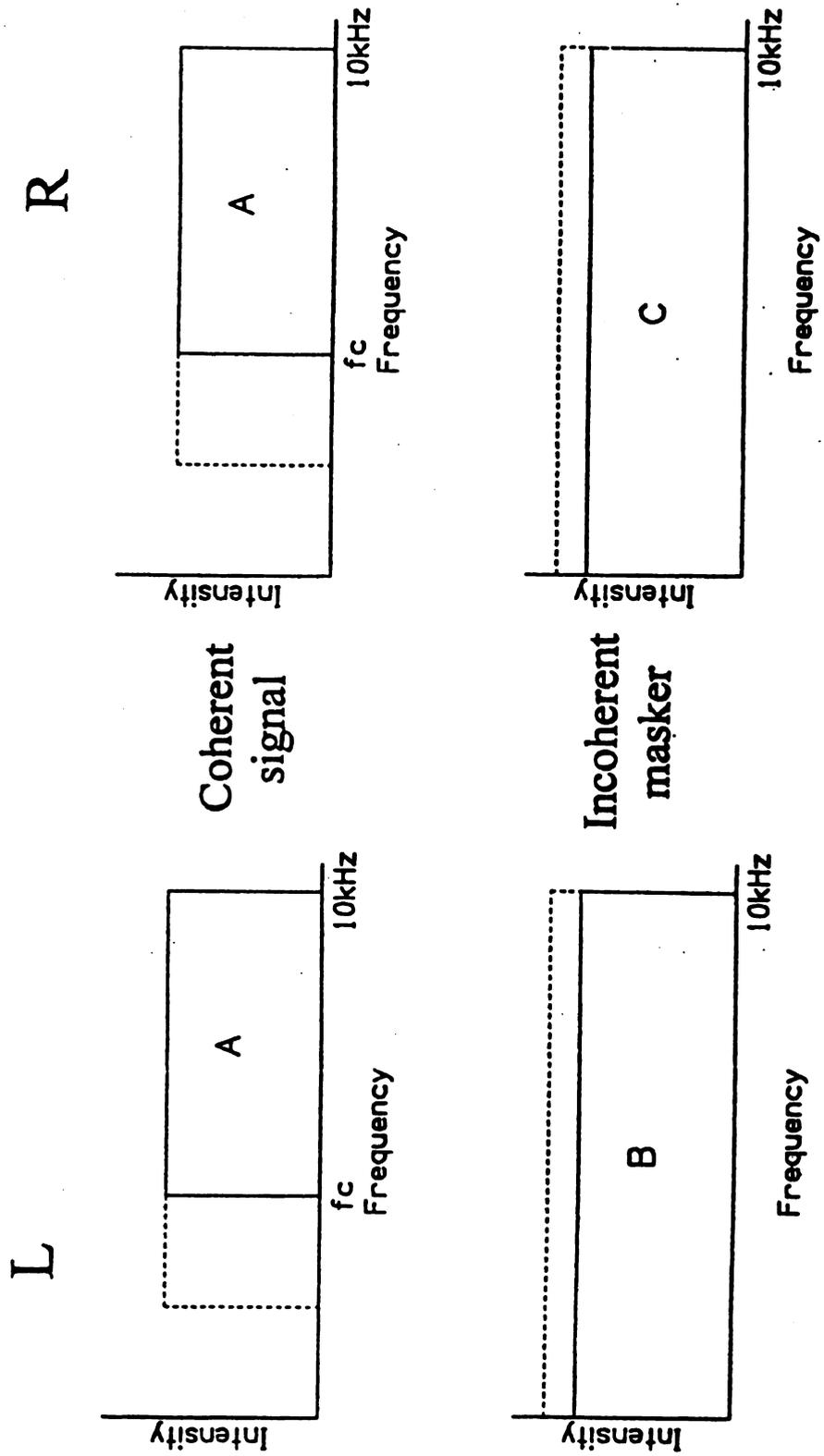


Figure III.11

Figure III.12. Results of Experiment 2: ITD discrimination as a function of high-pass cutoff. This graph shows the range of tested high-pass cutoff frequencies (seven values from 0 to 4000 Hz) along the x-axis. The y-axis indicates the interaural coherence necessary to correctly lateralize the internal image in 79% of trials. The five curves are the results for various ITDs (20 to 400 μ s) averaged over four listeners. The coherences were calculated ignoring incoherent power outside the target band.

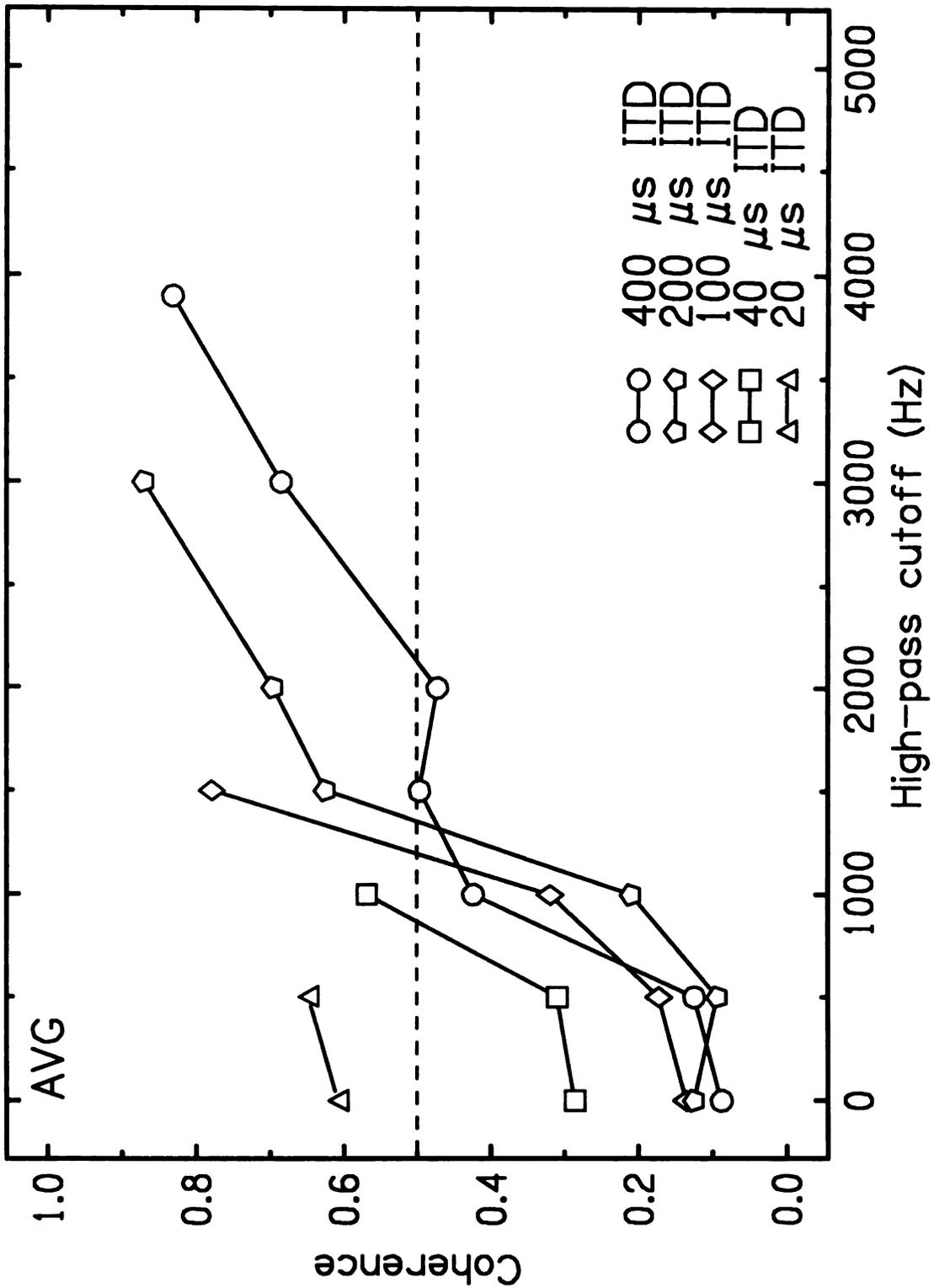


Figure III.12

Figure III.13. Results of Experiment 3: ITD discrimination of high-pass-filtered targets and maskers for listener W. This graph shows the range of tested high-pass cutoff frequencies (seven values from 0 to 4000 Hz) along the x-axis. These cutoffs were applied to both the coherent target noise and the incoherent masking noise. The y-axis indicates the interaural coherence necessary to correctly lateralize the internal image in 79% of trials. The square symbols mark thresholds for a 200- μ s ITD in Experiment 3, while the pentagons replot 200- μ s results from Experiment 2 for comparison.

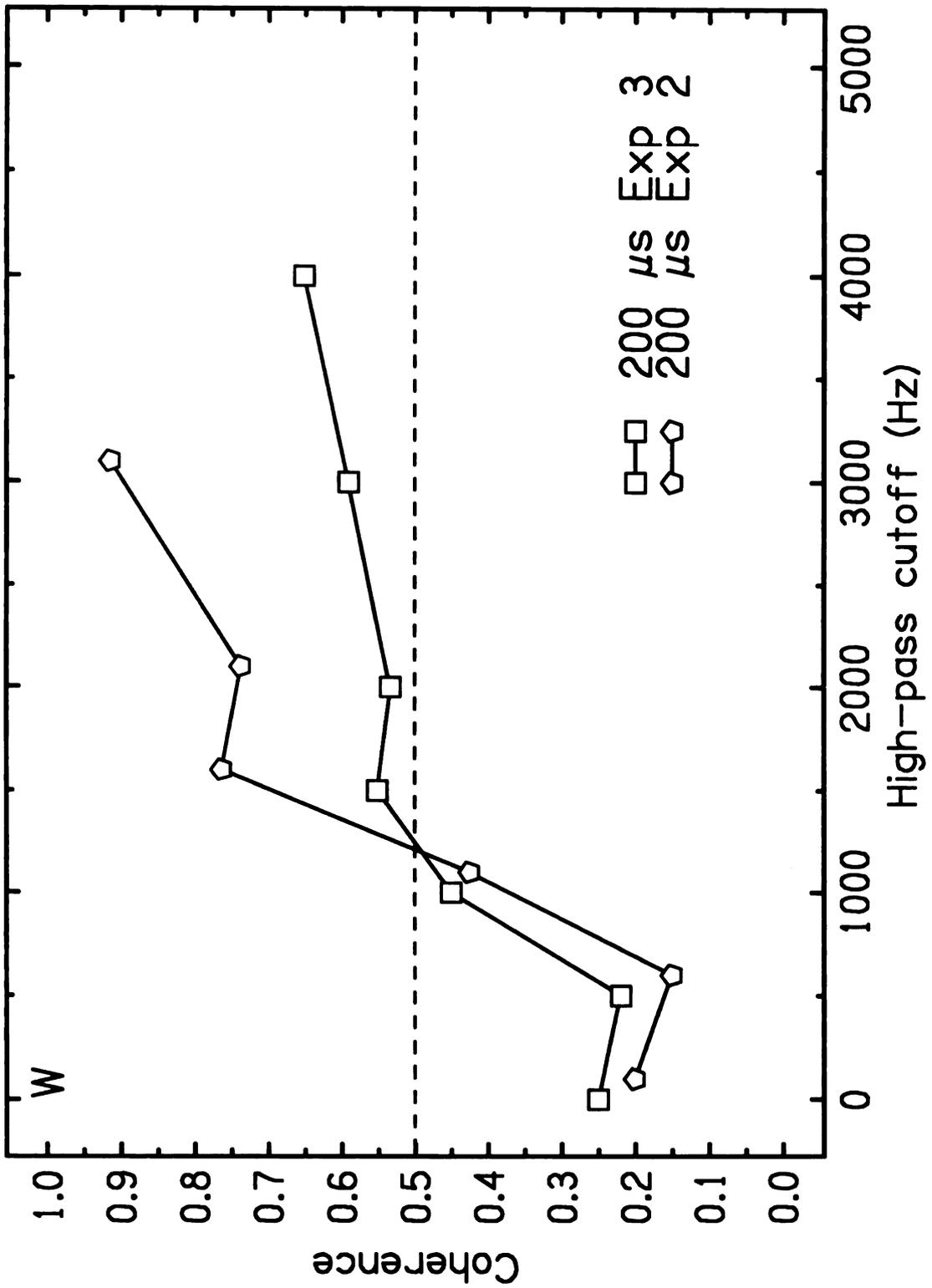


Figure III.13

Figure III.14. Results of Experiment 3: ITD discrimination of high-pass-filtered targets and maskers for listener X. This graph shows the range of tested high-pass cutoff frequencies (seven values from 0 to 4000 Hz) along the x-axis. These cutoffs were applied to both the coherent target noise and the incoherent masking noise. The y-axis indicates the interaural coherence necessary to correctly lateralize the internal image in 79% of trials. The square symbols mark thresholds for a 200- μ s ITD in Experiment 3, while the pentagons replot 200- μ s results from Experiment 2 for comparison.

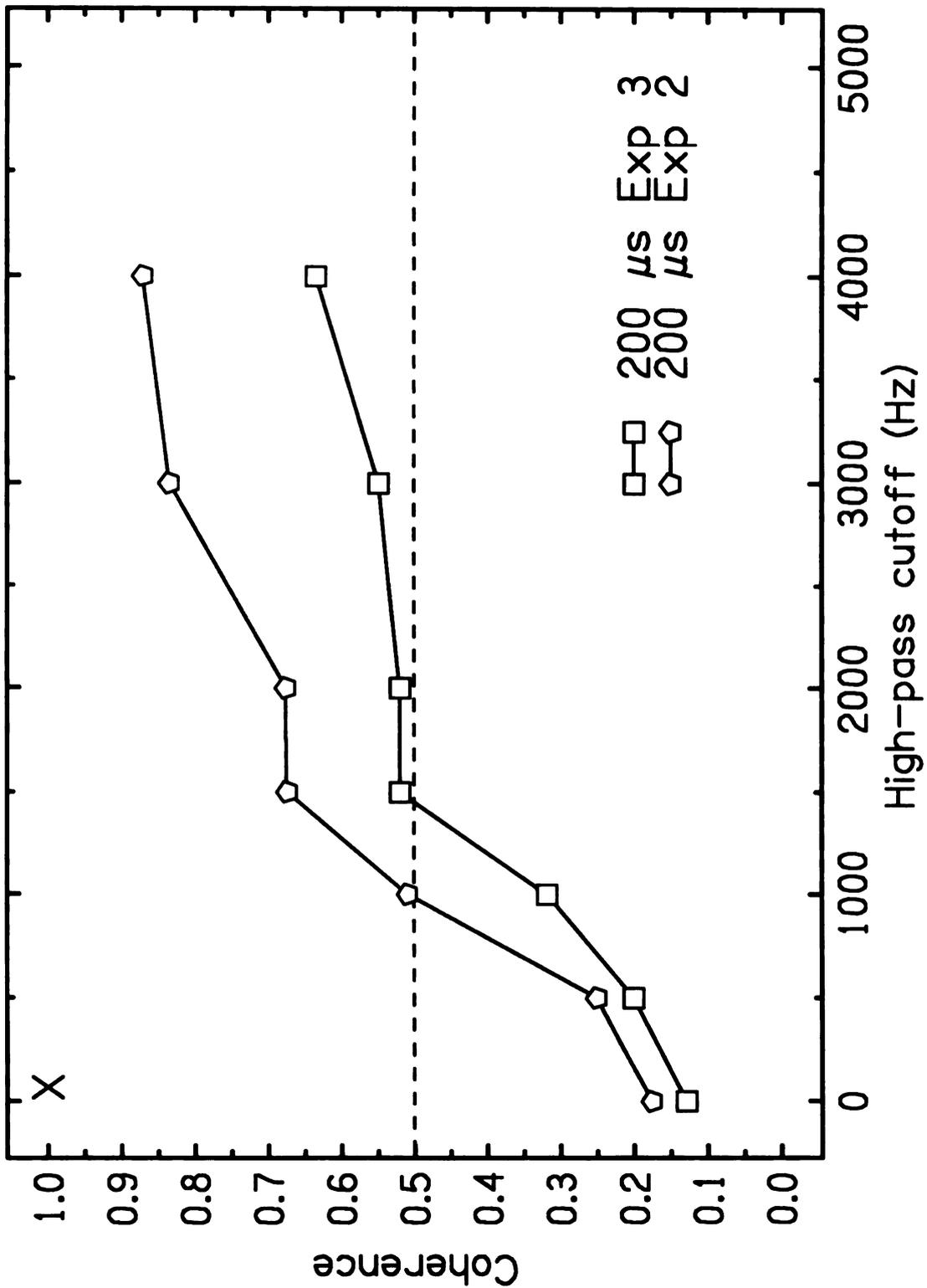


Figure III.14

Figure III.15. Results of Experiment 3: ITD discrimination of high-pass-filtered targets and maskers for listener Z. This graph shows the range of tested high-pass cutoff frequencies (seven values from 0 to 4000 Hz) along the x-axis. These cutoffs were applied to both the coherent target noise and the incoherent masking noise. The y-axis indicates the interaural coherence necessary to correctly lateralize the internal image in 79% of trials. The square symbols mark thresholds for a 200- μ s ITD in Experiment 3, while the pentagons replot 200- μ s results from Experiment 2 for comparison.

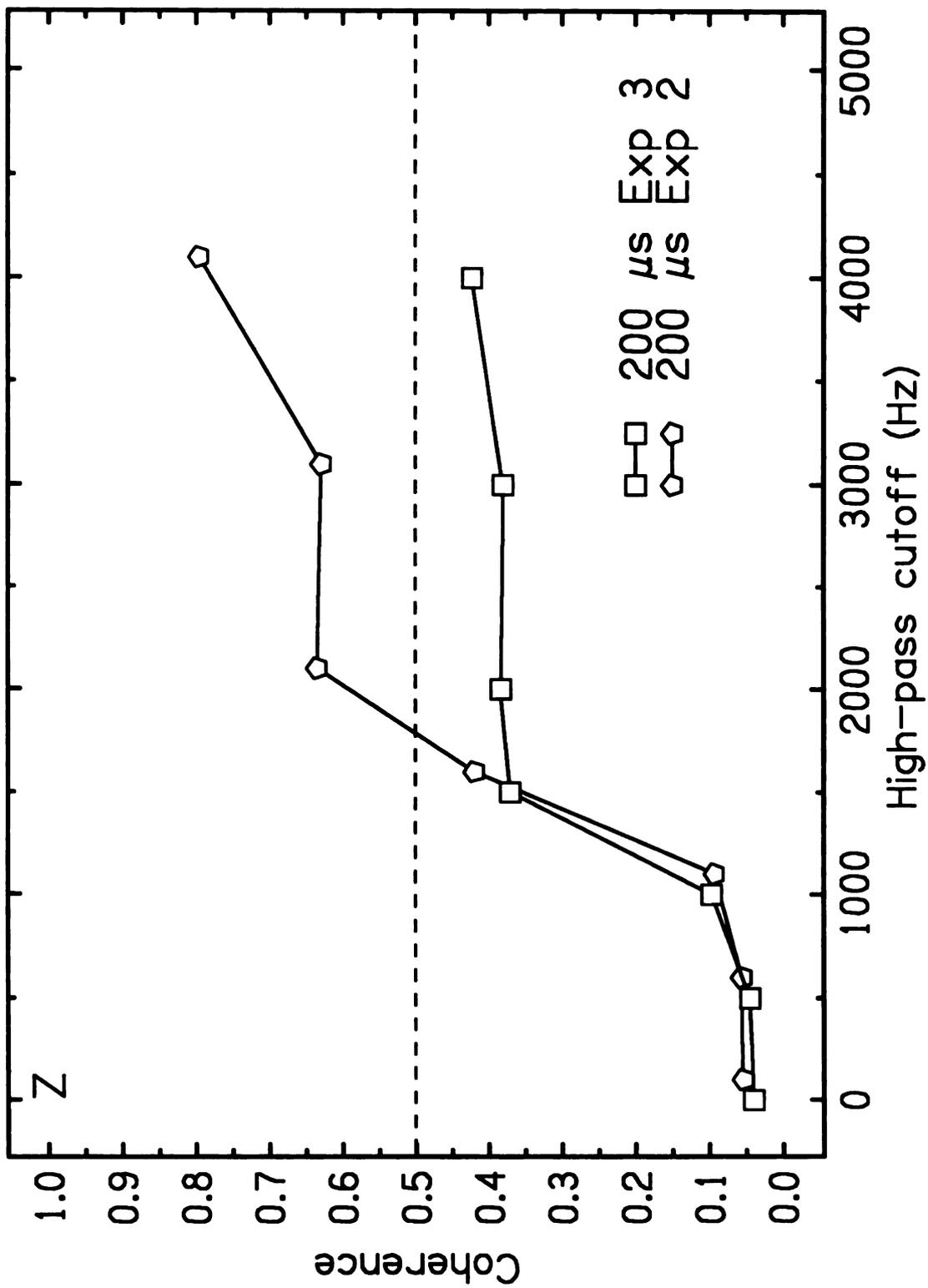


Figure III.15

Figure III.16. Results of Experiment 3: ITD discrimination of high-pass-filtered targets and maskers averaged over three listeners. This graph shows the range of tested high-pass cutoff frequencies (seven values from 0 to 4000 Hz) along the x-axis. These cutoffs were applied to both the coherent target noise and the incoherent masking noise. The y-axis indicates the interaural coherence necessary to correctly lateralize the internal image in 79% of trials. The square symbols mark thresholds for a 200- μ s ITD in Experiment 3, while the pentagons replot 200- μ s results from Experiment 2 for comparison. The dotted line connecting to the last Experiment 2 point indicates that it is the average of only two listeners' results; the third listener failed at that condition. Error bars on data points represent two standard deviations.

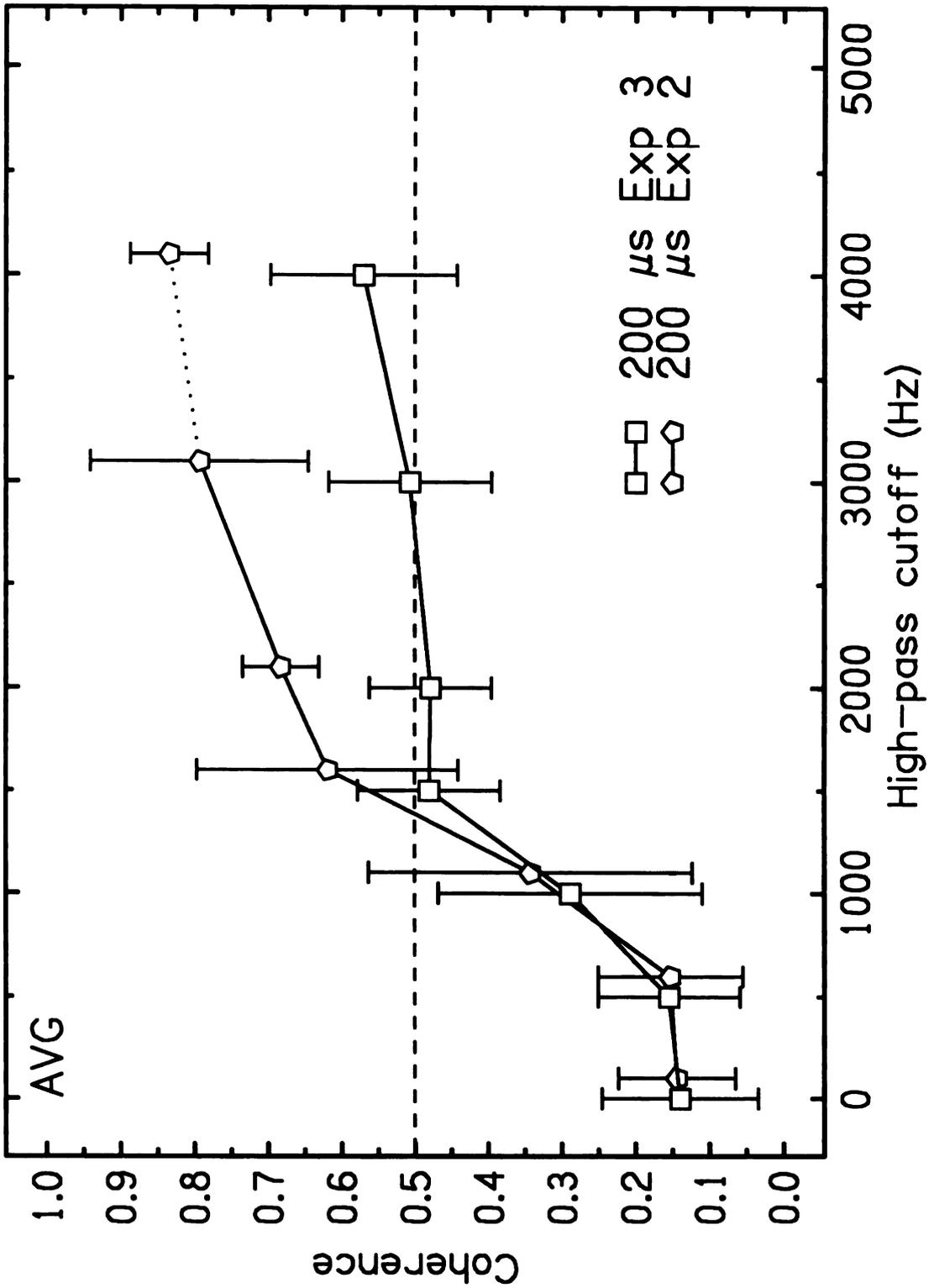
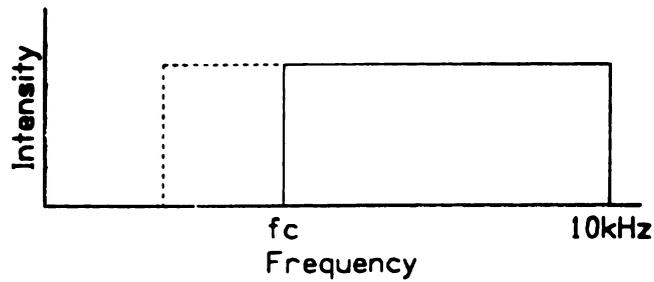


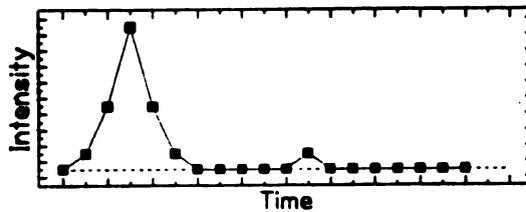
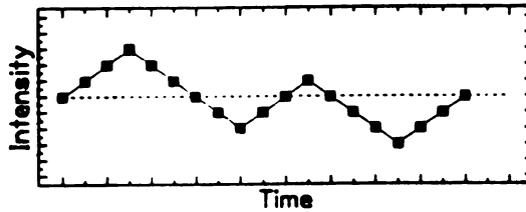
Figure III.16

Figure III.17. Simulating the actions of the auditory periphery in the model. There are three steps listed in this figure. The uppermost shows the generation of the stimulus according to experimental specifications (coherence, cutoff). The stimulus is separated into critical bands (with center frequencies falling in the range listed) for processing. Each band is rectified (center panel) and low-pass filtered (bottom panel, Bernstein and Trahiotis, 1996) before cross-correlation.

Generate stimulus within odd critical bands from 1 to 31
(124-7021 Hz)



Half-wave square-law rectify



Low-pass filtering

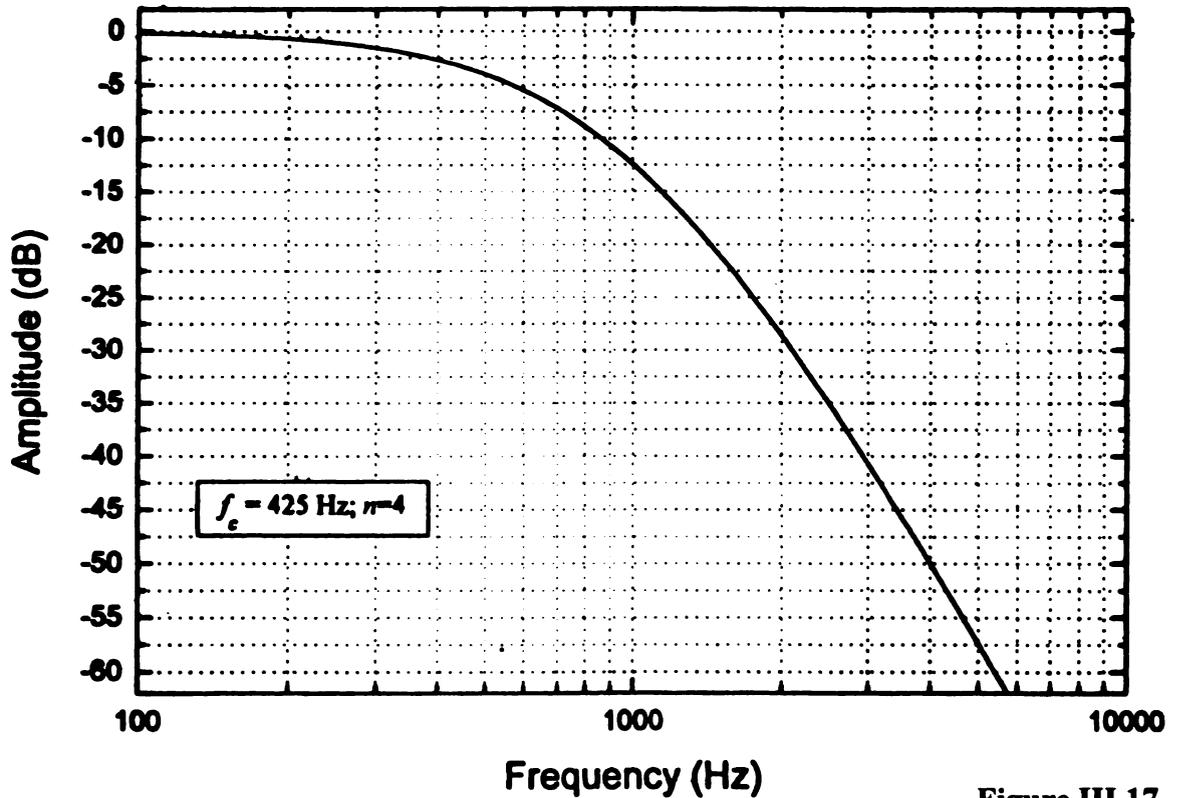


Figure III.17

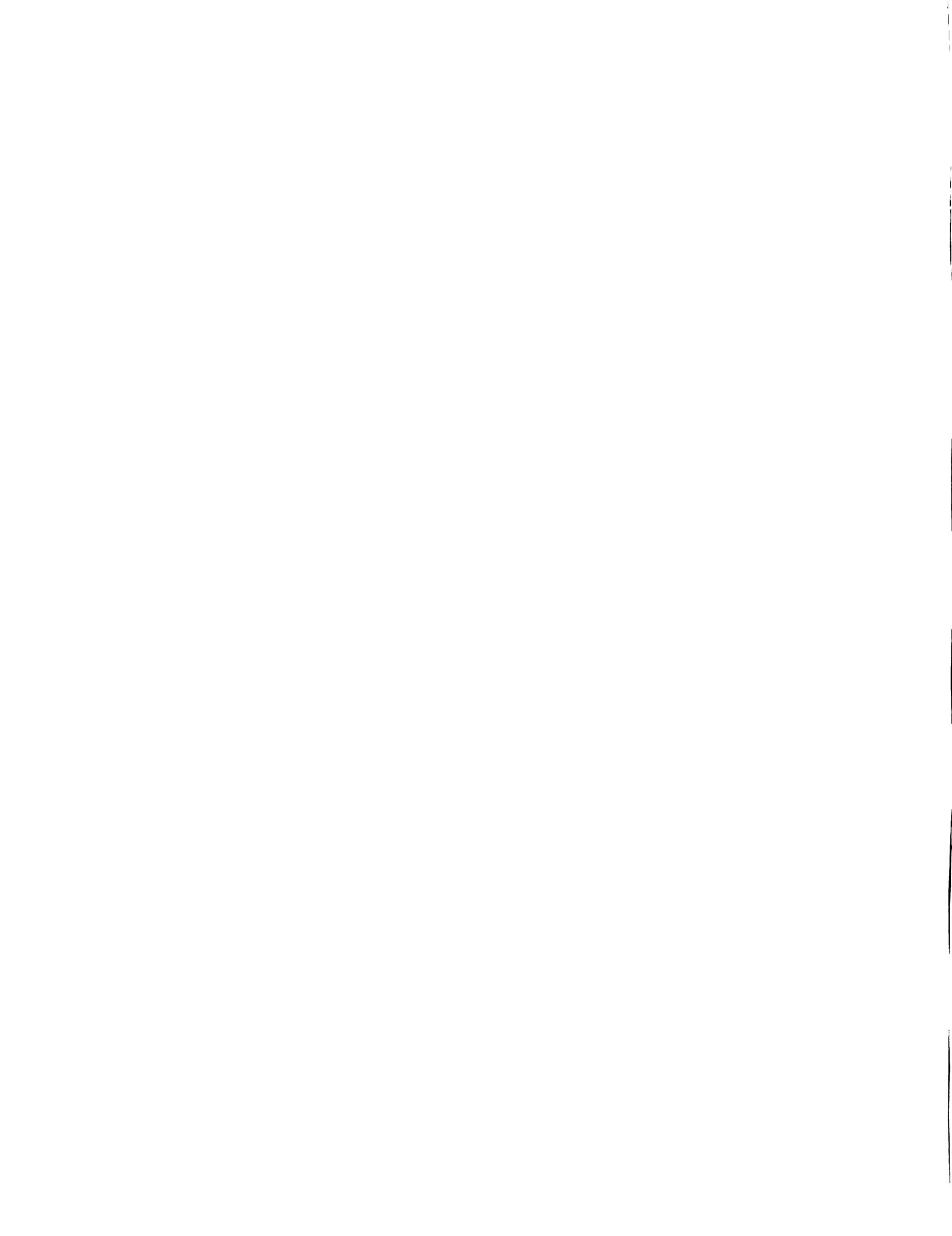
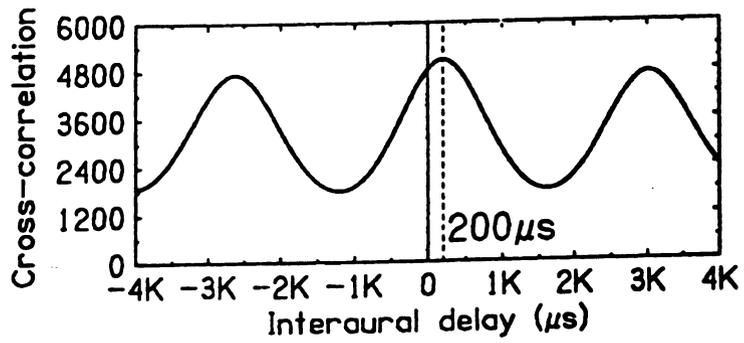
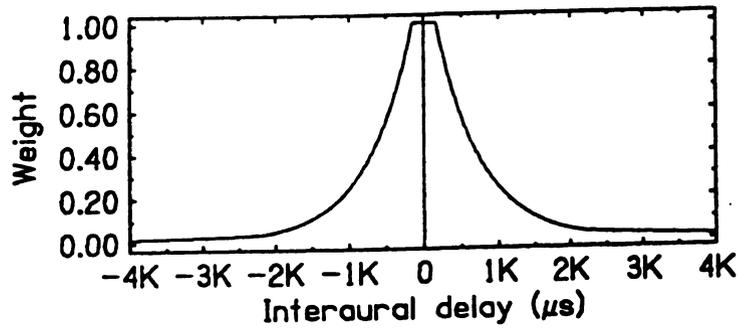


Figure III.18. Binaural processing in the model. Left and right channels of a critical band are cross-correlated without normalization (top panel) for delays between ± 4 ms. The peak falls at the experimental ITD of $200 \mu\text{s}$. To eliminate the contribution of larger delays, the cross-correlation values are weighted by Colburn's $P(\tau)$ function (center panel), which weights low delays heavily. The weighted correlation function (bottom panel, shown with un-weighted function for comparison) is used to calculate a centroid (labeled with an "x") along the delay axis.

Cross-correlation



Apply Colburn's $P(\tau)$ function to weight smaller delays



Calculate centroid of the weighted CC function

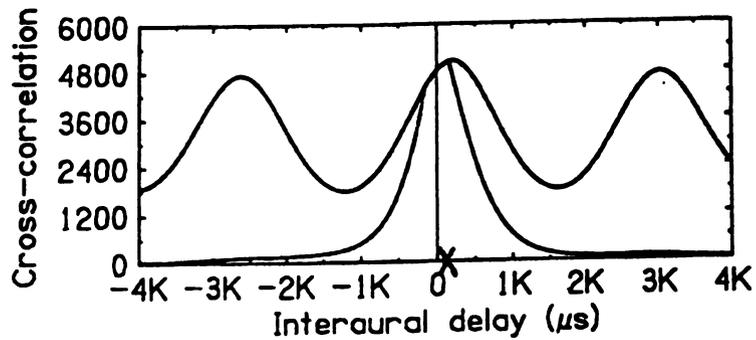


Figure III.18

Figure III.19. Model calculation results. The critical bands tested in the model are found on the x-axis, with corresponding center-frequencies along the top of the graph. The y-axis show the lateral position of the centroid in μs , with 0 being the center of the head and values above shifted off to the right. The points represent twenty model calculations performed for each critical band. The coherence value γ for this example is 0.5.

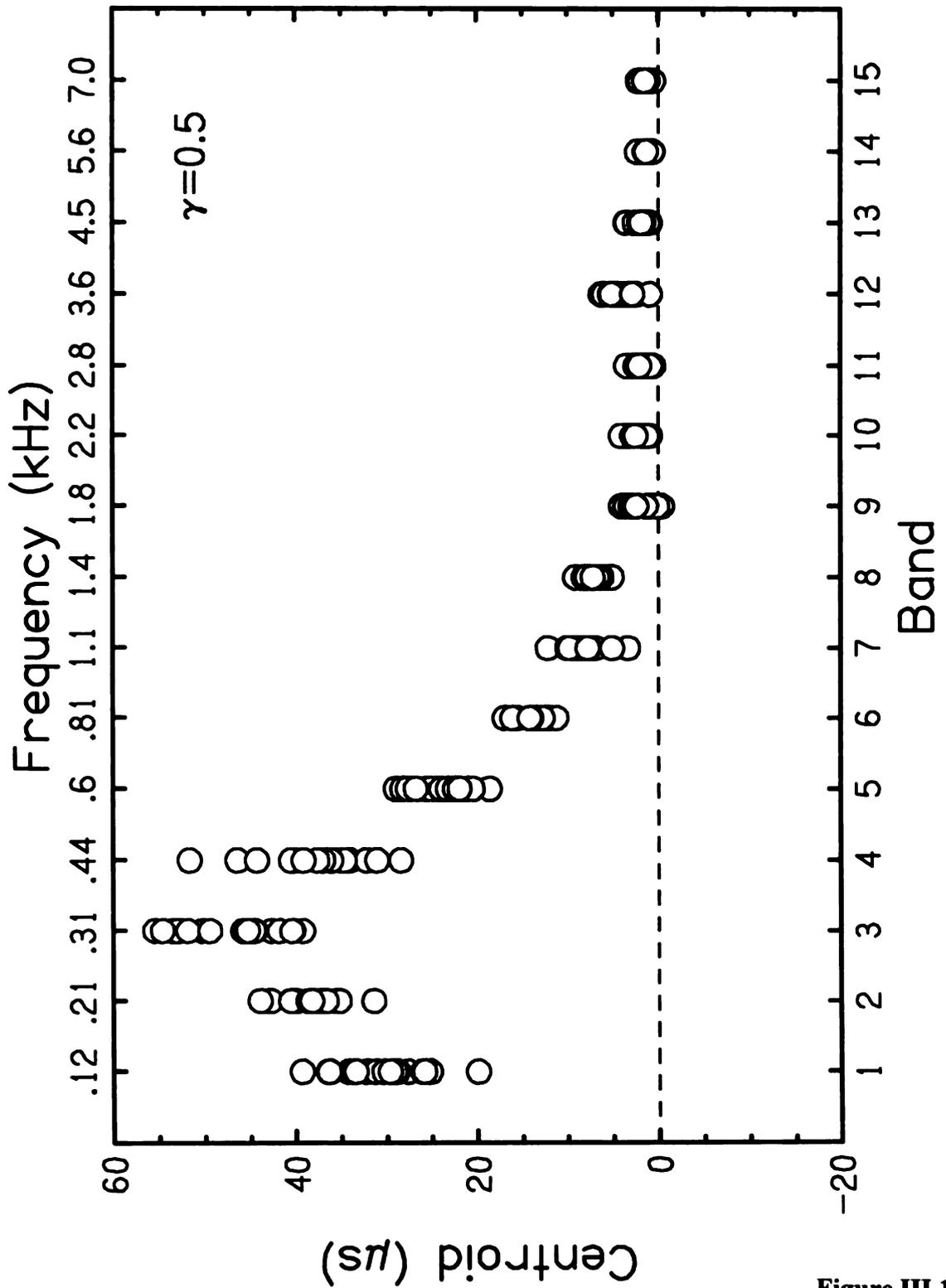


Figure III.19

Figure III.20. Model predictions. The range of tested high-pass cutoff frequencies (seven values from 0 to 4000 Hz) are along the x-axis. The y-axis indicates the interaural coherence necessary to correctly lateralize the internal image 79% of trials. The pentagon symbols mark thresholds for a 200- μ s ITD in Experiment 3. The closed triangles show thresholds predicted by the model without inclusion of neural noise. The open circles are thresholds predicted once estimated neural noise is added.

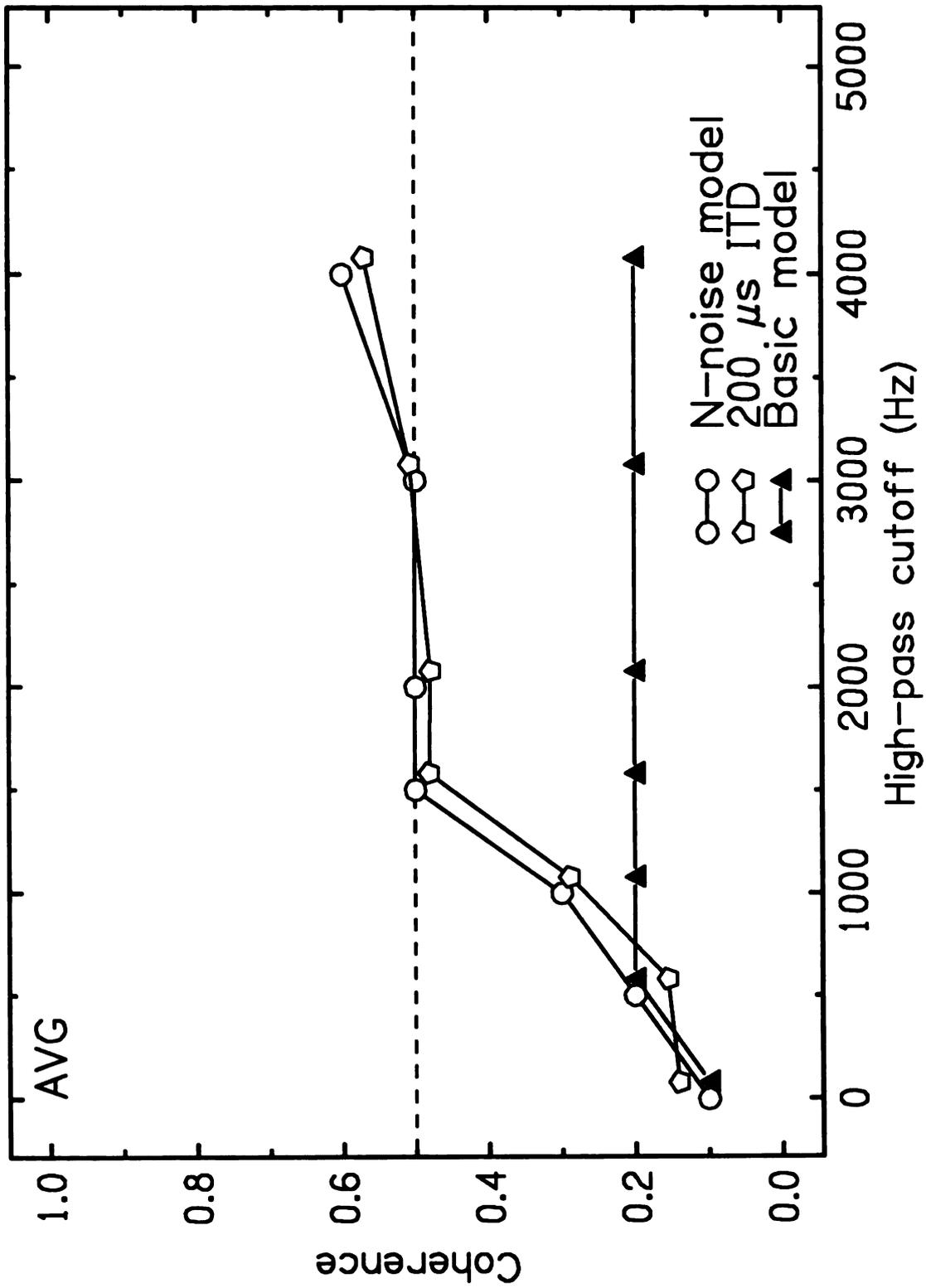


Figure III.20

Chapter IV: Intensity-based lateralization as a function of coherence

Introduction

The goal of the interaural level difference (ILD) experiments described in this chapter was to study how the human auditory system uses ILD information to localize a broadband noise stimulus. Specifically, they were designed to establish how ILD discrimination was affected by interaural coherence, if at all. The usefulness of interaural time difference (ITD), another binaural cue, increases with higher interaural coherences (Stern and Trahiotis, 1995). However, ILD and ITD are disparate cues, as evidenced by the way the superior olive processes them differently (Pickles, 1982) and the variability of time-intensity trading (Stern and Trahiotis, 1995). It is not clear *a priori* whether coherence values should influence ILD discrimination performance.

The above question readily applies to the problem of localization in rooms. While listeners can use both ITD and ILD cues in the presence of room reflections, there is always some interaural decorrelation. In a large lecture hall (reverberation time presumably less than a second), Lindevald and Benade found that signals to the two ears are uncorrelated for all frequencies above 500 Hz (1986). Binaurally uncorrelated signals provide no useful ITD cues, because they lack common temporal features in both ears. Thus, the ILD cue becomes paramount to sound localization for steady-state signals in highly-reflective rooms. Understanding the impact of decorrelation on ILD discrimination is quite a significant objective.

Models

When considering the question of coherence and ILD, several different models suggest themselves. The first of these models is dubbed the “level-meter”. This model assumes the existence of independent level meters that measure the power at both ears, and ILD is determined when some central processor calculates the difference between them. The level meter has no way to detect coherence, which is an aspect of the waveform. Instead, it calculates a power by integrating the signal over time, losing all waveform information as a result. Thus, the level-meter model predicts that ILD discrimination should be unaffected by changes in coherence.

The second model takes a less simplistic view of coherence. The “instantaneous ILD fluctuation” model notes that a binaurally coherent signal, time-shifted so the ITD is zero, will have the same ILD at any point along the signal waveform. This is because the left and right waveform variations are synchronized, and any disparity between them would be due to a level difference. However, any incoherence will cause power fluctuations in the two ears to grow more independent, leading to random variability in ILD from moment to moment. If the listener can track these “instantaneous” ILDs over time, then the fluctuations could lead to uncertainty in his or her judgments, and thus make the task more difficult. In terms of Signal Detection Theory (TSD, see General Appendix B), incoherence would actually increase the standard deviation σ , decreasing the sensitivity index d' . This model predicts that listeners should require lower thresholds for ILD lateralization for coherent noise (cross-correlation = 1) than for

incoherent (cross-correlation = 0) and anticorrelated (cross-correlation = -1) noises, and the amount of ILD variability determines those thresholds.

The third model concentrates on the intracranial images produced by stimuli. The auditory system interprets a perfectly-coherent noise as a compact image, where “compact” indicates a narrow Auditory Source Width (ASW). The movement of that image (due to a change in binaural cue) is readily identified because listeners can detect a clear lateral position for the noise. As the noise becomes less coherent, that image becomes less compact, increasing the ASW. At a coherence of 0, it spreads across the entire lateralization axis, “filling” the listener’s head. Instead of identifying the motion of a localized image, a listener would have to follow the image’s “center of mass” to make a lateralization judgement. For anticorrelated noise, as described below, the image is similar to that of incoherent noise, but doesn’t quite fill the whole head. The image model predicts listeners will have lower thresholds for stimuli that produce intracranial images of smaller ASW. Thus, listeners should perform best with coherent noise, worst with incoherent, and anticorrelated should fall in-between.

The relationship of ILD sensitivity and binaural coherence has been previously reported in unpublished work by Grantham and Ahlstrom (1982). They raised a question about the image model, suggesting that ASW may impact a listener’s ability to lateralize. When the stimuli were broadband (bandwidth presumably greater than one octave), listeners had consistently lower ILD thresholds for coherent noise (low ASW) than for incoherent noise (high ASW). Their results for narrowband stimuli (bandwidth of 0.4 octave) weren’t as clear, as two out of three listeners performed better with coherent noise at short durations (30 ms), but exhibited no significant effect due to coherence at

longer durations (500 ms). The experiments detailed in this chapter are meant to expand on their work with both similar and newly-introduced stimuli.

Methods

A series of experiments explored the ILD question by testing ILD thresholds for various stimuli. The stimuli included are as follows (with the coherence, or maximum cross-correlation, value γ indicated for each):

--Interaurally correlated noise ($\gamma = 1$). This stimulus was diotic (except for level), producing a compact image in the listener's head.

--Interaurally uncorrelated noise ($\gamma = 0$). This stimulus was generated by two independent noise sources, one for each ear, and formed a head-filling image.

--Interaurally anticorrelated noise ($\gamma = -1$). This stimulus was identical to the correlated noise above, except the signal to the right ear was inverted, causing the listener to hear images located at the left and right extremes that extended towards the center.

The ILD experiment employed a two-interval, two-alternative forced-choice (2I2AFC) task, in which the listener identified changes in lateralization. A reversal of ILD (e.g. from +2 dB to -2 dB) between the two intervals caused the internalized image to move, and listeners responded by indicating the apparent direction of motion.

Runs consisted of trials presented at a level of 60 dB SPL, with two 500-ms intervals per trial. Intervals were turned on by matched voltage-controlled amplifiers,

which were in turn directed by a single gating signal with rise/fall times of 30 ms. The simultaneous left and right-ear stimuli ensured that interaural onset differences would not affect lateralization. The total number of trials depended on the listener's performance during the run, as explained below. On the first interval of a trial, either the right or left channel was attenuated or amplified, with each possibility having equal *a priori* probability. At the start of a run, the level shift in one ear was ± 2 dB SPL. After a 500-ms delay, the second interval played, with levels for right and left ears reversed. This succession of ILDs would cause the stimulus image to move laterally in the listener's head. See an example chart of levels by interval in Figure IV.1, which simulates a trial where the image passes from right to left. The x-axis represents time, while the y-axis represents stimulus level in dB. The dashed line indicates the standard level of 60 dB. The figure exhibits two possible trials (a) and (b) that were equally likely to occur during the run, and were expected to present the listener with the same impression: that the image moved from right to left. The listener was then prompted to respond by pressing one of two buttons, indicating the direction of stimulus motion. The run would not continue until the listener had made a decision.

The ILD was controlled by a 3-down 1-up staircase (see General Appendix A). When the listener correctly identified the direction of ILD motion three trials in a row, the program reduced the ILD by an increment, making the "movement" smaller and more difficult to detect. However, only one incorrect answer caused the program to increase the ILD by an increment.

If the ILD was greater than 1.5 dB (as at the start of the run), the increment was 0.5—that is, the ILD would be increased or decreased by 0.5 dB according to the

staircase. If the ILD was between 1.5 and 1 dB, the increment was 0.2 dB, and reduced to 0.1 dB when ILDs fell below 1 dB. Since listeners spent the majority of a run with ILDs less than one, this arrangement allowed them to progress rapidly downward from the starting ILD of ± 2 dB, then provided high resolution in the region of interest. The lowest ILD possible in this experimental setup was 0.1 dB.

Each ILD run presented one of three perceptually-different stimuli, which produce distinct internal images: coherent noise leads to a compact image, incoherent noise is broad, and anticorrelated is somewhere in between. As mentioned in Appendix B, discrimination experiments showed that listeners could distinguish between them more than 75 percent of the time.

The level meter model predicts no difference in ILD thresholds for the above stimuli, due to its coherence insensitivity. In addition, if the ILD information is based simply on levels in each ear, then the lateralization task can be no more accurate than a monaural difference-limen-intensity (DLI) task. This result was previously suggested by von Békésy (1930). To provide a thorough test of the model, the experiments included one more stimulus:

--DLI (Difference Limen in Intensity) ($\gamma = 1$). The two ears received diotic noise.

The DLI stimulus was used in a task that required listeners to identify the louder of two intervals, rather than an ILD-induced motion. Runs of the DLI task found the smallest detectable difference in interval levels for each listener. Aside from the differences in the stimulus and the task, the DLI experiment followed the same procedure

as the lateralization experiments. A level chart for an example DLI run can be found in Figure IV.2, which is of the same format as Figure IV.1.

Runs featuring each of the four stimuli were intermingled during the experiment, ensuring that listeners were not exposed to the same stimulus on two successive runs. Due to prior tests in the lab, and the results of Grantham and Ahlstrom (1982), it was expected that any coherence-related effects would be small, and thus the experiments required a high amount of precision from the data. When a run resulted in an average ILD with standard deviation greater than 0.3 dB, it was redone to ensure a smaller variability. Following a few training runs, the final six runs for each listener at each condition (that met our above standard for variability) were used as data, and averaged together. Thus, each listener provided data in 24 runs. Each experiment took on unique modifications to the stimuli, as detailed below.

Listeners performed runs of the ILD experiment in a double-walled Acoustic Systems soundroom, wearing Sennheiser HD-480 headphones to hear the stimuli. The letters R, W, H, K, L, Z, and M will refer to the listeners in this experiment. K and W were females age 21 and 19; R, M, L and Z were males age 19 to 28, and H was a 60-year-old male. Runs lasted between two and four minutes.

Experiment 1: Broadband signal

Results

The first task used thermally-generated Gaussian noise that was both white and broadband, then low-pass filtered at 10 kHz (with a rolloff of -48 dB/octave).

Experimenters generated all necessary stimuli using two independent noise generators passing through two identical filters. This experiment involved listeners R, Z, H, K, M, and W.

Refer to Figure IV.3 for the data collected during the broadband experiment. The symbols corresponding to each stimulus type are listed in the legend. Results from the four tasks are grouped by listener, as identified by letter code along the x-axis. The mean threshold ILD in decibels (determined by the staircase targeting 79% correct) is along the y-axis. The length of the error bars is twice the standard deviation (n-1 weight). Note the shaded area below 0.15 dB on the plot. Any data within this region should be considered unreliable, because the experiment was limited to a minimum ILD of 0.1 dB and a minimum increment of 0.1 dB. Any listener who could discriminate that well would essentially cause the staircase to fail. None of the listeners ever ran into that “floor”—H reached it sporadically, but results from those runs showed that it wasn’t holding him back appreciably. Approximately 6.5% of all listener runs in this experiment were discounted due to a standard deviation higher than 0.3 dB.

Discussion

Looking at the broadband data, it is possible to make some general observations. The DLI value is the highest (and thus worst) threshold for every listener, but more significantly for R and Z at the left. R does best at the coherent task, while others perform best at the anticorrelated task. Incoherent results are consistently worse than for the other two binaural stimuli, but better than the DLI results. The thresholds averaged

across listeners for each stimulus are found in Table IV.1. They provide general evaluation of the relative difficulty of each task.

Table IV.1 - Broadband ILD thresholds according to stimulus type				
Stimulus	Anticorrelated	Coherent	Incoherent	DLI
Mean value	0.398	0.451	0.585	0.773

Threshold ILDs in dB from Experiment 1, averaged across listeners for each stimulus type.

In most cases, the large error bars made it difficult to estimate differences in performance. For this reason, the experimenters decided to perform a statistical study through analysis of variance (ANOVA). The ANOVA indicated that there was indeed a significant difference among the four conditions ($F(3,15) = 16.7, p < 0.01$). The results of post-hoc “general linear model univariate pairwise comparisons” for each combination of conditions can be found in Table IV.2. Those comparisons with a p-value that indicates significant difference (i.e. less than 0.05) are marked with an asterisk.

Table IV.2 - Broadband pairwise comparisons				
p-values	Anticorrelated	Coherent	Incoherent	DLI
Anticorrelated		0.501	0.008*†	0.002*†
Coherent			0.002*†	0.012*
Incoherent				0.052

The p-values for comparison between factor levels in experiment 1 (broadband 10 kHz)

* significant difference at the 0.05 level

† significant difference at the 0.05 level after Bonferroni correction

According to the pairwise comparisons, listener results were significantly different for all pairs of stimuli except for coherent/anticorrelated and incoherent/DLI. These relationships between the various stimuli indicate that coherence may, in fact, have an effect on ILD discrimination. Note, also, that the stimuli with images that didn’t fill the listeners’ heads (coherent and anticorrelated) were clearly easier to lateralize than

incoherent noise, but neither proved better than the other. This result agrees with Grantham and Ahlstrom's result for a coherent/incoherent comparison (1982).

Incoherent noise did not seem an improvement over the monaural DLI task.

However, the previous analysis might require some evaluation of the experimental goals. A 95% confidence interval may be sufficient for one comparison, but this experiment was testing for *any* difference between four sets of values. The null hypothesis stated that all four conditions were equivalent, and any one of the six comparisons could disprove that. It seems that stricter confidence intervals would be in order, and that is provided for by the Bonferroni correction (Dunn, 1961). By dividing the significance of each mean-difference pair by six (the number of comparisons), it essentially assigns a new alpha level of $0.05/6=0.008$. This is a very stringent requirement, and once the Bonferroni correction takes effect, the results change somewhat. In the case of this experiment, applying the correction eliminates the difference between coherent and DLI tasks. P-values that are significant under the Bonferroni correction are marked with a dagger symbol on Table IV.2. With these results, it is difficult to make a case for the level-meter model. It would seem that the models that deal with ASW and ILD fluctuation are best supported, since the uncorrelated noise was significantly more difficult to lateralize than the more compact stimuli.

Experiment 2: Low-pass filtered signal

Introduction and Method

The purpose of Experiment 1 was to test the effect of coherence, an ITD-based cue, on ILD discrimination. However, in the broadband regime, ITD tends to play the strongest part at lower frequencies. A well-established aspect of binaural hearing is that ITD coherence cues provide the most useful information below about 1500 Hz (Moore, 1989), even though the auditory system is sensitive to envelope coherence at higher frequencies (see Chapter II) where waveform cues are unusable (Henning, 1974). On the other hand, human ILD sensitivity is generally independent of frequency (Yost, 1981).

Because the stimulus in Experiment 1 was low-pass filtered at 10 kHz, most of its power was located in frequencies above the limit where ITD discrimination (and thus, coherence detection) was strongest. Thus, using the same noise as Experiment 1 and eliminating frequencies above 1 kHz by lowering the cutoff, Experiment 2 ensured that the stimulus fell entirely within ITD's most effective region. The Brickwall lowpass filter used had a -115 dB/octave rolloff. If the interaural coherence were to have an effect on the process of ILD discrimination, such a change in stimulus would afford it the best chance. Also, the switch to narrowband noise may eliminate the observed effect of coherence on ILD threshold, as reported in Grantham and Ahlstrom (1982). The listeners involved were R, Z, H, L, M and W.

Results and Discussion

Figure IV.4 plots data from the lowpass (1 kHz cutoff) experiment. The between-subject variation is much more pronounced for this task, which also proved more difficult overall than the broadband case (as shown by the consistently higher thresholds in Table IV.3 below compared to Table IV.1). The four listeners on the left seem to follow a pattern of low thresholds for coherent and anticorrelated, then worse for incoherent and DLI. However, M and W on the right are completely different. Overall, listeners seemed to do best with a coherent stimulus, and all but one had worse performance with an incoherent stimulus. This is a minor change from the broadband experiment, where anticorrelated noise produced the lowest threshold. Approximately 6.8% of all listener runs were discarded due to a standard deviation higher than 0.3 dB.

Stimulus	Anticorrelated	Coherent	Incoherent	DLI
Mean value	0.733	0.679	0.835	1.021

Threshold ILDs in dB from Experiment 2, averaged across listeners for each stimulus type.

Using the ANOVA to identify a pattern, the test finds a marginally-significant difference among the four conditions ($F(3,15) = 3.35$, $p = 0.045$). Pairwise comparisons show only one significant difference: between coherent and incoherent noises. The resultant p-values can be found in Table IV.4. Note that use of the Bonferroni correction (again, requiring a level of $0.05/6 = 0.008$) eliminates the only pairwise difference identified in this experiment.

Table IV.4 - Narrowband pairwise comparisons				
p-values	Anticorrelated	Coherent	Incoherent	DLI
Anticorrelated		0.294	0.199	0.163
Coherent			0.017*	0.070
Incoherent				0.239

The p-values for comparison between factor levels in experiment 2 (low-pass 1 kHz)

* significant difference at the 0.05 level

† significant difference at the 0.05 level after Bonferroni correction

This experiment provides support for the level-meter model, which predicts no significant differences between any two conditions. The results also agree with the conclusions of Grantham and Ahlstrom (1982), as suggested earlier.

Experiment 3: Low-pass filtered signal with randomized standard level

Introduction and Method

While Experiments 1 and 2 provided useful insight, there existed a possibility that the listeners were not completing the task as instructed. Referring to Figure IV.1, it becomes clear that the ILD “movement” can be identified monaurally. If the listener concentrated solely on the left ear, for instance, he or she could make a loudness comparison between intervals to make a decision. Granted, such a strategy would have a disadvantage when compared to binaural lateralization, because monaural level discrimination is subject to a greater variance than ILD discrimination. Thus, the monaural strategy would be less accurate. Nevertheless, it is important to establish what strategy the listeners chose. The final experiment was designed for comparison with Experiment 2, using the low-pass 1 kHz noise as defined above, but also introducing randomization of standard level between intervals. This strategy is similar to that of

Grantham (1984) and Koehnke *et al.* (1986), who were also trying to prevent monaural discrimination in their respective ILD sensitivity experiments. The randomization depended on an experimenter-set range value of 5, causing the standard level in each interval to vary among integer values within 5 dB of the 60 dB experimental norm. The distribution of selected levels was rectangular, meaning each possible choice had equal probability. This variation in standard level does not affect the relative ILD shifts.

In the example trials of Figure IV.5 (again, formatted as Figure IV.1), the levels are those of Figure IV.1 with a +2-decibel shift on the first interval, and a -1-decibel shift on the second. If the listener attempted to do the task with only information from the left ear, he or she would conclude that the image had moved from left to right. However, using binaural ILD information, it is clear that the image moved right to left. The introduction of a random standard has confounded the possible monaural cue. If listeners did not use monaural cues for the task in Experiment 2, then the results of Experiment 3 should correspond closely. If they still attempted to concentrate on a single ear, their thresholds would increase dramatically. Green (1988) estimated that a level randomization of X dB would necessarily require a threshold level of

$$L = X \left[1 - \sqrt{2 - 2P_C} \right]$$

with P_C representing the percentage of correct responses in a two-alternative task. Given the selected X of 5 dB, and P_C of 79% targeted by the staircase, Green would predict a threshold of 1.8 dB. Note that there was no change in the DLI task from the previous experiment, i.e. level randomization was not applied to DLI runs because it would have interfered with the task. Listeners for this experiment were R, Z, H, L, M and W—the same as for Experiment 2.

Results and Discussion

Data are summarized in Figure IV.6. Note how consistent the relative positions of coherent, anticorrelated and incoherent data were for the left four listeners. M and W once again do not fit the trend. They did, however, match all listeners but H by performing worst in the DLI case. Comparing the results for other (non-DLI) stimuli (as found in Table IV.5) to results without randomization from Experiment 2 (in Table IV.3), the pattern is similar. Thresholds for individual listeners both increased and decreased in the switch to random standard level. In addition, the thresholds do not approach the value of 1.8 dB predicted above. The evidence would suggest that listeners were indeed using binaural cues in Experiments 2 and 3. One curious result: although the DLI task was unchanged, some listeners' results were not consistent with the prior values. Listener R, in fact, achieved much higher thresholds that were too variable to be included in our results, thus his closed circle is off-plot. Approximately 9.2% of all listener runs were discarded due to a standard deviation higher than 0.3 dB.

Stimulus	Anticorrelated	Coherent	Incoherent	DLI
Mean values	0.766	0.756	0.880	1.152

Threshold ILDs in dB from Experiment 3, averaged across listeners for each stimulus type.

The ANOVA indicates a significant difference among the four thresholds ($F(3,15) = 6.53, p < 0.01$). The p -values for pairwise comparisons in Table IV.6 show that coherent and anticorrelated results were significantly different from DLI, and coherent data were significantly different from incoherent. These pairs are similar to those

identified in the broadband experiment. As with the Experiment 2 results, applying the Bonferroni correction eliminates all significant differences among thresholds.

p-values	Coherent	Anticorrelated	Incoherent	DLI
Anticorrelated		0.843	0.271	0.041*
Coherent			0.009*	0.029*
Incoherent				0.235

The p-values for comparison between factor levels in experiment 3 (low-pass, randomized level)

* significant difference at the 0.05 level

† significant difference at the 0.05 level after Bonferroni correction

Summary/Discussion

The inclusion of DLI as a test of the level-meter model

The level-meter model assumed that ILD detection was based solely on independent level measurements in each ear. If this were the case, interaural cross-correlation should have no effect on ILD thresholds. The model predicts equivalent thresholds for stimuli of all three correlations, as well as in the DLI condition. The reasons to include the radically-different DLI task in our comparison are as follows: in TSD, measurements are selected from a Gaussian probability density function (PDF). See the leftmost two plots of Figure IV.7 for example intensity measurements in left and right ears. Upon receiving a stimulus, the probability that the auditory system will judge its intensity to be a particular value is determined by a Gaussian curve. The most probable measured value in the left ear is marked μ_L (the location of the probability peak along the intensity axis), and σ_L is the intensity uncertainty caused by internal and

external noise. Similar variables are noted on the right-ear graph. It is important to mention here that there is no reason to assume that the intensity uncertainty is different between left and right ears.

$$\sigma = \sigma_L = \sigma_R$$

Note that μ_L is not equal to μ_R , indicating an ILD. In order to get the probability density for $ILD = L_R - L_L$, convolve the left and right ear curves to establish a new Gaussian of mean position $\Delta = \mu_R - \mu_L$ and standard deviation determined by that of the original measurement ($\sqrt{(\sigma^2 + \sigma^2)} = \sigma\sqrt{2}$) (Hartmann, 1997). The ILD decision variable d' is defined by these two variables that describe the Gaussian (Green and Swets, 1966) thus:

$$d' = \frac{\Delta}{\sigma\sqrt{2}}$$

The magnitude of the decision variable is directly related to the probability of a correct judgment. Increasing the mean distance from zero Δ or reducing the uncertainty σ will improve performance.

The experimental task was comprised of two intervals, with equal and opposite ILDs (note the center two plots on Figure IV.7). To model the interval comparison, convolve both ILD probability densities (with means Δ_1 and $\Delta_2 = -\Delta_1$, respectively) to establish the density for one trial (see the rightmost plot on Figure IV.7). The Gaussian's peak position is defined by $\Delta_1 - (-\Delta_1) = 2\Delta$, dropping the subscript for simplicity, and the standard deviation becomes $\sqrt{(2\sigma^2 + 2\sigma^2)} = 2\sigma$. The decision variable for one trial is:

$$d' = \frac{2\Delta}{2\sigma} = \frac{\Delta}{\sigma}$$

The d' for comparing the levels of two intervals in a monaural DLI task is $\Delta/\sigma\sqrt{2}$, for the same reasons as explained for individual ILD measurements above. But when the same task is done binaurally, with a diotic noise, the listener has more information with which to make a judgment. A paper by Jesteadt and Weir (1977) shows that the d' for the diotic case is higher than that of the monaural by a factor of $\sqrt{2}$, since the listener has twice as many stimulus channels. The d' improves to Δ/σ , the same as above for the ILD discrimination task. Therefore, an additional consequence of the level meter model suggests that the DLI task, using a diotic stimulus, should have thresholds equivalent to all the ILD stimuli.

Experimental inconsistency with the level-meter model

The thresholds of Experiments 1-3 all seem roughly comparable. Within each experiment, results from the four conditions did not differ more than 0.5 dB. To a first approximation, the level-meter model performed well. However, the presence of systematic and statistical differences argues against the level-meter model. While its predictions are nearly correct, the disparity between DLI and correlated thresholds is a glaring inconsistency. DLI thresholds tended to run higher than those for all ILD conditions, as evidenced in Tables IV.1, IV.3, and IV.5. The difference could be due to several factors missing from the model: additional neural processing applied to ILDs, the diverse natures of level and lateral discrimination, or perhaps other initial assumptions were erroneous. Additionally, smaller (but still interesting) differences in threshold were found between the various correlations. At any rate, it would seem that the level-meter model could use some adjustment.

The evolution from level-meter to loudness-meter

The level-meter model assumed that listeners only used power measurements to establish ILD, and because power is time-averaged, correlation makes no difference. However, the fluctuations introduced by anti- or uncorrelated noise are only truly eliminated if the averaging time is infinite. Work by Grantham and Ahlstrom (1982) revealed an indication of this effect. They found equivalent ILD thresholds for both correlated and uncorrelated noises when the stimulus duration was 500 ms, but when they shortened it to 30 ms, the uncorrelated noise resulted in a clearly higher threshold. Their results suggest that listeners can indeed suppress the variability of uncorrelated interaural measurements given sufficient averaging time. This leads to the consideration that thresholds for uncorrelated noise may be limited by the auditory system's inherent integration time, if the stimulus duration is long enough not to be a limiting factor (as in our experiments). Such information can account for the slight variation among our three experimental ILD thresholds, and harkens back to the "instantaneous ILD" model, which accurately predicted relative thresholds for coherent and incoherent noises. It is possible to correct the predictions of the level-meter model by including half-wave rectification, compression, and a realistic integration time for detection, all of which influence the perception of loudness. This effectively changes the model to a "loudness-meter."

Work by Hartmann and Constan (2002) on the loudness-meter resulted in a quantitative model that used the sensitivity index d' and the signal variance of each correlation condition to predict ILD thresholds. The results are included in Figure IV.8. This figure revisits data from Experiments 2 and 3, allowing straightforward comparison

for the model and all stimuli. Model-based calculations are indicated on the threshold scale by horizontal lines for each experimental condition. The dotted line represents predicted threshold levels if the model assumes a 200-ms integration time. The solid and dashed lines represent 300-ms and 400-ms durations, respectively. The closest agreement with Experiment 2 occurs for 300 ms, a value that agrees with Plomp and Bouman (1959) for similar low-frequency stimuli. The loudness-meter model also predicts that thresholds should decrease with increasing bandwidth, to 0.45 dB for a broadband correlated stimulus. Again, this value is in close agreement with our empirical results (see Table IV.1).

Conclusions

The level-meter model provided a close approximation to experimental results. The ILD thresholds across stimuli were quite close, and did not vary more than 0.5 dB. However, there was some subtle variation according to interaural correlations, indicating that the model could use some adjustment. The introduction of rectification, compression and integration time provided an interpretation of the signal, as seen by the auditory system. These modifications transformed the level meter to a loudness meter, which was able to account for the differences in threshold.

Appendix A: “The strange case of Listener C”

Listener K, a subject in Experiment 1, was unavailable for Experiment 2. This led to the search for a replacement, testing and rejecting three more listeners (males C and X, female J) whose 1-kHz results consistently exceeded the standard deviation limit of 0.3. This problem was unexpected for listener C, as he had produced some of the lowest thresholds at the 10-kHz condition. C had little difficulty with the 1-kHz condition, continuing to lateralize at low ILDs for all stimuli except correlated noise. The particular combination of correlated noise and narrow bandwidth proved more difficult, and C’s results varied wildly from run to run.

Interviews revealed that listener C perceived what he termed a “hole” along the lateralization axis, representing an absence of noise. The hole’s position and movements were always opposite those of the correlated image. Occasionally, listener C would become distracted by the hole, following its movements opposing the experimental stimulus. The distraction was intermittent; often, C would complete the majority of a run before the hole captured his attention. Thereafter, he would answer incorrectly on every trial, even at large ILDs. Experimenters reduced the stimulus duration, increased the stimulus volume, and introduced an ITD, but failed to eliminate the hole. Once the low-pass cutoff was increased to 2 kHz, the hole appeared less frequently, and vanished altogether when the cutoff exceeded 3 kHz. No other listener experienced this phenomenon, and we cannot currently explain its apparent existence.

Appendix B: Noise-type discrimination project

Introduction

As an aside to the ILD project, which measured listeners' ILD thresholds for N_0 , N_π , and N_u (the coherent, anticorrelated, and incoherent stimuli, respectively), an informal experiment attempted to verify the basic assumption prompting the project: that listeners could distinguish between the three kinds of noises, most likely on the basis of coherence. These stimuli must affect the binaural processing centers in very different ways, and there was every reason to believe that listeners could easily discriminate among them. The work of Pollack and Trittipoe (1959) provides evidence that listeners can do so, but indicates that the ability to discriminate may be very dependent on signal level, duration and frequencies. The letters H, X, J, C, R, L, and Z will refer to listeners involved in this experiment. J was a female age 42; R, C, X, L and Z were males age 19 to 28, and H was a 59-year-old male.

Experiment 1

The listener was required to identify the less coherent (according to ASW) of two intervals in a 2IFC task, while the two intervals presented a round-robin selection of coherent, anticorrelated and incoherent noise. Thus, on one trial the stimuli may have been coherent noise and anticorrelated noise, while the next trial included anticorrelated noise and incoherent noise. These noises were white, thermal, and Gaussian, ranging from 20 to either 1 kHz or 10 kHz, where they were low-pass filtered by Frequency Devices 901Fs with a -48 dB/octave rolloff. The experiment was run at both frequency

ranges because the results would be significant to both broad- and narrowband ILD experiments. Stimulus characteristics remained unchanged throughout the run, and the two stimuli for each trial were selected manually by the experimenter. Both intervals lasted 500 ms, with a 500 ms gap between. After hearing a trial, the listener recorded which interval seemed “broader” next to the corresponding trial number on a list. The run went on for 18 trials, including six of each stimulus pair. Listeners performed 3 runs at each frequency range. The subjects were H, J, and C.

The results are found in Table IV.7 below, which contains each listener’s discrimination results for each stimulus pair. The ratios listed in each cell show the number of correct responses versus the number of trials for that particular pair. Listeners were expected to have no trouble with the task, rating incoherent (Nu) noise as “broader” than anticorrelated ($N\pi$) and coherent (No), and anticorrelated noise as “broader” than coherent. The listeners met these expectations with virtually perfect scores for all combinations except those comparing coherent and anticorrelated noise, for which they were generally below the 75% correct detection threshold. This seemed counter-intuitive, since informal experiments suggest that it is easy to distinguish between No and $N\pi$ stimuli.

1K	No vs Nu	$N\pi$ vs Nu	No vs $N\pi$
H	18/18	17/18	16/18
J	18/18	18/18	13/18
C	18/18	17/18	4/18
10K	No vs Nu	$N\pi$ vs Nu	No vs $N\pi$
H	18/18	18/18	13/18
J	18/18	18/18	7/18
C	18/18	18/18	7/18

Results from the informal noise-type discrimination runs. Trials for each of the three stimulus pairs were intermingled during the run.

Experiment 2

The anomalous results of Experiment 1 prompted a more formal investigation with computer-controlled runs and automated response collection, in which the listener was asked to compare just two of the three stimuli. During a run of 40 trials, the listener would indicate which interval had a greater ASW by pressing one of two buttons on a response box. The left button corresponded to the first interval, and the right to the second interval. Additional listeners joined those that participated in Experiment 1. The subjects were H, X, J, C, R, L and Z.

Stimulus duration and frequency distribution were the same as the 10 kHz case described above, and the results are placed in Table IV.8 below. Again, all listeners performed excellently at No/Nu and $N\pi$ /Nu discrimination. In the case of the No/ $N\pi$ stimulus pair, H, X and C missed one or two of the first few trials, and then answered correctly for the remainder. Subsequent runs resulted in perfect scores (100% correct). Listeners R and L performed perfectly throughout. Listener J's overall score of 75% does not fully represent the interesting progression of her runs. She exhibited a longer learning process than other listeners, beginning with scores near 50% and gradually improving each run until she was consistently scoring 95% correct. However, when the set of No/ $N\pi$ runs was interrupted by an No/Nu run, she seemed to lose some of her previous accuracy on the next No/ $N\pi$ run. J would then regain more accuracy with each repeated run. Listener Z, while lacking the amount of data that J accumulated, also exhibited a learning process, though he did not achieve as high an accuracy. He did improve from 50% correct to about 75% correct in his final runs.

Table IV.8 - Formal noise-type discrimination			
	No vs Nu	N π vs Nu	No vs N π
H	40/40	40/40	158/160
X	40/40	40/40	119/120
J	80/80	119/120	450/600
C	40/40	40/40	77/80
R	40/40	40/40	40/40
L	40/40	40/40	40/40
Z	39/40	38/40	87/120

Results for the computer-controlled discrimination runs. Each run featured only one noise pair.

J's and Z's apparent uncertainty raised a possibility that the presence of incoherent noise in the rotation of Experiment 2 served to confuse listeners. On a relatively easy trial with coherent and incoherent noises, the listener would be distinguishing between a compact image and a head-filling noise. On the next No/N π trial, the distinction would not be nearly so clear, and the listener may have difficulty dealing with the range of coherence values. Similar "range effects" have been previously identified for tasks where the listener was asked to discriminate the more intense of two intervals (Purks *et al.*, 1980), and for ITD/ILD localization in large and small ranges (Koehnke and Durlach, 1989). Both articles showed that listeners could correctly identify very small changes in intensity or localization, respectively, as long as the ranges of intensities or differences were small. When presented with larger ranges, the listeners' just noticeable differences grew accordingly.

Experiment 3

To test whether the mixing of stimuli had indeed confused our listeners, Experiment 3 was based on the same format with one change: while one interval was set to be coherent, the other was randomly chosen as anticorrelated or incoherent. The

modification allowed the formal computer-controlled runs to mix up the stimuli as done in the first experiment, and Experiment 3 could study what effect, if any, was introduced by the coherent-incoherent trials. In addition, it allowed experimenters to observe confusion due to “range effects” on a trial-to-trial basis, instead of between runs. The results are summarized below, in Table IV.9. Perfect performance by every listener for No/Nu trials is expected, and therefore uninteresting. Discussion will concentrate solely on results for No/N π trials.

Listeners C, L and X were unaffected by the stimulus switch, continuing to discriminate coherent noise from anticorrelated as well as they did in Experiment 2. Listener J was able to answer correctly 64% of the trials, which was a considerable loss of accuracy from her previous 75% score in Experiment 2. Her results were paralleled by listener Z, who only managed 57% correct when he had neared 73% previously. Listeners H and R averaged only 80% correct for the coherent-anticorrelated trials, compared to near-perfect performance (99% and 100%, respectively) when incoherent noise was excluded in Experiment 2.

	No vs Nu	No vs N π
H	104/104	114/136
X	45/45	34/35
J	106/106	86/134
C	57/57	63/63
R	59/59	48/61
L	49/49	71/71
Z	83/83	44/77

Results for the noise-switched discrimination runs of Experiment 3. Each trial had an equal chance of comparing coherent-incoherent noise or coherent-anticorrelated noise.

The reduction in performance exhibited by the majority of listeners indicates a strong confusion effect. Indeed, the inclusion of a wide range of correlations on a trial-to-trial basis caused more difficulty than the run-to-run confusion identified in Experiment 2.

Conclusions

While this study began as a simple test, intended to show that listeners could easily distinguish between the stimuli used for ILD, it produced some interesting results. Listeners did eventually prove that they could distinguish between coherent and anticorrelated noises. See Table IV.10 for a summary of relevant results. Only in three cases did a listener perform worse than the 75% threshold. Those that demonstrated difficulty in discriminating No/N π stimulus pairs (specifically in Experiment 2) also improved with training.

Table IV.10 - Summary of noise-type experiments							
Listener	H	X	J	C	R	L	Z
Exp. 2 No vs Nu	100.0	100.0	100.0	100.0	100.0	100.0	97.5
Exp. 2 N π vs Nu	100.0	100.0	99.2	100.0	100.0	100.0	95.0
Exp. 2 No vs N π	98.8	99.2	75.0	96.3	100.0	100.0	72.5
Exp. 3 No vs Nu	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Exp. 3 No vs N π	83.8	97.1	64.2	100.0	78.7	100.0	57.1

Percent correct in each experimental case by listener

However, several listeners showed evidence that reflected the importance of range and previous stimulus on coherence discrimination. The declines in accuracy from

Experiment 2 to 3 were significant ($t(6) = 2.471$, $p = 0.048$), a sign that binaural systems may be confused by the variable ranges of stimulus coherence. The context-coding mode of sensory memory (Durlach and Braida, 1969) could explain the “range effect”.

Context-coding refers to a listener’s ability to compare the aspects of a stimulus with the overall context of stimuli given in an experiment. The relevant example from the noise-type discrimination study is the comparison of coherence values, or “diffusiveness” in the head. Such a comparison would suffer from internal noise, caused by imprecise memory of the context. According to Durlach and Braida, “...the amount of noise depends on the width of the context, larger widths leading to greater noise.” Thus, the “range effect” leads to confusion on N_o/N_π stimulus pairs when they are mixed with N_o/N_u pairs.

The most significant question that needs answering is: what implication do these results have for the ILD experiment? The changing amount of internal noise shouldn’t impact coherence-related ILD thresholds, which involves only one stimulus type per run. The original goal was to show that the three stimuli are binaurally distinct due to the differing auditory source widths of their intracranial images, as asserted in the Methods section earlier in this chapter. The listeners bore that prediction out, for the most part.

References

- Békésy von, G. "Zur Theorie des Hörens: Über das Richtungshören bei einer Zeitdifferenz oder Lautstärkeungleichheit der beidseitigen Schalleinwirkungen," *Phys. Z.* **31**, 824-838, 857-868 (1930); cited by Blauert (1983) page 162.
- Dunn, O.J. "Multiple comparisons among means," *J. Amer. Stat. Assn.* **56**, 52-64 (1961).
- Durlach, N.I. and Braida, L.D. "Intensity Perception. I. Preliminary Theory of Intensity Resolution," *JASA* **46**, 372-383 (1969).
- Grantham, D.W. "Interaural Intensity Discrimination: Insensitivity at 1000 Hz," *J. Acoust. Soc. Am.* **75**, 1191-1194 (1984).
- Grantham, D.W. and Ahlstrom, J.B. "Interaural intensity discrimination of noise as a function of center frequency, duration, and interaural correlation," *J. Acoust. Soc. Am.* **71**, S86 (1982).
- Green, D.M. *Profile Analysis Auditory Intensity Discrimination*, Oxford Psychology Series No. 13, Oxford, New York, pg 20.
- Green, D.M. and Swets, J.A. *Signal Detection Theory and Psychophysics*, J. Wiley, New York.
- Hartmann, W.M. *Signals, Sound, and Sensation*, American Institute of Physics, New York, 1997. p. 384.
- Hartmann, W.M. and Constan, Z.C. "Interaural level differences and the level-meter model," *J. Acoust. Soc. Am.*, manuscript accepted (2002).
- Henning, G.B. "Detectability of interaural delay in high-frequency complex waveforms," *J. Acoust. Soc. Am.* **55**, 84-90 (1974).
- Jesteadt, W. and Weir, C. "Comparison of monaural and binaural discrimination of intensity and frequency." *J. Acoust. Soc. Am.* **61**, 1599-1603 (1977)
- Koehnke, J. and Durlach, N.I. "Range effects in the identification of lateral position." *JASA* **86**, 1176-1178 (1989)
- Koehnke, J., Colburn, H.S. and Durlach, N.I. "Performance in several binaural-interaction experiments," *J. Acoust. Soc. Am.* **79**, 1558-1562 (1986).
- Lindevald, I.M. and Benade, A.H. "Two-ear correlations in the statistical sound fields of rooms," *J. Acoust. Soc. Am.* **80**, 661-664 (1986).

Moore, Brian C. J. *An Introduction to the Psychology of Hearing*. Academic Press, New York, 1989. p. 196

Pickles, J.O. *An Introduction to the Physiology of Hearing*. Academic Press, New York, 1982. pp. 173-177

Pollack, I. and Trittipoe, W.J. "Binaural Listening and Interaural Noise Cross Correlation," *J. Acoust. Soc. Am.* **31**, 1250-1252 (1959).

Purks, S.R., Callahan, D.J., Braida, L.D. and Durlach, N.I. "Intensity Perception. X. Effect of Preceding Stimulus on Identification Performance." *JASA* **67**, 634-637 (1980)

Stern, R.M. and Trahiotis, C. "Models of Binaural Interaction." *Hearing*. ed. B. Moore. Academic Press, New York, 1995. pp. 347-386.

Yost, W.A. "Lateral position of sinusoids presented with interaural intensive and temporal differences," *J. Acoust. Soc. Am.* **70**, 397-409 (1981).

Figure IV.1. Depiction of a two-interval ILD trial in Experiment 1. The dashed lines represent a level of 60 dB for conditions (a) and (b), which are equally likely. The levels of left and right ears appear under “L” and “R” for each interval. Both (a) and (b) diagram a right-left trial according to ILDs, but (a) establishes the ILD by intensifying one ear above 60 dB while (b) attenuates the other.

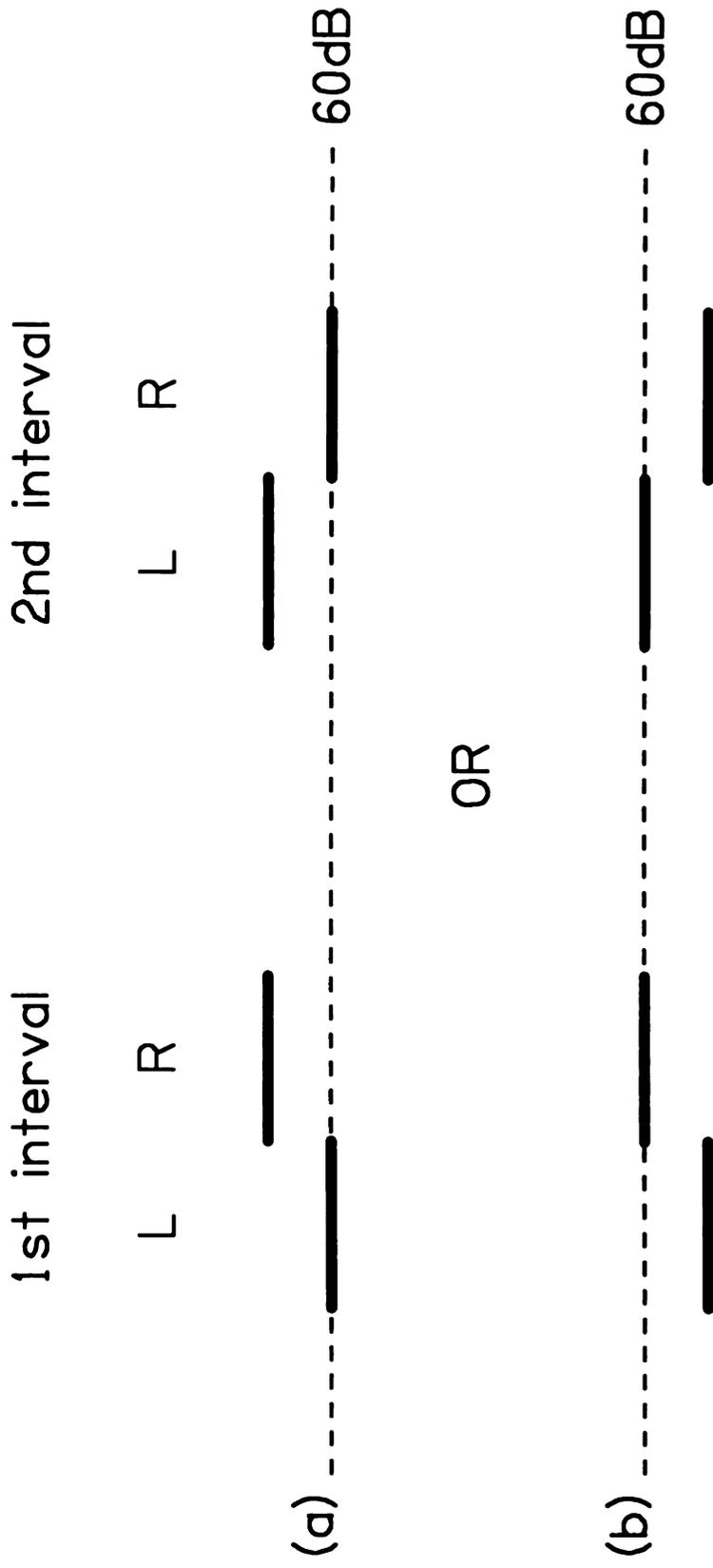


Figure IV.1

Figure IV.2. Depiction of a two-interval DLI trial in Experiment 1. The dashed lines represent a level of 60 dB for conditions (a) and (b), which are equally likely. The levels of left and right ears appear under “L” and “R” for each interval. Both (a) and (b) diagram a trial where the second interval is more intense, but (a) establishes a difference by intensifying the second interval above 60 dB while (b) attenuates the first interval.

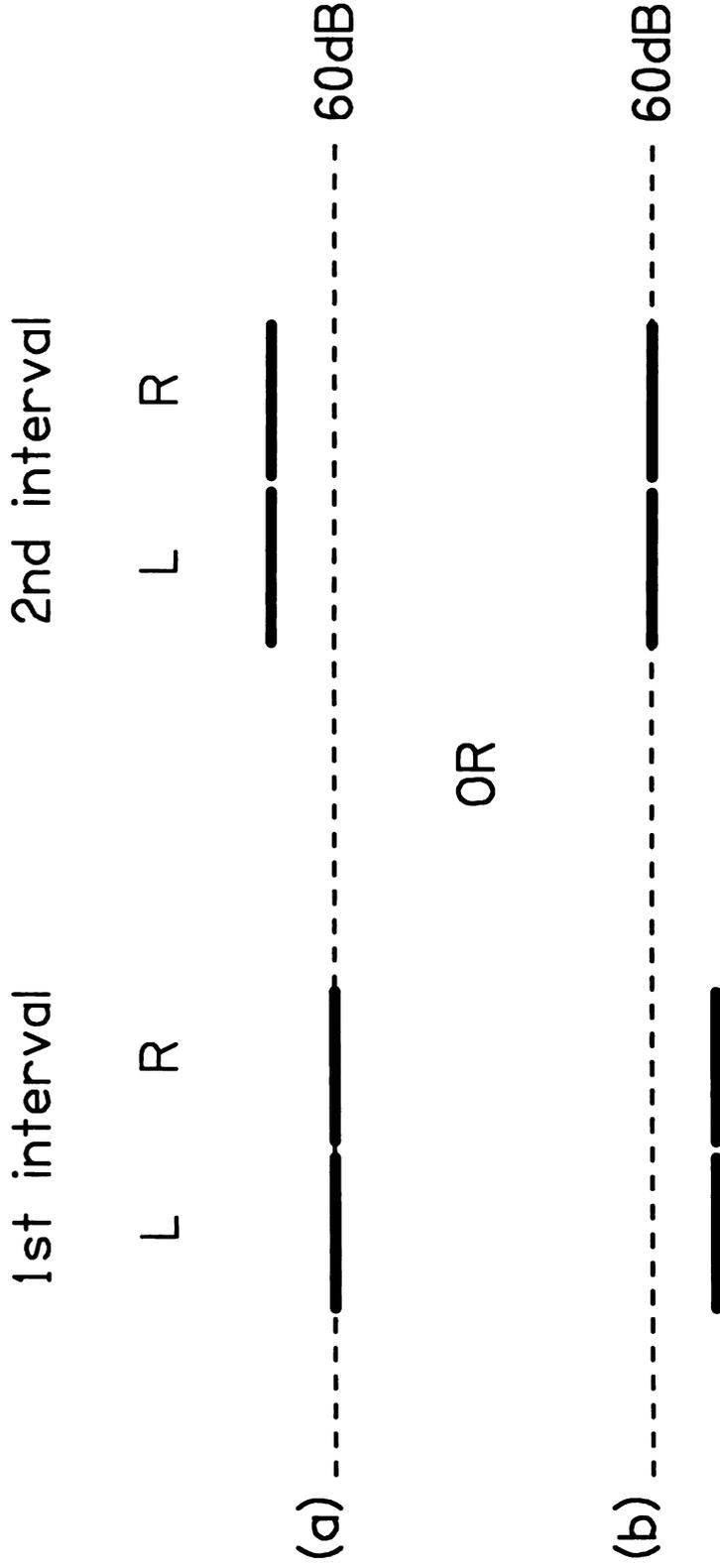


Figure IV.2

Figure IV.3. Results of Experiment 1: Broadband signal. The x-axis lists the six listeners that participated, and each features ILD discrimination thresholds (denoted along the y-axis) for broadband (20 Hz - 10 kHz) correlated, anticorrelated and uncorrelated noises, as well as DLI thresholds. The shaded bar below 0.15 dB represents an experimental limitation that invalidates results in that region. Error bars on data points represent two standard deviations.

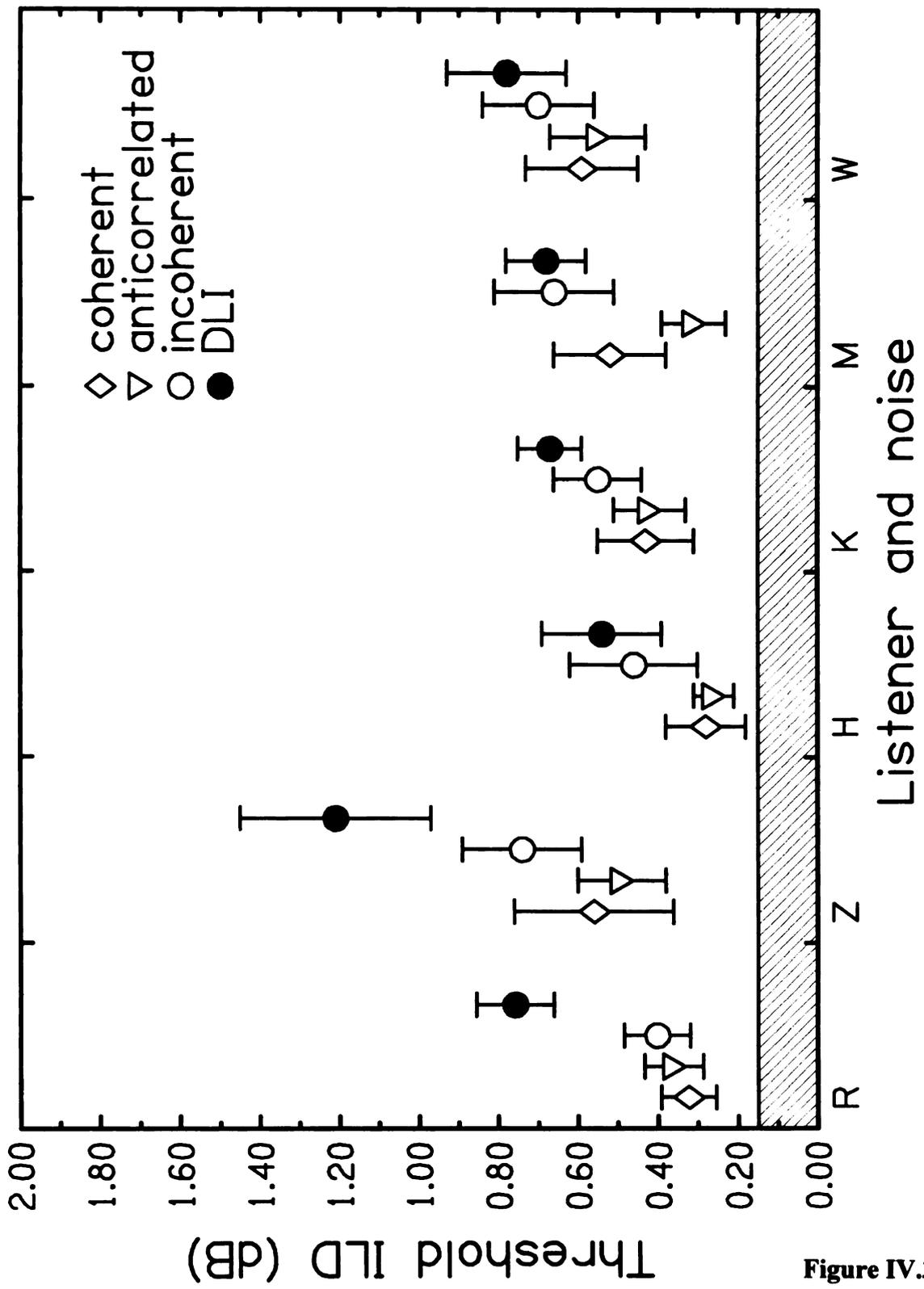


Figure IV.3

Figure IV.4. Results of Experiment 2: Low-pass filtered signal. The x-axis lists the six listeners that participated, and each features ILD discrimination thresholds (denoted along the y-axis) for narrowband (20 – 1000 Hz) correlated, anticorrelated and uncorrelated noises, as well as DLI thresholds. The shaded bar below 0.15 dB represents an experimental limitation that invalidates results in that region. Error bars on data points represent two standard deviations.

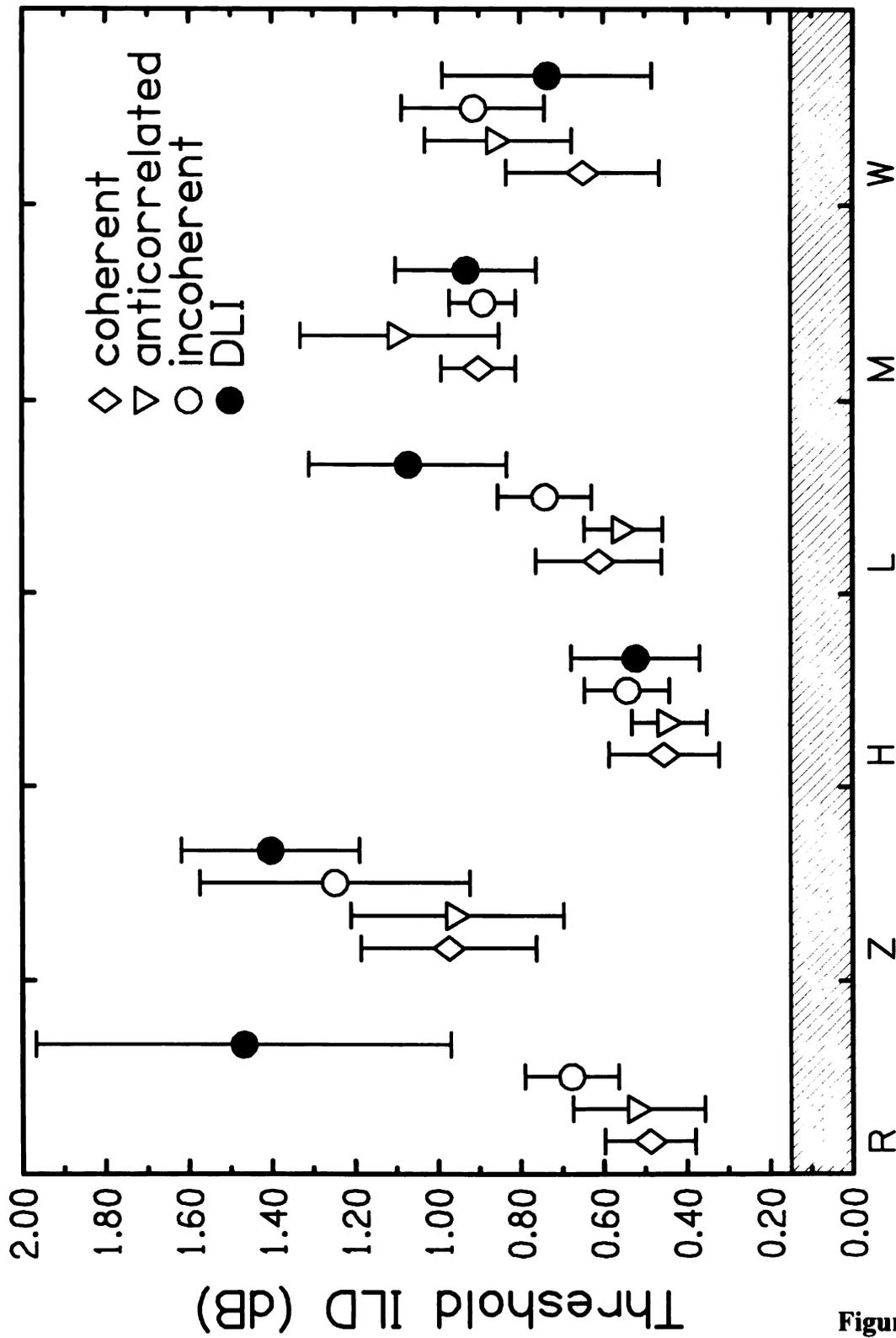


Figure IV.4

Figure IV.5. Depiction of a two-interval ILD trial in Experiment 3. The dashed lines represent a level of 60 dB for conditions (a) and (b), which are equally likely. The levels of left and right ears appear under “L” and “R” for each interval. Both (a) and (b) diagram a right-left trial according to ILDs, but (a) establishes the ILD by intensifying one ear above standard level while (b) attenuates the other. Intervals feature a randomized standard level, which lies within 5 dB of 60. This figure is identical to Figure 1 aside from a 62-dB standard for the first interval and a 59-dB standard for the second.

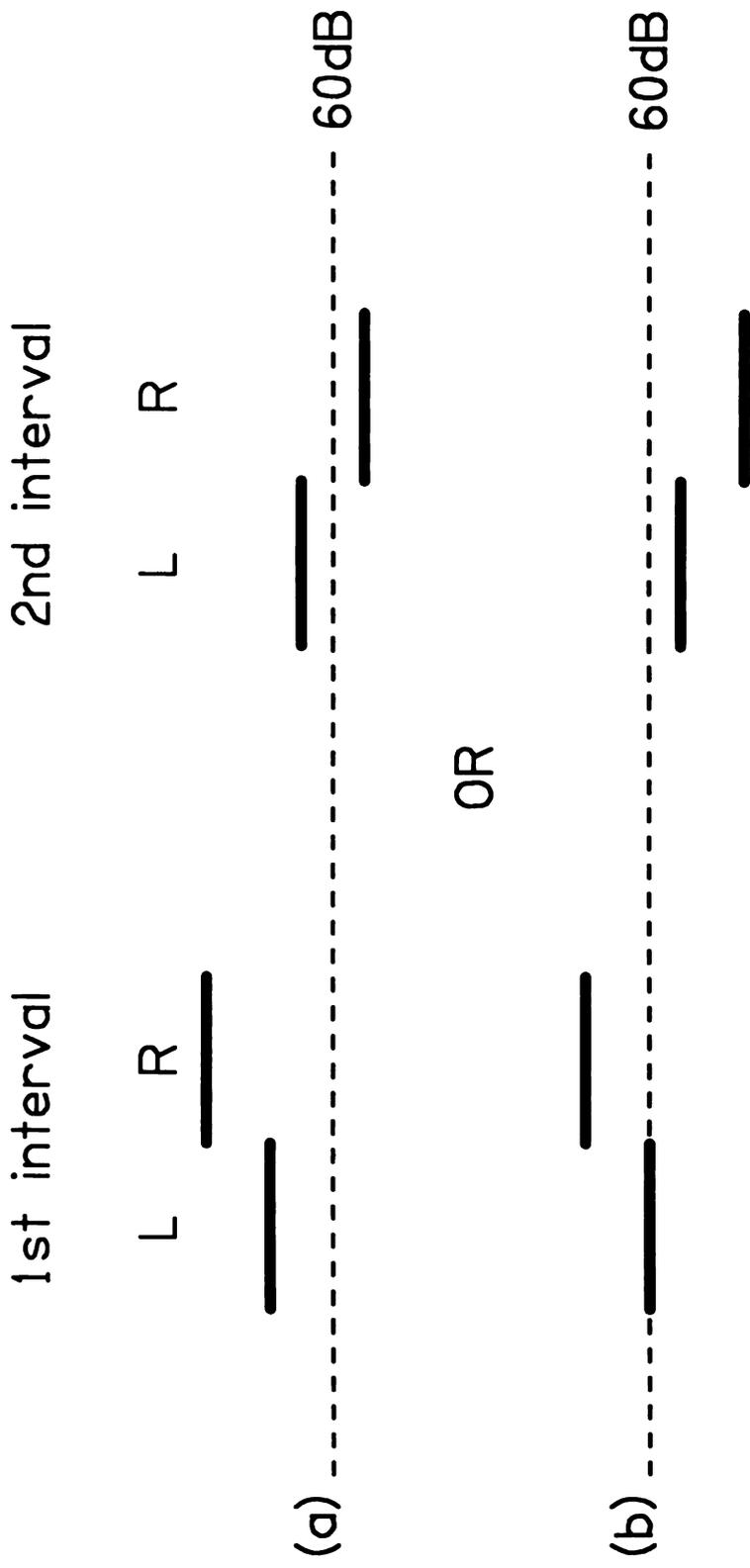
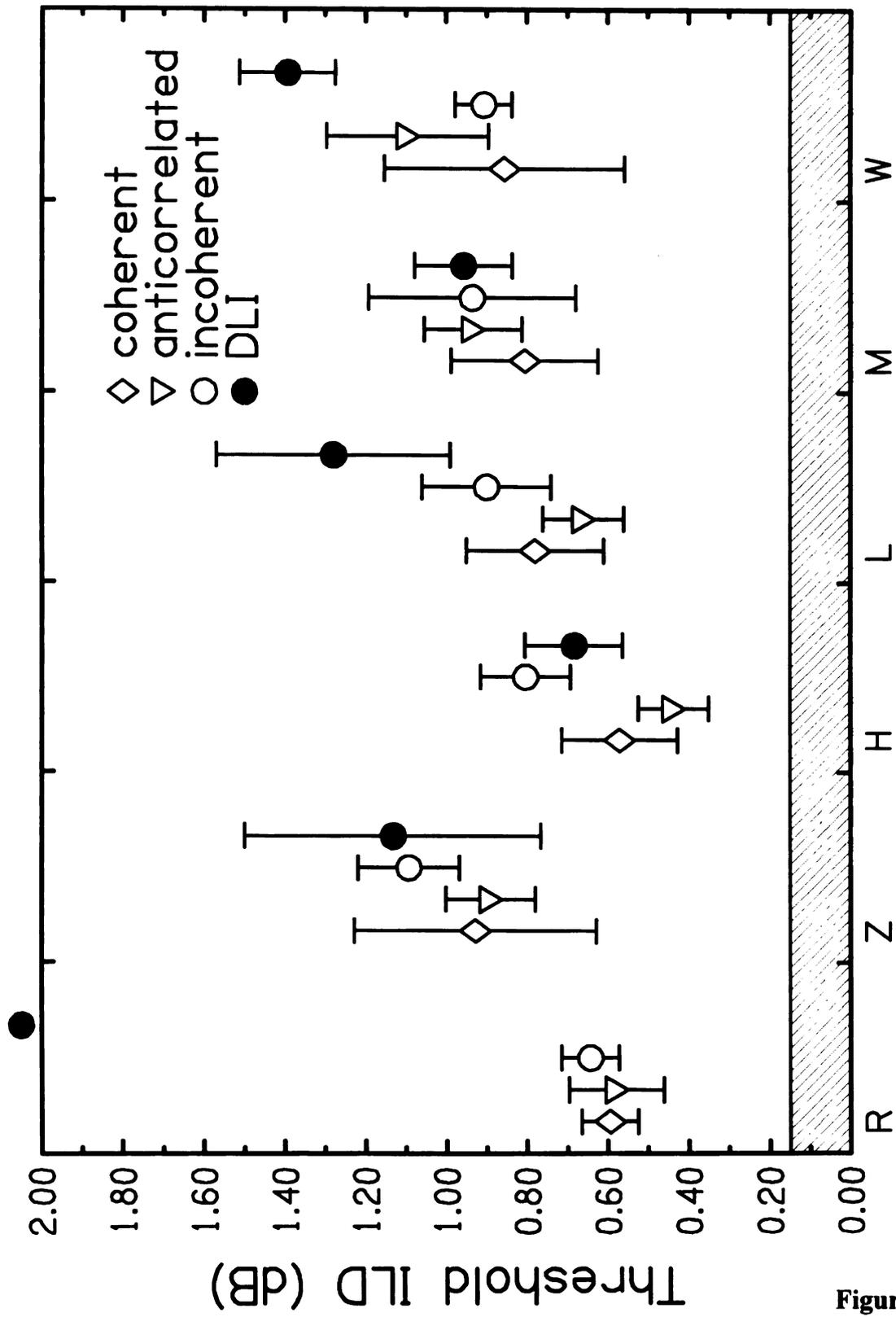


Figure IV.5

Figure IV.6. Results of Experiment 3: Low-pass filtered signal with randomized standard level. The x-axis lists the six listeners that participated, and each features ILD discrimination thresholds (denoted along the y-axis) for narrowband (20 – 1000 Hz) correlated, anticorrelated and uncorrelated noises, as well as DLI thresholds. The shaded bar below 0.15 dB represents an experimental limitation that invalidates results in that region. Error bars on data points represent two standard deviations.



Listener and noise

Figure IV.6

Figure IV.7. ILD discrimination according to signal detection theory (TSD). All five plots show detection as a probability density curve. The leftmost two represent the intensity measurements of a single interval in left and right ears. The signal is more intense in the right ear, and thus μ_R (the most probable measured intensity) is greater along the intensity axis than μ_L . The standard deviations of both curves (marked with dashed horizontal lines and σ symbols) are assumed to be equal. The top-center plot results from the difference between the leftmost curves, leading to a most probable ILD Δ_1 for the first interval. The bottom-center plot is the result of the second interval, with ILD reversed. The rightmost plot shows the probability of judging the ILD shift from right to left (the sum of all probabilities to the right of zero).

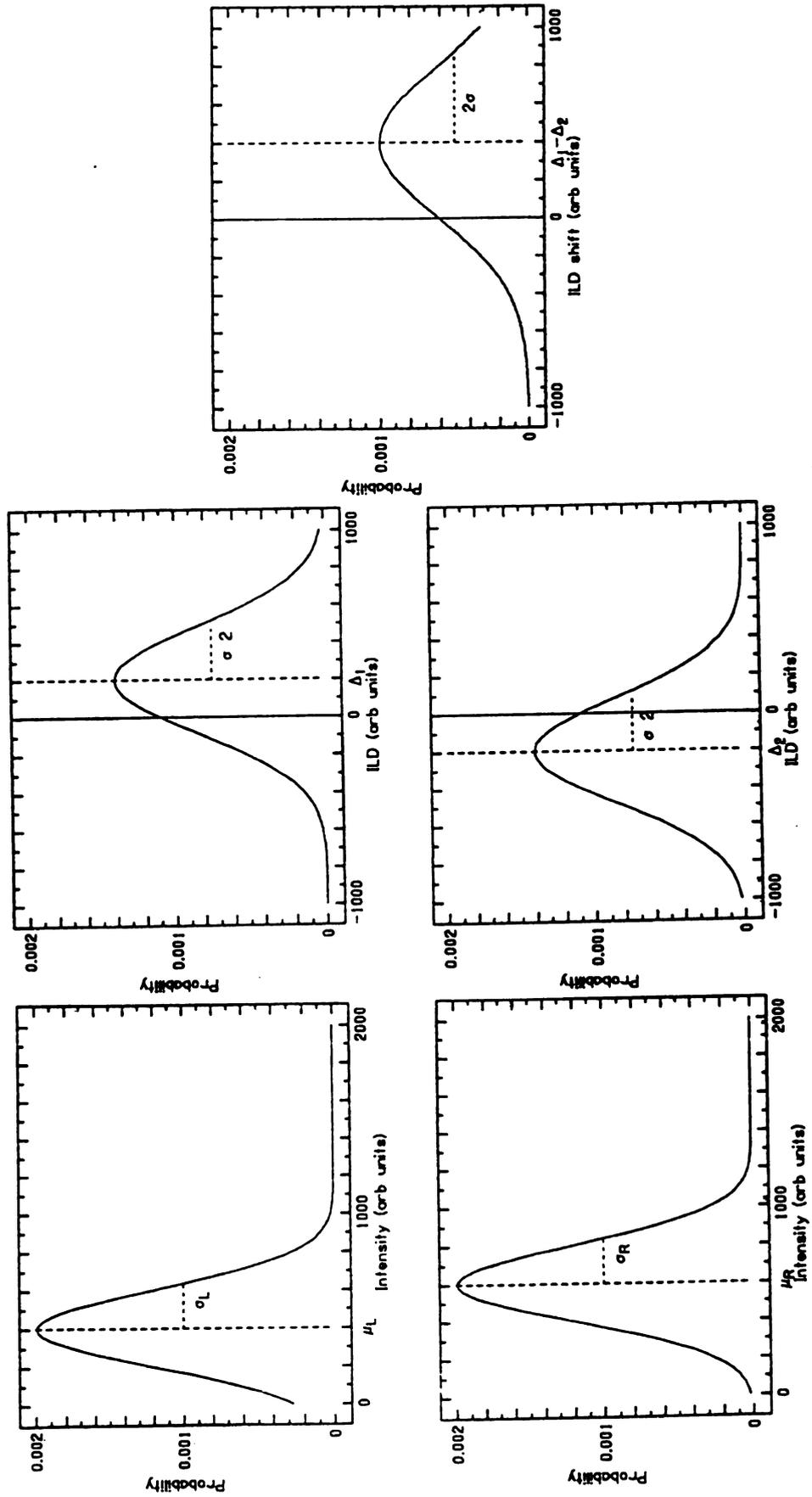


Figure IV.7

Figure IV.8. Loudness-meter model predictions compared to results from Exps. 2 and 3. The histograms show average threshold ILDs for all four stimulus types in Experiments 2 and 3. The dotted, solid and dashed lines represent predicted thresholds for integration times of 200, 300 and 400 ms, respectively. For the correlated condition, all three times result in the same prediction.

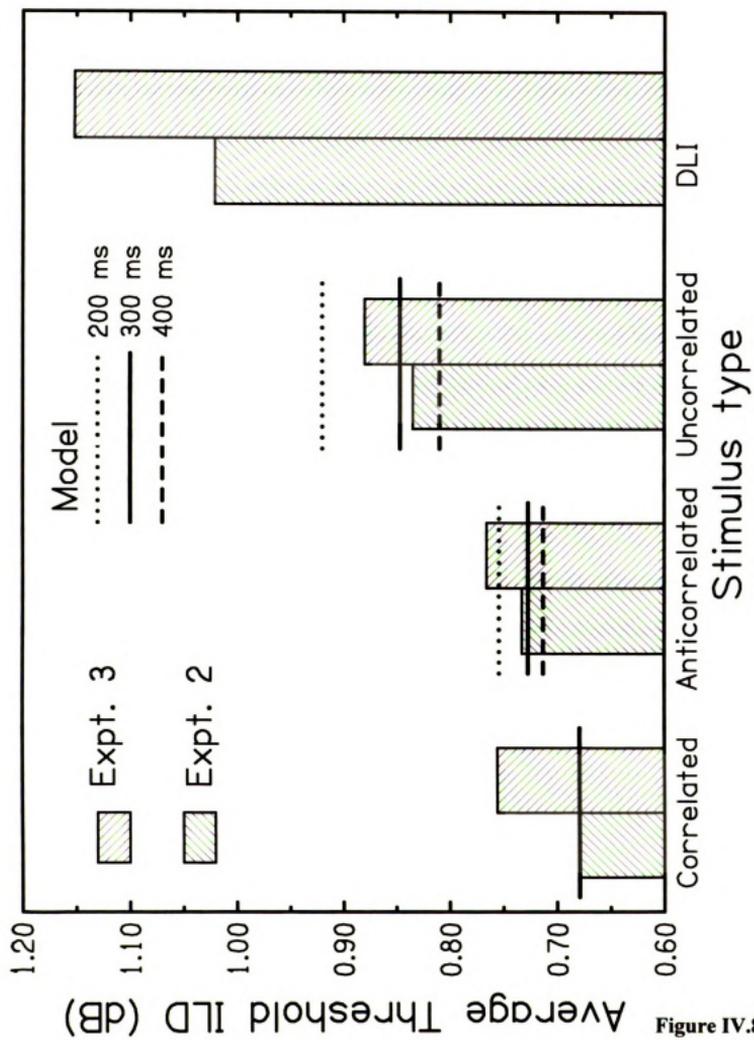
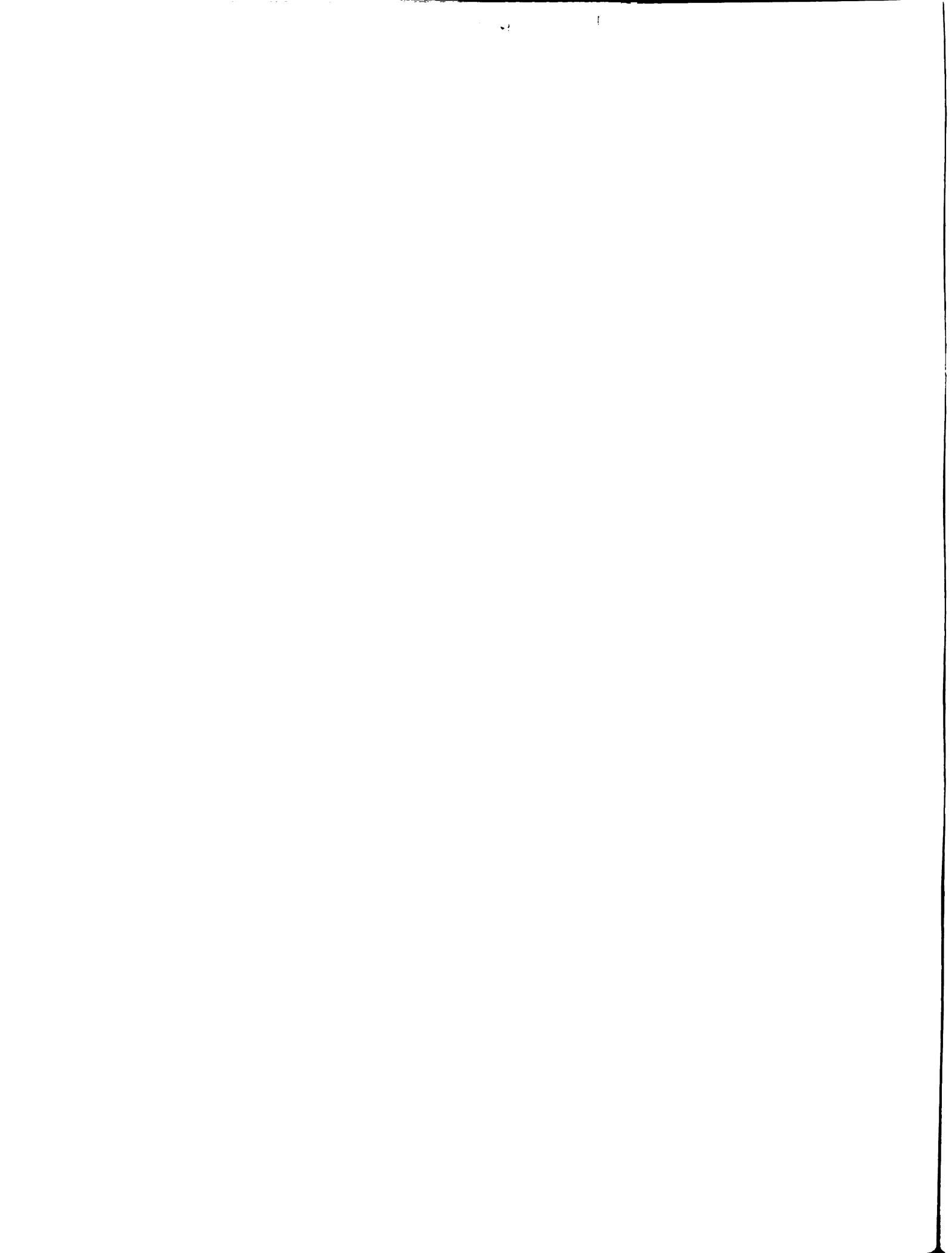


Figure IV.8



Introduction

Consider the real-world case of a sound arriving at a listener's ears. Refer to Figure V.1, borrowed from Kuhn (1987), for a simple diagram of this situation. This is a simplified model of the head, with ears placed at two points opposite each other on a sphere. Looking down on the head and an incident plane wave, it is clear that unless the listener is facing directly toward the source, one ear will always be caught in the "head shadow". The shadowed ear is not cut off from the signal, however; the diffraction around the head, as well as reflections from the shoulders and torso, provide an indirect passage for sound. In the interest of simplicity for all further analyses, this discussion will focus on the diffracted sound with the system including just a spherical head, and study the simplified setup of Figure V.1 in great detail.

The resulting stimulus at the shadowed ear has a particular set of phase and amplitude characteristics due to diffraction, which has been treated theoretically in the following equation for pressure induced by a plane wave on a spherical surface found in Kuhn (1977). It becomes useful when approximating the human head as a sphere with the ears at two opposing points (as in Figure V.1):

$$\left(\frac{P_i + P_s}{P_o} \right) \Big|_{r=a} \cong \left(\frac{1}{ka} \right)^2 \left[\sum_{n=0}^{n \max} \frac{i^{n+1} (2n+1) P_n(\cos \theta)}{j'_n(ka) - iy'_n(ka)} \right] \quad (1)$$

The left term is the sum of incident and scattered pressures, normalized by the free-field pressure. Argument k is the wave number for the relevant frequency (equal to 2π times the frequency divided by the speed of sound), a is the radius of the head (approximately 8 cm), n is the order to which this sum is calculated (for purposes of this analysis, $n_{max}=12$ was sufficient), and θ is the azimuthal angle of a point on the head (measured from the direction of the incident plane wave). Referring again to Figure V.1, angle θ is defined as $90 - \theta_{inc}$ for the right ear, and $90 + \theta_{inc}$ for the left. Functions P_n, j'_n and y'_n are the n th-order Legendre polynomials and the first derivative of the spherical Bessel and spherical Neumann functions, respectively.

When evaluated at the left and right ears, this equation leads to theoretical interaural phase shifts due to diffraction around the head, which, when divided by their respective frequency, yield Interaural Time Differences (ITDs). Those calculations are summarized in Figure V.2, which plots ITD versus frequency for incident angles of 10, 20, 30, 45, 60 and 90 degrees. The open circles on the left represent the low-frequency ITD limit for each angle. This limit is determined by reworking equation (1) with the assumption that $ka < 1$ (low frequencies have small k) (Kuhn, 1977):

$$\left(\frac{p_i + p_s}{p_o} \right) \Big|_{r=a} \cong 1 \pm i \left(\frac{3}{2} ka \right) \sin \theta_{inc} \quad (+ \text{ for near ear, } - \text{ for far ear}) \quad (2)$$

Since this value is complex, the pressure's phase at each ear can be found by simply taking the arctangent of real divided by imaginary parts. Thus, the ITD in the

low-frequency limit can be approximated by the phase difference between near and far (shadowed) ears:

$$ITD \cong \frac{3a}{c} \sin \theta_{inc} \quad (3)$$

where c is the speed of sound in air (343 m/s).

Proceeding with other features of Figure V.2, the dashed line delineates phase shifts of half a wavelength for each frequency, at which point it is impossible to establish which ear is leading or lagging. The black squares on the right represent a theoretical high-frequency limit, which has an approximate value given by:

$$ITD \cong \frac{2a}{c} \sin \theta_{inc} \quad (4)$$

Thus, it is clear that the high-frequency ITD is simply 2/3 that of the low-frequency ITD, and the squares are positioned according to that rule. The open triangles to the extreme right are predictions of the Woodworth formula

$$ITD \cong \frac{a}{c} [\theta + \sin \theta] \quad (5)$$

another approximation that ignores reflection, but accounts for opacity of the head. It is considered to be a close approximation for the ITD of high-frequency signals (Kuhn, 1977). The plots for various angles terminate at several different frequencies, with low

angles including data beyond 5 kHz and high angles only extending to 1.2 kHz. This is due to computational limitations.

The current study

With data from Kuhn's spherical-surface pressure equation, one can generate noises for left and right ears that demonstrate the correct distribution of ITDs (according to our calculation), and thus simulate dispersion of sound about the head with headphones. If the head-dispersed noise is treated as a slightly incoherent noise, it can be used in a coherence discrimination experiment (much like in Chapter II) that requires listeners to compare two intervals with the task of distinguishing the head-shifted stimuli from perfectly coherent noise (with uniform ITD over all frequencies). Theoretically, the listener could find it difficult to identify which was more incoherent, since he or she would be familiar with the particular incoherence introduced by the head.

The proposed experiment is much like Kulkarni's sensitivity experiment involving head-related transfer functions (HRTFs) (Kulkarni *et al.*, 1999). In that paper, listeners were asked to distinguish empirically-measured HRTFs from those that were reconstructed according to a model. The experiments reported here asked listeners to compare "head-shifted" stimuli modeled on Kuhn's theoretical sphere calculations with a constant-ITD coherent noise. Frequency-independent amplitudes are matched for both intervals. Note that Equation 1 can also derive Interaural Level Differences (ILDs) in addition to ITDs. However, the following experiments focused attention on the phase dispersion by eliminating all other cues, including level differences. Previous studies have also found that these head-related ILDs have little if any influence on the perceived

accuracy of an headphone-reproduced HRTF (Kulkarni *et al.*, 1999; Kistler and Wightman, 1992). Also, the model calculations allow isolation of head dispersion from pinnae cues and reflections off the shoulders and torso, which would necessarily be included in any measured HRTF. Thus, the discrimination task is based solely on the difference in phase shifts. The constant-ITD (frequency-independent delay) stimulus is the representation of perfect coherence under lab conditions, using headphones. Such a stimulus approximates what the auditory system would detect if the listener's head were transparent to sound. This chapter is intended to answer the question: are listeners sensitive to the incoherence introduced by head dispersion?

Experiment 1: Discrimination at two incident angles

Method

Experiment 1 featured two incident angles for the head-dispersed noise: 30 and 45 degrees. Both have an interesting ITD characteristic (refer to Figure V.2). Among the various angles presented, the 30-degree plot exhibits the steepest change in delays according to frequency (in the region of 800 Hz), while the 45-degree calculations exhibit the most extreme ITD shift within its range (from 563 μ s to 411 μ s). These two stimuli should be the most distinguishable from uniform ITD, and thus have the best chance of identification by listeners.

The experiment tested listeners at three different bandwidths, identified as broadband, narrowband, and narrow-narrowband. The purpose of each was to provide the listener with a different perspective on the head-shifted interval, and establish if one

of the bandwidths made it easier to identify. Since the head-shifted noise was produced digitally, the band-pass filtering was simple to achieve and absolute. The broadband (BB) case included all head-shifts as calculated by our spherical-head program, for a range of frequencies from 20 Hz to 3 kHz for both angles. Note that the 45-degree plot on Figure V.2 only includes data up to 2 kHz--this was remedied by assuming that the behavior after the last known ITD value was asymptotic, and filling the region from 2-3 kHz with that value. These upper limitations were likely of no consequence, since consensus in the literature agrees that ITDs above 1.5-2 kHz have little effect on the perceived image (Hartmann, 1996; Kulkarni *et al.*, 1999; Wightman and Kistler, 1992; Brungart *et al.*, 1999). Thus, the experimenters felt confident in performing this experiment without head-shift data above 3 kHz. Narrowband (NB) limits were defined by the region where ITDs went from a maximum to a minimum value, presenting listeners with only the band of greatest change. For the 30-degree data, that band was 400 Hz to 1200 Hz, while the 45-degree equivalent was 400-1800 Hz. Finally, narrow-narrowband (NNB) stimuli were limited to the steepest part of the ITD change, the region of greatest slope. These bands were 800-1000 Hz for 30-degree data, and 600-800 Hz for 45-degree data.

To check that the head-dispersed interval did indeed contain the appropriate ITD characteristics, as calculated according to frequency, it was tested with a digital delay line and an HP spectrum analyzer, both of which were independent of the Tucker-Davis equipment used to generate the stimulus. The input was generated by delaying the left-ear signal by $X \mu\text{s}$, inverting it and adding on the right-ear signal. The resultant spectrum, which was of equal amplitude, exhibited sharp dips at frequencies where the

head-shift ITD was $X \mu\text{s}$ in the right ear. This allowed the identification of several expected delays, and verified that the head-shifted stimulus was produced correctly.

To present a rigorous test of the perceptual similarity (or dissimilarity) of head-shifted vs. coherent stimuli, one must eliminate all non-coherence cues that could interfere with the planned task. The bandwidth and amplitude characteristics of both stimuli were matched. Also, one would naturally choose to eliminate position cues from the task by selecting an ITD for the coherent interval that placed it in an equivalent lateral position with the head-shifted interval. Experiment 1 actually used a range of ITDs, so that on each trial, the coherent noise had a random delay chosen from a set of 10 experimenter-defined values. Thus, the coherent noise would not be easy to identify simply because the listener recognized its lateral location after a few trials. For example, one head-dispersed interval was generated using the 30-degree data from 20-3000 Hz (the upper limit dictated by calculation limitations, as noted above), with equal amplitudes across the band. The experimental program provided a coherent-noise stimulus of identical bandwidth and amplitude, with an ITD selected from the set of 260, 280, 300, 320, 340, 360, 380, 400, 420, and 440 microseconds. Note that 260, 420 and 440 μs are all outside the array of ITDs included in the 30-degree plot. These conditions were included as a test, to establish how well the listeners could use a lateral cue.

Both constant-ITD and head-shifted intervals were generated digitally by a Tucker-Davis Technologies Array Processor. The processor constructed both in frequency space, setting the upper and lower band limits and filling between them with equal-amplitude random-phase noise. For a head-shifted interval, the processor would then introduce the pre-calculated interaural phases. For a constant-ITD interval, it

imposed one of 10 evenly-spaced ITD values, ranging from just below to just above the array of head-shift values for the corresponding incident angle and bandwidth. Below are the ITDs selected for the six different conditions:

30-degree, BB/NB: 260, 280, 300, 320, 340, 360, 380, 400, 420, 440 μ s

30-degree, NNB: 290, 298, 306, 314, 322, 330, 338, 346, 354, 362 μ s

45-degree, BB/NB: 400, 420, 440, 460, 480, 500, 520, 540, 560, 580 μ s

45-degree, NNB: 505, 510, 515, 520, 525, 530, 535, 540, 545, 550 μ s

The four listeners were each identified by a single letter.

- T, male age 24, normal hearing, no previous listening experience
- W, male age 60, common middle-age high-frequency hearing loss, extensive listening experience
- X, male age 26, normal hearing, previous experience in listening studies
- Z, male age 27, normal hearing, previous experience in listening studies

During runs, listeners sat in a double-walled Acoustic Systems soundroom listening with Sennheiser HD-480 II headphones. The experimental runs consisted of 100 two-interval forced-choice (2IFC) trials (ten for each coherent ITD value). On each trial, the program would present the listener with two 500-ms intervals in random order. One was a constant-ITD stimulus, and the other included head-shifts. Both intervals had rise/fall times of approximately 30 ms, and had simultaneous onsets/offsets in both ears. The lack of onset cues made it necessary for listeners to identify dispersion in the steady-

state sound, if possible. Also, this decision was justified by the work of Tobias and Schubert (1959), who found that onset ITDs are superseded by steady-state ITDs when localizing noises that exceed 125 ms in duration.

After presenting both intervals, the program prompted for a response, and the listener pushed one of two buttons to answer which interval was head-shifted. Listeners were instructed to identify the head-shifted stimulus by its incoherence, probably using the Auditory Source Width (ASW) of its image as was done in Chapter II's experiments. The ASW refers to diffuseness, or "broadness", of the intracranial image. Coherent stimuli are very compact, producing a small ASW, while incoherent stimuli result in a large ASW. Following several training runs, during which the program provided feedback indicating the correct answer after every trial, listeners participated in four runs at each combination of incident angle (30 and 45 degrees) and bandwidth (BB, NB, NNB).

Results

The experimental results are found in Figures V.3-10. The data are averaged over runs, showing percent of trials answered correctly (counted as selecting the head-shift interval) versus the ten ITDs applied to the coherent-noise stimulus. The bars on each point represent two standard errors of the mean. Standard error is defined as the standard deviation normalized by the square root of the number of data points minus one (in this case, we have four data points). Each graph includes three plots, one for each bandwidth condition. The horizontal bars labeled "BB/NB" and "NNB" range delineate the set of ITDs included in head-shifted stimuli of those bandwidths. The dashed vertical line

corresponds to the low-frequency ITD limit for sound dispersed about the head, which was represented by the open circles in Figure V.2. The dashed horizontal line indicates the 50% correct level (guessing), which is the performance expected if listeners could not distinguish between constant-ITD and head-shifted intervals. Figures V.3-6 are 30-degree results for listeners T, W, X, and Z, respectively. Figures V.7-10 are 45-degree results for listeners in the same order.

Taking all these complex data into account, there are two features of note. Listeners W and Z could indeed distinguish head-shifted intervals from constant-ITD, achieving very high success rates for ITDs that were close to or beyond the extremes for that particular head-shifted stimulus. However, when the constant-ITD fell nearer the center of the ITD range, performance reduced to guessing. Listeners T and X were not as successful at the task, and their results hovered near 50 percent. They did meet or exceed 75 percent for a few ITDs. X did so when the ITDs were 400 μ s in the 30-degree NB condition, 400 μ s in the 45-degree NB condition, and 400-440 μ s in the 45-degree BB condition. T did so for ITDs of 260, 300 and 320 μ s in the 30-degree BB condition, as well as 260, 420 and 440 μ s in the 30-degree NB condition. Note that all these ITDs are near the limits for their corresponding head-shifted stimuli.

Discussion

The results of Experiment 1 suggest the probability of a lateral cue. That is, some listeners may not have been performing the task using coherence discrimination, but rather learned to associate the head-shifted interval with a well-defined and absolute lateral location. The roving coherent-interval ITD only prevented the use of cues based

on relative lateralization. The range of ITDs was selected to include delays found within the head-shifted stimulus, with two or three lying outside that range. However, it seems that listeners W and Z tended to identify the head-shift image with a particular delay. If the ITD interval image was close to that delay, W and Z could do no better than guessing. However, when the ITD interval was significantly far (20-40 μ s) from that location, their rate of correct responses increased dramatically, reaching 100% once the delay had shifted 100 μ s or more (typically outside the range of delays found in the head-shift). Interviews with the listeners confirm that while they had been asked to discriminate based on “width” of the image, their training with the feedback led to lateral discrimination. Listeners X and T reported that they also concentrated on the lateral cue, but as their results show, they were not as successful at discriminating the image shifts. It is possible that they were not as astute as W and Z at discriminating lateral movement. They did, however, manage to distinguish the head-shifted interval from some constant ITDs, specifically those that were near the outer edges of the head-shift range.

Experiment 1 did show that the auditory system has difficulty discriminating head-related dispersion from a constant ITD. However, the experiment also yielded different information than originally expected, because some listeners were able to follow an alternative cue to dispersion: image movement. Listeners W and Z demonstrated an inability to distinguish a head-shifted image from a constant-ITD band when the two were at the same lateral position. If the listeners had identified an additional cue, such as the image’s ASW, then the loss of lateral position might have reduced their hit rate to 80%. Because performance reduced to guessing, Experiment 1 provides strong evidence of listeners’ insensitivity to head dispersion. Another interesting result was the apparent

lateral location of the head-shifted stimuli. Note that they tended to be well-removed from the low-frequency limit, the delay that characterizes the stimulus without dispersion. While the listeners were unable to successfully distinguish the head-shifted intervals, they did see some effect from the pattern of dispersed phase delays.

Experiment 2: Randomized angle

Introduction

The above results showed that listeners could not identify the head-shifted interval when its lateral position was near that of the constant-ITD interval. Unfortunately, that coincidence of location only occurred in 10-20% of trials, and listeners may have missed a subtle but useful cue while they were concentrating on the lateralization cue. Thus, the experimenters decided to design a new experiment that would truly eliminate the usefulness of lateral position, leaving listeners with only interaural correlation as a means for discrimination.

Method

The follow-up experiment used six different stimuli: a set of three head-shifted bands associated with various angles of incidence, and three corresponding constant-ITD bands. Each trial would select the two intervals randomly, one from each group so there would be a comparison between head-shifted and ITD stimuli. In this manner, the lateral positions of the two would be randomized, and provide no consistent cues for the listener. Aside from the roaming image positions, Experiment 2 was identical to Experiment 1,

with the same task: identify the head-shifted interval, possibly by detecting the ASW of the image due to incoherence.

The experimental design includes three angles for head-shifted stimuli based on their potential for success, and previous experience in Experiment 1. On Figure V.2, it is clear that 45-degree data exhibit the largest overall shift in ITD, producing the most decorrelation and the best chance for our listeners to distinguish from constant-ITD. Also, previous experimenters have found that Kuhn's model best approximates real HRTFs for a 45-degree angle of incidence (Wightman and Kistler, 1986). According to results from Experiment 1, when listeners attempted to identify the 45-degree head-shift stimulus, the ITD of greatest confusion (where they were reduced to guessing; see Figures V.7-10) was 520 μ s. This particular delay occurred in the head-shift data at a frequency of 727 Hz. It was decided, due to prior study in lateral discrimination and the steep changes in accuracy as ITD varied in Experiment 1, that 30 μ s was a detectable shift in lateralization. Thus, to establish two other angles of incidence that were distinct from 45 degrees, the spherical-head program was used to see which angles had an ITD of 490 and 550 μ s at 727 Hz, effectively bracketing the 45-degree image. Calculations showed that 42 degrees has 490 μ s delay at 727 Hz, and 49 degrees is delayed 550 μ s at 727 Hz. Thus, the angles of incidence selected were 42, 45, and 49 degrees. The BB, NB and NNB limits were the same as for the 45-degree case of Experiment 1.

Preliminary Experiment: "Equivalence of position"

To ensure the constant-ITD stimuli matched up perceptually with the three head angles, preventing any further identification of absolute position, listeners performed the

2IFC task with a slight change to the decision-making process. Instead of identifying the most incoherent interval, the subjects were asked to listen for lateral movement, and respond according to the direction that the image seemed to move between intervals (what was termed the left-right task). Assuming that the three head-shift images matched up laterally with the corresponding ITDs, the percent of responses that listeners judge the ITD interval to be left of the head-shift interval would be as shown in Table V.1 below. The three columns include the three possible constant-ITD intervals, while the three rows are the three possible head-shifted intervals. Each cell then represents a trial formed by one combination of intervals.

Table V.1 - Ideal lateralization results			
	ITD 1	ITD 2	ITD 3
Head-shift angle 1	50%	~25%	0%
Head-shift angle 2	~75%	50%	~25%
Head-shift angle 3	100%	~75%	50%

To interpret Table V.1, start with the diagonal. Since the head-shift and ITD images should be at the same lateral position, then on a trial where both seem to be lateralized at the same place, one would be equally likely to hear the ITD image to the left or right of the head-shifted image. Thus, it would sound further left about half the time. At the lower left corner is the result for all trials that include the leftmost ITD image and rightmost head-shift image. This combination should provide a clear lateral motion. The ITD lateral position should clearly be left of the head-shift image, and listeners should judge that way 100 percent of trials. Moving on to the table's upper right corner, it is composed of the rightmost ITD image and leftmost head-shift image. It is clear that, due to symmetries, the listener should judge the ITD interval to be further right for every one of these trials. The other cells of the table must necessarily have some

percentage that interpolates the predictions for the corners and diagonal. If these predictions hold, and the experimental stimuli are sufficiently randomized, then the overall results should amount to guessing (50%).

With this theory in mind, the experiment calibrated for each listener by running the above task. Calibration runs continued, with adjustments to the various ITD values, until the experimenter recorded two runs with results corresponding somewhat to the expected results (granted that lateral cues were eliminated) laid out in Table V.1.

Listeners T, X, W, and Z required various ITDs to achieve the goal of comparable lateralization, and the lateralization-task percentages for the resulting ITD values can be found in Table V.2. This table actually comprises twelve tables, arranged in four rows (according to listener) and three columns, one for each bandwidth condition: broadband (BB, from 20-3000 Hz), narrowband (NB, from 400-1800 Hz), and narrow-narrowband (NNB, from 600-800 Hz). Each table is labeled “lat %” because it contains the percentage of “ITD interval to the left” responses (hereafter called “ITD-left judgments”) for all combinations of trial stimuli. The individual tables have this format:

lat %	480	510	540	
42	50%	25%	0%	25%
45	75%	50%	25%	50%
49	100%	75%	50%	75%
	75%	50%	25%	

Trials consisted of a head-shifted band selected from the left column and a coherent band with one of the three ITD-shifts in the first row (in μ s). The numbers 42, 45, and 49 are the incidence angle in degrees represented by the head-shift stimulus. The ITD stimulus was presented randomly on the first or second interval. The listener’s percent of ITD-left responses for each combination of angle and ITD indicated on the

Table V.2. Experiment 2 lateralization data. This table comprises twelve tables, arranged in three columns (representing three bandwidth conditions) and four rows (one for each listener). The lateralization data are used as a tool to calibrate the ITDs for perfectly-coherent intervals (the three column headers in each table) so the images match lateral position with the three head-shifted stimuli (angles in degrees listed as the three row headers in each table). The overall performance at each ITD and angle is listed at the end of each column and row, respectively. The overall performance of a listener at one condition is listed below the individual tables.

table. The final row and column are the percent ITD-left judgments for trials totaled over a particular ITD or angle, respectively. They are not averages of the percentages found in this table, because each condition did not necessarily contain the same number of trials. The final result found under each table is the “Overall” ITD-left judgments. This value comes from the total number of ITD-left responses (in all nine conditions) divided by 198 trials.

All final calibrations were the best approximation established by the experimenter, though one might note that some are better than others. Listener Z meets the lateralization criterion quite neatly across BB, NB and NNB conditions. Listener W does well also, though his NB results are slightly different from the ideal. Listeners T and X have the most trouble with the left-right task. While their results show a general trend towards the theoretical goal, they seem to find less difference between the three head-shift images than the other two listeners, especially when limited to NNB frequencies. Note that T’s results aren’t any closer to the ideal of Table V.1 than X’s. In order to achieve the listed data, it was necessary for the experimenter to increase the differences between constant-ITDs to 35 μ s in the BB case, and 45 μ s for the other two bandwidths. Apparently, listener T was not as sensitive to ITD shift as listener X, much less W and Z. It is difficult to predict what effect T’s and X’s lower ITD sensitivity would have on the coherence-discrimination results, since they are not necessarily tied to lateral position. The goal was simply to ensure that the listeners couldn’t use lateral position as a cue, and it seems unlikely that T or X would be able to do so.

There is some additional support for the calibrated delays, provided by Experiment 1. Returning to Figures V.7-10, recall the constant-ITD delays that caused

the most confusion (reduced performance to guessing) with 45-degree head-shifted stimuli. Listeners did worst at approximately 520 μ s in the BB case, 510 μ s in the NB case, and 540 μ s in the NNB case. These delays coincide well with the calibrated values found in Experiment 2. One can be confident that the lateralization task has established dependable “effective ITDs” for the head-shifted stimuli.

Experimental task, Results and Discussion

Once the appropriate ITDs were established, listeners T, X, W, and Z participated in four runs of 99 trials apiece for three bandwidth conditions (BB, NB, NNB), for a total of twelve runs. Researchers gave them the task from Experiment 1: in a randomly-ordered trial, discriminate the head-shifted interval from the constant-ITD interval by any means possible. With the lateralization cue likely eliminated after calibration, it was expected they would depend solely on the stimulus dispersion. The actual experimental results can be found in Table V.3, which is laid out in the same manner as Table V.2. The “Overall” percent correct is the rate of correct responses for all 396 trials.

For a 2IFC task, the “detection threshold” is 75 percent correct; above that level, the experimenter can conclude that the listeners are not choosing randomly, but have a clear method of identifying the target interval. One can sum up the Experiment 2 data quite generally by checking the overall percents correct, and noting that no listener exceeds the guessing threshold at any bandwidth. The highest overall percent correct is 53. On closer examination, none of the individual head-shift rows or ITD columns shows performance over threshold, although they do range from 30 to 62 percent. Three individual head-shift/ITD combinations do indicate the possibility of non-random

Table V.3. Experiment 2 discrimination data. This table comprises twelve tables, arranged in three columns (representing three bandwidth conditions) and four rows (one for each listener). The percentages indicate the fraction of trial that listeners correctly discriminated one of three head-shifted stimuli (angles in degrees listed as the row headers in each table) from one of three perfectly-coherent intervals with a particular ITD (the column headers in each table). The overall performance at each ITD and angle is listed at the end of each column and row, respectively. The overall performance of a listener at a condition is listed below the individual tables.

Table V.3: Experiment 2 data

listener	BB		NB		NNB	
	ITD %		ITD %		ITD %	
listener Z	42	480	42	480	42	500
	45	510	45	510	45	530
	49	540	49	540	49	550
	57%	510	51%	510	52%	560
	Overall	51%	Overall	50%	Overall	48%
listener W	42	490	42	490	42	500
	45	520	45	520	45	530
	49	550	49	550	49	560
	62%	550	62%	550	62%	560
	Overall	53%	Overall	52%	Overall	50%
listener X	42	490	42	485	42	490
	45	520	45	515	45	520
	49	550	49	545	49	550
	57%	550	53%	545	50%	550
	Overall	52%	Overall	52%	Overall	48%
listener T	42	490	42	485	42	490
	45	520	45	515	45	520
	49	550	49	545	49	550
	57%	550	53%	545	50%	550
	Overall	52%	Overall	52%	Overall	48%

selection. They are located in the NNB condition for listener W, who was able to answer correctly 77 percent of trials for one combination, and 24 and 25 percent of trials for two other combinations. However, the isolated nature of these successful trials, combined with the lack of consistent performance over the entire NNB condition, support the conclusion that listener W couldn't reliably identify the head-shift interval. Generally, Experiment 2 leads to the conclusion that listeners are unable to distinguish between a head-shifted stimulus and one with a constant ITD over all frequencies. These findings are consistent with those of Hartmann and Wittenberg (1996) and Kulkarni *et al.* (1999).

Experiment 3: “Wrong” phase shifts

Introduction

The original goal of the experiment was to see if head-shifted stimuli were indistinguishable from constant ITD. Both Experiments 1 and 2 showed insensitivity to the dispersion introduced by the spherical-head model. This inability to discriminate head-shifted stimuli from constant-ITD could result from the small range of ITDs included in the head-shift profile, resulting in a slight incoherence that falls below discrimination threshold. Alternatively, there may be some “plausibility” of the spherical-head phase shifts. That is, the apparent similarity between constant-ITD and head-shift intervals might be due to listeners' experience with real-world head shifting of perfectly discrete sources. The presence of a head ensures that a broadband sound cannot be completely coherent. Thus, the auditory system might associate a slightly-incoherent

sound (that fits the head-shift profiles of Figure V.2) with a coherent image representing the source's distinct nature.

Method

To test whether the shape of the head-shift delays mattered in Experiment 2, or if any similar range of ITDs would also be indistinguishable from a constant-ITD stimulus, Experiment 3 implemented an “inverted” head-shift curve and tested in the broadband case. The stimuli were designed by starting with the values for 45-degree incidence as found in Figure V.2, and simply reversing the trend: where the original delays decrease, the inverted delays increase. Refer to Figure V.11, which is a simplified version of Figure V.2 including only the 45-degree shifts (solid line) and the new “wrong 45” inverted curve (dashed line). This produces a stimulus with the same ITD range as the head-shifts, but in an entirely different (and unnatural) pattern. The same process applied to the 42- and 49-degree stimuli as well, to produce three perceptually-separate “wrong head” images.

Aside from changes in the stimuli, the new experiment ran exactly the same as Experiment 2: beginning with calibrated runs to match up constant ITDs to the “wrong head” images. As one might expect, these ITDs were higher than those for the corresponding head-shift patterns. Compare the ITD values in the first row of Table V.4 with those in Table V.2. The same four listeners performed four runs of 99 trials to form the data set. Only the broadband case (20-3000 Hz) was used in this test. See Table V.4 for lateralization results (the format is identical to Table V.2 except that only broadband data were collected, and the expected results are in Table V.1).

Table V.4 - Wrong-head lateralization data									
Listener Z, BB					Listener W, BB				
lat %	530	560	590		lat %	530	560	590	
42	47%	0%	0%	14%	42	43%	25%	4%	24%
45	92%	53%	13%	52%	45	71%	46%	19%	44%
49	100%	100%	43%	82%	49	100%	85%	60%	83%
	82%	47%	18%			73%	51%	27%	
	Overall	49%				Overall	51%		
Listener T, BB					Listener X, BB				
lat %	500	545	590		lat %	530	565	600	
42	73%	39%	25%	42%	42	58%	47%	28%	44%
45	59%	50%	58%	56%	45	67%	57%	33%	50%
49	78%	80%	41%	66%	49	72%	52%	50%	59%
	69%	55%	40%			66%	52%	35%	
	Overall	55%				Overall	51%		

Results and Discussion

Experiment 3 results are found in Table V.5.

Table V.5 - Wrong-head discrimination data									
Listener Z, BB					Listener W, BB				
ITD %	530	560	590		ITD %	530	560	590	
42	53%	46%	47%	49%	42	21%	37%	42%	34%
45	60%	57%	59%	59%	45	55%	54%	58%	55%
49	61%	49%	63%	57%	49	52%	52%	53%	53%
	58%	51%	56%			44%	47%	50%	
	Overall	55%				Overall	47%		
Listener T, BB					Listener X, BB				
ITD %	500	545	590		ITD %	530	565	600	
42	53%	64%	38%	52%	42	55%	42%	59%	52%
45	67%	46%	46%	53%	45	43%	48%	59%	51%
49	54%	58%	58%	57%	49	41%	50%	56%	48%
	58%	56%	47%			47%	46%	58%	
	Overall	54%				Overall	50.25%		

Table V.5 follows the same format as Table V.3, due to the similarity with Experiment 2. Again, the overall percent correct shows little deviation from guessing. Only listener W showed performance that was appreciably different from 50 percent (21 percent) at one of the nine conditions. The results of the “wrong head” experiment seem to support the likelihood that, while listeners cannot distinguish between constant-ITD and head-shifted stimuli, it is likely because the distribution of diffraction-induced phase-shifts is not broad enough to exceed the just-noticeable-difference (jnd) for coherence.

Head-shift coherence, and delayed coherence discrimination

Introduction

Data from the spherical-head experiments made it clear that listeners could not distinguish between combinations of constant-ITD and head-shifted stimuli with equivalent lateral positions. Intrigued, the experimenters searched for an explanation of this result. Kulkarni *et al.* theorized that listeners established an “overall empirical ITD” for the complex stimulus using delays from the lower frequencies (below 2 kHz). Note that this ITD was set at a delay corresponding to the interaural cross-correlation peak. If both the “overall” ITD and the magnitude spectrum matched up in both intervals, then Kulkarni *et al.* argue they should be perceptually equivalent. The “Equivalence of position” experiment can establish the effective ITD for head-shifted images, allowing a test of this theory. That test is detailed in the section *Analysis of peak cross-correlation lag values*, using the results of *Coherence calculations* below.

Another possibility was that the combination of head-related phase shifts did not introduce enough incoherence to distinguish itself from the perfectly-coherent single-delay noise. The shifts were, in fact, limited to a narrow bandwidth, and bands above and below the sloped region (e.g. on the 30-degree plot of Figure V.2, frequencies from 400 to 1200 Hz) were perfectly coherent. This line of reasoning is explored in *Discussion-Lateral position* and the following sections.

Coherence calculations

To address both of the above explanations, the experimenters made recordings of the various stimuli featured in all three spherical-head experiments. A program using the Tucker-Davis Array Processor used one-second clips of the above recordings to calculate cross-correlations (CCs) over lag τ from -2 to $+2$ ms. Table V.6 indicates the conditions and bandwidths for which these calculations were performed by either a “Y” or a figure number for several plots representative of the various calculations.

Table V.6 - Stimulus cross-correlations			
Condition	BB	NB	NNB
30-degree head shift	Y	Y	Y
42-degree head shift	Y	Y	Y
45-degree head shift	Figure V.12	Figure V.13	Figure V.14
49-degree head shift	Y	Y	Y
42-degree wrong head	Y	N	N
45-degree wrong head	Figure V.18	N	N
49-degree wrong head	Y	N	N
520- μ s constant ITD	Figure V.15	Figure V.16	Figure V.17

Cross-correlations performed of experimental stimuli

The figures show cross-correlation value versus lag. One interesting feature of note: comparing the 45-degree BB Figure V.12 with the “wrong-head” 45-degree Figure V.18, the two curves appear similar, but reversed along the lag axis. This is due to the nature of the “wrong-head” stimulus, which uses an inverted 45-degree ITD curve. The relevant data are listed in the title at the top of each figure; see Figure V.12 for an example. The title shows the peak CC value that defines the stimulus coherence (maxCC:0.963) as well as the interaural lag of that peak (μ s:420). A summary of data from all CC plots is in Table V.7. The left column lists the stimulus condition, the middle contains the maximum cross-correlation obtained for that condition, and the right column indicates the interaural time lag where that maximum was located:

Table V.7 - CC peak magnitudes and lags		
Stimulus	Max CC	Lag τ (μs)
30-degree head shift, BB	0.982	300
30-degree head shift, NB	0.988	340
30-degree head shift, NNB	0.998	340
42-degree head shift, BB	0.971	400
42-degree head shift, NB	0.971	420
42-degree head shift, NNB	0.999	500
45-degree head shift, BB	0.963	420
45-degree head shift, NB	0.972	460
45-degree head shift, NNB	0.999	520
49-degree head shift, BB	0.961	460
49-degree head shift, NB	0.975	500
49-degree head shift, NNB	0.999	560
520- μ s constant ITD, BB	1.000	520
520- μ s constant ITD, NB	1.000	520
520- μ s constant ITD, NNB	1.000	520
“Wrong” 42-degree head shift	0.968	620
“Wrong” 45-degree head shift	0.964	660
“Wrong” 49-degree head shift	0.960	700

Analysis of peak cross-correlation lag values

First analysis of the data in Table V.7 must ensure that the lags are reasonable. Since the head-shifted stimuli contain mostly delays between the low- and high-frequency limits defined in Equations 3 and 4, one expects an effective ITD within that range. Assuming 8.75 cm for head-radius a , and 343 m/s for speed of sound c , calculations found:

Stimulus	Low-f limit	High-f limit
30-degree head shift, BB	382	255
42-degree head shift, BB	512	341
45-degree head shift, BB	541	361
49-degree head shift, BB	578	385
520- μ s constant ITD, BB	520	520
“Wrong” 42-degree head shift	512	635
“Wrong” 45-degree head shift	541	671
“Wrong” 49-degree head shift	578	710

The peak cross-correlation lag for each stimulus in Table V.7 lies between delays for the low- and high-frequency limits in Table V.8. Within the head-shift and “wrong-head” groups, the peaks occurred in approximately the same relative position between limits. In addition, the bandwidths selected for NB and NNB cases did tend to favor higher ITDs, which explains the peak’s migration along the lag axis as each head-shift stimulus was filtered into narrower bands. Thus, these lags corresponding to peak cross-correlations appear to meet expectations.

Discussion-Lateral position

The cross-correlation data were used to check the experimental results against the “overall empirical ITD” concept of Kulkarni *et al.* Table V.9 is a comparison of the

“overall” ITD predicted by the cross-correlation peak lag, and the calibrated ITDs (averaged over listeners) for constant-ITD stimuli established in the “Equivalence of position” experiment.

Table V.9 - Lag/ITD comparison		
Stimulus	Peak CC	Calibrated
42-degree head shift, BB	400	484
42-degree head shift, NB	420	479
42-degree head shift, NNB	500	495
45-degree head shift, BB	420	515
45-degree head shift, NB	460	514
45-degree head shift, NNB	520	529
49-degree head shift, BB	460	546
49-degree head shift, NB	500	545
49-degree head shift, NNB	560	563
“Wrong” 42-degree head shift	620	523
“Wrong” 45-degree head shift	660	558
“Wrong” 49-degree head shift	700	593

Let the 45-degree BB stimulus be an example. Taking the lag from Table V.7, the “overall empirical ITD” concept predicts a 420- μ s delay for the listener’s internal image. However, the corresponding constant-ITD stimulus established in Experiment 2, calibrated to match position with the 45-degree broadband signal, had a average delay of 515 μ s. This difference in delay certainly exceeds the ITD just-noticeable-difference of around 20 μ s at a 500 μ s delay (Colburn and Durlach, 1978). The “Max CC” lags and calibrated delays do match up more closely as the bandwidth decreases, which may be related to the narrowband stimuli’s higher coherences. The 45-degree NNB stimulus exhibits peak CC at 520 μ s, while the matching constant-ITD was set to 529 μ s. In the BB and NB conditions, the constant ITDs were higher than the corresponding peak CC lags, tending to approach the low-frequency limits of Table V.8. This may indicate that

listeners were giving more weight to ITDs at the lower frequencies, which is expected in light of results reported in Chapters II and III.

If one was to interpret the data of Table V.9 according to “overall empirical ITD,” assuming that the lateral position of the stimulus corresponded to the cross-correlation peak, the listeners should still have been able to distinguish head-shifted from constant-ITD intervals using lateral position, at least in the broadband case. The experimenters have some confidence that the lateral cue was indeed eliminated, due to the extensive calibration. Therefore, results do not appear to agree with the concept of “overall ITD” as proposed by Kulkarni *et al.*

Discussion-Correlation

It is left to examine the cross-correlation data for cues on the basis of coherence. First, the result that NNB correlations were highest among head-shifted stimuli seemed surprising, since the NNB condition was designed to have the highest dispersion of ITDs across frequency and thus expected to have the lowest cross-correlation peak. On further reflection, this result is predictable because in the extreme NNB limit (a band so small that only a sine tone is left) the maxCC must go to one. Also, Trahiotis *et al.* (2001) showed that for a broadband distribution of ITDs, passing the stimulus through narrowband filter led to better ITD discrimination, which was likely due to an increase in correlation.

Each trial, the auditory system made a coherence comparison between two stimuli. The constant-ITD stimulus was always 100% coherent (maxCC = 1). On the other hand, the various head-shift stimuli all had coherences less than one, allowing for a

possible cue. Granted, the narrow-narrowband cases showed peaks that should be indistinguishable from one (coherences ≥ 0.998). But previous experiments (see Chapter II) have shown that listeners are extremely sensitive to decorrelation. The pool of listeners exhibited just-noticeable-differences (jnds) as low as one percent, especially with respect to a coherence of one and a broadband stimulus. Gabriel and Colburn (1981) also discovered jnds of such magnitude. Thus, it would seem clear that at least the “wrong head” stimuli should be perceptually different from a constant ITD, based on decorrelations approaching four percent. There was no immediate reason to believe that BB and NB head-shift cases couldn’t be distinguished as well. According to current understanding of coherence sensitivity, listeners should have succeeded in discriminating head-shift intervals from constant ITD. It is possible that listeners’ coherence jnds were actually larger due to the imposed ITD in the vicinity of 500 μs —leading to a test of off-center coherence sensitivity with a final experiment.

Experiment 4: Off-axis coherence discrimination

Introduction and Method

While coherence discrimination is an easy task for images centered in the head, the auditory system loses resolution when the image’s interaural delay puts it off-center (Pollack and Trittipoe, 1959a&b). This concept is lent credence by the modeled high density of cross-correlation detectors at small delays, which becomes progressively sparser as the delay increases (Stern and Trahiotis, 1995). The “detector” neurons are crucial to ITD resolution, and thus, coherence discrimination. Indeed, Colburn and

Durlach (1978) state that once a coherent noise is delayed more than one ms (about the maximum delay one could expect for a real-world sound incident on the head), it “begins to lose its compactness”. In the region where lag τ is between 10 and 30 ms, the coherent noise is indistinguishable from an incoherent noise.

Experiment 4 repeated the study of Chapter II, testing decorrelation jnds for a broadband stimulus. The only change was the addition of a delay line to the left channel, so listeners found the intracranial image offset to the right. Three listeners (W and Z, who are previously mentioned and were subjects in Chapter II, and listener A, a 20-year-old male with normal hearing) participated in three experimental runs at each of four delays: 0, 500, 1000, and 4000 μ s. The 500- μ s condition is particularly interesting, because it applies to Experiments 2 and 3 wherein most ITDs were set near 500.

Results and Discussion

The results of these runs are found in Figure V.19, which shows that thresholds for listeners W and Z in zero-delay case (equivalent in every way to the BB case of the Chapter II task) matched well with their prior results. Then, as the interaural delay was increased, all listeners required a greater decorrelation to succeed at the task. The results lend some support to the explanation that decorrelation of an off-center image is more difficult to detect. They also make sense in light of the described “loss of compactness” for longer delays, and the well-known fact that coherence jnd’s increase rapidly as the reference correlation is reduced from one (Pollack and Trittipoe, 1959b). As initially postulated, coherence jnds are degraded at 500 μ s. However, they show that listeners could still distinguish a coherence of 0.96-0.98 from 1. The lowest coherence among all

the head-shift stimuli is 0.96, for the “wrong-head” 49-degree case. The results do not preclude our listeners from making a coherence discrimination, based on the sensitivity data of Experiment 4. The question remains: why did listeners fail to discriminate head-shift stimuli?

The Nature of Incoherence

The most likely explanation for the listeners’ inability to distinguish head-related dispersions from no dispersion (constant ITD) is the unusual makeup of the head-shift stimuli. Compare them to the slightly-incoherent signals used for the Chapter II experiment, in which listeners could identify small decorrelations. Those signals began with coherent noise, then had statistically independent noise of identical bandwidth added to one ear, reducing the overall coherence throughout the entire signal spectrum. By contrast, the broadband head-shift spectra were mostly coherent outside the “transition regions” of 400-1200 Hz for the 30-degree stimulus, and 400-1800 Hz for the 45-degree stimulus. Bands above and below these regions were set with relatively constant interaural delays.

Knowing that models of the auditory system separate the signal into critical bands for analysis (Hartmann, 1997), all head-shift bands outside the transition regions would exhibit correlation peaks near one. Of course, the NB and NNB conditions do not contain regions of constant-ITD. But the character of ITDs that are a smooth function of frequency (i.e. are not discontinuous and not random) is to produce very little decorrelation, as demonstrated by the NNB cases that had consistently high coherences.

The cross-correlations for individual critical bands likely resemble that for the NNB case. Thus, while the overall coherence for BB stimuli is low, each critical band probably has a coherence nearer one, which would preclude discrimination on that basis.

Summary and Conclusion

After presenting listeners with a range of theoretical head-shift stimuli (covering four different angles, three bandwidths, and one inversion), this study concludes that listeners cannot distinguish a signal with head-related dispersion from one that is perfectly coherent. They were insensitive to the decorrelation introduced by the presence of a spherical head. It was possible for listeners to discriminate based on lateral position, as Kulkarni *et al.* stated. Surprisingly, experiments found that the head-shift stimuli had coherences sufficiently different from one that they exceeded the known coherence jnds.

The experimenters explored reasons for the discovered insensitivity to incoherence, hypothesizing that coherence discrimination may be more difficult when the image is off-center. This hypothesis was proved correct, specifically at a delay of 500 μ s, which is representative of the ITDs involved in our experiments. However, the coherence jnds at that delay were still too small to explain the head-shift experiment results. A final postulation was that the problem dwelled within coherence measurement itself. The broadband coherences of head-shift stimuli were low, but cross-correlations within each critical band were much closer to one. Smoothly-varying ITDs do not produce as much decorrelation as the addition of statistically-independent noise, as

demonstrated by the NNB condition coherence values. Thus, it is clear that head-related dispersion does not introduce enough incoherence for listeners to distinguish it from a constant ITD. The analysis shows that listeners cannot identify head-related dispersion without the use of lateral position cues, simply because the change in coherence is too small for them to detect.

Appendix A

While the results of Experiment 4 did not reveal why listeners were unable to discriminate head-shift stimuli from constant ITD, they did raise a concern of their own. The experimenters found it unintuitive that listeners could apparently judge decorrelation even at extreme ITDs. As the ITD increases, there should be a delay at which the image is perceived as incoherent, thus making the task impossible. Listeners A and Z continued their runs to higher delays—Z could still reliably discriminate coherence at 12-ms delays, while A did so up to 8 ms. While these upper limits show that the discrimination task can be disrupted, they appear well beyond the expected boundary. Interviews with the listeners revealed that as the delay increased, they began to detect a pitch embedded in the noise. As the decorrelation cue gradually failed them, they used the pitch as a cue to complete the task. This phenomenon may be caused by dichotic repetition pitch, or DRP. According to Bilsen (1995), the DRP is detectable for correlated white noise with an ITD between 3 and 20 ms, which happens to be in the range of delays plotted in Figure V.18. Bilsen reports that many (but not all) listeners are able to identify a pitch in the noise, with a period determined by the delay time. DRP could explain the listeners' ability to perform the coherence discrimination task at high delays due to the presence or lack of a pitch, representing correlated and decorrelated stimuli, respectively.

References

- Bilsen, F.A. "What do dichotic pitch phenomena tell us about binaural hearing?" "Advances in Hearing Research," Proceedings of the 10th International Symposium on Hearing, ed. G.A. Manley *et al.* (1995).
- Brungart, D.S., Durlach, N.I., Rabinowitz, W.M. "Auditory localization of nearby sources II: Localization of a broadband source," *J. Acoust. Soc. Am.* **106**, 1956-1968 (1999).
- Durlach, N.I. and Colburn, H.S. "Binaural Phenomena," *Handbook of Perception* vol. 4. ed. E. Carterette, pp. 365-466 (1978).
- Gabriel, K.J. and Colburn, H.S. "Interaural correlation discrimination: I. Bandwidth and level dependence," *J. Acoust. Soc. Am.* **69**, 1394-1401 (1981).
- Hartmann, W.M. "Auditory Filters," Signals, Sound and Sensation, AIP Press (1997).
- Hartmann, W.M. and Wittenberg, A. "On the externalization of sound images," *J. Acoust. Soc. Am.* **99**, 3678-3688 (1996).
- Kistler, D.J. and Wightman, F.L. "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.* **91**, 1637-1647 (1992).
- Kuhn, G.F. "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.* **62**, 157-167 (1977).
- Kuhn, G.F. "Physical Acoustics and Measurements Pertaining to Directional Hearing," Directional Hearing, ed. W.A. Yost and G. Gourevitch, pp. 3-25 (1987).
- Kulkarni, A., Isabelle, S.K. and Colburn, H.S. "Sensitivity of human subjects to head-related transfer-function phase spectra," *J. Acoust. Soc. Am.* **105**, 2821-2840 (1999).
- Pollack, I. and Trittipoe, W.J. "Binaural listening and interaural noise cross correlation," *J. Acoust. Soc. Am.* **31**, 1250-1252 (1959). (a)
- Pollack, I. and Trittipoe, W.J. "Interaural noise correlations: Examination of variables," *J. Acoust. Soc. Am.* **31**, 1616-1618 (1959). (b)
- Stern, R.M. and Trahiotis, C. "Models of Binaural Interaction," Hearing ed. B. Moore, pp. 347-386 (1995).
- Tobias, J.V. and Schuber, E.R. "Effective onset duration of auditory stimuli," *J. Acoust. Soc. Am.* **31**, 1595-1605 (1959).

Trahiotis, C., Bernstein, L.R. and Akeroyd, M.A. "Manipulating the straightness and curvature of patterns of interaural cross correlations affects listeners' sensitivity to changes in interaural delay," J. Acoust. Soc. Am. **109**, 321-330 (2001).

Wightman, F.L. and Kistler, D.J. "Headphone simulation of free-field listening," J. Acoust. Soc. Am. **85**, 858-867 (1986).

Wightman, F.L. and Kistler, D.J. "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648-1661 (1992).

Figure V.1. Depiction of a human head and an incoming sound wave. The wave is assumed to be planar and incident at azimuthal angle θ_{inc} from the forward direction. The “directly irradiated ear” receives the sound without interference, while the “shadowed ear” is separated from the sound by the head. The radius of the head is labeled “a”. (Kuhn, 1987)

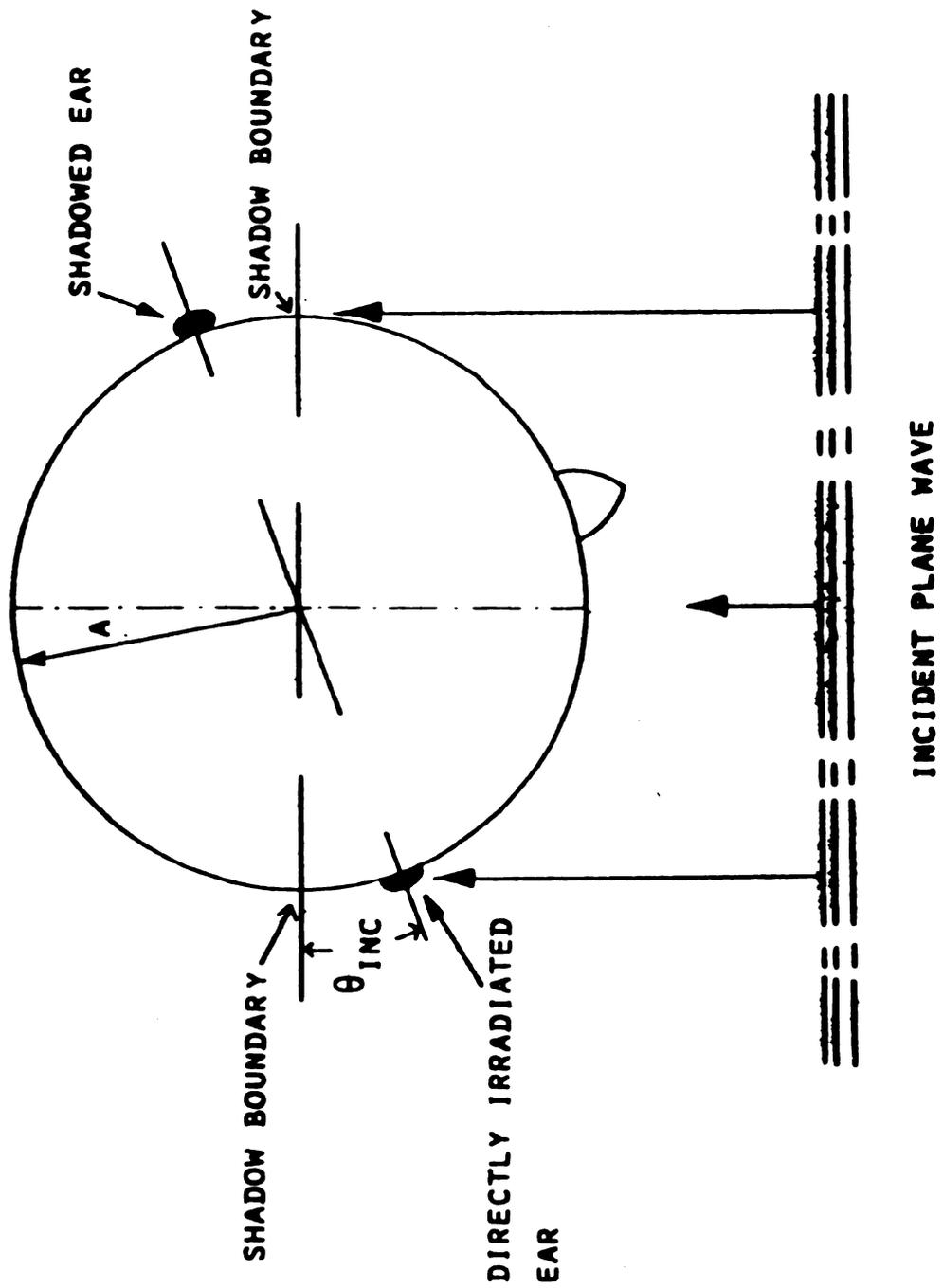


Figure V.1

Figure V.2. Head-related dispersion according to Kuhn's spherical head model. This graph plots the ITD (y-axis) as a function of log frequency for sound incident on a spherical head from various angles. The dashed line indicates ITDs that equal one-half period for a particular frequency. The open circles on the left are the resultant ITDs from a low-frequency approximation. The closed squares on the right are a high-frequency approximation which is simply two-thirds the ITD values of the open circles. The open triangles further right are the predictions of the Woodworth formula.

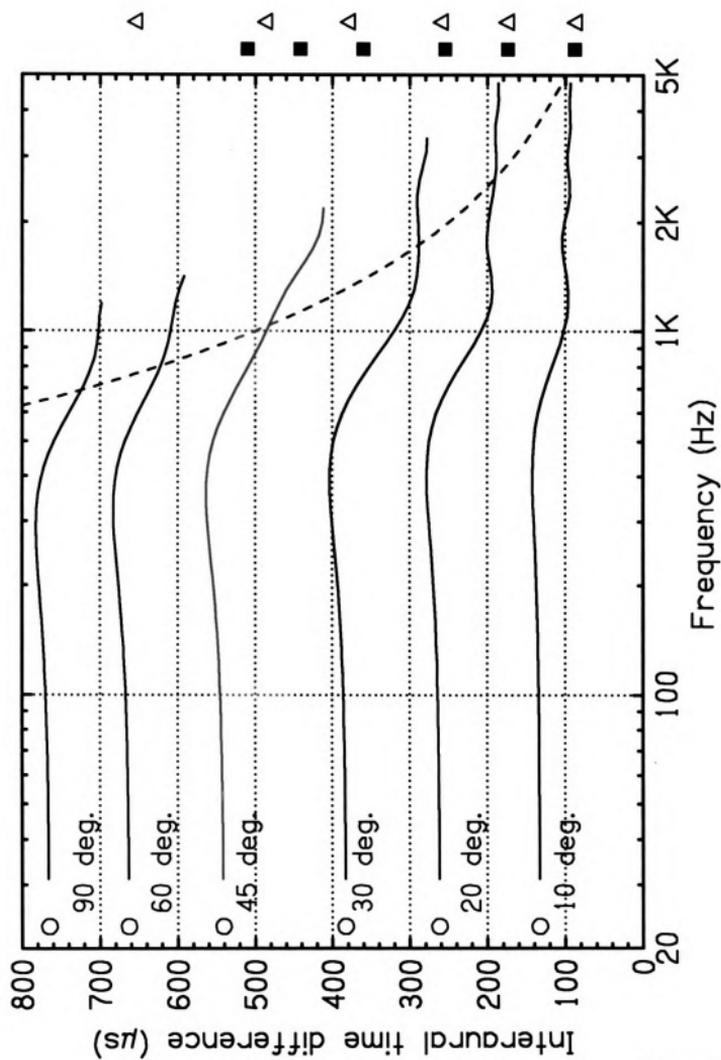


Figure V.2

Figure V.3. Results of Experiment 1: Discrimination at 30 degrees. This graph plots the percentage of trials in which listener T correctly identified the head-shifted stimulus versus the ITD of the perfectly-coherent interval (ten values from 260 to 440 μ s). The three curves represent results from broadband (BB), narrowband (NB), and narrow-narrowband (NNB) cases. The dashed horizontal line at 50 percent correct indicates guessing level. The thin horizontal line marks the range of ITDs included in the BB and NB head-shifted stimuli, while the thick line indicates those ITDs included in the NNB head-shifted stimulus. The dashed vertical line marks the low-frequency limit ITD for a sound incident at 30 degrees. Error bars on data points represent two standard errors.

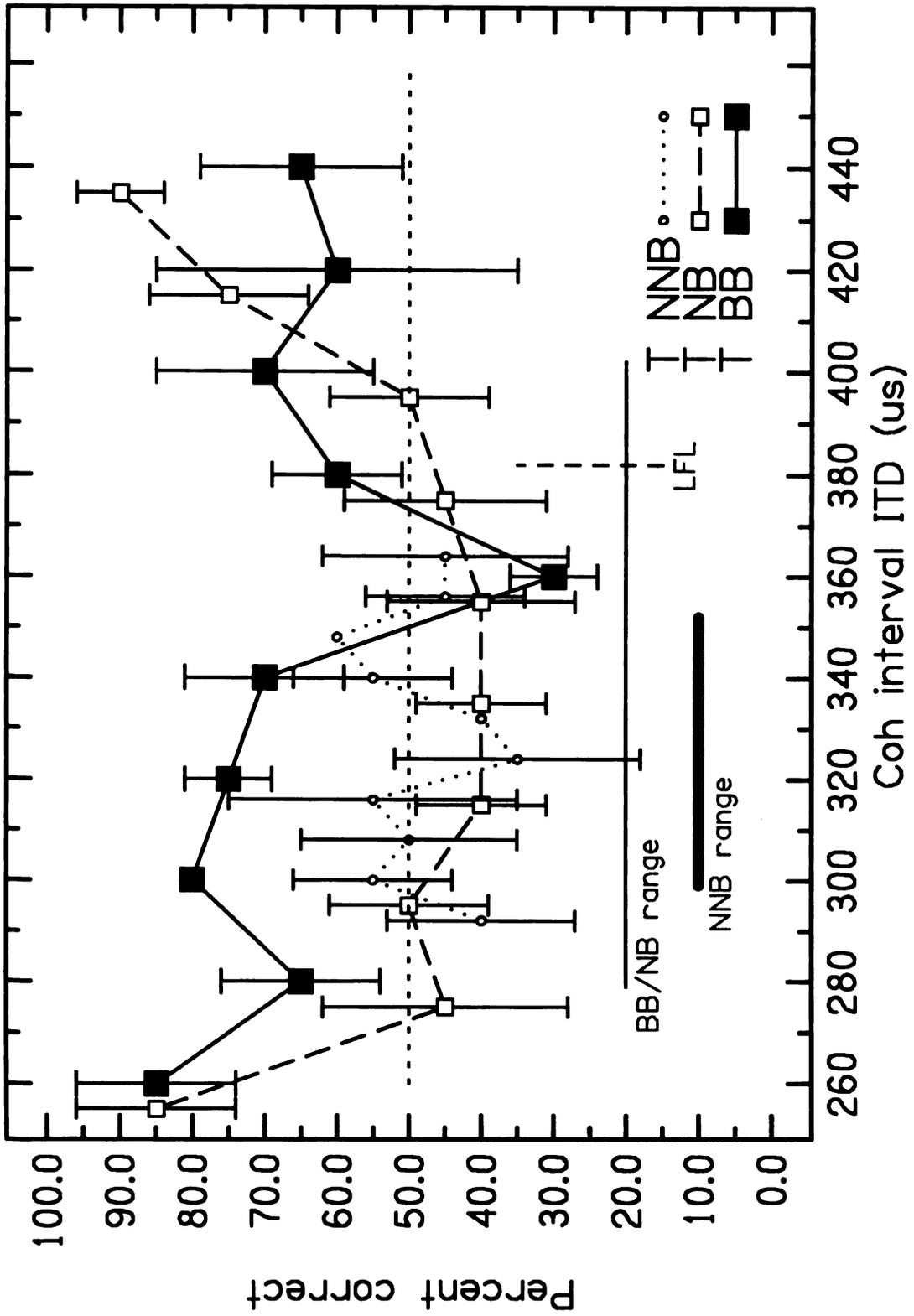


Figure V.3

Figure V.4. Results of Experiment 1: Discrimination at 30 degrees. This graph plots the percentage of trials in which listener *W* correctly identified the head-shifted stimulus versus the ITD of the perfectly-coherent interval (ten values from 260 to 440 μ s). The three curves represent results from broadband (BB), narrowband (NB), and narrow-narrowband (NNB). The dashed horizontal line at 50 percent correct indicates guessing level. The thin horizontal line marks the range of ITDs included in the BB and NB head-shifted stimuli, while the thick line indicates those ITDs included in the NNB head-shifted stimulus. The dashed vertical line marks the low-frequency limit ITD for a sound incident at 30 degrees. Error bars on data points represent two standard errors.

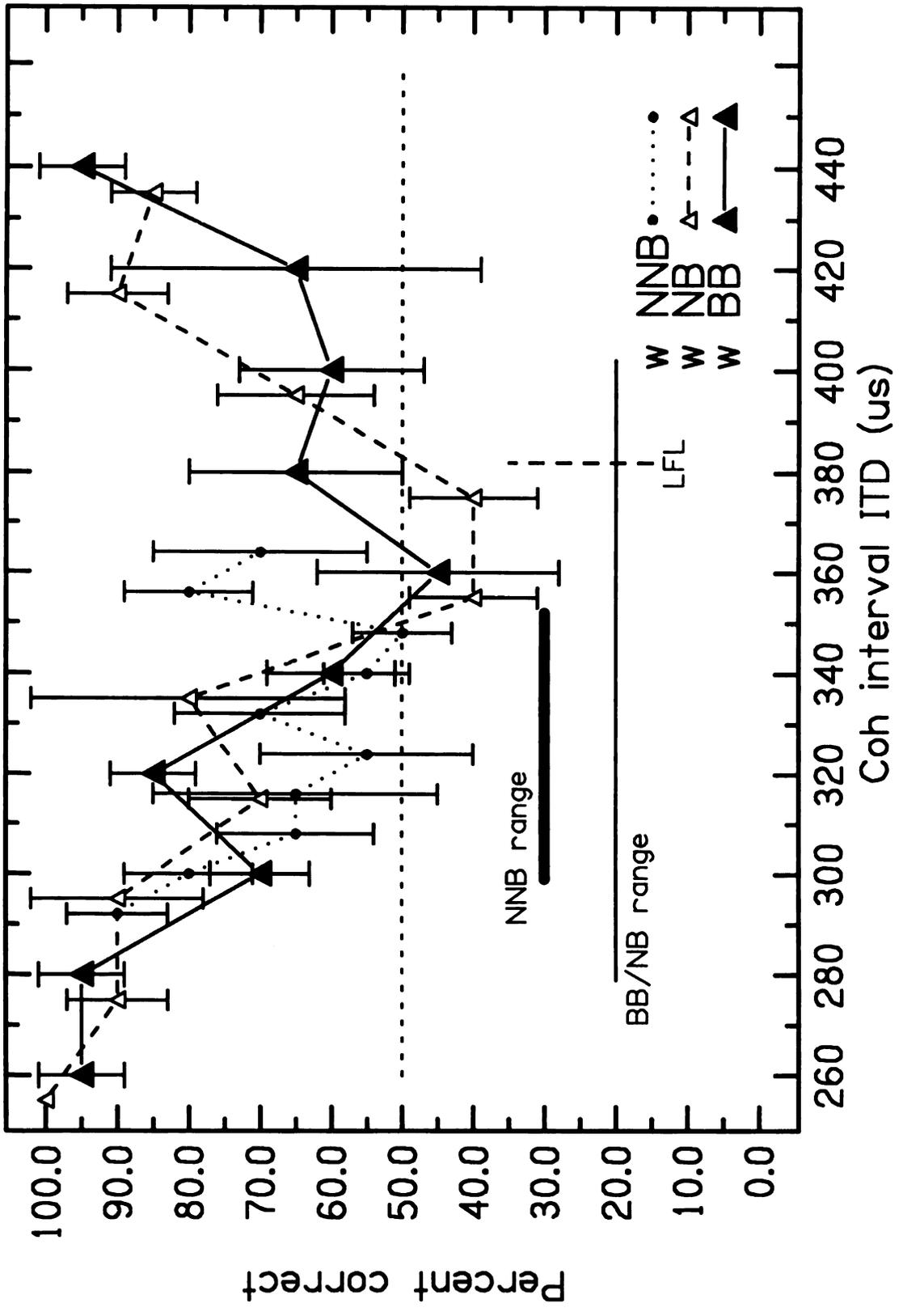


Figure V.4

Figure V.5. Results of Experiment 1: Discrimination at 30 degrees. This graph plots the percentage of trials in which listener X correctly identified the head-shifted stimulus versus the ITD of the perfectly-coherent interval (ten values from 260 to 440 μ s). The three curves represent results from broadband (BB), narrowband (NB), and narrow-narrowband (NNB). The dashed horizontal line at 50 percent correct indicates guessing level. The thin horizontal line marks the range of ITDs included in the BB and NB head-shifted stimuli, while the thick line indicates those ITDs included in the NNB head-shifted stimulus. The dashed vertical line marks the low-frequency limit ITD for a sound incident at 30 degrees. Error bars on data points represent two standard errors.

photo
analysis
of the
low-
frequency
NB has
lead-
or a star

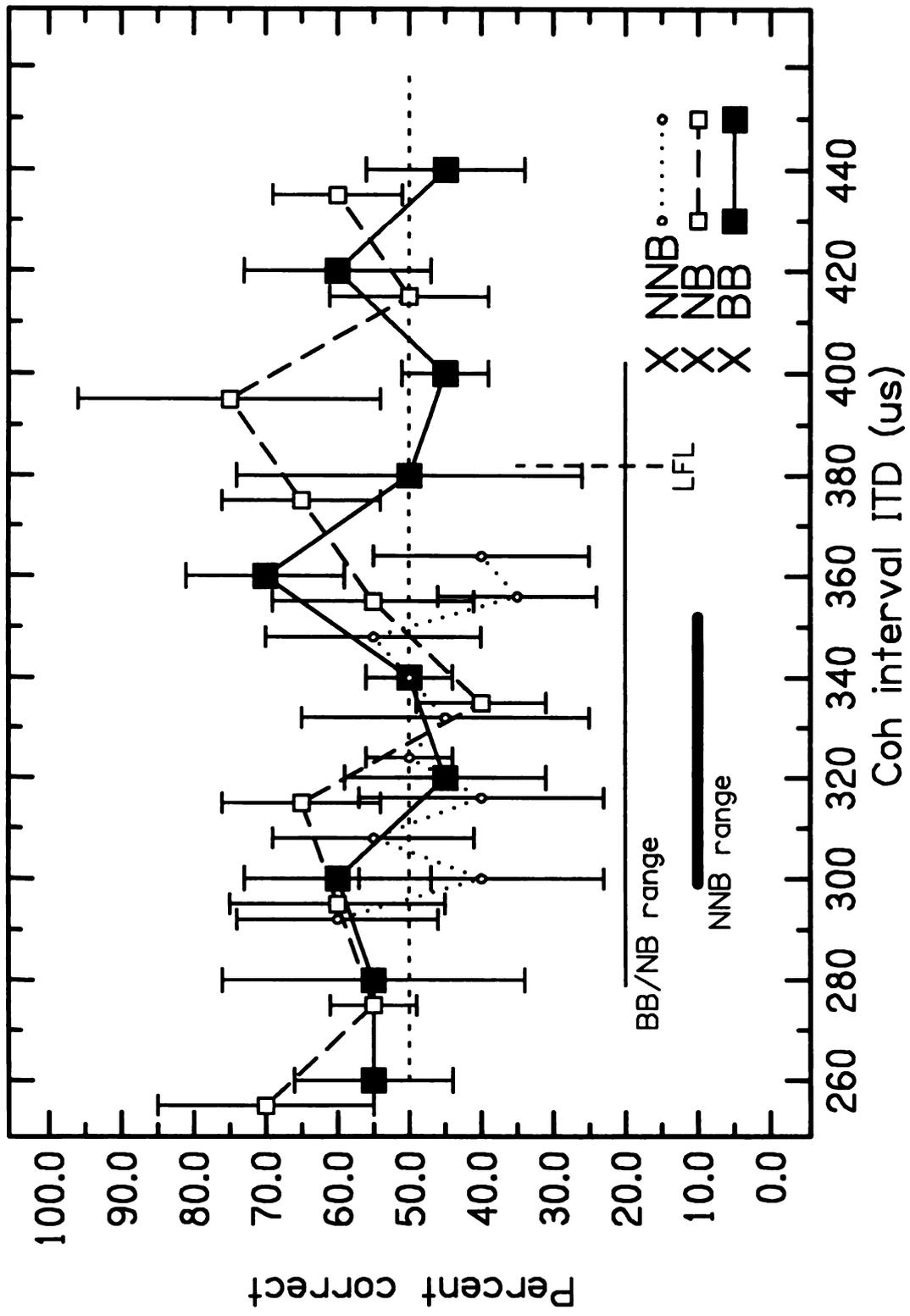


Figure V.5

Figure V.6. Results of Experiment 1: Discrimination at 30 degrees. This graph plots the percentage of trials in which listener Z correctly identified the head-shifted stimulus versus the ITD of the perfectly-coherent interval (ten values from 260 to 440 μ s). The three curves represent results from broadband (BB), narrowband (NB), and narrow-narrowband (NNB). The dashed horizontal line at 50 percent correct indicates guessing level. The thin horizontal line marks the range of ITDs included in the BB and NB head-shifted stimuli, while the thick line indicates those ITDs included in the NNB head-shifted stimulus. The dashed vertical line marks the low-frequency limit ITD for a sound incident at 30 degrees. Error bars on data points represent two standard errors.

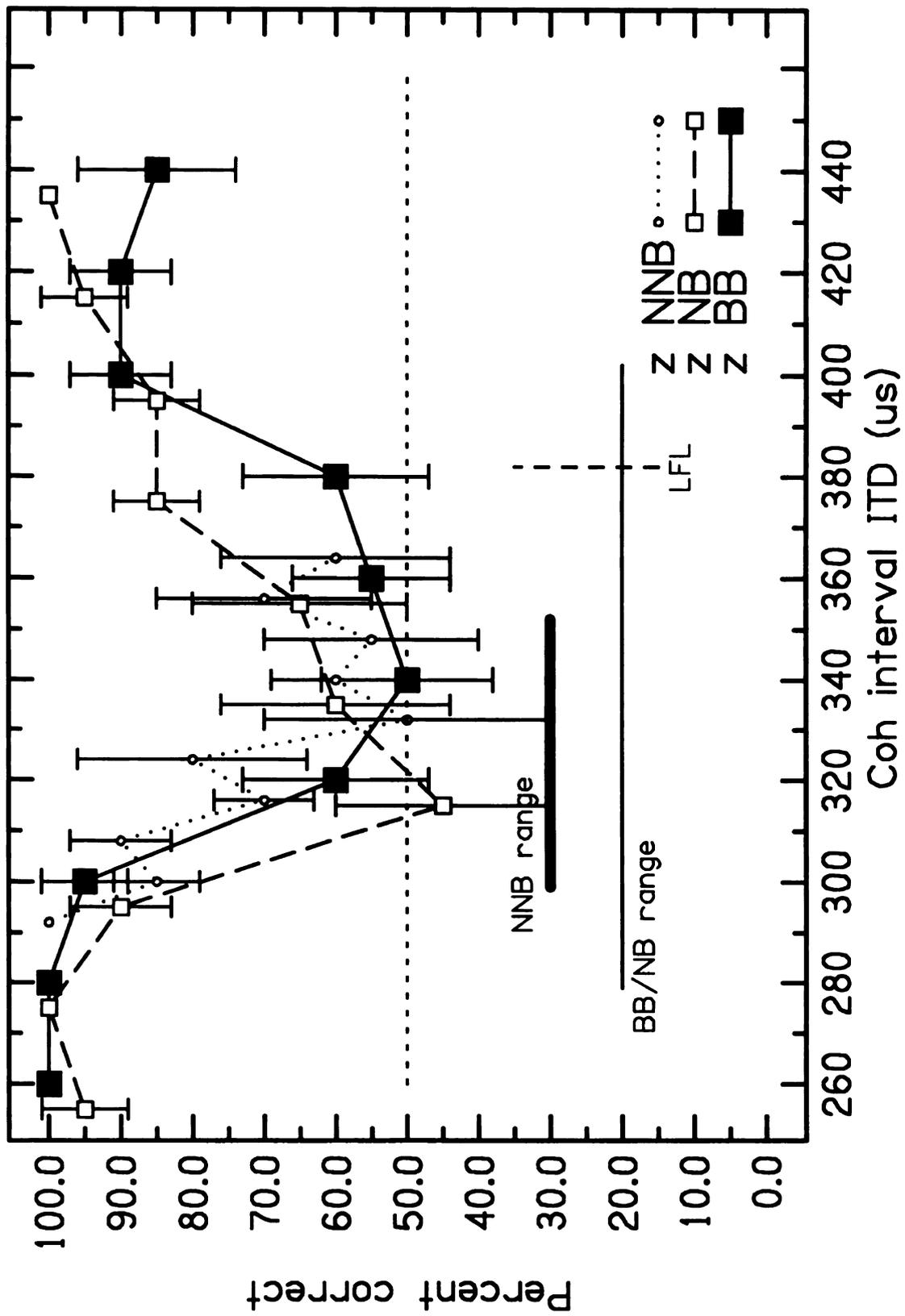


Figure V.6

Figure V.7. Results of Experiment 1: Discrimination at 45 degrees. This graph plots the percentage of trials in which listener T correctly identified the head-shifted stimulus versus the ITD of the perfectly-coherent interval (ten values from 400 to 580 μ s). The three curves represent results from broadband (BB), narrowband (NB), and narrow-narrowband (NNB). The dashed horizontal line at 50 percent correct indicates guessing level. The thin horizontal line marks the range of ITDs included in the BB and NB head-shifted stimuli, while the thick line indicates those ITDs included in the NNB head-shifted stimulus. The dashed vertical line marks the low-frequency limit ITD for a sound incident at 45 degrees. Error bars on data points represent two standard errors.

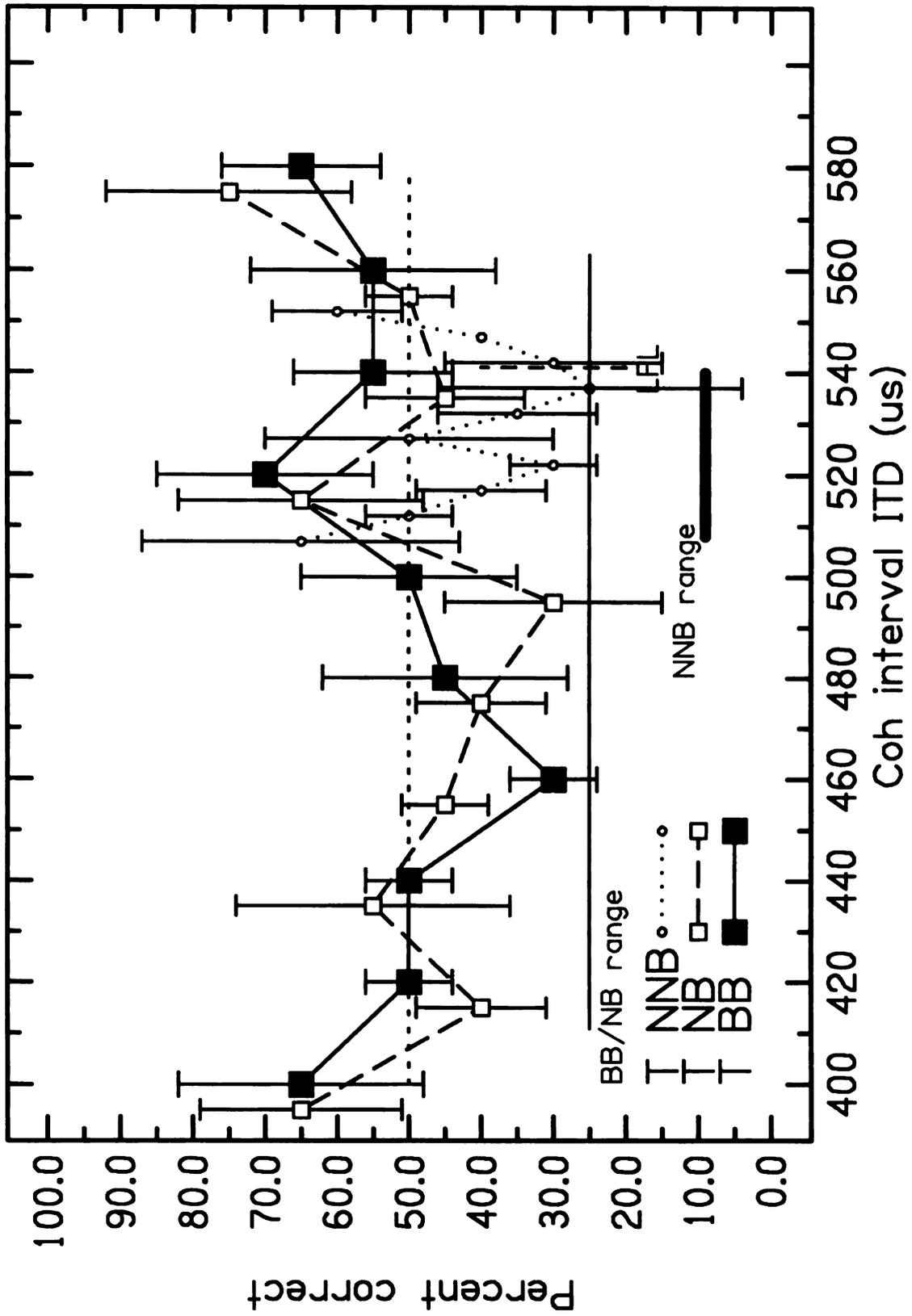


Figure V.7

Figure V.8. Results of Experiment 1: Discrimination at 45 degrees. This graph plots the percentage of trials in which listener W correctly identified the head-shifted stimulus versus the ITD of the perfectly-coherent interval (ten values from 400 to 580 μ s). The three curves represent results from broadband (BB), narrowband (NB), and narrow-narrowband (NNB). The dashed horizontal line at 50 percent correct indicates guessing level. The thin horizontal line marks the range of ITDs included in the BB and NB head-shifted stimuli, while the thick line indicates those ITDs included in the NNB head-shifted stimulus. The dashed vertical line marks the low-frequency limit ITD for a sound incident at 45 degrees. Error bars on data points represent two standard errors.

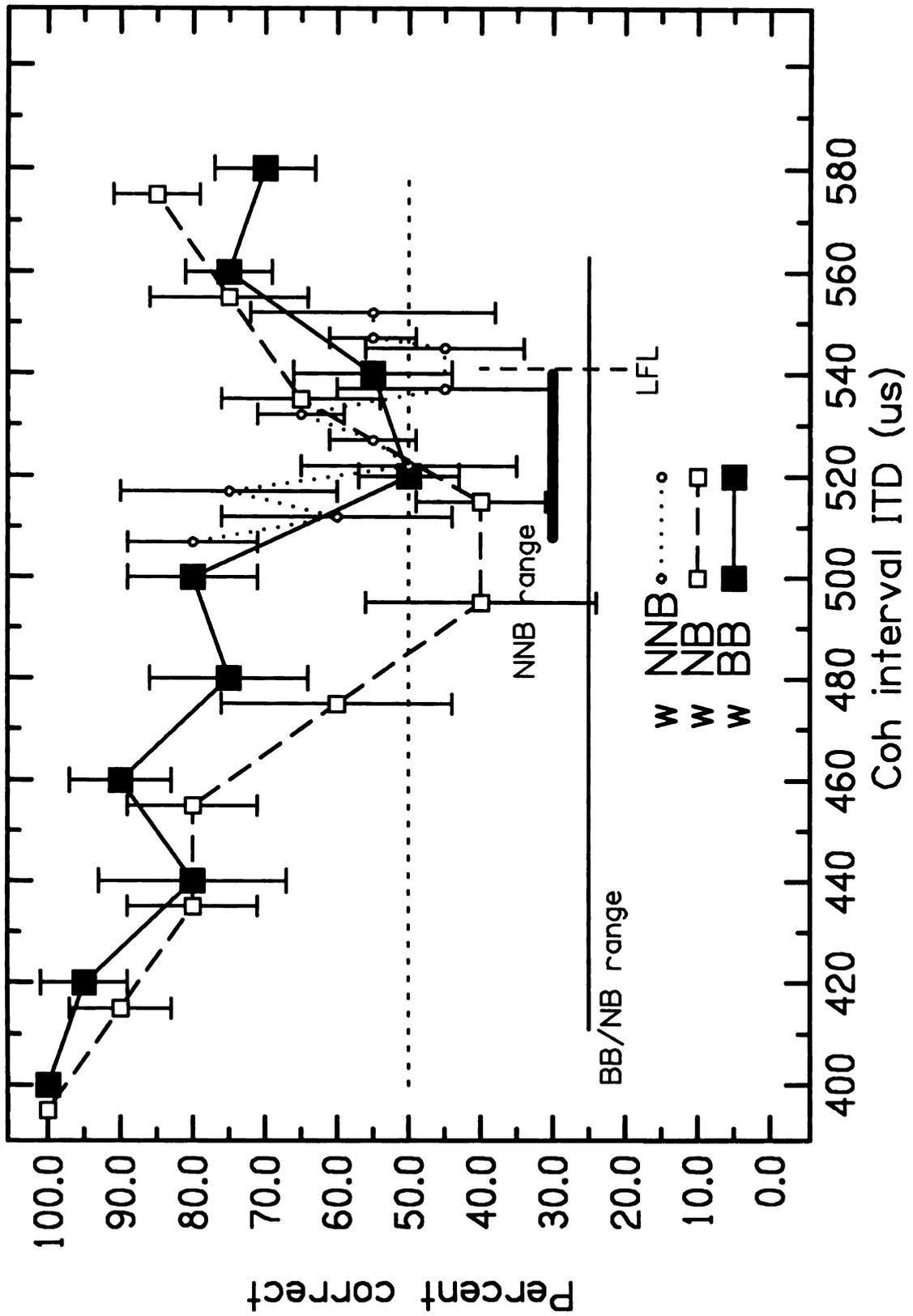


Figure V.8

Figure V.9. Results of Experiment 1: Discrimination at 45 degrees. This graph plots the percentage of trials in which listener X correctly identified the head-shifted stimulus versus the ITD of the perfectly-coherent interval (ten values from 400 to 580 μ s). The three curves represent results from broadband (BB), narrowband (NB), and narrow-narrowband (NNB). The dashed horizontal line at 50 percent correct indicates guessing level. The thin horizontal line marks the range of ITDs included in the BB and NB head-shifted stimuli, while the thick line indicates those ITDs included in the NNB head-shifted stimulus. The dashed vertical line marks the low-frequency limit ITD for a sound incident at 45 degrees. Error bars on data points represent two standard errors.

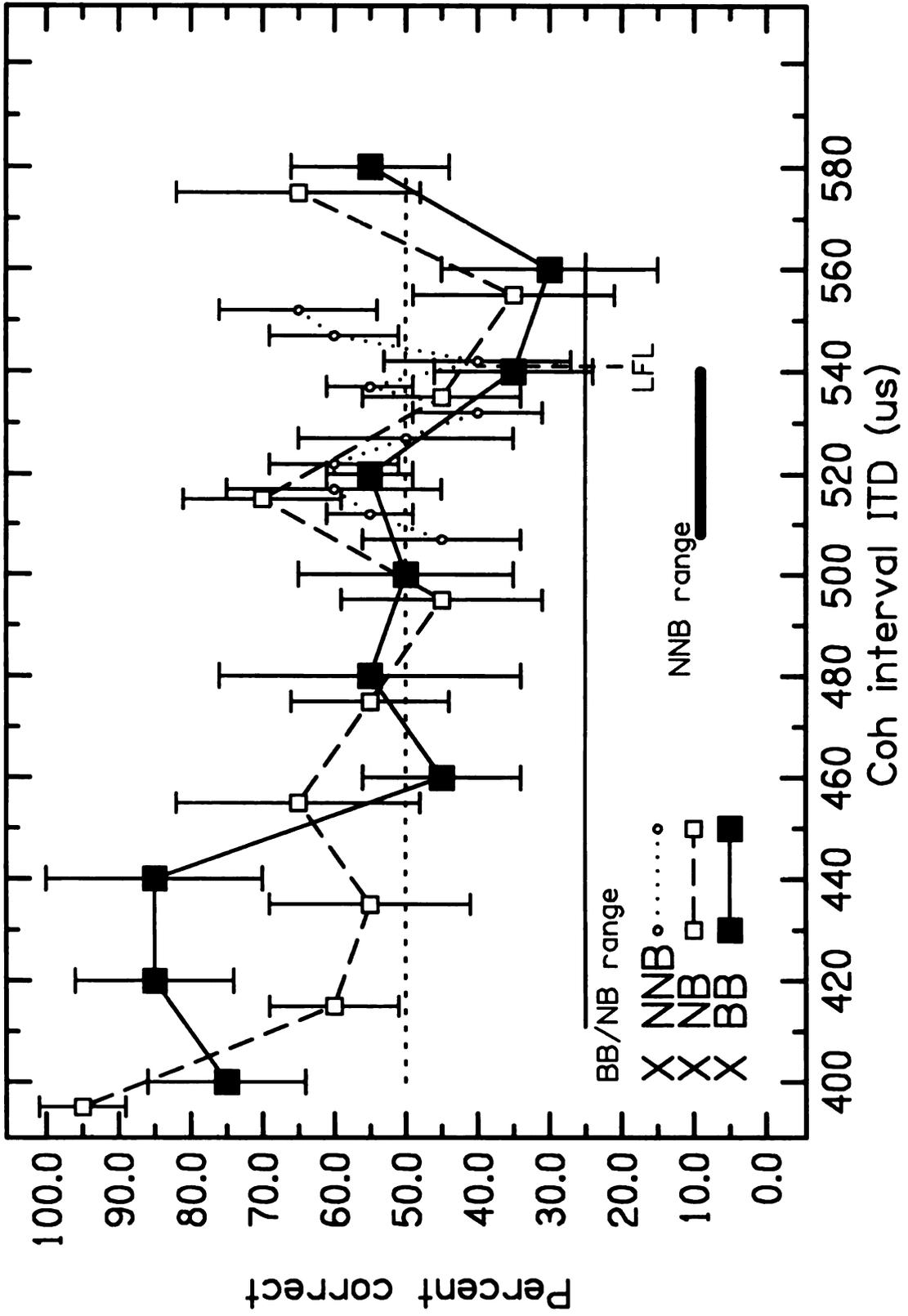


Figure V.9

Figure V.10. Results of Experiment 1: Discrimination at 45 degrees. This graph plots the percentage of trials in which listener Z correctly identified the head-shifted stimulus versus the ITD of the perfectly-coherent interval (ten values from 400 to 580 μ s). The three curves represent results from broadband (BB), narrowband (NB), and narrow-narrowband (NNB). The dashed horizontal line at 50 percent correct indicates guessing level. The thin horizontal line marks the range of ITDs included in the BB and NB head-shifted stimuli, while the thick line indicates those ITDs included in the NNB head-shifted stimulus. The dashed vertical line marks the low-frequency limit ITD for a sound incident at 45 degrees. Error bars on data points represent two standard errors.

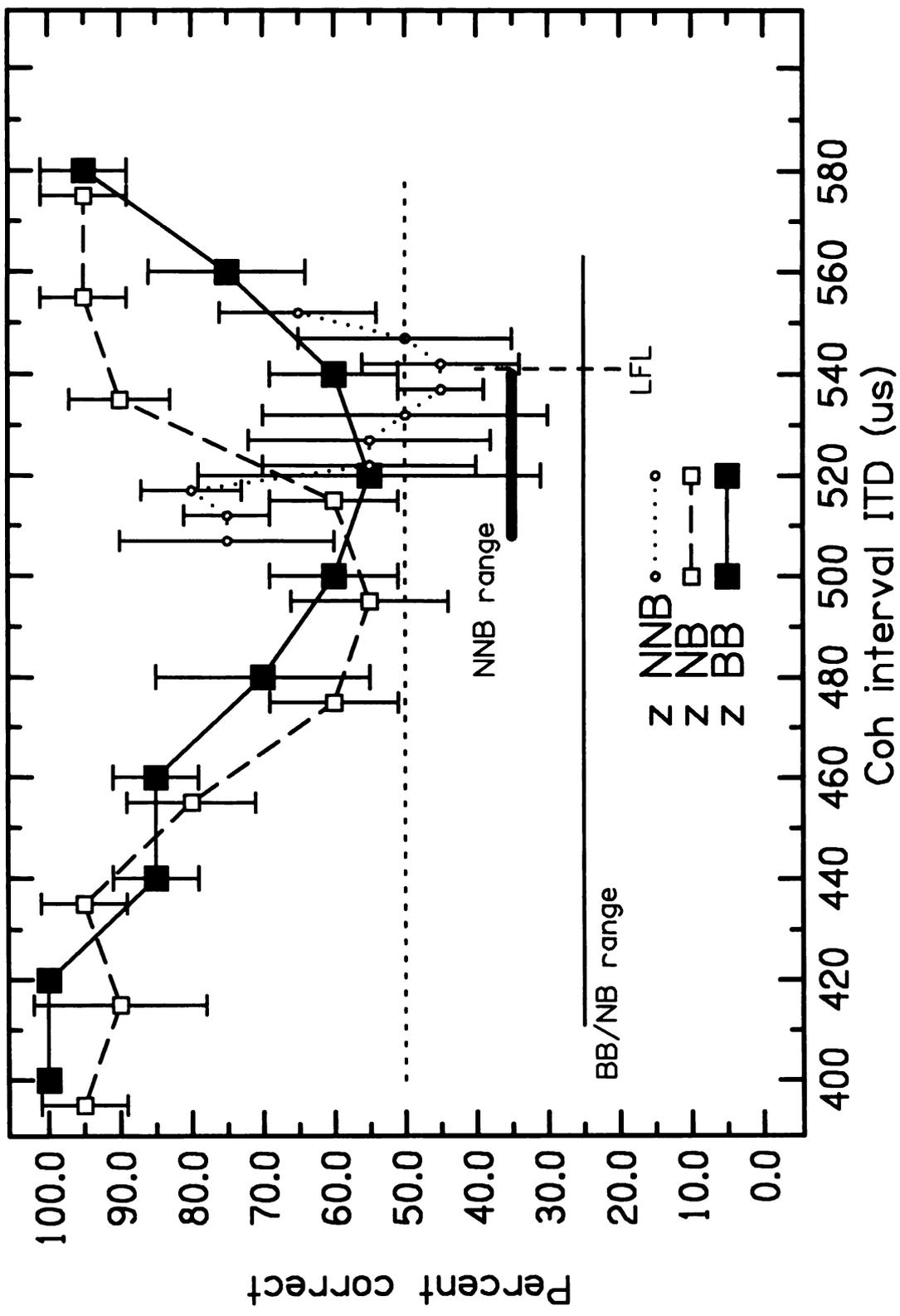


Figure V.10

Figure V.11. Head-related dispersion according to Kuhn's spherical head model, and the inverse. This graph plots the ITD (y-axis) as a function of log frequency for sound incident on a spherical head from a 45-degree angle (solid line) and also includes frequency-dependent ITDs that have an opposite shift ("wrong 45" dashed line). The open circle on the left is the resultant ITD from a low-frequency approximation. The closed square on the right is a high-frequency approximation which is simply two-thirds the ITD value of the open circles. The open triangle further right is the prediction of the Woodworth formula.

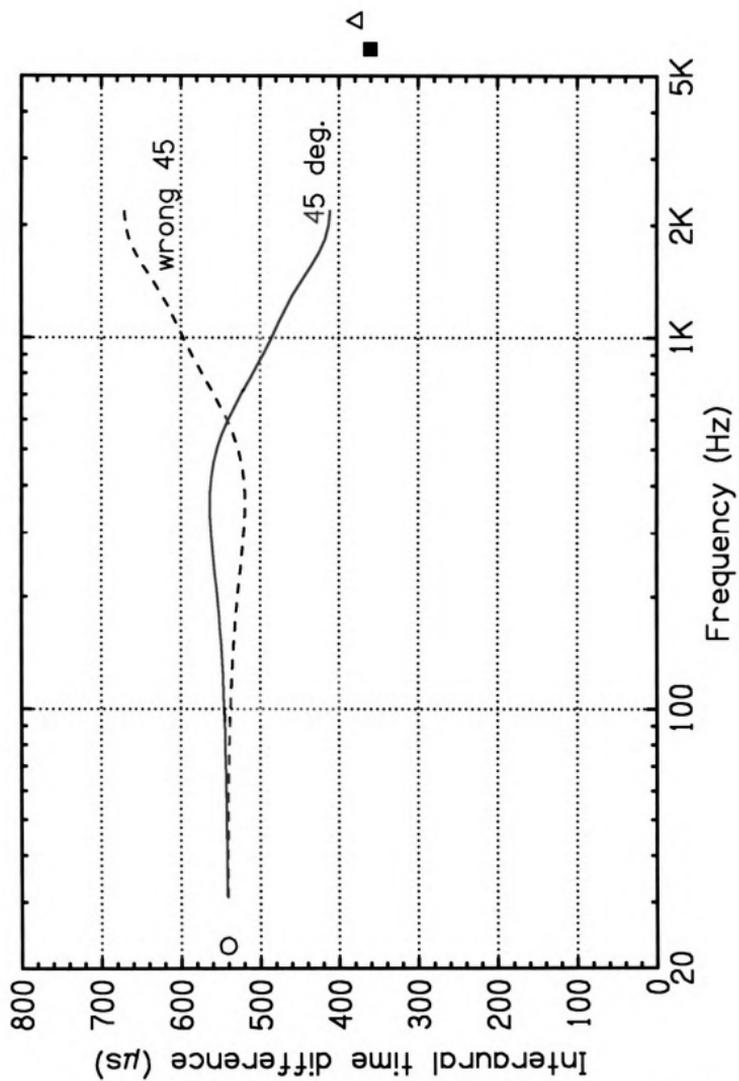


Figure V.11

Figure V.12. Cross-correlation of broadband model head-dispersed noise at a 45-degree angle. The cross-correlation of the experimental 45-degree BB stimulus is plotted versus internal delay τ . The values "maxCC" and "us" at top indicate the peak value of the correlation function (i.e. coherence) and the location of that peak along the delay axis, respectively.

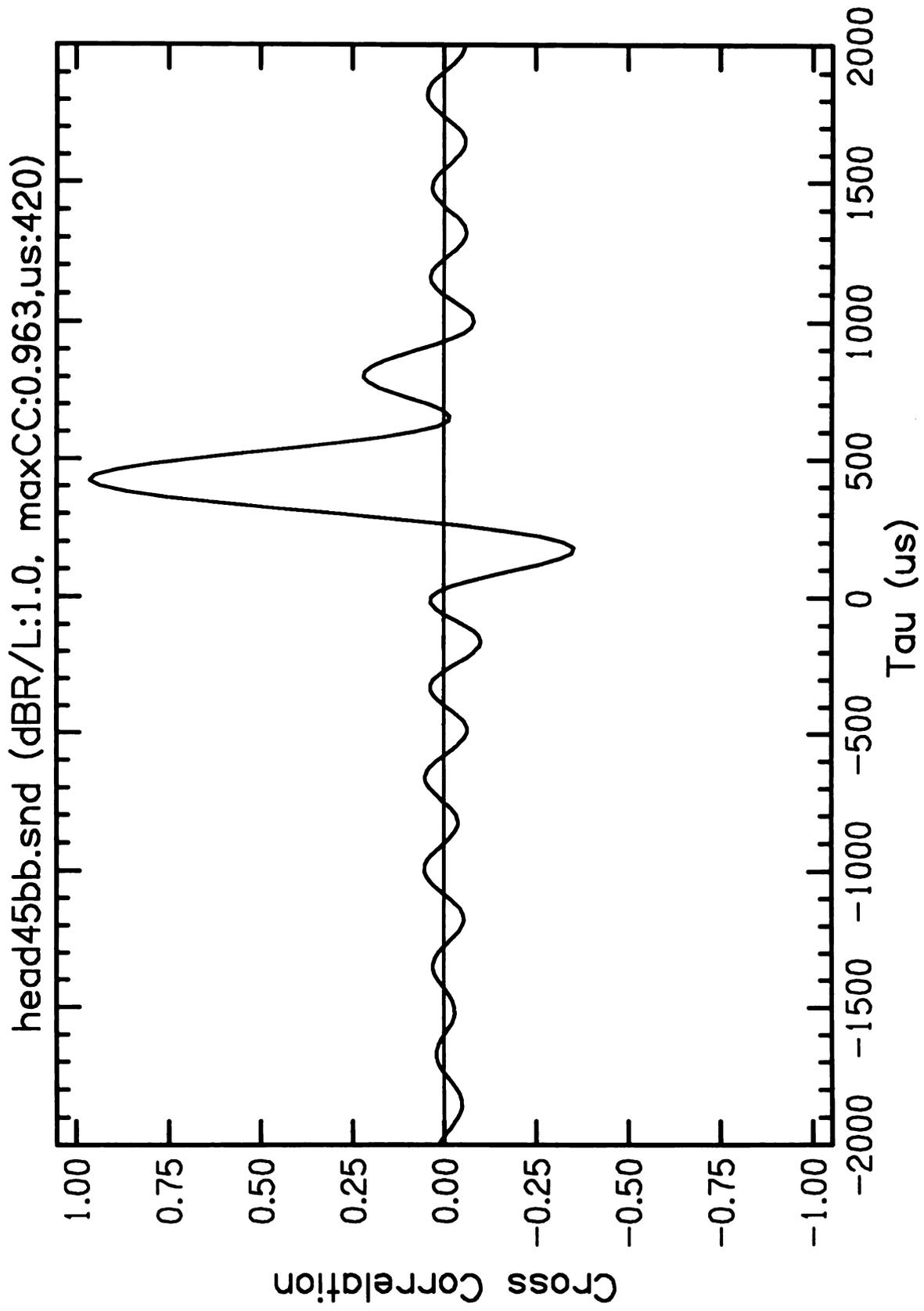


Figure V.12

Figure V.13. Cross-correlation of narrowband model head-dispersed noise at a 45-degree angle. The cross-correlation of the experimental 45-degree NB stimulus is plotted versus internal delay τ . The values "maxCC" and "us" at top indicate the peak value of the correlation function (i.e. coherence) and the location of that peak along the delay axis, respectively.

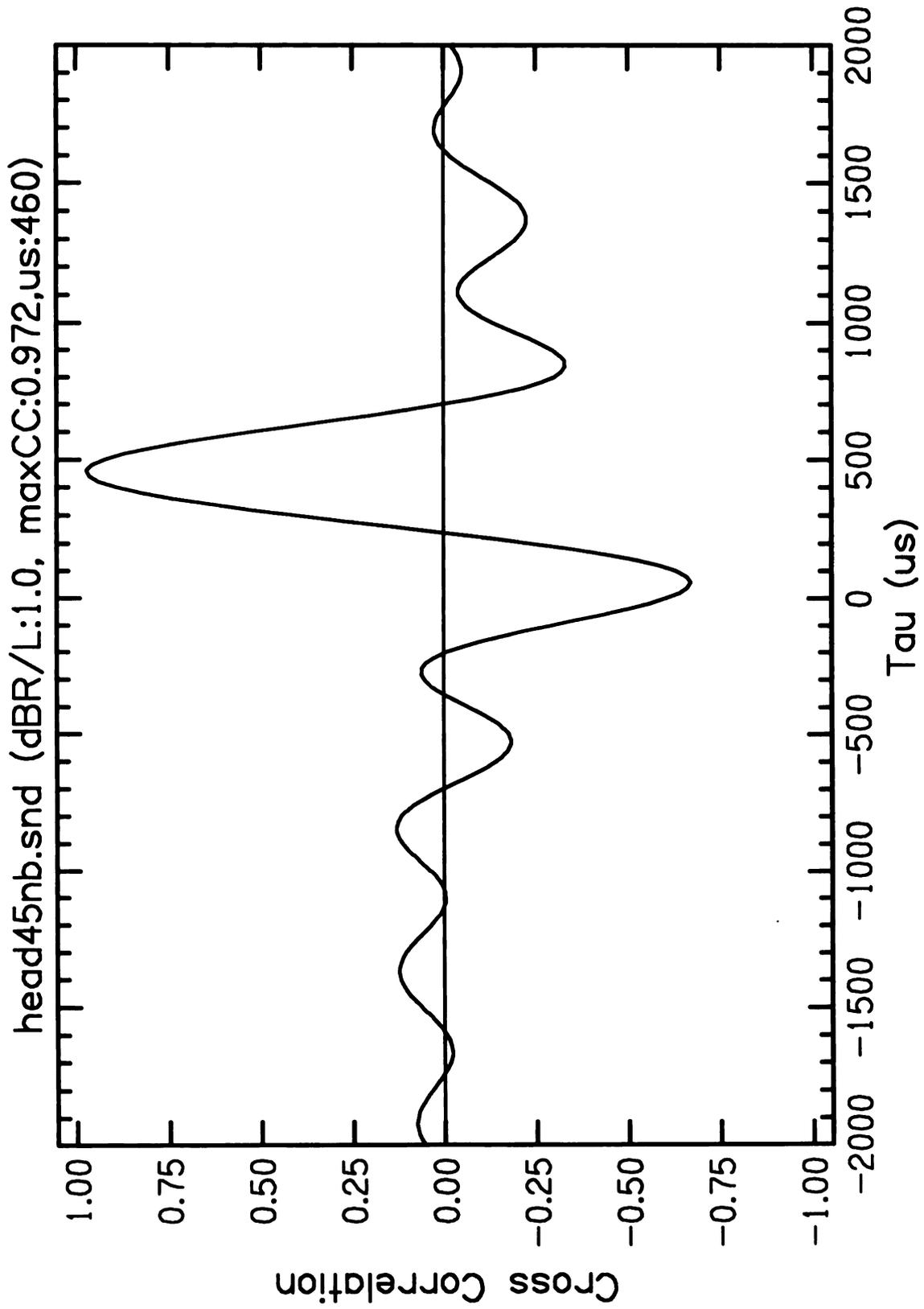


Figure V.13

Figure V.14. Cross-correlation of narrow-narrowband model head-dispersed noise at a 45-degree angle. The cross-correlation of the experimental 45-degree NNB stimulus is plotted versus internal delay τ . The values "maxCC" and "us" at top indicate the peak value of the correlation function (i.e. coherence) and the location of that peak along the delay axis, respectively.

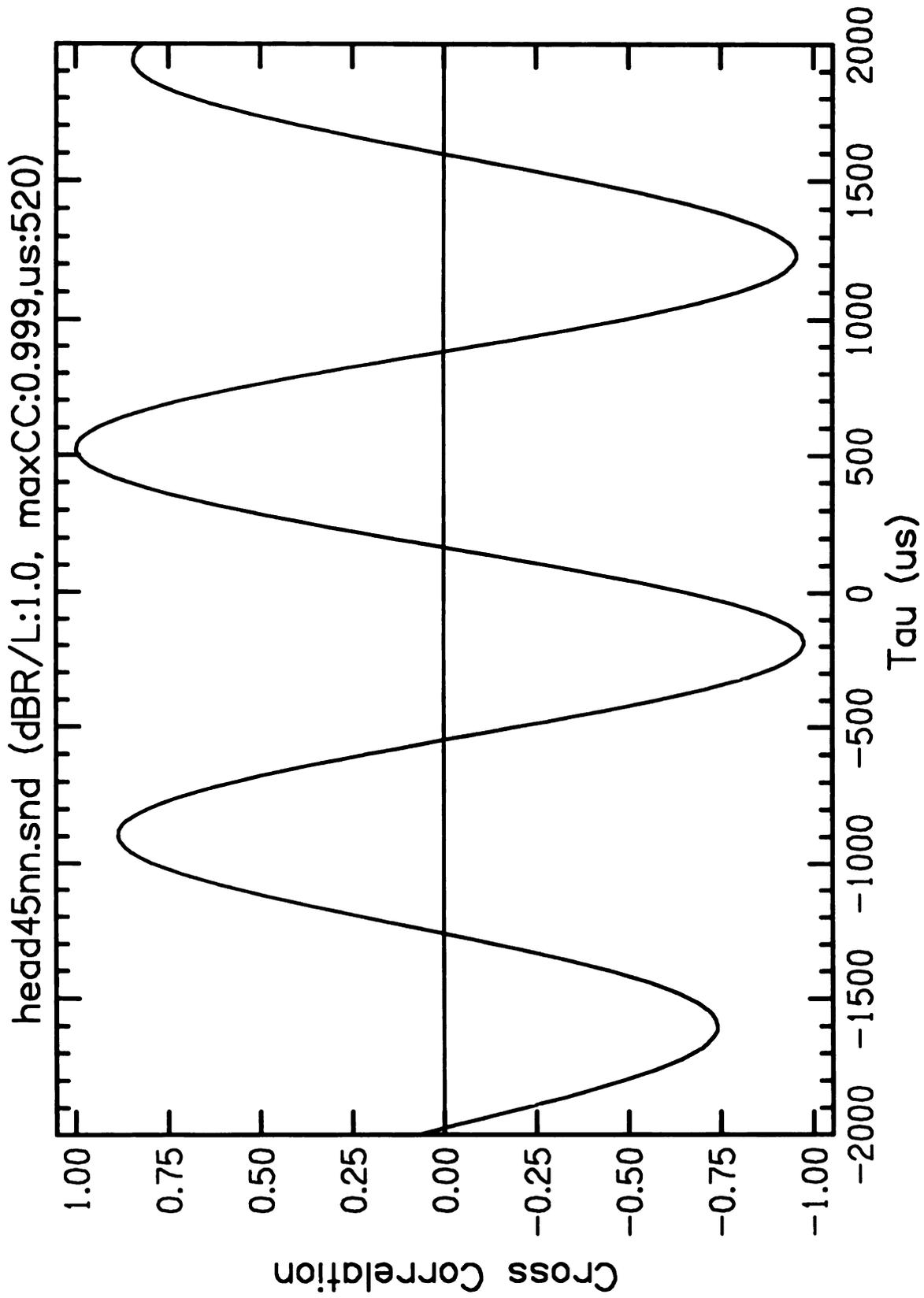


Figure V.14

Figure V.15. Cross-correlation of broadband coherent noise at 520 μs ITD. The cross-correlation of the experimental 520 μs -ITD BB coherent stimulus is plotted versus internal delay Tau. The values "maxCC" and "us" at top indicate the peak value of the correlation function (i.e. coherence) and the location of that peak along the delay axis, respectively.

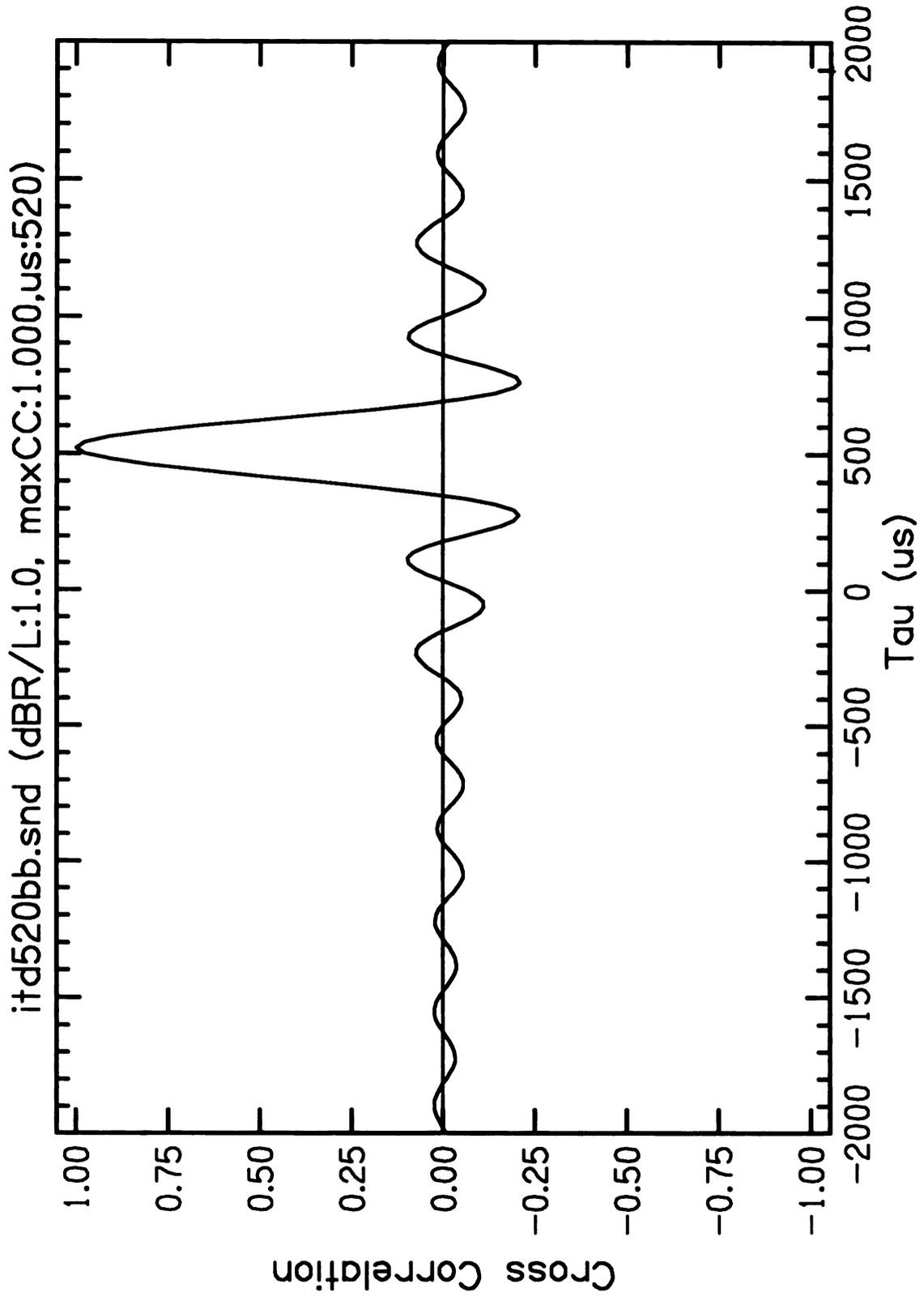


Figure V.15

Figure V.16. Cross-correlation of narrowband coherent noise at 520 μs ITD. The cross-correlation of the experimental 520 μs -ITD NB coherent stimulus is plotted versus internal delay Tau . The values "maxCC" and "us" at top indicate the peak value of the correlation function (i.e. coherence) and the location of that peak along the delay axis, respectively.

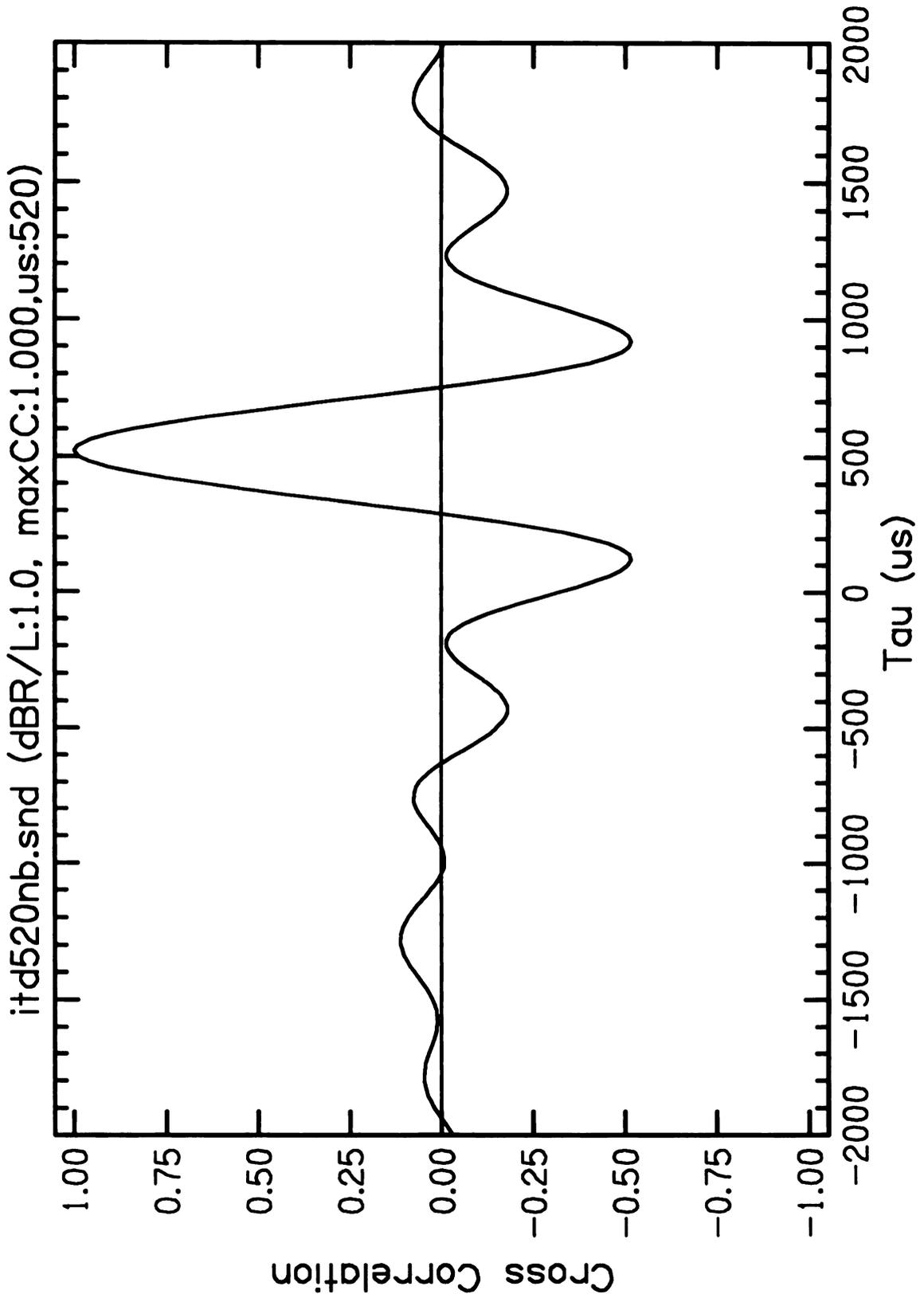


Figure V.16

Figure V.17. Cross-correlation of narrow-narrowband coherent noise at 520 μs ITD. The cross-correlation of the experimental 520 μs -ITD NNB coherent stimulus is plotted versus internal delay Tau. The values "maxCC" and "us" at top indicate the peak value of the correlation function (i.e. coherence) and the location of that peak along the delay axis, respectively.

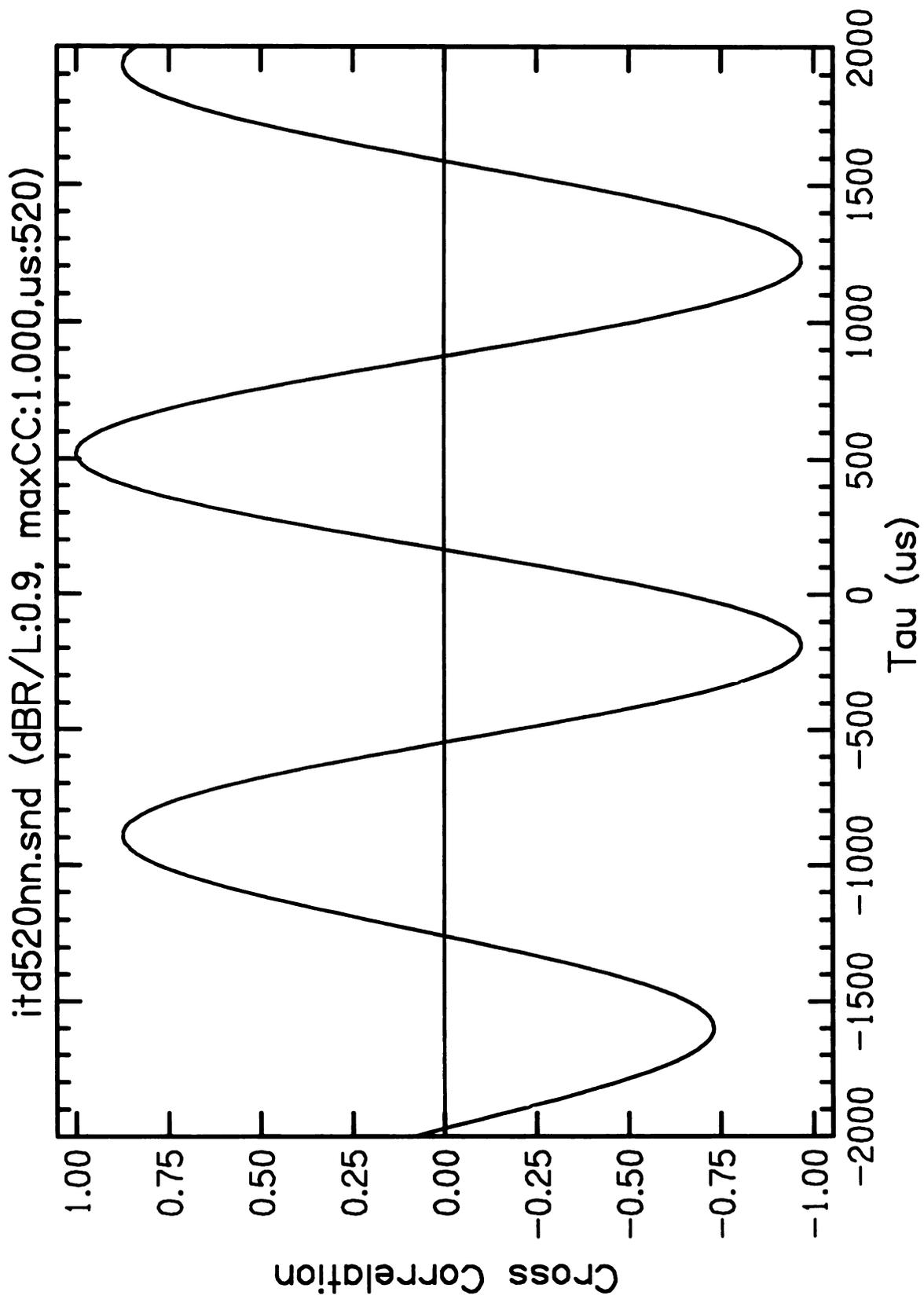


Figure V.17

Figure V.18. Cross-correlation of broadband “wrong-head” noise at a 45-degree angle. The cross-correlation of the experimental 45-degree BB wrong head stimulus is plotted versus internal delay τ . The values “maxCC” and “us” at top indicate the peak value of the correlation function (i.e. coherence) and the location of that peak along the delay axis, respectively.

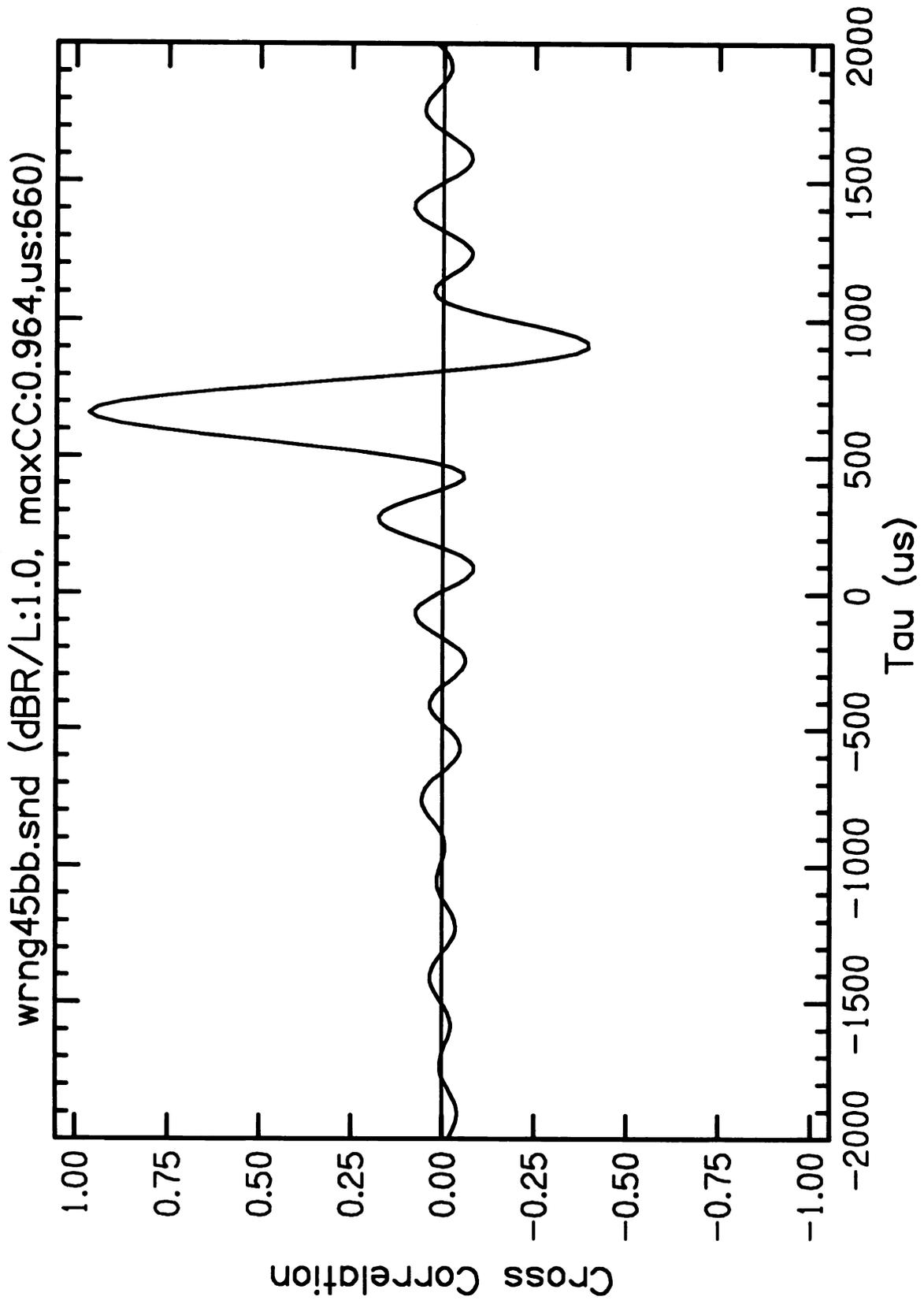


Figure V.18

Figure V.19. Results of Experiment 4: Off-axis coherence discrimination. Threshold coherences necessary to discriminate from perfect coherence are plotted versus log Delay (μs) in the left ear. Data from three listeners appear in this graph. Error bars on data points represent two standard deviations.

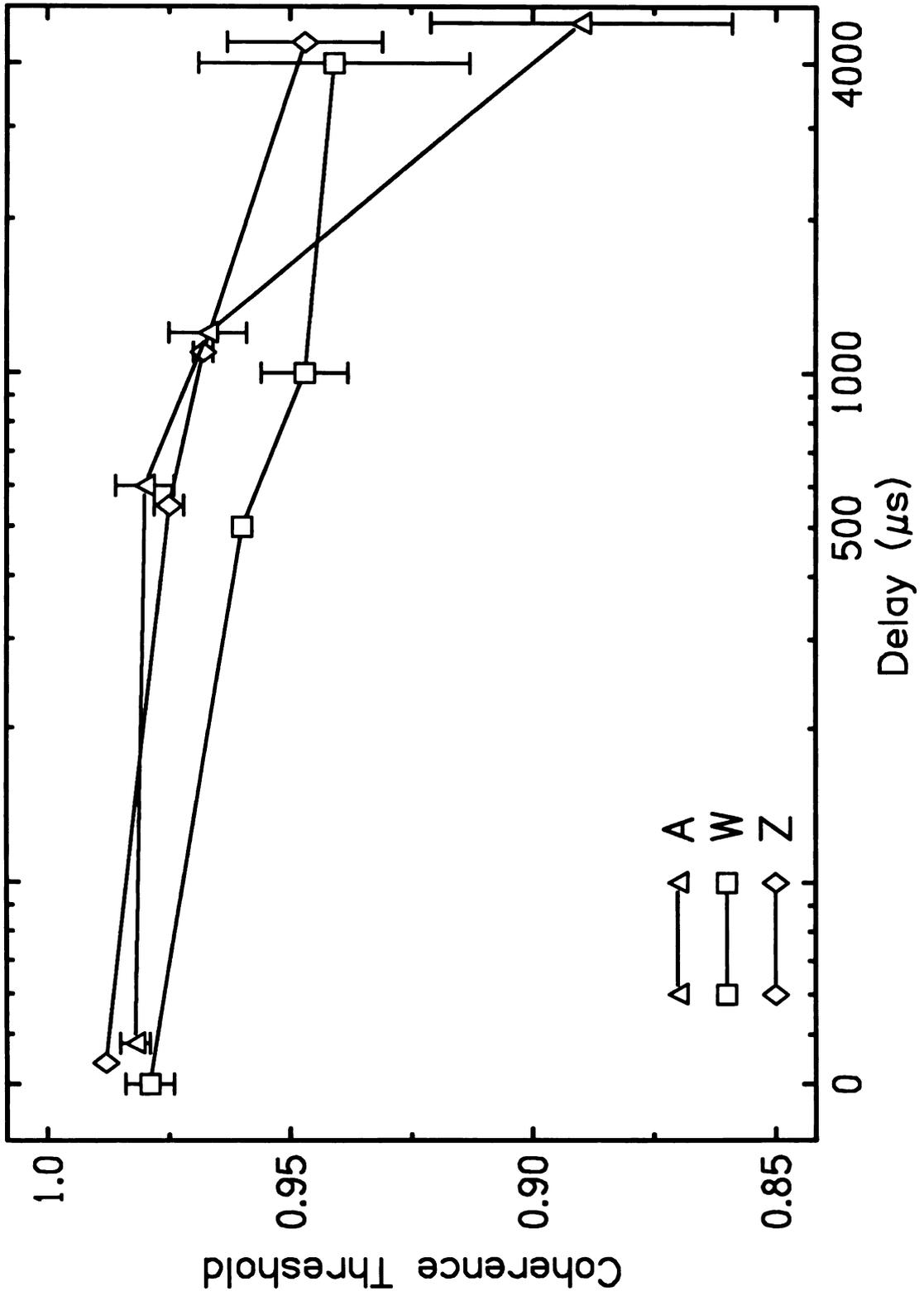


Figure V.19

GENERAL APPENDICES

GENERAL APPENDIX A

The Staircase Method

Unlike the method of constant stimuli, where listeners make several comparisons between two unchanging signals, the adaptive staircase method of Levitt (1977) provides an adapted stimulus. This method allows experimenters to make meaningful comparisons of various conditions with fewer data sets.

The adaptive strategy is quite simple. Following a two-interval forced-choice trial (e.g. a level comparison where the listener is asked to identify the louder interval), the experimental program checks whether the response was correct. If not, the program will alter one aspect of the stimulus, the “staircase variable”, to make the task easier (increasing the level difference between intervals). This change is called moving “up the staircase”. If correct, the program does not change the staircase variable unless it was the third correct response in a row. Then, the program makes the task more difficult (decreasing the level difference), termed moving “down the staircase”. The staircase variable and the amount it changes (the “step size”) depend on the nature of the experiment.

This method of stimulus variation tends to target a difficulty level for which the listener is equally likely to move up or down the staircase. Thus we can find the probability of one correct answer at this level:

$$P(\text{up}) = P(\text{down}) = \frac{1}{2}$$

$$P(\text{down}) = P(\text{correct})^3$$

$$P(\text{correct}) = \left(\frac{1}{2}\right)^{\frac{1}{3}} = 0.7937$$

In this manner, the program identifies the listener's 79.4% correct level for a particular stimulus condition. Making an effective measurement of that level requires a reasonable understanding of the staircase behavior. For instance, the listener may require some time before settling into a region of consistent performance. Even then, the staircase may vary widely. To accurately estimate the level of the staircase variable leading to 79% correct performance, the program keeps track of "turnarounds", or values of the staircase variable where the listener changes staircase direction (from moving up to moving down, and vice versa). At each turnaround, the value of the staircase variable (in our example, intensity) is recorded, and the run continues until an experimenter-set quantity of turnarounds is completed. This requirement means that runs have no preset number of trials. The first few turnarounds are discarded to allow for early variability, and the average of the remainder establishes the 79% correct level at that condition. The (n-1) standard deviation of those turnaround values measures uncertainty in the result.

By varying the stimulus during the run, the staircase method allows experimenters to effectively test multiple conditions at once. The method of constant stimuli would require runs at each value of the staircase variable to estimate the listener's 79% correct level. Even then, the estimation would likely be less accurate than the average taken in a staircase run, which is directly comparable with results from other staircase runs featuring different experimental conditions.

GENERAL APPENDIX B

Signal Detection Theory

The practice of psychoacoustics has been greatly influenced by signal detection theory (TSD). The generality of TSD's approach, combined with a foundation in probability statistics, provides a necessary framework for understanding the implications of how the auditory system responds to stimuli. TSD also allows researchers to compare the results of various experimental tasks on common mathematical ground.

TSD is based on signal-to-noise ratios, where the noise can either be added by the experimenter, or inherent to the system. To apply TSD to psychoacoustics, we refer to the work of Green and Swets (1966). They define a decision variable d' as a signal-to-noise ratio: the magnitude of the auditory system's response to a signal divided by the amount of standard deviation introduced by the response to included noise. Both responses are measured along some internal coordinate that is a quantitative representation of the task at hand. See the top half of Figure A.1, borrowed from Hartmann (1997). This graph shows the response of the auditory system in a two-interval forced-choice (2IFC) task, where the listener is asked to identify the interval that includes the signal. The decision coordinate is r , which could represent interaural coherence (as in an MLD experiment) or a more basic quantity like a coincidence counter output. The leftmost probability density curve (f_N) represents how the auditory system is stimulated by the noise-only interval, with a peak position we define as zero and a standard deviation (due to external or internal noise) of σ . The rightmost probability density curve

(f_{SN}) represents the signal+noise interval, which has shifted the peak along r by an amount μ . We assume that the introduction of the signal has no effect on the size of the uncertainty, so the standard deviation is unchanged.

The listener will choose the interval with a greater r -value, which is determined by samples from each probability function. If $\Delta r = r_{SN} - r_N$ is greater than zero, then the listener will select the correct interval. The probability of this occurrence is determined by the difference (essentially a convolution) of the two probability density curves, establishing a third ($f_D = f_{SN} - f_N$). The new curve is shown in the bottom half of Figure A.1, also from Hartmann (1997). The peak probability for Δr is located at μ , which is expected because the most likely values for r_{SN} and r_N are μ and zero, respectively. Standard deviation σ_D is determined by the nature of convolved distributions, such that $\sigma_D = \sqrt{(\sigma^2 + \sigma^2)} = \sigma\sqrt{2}$ (Hartmann, 1997). The area of the shaded region represents the probability of a positive Δr , which would lead to a correct response. The d-prime for such a task is

$$d' = \frac{\mu}{\sigma\sqrt{2}}$$

To understand how SDT relates to experiments in the lab, we can associate d-prime with a corresponding percentage of correct judgments (P_C). For the above example, the P_C equals the cumulative normal (the shaded area) for all $\Delta r > 0$. If μ increases (raising d-prime), clearly the area in the positive region will increase, leading to a higher P_C . Conversely, if the standard deviation increases (lowering d-prime), then more of the curve will extend below zero, decreasing P_C . For example, the staircase method described in these appendices targets a P_C of 79.4%, corresponding to a d-prime of 1.16. When considering the 2IFC task shown in Figure A.1, a d-prime of one leads to

a P_C of 76%. However, in a task where the listener is given three alternatives (3IFC), that same d' yields a P_C of 63%. Thus, if one were interested in comparing detection results from 2- and 3IFC tasks, one would have to consider those conditions where performance was 76 and 63 percent correct respectively. This method extends to other tasks as well, giving the experimenter a powerful tool in data analysis and evaluation.

References

Green, D.M. and Swets, J.A. *Signal Detection Theory and Psychophysics*, J. Wiley, New York.

Hartmann, W.M. *Signals, Sound and Sensation*. AIP press, Woodbury, New York, 1997. p. 542

Levitt, H. "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467-477 (1971).

Figure A.1. Graphical representation of signal detection theory (TSD). (Hartmann, 1997)

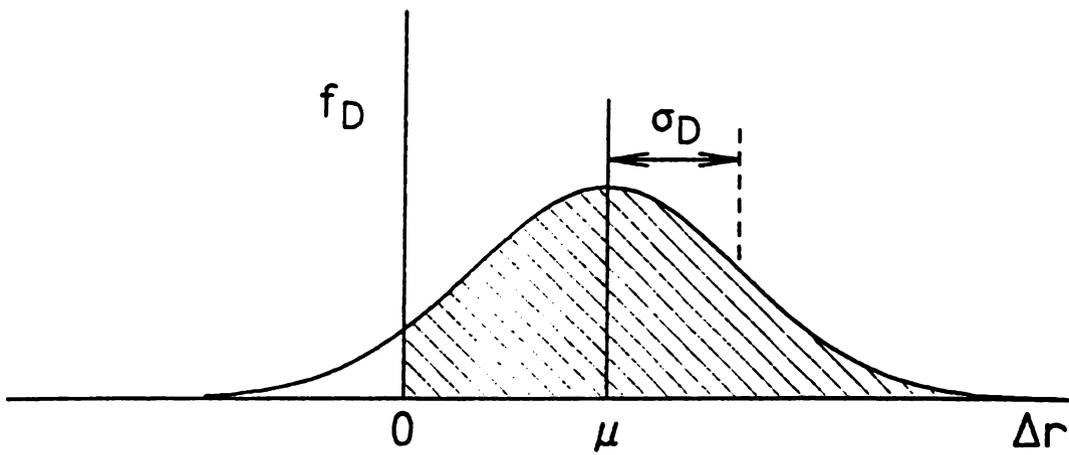
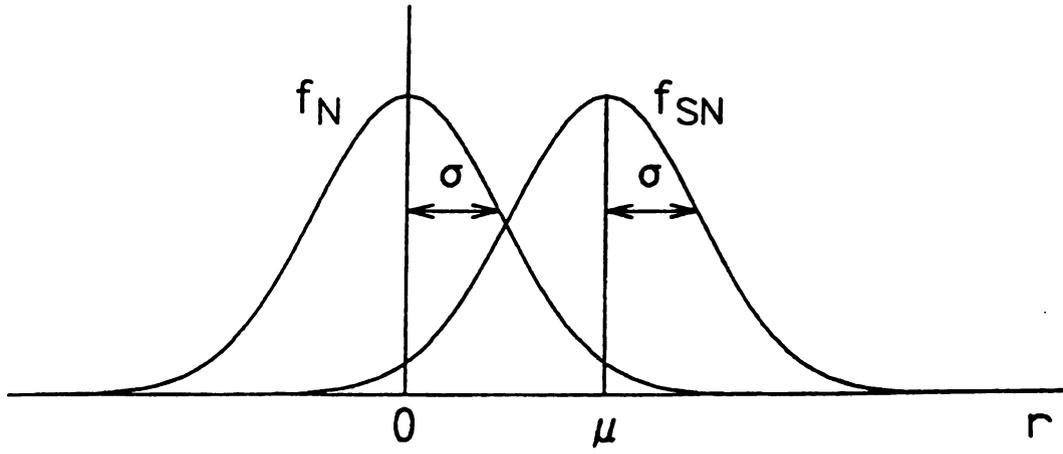


Figure A.1