

LIBRARY
Michigan State
University

This is to certify that the
dissertation entitled

Speech-on-Speech Masking in a Front-Back Dimension and
Analysis of Binaural Parameters in Rooms Using MLS
Methods

presented by

Neil L. Aaronson

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Physics


Major Professor's Signature

28 July 2008
Date

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

ABSTRACT

SPEECH-ON-SPEECH MASKING IN A FRONT-BACK DIMENSION AND ANALYSIS OF BINAURAL PARAMETERS

SPEECH-ON-SPEECH MASKING IN A FRONT-BACK DIMENSION AND ANALYSIS OF BINAURAL PARAMETERS IN ROOMS USING MLS METHODS

Neil L. Aaronson

By

Neil L. Aaronson

This dissertation deals with questions related to the problem of speech sound source localization in rooms. The first part of the dissertation is based on physical measurements made in a room. The second part is based on measurements performed relevant to a specific phenomenon occurring in the front-back dimension and the third part is based on measurements from unreflected speech. Distraction was measured by comparing the target speech, a loudspeaker directly in front of the listener, with a loudspeaker directly behind the listener. Steps were taken to minimize all delays within ± 32 μ s, a release from the target speech occurred. This suggested that the listener's perception of the target speech was based on spatial cues.

A DISSERTATION

Submitted to
Michigan State University

in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Physics

2008

In moving on to physical measurements in rooms, a method was sought for simultaneous measurement of room parameters such as room transfer function (RTF) and reverberation time (RT_{60}), and binaural parameters such as interaural time difference (ITD), interaural level difference (ILD), and the interaural cross-correlation function and coherence. Chapter 2 involves investigations into the usefulness of maximum length sequences (MLS) for these purposes. Comparisons to random telegraph noise (RTN) show that MLS performs better in the measurement of stationary and room transfer functions, RTF, and RT_{60} by an order of magnitude

ABSTRACT

SPEECH-ON-SPEECH MASKING IN A FRONT-BACK DIMENSION AND ANALYSIS OF BINAURAL PARAMETERS IN ROOMS USING MLS METHODS

By

Neil L. Aaronson

This dissertation deals with questions important to the problem of human sound source localization in rooms, starting with perceptual studies and moving on to physical measurements made in rooms. In Chapter 1, a perceptual study is performed relevant to a specific phenomenon — the effect of speech reflections occurring in the front-back dimension and the ability of humans to segregate that from unreflected speech. Distracters were presented from the same source as the target speech, a loudspeaker directly in front of the listener, and also from a loudspeaker directly behind the listener, delayed relative to the front loudspeaker. Steps were taken to minimize the contributions of binaural difference cues. For all delays within ± 32 ms, a release from informational masking of about 2 dB occurred. This suggested that human listeners are able to segregate speech sources based on spatial cues, even with minimal binaural cues.

In moving on to physical measurements in rooms, a method was sought for simultaneous measurement of room characteristics such as impulse response (IR) and reverberation time (RT_{60}), and binaural parameters such as interaural time difference (ITD), interaural level difference (ILD), and the interaural cross-correlation function and coherence. Chapter 2 involves investigations into the usefulness of maximum length sequences (MLS) for these purposes. Comparisons to random telegraph noise (RTN) show that MLS performs better in the measurement of stationary and room transfer functions, IR, and RT_{60} by an order of magnitude

in RMS percent error, even after Wiener filtering and exponential time-domain filtering have improved the accuracy of RTN measurements.

Measurements were taken in real rooms in an effort to understand how the reverberant characteristics of rooms affect binaural parameters important to sound source localization. Chapter 3 deals with interaural coherence, a parameter important for localization and perception of auditory source width. MLS were used to measure waveform and envelope coherences in two rooms for various source distances and 0° azimuth through a head-and-torso simulator (KEMAR). A relationship is sought that relates these two types of coherence, since envelope coherence, while an important quantity, is generally less accessible than waveform coherence. A power law relationship is shown to exist between the two that works well within and across bands, for any source distance, and is robust to reverberant conditions of the room.

Measurements of ITD, ILD, and coherence in rooms give insight into the way rooms affect these parameters, and in turn, the ability of listeners to localize sounds in rooms. Such measurements, along with room properties, are made and analyzed using MLS methods in Chapter 4. It was found that the pinnae cause incoherence for sound sources incident between 30° and 90° . In human listeners, this does not seem to adversely affect performance in lateralization experiments.

The cause of poor coherence in rooms was studied as part of Chapter 4 as well. It was found that rooms affect coherence by introducing variance into the ITD spectra within the bands in which it is measured. A mathematical model to predict the interaural coherence within a band given the standard deviation of the ITD spectrum and the center frequency of the band gives an exponential relationship. This is found to work well in predicting measured coherence given ITD spectrum variance. The pinnae seem to affect the ITD spectrum in a similar way at incident sound angles for which coherence is poor in an anechoic environment.

This work is dedicated to my loving parents, who never stopped

© 2008

supporting me in me.

Neil L. Aaronson

All Rights Reserved

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the many people who have aided and supported me in the course of my graduate studies. I would like to especially thank my adviser, Dr. William M. Hartmann, who has put up with me for longer than anyone but a blood relative should have to. He has gone to epic lengths to support my studies and shape me from a student into a scientist. His patience, calm, and sense of humor have been remarkable, especially that time I made fun of him in front of everyone at the Binsural Basin. I have found in him not just an adviser or a teacher, but a role model.

This work is dedicated to my loving parents, who never stopped supporting and believing in me. I would like to thank Dr. Brad Atland for his gracious support and advice in all aspects of my studies and professional development, especially in generously allowing me to use his lab pretty much as I please. I also thank Dr. Barbara Shinn-Cunningham, Dr. Richard Freyman, and Dr. Leslie Bernstein for their guidance and expertise in our many discussions. I would like to thank Alex Azima for guiding me in my development as a teaching professor.

I would be remiss not to thank my undergraduate mentors, the physics faculty at the College of New Jersey. I especially thank Dr. Ronald Gleeson, Dr. Raymond Pfeiffer, and Dr. Thulasi Wickramasinghe. They are responsible not just for my early development as a scientist, but for convincing me to become a physicist in the first place. Through them I gained a love and fascination for physics that made me want to go on to graduate school.

Students such as myself are often confounded by the intricacies of academic bureaucracy. Too often forgotten are those who guide us through the details we usually forget about later when recounting the trek toward academic accomplishment. I owe special thanks to Mrs. Carole Calu and Mrs. Debbie Simmons for being absolutely amazing secretaries.

Many of my friends **ACKNOWLEDGMENTS** New Jersey, where I

had lived for the first 22.8 years of my life, to Michigan for graduate school was

a d It is a pleasure to acknowledge the many people who have aided and supported me in the course of my graduate studies. I would like to especially thank my adviser, Dr. William M. Hartmann, who has put up with me far longer than anyone but a blood relative should have to. He has gone to epic lengths to support my studies and shape me from a student into a scientist. His patience, calm, and sense of humor have been remarkable, especially that time I made fun of him in front of everyone at the Binaural Bash. I have found in him not just an adviser or a teacher, but a role model. I'd like to thank my dissertation committee members not previ-

ous I would like to thank Dr. Brad Rakerd for his gracious support and advice in all aspects of my studies and professional development, especially in generously allowing me to use his lab pretty much as I please. I also thank Dr. Barbara Shinn-Cunningham, Dr. Richard Freyman, and Dr. Leslie Bernstein for their guidance and expertise in our many discussions. I would like to thank Alex Azima for guiding me in my development as a teaching professor.

I would be remiss not to thank my undergraduate mentors, the physics faculty at the College of New Jersey. I especially thank Dr. Ronald Gleeson, Dr. Raymond Pfeiffer, and Dr. Thulsi Wickramasinghe. They are responsible not just for my early development as a scientist, but for convincing me to become a physicist in the first place. Through them I gained a love and fascination for physics that made me want to go on to graduate school.

Students such as myself are often confounded by the intricacies of academic bureaucracy. Too often forgotten are those who guide us through the details we usually forget about later when recounting the trek toward academic accomplishment. I owe special thanks to Mrs. Carole Calu and Mrs. Debbie Simmons for being absolutely amazing secretaries.

Many of my friends and family know that moving from New Jersey, where I had lived for the first 22.8 years of my life, to Michigan for graduate school was a difficult thing for me to do. I would like to thank those who reminded me that I was missed back home, especially my parents, Rae, Mom-Mom, Joe, Marty, Ed, Rob, Jen, Steve, Jamie, T.J., Ian, Lisa, Kevin, and Kristin. Similarly, I must thank the dear friends I have made while in Michigan who have managed to convince me that the Midwest really isn't so bad, even though it's hard to find good Italian food: Kim, Erin, Carol, Luke, Catherine, Ana, Johanna, Dan, the other Dan, Shannon, Sean, Chris, Josh, and the enigmatic Steve.

Finally, I would like to thank my dissertation committee members not previously mentioned for their helpful advice and genuine interest in my education: Dr. Kirsten Tollefson, Dr. Norman Birge, Dr. S.D. Mahanti, and Dr. Wayne Repko.

1.1	Introduction	15
1.2	1.2.1 Front-Only baselines	19
1.2	1.2.2 Front-Back ADD experiment	20
1.2	1.2.3 Results and Analysis	20
1.3	Experiment 2: Front-Back Presentation with Speech-Shaped Noise	23
1.3.1	Maskers	23
1.3.1	Experimental setup	24
1.3.2	Stimuli	24
1.3.3	Results and analysis	26
1.4	Experiment 3: Front-Front Presentation with Speech Distracters	29
1.4.1	Experimental design	29
1.4.2	Results and discussion	30
1.4.3	Delay-and-add filtering	32
1.5	Experiment 4: Spectral Structure - A Follow-Up to Experiments 1 and 3	33
1.5.1	Experimental design	35
1.5.2	Results and analysis	37
1.6	Summary	38
2	The MLS Method and Applications to Binaural Measurements in Rooms	44
2.1	Introduction	44
2.2	Capabilities and Limitations	47
2.3	Fundamental Equations and Concepts	48
2.3.1	Interaural Level Difference Measurements	48

TABLE OF CONTENTS

2.3.2	Cohere	49
2.3.3	Impulse Response and Transfer Function Measurements	51
2.3.4	Reverberation Time Measurements	53
LIST OF TABLES		xi
LIST OF FIGURES		xiv
0	An Introduction to Spatial Hearing and Speech Segregation	1
1	Release From Informational Masking in the Front-Back Dimension	11
1.1	Introduction	11
1.2	Experiment 1: Front-Back Presentation with Speech Distracters	15
1.2.1	Listeners	15
1.2.2	Anechoic room and experimental layout	15
1.2.3	Listener's chair	16
1.2.4	Loudspeaker alignment	17
1.2.5	Stimuli	17
1.2.6	The Task	18
1.2.7	Front-Only baselines	19
1.2.8	Front-Back ADD experiment	20
1.2.9	Results and Analysis	20
1.3	Experiment 2: Front-Back Presentation with Speech-Shaped Noise	23
1.3.1	Maskers	23
1.3.1	Experimental setup	24
1.3.2	Stimuli	24
1.3.3	Results and analysis	26
1.4	Experiment 3: Front-Front Presentation with Speech Distracters	29
1.4.1	Experimental design	29
1.4.2	Results and discussion	30
1.4.3	Delay-and-add filtering	32
1.5	Experiment 4: Spectral Structure - A Follow-Up to Experiments 1 and 3	33
1.5.1	Experimental design	35
1.5.2	Results and analysis	37
1.6	Summary	38
2	The MLS Method and Applications to Binaural Measurements in Rooms	44
2.1	Introduction	44
2.2	Capabilities and Limitations	47
2.3	Fundamental Equations and Concepts	48
2.3.1	Interaural Level Difference Measurements	48

2.3.2	Coherence and Interaural Time Difference Measurements	49
2.3.3	Impulse Response and Transfer Function Measurements	51
2.3.4	Reverberation Time Measurements	53
2.4	Verification of the value of the MLS method	56
2.5	Experiment 5: Comparison to Random Telegraph Noise	59
2.5.1	Methods	60
2.5.2	Measurements using MLS and RTN	61
2.5.3	Error calculation	62
2.5.4	Results	63
2.6	Experiment 6: Error Reduction Techniques for RTN Measurements	66
2.6.1	Wiener filtering	69
2.6.2	Exponential windowing	70
2.6.3	Conclusions	74
2.7	Experiment 7: MLS and RTN measurements in a real room	75
2.7.1	Methods	76
2.7.2	Results	77
2.7.3	Conclusions	80
3	Experiment 8: Measurements of Envelope and Waveform Interaural Coherences With KEMAR	82
3.1	Computer Simulation of Noises and Coherences	85
3.2	Methods	91
3.3	Analysis of Results Within Bands	93
3.4	Analysis of Data Combined Across Bands	109
3.4.1	Trimming Data Based on Centrality of Waveform Coherences	112
3.5	Conclusions	116
4	Measurement of Binaural Properties in Rooms as a Function of Azimuth	120
4.1	Experiment 9: Measurements of Binaural Properties as a Function of Horizontal Angle in Anechoic Conditions	120
4.1.1	Methods	123
4.1.2	ITD Results	126
4.1.3	ILD Results	132
4.1.4	Coherence Results	137
4.1.5	Incoherence due to Pseudorandom Dispersion	158
4.1.6	Comparison of KEMAR Ears	163
4.2	Further Experiments on the Coherence at 45°	169
4.2.1	Variations on Methods of Coherence Measurement in KEMAR	169
4.2.2	Localizability of Noise Bands at 45° by Human Listeners	175
4.2.3	Measurements in KEMAR and Human Listeners with Probe Microphones	181

4.2.4	Localization of noises with coherence as measured in human listeners	187
4.3	Experiment 10: Measurements of Binaural Parameters and Room Characteristics in a Highly Reverberant Environment	192
4.3.1	Methods	193
4.3.2	Coherence Results	194
4.3.3	Reverberation Times	196
4.3.4	ITD Results	199
4.3.5	ILD Results	203
4.4	Experiment 11: Measurements of Binaural Parameters and Room Characteristics in a Normal Room	205
4.4.1	Methods	208
4.4.2	Coherence Results	208
4.4.3	Reverberation Times	212
4.4.4	ITD Results	214
4.4.5	ILD Results	219
4.5	Summary and Conclusions	222
APPENDICES		225
A	Test for Interaural Differences	225
B	Separated-Source Presentation with Speech Distracters	228
C	Wiener filtering	231
D	Guide to Acronyms	234
BIBLIOGRAPHY		236
3.1	Values of the power parameter π from a fit of the coherence data within each 1/3-octave band to Equation 3.5 for randomly generated Gaussian noises. The bounds about each π value are 95% confidence intervals. As expected, the RMS error decreases as the center frequency, and thus the width of the noise band, increases. The errors indicate a general trend toward a better fit to Equation 3.5 with increasing frequency.	90
3.2	Values of the power parameter π with 95% confidence intervals for the data in each 1/3-octave band measured in room 10B at distances of 0.5 and 1.0 m. These values were found from the best fit of the data within that band to Equation 3.5, and the RMS differences between the data and the best fitting line are given.	103

LIST OF TABLES

1.1	Experiment 4: Percentage of correct responses (plus or minus one standard deviation) for each listener by condition. (a) Front-only baseline condition. (b) and (c) Simulated delay-and-add filtering with a 2 ms delay. (d) Front-Front condition with 2 ms delay - data taken from Experiment 3. Averages and standard deviations for each listener are calculated across three runs. Averages and standard deviations across listeners are shown in the bottom row.	37
2.1	Average percents RMS error, plus and minus one standard deviation, for single measurements of the TF at each order j (corresponding to the length of the signal). The header label of each column indicates the type of signal used (MLS or RTN). Quantities under the $R_w^{(j)}$ correspond to RTN measurements made including <i>a posteriori</i> Wiener filtering of the IR, and quantities under the $R_e^{(j)}$ correspond to RTN measurements made with exponential windowing in time of the IR.	64
2.2	Percents error calculated from the average TF calculated across nine trials using various methods and signals. The header label of each column indicates the type of signal used (MLS or RTN). Quantities under the $R_w^{(j)}$ correspond to RTN measurements made including <i>a posteriori</i> Wiener filtering of the IR, and quantities under the $R_e^{(j)}$ correspond to RTN measurements made with exponential windowing in time of the IR.	65
3.1	Values of the power parameter n from a fit of the coherence data within each 1/3-octave band to Equation 3.5 for randomly generated Gaussian noises. The bounds about each n value are 95% confidence intervals. As expected, the RMS error decreases as the center frequency, and thus the width of the noise band, increases. The errors indicate a general trend toward a better fit to Equation 3.5 with increasing frequency.	90
3.2	Values of the power parameter n with 95% confidence intervals for the data in each 1/3-octave band measured in room 10B at distances of 0.5 and 1.0 m. These values were found from the best fit of the data within that band to Equation 3.5, and the RMS differences between the data and the best fitting line are given.	103

3.3	Values of the power parameter n with 95% confidence intervals for the data in each 1/3-octave band measured in room 10B at distances of 3.0 and 5.0 m. These values were found from the best fit of the data within that band to Equation 3.5, and the RMS differences between the data and the best fitting line are given.	104
3.4	Values of the power parameter n with 95% confidence intervals for the data in each 1/3-octave band measured in a reverberant room at distances of 0.5 and 1.0 m. These values were found from the best fit of the data within that band to Equation 3.5, and the RMS differences between the data and the best fitting line are given.	105
3.5	Values of the power parameter n with 95% confidence intervals for the data in each 1/3-octave band measured in a reverberant room at distances of 3.0 and 5.0 m. These values were found from the best fit of the data within that band to Equation 3.5, and the RMS differences between the data and the best fitting line are given.	106
3.6	Summary of average values of the power parameter n calculated across fits to Equation 3.5 within 1/3-octave bands in both room 10B and the reverberant room at each distance. Values of n within each room that are not significantly different from each other are grouped by brackets.	109
3.7	Values of the power parameter n and the RMS error for the best fit of Equation 3.5 to envelope versus waveform coherence data combined, at each position, across all 1/3-octave bands.	110
3.8	95% confidence intervals about the mean of the power parameter n for different source distances in room 10B and the reverberant room. The parameter n was found by nonlinear regression of Equation 3.5 to data combined across 1/3-octave bands after points with waveform coherences $\gamma_w < \psi$ or $\gamma_w > (1 - \psi)$ were removed. Thus, these values of n and the related RMS errors are determined only by data with abscissae near 0.5. Increasing values of ψ remove increasingly many points near the boundary values of γ_w . In the case of a 1.0 m source distance, most data points are grouped on the high end of the best-fit curve (see Figure 3.14), and a cutoff of $\psi = 0.25$ left too few points in the data set to perform a regression.	115
4.1	Coherence and ITD of 500-Hz noise bands recorded through KEMAR at incident angles of $\pm 45^\circ$ in an anechoic environment.	176

4.2	Just-noticeable differences (JNDs) for coherent and incoherent noises moving to the left or to the right. JNDs for sounds moving to the left are on the left side of the table and JNDs for sounds moving to the right are on the right side of the table. One degree on the table corresponds to an ITD of $5.1 \mu\text{s}$. Incoherent sounds at 45° had an interaural coherence of 0.7303 and those at -45° had a coherence of 0.6865. These JNDs tend to be significantly larger for the incoherent noises than for coherent noises, indicating greater difficulty of listeners to localize incoherent noises under otherwise identical conditions.	182
4.3	Interaural coherences in the 500-Hz band from signals measured with probe microphones in KEMAR and human listeners E, N, and Z.	188
4.4	Just-noticeable differences (JNDs) for coherent and incoherent noises moving to the left or to the right. JNDs for sounds moving to the left are on the left side of the table and JNDs for sounds moving to the right are on the right side of the table. Incoherent sounds at 45° and -45° had coherences for each listener identical to that measured in that listener with probe microphones. The specific coherences for each listener can be found in Table 4.3. There is no significant difference in JNDs for the coherent and incoherent noises.	191
B.1	Separated-source Experiment: Percentage of correct responses (plus or minus one standard deviation) for each listener by condition. (a) Front-Only baseline condition. (b) Front-Only condition with +4 dB target level relative to a single distracter. (c) Target in front and distracters in back. Averages and standard deviations for each listener are calculated across five runs. Averages and standard deviations across listeners are shown in the bottom row.	230
1.6	Individual and average results of Experiment 2, the Front-Back experiment with noise distracters, similar in form to Figure 1.4, as a plot of percent correct versus delay, τ . For listener N, who showed no variation within three runs in performance of the boosted Front-Only condition, a dashed straight line represents the average percent correct.	27
1.7	Arrangement of speakers relative to a listener for Experiment 3. The delay between speakers is represented by τ	30
1.8	Individual and average results of Experiment 3, the Front-Front experiment with speech distracters, similar in form to Figure 1.4, as a plot of percent correct versus delay, τ	31

LIST OF FIGURES

- 1.1 An overview of the extent to which release from informational and energetic masking was found to occur in the horizontal plane across a wide range of delays by Rakerd *et al.* [76]. The release from informational masking is approximately 10 dB in magnitude throughout the range of delays from -32 ms to +32 ms. The magnitude of the release from energetic masking is only about one-third as strong (approximately 4 dB) and is limited to a range of delays from -2 ms to +2 ms. 14
- 1.2 Arrangement of speakers relative to a listener for Experiment 1. The delay between speakers is represented by τ , positive if the distracters from the back loudspeaker lead those from the front and negative if the distracters from the front loudspeaker lead those from the back. 16
- 1.3 A schematic of the full set of CRM sentences that were possible in this study. The listener is instructed to listen for "Laker." The target talker always speaks this call sign. 18
- 1.4 Results of Experiment 1, the Front-Back experiment with speech distracters, for each of the four listeners and the average across listeners as a plot of percent correct versus delay, τ 22
- 1.5 Average amplitude spectra of four different female voices. Each panel represents the average spectrum for one talker's voice, averaged over five utterances. The vertical axis has arbitrary amplitude units, the same for all four. 25
- 1.6 Individual and average results of Experiment 2, the Front-Back experiment with noise distracters, similar in form to Figure 1.4, as a plot of percent correct versus delay, τ . For listener N, who showed no variation within three runs in performance of the boosted Front-Only condition, a dashed straight line represents the average percent correct. 27
- 1.7 Arrangement of speakers relative to a listener for Experiment 3. The delay between speakers is represented by τ 30
- 1.8 Individual and average results of Experiment 3, the Front-Front experiment with speech distracters, similar in form to Figure 1.4, as a plot of percent correct versus delay, τ 31

- 1.9 Solid lines show the theoretical amplitude response of an ideal delay-and-add filter with delay $\tau = 2$ ms. Dips occur at 250 Hz and every additional 500 Hz thereafter. Peaks occur at every integer multiple of 500 Hz. Superimposed in dashed lines is the average amplitude spectrum of one of the four female CRM voices, which is taken from the upper left panel of Figure 1.5. 34
- 1.10 The top graph shows the amplitude response of a filter imitating the first dip in a delay-and-add filter with delay $\tau = 2$ ms. This is a FIR filter of order 336. The bottom graph shows the amplitude response of a filter imitating the second dip in a delay-and-add filter with delay $\tau = 2$ ms. This is a FIR filter of order 392. 36
- 2.1 A linear-feedback shift register with five bits and taps at bits 1, 2, and 4. The bits a_i can take only binary values. An XOR gate performs modulo 2 addition of the output bit, a_0 and a_1 , the result of which becomes one input for the next XOR gate, which "taps" a_2 for its other input. This is then fed as an input to the XOR gate which taps a_4 . The last XOR gate in the sequence feeds its output back to the first bit, a_4 . The resulting output sequence, $s[n]$ will be a MLS of order 5. 45
- 2.2 Cross-correlations between signals measured in some 1/3-octave bands in an anechoic environment between signals measured in the left and right ears of a KEMAR in an anechoic environment. The KEMAR was facing a direction 45° from the direction of the incident sound. A vertical dashed line indicates the location of the peak cross-correlation (the waveform coherence) in each panel, which occurs at the ITD. The value of the ITD in each 1/3-octave band shown is indicated by the number in the lower-right corner of each panel. 52
- 2.3.1 Measurements of the 60-dB reverberation time in the ear space 53
- 2.3 A simple depiction of the MLS measurement method. A MLS, x is played through a loudspeaker in a room with transfer function h . The signal recorded by the receiver is y . The recorded signal y is related to the MLS x via the transfer function by $y = h * x$ 53
- 2.4 The first 400 ms of the integrated impulse decay curve (IIDC) for a small, roughly square room (of dimensions 5.7 m by 4.5 m by 3.5 m high) is shown as a solid line. A dashed line shows a linear trend line fit to the first decay mode of the IIDC. The fit line reaches a level of -60 dB at a time of 323 ms, thusly indicating the 60-dB reverberation time, RT_{60} , of the room. 55

2.5	The impulse response of the electronic 1/3-octave band equalizer as measured by a MLS of order 19.	57
2.6	The transfer function of the electronic 1/3-octave band equalizer as measured by a MLS of order 19. Numbers 1 through 6 count the number of valleys and shoulders over the audible range.	58
2.7	The accepted TF, plotted as a dashed line, and the TF as measured by a MLS of order 9, plotted as a solid line. There is almost no difference between the two at most frequencies, and thus the order-9 MLS measurement obscures the dashed line corresponding to the accepted TF.	67
2.8	The accepted TF, plotted as a dashed line, and the average TF measured by nine RTN of orders as indicated in each panel, plotted as a solid line. The jaggedness of the TF as measured by the RTN tends to obscure the dashed line.	68
2.9	The accepted impulse response, plotted as a continuous line, and the exponential window in time, plotted with a dashed line. The exponential window shown here has a rate parameter of $b = 0.05$ and starts its decay at time $\tau_e = 1.5$ ms.	72
2.10	The TF of a small, roughly square office of dimensions 5.7 m by 4.5 m by 3.5 m high. In each case, the TF was calculated from the average IR calculated over five measurements. The top panel shows the TF as measured by MLS. The middle panel shows the TF as measured by RTN. The final panel shows the TF calculated when the average IR measured by RTN was filtered by a Wiener filter before the TF was calculated.	78
2.11	Measurements of the 60-dB reverberation time in the same space as that referred to in Figure 2.10. Open circles represent times calculated from MLS measurements, exes mark times calculated from RTN measurements, and open diamonds mark times calculated from Wiener-filtered RTN measurements.	79
3.7	Plots of waveform coherence versus envelope coherence as measured through KEMAR in room 1011 at a distance of 5.0 m to the loudspeaker, in a manner similar to that of Figure 3.6	97

3.1	A frequency histogram showing the distribution of the amplitude of samples in an order-18 MLS after the signal was passed through a gammatone filter with center frequency 500 Hz and a 1/3-octave bandwidth. The histogram is enveloped by a Gaussian distribution with zero mean, standard deviation 0.229, and scaled by a factor of 0.067, shown as a dashed solid line. This attests to the Gaussian nature of the filtered MLS.	86
3.2	Plots of envelope coherence versus waveform coherence for simulated Gaussian noise pairs in six different 1/3-octave bands. Each band contains 1001 coherence pairs, each of which is a single data point. A best fitting line to Equation 3.5 is shown as a thick solid line in each plot, though it may be obscured by the data points. The value of the power parameter n for each set of points is shown in each panel.	89
3.3	RMS Error as a function of center frequency. The trend line shows a decrease in RMS Error proportional to $1/f^{0.35}$	91
3.4	Plots of waveform coherence versus envelope coherence as measured through KEMAR in room 10B at a distance of 0.5 m to the loudspeaker. Each panel shows the data measured in a particular 1/3-octave band with center frequency indicated in the top left corner of the panel. A best fit line is shown in each panel, which fits the data within the indicated band to Equation 3.5. The best fit value of the power parameter n for the data in each band is shown in the upper left of each panel. Note that the horizontal and vertical scales differ in each panel.	94
3.5	Plots of waveform coherence versus envelope coherence as measured through KEMAR in room 10B at a distance of 1.0 m to the loudspeaker, in a manner similar to that of Figure 3.4	95
3.6	Plots of waveform coherence versus envelope coherence as measured through KEMAR in room 10B at a distance of 3.0 m to the loudspeaker, in a manner similar to that of Figure 3.4	96
3.7	Plots of waveform coherence versus envelope coherence as measured through KEMAR in room 10B at a distance of 5.0 m to the loudspeaker, in a manner similar to that of Figure 3.4	97

3.8	Plots of waveform coherence versus envelope coherence as measured through KEMAR in a reverberant room at a distance of 0.5 m to the loudspeaker, in a manner similar to that of Figure 3.4	98
3.9	Plots of waveform coherence versus envelope coherence as measured through KEMAR in a reverberant room at a distance of 1.0 m to the loudspeaker, in a manner similar to that of Figure 3.4	99
3.10	Plots of waveform coherence versus envelope coherence as measured through KEMAR in a reverberant room at a distance of 3.0 m to the loudspeaker, in a manner similar to that of Figure 3.4	100
3.11	Plots of waveform coherence versus envelope coherence as measured through KEMAR in a reverberant room at a distance of 5.0 m to the loudspeaker, in a manner similar to that of Figure 3.5	101
3.12	RMS Error as a function of 1/3-octave band center frequency for coherences measured in the reverberant room at a source distance of 3.0 m.	108
3.13	A plot of values of n measured in rooms 10B (above) and the reverberant room (below) as a function of 1/3-octave band center frequency. Different lines are plotted for different source distances. There are no clear trends evident in the trends across either frequency or distance to the source.	110
3.14	Combined envelope versus waveform coherence data across all 1/3-octave bands measured in rooms 10B and the reverberant room at each source distance. Trend lines show the best fit of the data in each panel to Equation 3.5, and the value of the fitting parameter n is shown in the upper-left corner of each panel.	111
3.15	Plots of Equation 3.5 with three different values of the power parameter, n . There is very little difference between the curves for these values of n	113
3.16	Plots of the combined data at a source distance of 5.0 m in room 10B and in the reverberant room. Each panel shows the data remaining after points with waveform coherence $\gamma_w < \psi$ and those with $\gamma_w > (1 - \psi)$ have been removed. The best fit of Equation 3.5 to the remaining data is shown as a solid line and the power parameter n is given in each panel.	114

3.17	The 95% Bound curve (closed circles) gives the absolute error in envelope coherence for which 95% of the simulated coherences deviate by that amount or less from the coherence predicted by Equation 3.6. The Mean curve (open circles) gives the mean absolute error relative to Equation 3.6 for simulated envelope coherences. Both curves are third-order polynomials fitted to the absolute errors of the binned data.	119
4.1	A schematic view of the alignment of the KEMAR head relative to the fixed loudspeaker and some of the angles θ , indicating the direction of the KEMAR nose. When the KEMAR is rotated in place and the loudspeaker remains stationary, the angles of incidence are positive for clockwise rotations and negative for counterclockwise rotations.	124
4.2	A depiction of the locations of the KEMAR and loudspeaker relative to the surroundings of the anechoic room in Positions 1 and 2. In each position, the relative separation of the KEMAR and loudspeaker remains the same, though their positions within the room change.	126
4.3	ITD in milliseconds as a function of angle in 1/3-octave bands, plotted in the same format as Figure 4.11. ITDs for the remaining higher frequency bands are shown in Figure 4.4.	128
4.4	ITD in milliseconds (ms) as a function of angle in 1/3-octave bands, plotted in the same format as Figure 4.12. Closed symbols represent original data before corrections for period errors were made. ITDs for lower frequency bands are shown in Figure 4.3.	129
4.5	Head radii a as a function of frequency, calculated by fitting Kuhn's low-frequency approximation (Equation 4.3) and the Woodworth formula (Equation 4.5) to ITD data in 1/3-octave bands. The results of fitting the Woodworth equation are shown by closed triangles and the results of the low-frequency approximation are represented by open circles. The low-frequency approximation predicts a head radius of about 8.0 cm in the low-frequency bands, and this is in close agreement with the predictions of the Woodworth formula in high-frequency bands.	130

4.6	The linear parameter b as a function of frequency, calculated by fitting Equation 4.6 to ITD data for frequencies at and above 2500 Hz. This parameter describes the relative contribution of the linear term θ in Woodworth's equation (Equation 4.5) when the head radius is fixed to be 7.25 cm.	132
4.7	ITD data from the 8000 Hz 1/3-octave band is plotted across frequency as open circles. A fit of Equation 4.5 to this data, shown by the solid line, is made when the fitting parameter (the head radius) is $a = 8.2$ cm. The dashed line shows a fit to this data of Equation 4.6, in which the head radius is set to $a = 7.25$ cm, and the fitting parameter is found to be $b = 1.22$	133
4.8	ILD in decibels (dB) as a function of angle in 1/3-octave bands, plotted in the same format as Figure 4.11. ILDs for the remaining higher frequency bands are shown in Figure 4.9. In each panel, the vertical scale changes, as shown by axes labels on the left and right sides of the figure.	135
4.9	ILD in decibels (dB) as a function of angle in 1/3-octave bands, plotted in the same format as Figure 4.12. ILDs for lower frequency bands are shown in Figure 4.8. In each panel, the vertical scale changes, as shown by axes labels on the left and right sides of the figure.	136
4.10	Coherence γ as a function of frequency for positive angles in 30° increments. The top panel shows measurements made at Position 1, and the bottom panel shows measurements made at Position 2 in the anechoic environment.	138
4.11	Coherences as a function of angle in 1/3-octave bands for each of two different positions in the same anechoic room. Four frequency bands are shown per plot, with results for position 1 on the left and for position 2 on the right. Shown in this plot are coherences for frequency bands from 80 to 1000 Hz. Filled-in symbols represent original data before corrections for period errors were made. Coherences for the remaining higher frequency bands are shown in Figure 4.12.	139
4.17	A linearly varying ITD over some finite frequency band with center frequency ω_0 and bandwidth 2Δ . The parameter ζ describes the maximum difference between the ITD at any given frequency and the average ITD, T_0	149

- 4.12 Coherences as a function of angle in 1/3-octave bands for each of two different positions in the same anechoic room. Four frequency bands are shown per plot, with results for position 1 on the left and for position 2 on the right. Shown in this plot are coherences for frequency bands from 12500 to 16000 Hz. Filled-in symbols represent original data before corrections for period errors were made. Coherences for the lower frequency bands are shown in Figure 4.11. 140
- 4.13 In open circles, a plot of average difference in coherence between those measured at 0° and 45° across 1/3-octave frequency bands from 160 to 3150 Hz in anechoic conditions. The trend line and its accompanying equation represent a nonlinear least squares regression to the data. 141
- 4.14 Coherences in a 500 Hz 1/3-octave band measured in KEMAR for sound at incident angles of $\theta = \{-45^\circ, 0^\circ, +45^\circ\}$. Symbols represent averages over five measurements and error bars extend one standard deviation in each direction. Square symbols represent the "true" interaural coherence, measured between unaltered signals recorded in the left and right ears. Circles represent coherences when the ILD was made to be constant within the frequency band, and thus are representative of the effect of varying ITDs within the band. Triangles represent coherences when the ITD was made to be constant within the band, and thus are representative of the effect of varying ILDs within the band. 144
- 4.15 The ITD spectra in the 10000 Hz-centered band (left plots) and the 500 Hz-centered band (right plots). The three sets of plots show the variation in the ITDs within the indicated 1/3-octave frequency band when the KEMAR is oriented -45° , 0° , and $+45^\circ$ relative to the direction of the sound source. At $\pm 45^\circ$, the large variability of the ITDs in the 500-Hz band may lead to significant incoherence. 146
- 4.16 A plot of the cross-correlation function as described by Equation 4.11 with a center frequency of 1000 Hz, a 1/3-octave bandwidth, and an ITD of 0.27 ms. The dashed line shows the sinc function enveloping the cosine. 149
- 4.17 A linearly varying ITD over some finite frequency band with center frequency ω_0 and bandwidth 2δ . The parameter ξ describes the maximum difference between the ITD at any given frequency and the average ITD, T_0 149

4.18	Plots of the Fresnel integrals $C(z)$ (solid line) and $S(z)$ (dashed line).	151
4.19	The ITD shift ΔT and the coherence γ for various bands with center frequencies indicated on the vertical axis and flat 1/3-octave bandwidths as a function of the dispersion parameter ξ . Data points for various values of ξ are shown in open symbols (circles for ΔT and diamonds for γ), and best fitting curves to Equations 4.19 and 4.20 are plotted in solid lines. As either ξ or center frequency increases, these equations do not hold, and so no such fit lines were plotted for the highest frequency band. In principle, only one plot is necessary because γ is a function of $\delta \xi$ (see Equation 4.23).	153
4.20	Best fit parameters a , b , and m across band center frequency. These parameters fit ITD shift data and coherence data to equations of the form of Equations 4.19 and 4.20.	154
4.21	The parameter a relating ΔT to the dispersion parameter ξ as a function of the bandwidth parameter β . An approximate relationship between a and β is given by Equation 4.24 to be $a = \beta/3$.	157
4.22	Coherence in various 1/3-octave bands as a function of the standard deviation of the noise added to the ITD of the signals. Best fitting curves corresponding to Equation 4.26 are drawn with the data.	160
4.23	Measurements of the ITD taken at 0° (circles) and 180° (triangles) in an anechoic environment in 24 1/3-octave bands from 80 to 16000 Hz. Measurements were taken at two different positions in the anechoic room, each with the same separation between KEMAR and loudspeaker. Measurements taken at position 1 are indicated by open symbols and are shown in the top panel. Measurements taken at position 2 are indicated by filled symbols and are shown in the bottom panel.	165
4.24	Measurements of the interaural coherence taken at 0° (circles) and 180° (triangles) in an anechoic environment in 24 1/3-octave bands from 80 to 16000 Hz. Measurements were taken at two different positions in the anechoic room, each with the same separation between KEMAR and loudspeaker. Measurements taken at position 1 are shown in the top panel, and measurements taken at position 2 are shown in the bottom.	166
	Thresholds for discrimination are plotted as solid horizontal grid lines. A shift of one sample corresponds to a perceptual shift of approximately one degree.	180

- 4.25 Measurements of the ILD taken at 0° (circles) and 180° (triangles) in an anechoic environment in twenty-four $1/3$ -octave bands from 80 to 16000 Hz. Measurements were taken at two different positions in the anechoic room, each with the same separation between KEMAR and loudspeaker. Measurements taken at position 1 are indicated by open symbols of either type and are located in the top panel. Measurements taken at position 2 are indicated by filled symbols and are located in the bottom panel. 168
- 4.26 Coherences measured in $1/3$ -octave bands in an anechoic environment using several variations of the experimental apparatus to record the signals in both ears. Each variation is labeled with a different letter and is plotted separately with different symbols. The variations are A: measurements through headphone-mounted microphones placed on KEMAR; B: recordings through KEMAR ears but with pinnae removed; C: as variation B, but with cardboard "ear canals;" D: as variation B, but with only the left pinna removed; E: as variation B, but with only the right pinna removed; F: recordings through microphones mounted on a JVC foam dummy head; G: recordings made through KEMAR normally with both pinnae present. 172
- 4.27 Coherences measured in $1/3$ -octave bands in an anechoic environment normally through KEMAR with pinnae in place (variation G, + symbols), and though a KEMAR head, detached from the KEMAR torso (variation H, up-arrow symbols). 174
- 4.28 Percentage of "right" responses as a function of delay for noises at an angle of 45° for four different listeners. Responses corresponding to coherent noises are plotted with filled circles. Responses corresponding to incoherent noises are plotted with open circles. Thresholds for discrimination are plotted as solid horizontal grid lines. A shift of one sample corresponds to a perceptual shift of approximately one degree. 179
- 4.29 Percentage of "right" responses as a function of delay for noises at an angle of -45° for four different listeners. Responses corresponding to coherent noises are plotted with filled circles. Responses corresponding to incoherent noises are plotted with open circles. Thresholds for discrimination are plotted as solid horizontal grid lines. A shift of one sample corresponds to a perceptual shift of approximately one degree. 180

- 4.30 Coherences measured in three human listeners and a KEMAR at incident angles of sound of $\pm 45^\circ$ and 0° . Averages and standard deviations are calculated over three measurements made at each angle. Error bars span one standard deviation in each direction. Error bars smaller than the size of the data points are invisible. 184
- 4.31 Coherences measured at 45° from recordings made through KEMAR Zwislocki couplers and electronics (+ symbols) and through ER-7 probe microphones (open squares). 186
- 4.32 Percentage of "right" responses as a function of delay for noises at an angle of 45° for four different listeners. Noises presented to each listener had interaural coherences equal to that measured in that particular listener's ears in the 500 Hz 1/3-octave band. Responses corresponding to coherent noises are plotted with filled circles. Responses corresponding to incoherent noises are plotted with open circles. Thresholds for discrimination are plotted as solid horizontal grid lines. A shift of one sample corresponds to a perceptual shift of approximately one degree. 189
- 4.33 Percentage of "right" responses as a function of delay for noises at an angle of -45° for four different listeners. Noises presented to each listener had interaural coherences equal to that measured in that particular listener's ears in the 500 Hz 1/3-octave band. Responses corresponding to coherent noises are plotted with filled circles. Responses corresponding to incoherent noises are plotted with open circles. Thresholds for discrimination are plotted as solid horizontal grid lines. A shift of one sample corresponds to a perceptual shift of approximately one degree. 190
- 4.34 Average coherences across all measurements in 1/3-octave bands made in a reverberant room at a distance of 3.05 m from the sound source. Error bars extend one standard deviation in each direction. 195
- 4.35 The ITD spectrum of the 1/3-octave band centered at 1000 Hz, recorded through KEMAR in the highly reverberant environment. The interaural waveform coherence within this band is $\gamma = 0.2453$. The standard deviation in ITD across the band is $\sigma = 251$ ms, which gives a coherence of $\langle \gamma \rangle = 0.2883$ when Equation 4.31 is employed. 197

4.36	60-dB reverberation times, averaged across all measurements from both ears of a KEMAR in 1/3-octave bands, made in a reverberant room. Error bars extend one standard deviation in each direction and indicate excellent agreement across different positions in the room and between the two ears.	198
4.37	ITDs measured between two ears of a KEMAR in a reverberant room. Data points indicate average measured values, with error bars spanning one standard deviation in each direction. Averages are calculated across 30 measurements, made at six different positions in the reverberant room, and standard deviations are calculated across average values from each position. At each position, the KEMAR and sound source were separated by 3.05 m with the KEMAR facing the sound source. Small dots indicate the ITD of period errors that occurred in at least one trial. Dashed lines indicate the theoretical ITD for one, two and three period errors respectively as a function of frequency.	200
4.38	ITDs measured between two ears of a KEMAR in a reverberant room, presented in a manner identical to that of Figure 4.37. ITDs are calculated by finding the lag of the peak in a cross-correlation function between the signals recorded in the left and right ears, with a maximum lag of ± 2 ms.	202
4.39	ILDs measured between two ears of a KEMAR in a reverberant room. Data points indicate average measured values, with error bars spanning one standard deviation in each direction. Averages are calculated across 30 measurements, made at five different positions in the reverberant room, and standard deviations are calculated across average values from each position. At each position, the KEMAR and sound source were separated by 3.05 m with the KEMAR facing the sound source.	203
4.40	ILDs measured in a reverberant room at six different positions. The data points for each position are labeled with different symbols. Error bars extending one standard deviation in each direction are all smaller than the size of the data points, and so are not shown.	204
4.41	Average coherences across all measurements in 1/3-octave bands made in room 10B. Error bars extend one standard deviation in each direction.	209

- 4.42 A plot of waveform coherence γ versus the ratio of direct to total sound intensity in room 10B. Direct sound intensities were measured in an anechoic environment and total intensities were measured in room 10B. The location of each point is found from the coherence and intensities measured in one of nineteen 1/3-octave bands with ISO center frequencies between 160 Hz and 10 kHz. Points from bands with center frequencies less than or equal to 1 kHz are marked as solid circles, and points from bands above 1 kHz are marked as open circles. This plot is generated from data pooled across all frequency bands and measurements taken at several locations in the room. The linear trend line is shown and the associated Pearson product-moment correlation coefficient r is noted. 210
- 4.43 60-dB reverberation times, averaged across all measurements from both ears of a KEMAR in 1/3-octave bands, made in room 10B. Error bars extend one standard deviation in each direction. Small error bars in most bands indicate excellent agreement across the two ears and across different positions in the room. Errors at low frequencies are similar to those seen in Figure 4.36 for the reverberant room. 213
- 4.44 ITDs measured between two ears of a KEMAR in room 10B. Data is presented in a manner like that of Figure 4.37. Averages are indicated by open circles, with error bars spanning one standard deviation in each direction. Small dots indicate the ITD of period errors that occurred in at least one trial. Dashed lines indicate the theoretical ITD for one, two and three period errors respectively as a function of frequency. 215
- 4.45 ITDs measured between two ears of a KEMAR in room 10B, presented in a manner identical to that of Figure 4.44. ITDs are calculated by finding the lag of the peak in a cross-correlation function between the signals recorded in the left and right ears, with a maximum lag of ± 2 ms. 216
- 4.46 Average ITDs (open circles) plus and minus one standard deviation, and sizes of the standard deviations (closed symbols) measured in room 10B. ITDs are measured in a 800 Hz band with a bandwidth as indicated on the abscissa as a fraction of an octave. ITDs approach the expected value for sounds incident at 0° of 0 ms and the variances across measurements decrease as the bandwidth is increased. 218

- 4.47 ILDs measured between two ears of a KEMAR in room 10B. The presentation of data is the same as that used in Figure 4.39. Data points indicate average measured values, with error bars spanning one standard deviation in each direction. This represents the average of the points shown in Figure 4.48, with error bars extending one standard deviation in each direction. The trend in ILDs is similar to that found in the reverberant environment. 220
- 4.48 ILDs measured in room 10B at six different positions. This figure is akin to Figure 4.40 for ILDs measured in the reverberant room. The data points for each position are labeled with different symbols. Error bars smaller than the size of the data points are not shown. 221

and Speech Segregation

The present chapter is meant to serve as a short, informal introduction to the science of sound localization and its role in speech segregation, and an overview of what is covered in this thesis. To encourage a flowing and continuous narrative, it is free of citations, as those are more appropriate and helpful in the context of the following chapters. All the important information contained within this chapter is reiterated, albeit in a more formal manner, in other parts of this thesis, where it is also thoroughly cited. Readers unfamiliar with or new to the science of binaural hearing would benefit from reading this chapter, while those familiar with the field can safely move on to Chapter 1.

In our everyday lives, we experience many different auditory environments and face the challenges of listening and understanding the soundscape around us. A common experience is one in which people must comprehend and attend to speech in an enclosed space (i.e. a room). Often, we are faced with the task of attending to the speech of a particular person while filtering out other speech signals in the room. This situation, often referred to as a "cocktail party," is one that has been studied by the field of psychoacoustics. The narrative that follows

is meant to describe a "cocktail party" environment. It will introduce some of the terminology common in the field of psychoacoustics and describe how our auditory system localizes different kinds of sound.

Imagine that you are just arriving at a swanky cocktail party at the new Martin Dalar exhibit. Only ten minutes late, you are one of the first to arrive. As you wander aimlessly around the room, casually searching for someone you know,

CHAPTER 0

An Introduction to Spatial Hearing

and Speech Segregation

You were able to turn directly to the waiter as soon as he spoke to you though you had not seen him there before. You knew from the sound right where he was. The present chapter is meant to serve as a short, informal introduction to the science of sound localization and its role in speech segregation, and an overview of what is covered in this thesis. To encourage a flowing and continuous narrative, it is free of citations, as those are more appropriate and helpful in the context of the following chapters. All the important information contained within this chapter is reiterated, albeit in a more formal manner, in other parts of this thesis, where it is also thoroughly cited. Readers unfamiliar with or new to the science of binaural hearing would benefit from reading this chapter, while those familiar with the field can safely move on to Chapter 1.

In our everyday lives, we experience many different auditory environments and face the challenges of listening and understanding the soundscape around us. A common experience is one in which people must comprehend and attend to speech in an enclosed space (i.e. a room). Often, we are faced with the task of attending to the speech of a particular person while filtering out other speech signals in the room. This situation, often referred to as a "cocktail party," is one that has been studied by the field of psychoacoustics. The narrative that follows

is meant to describe a "cocktail party" environment. It will introduce some of the terminology common in the field of psychoacoustics and describe how our auditory system localizes different kinds of sound.

Imagine that you are just arriving at a swanky cocktail party at the art museum to celebrate the new Martin Dulak exhibit. Only ten minutes late, you are one of the first to arrive. As you wander aimlessly around the room, casually searching for someone you know, you are surprised by someone asking you a question. "Can I get you something to drink?" Just two meters off to your right stands a waiter. Sounds arriving in your left and right ears are sounds with high coherence. You were able to turn directly to the waiter as soon as he spoke to you though you had not seen him there before. You knew from the sound right where he was. How are you able to do this? Our auditory system uses several sources of information (often referred to as cues) in the task of localization. The waiter was standing off to your right, and so the sound of his voice arrived at your right ear before it arrived at your left. This difference in arrival time is called the *interaural time difference* (ITD). This cue is especially important for low-frequency sounds (below about 2 kHz). Also, because your head is an obstruction for sound arriving at your left ear, the level (or intensity, or amplitude) of the sound will be less at your left ear than it was at your right ear. This difference in level is called the *interaural level difference* (ILD). Your head is a greater barrier as sounds increase in frequency, and so the ILD cue is quite important for high-frequency sounds (above about 2.5 kHz). Also, note that the waiter was only a couple of meters away when he caught your attention, so reflections from the room were not especially distracting. You thank the waiter and order a Manhattan with Maker's Mark. As you are waiting for the waiter to return with your drink, you hear a soft tone ring out from the front of the room and looking

around you, see a horn player warming up near the other side of the room. The player holds the note for a few moments. Apparently, the entertainment for the evening is a brass band.

The horn player is on the other side of the room. Most ITD and ILD cues are unreliable in this case because of reflections in the room. Your auditory system encounters many copies of the same sound. This can lead to a sense that the sound is not localized in one compact point in space, but instead seems spread out — a perception of “auditory source width” (ASW). ASW is related to the interaural waveform coherence, γ_w , which is a physical quantity that describes how similar the signals arriving in your left and right ears are. Sounds with high coherence (near 1) generally seem compact, whereas sounds with lesser coherence tend to sound spread out in space. The reverberant characteristics of rooms can affect the coherence of sounds, as the room does to the horn sound in the example above. The way in which rooms do this is explored as part of Chapter 4 of this thesis.

How then were you able to localize the sound of the horn player at the other side of the room at all? The sound traveling directly from the sound source to your ears (the direct sound) is usually what your auditory system listens to for localization purposes. The direct sound follows the shortest route from the sound source to your ears, and so it arrives first. You also register reflections of the direct sound from the walls, floor, ceiling, and other surfaces within the room, and these come later in time. Reflections are discounted by the auditory for purposes of sound localization. Most of the time, you will not be confused about the location of the sound source because our auditory system suppresses these reflections and pays attention to only the first instance of the sound it hears — the direct sound. This phenomenon is called the “precedence effect” or the “law of the first wavefront.”

The horn player stops warming up his instrument, and as he does, you become aware of a noise you had not noticed before. It is

distant, but you are able to localize it far to your left. You eventually realize that it is the sound of water running out of a faucet from the sink at the bar. You didn't hear the faucet running until the horn stopped playing, which means you didn't hear its onset. Thus, the precedence effect cannot help you. How then were you able to localize the sound of the water at all? Because the sound was broadband (had many frequency components), you were able to still make use of ITD cues in the *envelope* of the sound. This relies on detection of the interaural envelope coherence, γ_e — the coherence between the envelopes of the signals arriving in your left and right ears. The envelope coherence is important if one is to make use of ITD cues above about 1.5 kHz. The relationship between envelope and waveform coherence is the subject of Chapter 3.

The bartender turns off the faucet, and you then notice yet another sound you had not picked up on before. It's some sort of continuous whine, like from a failing old computer monitor or a squeaky motor. It seems to be coming from someplace off to your left. You decide to try to locate it, but as you move off to your left, you become uncertain that the sound really was coming from that direction. Listening again, it seems off to your right now, and maybe somewhere above you. Turning your head in that direction and listening as carefully as you can, you find that you still can't quite pinpoint the source of the noise.

Why can't you localize the whining sound when so far you've been able to localize every other source by some means or another? You did not hear the onset of the sound, so the precedence effect can't help you. The source is distant, so reflections due to the room are prevalent and ITD and ILD cues fail. Also, the source is spectrally sparse (made up of few frequency components), and so envelope cues

fail as well. The noise must spectrally overlap the speech. That means that the noise must have the same frequency components as the speech. From across the room, you hear your name called by a voice you recognize. Turning until you believe you are facing the direction of the voice, you head to a small group of people. Your friend is among them, a fact that you become more certain of as you approach. Your friend welcomes you to the event and introduces you to a financial analyst, an eccentric artist, and a famous opera singer, all at least as swanky as yourself. As you greet them, you take note of the increasing number of people filling the lobby. You enter into an entertaining conversation with this interesting bunch of people. The noise level in the room is increasing, as more people arrive. A few minutes later, the brass band starts playing. You find that you have no trouble maintaining conversation with your companions, though some of the background sounds you could pick out before you can no longer hear. Eventually, you come to realize that the ambient noise is louder than the voices of the people with whom you are speaking, and yet you are still able to understand their speech, and they seem to be able to understand yours. Why do you have no trouble with understanding the conversation even though the brass band is louder? When the brass band started playing, it became harder to hear certain sounds in the room, such as the clinking of glasses and the ongoing series of cellular phone ring tones. The task of understanding the speech around you became a bit more difficult, but was not much more difficult than it was before the band started playing, though no one started talking any louder than they had been. Ambient speech and noise can both distract you from the person you are trying to pay attention to (the "target talker"), but they do so in different ways. In order for ambient noise to interfere with your ability to hear and understand

speech, it first must spectrally overlap the speech. That means that the noise must have power in the same part of the frequency spectrum as the speech. For instance, a noise that only has power in very low frequencies, say between 10 and 100 Hz, is unlikely to interfere with speech because the power in most male speech is above 100 Hz. If instead the noise was broadband and contained power in all the frequencies between, for instance, 100 and 2000 Hz, then its spectrum would overlap that of either male or female speech. In that case, the noise could make the speech impossible to hear. This is known as "energetic masking."

You become engrossed in a conversation with the financial analyst. As you are trying to figure out exactly what it is a financial analyst does, you suddenly become aware of the raised voices of the opera singer and the artist. They are nearly face-to-face, talking at the same time, gesticulating wildly. You try to figure out what they are saying, but have no luck. Eventually, they tire of their argument. The artist turns to your friend to complain about the narrow-minded opera singer, and the opera singer turns to the financial analyst to similarly complain about the idiosyncratic artist. In turn, you are able to listen to the ranting of either the opera singer or artist and are just getting the notion that the argument was about the purity of various art forms and their value in society.

Something similar is happening here to the case of energetic masking because there is likely spectral overlap between the speech of the two, especially if they are of the same gender. If you try to pay attention to just one of them, the task is very difficult. There is something else happening here that is much different, however. While broadband noise will make speech more difficult to understand, these two people talking at the same time makes it very difficult to understand anything at all of what either one of them are saying. In fact, if we wanted to make the noise

completely mask the speech (such that you could not understand what was being said), it would have to be quite a lot louder than the speech. The artist and opera singer, however, can talk at the same level and you won't be able to understand either one of them. This is an example of speech-on-speech masking, of which "informational masking" is a part. The speech of the artist player does an effective job of masking the speech of the opera singer because the masker (the distracting speech) contains information. In fact, it doesn't even have to be information that is meaningful to you; if one of the two talkers were speaking in a language you didn't understand, the task would be nearly as difficult. The effects of speech-on-speech masking is the basis for the studies presented in Chapter 1 of this thesis.

Just as the opera singer and artist are cooling down from their heated debate, a voice from far behind you begins speaking. It is the curator, welcoming everyone and beginning his speech. You are clearly able to identify that the sound was coming from directly in back of you. You turn around and politely listen.

How did you know that the sound was coming from directly behind you? For a source located directly in back of you, the expected ILD and ITD are both zero. This is the same for a source in front of you, or directly above you, or in any way an equal distance from both of your ears. How can you tell the difference between them? Such sound sources lie in an imaginary plane that bisects your body, separating left from right, called the median sagittal plane (MSP). Such sound sources are localizable thanks to the behavior of your pinnae (the part of your ear that sticks out from the side of your face). Your pinnae are asymmetric and rather oddly shaped. Their odd shape allows them to respond differently to sounds based on the location of the sound in the MSP. Sounds located directly in front of you are filtered by the pinnae differently than sounds coming from directly behind you. Pinnae are thought to be especially effective when sounds have power in frequency

components above about 6 kHz. MSP localization is also facilitated by various anatomical differences between the left and right sides of your body such as asymmetries in the shape of your head. Thus, you benefit from a third type of cue: a spectral cue.

The Cocktail Party Effect

In the narrative described above, many people are talking simultaneously, but you are able to carry a conversation with one individual. Even if the sound level of the surrounding din of “distracting speech” is significantly higher than that of the person with whom you are conversing, you have no problem understanding what they are saying (unless the noise becomes very loud or the crowd becomes tightly packed). Your ability to do this is often called the “cocktail party effect.”

The cocktail party effect depends on the ability of a listener to determine where a sound is coming from (localization) and to separate that sound source perceptually from any other distracting sources of speech (segregation). There are other ways for the cocktail party effect to take place as well, all of which depend on some fundamental difference between the distracting and target speech. Spectral, and bandwidth differences may all contribute to the cocktail party effect. The cocktail party effect is manifested even when the differences between the voices we are trying to segregate are due to voice quality, pitch, timbre, and even the logical syntax of the speech (e.g. we can segregate talkers based on what they are talking about). When the opera singer and the artist were talking face-to-face, you became unable to separate their voices by any of these means, and so your ability to understand what they were saying diminished. When their voices were separated in space, even when they were both speaking at the same time, you were able to pay attention to one or the other because you could localize their voices in space, filtering out sounds localized elsewhere.

Acoustical Measurements

So far, many aspects of binaural hearing have been discussed in this introductory chapter. Quantities such as coherence, ILD, and ITD can all be measured by performing calculations on signals recorded at the ears. If we are to understand the ways in which rooms and our anatomy affect sounds, we indeed must take such measurements. Our auditory anatomy analyzes sounds by spectrally breaking the sound up into approximately 1/3-octave bands, which are fed independently to a central processor. We wish to imitate this method of analysis.

The most efficient method of completely characterizing the acoustics of a room is with a broadband noise that has a flat frequency spectrum (i.e. a white noise). One particular class of noises that fit this description are called “maximum length sequences” (MLS). MLS have a perfectly flat frequency spectrum, and we can easily create different MLS that are independent of one another. MLS also have several other properties that are desirable in the course of making acoustical measurements. Details on the generation and properties of MLS, as well as a comparison of MLS to other types of noise is the subject of Chapter 2 of this thesis.

Organization

This thesis is organized into four chapters. Chapter 1 deals with a perceptual experiment involving speech-on-speech masking when both target and distracting speech is located in the median sagittal plane. Chapter 2 treats the issue of acoustical measurements with MLS by qualitatively comparing the performance of MLS and random telegraph noise (RTN) in making room and binaural measurements. Chapter 3 moves on to physical measurements of interaural waveform and envelope coherences, and their relationship in rooms. In Chapter 4, measurements of ITD, ILD, and interaural waveform coherence are taken in rooms alongside of the reverberant characteristics of the rooms, and the ways in which rooms affect those

binaural parameters is studied.

CHAPTER 1

Release From Informational Masking in the Front-Back Dimension

1.1 Introduction

In a room full of talking people, one can focus on and converse effectively with a particular person, despite the large volume of speech going on elsewhere. The effectiveness of this "cocktail party effect" depends on a number of factors, including the differences between the various talkers' voices (such as pitch and loudness), the content of the messages being spoken, and the talkers' locations relative to the listener [21, 26, 72, 97]. When several talkers are near to each other, the speech of any one of them can be hard to understand, but the intelligibility of a single target talker improves if the distracting talkers are separated in space from the target. Thus, spatial separation gives rise to a release from masking of the target speech [26].

Freyman and colleagues achieved a release from masking by creating a shift in the perceived location of the distracting speech [41, 38, 39]. As a baseline condition, a single loudspeaker directly in front of the listener presented the speech of multi-

ple talkers. The listener's task was to attend to a single target talker and to ignore all others, which were the distracters. A second loudspeaker was placed an equal distance from the listener's head, but to the listener's right. This loudspeaker produced a copy of the distracters, shifted forward in time by 4 ms so as to lead the

CHAPTER 1 loudspeaker. Although the addition of this second loudspeaker increased the overall level of the distracters, the intelligibility of the target speech was greatly improved. This experimental design shall be referred to as an

Release From Informational Masking in the Front-Back Dimension

Freyman *et al.* [41, 38] also showed that when the distracting speech was re- such release from masking occurred. This result was interpreted as a fundamental difference between energetic masking, usually thought of as noise of some sort that overwhelms

1.1 Introduction

such release from masking, usually thought to cause confusion between target and masker [61, 64, 92].

In a room full of talking people, one can focus on and converse effectively with a particular person, despite the large volume of speech going on elsewhere. The effectiveness of this "cocktail party effect" depends on a number of factors, including the differences between the various talkers' voices (such as pitch and loudness), the content of the messages being spoken, and the talkers' locations relative to the listener [21, 26, 72, 97]. When several talkers are near to each other, the speech of any one of them can be hard to understand, but the intelligibility of a single target talker improves if the distracting talkers are separated in space from the target. Thus, spatial separation gives rise to a release from masking of the target speech [26].

distracters [41, 23]. Freyman and colleagues achieved a release from masking by creating a shift in the *perceived* location of the distracting speech [41, 38, 39]. As a baseline condition, a single loudspeaker directly in front of the listener presented the speech of multi-

ple talkers. The listener's task was to attend to a single target talker and to ignore all others, which were the distracters. A second loudspeaker was placed an equal distance from the listener's head, but to the listener's right. This loudspeaker produced a copy of the distracters, shifted forward in time by 4 ms so as to lead the distracters in the front loudspeaker. Although the addition of this second loudspeaker increased the overall level of the distracters, the intelligibility of the target speech was greatly improved. This experimental design shall be referred to as an added-delayed-distracter (ADD) experiment.

Freyman *et al.* [41, 39] also showed that when the distracting speech was replaced by distracting noise (i.e. by an energetic masker), no such release from masking occurred. This result was interpreted as a fundamental difference between energetic masking, usually thought of as noise of some sort that overwhelms or suppresses the target, and informational masking, usually thought to cause confusion between target and masker [61, 64, 92].

A delay in which the added distracters lead the distracters presented with the target is defined here as a positive delay. A delay in which the added distracters lag will be defined as negative. Freyman *et al.* [41] found a significant release from masking in an ADD experiment conducted with both positive and negative delays, though with a negative delay (-4 ms) the auditory image of the distracters was near the target, in front. A small perceptual difference in the locations of the target and distracter speech, along with a relatively diffuse auditory image of the distracters due to interaural disparities in the two-loudspeaker presentation apparently provided an adequate binaural basis for differentiating the target talker from the distracters [41, 23].

In studies such as those by Brungart *et al.* [23] and Rakerd *et al.* [76], the ADD experiment of Freyman and colleagues was thought of as a model for a single reflection of distracting speech in a room. This interpretation led to experiments

in which the delay was varied over a wider range of positive and negative values than had been previously explored. This study revealed a release from masking for all positive and negative delays within the 50-ms region delimited by Haas [46], outside of which Haas reported that delayed speech was perceived as an echo. For delays in the region of echoes, the masking release disappeared. A similar experiment employing speech-shaped noises as maskers showed release from energetic masking only for short delays between plus and minus 2 ms, and even then with considerably reduced release. The range and degree of informational and energetic masking release found in the ADD experiment by Rakerd *et al.* [76] are depicted in a general way in Figure 1.1. These results supported the contention that the release from masking in the speech distracter condition is different from a release from energetic masking. They also showed that the masking release found in the ADD paradigm cannot be explained only as a relocation of the distracter caused by the precedence effect [46, 66, 91] because substantial release occurs for a large range of negative values of the delay. For large negative delays, the precedence effect would perceptually shift the distracters to the physical location of the target, providing no perceived spatial separation between the two. It has been suggested that spatial effects other than spatial separation or timbre cues may lead to a release from masking for negative delays [23, 6].

Thus far, the release from masking in ADD experiments has been explained in terms of binaural cues and delay-and-add filtering [22]. In the present study, sound sources are moved to the median sagittal plane, where localization cues are mainly spectral in nature and binaural cues are minimized [15]. The present study investigated the ADD paradigm for a front-back geometry where special steps were taken to minimize binaural cues.

1.2 Experiment 1: Front-Back Presentation with Speech Distracters

Experiment 1 was designed to measure release from masking in an ADD experiment where the sound sources were directly in front or in back of the listener. The goal was to study release from informational masking when binaural cues are reduced as far as possible.

Locations that can be discriminated by release from informational masking differences [15]. These locations are easily confused by normal hearing listeners in broadband localization experiments [6]. However, the use of the front-back geometry will sometimes be limited to a condition that has been used to emphasize the unimportance of release from informational masking in the case where the unimportance of release from informational masking was similar to those used in the present study (see the discussion by Rakerd *et al.* [76]).

1.2.1 Listeners

Four listeners, three male (listeners I, N, and S) and one female (listener K), partic-

Figure 1.1. An overview of the extent to which release from informational and energetic masking was found to occur in the horizontal plane across a wide range of delays by Rakerd *et al.* [76]. The release from informational masking is approximately 10 dB in magnitude throughout the range of delays from -32 ms to +32 ms. The magnitude of the release from energetic masking is only about one-third as strong (approximately 4 dB) and is limited to a range of delays from -2 ms to +2 ms.

Testing took place in an anechoic chamber, 3.8 m wide by 4.3 m long by 2.4 m high (IAC 107840). A listener was seated near the center of the chamber in a special chair, described below. One loudspeaker was placed directly in front of the listener, at ear height, 1.5 meters from the center of the listener's head. Another loudspeaker was placed directly behind, also at ear height and 1.5 meters from the

1.2 Experiment 1: Front-Back Presentation with Speech Distracters

Experiment 1 was designed to measure release from masking in an ADD experiment where the sound sources were directly in front or in back of the listener. The goal was to study release from informational masking when binaural cues are reduced as far as possible in free field. Directly front and back are two locations that can be discriminated by listeners because of spectral differences [15]. These locations are rarely confused by normal hearing listeners in broadband localization experiments [93]. In what follows, the conditions of the front-back geometry will sometimes be referred to as median sagittal plane (MSP) conditions to emphasize the unimportance of interaural differences. Methods used here were similar to those used in the previous study of masking release by Rakerd *et al.* [76].

1.2.1 Listeners

Four listeners, three male (listeners B, N, and S) and one female (listener K), participated in the experiment. Listeners K, N, and S were in their mid-twenties; listener B was 52. All four listeners had normal hearing (pure tone thresholds ≤ 15 dB HL at 0.5, 1, 2, and 4 kHz). Listener N is the author of this dissertation.

1.2.2 Anechoic room and experimental layout

Testing took place in an anechoic chamber, 3.0 m wide by 4.3 m long by 2.4 m high (IAC 107840). A listener was seated near the center of the chamber in a special chair, described below. One loudspeaker was placed directly in front of the listener, at ear height, 1.5 meters from the center of the listener's head. Another loudspeaker was placed directly behind, also at ear height and 1.5 meters from the

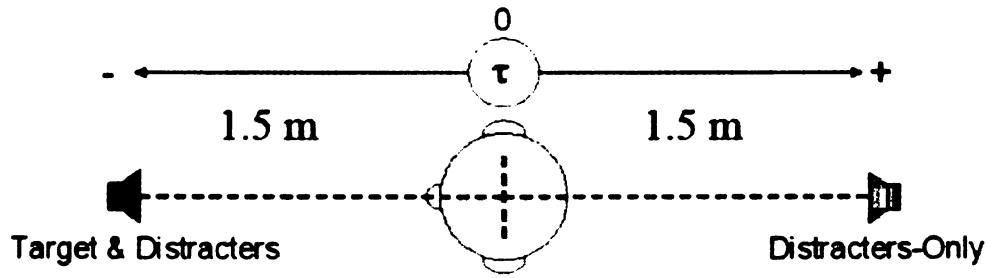


Figure 1.2. Arrangement of speakers relative to a listener for Experiment 1. The delay between speakers is represented by τ , positive if the distracters from the back loudspeaker lead those from the front and negative if the distracters from the front loudspeaker lead those from the back.

head. This layout is shown in Figure 1.2. It is referred to here as the *Front-Back geometry*.

1.2.3 Listener's chair

Rigorous measures were taken to prevent head motion and to ensure that each loudspeaker was equally distant from the listeners' ears. A wooden bite bar, 53 centimeters long, was attached to the chair, running parallel to the back of the chair. This bar was given a dark center line around its circumference at the center of its length and aligned approximately with the center of the chair. To ensure a constant alignment of the head, listeners were instructed to bite lightly on the bar and to maintain contact throughout the test facing the front loudspeaker. Prior to the test, listeners aligned the center line of their top incisors with the center line drawn on the bar using a small hand mirror. Given this procedure, it is reasonable to expect that each listener's ears were symmetrically located with respect to the loudspeakers with a deviation less than 5° . Simple calculations based on the geometry of the setup relate this angle to a corresponding interaural time difference (ITD) of less than $65 \mu\text{s}$.

1.2.4 Loudspeaker alignment

The loudspeaker azimuths were aligned individually with the goal of minimizing interaural differences. Two small microphones were attached to the bite bar, one at each end. A sine tone was presented from the loudspeaker to be aligned. The outputs of the two microphones were observed simultaneously on a dual-channel oscilloscope outside the anechoic room, and the loudspeaker position was adjusted until the oscilloscope traces showed two sine tones with the same phase. A low frequency was used initially, and then successively higher frequencies were used for finer adjustments. After the final adjustment with a tone of 10 kHz, the expected maximum error in angle was less than 0.5° , assuming the phases of the tones could be aligned to within an eighth of a period. This error in angle corresponds to a maximum difference in arrival time of the sound between the left and right ends of the bar of less than $13\ \mu\text{s}$.

The loudspeaker distances were determined with a tape measure. This method had an expected error of less than 1.5 cm, corresponding to a difference in arrival time to the ears between the front and back loudspeakers of less than $44\ \mu\text{s}$. In a delay-and-add filter, this delay corresponds to a dip in the spectrum at over 11 kHz - well above the frequencies used in this experiment. A test of the perceptual effects of misalignment appears in Appendix A. This test indicated that the experiments adequately suppressed binaural difference cues, barring those inherent in the asymmetries of the listeners' heads.

1.2.5 Stimuli

The stimuli used for both targets and distracters were sentences taken from the Coordinate Response Measure (CRM) corpus [19]. For this experiment, each stimulus consisted of three female voices (the target and two distracters) issuing commands

{Talker}:	“Ready	{call sign}	go to	{color}	{number}	now.”
Target	“Ready	Laker	go to	{ Blue Red White Green }	{ One Two Three Four }	now.”
Distracter 1 } Distracter 2 }	“Ready	{ Hopper Ringo Charlie }	go to			now.”

Figure 1.3. A schematic of the full set of CRM sentences that were possible in this study. The listener is instructed to listen for “Laker.” The target talker always speaks this call sign.

that follow the format “Ready {call sign}, go to {color} {number} now.” A chart of call signs, colors, and numbers allowed in these experiments is given in Figure 1.3. The target talker always used the call sign “Laker.”

The voices of the three talkers were randomly chosen from among the four female voices available in the CRM. In any given stimulus, no two talkers shared any of the attributes of call sign, color, number, or individual female voice. With four colors and four numbers, the chance of guessing correctly becomes 1 in 16, or approximately 6 %.

1.2.6 The Task

Listeners were instructed to listen for the call sign “Laker” on each trial and to determine the color/number combination in the associated sentence. A stimulus with the “Laker” call sign was presented from the front on every trial. An LCD computer monitor was mounted below the front loudspeaker. To help minimize any head motion, the listeners used a wireless gyroscopic mouse to control a pointer on the monitor without the need for a mouse pad or other surface. The listeners responded to each stimulus by clicking on the appropriately numbered button within the field of the appropriate color on the monitor screen. A response

was considered correct only if the selected number and the color were both correct.

A single run consisted of five practice trials without feedback followed by 30 test trials. The listeners went through three runs for each experimental condition, with the order of runs randomized differently for each subject.

1.2.7 Front-Only baselines

In a condition referred to here as the *Front-Only baseline* (FO), the target and the distracters were presented exclusively from the front loudspeaker, with the level of each talker's speech fixed at 65 dB SPL. Thus, the level of the target was 0 dB relative to the level of each distracter. For a second reference point, the Front-Only baseline experiment was repeated with the level of the target talker raised to +4 dB relative to the FO condition. This +4 dB signal-to-noise ratio (SNR) reference (the FO+4 dB condition) provided a way to express performance changes using a dB scale.

As this experiment was designed to search for an effect over a wide range of delays, it was desired that all listeners start with similar baseline performance, based on the expectation that similar performance would lead to comparably useful results over a wide range of delays for all listeners. In a pilot test, it was found that three of the four listeners performed the baseline test near 30 % correct, which was well above chance (6 %). Listener K also performed above chance, but less well than the other listeners (~ 15 % correct). The difference in baseline performance was eliminated by increasing the target level by 2 dB for listener K. Data for listener K below reflect this 2 dB increase. For listener K, the condition where the target is at 67 dB and each distracter is at 65 dB is the "0 dB SNR" condition, and other values of SNR will be referenced to this baseline. For instance, the +4 dB condition corresponds to a target at a level of 71 dB and each distracter at 65 dB for listener K. Thus, values of SNR will be reported relative to the initial level of the

target, which is 65 dB for listeners B, N, and S; and 67 dB for listener K.

1.2.8 Front-Back ADD experiment

For Front-Back ADD tests, the target and the distracters were presented from the front loudspeaker as for the Front-Only baseline tests. In addition, the distracters were presented from the back loudspeaker using the same level as in front (65 dB). There was a delay, τ , between the front and back distracters. The delay was varied over a wide range over the course of the experiment. The set of delays employed was: $\tau = \{\pm 32, \pm 8, \pm 2, \pm 0.5, 0\}$ ms. Positive delays indicate that the back loudspeaker (distracters only) led the front. Negative delays indicate that it lagged. Zero delay corresponds to synchrony in the presentation of the distracters from the two loudspeakers.

1.2.9 Results and Analysis

Figure 1.4 shows the results of the Front-Back experiment for each listener, with the average across listeners given in the bottom panel. For the individual subjects, percent correct scores, averaged over three runs of 30 trials each, are given as a function of the delay, τ . Error bars represent the standard deviation over runs and are one standard deviation wide in each direction. A diagonally-hatched rectangular stripe near the bottom of each panel shows the subject's average results for the Front-Only (single speaker) geometry, where the level of the target equals the level of each distracter voice (0 dB SNR). The width of the stripe represents a 95% confidence interval about the average baseline score. Another stripe (drawn with vertical hatch lines) is given for the Front-Only test in which SNR was set to +4 dB (see right-hand axis of the figure). The bottom panel of Figure 1.4 shows percent correct scores averaged across the four subjects. In that panel, error bars and the

95% confidence intervals for Front-Only tests are based on the standard deviation over subjects.

Analysis of variance (ANOVA) showed that Front-Back scores for all τ , $-32 \leq \tau \leq +32$ ms, were significantly greater than the Front-Only baseline score ($p < 0.05$) except for $\tau = +8$ ms, where the score approached, but did not reach, significance ($p = 0.08$). Thus, there was strong evidence of release from masking in the Front-Back geometry for delays spanning the full range of delays tested.

For each listener, the measured difference in percent correct scores between the 0-dB and 4-dB Front-Only reference conditions corresponded to a release of 4 dB (Section G). These benchmark values were used to estimate the amount of unmasking in decibels on Front-Back tests by linear interpolation or extrapolation, as shown on the right-hand axes on Figure 1.4. The delay times of $\tau = \{+2, -2\}$ ms showed the greatest release. Listeners displayed, on average, a release of 3.5 dB for $\tau = 2$ ms, and a release of about 2.5 dB for $\tau = -2$ ms. Measurable release of at least 1 dB extended out to $\tau = \pm 32$ ms.

These results agree with the experiments by Rakerd *et al.* [76] for sources in the horizontal plane, which found masking release for delays as long as 32 ms, but not for 64 ms. Figure 1.4 shows that the release in the front-back ADD experiment is smaller at 32 ms than at shorter delays. Therefore, it seems reasonable to conclude that about the same range of delays elicits a release from masking in both the horizontal and the median sagittal planes. The 50-ms speech echo boundary found by Haas [46], which Rakerd *et al.* held responsible for the loss of masking release for long delays, is apparently insensitive to the source geometry. However, at every value of the delay, the average release from masking is smaller in the MSP than in the horizontal plane. In the horizontal plane with two distracters, a maximum release of 11 dB was found, compared to about 4 dB in the present experiment in the MSP.

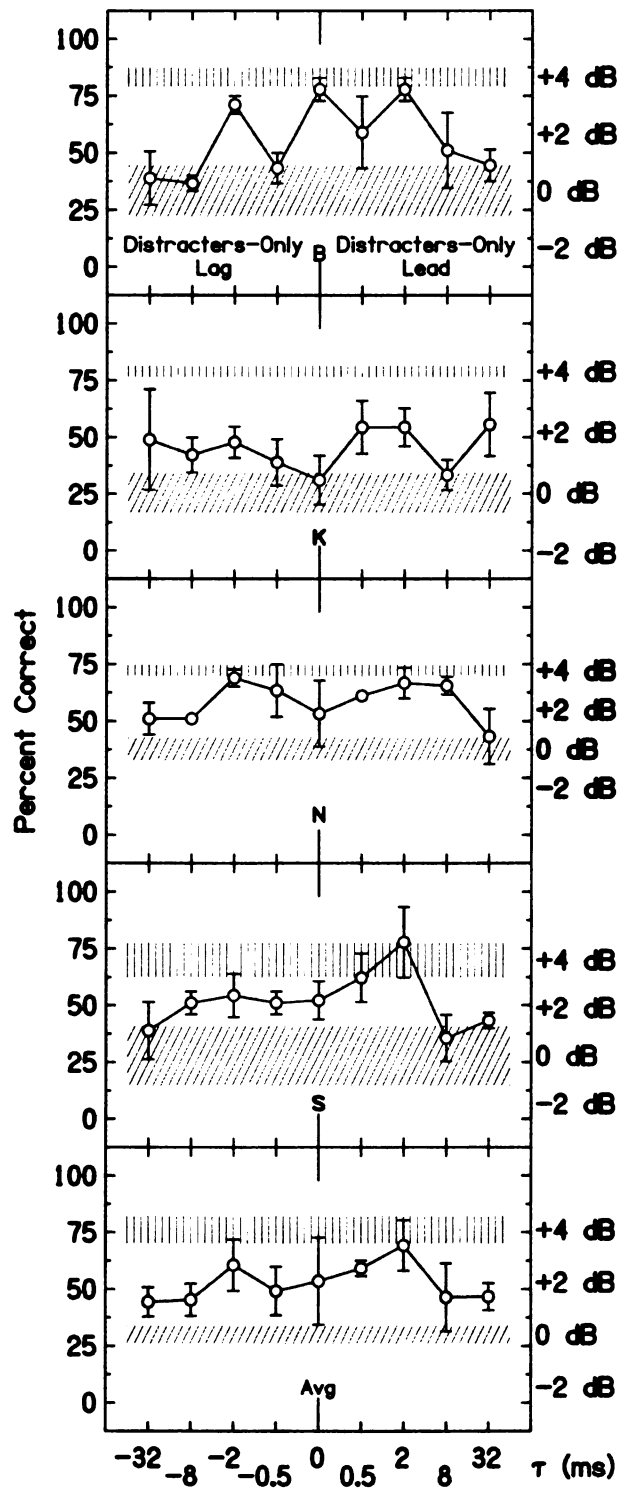


Figure 1.4. Results of Experiment 1, the Front-Back experiment with speech distracters, for each of the four listeners and the average across listeners as a plot of percent correct versus delay, τ .

It is likely that the release from masking seen in this experiment is aided by the ability of listeners to localize sounds in the MSP. Listeners use spectral structure in various frequency bands to localize sounds in the MSP [16, 15]. Roffler and Butler [77] showed that effective localization in the MSP is assisted by broadband stimuli with components above 8000 Hz, and fails for such stimuli without components above 2000 Hz. Because the CRM stimuli are low-pass filtered at 8000 Hz, energy in the 2000–8000 Hz range is present to contribute to front-back localization. A brief follow-up experiment, in which speech distracters are presented in the back loudspeaker only while the target remains in the front, appears in Appendix B and seems to corroborate these conclusions.

1.3 Experiment 2: Front-Back Presentation with Speech-Shaped Noise Maskers

The masking of a signal such as speech by continuous noise, where the spectral content of the noise overlaps that of the speech, is referred to as energetic masking (EM). This type of masking is mainly attributed to the spectral overlap of a target signal with distracting noise. In the masking of speech on speech, energetic masking occurs because of spectral overlap of the target signal and distracting speech, resulting in competition between the target and masker in the periphery of the auditory system. By contrast, informational masking (IM) does not necessarily involve the overlap of signals in the auditory periphery [61]. IM is thus often equated with central masking [35]. IM in competing speech signals is the result of the difficulty a listener experiences in trying to distinguish target speech from distracting speech. EM may be an important component of the masking of speech by speech, and therefore it is studied in this experiment.

Previous ADD studies examining masking release in the horizontal plane have

found that release from EM occurs for very brief delays ($-0.5 \text{ ms} \leq \tau \leq +0.5 \text{ ms}$), and that EM release becomes negligible for longer delays [41, 38, 23, 76]. In the horizontal plane, Rakerd *et al.* [76] found an average release from EM of 2 dB in the very brief delay range. The present study examined the delay dependence of the EM release in the MSP, specifically for the Front-Back geometry. Following previous studies, maskers used to test for EM release were speech-spectrum shaped noises, with spectra matched to those of speech maskers.

1.3.1 Experimental setup

To compare the unmasking shown in Experiment 1 with energetic unmasking, the same ADD experiment was performed, but with continuous speech-spectrum-shaped noise distracters rather than speech distracters. The target speech stimuli remained unchanged from Experiment 1 as did all the other experimental features. The four listeners who participated in this experiment were the same as in Experiment 1.

1.3.2 Stimuli

All distracters from the CRM Corpus (i.e. those voices which spoke call signs other than “Laker”) were modified to derive equivalent speech-shaped noise samples, forming a “Noise Corpus.” To do this, a discrete Fourier transform was applied to each individual speech file in its entirety. Then, each complex spectral component was multiplied by $e^{i\phi}$, where ϕ was a random variable uniformly distributed on the range $[-\pi, \pi]$. Thus, the phases of the frequency spectrum were randomized while the amplitudes remained unchanged. An inverse Fourier transform converted the modified spectrum back to the time domain. All listeners agreed that the result sounded like a swarm of bees, at roughly the same pitch as the origi-

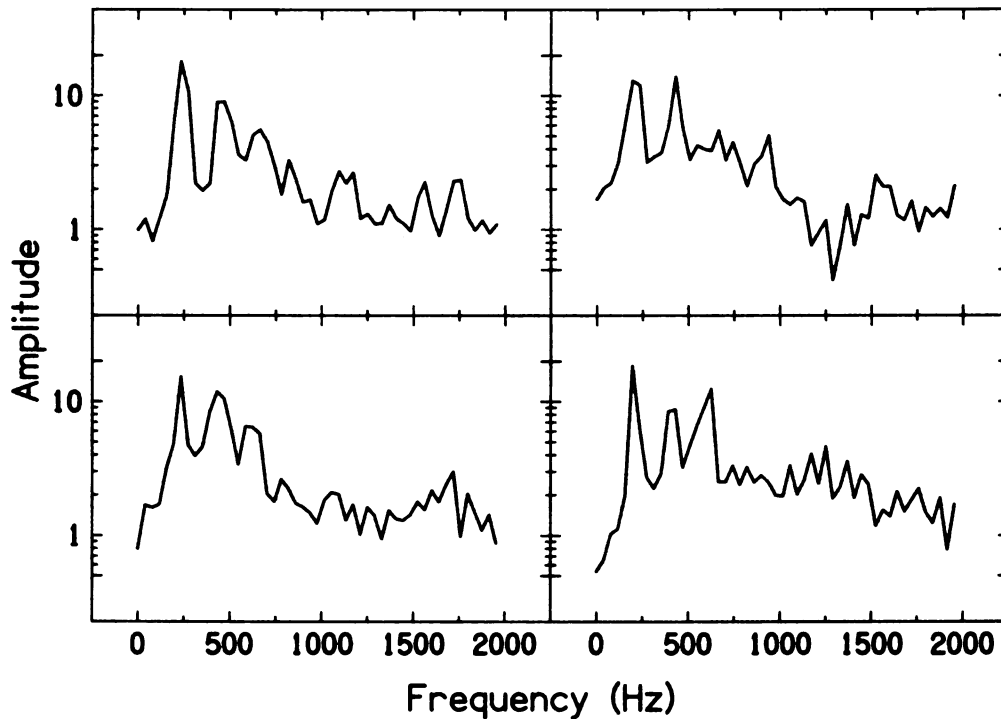


Figure 1.5. Average amplitude spectra of four different female voices. Each panel represents the average spectrum for one talker's voice, averaged over five utterances. The vertical axis has arbitrary amplitude units, the same for all four.

nal voices. Figure 1.5 shows an amplitude spectrum, averaged over five randomly chosen sentences, for each of the four talkers in the CRM. Each spectrum has a strong peak near 200 Hz indicating the fundamental component.

Pilot testing in the Front-Only condition showed that at a SNR of -10 dB, baseline performance with noise maskers was similar to that seen with speech distracters in Experiment 1. Front-Only baseline tests were therefore conducted for all listeners at -10 dB SNR. For a second reference point the speech level was increased to -6 dB SNR.

To test for release from energetic masking in the front-back geometry, both the target speech and two noise maskers were presented from the front loudspeaker, with the level of the target at -10 dB relative to each masker. Only the noises were presented from the back loudspeaker. Conditions with different values of the

delay, τ , were repeated as in Experiment 1.

1.3.3 Results and analysis

Figure 1.6 shows the results of the experiment with noise maskers. All listeners showed masking release (relative to the baseline) at $\tau = \pm 0.5$ ms, and these eight points (four listeners, two delays) were significantly statistically different from the baseline (one-sample t -test, $p < 0.05$). All listeners, except for K, exhibited the greatest release for $\tau = 0$. Listeners N and B displayed a release of about 4 dB for $\tau = 0$. Other listeners showed lesser release, and K showed a local minimum in performance.

The results for Experiment 2, averaged over all four listeners, are shown in the bottom panel of Figure 1.6. Significant release from masking occurred only for every value of τ in the region $|\tau| \leq 0.5$ ms ($p < 0.05$) and for no other value. The average release from energetic masking at $\tau = 0$ was about 2.5 dB. A release of 2.5 dB was also found for $\tau = 0.5$ ms, and a release of 1.5 dB was found for $\tau = -0.5$ ms. There was no evidence of masking release for delays outside the range from -0.5 ms to $+0.5$ ms. For $\tau = \pm 32$ ms and $\tau = \pm 8$ ms, the average performance was, in fact, significantly below the performance in the baseline condition ($p < 0.05$), presumably because the masking power in Front-Back trials was double that for Front-Only trials. These results agree with the horizontal plane experiments by Rakerd *et al.* [76], which searched for masking release in the range from -32 ms to $+32$ ms, but found significant energetic masking release only for delays near zero, specifically within the range from -0.5 ms to $+0.5$ ms and for -2.0 ms.

The results of Experiment 2 (Figure 1.6) on energetic masking can be compared with the results of Experiment 1 (Figure 1.4) using speech distracters. The following points are notable: (1) The range of delays for which the release appears in

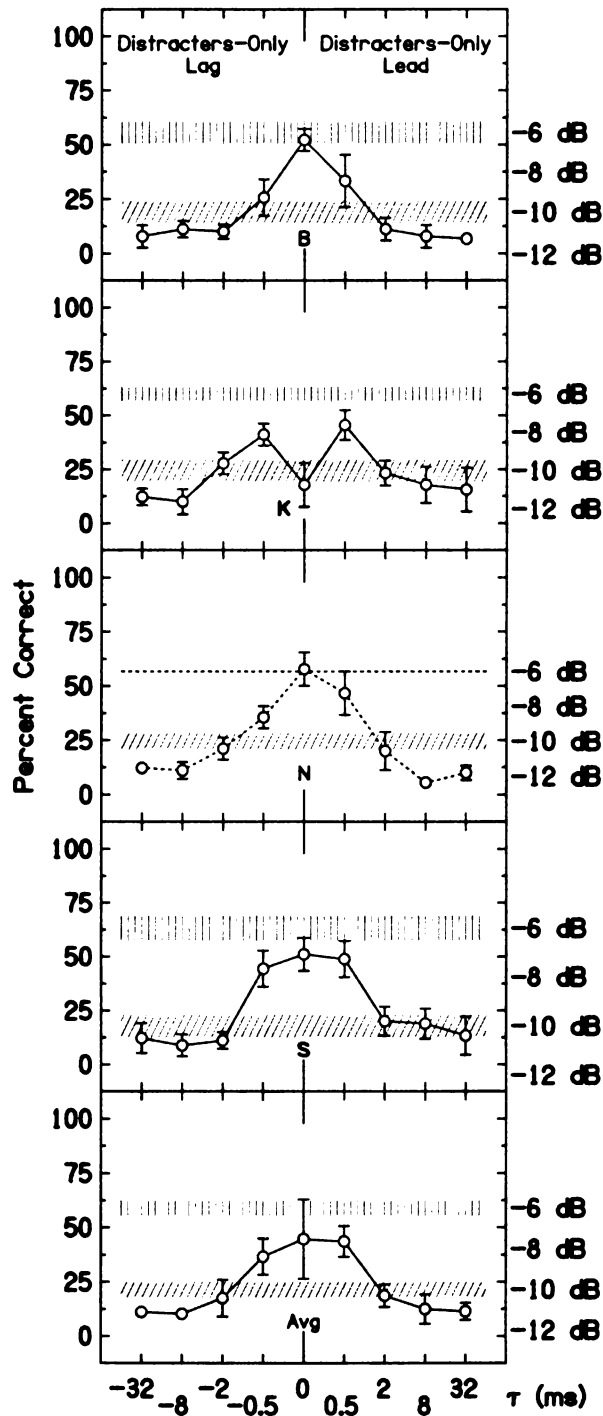


Figure 1.6. Individual and average results of Experiment 2, the Front-Back experiment with noise distracters, similar in form to Figure 1.4, as a plot of percent correct versus delay, τ . For listener N, who showed no variation within three runs in performance of the boosted Front-Only condition, a dashed straight line represents the average percent correct.

Experiment 2 (-0.5 to 0.5 ms) was far more limited than in Experiment 1. (2) No release occurred in Experiment 2 at $\tau = \pm 2$ ms, where Experiment 1 showed the greatest release. (3) The release at $\tau = \{-0.5, 0.5\}$ ms was statistically the same in Experiments 1 and 2 ($p > 0.716$ in a two-sample t -test of zero difference).

It is reasonable to conjecture that the release seen in Experiment 2 at $\tau = 0$ and the release seen at ± 0.5 ms occurred for different reasons. It is possible that the release from energetic masking at $\tau = 0$ is a localization effect, where the noise maskers are localized separately from, or are simply more diffuse than the target speech because they are identically presented from different locations. The release from energetic masking at $\tau = \{-0.5, 0.5\}$ ms could be attributed to delay-and-add (comb) filtering [47]. This may be the origin of some of the masking release with speech distracters seen in Experiment 1 as well.

Because of delay-and-add filtering, the masking noise has a broad spectral valley from 0 to 2 kHz, centered at 1 kHz. In an informal ADD experiment where listeners adjusted the delay of noise maskers, a delay near 0.5 ms was found to be particularly effective in unmasking target female speech. The conjecture that different mechanisms lead to release for the different delays might explain why all listeners exhibited a release at $\tau = \{-0.5, 0.5\}$ ms, but, unlike the other listeners, listener K failed to show a release at $\tau = 0$. According to the conjecture, all listeners were able to utilize the spectral mechanism that leads to a release at $\tau = \pm 0.5$ ms, but listener K failed to make use of the localization mechanism that other listeners used to achieve a release from masking at $\tau = 0$. Implications of these results are further addressed in section B of the Summary for this chapter.

1.4 Experiment 3: Front-Front Presentation with Speech Distracters

Interaural differences were minimized in Experiment 1, which means that the masking release for speech distracters found there was likely due to spectral effects. These spectral effects may have had either or both of two origins. One possibility is that they arose from the spatial nature of the Front-Back layout. The head-related transfer functions (HRTFs) for sources in front and back are quite different [16], and listeners may have gained some localization information from their HRTFs. The localization information in turn may have substantially mediated speech masking release.

The other possibility is that spectral effects may have arisen from delay-and-add filtering. Experiment 3 was conducted to separate out the contributions of HRTFs and delay-and-add filtering. To do this, a new layout was established, referred to here as the “Front-Front” geometry, which deprived the listeners of any front-back spatial cues but retained spectral cues caused by delay-and-add filtering.

1.4.1 Experimental design

The layout in Experiment 3 was the same as in Experiment 1, except that the back loudspeaker was placed on top of the front loudspeaker so as to make them nearly collocated. This arrangement, the Front-Front geometry (FF), is depicted in Figure 1.7. The loudspeaker alignment, method of data collection, and the target and masking stimuli were the same as in Experiment 1.

The distance between the centers of the two speakers was 8.9 cm, which corresponded to a difference in vertical angle of 3.3° . Therefore, the arrival time discrepancy between the the signals from the two speakers was less than $10\ \mu\text{s}$, producing

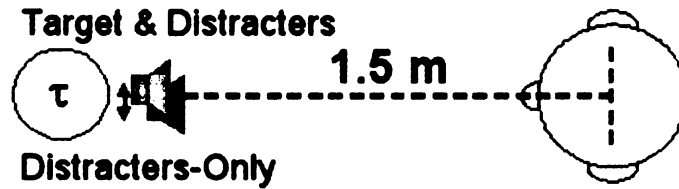


Figure 1.7. Arrangement of speakers relative to a listener for Experiment 3. The delay between speakers is represented by τ .

a negligible effect on the delay-and-add filtering.

The four listeners were the same as for Experiment 1. The task, stimulus set, and set of delays remained the same as well. The Front-Only test was repeated here to provide a measurement of baseline performance contemporaneous with the Front-Front ADD test.

1.4.2 Results and discussion

Figure 1.8 shows the results for the Front-Front ADD test. This figure is in every way parallel to Figure 1.4 for the Front-Back ADD test.

The shape of the functions in Figure 1.8 was remarkably similar across the four listeners. All listeners exhibited substantial release from masking for the delays of $\tau = \{+2, -2\}$ ms. This was most dramatically demonstrated by listener N, who showed nearly 6 dB of release for $\tau = 2$ ms. The smallest release seen at plus or minus 2 ms was for listener K who showed a release of 2.5 dB at $\tau = -2$ ms. No listener showed any evidence of substantial release for any other delay.

Another point of agreement among listeners is that they all performed worse than baseline for a delay of $\tau = 0$. In that condition, the distracter presentations perfectly coincided at the leading and lagging speakers and were therefore 6 dB more intense than at baseline.

The average results in the bottom panel of Figure 1.8 clearly show significant

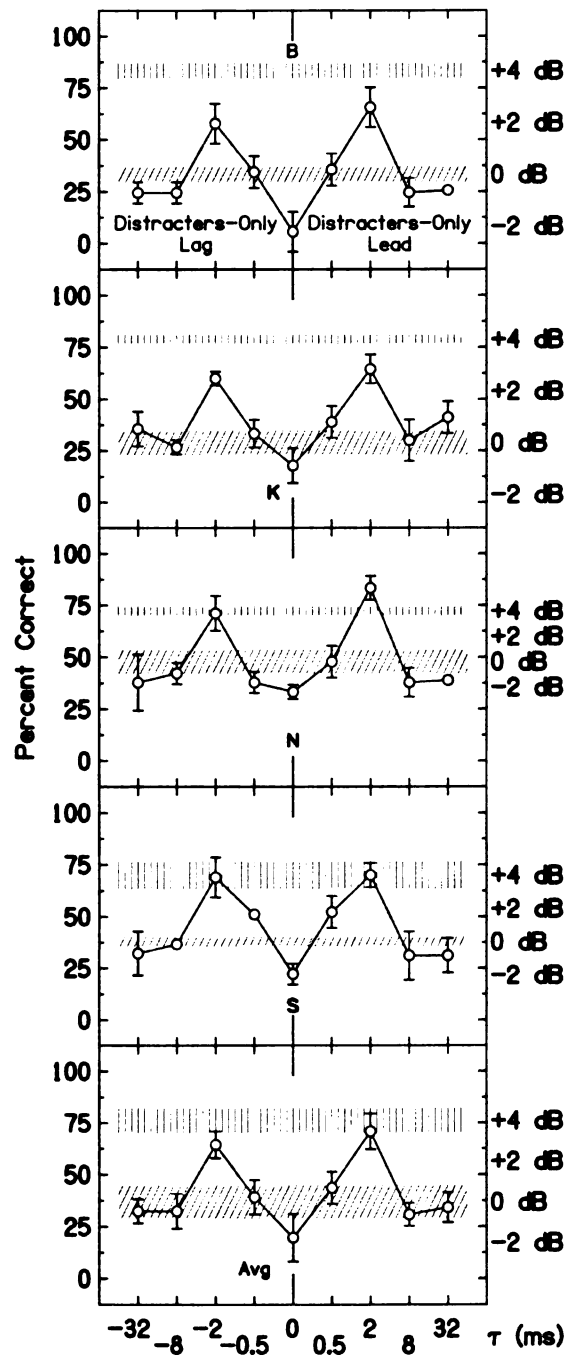


Figure 1.8. Individual and average results of Experiment 3, the Front-Front experiment with speech distractors, similar in form to Figure 1.4, as a plot of percent correct versus delay, τ .

release from masking at $\tau = \pm 2$ ms and no release at any other delay value. The peak in performance for $\tau = 2$ ms corresponds to roughly a 3.5 dB release in masking, while the peak for $\tau = -2$ ms corresponds to a release of nearly 3.0 dB. These two decibel values are essentially the same, indicating symmetry about $\tau = 0$.

1.4.3 Delay-and-add filtering

The improved performance for $\tau = \pm 2$ ms in Experiment 3 is very likely due to delay-and-add filtering of the distracters in this experimental setup. The transfer function of a delay-and-add filter with a delay of 2 ms is shown in Figure 1.9, superimposed on the average spectrum of a typical female voice from the CRM. For a delay of 2 ms, peaks occur at integer multiples of 500 Hz, and dips occur at 250 Hz and every additional 500 Hz thereafter.

The first dip in the delay-and-add spectrum for a delay of $\tau = \pm 2$ ms may be especially important. As shown in Figure 1.9, this dip is close to the average fundamental frequency of female voices, and so would remove energy from the fundamental frequency of distracters. It may be that filtering in this way introduced timbre differences that helped the listeners distinguish between the target talker and the distracters, leading to a release from masking of the target. The release seen at 2 ms in Experiment 3 (no spatial cues) was as large as the release at 2 ms in Experiment 1 (front-back spatial cues). It seems likely that the peaks at ± 2 ms seen in Experiment 1 were the result of the delay-and-add filtering effect, as made evident in Experiment 3, but spatial cues may also have been present in Experiment 1.

By contrast, no release was seen at $\tau = \pm 2$ ms with noise maskers (Experiment 2) though presumably delay-and-add filtering had a similar effect on the spectrum of the maskers in that experiment. Delay-and-add filtering of the noise maskers removes power from certain parts of the spectrum but adds power to other parts of

the spectrum. Therefore, delay-and-add filtering is not expected to lead to a release from energetic masking unless the delay is strategically selected. Unmasking the fundamental frequency of the target also does not lead to better intelligibility of the target speech, since the relevant information in speech is not contained in the fundamental formant.

Support for the notion that a dip at the fundamental frequency caused by delay-and-add filtering contributes to a release from masking by speech is found in the results of Brungart *et al.* [22]. In that study, a similar experiment to the present one was performed in a virtual auditory environment with a single *male* speech distracter and a target taken from the CRM corpus (this was referred to as the F-FF condition). Brungart *et al.* found a peak in release from masking for a delay of 4 ms, distinct from no release at 2 and 16 ms. This peak is similar to the peak seen in the present experiment for 2 ms, distinct from no release seen for 0.5 and 8 ms. A delay-and-add filter with a delay of 4 ms has a first dip at 125 Hz, close to the expected fundamental frequency for a male talker, as used by Brungart *et al.* By comparison, a delay of 2 ms places the first dip at 250 Hz, near the expected fundamental frequency for female talkers, as used in the present experiments. Brungart *et al.* also reported release at delays of 0, 0.25, 0.5, and 1 ms of magnitudes similar to that of the release at 4 ms, whereas no release for 0 or 0.5 ms was found in the present experiment.

1.5 Experiment 4: Spectral Structure - A Follow-Up to Experiments 1 and 3

The results of Experiment 3 strongly suggest delay-and-add filtering as an explanation for the release from masking for a delay of $\tau = 2$ ms when speech distracters are used. Experiment 4 tested this preliminary conclusion and tested the conjec-

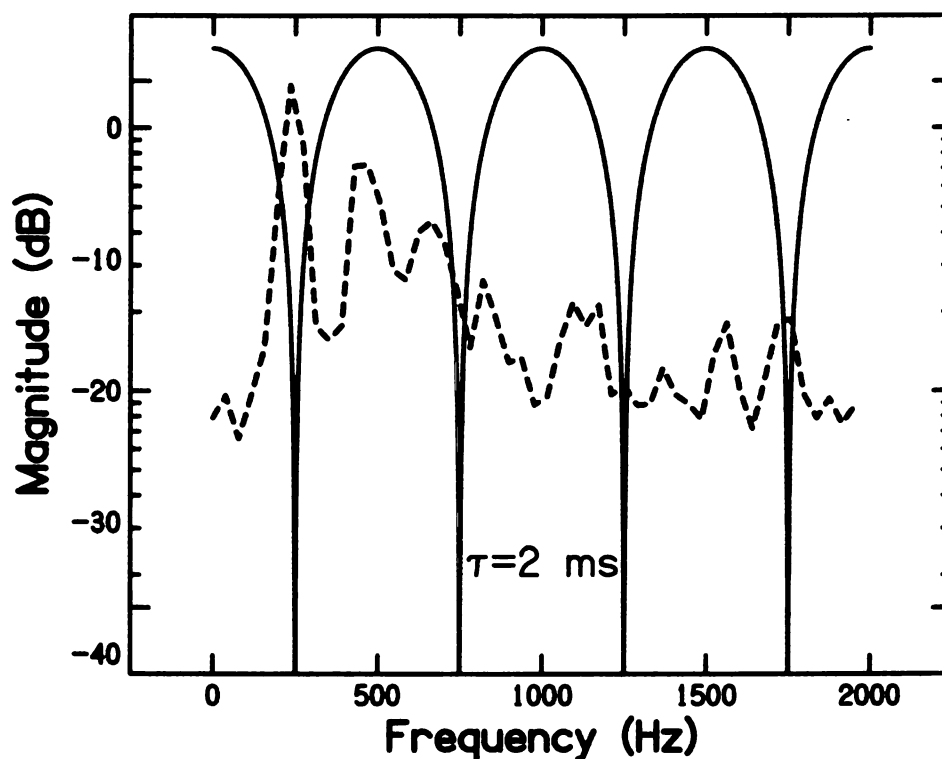


Figure 1.9. Solid lines show the theoretical amplitude response of an ideal delay-and-add filter with delay $\tau = 2$ ms. Dips occur at 250 Hz and every additional 500 Hz thereafter. Peaks occur at every integer multiple of 500 Hz. Superimposed in dashed lines is the average amplitude spectrum of one of the four female CRM voices, which is taken from the upper left panel of Figure 1.5.

ture that attenuation of the fundamental component of the distracters, near 250 Hz, was responsible for the masking release.

1.5.1 Experimental design

A digital filter was designed to “mimic” the first dip in the amplitude response of a delay-and-add filter with delay $\tau = 2$ ms. This was a finite impulse response (FIR) filter of order 336, with 0.04% ripple in the passband. The amplitude response of this filter is shown in Figure 1.10a. The dip in the filter’s amplitude response was centered at 250 Hz and had a depth of approximately 32 dB. A second filter was designed with a dip at 750 Hz in order to mimic only the second dip in the delay-and-add filter. This filter (Figure 1.10b) was FIR of order 392, with 0.02% passband ripple, and a dip at 750 Hz of 39 dB. By comparison, the first two spectral dips actually measured for the front-front geometry in the anechoic room (through a free-standing condenser microphone) used for these experiments occurred at 250 Hz and 750 Hz, with depths of 22 dB and 12 dB respectively.

The entire CRM stimulus set was processed with each digital filter to create two separate filtered corpora, one with energy removed at 250 Hz, the other with energy removed at 750 Hz. These stimuli were then used individually as the distracters in two separate Front-Only experiments. For each experiment, listeners completed three runs consisting of five practice trials (without feedback) followed immediately by 30 test trials. To simulate the effective target-to-distracter level difference of Experiment 1, wherein both loudspeakers were actively producing distracting speech, the level of the target talker was reduced by 6 dB. The listeners in this experiment were the same as in all previous experiments.

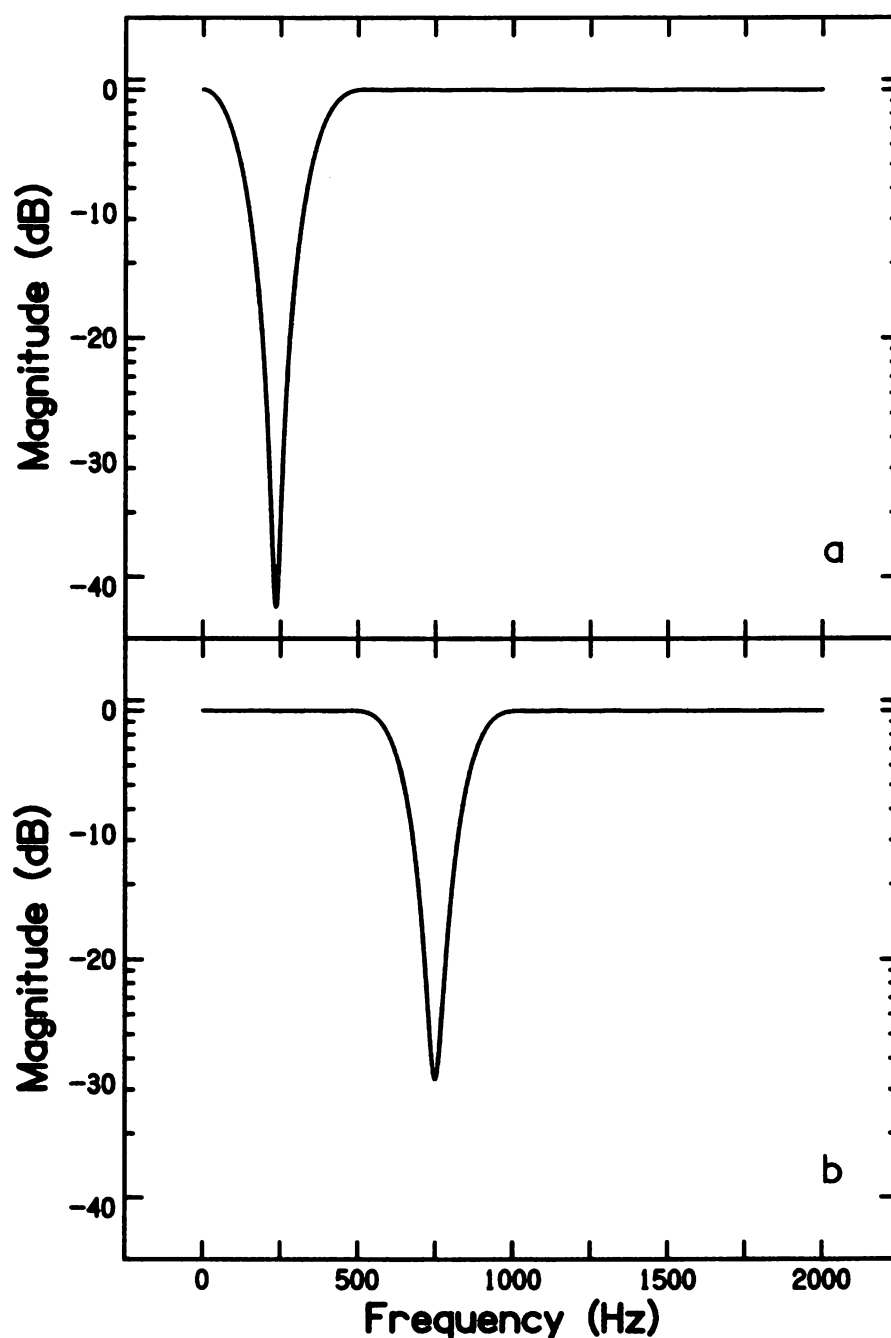


Figure 1.10. The top graph shows the amplitude response of a filter imitating the first dip in a delay-and-add filter with delay $\tau = 2$ ms. This is a FIR filter of order 336. The bottom graph shows the amplitude response of a filter imitating the second dip in a delay-and-add filter with delay $\tau = 2$ ms. This is a FIR filter of order 392.

Listener	(a) F-O 0 dB	(b) F-O 750 Hz	(c) F-O 250 Hz	(d) FF
B	33 $\pm$ 3%	24 $\pm$ 8%	60 $\pm$ 3%	66 $\pm$ 9%
K	29 $\pm$ 5%	33 $\pm$ 7%	51 $\pm$ 8%	64 $\pm$ 7%
N	48 $\pm$ 5%	47 $\pm$ 0%	64 $\pm$ 7%	83 $\pm$ 6%
S	38 $\pm$ 2%	20 $\pm$ 0%	59 $\pm$ 8%	70 $\pm$ 6%
Avg	37 $\pm$ 8%	31 $\pm$ 11%	57 $\pm$ 9%	71 $\pm$ 9%

Table 1.1. Experiment 4: Percentage of correct responses (plus or minus one standard deviation) for each listener by condition. (a) Front-only baseline condition. (b) and (c) Simulated delay-and-add filtering with a 2 ms delay. (d) Front-Front condition with 2 ms delay - data taken from Experiment 3. Averages and standard deviations for each listener are calculated across three runs. Averages and standard deviations across listeners are shown in the bottom row.

1.5.2 Results and analysis

The results of Experiment 4 are shown in Table 1.1 for each listener and for the average across listeners. “FF” refers to the results of Experiment 3, speech-on-speech masking, with a delay of $\tau = 2$ ms, where the greatest release occurred. “F-O 250 Hz” refers to the Front-Only test wherein the distracters have a notch at 250 Hz per the first dip in a delay-and-add filter with delay $\tau = 2$ ms, and similarly for the condition labeled “F-O 750 Hz”. The “F-O 0 dB” condition refers to the baseline condition where target and unfiltered distracters are presented from the front loudspeaker only.

When the filter dip was at 250 Hz, all listeners showed an improvement in performance over the Front-Only baseline, F-O 0 dB. The average magnitude of the release was 2 dB. When the dip was at 750 Hz, none of the listeners showed any improvement in performance, and listener S performed consistently worse than in the Front-Only baseline condition.

The release found for distracters with a notch at 250 Hz supports the idea that delay-and-add filtering was responsible for the release from masking demonstrated in Experiment 1. The lack of release found for distracters with a notch at

750 Hz suggests that only the first dip in the spectrum of delay-and-add filtering is responsible for this effect. However, since performance in this experiment in no case reached the level found in the Front-Front geometry for delay $\tau = 2$ ms, this first spectral dip may not be solely responsible for the release in masking shown in Experiment 3. An alternative interpretation is that the 6-dB target reduction used in Experiment 4 may have been an underestimation of the effective SNR ratio in Experiment 3.

Experiment 4 suggests that the release from masking seen in Experiment 3 uniquely at delays of ± 2 ms was an energetic effect, the result of eliminating the fundamental component of the distracting speech. A report by Freyman *et al.* [40] supports this conclusion. Those authors began with a baseline condition wherein target and distracters were both high-pass filtered so as to remove the first few harmonics. When the fundamental was added back to the target, performance improved by about 2 to 3 dB. The experiment by Freyman *et al.* is quite similar to Experiment 4 in that it demonstrated a release from masking that takes place because of the lack of energy in the fundamental of the distracters compared to that of the target.

1.6 Summary

This chapter describes added-delayed distracter (ADD) experiments in a front-back geometry, where care was taken to minimize interaural differences. Experiment 1 was the main experiment. It tested the ability of listeners to segregate target speech from distracting speech presented directly ahead of them when an additional, time-shifted copy of the distracting speech was presented behind them. Thus, this experiment continued a paradigm begun by Freyman *et al.* [41], extending it into the median sagittal plane. This experimental arrangement also simu-

lated an acoustical situation with a listener standing in front of a wall conversing with a nearby target talker, and with intense distracting talkers in the distance. With distant distracting talkers, effectively on the midline, the direct sound and the reflection from the wall in back can be comparably intense. Subsequent experiments were done in order to gain insight into the results of Experiment 1. Experiment 2 measured energetic masking in a front-back ADD experiment like Experiment 1. Experiments 3 and 4 examined the role of delay-and-add filtering by eliminating the spatial component of the release seen in Experiment 1.

A. Speech distracters: Experiment 1 showed that listeners experience a release from masking in an ADD experiment in the MSP for all delays between -32 and $+32$ ms. The magnitude of release was on the order of 2-4 dB (see Figure 1.4), and the peak release occurred for delays of $\tau = \pm 2$ ms. Comparison to previous ADD studies in the horizontal plane [76] shows that release occurs over a similar range of delays in both planes, though the magnitude of the release in the MSP is significantly less than that seen in the horizontal plane (8-11 dB). Greater release in the horizontal plane is not surprising, since it seems likely that the segregation task could only be helped by binaural cues and better-ear acoustical effects, which were minimized in the present MSP experiment. That this release occurs for delays as long as ± 32 ms is an important result of this experiment. Though small binaural differences due to asymmetries in anatomy and orientation may have contributed to the unmasking seen in this experiment, binaural discrepancies cannot alone account for these results. Further, performance was approximately independent of the sign of the delay. It did not matter which distracter led, either the front (coincident with the target) or the back, even for the longest delays.

B. Noise maskers: Experiment 2 was identical to Experiment 1 except that continuous noise maskers were used in place of speech distracters. In contrast to Experiment 1, the results of Experiment 2 (Figure 1.6) showed a release that occurs

for only short delays, $\tau = 0$ and $\tau = \pm 0.5$ ms. The magnitude of EM release and the range of delays over which it occurred were similar to results in the horizontal plane [38, 76].

The results for noise maskers in Experiment 2 were also similar to the results of a previous study with noise using HRTFs and headphones to simulate a MSP geometry [23]. In that study, which used both male and female target voices, a significant release from EM was found for delays of $\tau = \{-0.5, +1.0, +2.0\}$ ms. Brungart *et al.* [23] explained that poor performance with a noise masker in an ADD experiment is expected, since the second presentation of the masker adds noise power to the masker, but that some delays may lead to a release due to delay-and-add filtering. Short delays in a delay-and-add filter lead to broad spectral valleys through which a listener may perceive an unfiltered target.

The release from energetic masking can be understood from some combination of several effects. One effect is delay-and-add filtering. A delay of $\tau = \pm 0.5$ ms in the masking noise leads to a broad valley centered on 1000 Hz, giving the listener improved access to an important spectral region for the target speech. This is a plausible explanation for the release demonstrated by all listeners in Experiment 2 at $\tau = \pm 0.5$ ms. However, at $\tau = 0$, no delay-and-add filtering occurs, and the release seen in Experiment 2 at this delay must have some other explanation.

A second possibility is that release from energetic masking at $\tau = 0$ ms, or for the entire range, $|\tau| \leq 0.5$ ms, is a localization effect, caused by summing localization. Summing localization occurs for delays less than 1 ms, and is known to occur in the MSP [67]. Summing localization may shift the perceived location of the maskers away from the target, probably making the maskers more diffuse, leaving only the target with a clear localization. This effect may also account for some of the overall release from masking at small delays seen in the front-back ADD conditions (Figures 1.4 and 1.6). It is an effect that is absent in the front-front ADD

experiment (Figure 1.8).

The latter possibility involving localization is not initially expected. As applied to Experiment 2, that possibility requires that summing localization produces a larger release from energetic masking than is produced by the law of the first wavefront. That law is a part of the localization precedence effect. It says that the location of the leading source dominates. The problem is that the law of the first wavefront for broadband noise, for instance at a delay of 4 ms, is very strong. However, other explanations for the observed release from energetic masking at $\tau = 0$ are difficult to find. One possibility is that spectral and phase differences in HRTFs for sources in front versus those in back of the listener may result in some cancellation.

C. The role of delay-and-add filtering: Experiments 3 and 4 examined the role of delay-and-add filtering in the results of Experiment 1 by removing the spatial aspect of the ADD experiment. In Experiment 3, the back and front loudspeakers from Experiment 1 were placed together in front so that the target, distracters, and added distracters were collocated. Otherwise, Experiment 3 was identical to Experiment 1. In this way, the effect of delay-and-add filtering was separated from spatial effects. The results (Figure 1.8) show a release only for delays of $\tau = \pm 2$ ms, matching the delays at which peak release occurred in Experiment 1. Experiment 4 used digital signal processing techniques to show that the release for $\tau = \pm 2$ ms likely occurs because the first notch in the delay-and-add filter occurs near the fundamental frequency of the distracting speech. Severely attenuating their fundamentals cause the distracters to be distinguishable from the target speech by their unusual timbre.

D. Symmetry: The results of all the experiments showed a notable symmetry about the zero-delay condition. This symmetry speaks well of the quality of the alignment of the front and back loudspeakers and the listener in the geometry of this

experiment. In Experiment 1, with a speech target and speech distracters, the average data shown in Figure 1.4 were approximately symmetrical for all delays, 0.5 to 32 ms. Symmetry between positive and negative delays is not expected given a model in which localization by the precedence effect is the dominant effect. However, for every delay showing appreciable release, the release was slightly stronger when the source in back led the source in front (positive delay). This small asymmetry might be attributed to the localization precedence effect.

E. Implications

1. Previously, it was shown that a release from informational masking occurs in ADD experiments in the horizontal plane [23, 41, 38, 40, 76], and binaural cues were held primarily responsible for this masking release. It has been shown now that such a release occurs as well in the MSP when binaural cues are minimized to the extent that they are expected to be unimportant. When distracters are presented from in front and in back the greatest release from masking occurs for a delay of ± 2 ms. Experiments with distracters only in front show that this peak owes much of its importance to delay-and-add filtering. Apart from that, the front-back experiment shows release from masking for a wide range of delays, at least out to 32 ms, and this release is likely caused by the ability of listeners to localize the distracters (or to delocalize them) using the localization cues that are available in the MSP, namely spectral cues. Localization in the MSP is weaker than in the horizontal plane, and thus it was not surprising to find that the release that can be achieved in an ADD condition is smaller in the MSP than in the horizontal plane. The presence of the ADD effect out to long delays, both positive and negative, reveals a behavior similar to that noticed in the horizontal plane [76] and suggests that a similar general mechanism is at work to achieve a release from informational masking in both cases.

2. In many previous studies, the precedence effect [67, 66] has been cited as the

main mechanism by which release from informational masking is achieved in ADD experiments. By moving the *perceived* location of the distracters away from the target (i.e. for positive delays) localization of the distracters separately from the target allows for the perceptual segregation of the two. However, for large negative delays, the precedence effect will place the perceived location of the distracters very near the target. By the action of the precedence effect on localization alone, this should result in no benefit to the listener trying to hear out the target speech, and yet this study, as well as studies in the horizontal plane by Brungart *et al.* [23] and Rakerd *et al.* [76], show a release from informational masking for an equally large range of negative delays. Furthermore, in both the horizontal plane and the MSP, the magnitude of the release is similar for both positive and negative delays. Although the localization precedence effect is undoubtedly at work in ADD experiments, it appears to be only one of several contributors to the ADD effect.

CHAPTER 2

The MLS Method and Applications to Binaural Measurements in Rooms

2.1 Introduction

In order to make measurements of binaural quantities such as interaural time difference (ITD), interaural level delay (ILD), and interaural coherence (γ) in rooms, an accurate, reliable, and repeatable method of measurement must be utilized. The use of maximum length sequences (MLS) has the potential not only to measure binaural quantities accurately, but also to measure certain properties of the room itself such as impulse response (IR) and 60-dB reverberation time (RT_{60}). A program was written to use MLS to simultaneously calculate both binaural quantities and properties of the room from a single measurement.

The MLS is a binary signal (usually, in acoustics, a series whose values are 1 or -1 rather than 1 or 0) which has a length of $N = 2^j - 1$. This length corresponds to exactly one period of the MLS. The parameter j may be any positive integer, known as the “order” of the MLS. Any MLS has a perfectly flat frequency spectrum, and the MLS has an autocorrelation function that is a delta function with a small DC

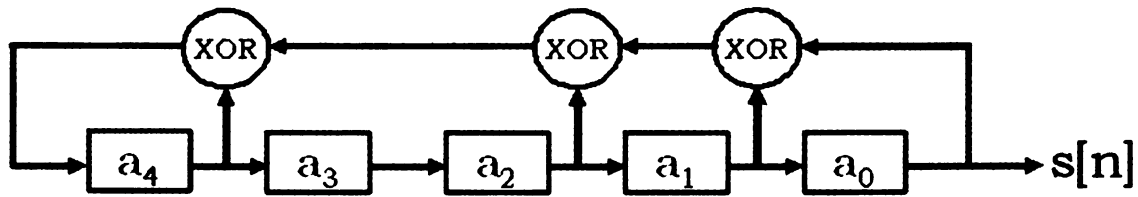


Figure 2.1. A linear-feedback shift register with five bits and taps at bits 1, 2, and 4. The bits a_i can take only binary values. An XOR gate performs modulo 2 addition of the output bit, a_0 and a_1 , the result of which becomes one input for the next XOR gate, which “taps” a_2 for its other input. This is then fed as an input to the XOR gate which taps a_4 . The last XOR gate in the sequence feeds its output back to the first bit, a_4 . The resulting output sequence, $s[n]$ will be a MLS of order 5.

shift [44, 80]. The usefulness of these properties of MLS will be described below.

A MLS can be generated by a linear feedback shift register (LFSR), which is an arrangement of connected flip flops where the input bit is a linear function of its previous state. Consider a LFSR with five bits, as depicted in Figure 2.1. Each bit, a_i starts with a random value, 1 or 0, where the only invalid set of initial values is that for which all bits start with 0. At each step, the values from the a_i th bit are moved to the a_{i-1} th bit. The sequence $s[n]$ is generated step by step by taking the value of the last bit in the register (as depicted, this is a_0). At each step, the output of the last bit is summed modulo 2 with the value held in at least one other bit. This other bit is said to have a “tap,” which takes its value and feeds it to one input of a modulo 2 summer (XOR gate). This result flows to other taps as depicted in Figure 2.1 to generate a new input for the first bit. This process is repeated $2^j - 1$ times, where the order j equals the number of bits in the LFSR. The resulting series, $s[n]$ is the desired MLS.

For lower orders of MLS, there may be as few as a single tap required in the LFSR. In fact, the order 5 MLS is the lowest order of MLS that can be generated with a tap on more than one bit (other than the last). Also, for orders of MLS of 5 or above, the set of bits that can be tapped is not unique - there is more than one

possible set of locations for the taps that will still yield a valid MLS at the output. Higher orders of MLS can be generated with a number of possible sets of locations for taps [28]. For instance, an order-10 MLS can be made with a single tap on bit a_3 , or with three taps that can be placed in one of ten different ways, or with five taps placed in one of 14 different ways, or with seven taps placed in one of five different ways. For a 17-bit LFSR, there are 1348 ways to place a set of ten taps such that the output is a MLS [68]. The locations of the taps are critical in the generation of MLS by this method. A tap that is out of place will result in an output sequence $s[n]$ that is not a MLS.

The locations of the taps are related to the existence of irreducible primitive polynomials over the Galois Field $GF(2)$ whose coefficients are only 1 or 0, corresponding to the existence or absence of a tap, respectively [44]. The Galois field is one that contains only finitely many elements. The polynomials are of a degree equal to the number of bits in the register. For the register shown in Figure 2.1, the related primitive polynomial is $x^5 + x^4 + x^2 + x + 1$, which corresponds to a tap after bits a_0 , a_1 , a_2 , and a_4 since the coefficients of x^0 , x^1 , x^2 , and x^4 are 1. Fortunately, these primitive polynomials are well known to very high order [68], and so the appropriate tap locations can be readily found.

MLS has been often compared to other methods for making acoustical measurements such as swept-sine techniques [36], impulse burst techniques [90], and time-delay spectroscopy [90]. Each technique has advantages and disadvantages, but MLS remains a standard for comparison. MLS has been noted for good distortion immunity in the resulting impulse responses [33]. The fact that different MLS of the same or different orders are relatively uncorrelated has been used for applications relating to the simultaneous measurement of several acoustical channels [96, 95].

2.2 Capabilities and Limitations

Generation and use of MLS is facilitated by modern computer technology. A computer program was designed to generate and play a MLS through Tucker-Davis Technology (TDT) System 3 RP2.1 hardware to an external loudspeaker. The hardware then simultaneously reads two input channels, which are likely connected to microphones, such as those in the ears of a Knowles Electronics Manikin for Acoustic Research (KEMAR) [24]. The signals from each channel were stored in a serial buffer on the TDT, dumped into the program, and may then be analyzed and compared off-line in various ways.

MLS of up to order 20 may be generated and played (which gives a MLS of length 1,048,575 samples), and playing at sampling rates dictated by the TDT hardware (approximately 6, 12, 25, 50, 100, or 200 kHz). At the present time, the order of MLS used in measurement is limited only by the extent of the on-board memory of the TDT hardware. The MLS may be “frozen” by starting with all ones in each bit of the bit-shift-register. Doing so will result in the same MLS each time, thus making it possible to repeat an experiment using the same MLS. It is also possible to pre-generate MLS signals and load them in at a later time, saving the significant amount of time required to generate higher order MLS. “Silent signals” can also be generated of MLS-equivalent lengths, which are useful in measuring background and internal noise levels.

Once a pair of signals was acquired, various forms of analysis were performed. For some calculations of binaural parameters, it was first useful to filter the signals into 1/3-octave bands. This was accomplished by convolving the signals with a bank of gammatone filters [47]. The gammatone filters have center frequencies that can be defined by the user, but default to ISO standard center frequencies for 1/3-octave bands [47] from 80 to 20000 Hz (thus covering eight octaves). For the gammatone filter, an order of $\eta = 4$ and Cambridge 1/3-octave bandwidths [71]

were used. These parameters were selected as such because they are believed to give a good approximation to the shape of auditory critical band filters [43].

2.3 Fundamental Equations and Concepts

Certain equations are fundamental to the study of binaural hearing. They mathematically describe the binaural parameters that are understood to be most important in matters of sound localization. Others describe important acoustical characteristics of rooms, which are important because the characteristics of a listening environment impact binaural parameters. This section will provide a common mathematical basis for measurements and calculations presented in the rest of the body of this dissertation.

2.3.1 Interaural Level Difference Measurements

The level difference between two signals of length N is computed by comparing the average power in each signal. Taking, for example, the signals to be from the left and right ears of a KEMAR, and labeling them x_L and x_R respectively, these powers, P_L and P_R , may be written as:

$$\begin{aligned} P_L &= \frac{1}{N} \sum_{t=1}^N x_L^2[t] \\ P_R &= \frac{1}{N} \sum_{t=1}^N x_R^2[t] \end{aligned} \tag{2.1}$$

With the power of the signals in each ear, the interaural level difference (ILD) can be computed as:

$$ILD = 10 \log_{10} (P_R / P_L) \tag{2.2}$$

2.3.2 Coherence and Interaural Time Difference Measurements

Cross-correlation functions can be computed, within or without a user-defined limit of the maximum lag (τ_m), on the entire signals (broadband) and on signals filtered into 1/3-octave bands. Here, the lag is defined as a circular shift in time of one signal relative to the other signal against which the cross-correlation function is being computed. The coherence, γ , is estimated by locating the maximum value of the cross-correlation function between the two signals. The interaural time difference (ITD) is estimated by finding the lag at which the maximum of the cross-correlation function occurs.

For high frequency bands, the cross-correlation function is expected to vary rapidly, reflecting the periodicity of a sinusoid at that band's center frequency. It has thus been hypothesized that the auditory periphery calculates ITDs in higher frequency bands by looking at the cross-correlation of the envelope of the signals in the left and right ears. This envelope of each signal can be calculated by taking the magnitude of the analytic signal. The envelope coherence, γ_e , may be found from the envelope of the signals in the same way that the coherence γ is found from the signals themselves using the waveform cross-correlation $c_{x_L x_R}$. Again labeling the signals x_L and x_R , the equations for these calculations are:

$$c_{x_L x_R} [\tau] = \frac{\sum_{t=0}^N x_L [t + \tau] x_R^* [t]}{N \sqrt{P_L P_R}} \quad (2.3)$$

$$\gamma = \text{Max}_{-\tau_m \leq \tau \leq \tau_m} \{c_{x_L x_R} [\tau]\} \quad (2.4)$$

$$\gamma_e = \text{Max}_{-\tau_m \leq \tau \leq \tau_m} \{c_{|x_L + i\mathcal{H}(x_L)| |x_R + i\mathcal{H}(x_R)|}\} \quad (2.5)$$

Here, $\mathcal{H}(x_L)$ and $\mathcal{H}(x_R)$ are the Hilbert transforms of the left and right signals, respectively.

$$ITD = \tau_{max} \quad \text{where} \quad c_{x_L x_R}[\tau_{max}] = \gamma \quad (2.6)$$

The limit for the range of ITDs in the “physiological range” is from -1 ms to $+1$ ms [9]. Given the radius of the head, this is thought to be on the order of the maximum time difference for sounds arriving at the ears in free field. When making measurements on a KEMAR head, it is appropriate to set $\tau_m = 1$ ms, though it should be noted that humans are quite capable of correctly discriminating larger ITDs [79, 18]. The number of samples within this time range in the cross-correlation function $c_{x_L x_R}$ is then limited by the sampling frequency. For instance, a sampling frequency of 50 kHz would yield 100 samples over a time of 2 ms, while a sampling rate of 100 kHz would yield twice as many samples over the same time span. The sampling frequency thus limits the resolution of the cross-correlation function.

For any sampling rate F_s , the time between subsequent samples of data is $t_s = 1/F_s$. Consider then measuring the waveform coherence, γ . The greatest error would occur when two samples are taken equidistant from either side of the true peak in the cross-correlation function that yields the coherence. The maximum error in ITD that could occur is then $\tau = \frac{1}{2F_s}$. Thus, it is often preferable to use high sampling rates in order to achieve high accuracy in measurements of the coherence and ITD.

Examples of the cross-correlations in certain 1/3-octave bands are shown in Figure 2.2, labeled with the center frequencies of the bands they represent. These cross-correlations for lags between $+1$ and -1 ms were calculated given the signals arriving in the ears of a KEMAR that was facing 45° off-axis from the direction of the sound source. For these measurements, the KEMAR was placed in an anechoic environment and the stimulus was an order 18 MLS played through a Mackie studio monitor situated 1.83 m from the KEMAR at ear level. Details about this

equipment will be described later. Waveform coherences γ in each band may be found by finding the height of the peak in the relevant cross-correlation function. The value of the cross-correlation function at its peak is the waveform coherence, and the time at which that peak occurs is the ITD. The ITD for each 1/3-octave band shown is indicated by the time given in the lower left corner of each panel.

2.3.3 Impulse Response and Transfer Function Measurements

Certain characteristics of a room such as the impulse response (IR) and transfer function (TF) may be measured, and these are important to understanding the acoustical characteristics of the room. To show the mathematical formalism of the method employed for taking such measurements, it is necessary to make use of the autocorrelation function of a signal y , R_{yy} , where:

$$R_{yy}[\tau] = \sum_{n=1}^N y[n] y[n - \tau]$$

The impulse response of the room may be found thanks to the autocorrelation property of the MLS of length N , x :

$$R_{xx}[\tau] = \begin{cases} N, & \tau = 0 \\ -1, & \tau \neq 0 \end{cases}$$

Thus, $R_{xx}[\tau] \approx N\delta(\tau)$. With this, the impulse response of the room at each ear, h_L and h_R may be deconvolved from the measured signals by convolving (*) each measured signal with the MLS. If y is the signal at the receiver, x is the MLS, and h is the IR of the system (see Figure 2.3), then:

$$\begin{aligned} y &= h * x \\ y * x &= h * x * x \\ &= h * N\delta \\ &= Nh \end{aligned}$$

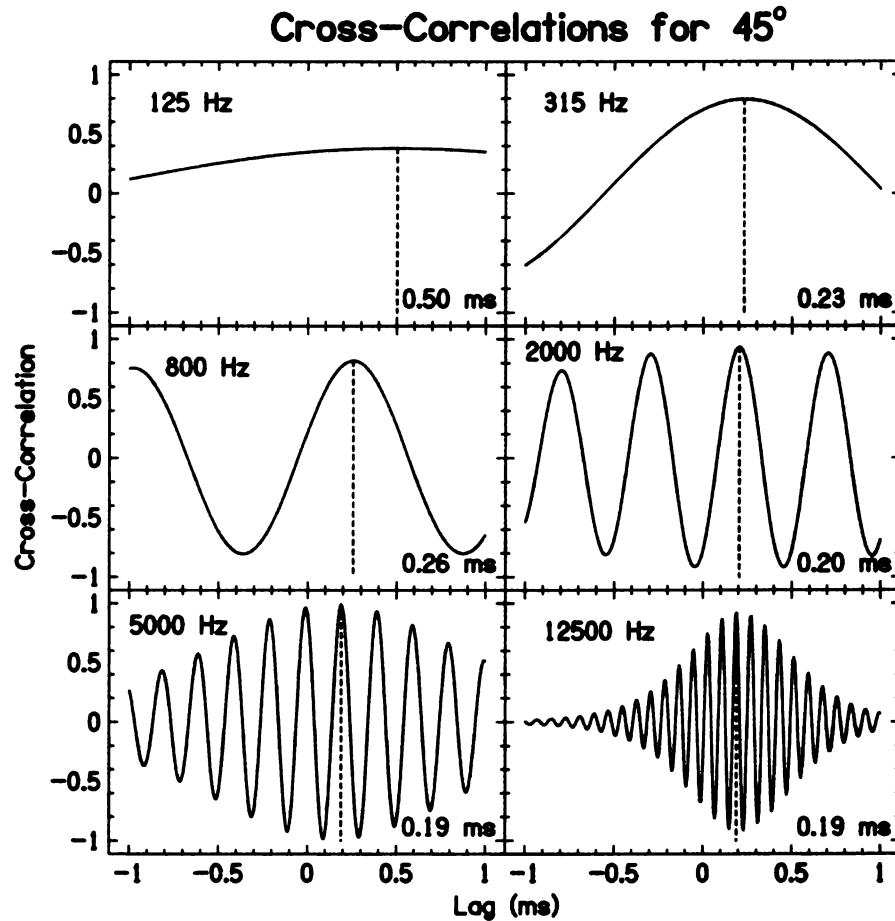


Figure 2.2. Cross-correlations between signals measured in some 1/3-octave bands in an anechoic environment between signals measured in the left and right ears of a KEMAR in an anechoic environment. The KEMAR was facing a direction 45° from the direction of the incident sound. A vertical dashed line indicates the location of the peak cross-correlation (the waveform coherence) in each panel, which occurs at the ITD. The value of the ITD in each 1/3-octave band shown is indicated by the number in the lower-right corner of each panel.

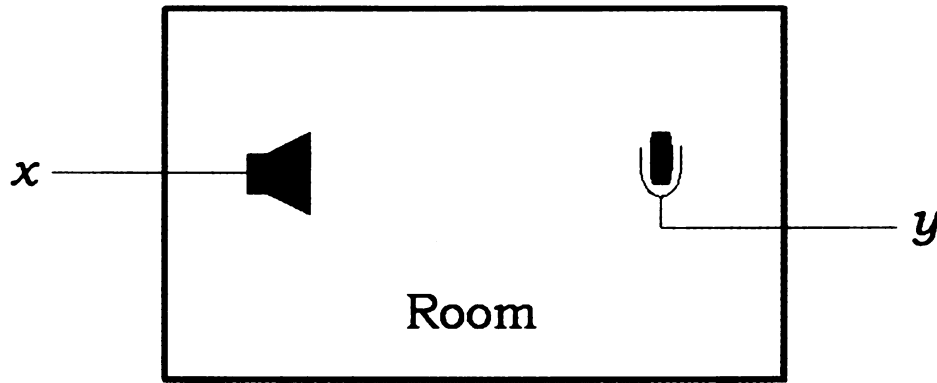


Figure 2.3. A simple depiction of the MLS measurement method. A MLS, x is played through a loudspeaker in a room with transfer function h . The signal recorded by the receiver is y . The recorded signal y is related to the MLS x via the transfer function by $y = h * x$.

It can be seen then that the result of convolving the measured signal with the MLS used to measure the signal is a scaled version of the IR, h . With the IR at hand, the transfer function of the system is only a Fourier transform away.

Additionally, multiple trials may be performed, after which the averages and standard deviations across trials for the reverberation times, ILD, ITD, and γ may be calculated. Average impulse responses, $\bar{h}_L$ and $\bar{h}_R$, formed by averaging the impulse responses across trials in the left and right ears respectively are calculated. Single impulse responses may be derived from averaging the signals in each ear respectively across trials and then calculating an IR from the average signal, $h_{L\text{avg}}$ and $h_{R\text{avg}}$. Such methods are particularly useful when making measurements in noisy environments.

2.3.4 Reverberation Time Measurements

Reverberation time is the amount of time required for sound in a room to decay in level by a certain amount. Often cited in practice is the time required for a sound to decay by 60 dB. This is RT_{60} , the 60-dB reverberation time. This time is usually

defined for broadband signals, but may also be computed at specific frequencies or in frequency bands.

Given the impulse response of the system, the reverberation times for the room may then be computed. This is done by using a method of reverse integration developed by Schroeder [81]. Schroeder makes use of the autocorrelation property of long white Gaussian noises (noises with instantaneous amplitudes that are normally distributed and long-term amplitude spectra that are flat across all frequencies) to show that if such a signal is used to excite a reverberant system, then the ensemble average squared signal at the receiving point, $\langle s^2(t) \rangle$, is related to the square of the impulse response, h^2 , by:

$$\langle s^2(t) \rangle = N \int_t^\infty h^2(t') dt' \quad (2.7)$$

Here, N is the noise power per unit bandwidth of the signal. In practice, the MLS used to measure the system is finite in length, and the length of the IR measurement will be equal to that of the MLS. After this time, the IR is assumed to be zero, and thus the upper bound on the integral in Equation 2.7 is set equal to the duration of the MLS.

The level of the left hand side of Equation 2.7 as a function of time is computed relative to the maximum of the function (which must occur at time $t = 0$). The resulting function is the “integrated impulse decay curve” (IIDC) of the system, and tends to have a bi-linear (elbow) shape as depicted in Figure 2.4. The reverberation time can be computed from this function by fitting a straight line to the first linear part of the decay. Schroeder’s method of calculating RT_{60} may be performed on the unfiltered IRs to find the broadband value of RT_{60} . Or, to find RT_{60} in 1/3-octave bands, the IR can first be filtered with the gammatone filter bank and then the reverse integration is performed. The first 400 ms of the broadband IIDC of a small, roughly square room, as measured using an order 19 MLS, is shown in Figure 2.4.

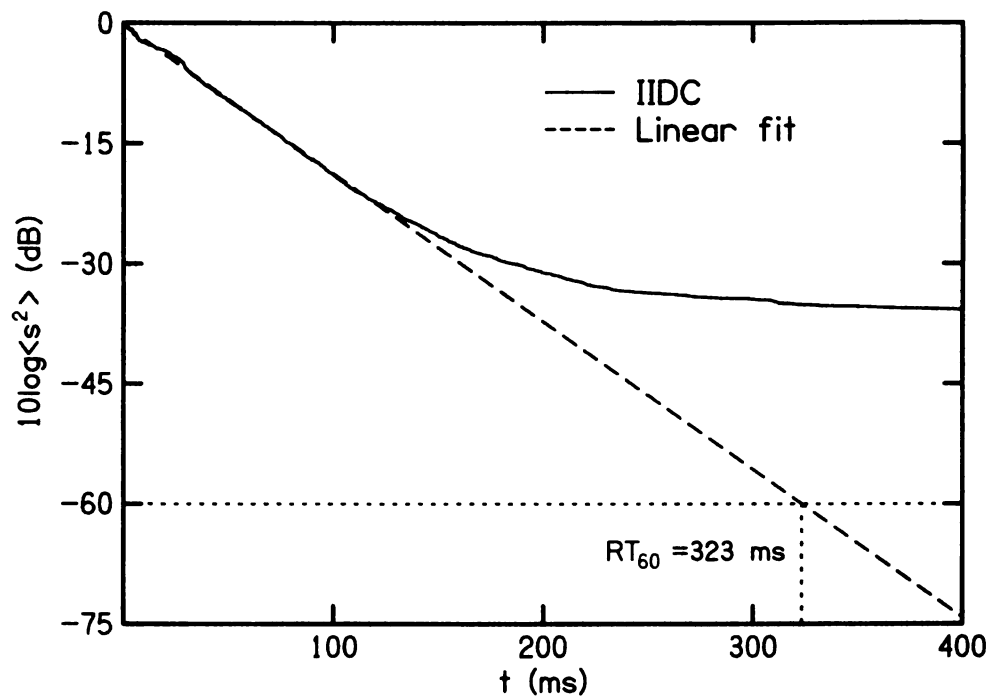


Figure 2.4. The first 400 ms of the integrated impulse decay curve (IIDC) for a small, roughly square room (of dimensions 5.7 m by 4.5 m by 3.5 m high) is shown as a solid line. A dashed line shows a linear trend line fit to the first decay mode of the IIDC. The fit line reaches a level of -60 dB at a time of 323 ms, thusly indicating the 60-dB reverberation time, RT_{60} , of the room.

Schroeder several years later described the usefulness of MLS in measuring IR and RT_{60} in rooms [83]. In that paper, he suggested that MLS are superior to finite-length white Gaussian noises because their autocorrelation property allows for a better measurement of the IR. The deconvolution of the MLS from the room IR could easily be accomplished with the use of a fast Fourier transformation (FFT). Additionally, he noted that if a single MLS would suffice in repeated measurements and Fourier techniques were utilized in deconvolving the MLS from the IR, then storage requirements would only have to be met for the MLS itself and the phase angles of its Fourier coefficients (since the coefficients all have equal magnitude).

Schroeder also mentioned the signal-to-noise ratio benefit of using MLS. He noted that it is great enough that if repeated measurements were taken of the IR in a system, that the IR (and therefore RT_{60}) could be accurately measured even in the presence of louder noise, as long as that noise is incoherent with the MLS. For example, this technique could be used even in a concert hall during a musical performance. During a musical recital of modest duration, hundreds of measurements of the IR of the recital hall could be made and averaged at the end. Assuming the music is incoherent with the MLS, the necessary level of the MLS could be so low as to be inaudible. Schroeder reported doing just this in a lecture hall while a lecture was taking place [83].

2.4 Verification of the value of the MLS method

As a simple test of the usefulness of the method, a MLS of relatively high order (19) was played through the electronic filter at a sampling rate of 97,656.25 Hz and the impulse response (IR) and transfer function (TF) were derived as described previously. The IR derived from this measurement is shown in Figure 2.5, and

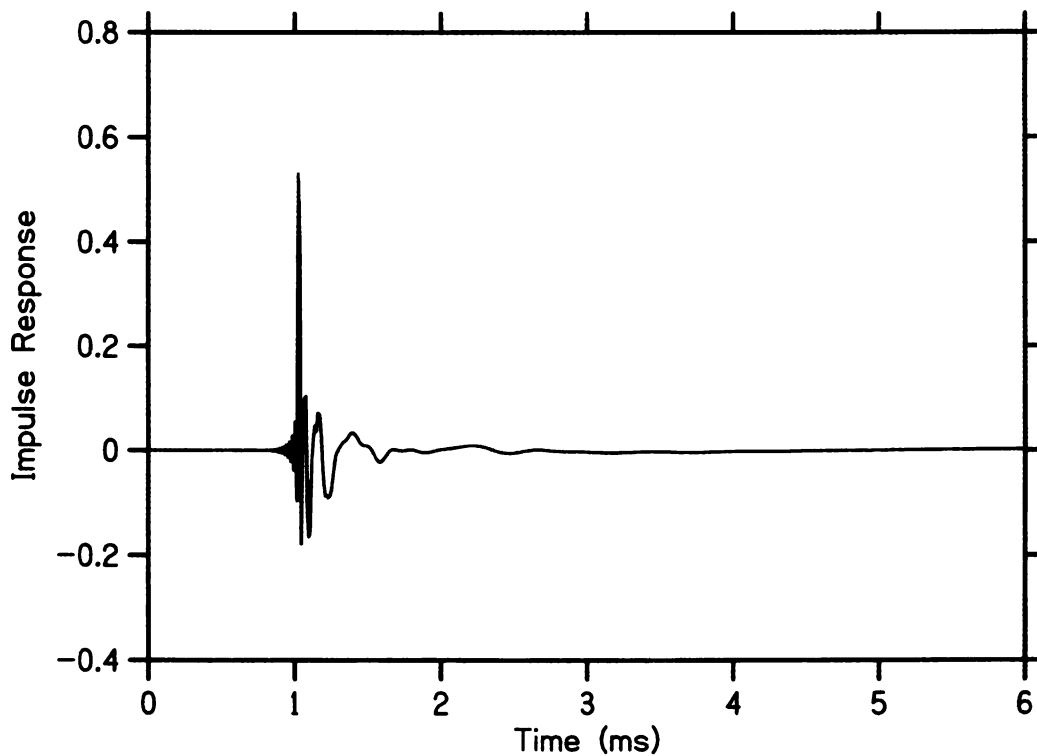


Figure 2.5. The impulse response of the electronic 1/3-octave band equalizer as measured by a MLS of order 19.

the TF in Figure 2.6. The TF was then measured again using a more conventional method — employing an analog Hewlett Packard 3580A spectrum analyzer with tracking oscillator. To the eye, there was no discernible disagreement between the TF derived from the MLS measurement and that given by the analog spectrum analyzer. Thus, the MLS method seems to be accurate.

As there are no standards for comparing the accuracy of different signals in measuring quantities such as TF, one was developed for use in this study. The TF given by an order 19 MLS, averaged over 10 measurements, each with a different MLS of order 19, was used in further experiments regarding the same electronic filter as a reference point against which all other measurements (using lower-order sequences) could be compared for the purpose of determining accuracy. As the order 19 MLS is of a length that is near the operational limitations of the hard-

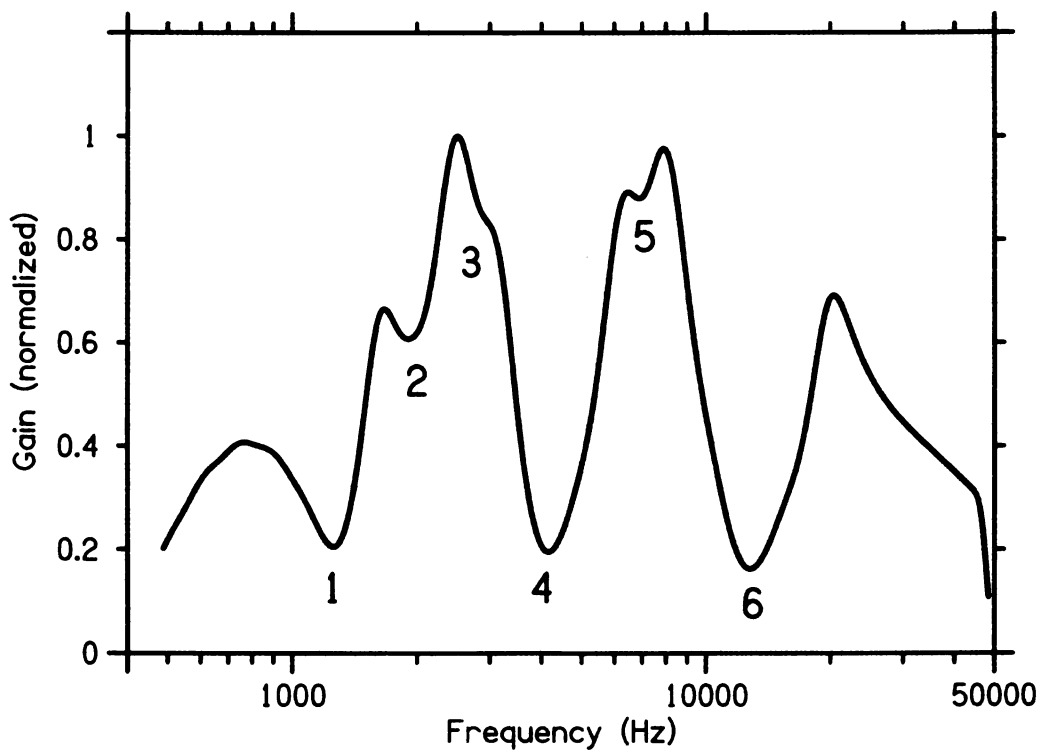


Figure 2.6. The transfer function of the electronic 1/3-octave band equalizer as measured by a MLS of order 19. Numbers 1 through 6 count the number of valleys and shoulders over the audible range.

were being used (due to physical memory limitations) and measurements using it agreed well with the analog spectrum analyzer measurement, it was used as the standard TF. Also, it is expected that any random noise should converge to the same result for the TF in the long-sequence limit, and so if measurements using, for example, random telegraph noise (RTN) do not approach the standard, then there will later be reason to question the validity of the MLS measurement.

2.5 Experiment 5: Comparison to Random Telegraph Noise

A simple test of the effectiveness of the program as well as the efficiency of the MLS method was carried out. M.R. Schroeder initially advocated the use of stationary white noise in the measurement of reverberation times due mostly to the fact that the autocorrelation of such signals is a delta function scaled by the length of the signal [81]. However, even the best stationary white noises only exhibit this characteristic in the long-signal limit. For signals of a more practical (finite) length, the autocorrelation of stationary white noise is not quite a delta function, and there is randomness in its autocorrelation at non-zero lags. This then becomes a source of error in measurements such as that of the impulse response that depend on the sifting property of the delta function when stationary white noise is used. However, this error may be made vanishingly small with a sufficiently long noise.

In principle, the MLS is more reliable and consistent than RTN, especially for short durations, in the sense that the non-zero-lag components of its autocorrelation (when normalized by the signal length) are all identically $-\frac{1}{N}$, where N is the total length of the MLS signal, for any order MLS. For MLS of even modest orders, the magnitude of the non-zero-lag components of its autocorrelation approach zero, and thus the autocorrelation approaches a true delta function quite

quickly. For example, a MLS of order $j = 7$ has a length of $N = 2^j - 1 = 127$ samples, leading to an autocorrelation function with a peak at lag zero of height 1 and a magnitude at non-zero lags of $-\frac{1}{127}$. It is thus expected that MLS will lead to more reliable estimation of the impulse response and properties derived from the impulse response such as RT_{60} , of a given system.

At the time of its development, the MLS method was heralded as being computationally efficient [83, 20]. A MLS can be deconvolved from the impulse response (IR) by means of the fast Hadamard-Walsh transform (FHT) [28]. Since both the MLS and the kernel of the FHT are made of elements of only 1 or -1 , this computation requires only addition operations. The modern era of computing has no such lack of computing power or storage space as Schroeder dealt with in his day, however; signals of almost arbitrary length may be readily generated, played, and simultaneously recorded using modern computers. Any random noise, in the long signal limit, shares the same autocorrelation property of the MLS, and so it becomes reasonable to question the value of using MLS.

In this experiment, both MLS and random telegraph noise (RTN) were used to perform similar tasks and their results and effectiveness were compared. RTN, like MLS, is a pseudorandom sequence consisting only of the numbers 1 and -1 . In this experiment, measures were taken to ensure that the MLS and RTN had similar statistical properties, and these are further discussed below.

2.5.1 Methods

In this experiment, MLS and RTN were both used to measure the impulse response and transfer function of a McLelland model GE-31 one-third-octave band graphic equalizer, whose sliders were set such that there were six valleys and shoulders in the transfer function over the audible range. For each trial, a MLS was generated with random initial inputs in the bit-shift register. A RTN sequence was then

generated by “scrambling” the MLS. This was done by taking the MLS and exchanging the value at each position with the value at some other random point in the MLS. This randomization process went through two full iterations. The RTN thus had the same length, same number of 1s and -1 s, and same RMS power as the original MLS, but did not share the same autocorrelation property as the original MLS. However, for very long MLS, the autocorrelation of the corresponding RTN, like that of the MLS itself, will approximate a delta function.

MLS and RTN signals were played by the program three times, with no pause between successive plays. The program began recording concurrently with the onset of the second repetition of the signal. Doing this allowed the system to “stabilize” before data was taken. This also allowed for any delay between input and output of the system. In an electronic filter, this may come in the form of the inherent delay of the filter; in a real room, this appears in the time it takes the sound to travel from the source to the receiver. The program may then actually record most, but not the beginning of, the second repetition of the signal, and a small part of the beginning of the third repetition. Fortunately, this corresponds to a circular shift in the signal itself, under which the properties of the MLS and RTN are unchanged. Analysis of recorded signals using the methods described above depends on this continuity under circular shifting.

2.5.2 Measurements using MLS and RTN

In this part of the experiment, a particular order of MLS and the same order RTN were played through the filter, and the response was recorded at the filter’s output. MLS and RTN of orders 9 through 18 were tested. This was repeated eight more times for a total of nine measurements with different MLS and nine measurements with different RTN each time for each order. The impulse response was calculated each time and transformed to find the TF for each calculated impulse response. All

TFs, including the standard, were computed from the impulse response using a 4096-point FFT. Each of these was compared to the standard TF, and an estimate of the RMS percent error based on the point-by-point difference between the two was calculated. Averages and standard deviations of percents error were computed across sets of nine trials for both MLS and RTN.

2.5.3 Error calculation

This section describes the method of calculation of the RMS percent error between a measured TF and the standard TF. First, let ${}^kM_i^{(j)}$ be the i th frequency bin of the TF measured by a MLS of order j . This also refers to the TF measured on the k th trial, out of n total trials. All such transfer functions are individually normalized to a maximum of 1. Then, the i th element of an average TF across all n trials (here $n = 9$) with a MLS of order j is given by:

$$M_i^{(j)} = \frac{1}{n} \sum_{k=1}^n {}^kM_i^{(j)}$$

$M_i^{(j)}$ is also normalized to a maximum value of 1. Note that the standard for comparison described above is notated as $M^{(19)}$.

The RMS percent error across all n MLS of order j , $R^{(j)}$, is the RMS average sum squared error divided by the area under the standard TF, S_{19} .

$$R^{(j)} = 100 \times \sqrt{\frac{1}{n} \sum_{k=1}^n \sum_i \left({}^kM_i^{(j)} - M_i^{(19)} \right)^2} / S_{19}$$

Similarly, the RMS percent error between the standard TF and the average measured TF from all n MLS of order j is:

$$Q^{(j)} = 100 \times \sqrt{\sum_i \left(M_i^{(j)} - M_i^{(19)} \right)^2} / S_{19}$$

Here, as mentioned previously, S_j is the root-sum-squared area under the standard TF derived from measurements with MLS of order 19:

$$S_{19} = \sqrt{\sum_i \left(M_i^{(19)}\right)^2}$$

The $R^{(j)}$ can be thought of as the average RMS percent error expected from a single measurement using a MLS of order j , and $Q^{(j)}$ is then the RMS percent error expected from the average of n independent MLS measurements. It can be shown that the quantities $R^{(j)}$ and $Q^{(j)}$ are related to the variance across n trials, $V^{(j)}$ by:

$$V^2 = R^2 - Q^2$$

These equations hold for measurements made with noise as well. For error measurements made with RTN and *a posteriori* Wiener filtering, a subscript w is added, and for RTN measurements made with exponential filtering in time, a subscript e is added. Note that the standard of comparison, however, remains $M^{(19)}$.

2.5.4 Results

The relevant percent RMS errors, noted plus and minus one standard deviation where appropriate, are given in Tables 2.1 and 2.2. First, note that, at a sampling rate of 97,656.25 Hz, the shortest MLS has a playing time of approximately 5.24 ms. “Aliasing” of the impulse response (IR) occurs when the noise used to measure the system is shorter than the impulse response itself. The result of calculating the IR is that the IR seems to repeat itself in time. The signals used in this study are already significantly longer than the bulk of the impulse response of the filter, as shown in Figure 2.5. Thus, little aliasing of the IR is expected.

Across nine MLS measurements, the average RMS percent error, $R^{(j)}$, was at most 0.76%. Also, $R^{(j)}$ was within one standard deviation of the RMS percent error inherent in the average TF, $Q^{(j)}$ at all orders. This suggests that there is little

advantage to be gained in taking multiple measurements with MLS at any order. In general, there is a trend towards lower RMS percent errors with increasing order.

Order (j)	MLS $R^{(j)}$	RTN $R^{(j)}$	RTN _w $R_w^{(j)}$	RTN _e $R_e^{(j)}$
9	0.76 ± 0.03	71 ± 5	67 ± 5	35 ± 8
10	0.39 ± 0.04	76 ± 6	75 ± 5	25 ± 6
11	0.37 ± 0.05	76 ± 3	69 ± 5	20 ± 7
12	0.38 ± 0.016	78 ± 4	60 ± 5	14 ± 3
13	0.076 ± 0.0019	80 ± 3	60 ± 3	10 ± 1.6
14	0.0533 ± 0.0008	70 ± 4	39 ± 3	9 ± 1.8
15	0.045 ± 0.002	60 ± 4	31 ± 6	8 ± 2
16	0.0643 ± 0.002	47 ± 7	18 ± 3	6 ± 1.5
17	0.0276 ± 0.0008	37 ± 3	13 ± 4	6.0 ± 1.0
18	0.0301 ± 0.0007	25 ± 4	12 ± 2	5.4 ± 0.6

Table 2.1. Average percents RMS error, plus and minus one standard deviation, for single measurements of the TF at each order j (corresponding to the length of the signal). The header label of each column indicates the type of signal used (MLS or RTN). Quantities under the $R_w^{(j)}$ correspond to RTN measurements made including *a posteriori* Wiener filtering of the IR, and quantities under the $R_e^{(j)}$ correspond to RTN measurements made with exponential windowing in time of the IR.

At all orders, measurements with RTN exhibited significantly greater percent RMS errors (often by a factor of greater than 100) than did measurements with MLS. Averaging the response across trials and then calculating a TF gave a significant improvement in the performance of the RTN measurements. The average improvement (measured as the difference of errors for RTN, $R^{(j)} - Q^{(j)}$, found in the second columns of Tables 2.1 and 2.2) was 35.2%. This shows that repeated measurements are certainly necessary at any order at least up to 18 (which corresponds to a signal length of 2.68 s) when using RTN.

Even in the best-case scenario - an averaged TF measurement using order 18 RTN - there is significantly more error ($Q^{(18)}$) than that given by even a sin-

Order (j)	MLS $Q^{(j)}$	RTN $Q^{(j)}$	RTN _{w} $Q_w^{(j)}$	RTN _{e} $Q_e^{(j)}$
9	0.755	42.0	26.7	11.6
10	0.360	52.6	31.3	12.7
11	0.374	33.6	39.1	9.69
12	0.384	42.0	26.2	5.77
13	0.0738	40.0	20.2	6.82
14	0.0527	33.0	11.5	6.72
15	0.0447	21.3	8.52	7.76
16	0.0642	17.2	6.41	5.49
17	0.0274	13.6	3.07	5.69
18	0.0300	8.28	2.53	5.12

Table 2.2. Percents error calculated from the average TF calculated across nine trials using various methods and signals. The header label of each column indicates the type of signal used (MLS or RTN). Quantities under the $R_w^{(j)}$ correspond to RTN measurements made including *a posteriori* Wiener filtering of the IR, and quantities under the $R_e^{(j)}$ correspond to RTN measurements made with exponential windowing in time of the IR.

gle MLS of order 9 ($R^{(9)}$). To see this, consider Figures 2.7 and 2.8. It is clear that the measurement from a single order 9 MLS (plotted as a solid line in Figure 2.7) shows almost perfect agreement with the reference accepted TF (plotted as a dashed line). The only serious discrepancies occur at low frequencies, perhaps because such a short signal (a MLS of order 9 sampled at the current sampling frequency is just over 5 ms long) does not capture enough oscillations of low frequency components of the spectrum to make an accurate measurement of them. Figure 2.8 shows a similar picture for the average measurement of the TF taken from nine RTN measurements of orders from 13 to 18. Clearly, the RTN measurements improve significantly at higher orders. However, it is plain to see that although the RTN have provided perhaps an acceptable representation of the TF at higher orders, its jaggedness makes it far less accurate than a MLS of even modest order.

The results shown in this experiment confirm predictions made on the basis of statistical properties of both MLS and RTN. The autocorrelation property of the MLS is such that it looks more like an ideal delta function than that of a RTN at almost any order at which the signals are of a practical length. There are, however, noise reduction techniques which may prove useful in processing the IR given by RTN measurements, and these may lead to much improved results. It may also help to assume certain properties of the IR as well, such as the expectation that the IR should decay rather quickly.

2.6 Experiment 6: Error Reduction Techniques for RTN Measurements

Some improvement in the performance of RTN in measuring the TF may be possible if certain assumptions are made about the nature of the IR of the system being

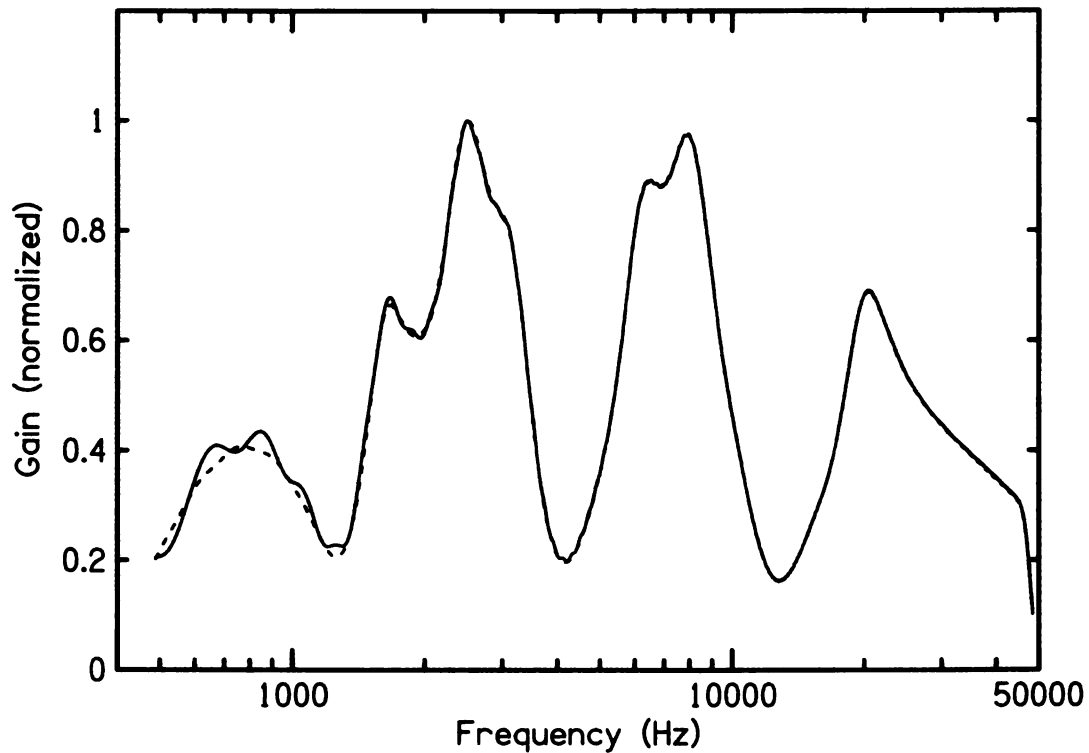


Figure 2.7. The accepted TF, plotted as a dashed line, and the TF as measured by a MLS of order 9, plotted as a solid line. There is almost no difference between the two at most frequencies, and thus the order-9 MLS measurement obscures the dashed line corresponding to the accepted TF.

measured. If the results are such that the error rates are closer to those demonstrated by MLS techniques, then the RTN may still be a viable method of measurement.

A simple assumption that can be made regarding the nature of the IR is that the IR should disappear after a long time. For instance, Figure 2.5 shows that the IR of the electronic filter measured disappears by about 6 ms. The signal used to measure it, an order 19 MLS, was 5.37 s long, and so the entire measurement of the IR also extends to 5.37 s. In the tail of the IR then, we expect nearly no signal to exist. For example, the RMS signal level for the last second of the IR shown in Figure 2.5 is on the order of 10^{-9} . During the last second of a single RTN measurement of order 18, the RMS signal level is on the order of 10^{-6} , three orders of magnitude

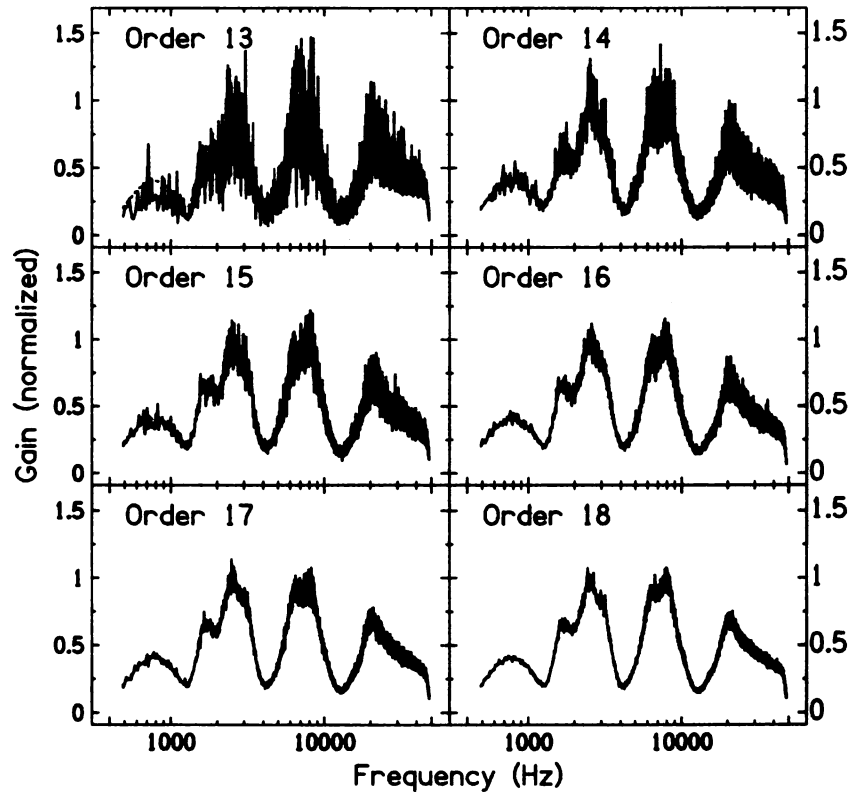


Figure 2.8. The accepted TF, plotted as a dashed line, and the average TF measured by nine RTN of orders as indicated in each panel, plotted as a solid line. The jaggedness of the TF as measured by the RTN tends to obscure the dashed line.

larger than the standard. The RMS signal level in the averaged IR calculated over nine RTN measurements is less than one order of magnitude - about a factor of three - smaller.

This noise in the tail of the measured IR may be at least partially responsible for the jaggedness seen in the average TF measured by the order 18 RTN and the otherwise gross errors at lower orders. Here, two fast methods are presented that target this potential source of error - Wiener filtering and exponential windowing in time. The Wiener filter is a widely-used method for reducing noise power in a signal and is based on statistical methods rather than spectral analysis. The exponential window is an *a posteriori* filter that operates on the measured IR in time in a way that forces the IR to go to zero at high time. First, the effect of Wiener filtering

was examined.

2.6.1 Wiener filtering

The Wiener filter is a method of removing noise from a signal based on statistical methods. The properties of the non-causal Wiener filter are well-understood (see, for example [85, 89]) and the main concepts are outlined in Appendix C. The Wiener filter requires information about the noise and signal. Assuming the noise and signal to be uncorrelated, it suffices to know the spectral power of the noise.

Methods

The noise power was computed as the variance of the signal in the last quarter of the IR. With this, the Wiener filter operated on the entire IR to yield a filtered IR. This IR was then treated as in Experiment 5 to find the TF and the RMS percent error, $r_w^{(j)}$ and $w_w^{(j)}$, for all previously studied orders j (9 through 18).

Results

The average RMS percent errors are shown in Tables 2.1 and 2.2. The column labeled $r_w^{(j)}$ shows average RMS percent errors, plus and minus one standard deviation, for any given single RTN measurement when Wiener filtering is applied *a posteriori* to the derived IR, before the TF is obtained. The column labeled $q_w^{(j)}$ similarly shows the RMS percent error when Wiener filtering is applied to the average IR measured by RTN across all nine trials.

For RTN of order 11 and higher, $r_w^{(j)}$ is significantly less than $r^{(j)}$ ($p < 0.01$ in a paired t-test of absolute differences). By the same standard, $q_w^{(j)}$ is significantly less than $q^{(j)}$. On average, Wiener filtering decreases the RMS percent error in the averaged TF measurement by about 13%.

Without Wiener filtering, a best RMS percent error among the average TF measurements with RTN was $Q^{(18)} = 8.28\%$ for RTN of order 18. A similar RMS percent error was achieved after Wiener filtering for RTN of order 15, hence with a signal one eighth as long. At order 18, a single RTN measurement after Wiener filtering resulted in, on average, $R_w^{(18)} = 12 \pm 2\%$ error, and the averaged TF had only $Q_w^{(18)} = 2.53\%$ error. Though this is still over three times the average error of a single MLS measurement at order 9, it is a substantial improvement over measurements using RTN without Wiener filtering.

2.6.2 Exponential windowing

As was noted above, the impulse response of a real room or filter is expected to go to zero after a long time. The presence of noise in the system or the measurement of a system with noise whose autocorrelation function is not very close to a delta function will leave noise in the long-time tail of the impulse response and this is a source of error when calculating the TF or RT_{60} . A potential filter for reducing the noise in the long-time part of the IR would be one which exponentially attenuates the IR after a certain amount of time. This would leave the bulk of the structure in the IR unaffected, while forcing the later parts of the IR to die away. The main difficulty with this method is that it requires some amount of *a priori* information that can only be gathered *a posteriori*. That is, the filter needs to know at what time to begin the decay, but this requires some knowledge of the yet-unmeasured IR. Fortunately, since the main function of the window is to force the parts of the IR at long times to decay, the exact time at which the window begins is not critical. It is only necessary to leave the characteristic structure of the IR intact.

Methods

Whereas the Wiener filter required knowledge of the noise power in the tail of the IR, the exponential window needs to know how long to let the IR pass before beginning the exponential decay. This time must in practice be approximated, and a good guess could be made by examining the IR derived from a relatively low-order RTN measurement. Calling this estimated value τ , the form of the exponential window as a function of time, $e(t)$ is then:

$$e(t) = \begin{cases} 1 & \text{for } t \leq \tau_e \\ e^{-b(t-\tau)} & \text{for } t > \tau_e \end{cases}$$

Here, τ_e is the user-defined time at which exponential windowing begins. The parameter b is a rate parameter which affects the rate of decay of the exponential window in time. In practice, an initial rough measurement of the IR must be made first in order to choose τ such that most of the information in the IR occurs before τ . For example, if the IR is seen to almost entirely disappear by a time of 0.35 s, then it is preferable to set $\tau_e = 0.35$. The choice of the rate parameter b is much more arbitrary, as it merely affects the rate at which the exponential window goes to zero. Very large values of b will lead to an exponential window that looks like a square window in time from $t = 0$ to $t = \tau_e$. An example of an exponential window tailored for the system measured in Experiments 1 and 2 is depicted in Figure 2.9 with $b = 0.05$ and $\tau_e = 1.5$ ms. These values were selected because they performed very well in generating an accurate IR.

Results

The RMS percents error for TFs measured after exponential windowing are shown in the bottom-rightmost parts of Tables 2.1 and 2.2. The average RMS percent errors, plus and minus one standard deviation, for each individual measurement us-

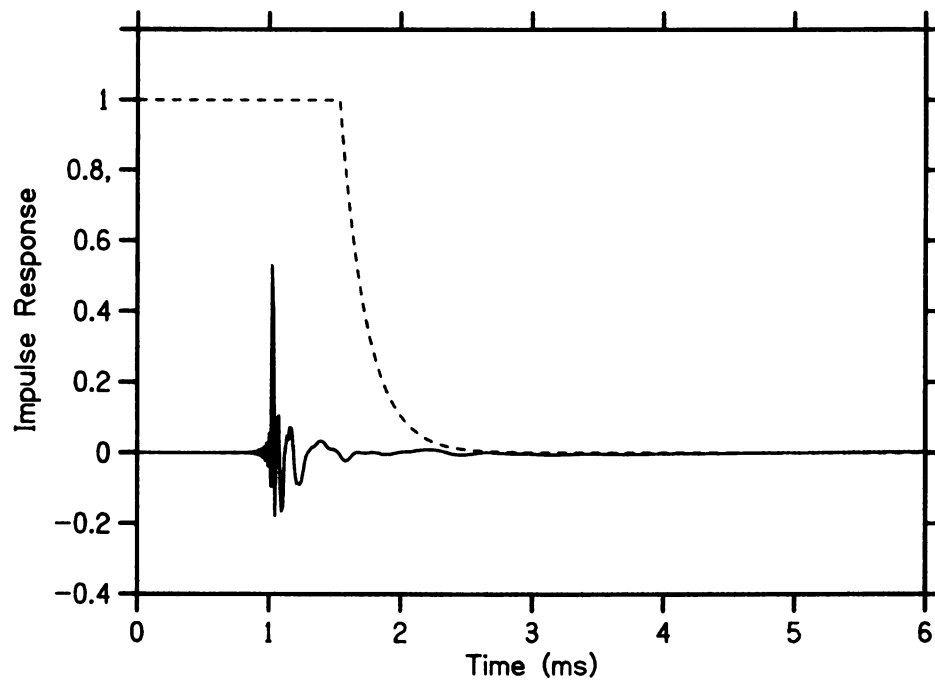


Figure 2.9. The accepted impulse response, plotted as a continuous line, and the exponential window in time, plotted with a dashed line. The exponential window shown here has a rate parameter of $b = 0.05$ and starts its decay at time $\tau_e = 1.5$ ms.

ing RTN and applying an exponentially decaying window in time are given in the $R_e^{(j)}$ column. The RMS percent errors for the average TF formed from applying the exponential window to the IR derived from the average of all nine measurements using RTN are in the column labeled $Q_e^{(j)}$.

Most notably, the exponential time window significantly decreased the average error of a single measurement using RTN at all orders tested when compared to the RTN measurement without filtering or with Wiener filtering (compare the $R_e^{(j)}$ column to the $R^{(j)}$ and $R_w^{(j)}$ columns for the cases of no filtering and Wiener filtering, respectively). The fact that, for individual measurements, the exponential window performed better than Wiener filtering is likely due to the Wiener filtering destroying some fine structure in the low-time part of the IR. The exponential window in time left the main body of the IR unchanged while only diminishing the long-time part, which is physically expected to go to zero.

In all cases, the error in the average TF after the exponential window ($Q_e^{(j)}$) was lower than the error without filtering ($Q^{(j)}$). However, it did not perform better than the Wiener filter ($Q_w^{(j)}$) at high orders. At high orders, the power of the signal remaining in the long-time part of the IR is quite small, especially in the IR averaged over all nine trials. The Wiener filter is adaptive to the noise power, and so it is weaker for measurements with high order signals, where the noise power is expected to be weaker as the autocorrelation of the RTN more closely approximates a delta function. A weaker Wiener filter is less likely to destroy fine structure in the low time part of the IR that it might otherwise “mistake” for noise. At high orders, the exponential window, however, has less and less of an effect, since the long-time noise that it is designed to dampen is already quite small.

2.6.3 Conclusions

The MLS technique has not only proven to be an effective and accurate method of measuring the acoustical characteristics of stationary systems, but it has been shown to be far more accurate than even long RTN signals. Assuming the signals are longer than the IR of the system being measured, even a single short measurement with MLS more accurately measures the IR and TF of a stationary system than a RTN several times longer, even if the average of several RTN measurements is used.

Filtering the IR derived from the RTN measurements has proved to be quite useful in reducing the error in the resulting TF. Both Wiener filtering and the application of an exponentially decaying window in time to the IR were quite successful at improving the TF measurements. Which one should be used with RTN measurements then depends on the situation. If the signals to be used are long (here, longer than about 1.3 s) and a measurement of the IR is to be made by averaging the response from multiple trials, then Wiener filtering appears to be advantageous. In all other cases, the exponential window filter in time has the advantage. In no case do either of these filters improve the RTN measurements to such a degree that they are more accurate on average than even a single MLS measurement of the smallest length considered. However, these filtering methods may still be useful when only RTN data is available.

In most studies of rooms, it is customary to require that the signals be of such a length that they are longer than RT_{60} for the room. This prevents aliasing of the important parts of the IR that contribute to the calculation of RT_{60} . In even moderately reverberant rooms, this may be as long as 1 s. An order 17 MLS, of length $2^{17} - 1 = 131071$ at the current sampling frequency of 97656.25 Hz (set by the TDT hardware) has a length of 1.3 s. Since an order 16 MLS will be almost exactly half as long, an order 17 MLS is the shortest MLS at this sampling rate

useful for measuring such rooms without aliasing the IR. Also, real rooms tend to be, at best, pseudo-stationary. The acoustical characteristics of real rooms are dependent, for example, on atmospheric conditions such as relative humidity and temperature [9]. Thus, measuring real rooms necessitates repeated measurements and averaging. In such cases, Wiener filtering seems to be the best option. In some rooms, where RT_{60} may be very long, it may be difficult to generate and work with a MLS of sufficient order, especially if a high sampling rate is also required. In that case, it may suffice to fall back on RTN techniques, assisted by the filtering techniques presented here. These techniques may also be useful for cleaning up existing RTN data.

No amount of filtering or correction seem to make the TF measured by RTN as smooth or as accurate as that measured by MLS. This alone should justify the small amount of extra processing time inherent in forming the MLS. The findings developed in the process of filtering the RTN may also be used on MLS measurements in real systems where there is additive noise.

2.7 Experiment 7: MLS and RTN measurements in a real room

Real rooms are a useful example of systems that are expected to have a certain amount of “noise” in them. The acoustical properties of rooms are affected by noise sources such as ducts and fans, and these properties may change over time with a plethora of factors such as the room’s contents, temperature, and humidity. The purpose of this experiment was to compare the performance of MLS and RTN in the measurement of a real room. The measurement of real rooms usually involves more noise inherent in the system than what exists in a typical stationary filter. Given inherent noise in the system, one may once again question the necessity of

the MLS if the presence of significant noise disrupts its beneficial characteristics. That is, if there is noise inherent in the system, will it matter which test signal is used? This experiment is intended to address that question.

2.7.1 Methods

Measurements were made with order 19 MLS and RTN in a roughly square room, measuring approximately 5.7 m by 4.5 m by 3.5 m high (this room is identified as 172GH, and is the same room whose broadband RT_{60} is shown in Figure 2.4). The Schroeder frequency [82] of a room is that approximate frequency above which there are many overlapping normal modes and the room will respond in a reverberant way. Below this frequency the room will tend to produce discrete, well-separated normal modes. This frequency in Hz, f_s , is given by:

$$f_s = 2000 \sqrt{\frac{RT_{60}}{V}} \quad (2.8)$$

Here, V is the volume of the room and the leading coefficient of 2000 has units of $\sqrt{m^3/s^3}$. For this room, the Schroeder frequency was approximately $f_s = 120$ Hz. Results of measurements made in frequency bands centered below this frequency may be unreliable as there exists the possibility that the microphone was located near a node of a standing wave within that low frequency band. Above this frequency, measurements should be considered reliable since it is unlikely that standing waves in the room dominate frequency bands centered above the Schroeder frequency.

RTN measurements were made with and without *a posteriori* Wiener filtering. The IR was calculated as the average over all nine trials. The TF and RT_{60} in 1/3-octave bands were then calculated from this IR. Thirty 1/3-octave bands were used from 20 to 16,000 Hz according to ISO standards [47].

Signals were played through a Mackie HR824 high-accuracy studio monitor

and recorded by a Shure KSM32 directional condenser microphone. The source and receiver were separated by approximately 2.6 m, and were turned so as to “face” one another. Signals were played at a level of 91 dB SPL, and the level of the background noise inherent in the room was approximately 44 dB SPL (A-weighting).

2.7.2 Results

The TFs measured via all three methods (MLS, RTN, and RTN with Wiener filtering applied) can be seen in Figure 2.10. The TF as measured by MLS and RTN appear to be quite similar. The shapes are quite similar, though perhaps the TF measured by RTN is a bit more jagged at low frequencies. The TF measured by RTN after Wiener filtering was applied looks quite different, however. Of particular note is the peak near 26 kHz, whose magnitude the Wiener filtering reduces. Also, the TF as measured after Wiener filtering seems to be much more jagged at low frequencies up to about 10 kHz.

A plot of RT_{60} measured in 1/3-octave bands by all three methods is shown in Figure 2.11. Open circles represent MLS measurements, exes represent RTN measurements, and open diamonds represent Wiener-filtered RTN measurements. Measurements of RT_{60} derived from RTN measurements without Wiener filtering were quite erratic, and are likely unreliable, especially given estimations of long RT_{60} at high frequencies where the room is expected to become more anechoic. RTN measurements nearly agree with MLS and Wiener-filtered RTN measurements for the measurements at 80, 100, 125, and 315 Hz but otherwise return higher-than-expected values of RT_{60} .

Wiener filtering resulted in RTN measurements that largely agreed with the MLS measurements, though there are some significant differences at low frequencies. The Wiener filtering likely makes RTN estimations of RT_{60} agree with those

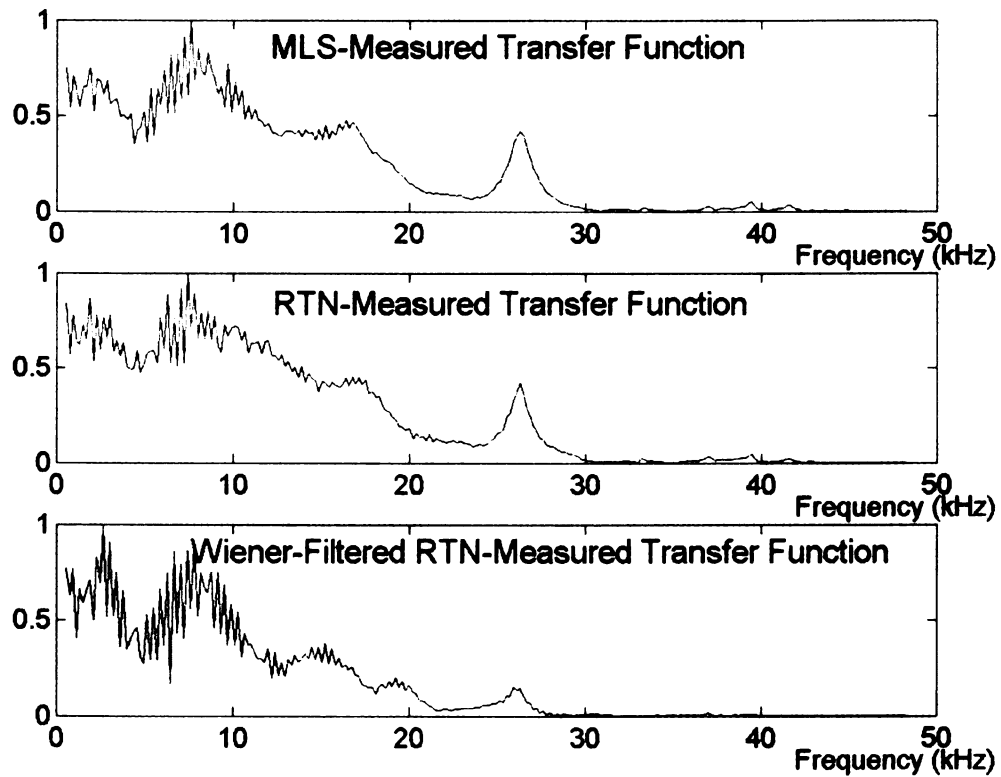


Figure 2.10. The TF of a small, roughly square office of dimensions 5.7 m by 4.5 m by 3.5 m high. In each case, the TF was calculated from the average IR calculated over five measurements. The top panel shows the TF as measured by MLS. The middle panel shows the TF as measured by RTN. The final panel shows the TF calculated when the average IR measured by RTN was filtered by a Wiener filter before the TF was calculated.

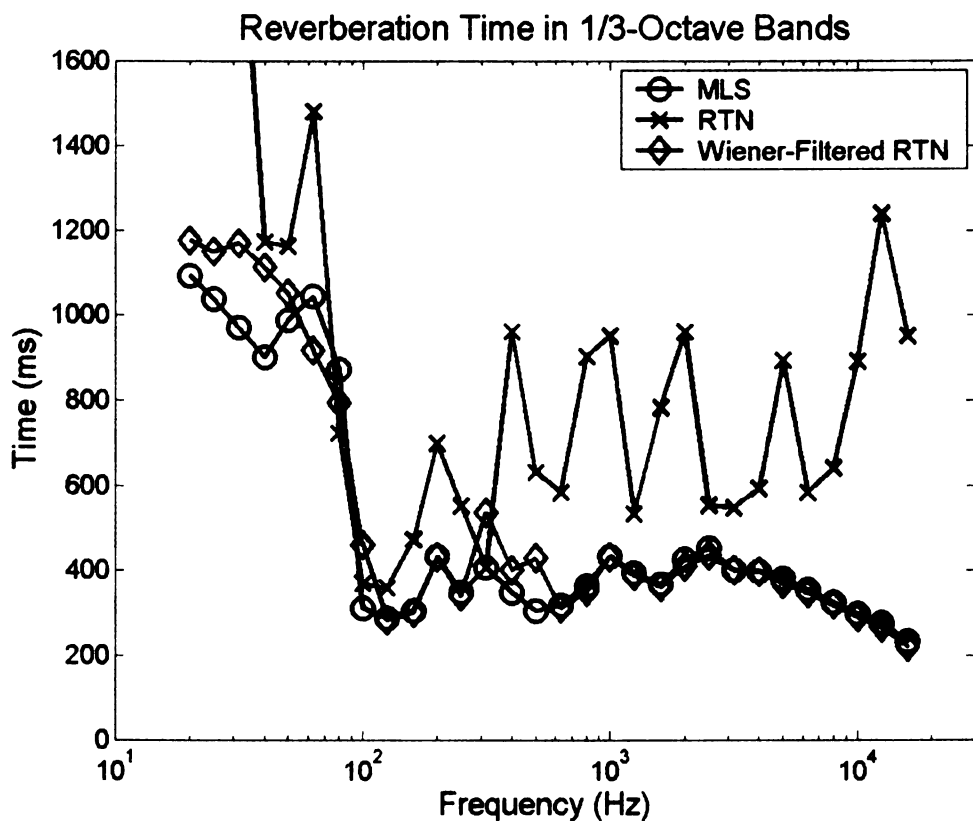


Figure 2.11. Measurements of the 60-dB reverberation time in the same space as that referred to in Figure 2.10. Open circles represent times calculated from MLS measurements, exes mark times calculated from RTN measurements, and open diamonds mark times calculated from Wiener-filtered RTN measurements.

made with MLS because both MLS and the Wiener-filtered RTN measurements have relatively little noise in the high-time part of the IR. Noise power in the high-time part of the IR will distort measurements of RT_{60} made via Schroeder reverse integration, decreasing the decay slope and thus increasing the estimate of RT_{60} . However, fine structure in the IR, which the Wiener filtering process tends to “smooth out,” is not very important when calculating RT_{60} by the method of reverse integration because small changes point by point will not significantly affect the resulting reverberation time, as long as the overall structure of the IR is preserved.

2.7.3 Conclusions

The Wiener-filtered RTN and MLS agree closely when it comes to calculating RT_{60} . This agreement is likely due to the fact that both methods yield relatively little noise in the long-time component of the IR. Though the Wiener filtering procedure destroys some fine structure information in the RTN-measured IR, this information is not very important in calculating RT_{60} . The effect of this loss is seen, however, in the calculation of the TF. In this case, the RTN-measured TF is closer to that given by the MLS than does the Wiener-filtered RTN. This is again likely because the Wiener-filtering procedure smooths out some parts of the IR that are not noise, but structure. It would seem then that while Wiener filtering may be beneficial in measuring RT_{60} , it is wholly detrimental if applied before the TF is calculated.

Given the results of this chapter, it is hard to deny the overall usefulness of MLS when compared to RTN. The MLS performs better than RTN at measuring impulse responses, transfer functions, and reverberation times of rooms, and does so using signals of shorter length. Even when using filtering techniques to improve the accuracy of RTN measurements, the performance of MLS is more accurate and more consistent. In the experiments that follow, MLS will be used exclusively for mea-

surements of rooms and binaural properties. Of particular interest in the following chapter will be measurements of interaural coherence.

CHAPTER 3

Experiment 8: Measurements of Envelope and Waveform Interaural Coherences With KEMAR

Measurements of the waveform coherence (Equation 2.4) are relevant in certain aspects of acoustics, and waveform coherence is often quoted in studies of rooms [8, 27, 30, 55, 58, 63]. It has increasingly become an important measure in studies of both human and animal hearing (for example, [13, 14, 12, 29, 31, 34, 42, 45, 51, 60, 84, 88]). Coherence is related to the degree to which listeners can make use of ITDs in localizing sounds [59] and the apparent auditory source width [17, 73, 4, 74]. It is of interest to make measurements of these values in real listening environments. As center frequency increases in fixed-bandwidth analysis, so does the rate of oscillation of the cross-correlation function (this can be seen, for example, in Figure 2.2). At high frequencies, the ear is insensitive to interaural differences in the fine structure of signals, but differences in the envelopes of the signals become important [54, 69, 86]. To see why this is so, consider a calculation of the interaural phase difference (IPD). The ITD and IPD are simply related by the

angular frequency, $\omega = 2\pi f$:

$$\text{ITD} = \frac{\text{IPD}}{\omega}$$

The ITD at a given angle of incidence is roughly constant across frequency, though it tends to decrease slightly with increasing frequency. Thus, as frequency increases, so must the IPD. Writing the ITD as $\text{ITD} = a \sin\theta$ where a is a constant and θ is the angle of incidence,

$$\text{IPD} = 2\pi f a \sin\theta.$$

At some frequency (f), the IPD will equal π . There is then the possibility of a 180° phase error, leading to an ambiguity in whether the angle of incidence is θ or $-\theta$. The lag at which waveform coherence occurs thus becomes an unreliable estimator of ITD. The envelope coherence is then a better estimator at high frequencies. A calculation of the cross-correlation of the envelope of the signals in the left and right ears leads to a measurement of the envelope coherence (Equation 2.5). Measurement of ITD likely depends on detection of the envelope coherence, especially as the auditory system becomes insensitive to the waveform structure at higher frequencies [100, 98]. The envelope coherence is thought to dominate auditory processing of ITDs above 1300 Hz, while waveform coherence dominates below about 1000 Hz. At other frequencies, the trade-off between timing estimates made in the waveform and the envelope is not well understood.

The waveform coherence is relatively straightforward to measure, involving direct manipulation of the measured signals. Measuring the envelope coherence is more complex, and requires transformation of the signals. Therefore, it would be helpful to establish some constraints on the possible values of the envelope coherence based on measured values of the waveform coherence. To this end, MLS signals were presented to a KEMAR in both a normal listening environment and a highly reverberant space. Recordings of the signals were taken from the KEMAR

ears, and the interaural coherences in both the waveforms and the envelopes were calculated in 1/3-octave bands. The coherence in the waveform and in the envelope of any two signals can be calculated as the maximum of the cross-correlation between those signals or between the envelopes of those signals, respectively. In this experiment, the calculation of the cross-correlation functions was limited to a lag of ± 1 ms.

The average cross-correlation between “uncorrelated” signals is zero, but not so for the cross-correlation of the envelope of those signals, $E(t)$. Since the envelope of a signal involves only positive values, the cross-correlation between two envelopes will always be greater than zero. Van de Par and Kohlrausch [87] note that for a Gaussian noise, the probability density function (PDF) of the envelope of the noise is Rayleigh distributed. Then:

$$\frac{\langle E \rangle}{\sqrt{\langle E^2 \rangle}} = \frac{1}{2} \sqrt{\pi} \quad (3.1)$$

Here, angled brackets denote expectation values in time. Let γ_w be the *waveform* coherence and γ_e be the *envelope* coherence of two otherwise uncorrelated (i.e. $\gamma_w = 0$) Gaussian noises. Let $c_{E_L E_R}$ be the cross-correlation between the envelopes of these noises. Since Equation 3.1 holds for both noises,

$$c_{E_L E_R} = \left(\frac{1}{2} \sqrt{\pi} \right)^2 = \frac{\pi}{4} \quad (3.2)$$

Thus, for uncorrelated noises, $\gamma_e = \frac{\pi}{4}$.

Bernstein suggested a power law relationship between the waveform and envelope coherence [11]. A relationship between envelope coherence and waveform coherence may be expected to have the form:

$$\gamma_e = \frac{\pi}{4} + b \gamma_w^n \quad (3.3)$$

Note that n is not necessarily an integer. This equation allows for $\gamma_e = \frac{\pi}{4}$ when $\gamma_w = 0$. Since both γ_w and γ_e must be bounded on a range from 0 to 1, it is expected that $b = 1 - \frac{\pi}{4}$. Bernstein [11] fitted generated Gaussian noise signals to an

equation of the form of Equation 3.3 without constraining the value of parameter b and found fitting parameters of $b = 0.2142$ and $n = 2.2$. Note that $1 - \frac{\pi}{4} = 0.2146$ is quite close to Bernstein's leading multiplicative constant (0.19% difference).

The results of van de Par and Kohlrausch [87] should hold true for measurements made with white Gaussian noise. The current experiment involves the measurement of coherence in 1/3-octave bands through KEMAR in various acoustical environments using MLS. Though MLS was used and not white Gaussian noise, the distribution of amplitudes of a MLS after being passed through a gammatone filter appear to be Gaussian, as shown in Figure 3.1. Moreover, this experiment was performed in a real room, and so some deviation from Equation 3.3 is expected. The goal of this experiment was to compare the coherence relationships measured in real rooms with those found using idealized noise generated on a computer and to arrive at a general equation for predicting the envelope coherence of two signals given their waveform coherence.

3.1 Computer Simulation of Noises and Coherences

It is possible to create a simple but accurate simulation that will provide a basis for the expected relationship between waveform and envelope coherences. Gaussian white noises can be generated on a computer and manipulated such that their waveform coherence is approximately equal to some desired value. Their envelope coherence can then be calculated per Equation 2.5. The process of generating these noises begins with two independent, identically-distributed Gaussian white noises, n_o and n_L . The signal n_L represents the signal in the left ear. The signal in the right ear, n_R is then generated using:

$$n_R = \sqrt{1 - \alpha^2} n_o + \alpha n_L \quad (3.4)$$

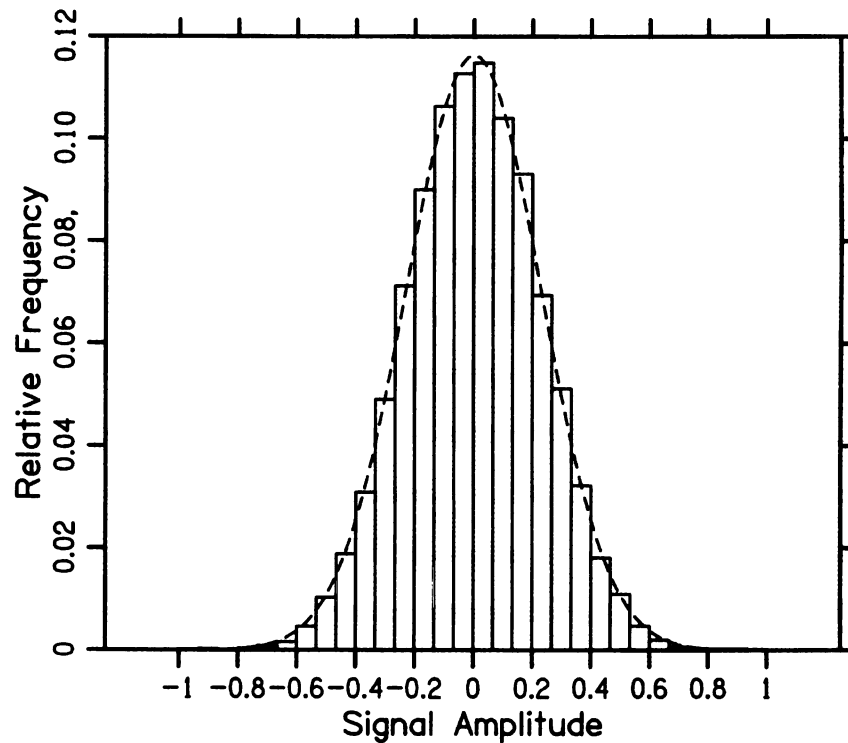


Figure 3.1. A frequency histogram showing the distribution of the amplitude of samples in an order-18 MLS after the signal was passed through a gammatone filter with center frequency 500 Hz and a 1/3-octave bandwidth. The histogram is enveloped by a Gaussian distribution with zero mean, standard deviation 0.229, and scaled by a factor of 0.067, shown as a dashed solid line. This attests to the Gaussian nature of the filtered MLS.

The parameter α is a constant, bounded on the range of $[0, 1]$. Note that a scaled Gaussian random variable is still Gaussian, and the sum of two Gaussian random variables is also Gaussian. Thus, n_R is Gaussian distributed. The cross-correlation function of Equation 2.3 between n_L and n_R can be written as:

$$c_{n_L n_R} = \frac{\int_0^T n_R(t) n_L(t + \tau) dt}{\sqrt{\int_0^T n_R(t)^2 dt \int_0^T n_L(t)^2 dt}}$$

The integrals in the cross-correlation function can be simplified if it is assumed that n_o and n_L are uncorrelated. It is simple then to show that:

$$c_{n_L n_R} = \frac{\alpha \int_0^T n_L(t) n_L(t + \tau) dt}{\int_0^T n_L(t)^2 dt}$$

Thus the cross-correlation function becomes simply the autocorrelation function of n_L , multiplied by α . The waveform coherence is the maximum value of this function. Since the normalized autocorrelation of any real signal is 1, it becomes clear that α is the desired waveform coherence. Thus, by choosing α , one can approximately specify the waveform coherence of n_L and n_R . It should be noted that in practice, n_o and n_L will not be completely uncorrelated if generated randomly. However, in this simulation, noises with a wide range of coherences will be generated, and the exact value of the coherence for each individual noise pair is unimportant.

To create a set of noise pairs with a wide range of coherences thoroughly covering the range of possible waveform coherences, 1001 values of α were used between 0 and 1. For each value of α , a pair of Gaussian noises were generated and treated in the manner described above to create a “left” and a “right” signal. These noises were filtered by 1/3-octave band gammatone filters with ISO center frequencies between 160 and 10000 Hz and their waveform and envelope coherences were calculated within each band. Then, a nonlinear regression was performed on the coherence data, fitting them to an equation of the form of Equation 3.3 but with

a restricted multiplicative constant:

$$\gamma_e = \frac{\pi}{4} + \left(1 - \frac{\pi}{4}\right)\gamma_w^n \quad (3.5)$$

The nonlinear regression yielded a best-fit value for the power parameter n . A good measure of how well the data fit Equation 3.5 within bands is the RMS error between the data and the fitting equation.

These simulations resulted in a set of 1001 waveform-envelope coherence pairs per 1/3-octave band, and a best fit to Equation 3.5 was found. The data and best fitting lines for six representative bands are shown in Figure 3.2. A complete set of values for the power parameter n and its respective RMS error is shown in Table 3.1.

Two are particularly worth noting in the data gathered from the simulated Gaussian noise pairs. First, the value of n does not vary significantly across frequency bands and has an average value of about $n = 2.11$. This value is quite close to that determined by Bernstein [11] ($n = 2.2$). A derivation of the analytic relationship between the envelope and waveform coherences is difficult to obtain. Van de Par and Kohlrausch made progress in this derivation (though it was not their original intent), which led to the $\pi/4$ factors in Equation 3.5 [88]. Further progress in arriving at an analytical derivation of this relationship is complicated by the nature of the integrals, which involve the absolute value function when calculating the envelope coherence (see Equation 2.5). It is tempting then to theorize that perhaps a value of exactly $n = 2$ might result from an analytic derivation of the envelope coherence as a function of waveform coherence. Given the results of the simulated coherences however, the value of n is statistically almost certainly not 2 (one-sample t -test, $N = 19$, $t = 32.75$, $p < 0.001$).

The spread of the data about the best fit curve is larger at low values of the waveform coherence. The extent of this deviation is related to the RMS error. Both Figure 3.2 and Table 3.1 show that the deviation of low-coherence points from the

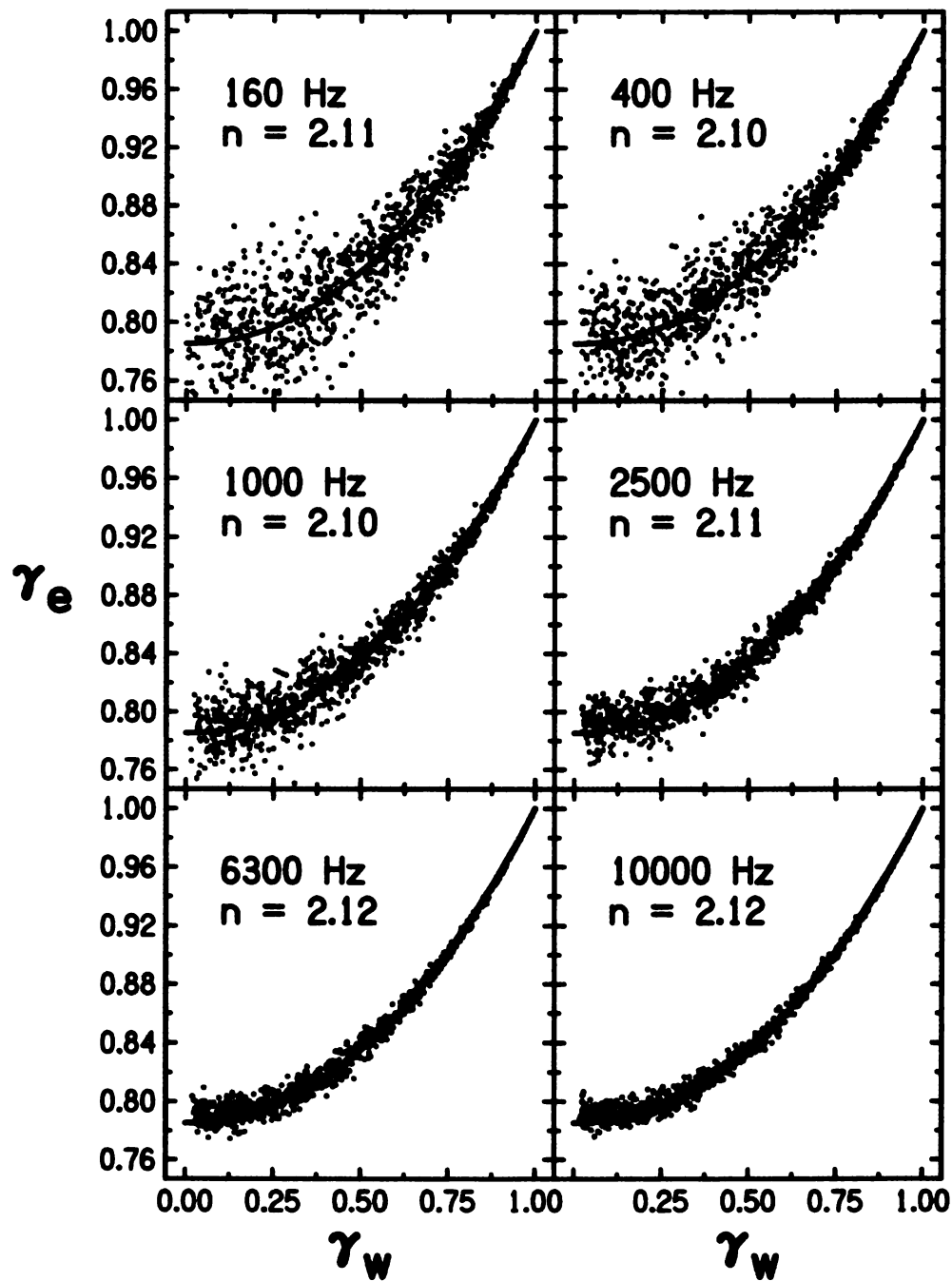


Figure 3.2. Plots of envelope coherence versus waveform coherence for simulated Gaussian noise pairs in six different 1/3-octave bands. Each band contains 1001 coherence pairs, each of which is a single data point. A best fitting line to Equation 3.5 is shown as a thick solid line in each plot, though it may be obscured by the data points. The value of the power parameter n for each set of points is shown in each panel.

Frequency (Hz)	n	RMS Error
160	2.11 ± 0.048	0.020
200	2.08 ± 0.045	0.019
250	2.08 ± 0.040	0.017
315	2.11 ± 0.039	0.016
400	2.10 ± 0.038	0.016
500	2.12 ± 0.035	0.014
630	2.10 ± 0.032	0.013
800	2.11 ± 0.029	0.012
1000	2.10 ± 0.027	0.011
1250	2.12 ± 0.024	0.011
1600	2.10 ± 0.023	0.0095
2000	2.13 ± 0.021	0.0086
2500	2.11 ± 0.019	0.0079
3150	2.13 ± 0.017	0.0069
4000	2.11 ± 0.015	0.0062
5000	2.13 ± 0.014	0.0057
6300	2.12 ± 0.012	0.0052
8000	2.12 ± 0.011	0.0046
10000	2.12 ± 0.011	0.0045
Avg.	2.111 ± 0.0071	0.011

Table 3.1. Values of the power parameter n from a fit of the coherence data within each 1/3-octave band to Equation 3.5 for randomly generated Gaussian noises. The bounds about each n value are 95% confidence intervals. As expected, the RMS error decreases as the center frequency, and thus the width of the noise band, increases. The errors indicate a general trend toward a better fit to Equation 3.5 with increasing frequency.

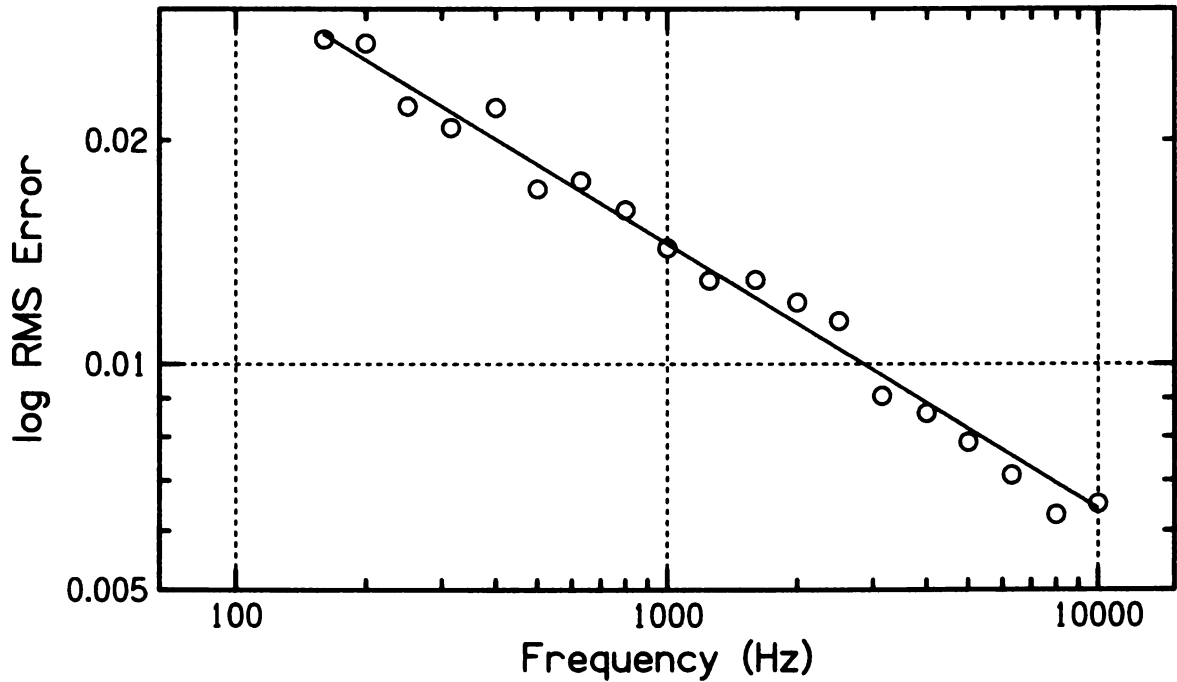


Figure 3.3. RMS Error as a function of center frequency. The trend line shows a decrease in RMS Error proportional to $1/f^{0.35}$.

best fit curve decreases with increasing frequency. The RMS error has a high negative correlation with the center frequency of the band ($r = -0.814$). The RMS error as a function of center frequency is plotted in Figure 3.3, and can be fit to an equation of the form:

$$\text{RMS Error} = \frac{0.17}{f^{0.35}}$$

The results of these simulated coherences forms a basis for comparison to coherences measured in real rooms.

3.2 Methods

To compare the relationship between waveform and envelope coherences in a real room, a KEMAR manikin and Mackie HR824 studio monitor, seated on a padded chair so as to be at the ear level of the KEMAR, were situated in a room. The Mackie

HR824 is a powered, high-accuracy, two-way loudspeaker with an especially flat response curve over a wide frequency range, chosen for these properties. Two rooms were used. One was a fairly open lab space, 6.5 m by 7.5 m by 4.5 m high (this is room 10B, cited in Hartmann, *et al.* [51]). The broadband RT_{60} of room 10B was approximately 0.8 s. The other room was a large reverberant chamber, with dimensions 7.67 m by 6.35 m by 3.58 m high (the same reverberation room as used by Hartmann, *et al.* [51]). The broadband RT_{60} of the reverberant room was approximately 2.0 s.

A set of four distances between the KEMAR ears and the loudspeaker was used: 0.5, 1.0, 3.0, and 5.0 meters. The KEMAR and loudspeaker were always situated so as to be “facing” each other. For each measurement, a different order-18 MLS was played through the loudspeaker at a sampling rate of 195,312.5 Hz (set by the TDT hardware). The response was recorded through the ears of the KEMAR on two channels (left and right). Then, the interaural waveform and envelope coherences between the left and right ears were calculated in nineteen 1/3-octave bands from 80 to 10,000 Hz with ISO standard center frequencies. This constituted a single trial at a given distance. In subsequent trials, the procedure was repeated, but the KEMAR and loudspeaker were moved to different places within the room, maintaining the same fixed distance between them.

For each distance, 11 trials were taken, each at a different place within the room. This resulted in 11 measurements of interaural coherence within each 1/3-octave band for each room for a total of $19 \times 11 = 209$ pairs of measurements of envelope and waveform interaural coherence. Nonlinear regression to Equation 3.5 was then performed in a manner similar to that done for the simulated noises. This was done for each distance within each room both within and across bands.

Errors in the measurement of waveform coherence occurred when the peak in the cross-correlation function did not occur at the “correct” ITD. Problems of this

kind occurred most often at high frequencies, where multiple peaks in the cross-correlation function have similar heights. Also, as noted in the introductory passage on the cross-correlation function and coherence measurements, the temporal resolution of the cross-correlation function is limited by the sampling frequency. At higher frequencies the true peak of the cross-correlation function is more likely to fall in between samples and falls off fast enough that a nearby peak, which has a similar amplitude and may occur closer to a sample, may be mistaken for the peak of the function. This is known as a “period error” since a peak one or more periods away from the true peak is the largest. In practice, the difference between the amplitudes of the measured peak and the true peak in the case of a period error was very small. They were detected and corrected for by looking at the estimate of ITD that comes from such measurements and determining if it is very different than the estimate that was given by the next lower frequency band. This relies on the lowest frequency bands being free of such errors, but period errors are not possible in low-frequency bands since the lag of the cross-correlation function is limited to ± 1 ms. Such errors were far fewer in the measurement of the envelope coherence.

3.3 Analysis of Results Within Bands

For each distance, in each room, and within each 1/3-octave band, the behavior of the relationship between waveform and envelope interaural coherence varied significantly across different locations of the KEMAR and loudspeaker. Plots of the data taken in twelve of the nineteen bands for each distance (0.5, 1.0, 3.0, and 5.0 meters) in each room (10B and the reverberant room) are shown in Figures 3.4 through 3.11.

Equation 3.5 fitted the data remarkably well in most cases. There are, however, large deviations of the data from the predicted trend as distance from the loud-

10B at 0.5 m

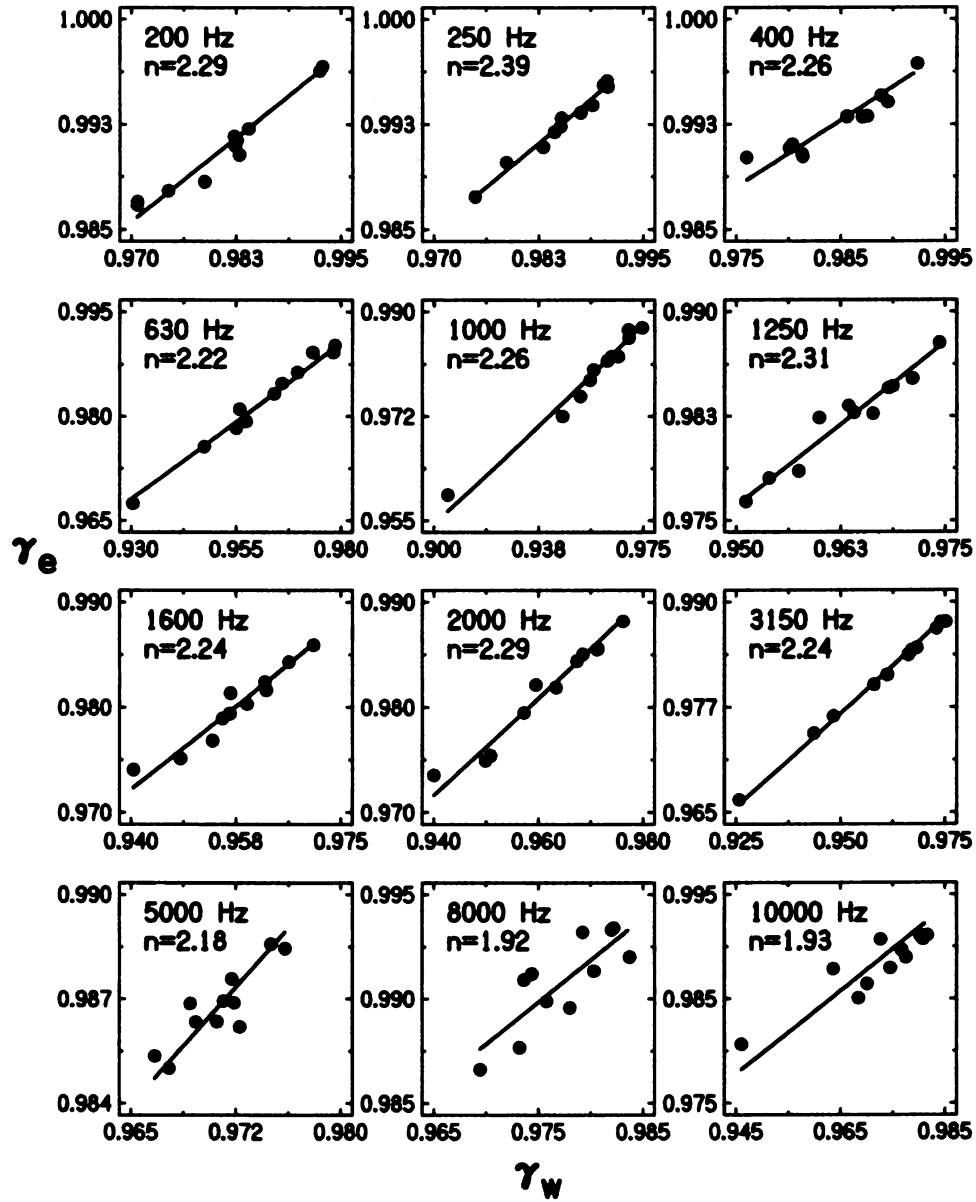


Figure 3.4. Plots of waveform coherence versus envelope coherence as measured through KEMAR in room 10B at a distance of 0.5 m to the loudspeaker. Each panel shows the data measured in a particular 1/3-octave band with center frequency indicated in the top left corner of the panel. A best fit line is shown in each panel, which fits the data within the indicated band to Equation 3.5. The best fit value of the power parameter n for the data in each band is shown in the upper left of each panel. Note that the horizontal and vertical scales differ in each panel.

10B at 1 m

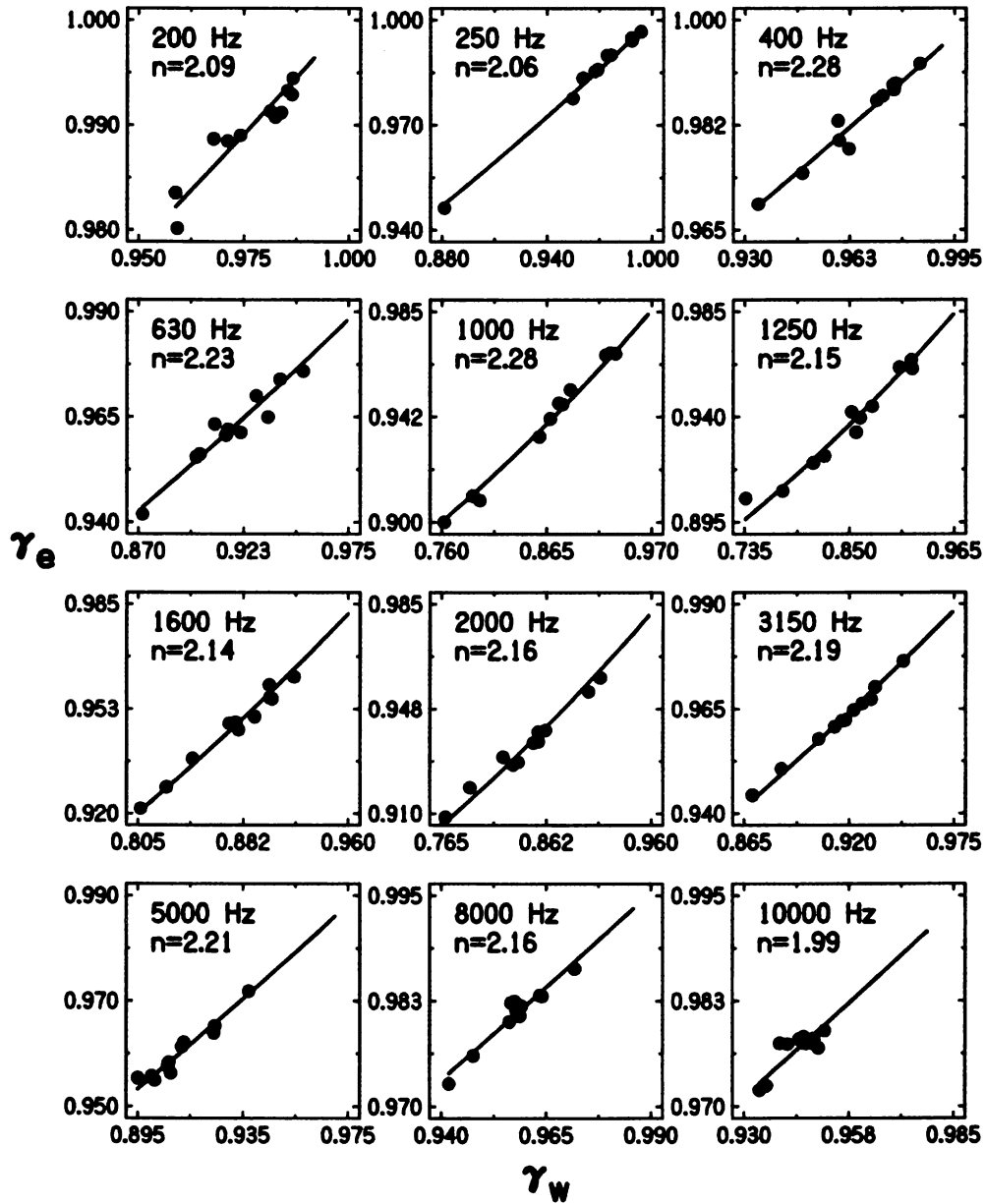


Figure 3.5. Plots of waveform coherence versus envelope coherence as measured through KEMAR in room 10B at a distance of 1.0 m to the loudspeaker, in a manner similar to that of Figure 3.4

10B at 3 m

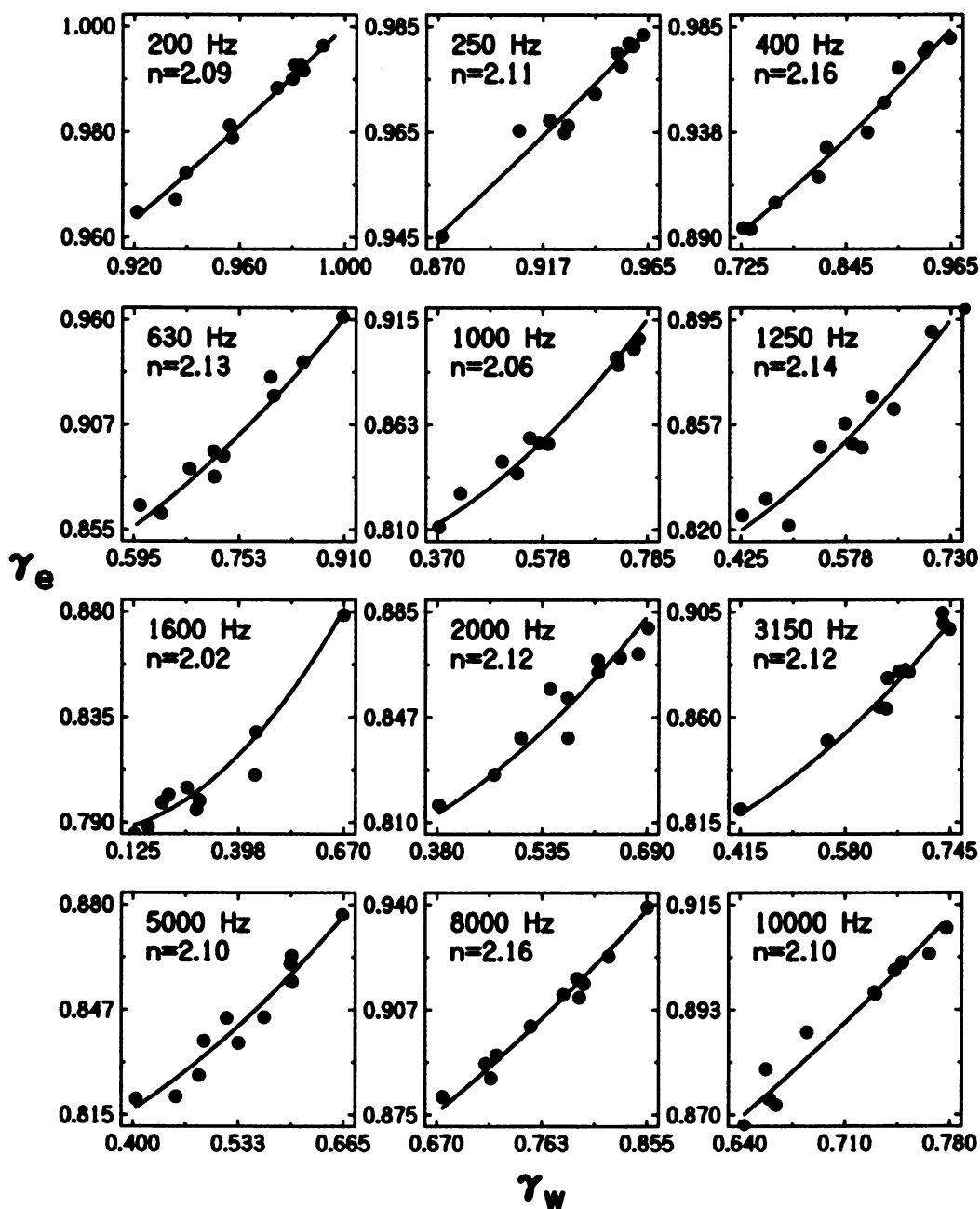


Figure 3.6. Plots of waveform coherence versus envelope coherence as measured through KEMAR in room 10B at a distance of 3.0 m to the loudspeaker, in a manner similar to that of Figure 3.4

10B at 5 m

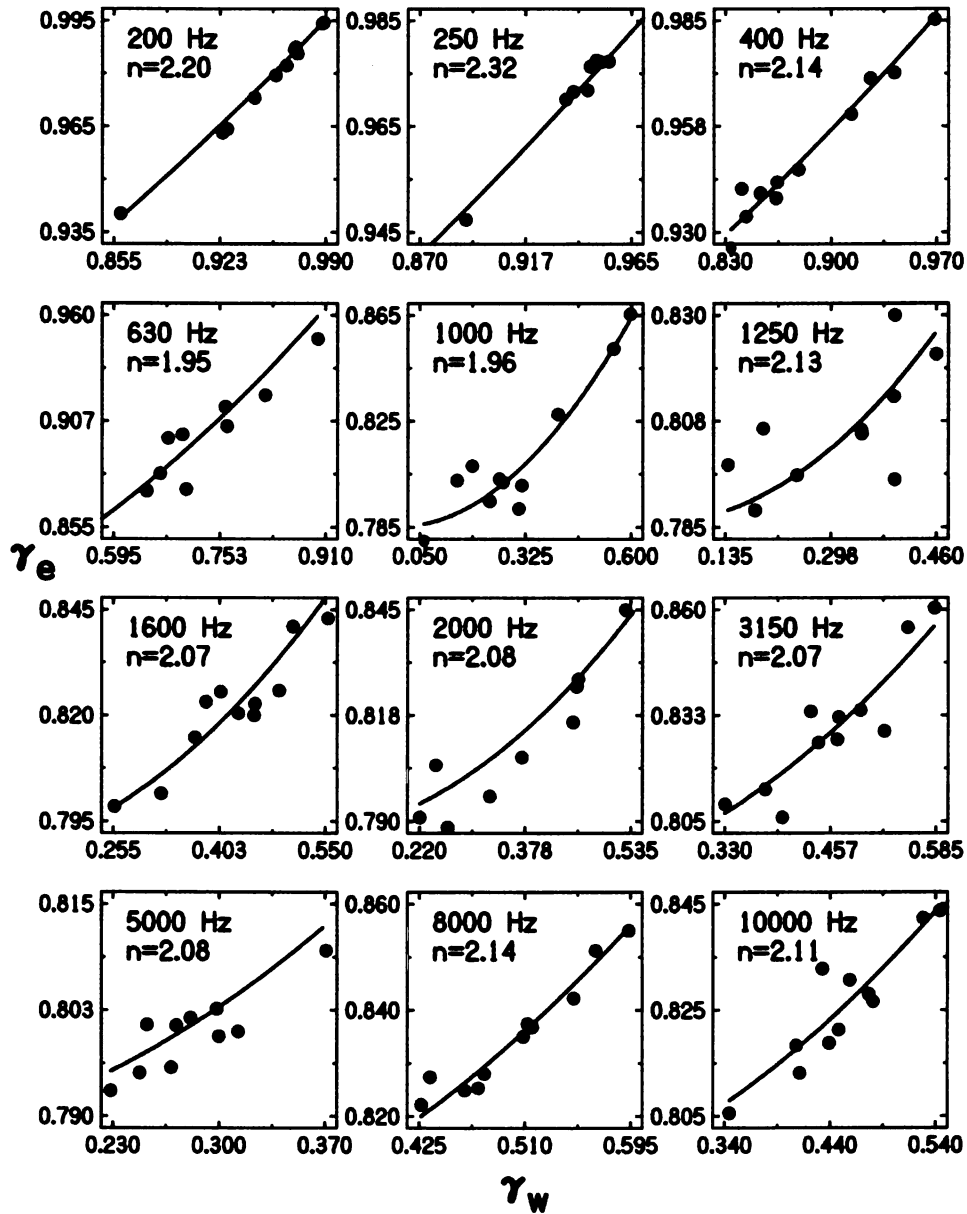


Figure 3.7. Plots of waveform coherence versus envelope coherence as measured through KEMAR in room 10B at a distance of 5.0 m to the loudspeaker, in a manner similar to that of Figure 3.4

RR at 0.5 m

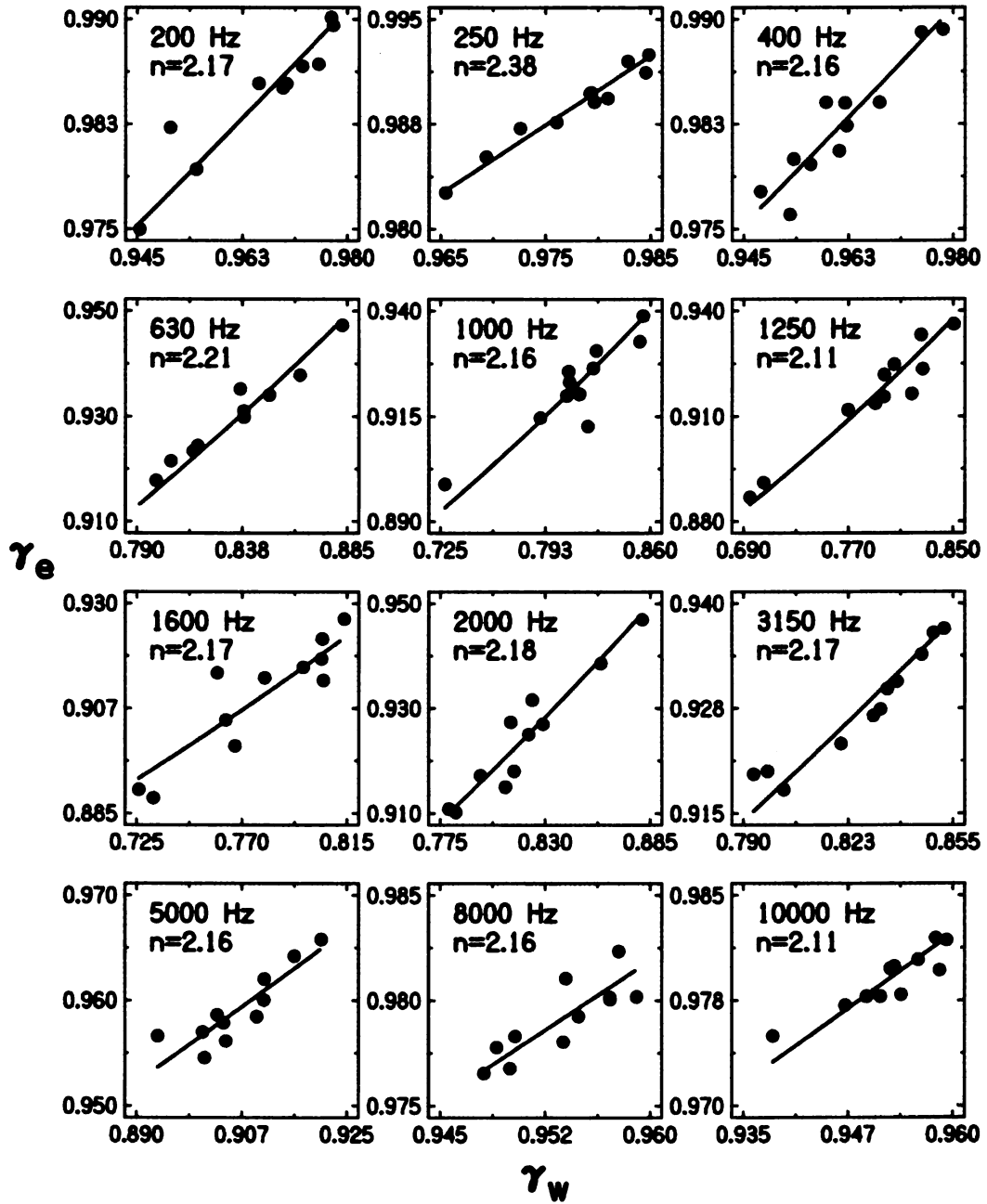


Figure 3.8. Plots of waveform coherence versus envelope coherence as measured through KEMAR in a reverberant room at a distance of 0.5 m to the loudspeaker, in a manner similar to that of Figure 3.4

RR at 1 m

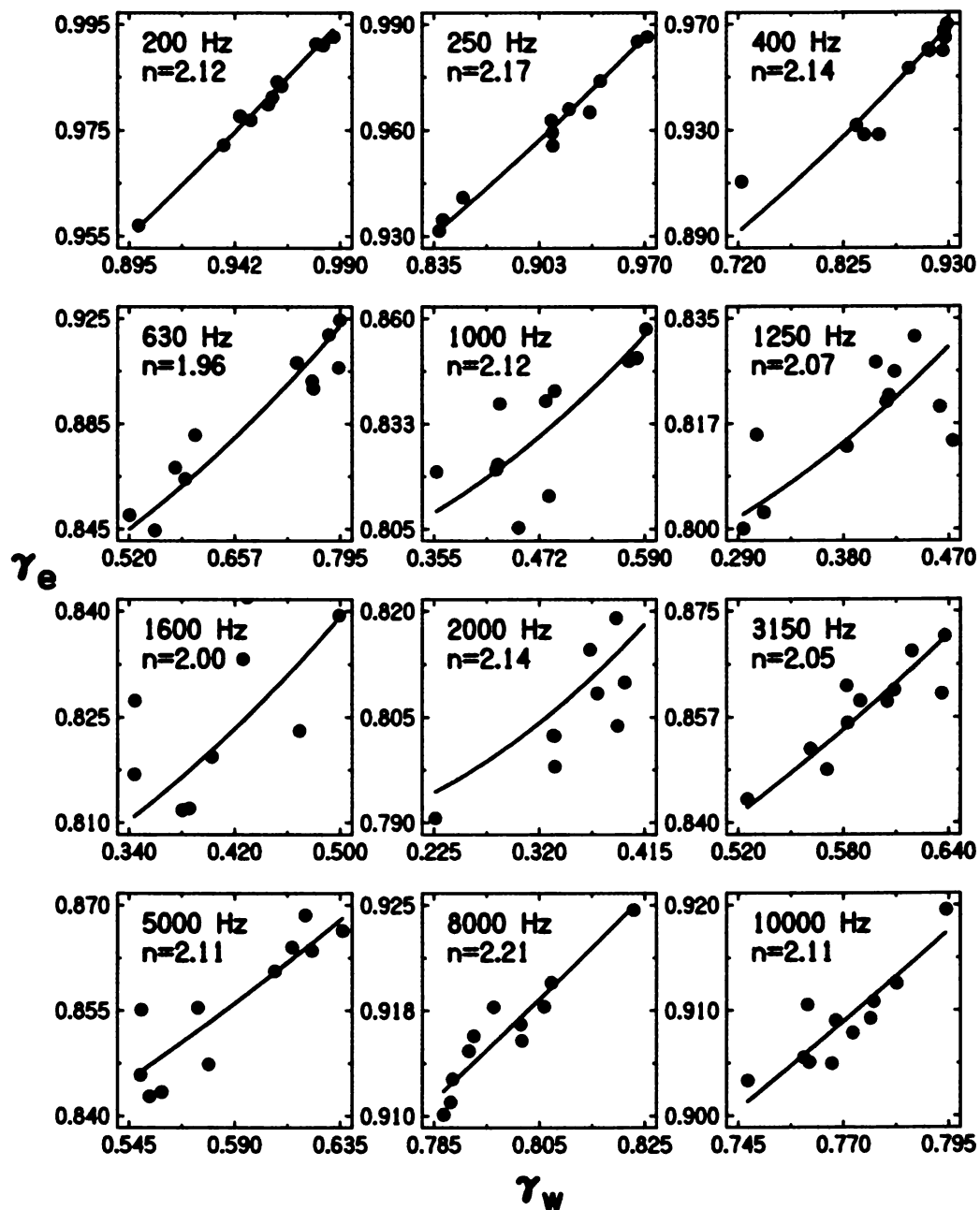


Figure 3.9. Plots of waveform coherence versus envelope coherence as measured through KEMAR in a reverberant room at a distance of 1.0 m to the loudspeaker, in a manner similar to that of Figure 3.4

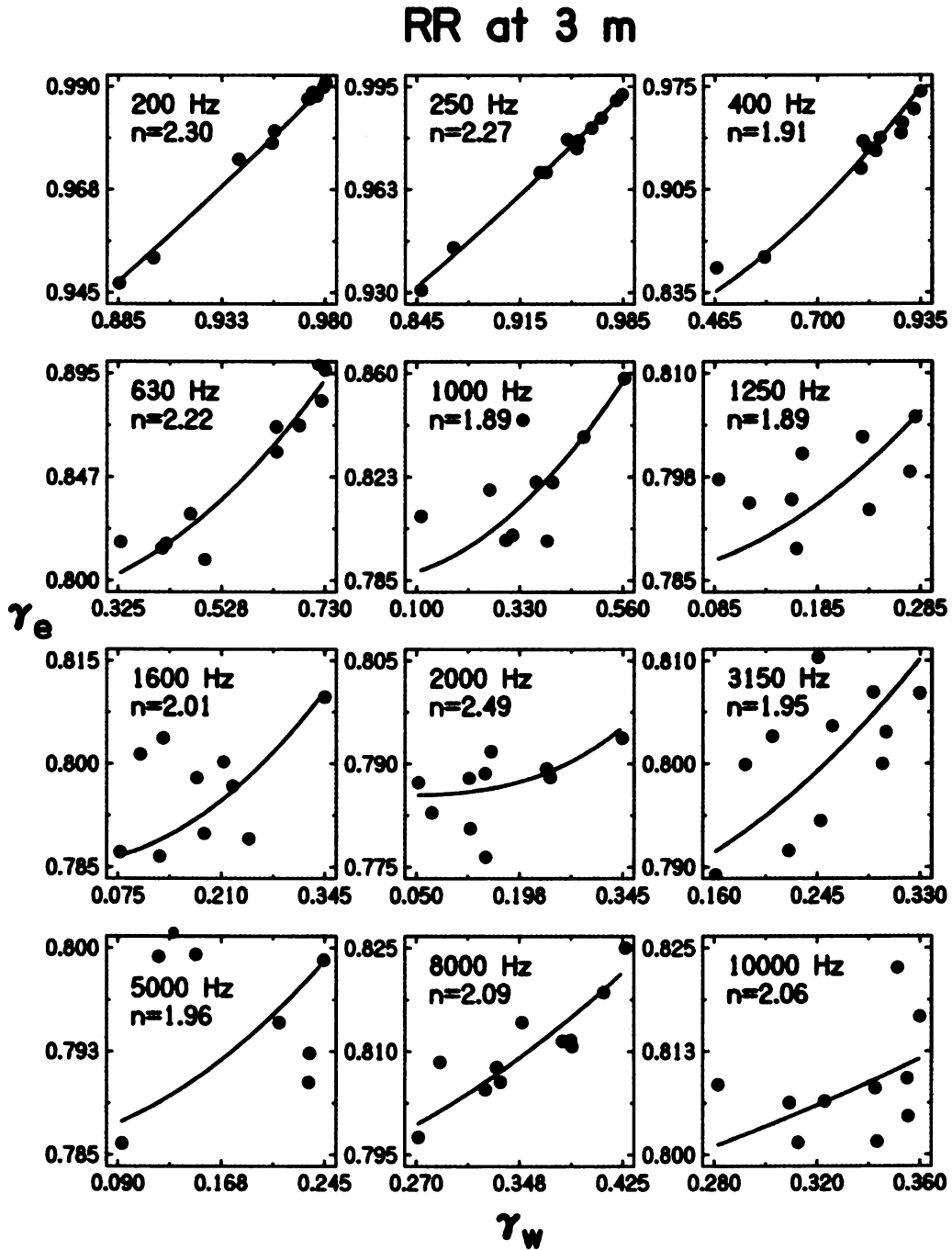


Figure 3.10. Plots of waveform coherence versus envelope coherence as measured through KEMAR in a reverberant room at a distance of 3.0 m to the loudspeaker, in a manner similar to that of Figure 3.4

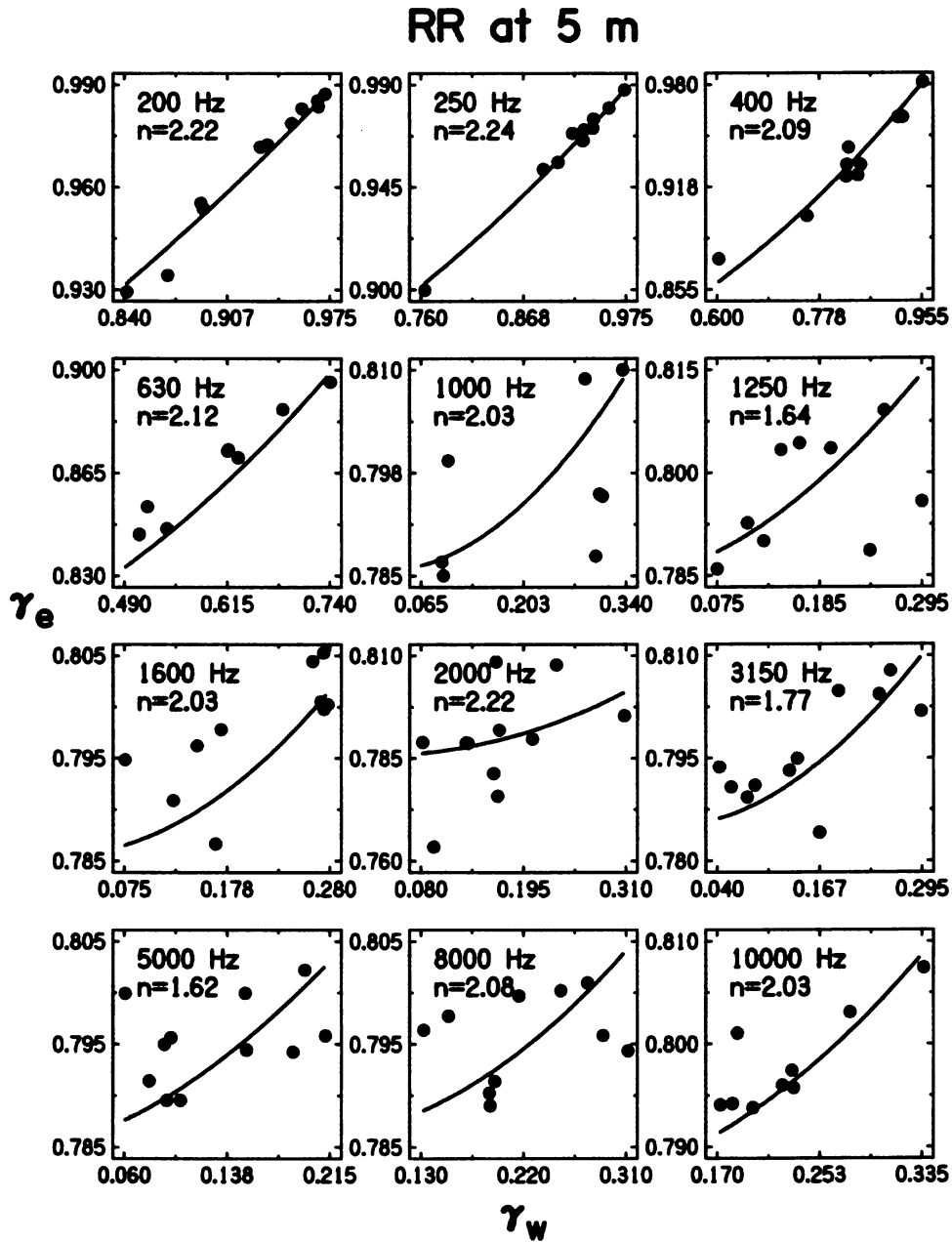


Figure 3.11. Plots of waveform coherence versus envelope coherence as measured through KEMAR in a reverberant room at a distance of 5.0 m to the loudspeaker, in a manner similar to that of Figure 3.5

speaker and reverberation times increase. In room 10B, the data deviate far from the best fit of Equation 3.5 in some bands (e.g. 1250 Hz) when the distance is increased to 5.0 m (Figure 3.7). In the reverberant room, some data differ noticeably from the fitting equation even at the relatively short distance of 1.0 m. Deviation from the fitting equation tends to be greater in situations where the waveform coherence is small, as is more often the case in reverberant environments. This can be seen in the reverberant room data in, for example, a distance of 5.0 m in every 1/3-octave band above 630 Hz.

The values of the power parameter n and the RMS error of the fit for each condition and each 1/3-octave band can be found in Tables 3.2 through 3.5. Several things can be noted immediately. First, most values of the RMS error are quite small. On average, percents error increase with distance from the sound source and are larger in the reverberant room. The largest single RMS error, 0.016, was measured in the 800 Hz band in room 10B at a source distance of 5.0 m. On average, the largest errors occur in the reverberant room at a distance of 5.0 m. In general, even the largest deviations between measured data and curve fits are relatively small.

RMS errors, and thus deviations of the data from the model of Equation 3.5, tend to get larger as the distance to the source increases. In room 10B, RMS errors are significantly larger at 3.0 m than at 1.0 m (one-sided, two-sample t-test, $t = -5.61$, $df = 24$, $p < 0.001$) and larger at 1.0 m than at 0.5 m ($t = -3.64$, $df = 20$, $p = 0.001$). However, there is no significant difference between the errors at 3.0 m and at 5.0 m. In the reverberant room, the errors are smaller at a distance of 0.5 m than at 1.0 m ($t = -3.52$, $df = 28$, $p = 0.001$), but there is no significant difference between the errors at distances of 1.0, 3.0, and 5.0 m (One-way ANOVA, $F(2, 56) = 0.36$, $p = 0.700$). This is likely because errors do not increase appreciably in highly reverberant environments once the sound source is in the diffuse sound field. The

Room 10B				
Distance	0.5 m		1.0 m	
Frequency (Hz)	n	RMS Error	n	RMS Error
160	2.03 ± 0.089	0.00056	2.09 ± 0.17	0.0011
200	2.29 ± 0.13	0.00073	2.06 ± 0.19	0.0014
250	2.39 ± 0.078	0.00036	2.22 ± 0.0940	0.0011
315	2.07 ± 0.13	0.000048	2.25 ± 0.16	0.0012
400	2.26 ± 0.16	0.00078	2.28 ± 0.14	0.0014
500	2.28 ± 0.068	0.00048	2.08 ± 0.049	0.00058
630	2.22 ± 0.078	0.00089	2.23 ± 0.12	0.0026
800	2.31 ± 0.12	0.0011	2.07 ± 0.11	0.0045
1000	2.26 ± 0.11	0.0014	2.28 ± 0.083	0.0026
1250	2.31 ± 0.073	0.00077	2.15 ± 0.12	0.0040
1600	2.24 ± 0.077	0.00094	2.14 ± 0.067	0.0020
2000	2.29 ± 0.085	0.00096	2.16 ± 0.077	0.0027
2500	2.24 ± 0.024	0.00035	2.26 ± 0.057	0.0019
3150	2.24 ± 0.036	0.00042	2.19 ± 0.039	0.00087
4000	2.18 ± 0.051	0.00047	2.22 ± 0.055	0.0010
5000	2.18 ± 0.071	0.00059	2.21 ± 0.054	0.0012
6300	2.15 ± 0.094	0.00091	2.13 ± 0.052	0.0012
8000	1.92 ± 0.18	0.0012	2.16 ± 0.079	0.00096
10000	1.9 ± 0.20	0.0018	1.99 ± 0.086	0.0013
Avg.	2.20 ± 0.062	0.00080	2.17 ± 0.040	0.00178

Table 3.2. Values of the power parameter n with 95% confidence intervals for the data in each 1/3-octave band measured in room 10B at distances of 0.5 and 1.0 m. These values were found from the best fit of the data within that band to Equation 3.5, and the RMS differences between the data and the best fitting line are given.

Room 10B				
Distance	3.0 m		5.0 m	
Frequency (Hz)	n	RMS Error	n	RMS Error
160	2.1 ± 0.23	0.0020	2.11 ± 0.12	0.0016
200	2.25 ± 0.12	0.0014	2.20 ± 0.14	0.0022
250	2.11 ± 0.13	0.0024	2.32 ± 0.075	0.0013
315	2.13 ± 0.15	0.0034	2.00 ± 0.14	0.0029
400	2.16 ± 0.12	0.0042	2.14 ± 0.11	0.0033
500	1.94 ± 0.21	0.0098	1.86 ± 0.11	0.0051
630	2.13 ± 0.17	0.0080	2.0 ± 0.20	0.010
800	1.99 ± 0.13	0.0064	2.1 ± 0.37	0.016
1000	2.06 ± 0.11	0.0052	2.0 ± 0.23	0.0081
1250	2.14 ± 0.13	0.0065	2.1 ± 0.37	0.012
1600	2.0 ± 0.28	0.0096	2.1 ± 0.20	0.0086
2000	2.12 ± 0.13	0.0065	2.1 ± 0.30	0.012
2500	2.03 ± 0.19	0.0088	2.0 ± 0.27	0.0064
3150	2.12 ± 0.14	0.0069	2.07 ± 0.15	0.0070
4000	2.12 ± 0.076	0.0039	2.19 ± 0.083	0.0034
5000	2.10 ± 0.089	0.0044	2.1 ± 0.21	0.0059
6300	2.07 ± 0.094	0.0050	2.13 ± 0.16	0.0050
8000	2.16 ± 0.057	0.0026	2.14 ± 0.053	0.0025
10000	2.1 ± 0.062	0.0031	2.11 ± 0.11	0.0050
Avg.	2.10 ± 0.027	0.0053	2.09 ± 0.050	0.0063

Table 3.3. Values of the power parameter n with 95% confidence intervals for the data in each 1/3-octave band measured in room 10B at distances of 3.0 and 5.0 m. These values were found from the best fit of the data within that band to Equation 3.5, and the RMS differences between the data and the best fitting line are given.

Reverberant Room				
Distance	0.5 m		1.0 m	
Frequency (Hz)	n	RMS Error	n	RMS Error
160	2.2 ± 0.28	0.0028	2.10 ± 0.11	0.0010
200	2.17 ± 0.17	0.0018	2.12 ± 0.080	0.0011
250	2.38 ± 0.093	0.00062	2.17 ± 0.11	0.0027
315	2.24 ± 0.13	0.00093	2.15 ± 0.16	0.0028
400	2.16 ± 0.13	0.0015	2.1 ± 0.25	0.0073
500	2.30 ± 0.13	0.0029	2.32 ± 0.15	0.0051
630	2.21 ± 0.087	0.0033	1.96 ± 0.17	0.0090
800	2.10 ± 0.15	0.0050	2.2 ± 0.27	0.012
1000	2.16 ± 0.12	0.0049	2.1 ± 0.23	0.011
1250	2.11 ± 0.10	0.0023	2.07 ± 0.19	0.0077
1600	2.17 ± 0.12	0.0052	2.0 ± 0.23	0.010
2000	2.18 ± 0.085	0.0033	2.1 ± 0.33	0.012
2500	2.20 ± 0.073	0.0032	2.08 ± 0.19	0.0089
3150	2.17 ± 0.053	0.0020	2.05 ± 0.071	0.0038
4000	2.16 ± 0.033	0.00098	2.12 ± 0.10	0.0053
5000	2.16 ± 0.066	0.0016	2.11 ± 0.080	0.0042
6300	2.17 ± 0.036	0.00077	2.14 ± 0.084	0.00043
8000	2.16 ± 0.074	0.00097	2.21 ± 0.033	0.0014
10000	2.11 ± 0.078	0.0011	2.11 ± 0.046	0.0021
Avg.	2.19 ± 0.032	0.0025	2.12 ± 0.040	0.0059

Table 3.4. Values of the power parameter n with 95% confidence intervals for the data in each 1/3-octave band measured in a reverberant room at distances of 0.5 and 1.0 m. These values were found from the best fit of the data within that band to Equation 3.5, and the RMS differences between the data and the best fitting line are given.

Reverberant Room				
Distance	3.0 m		5.0 m	
Frequency (Hz)	n	RMS Error	n	RMS Error
160	2.0 ± 0.25	0.0030	2.1 ± 0.22	0.0045
200	2.30 ± 0.081	0.0012	2.22 ± 0.16	0.0034
250	2.27 ± 0.092	0.0017	2.24 ± 0.11	0.0026
315	2.0 ± 0.27	0.0077	2.16 ± 0.10	0.0028
400	1.9 ± 0.21	0.0085	2.1 ± 0.21	0.0074
500	2.36 ± 0.18	0.0076	2.1 ± 0.24	0.011
630	2.2 ± 0.21	0.0094	2.1 ± 0.24	0.012
800	1.8 ± 0.22	0.0096	2.0 ± 0.34	0.014
1000	1.9 ± 0.36	0.015	2.0 ± 0.44	0.011
1250	1.9 ± 0.48	0.012	1.8 ± 0.45	0.016
1600	2.0 ± 0.43	0.0081	2.0 ± 0.46	0.0099
2000	2.2 ± 0.45	0.0080	2.2 ± 0.51	0.011
2500	2.1 ± 0.49	0.0079	2.1 ± 0.30	0.0056
3150	1.95 ± 0.19	0.0056	1.8 ± 0.23	0.0056
4000	1.8 ± 0.23	0.0062	1.9 ± 0.37	0.0078
5000	2.0 ± 0.40	0.0081	1.6 ± 0.22	0.0054
6300	2.0 ± 0.30	0.0049	1.6 ± 0.30	0.0076
8000	2.09 ± 0.093	0.0034	2.1 ± 0.36	0.0074
10000	2.06 ± 0.17	0.0059	2.03 ± 0.19	0.0041
Avg.	2.04 ± 0.080	0.0070	2.01 ± 0.090	0.0079

Table 3.5. Values of the power parameter n with 95% confidence intervals for the data in each 1/3-octave band measured in a reverberant room at distances of 3.0 and 5.0 m. These values were found from the best fit of the data within that band to Equation 3.5, and the RMS differences between the data and the best fitting line are given.

distance from the detector at which the diffuse field “begins” varies with room, and is expected to scale inversely with reverberation time. In the diffuse field, the sound intensity no longer depends on the source distance to the detector, interaural coherence is generally poor, and localization is quite difficult if not impossible. A reasonable expectation is that the distribution of coherences also does not change appreciably in the diffuse field, and so the related RMS error should plateau at some maximum value, which should depend on the room.

For any given source distance, the RMS error tends to be smallest at both the lowest and the highest frequencies, and is greatest at some intermediate frequency. For example, Figure 3.12 plots the RMS error within 1/3-octave bands across center frequency for the reverberant room at a distance of 3.0 m. A low RMS error in any particular band is due to the tendency for coherences to be large within that band since Equation 3.5 tends to agree uniformly well with high waveform coherences. At low frequencies, the ears may not be in a diffuse sound field, and so waveform coherences are high. At high frequencies, large waveform coherences tend to occur because the room is more anechoic than at lower frequencies.

Average values of the power parameter n and of the RMS error calculated across bands at each source distance for each room are shown in the bottom rows of Tables 3.2 through 3.5. In room 10B, there is no significant difference between the average value of n found at a distance of 0.5 m and that found at 1.0 m (two-sided, two-sample t-test, $t = 0.93$, $df = 30$, $p = 0.357$), and there is no significant difference between 3.0 m and 5.0 m in this respect ($t = 0.37$, $df = 31$, $p = 0.711$). However, the values of n measured at 0.5 m and 1.0 m are significantly larger on average than those measured at 3.0 m and 5.0 m ($t = 4.08$, $df = 70$, $p < 0.001$). Similarly, neither the mean difference between n values in each band measured at 0.5 m and at 1.0 m (two-sided, paired t-test, $t = 1.07$, $p = 0.297$), nor the differences between 3.0 m and 5.0 m ($t = 0.52$, $p = .304$) are statistically significant.

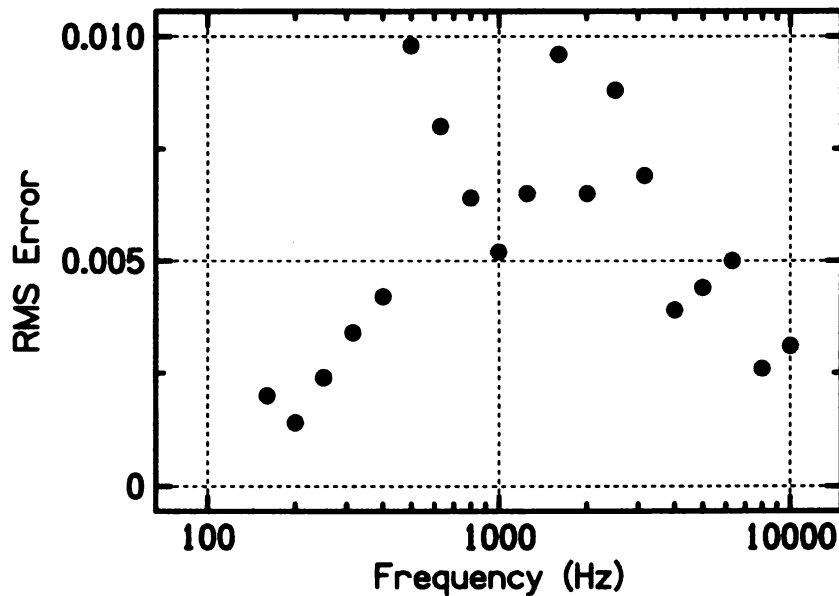


Figure 3.12. RMS Error as a function of 1/3-octave band center frequency for coherences measured in the reverberant room at a source distance of 3.0 m.

However, the n values at 1.0 m are significantly larger than those at 3.0 m ($t = 3.02$, $p = 0.004$). In the reverberant room, the average value of n across bands is significantly larger at 0.5 m than at 1.0 m ($t = 2.63$, $df = 34$, $p = 0.006$) and larger at 1.0 m than at 3.0 m ($t = 1.93$, $df = 25$, $p = 0.033$), but there is no difference on average between 3.0 m and 5.0 m ($t = 0.51$, $df = 35$, $p = 0.308$). These averages and the statistical significance are summarized in Table 3.6, where values of n that are not significantly different within each room are grouped by brackets.

Except for the difference on average between 0.5 m and 1.0 m, the tendency of n to decrease with increasing source distance is the same as in room 10B. It should be noted, however, that this is an *average* tendency across bands and is not necessarily true for individual bands. Consider, for example, the behavior of the 500 Hz band in room 10B. The value of n measured in that band monotonically decreases with increasing source distance. The 10,000 Hz band in room 10B, however, shows n values that consistently increase with increasing source distance. Some bands, such as the 250 Hz band, neither increase or decrease monotonically with chang-

Distance	10B	RR
0.5	2.20 ± 0.062	2.19 ± 0.032
1.0	2.17 ± 0.040	2.12 ± 0.040
3.0	2.10 ± 0.027	2.04 ± 0.080
5.0	2.09 ± 0.050	2.01 ± 0.090

Table 3.6. Summary of average values of the power parameter n calculated across fits to Equation 3.5 within 1/3-octave bands in both room 10B and the reverberant room at each distance. Values of n within each room that are not significantly different from each other are grouped by brackets.

ing source distance. Similarly, there seems to be no clear trend in the values of n with band center frequency. This point is exemplified by Figure 3.13. These plots show values of n across frequency and source distance for each room, and betray no apparent trends in the value of n either across frequency or source distance.

3.4 Analysis of Data Combined Across Bands

If the data across all 1/3-octave bands for a given source distance are combined, they form a more complete picture of the relationship between waveform and envelope coherence. By pooling all data points in all bands and fitting Equation 3.5 to the combined data, the overall relationship between waveform and envelope coherence can be examined for each room at each source distance. These combined data and best fitting curves are shown in Figure 3.14, and the best fitting n values and RMS percents error are given in Table 3.7. It may be noted that the values of n estimated from the combined data differ little between source distances of 3.0 m and 5.0 m, and are also nearly identical in both room 10B and the reverberant room.

Figure 3.14 clearly shows that in more reverberant conditions and at greater source distances, the waveform coherences decrease on average, spreading out over the range of values shown along the abscissa. The data deviate from the best

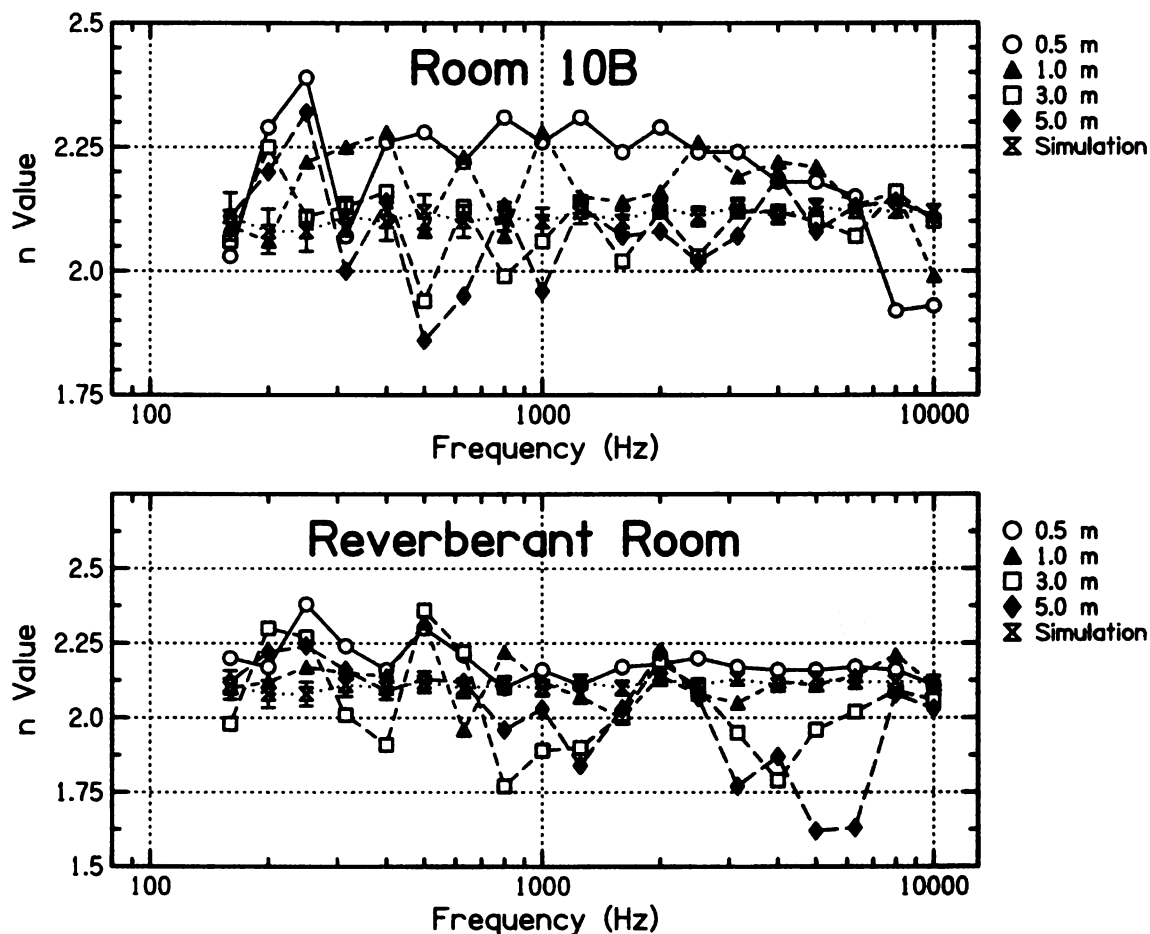


Figure 3.13. A plot of values of n measured in rooms 10B (above) and the reverberant room (below) as a function of 1/3-octave band center frequency. Different lines are plotted for different source distances. There are no clear trends evident in the trends across either frequency or distance to the source.

	10B		Reverb	
Distance	n	RMS Error	n	RMS Error
0.5 m	2.22 ± 0.023	0.0011	2.17 ± 0.020	0.0030
1.0 m	2.17 ± 0.020	0.0024	2.11 ± 0.035	0.0072
3.0 m	2.09 ± 0.027	0.0061	2.07 ± 0.060	0.0088
5.0 m	2.07 ± 0.042	0.0079	2.07 ± 0.068	0.0092

Table 3.7. Values of the power parameter n and the RMS error for the best fit of Equation 3.5 to envelope versus waveform coherence data combined, at each position, across all 1/3-octave bands.

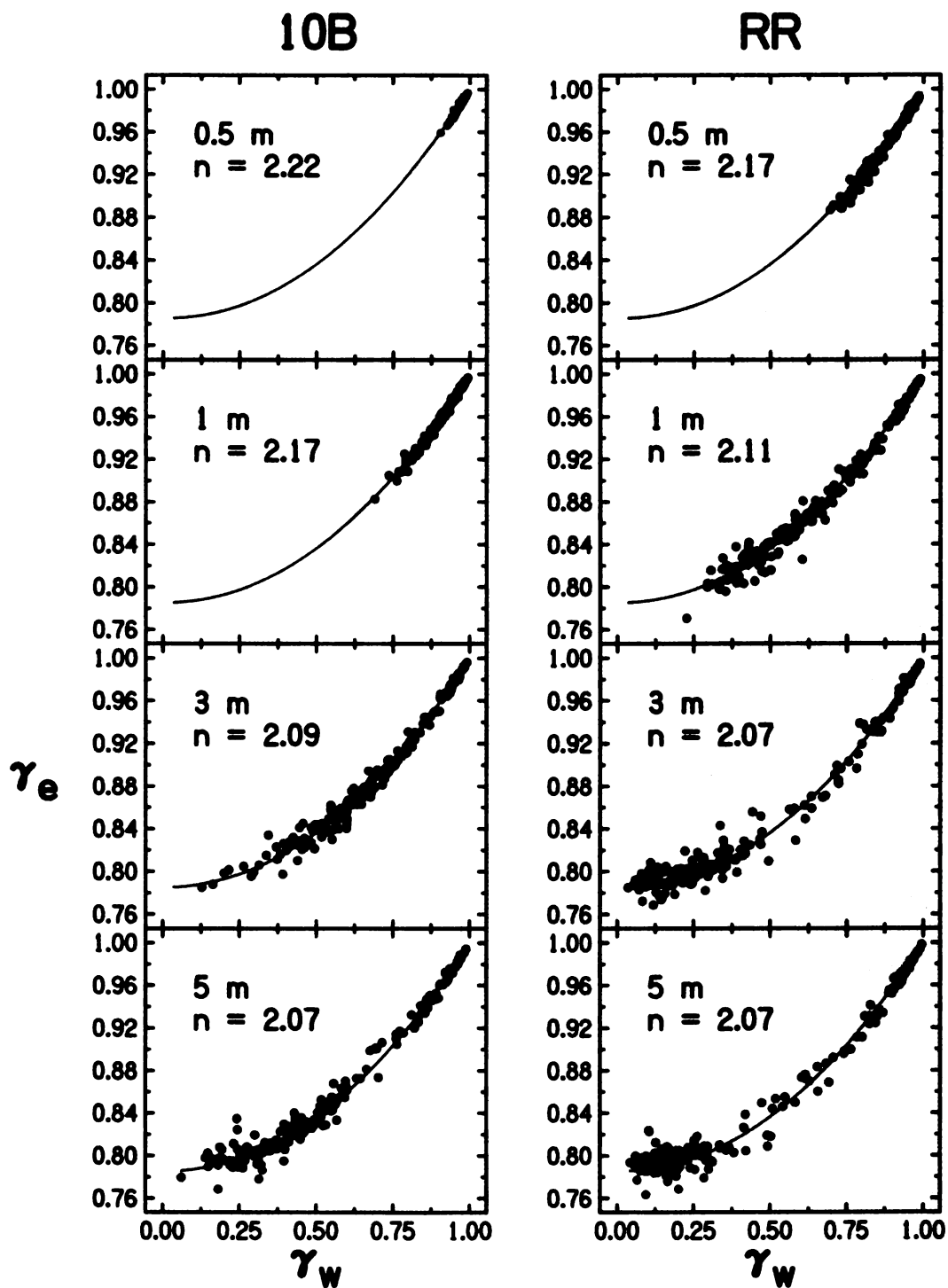


Figure 3.14. Combined envelope versus waveform coherence data across all 1/3-octave bands measured in rooms 10B and the reverberant room at each source distance. Trend lines show the best fit of the data in each panel to Equation 3.5, and the value of the fitting parameter n is shown in the upper-left corner of each panel.

fit to Equation 3.5 more at lower waveform coherences than at greater coherences, especially in the reverberant room at source distances of 3.0 and 5.0 m. In some cases at very low waveform coherence, the envelope coherence even drops below the theoretical minimum of $\frac{\pi}{4}$.

3.4.1 Trimming Data Based on Centrality of Waveform Coherences

It must be noted that across all conditions tested in this experiment, the actual value of n is rather subtle in its effect on Equation 3.5. Consider, for example, Figure 3.15, in which Equation 3.5 is plotted for three different values of the power parameter n . The values of n chosen for Figure 3.15 are roughly indicative of the range of n values found for regression to combined data per Figure 3.14. Clearly, the difference between the curves for the range of the power parameter from $n = 2.0$ to $n = 2.2$ is small over the range of possible waveform coherences (0 to 1).

The greatest difference between the curves shown in Figure 3.15 occurs around $\gamma_w = 0.62$ (although this varies slightly depending on exactly which two curves are being compared). For instance, the greatest difference in envelope coherence between the curves for $n = 2.0$ and $n = 2.2$ is $\Delta\gamma_e = 0.0075$. Near $\gamma_w = 0$ and $\gamma_w = 1$, there is an even smaller difference between the curves. Therefore, data with waveform coherences near the boundaries provide little reliable information for determining the best fitting value of n for any given set of data because these points could potentially fit well to curves with widely varying values of n .

A more reliable value of power parameter n may be obtained for any given data set if only the data with waveform coherences near the middle of the possible range (near $\gamma_w = 0.5$) are used in the regression to Equation 3.5. A cutoff parameter ψ may be defined and the data treated such that any data points with $\gamma_w < \psi$ or $\gamma_w > (1 - \psi)$ are not included in the regression procedure. The data

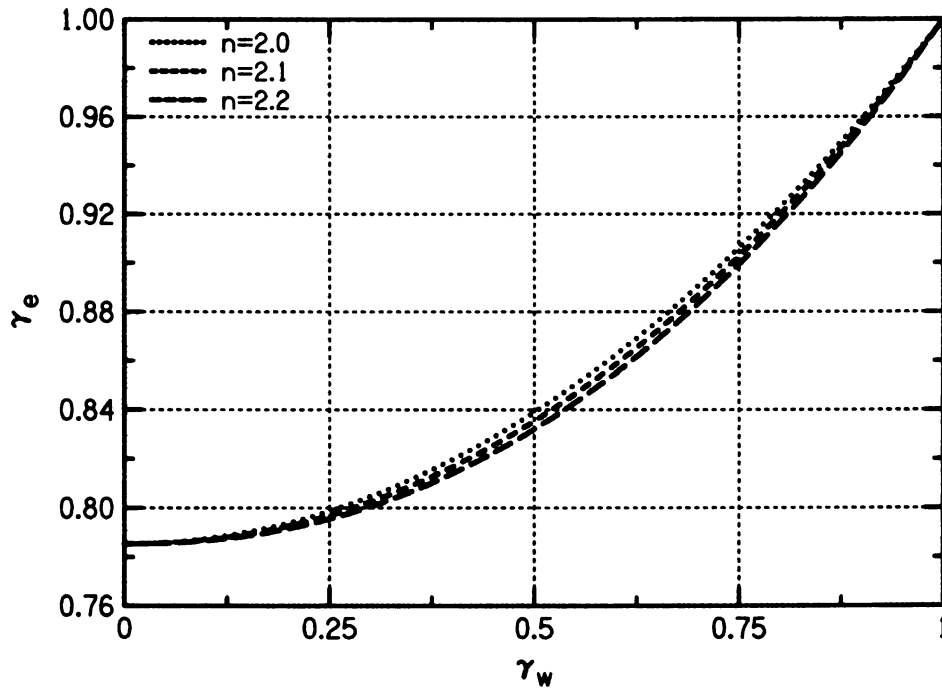


Figure 3.15. Plots of Equation 3.5 with three different values of the power parameter, n . There is very little difference between the curves for these values of n .

that contribute to the regression are then only those near the middle of the curve. For instance, a cutoff parameter of $\psi = 0.1$ would remove both the top and bottom 10% of the allowable values of γ_w . A cutoff of $\psi = 0$ would reject no data and thus return a regression identical to those shown in Figure 3.14. This is depicted for increasing values of ψ in Figure 3.16. Table 3.8 shows the 95% confidence intervals about the mean for values of the power parameter n given the source distance for room 10B and the reverberant room respectively as a function of ψ .

When only the central data points are used in the regression, the combined data across 1/3-octave bands tends toward a fit to Equation 3.5 with $n \simeq 2.09$, independent of source distance and room. Though the width of the related confidence intervals and the RMS percents error increase with increasing ψ , this is likely due to the fact that the best fitting points (those close to $\gamma_w = 1$) have been eliminated from the fit. However, due to the small differences between the values of Equa-

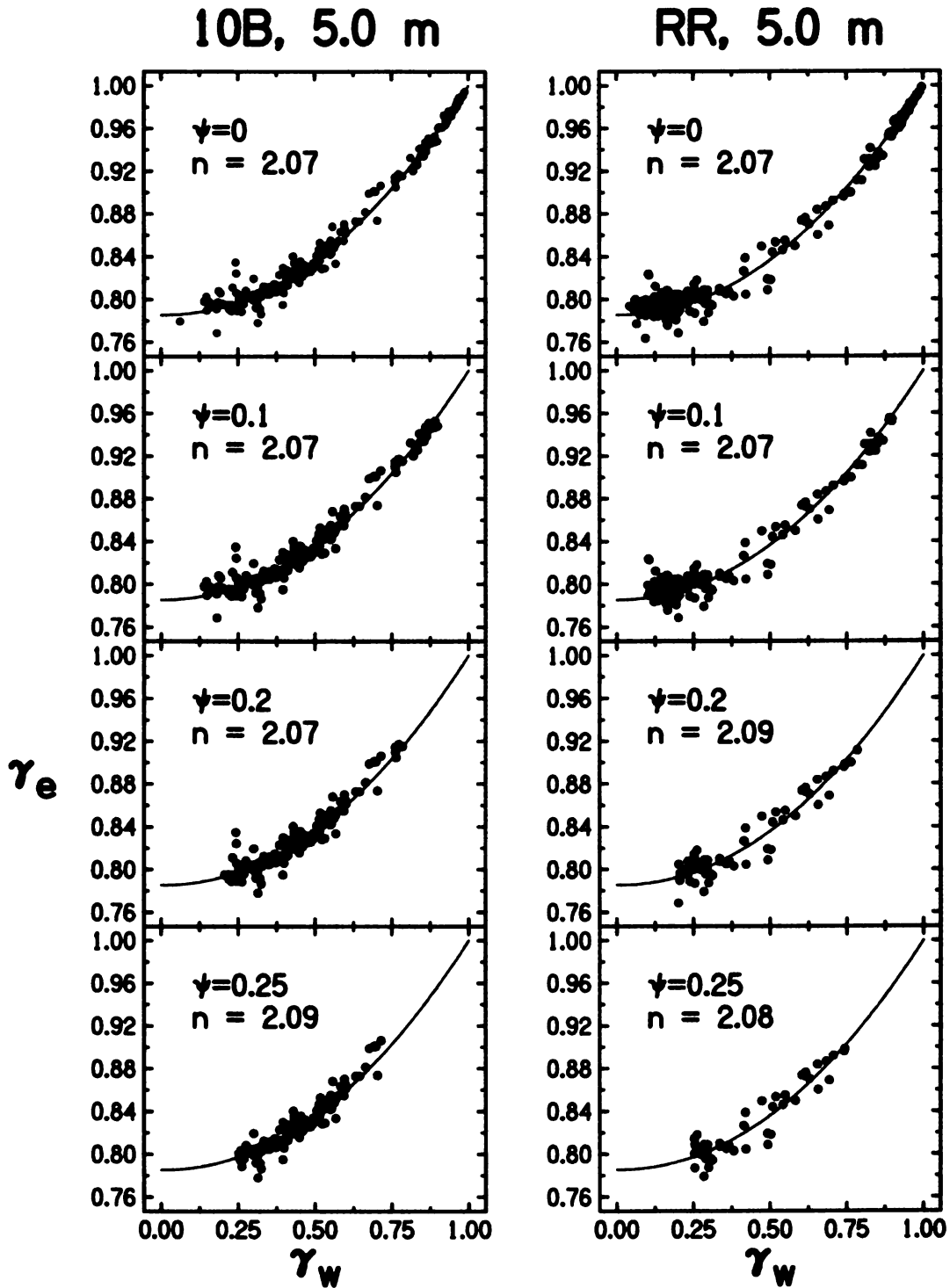


Figure 3.16. Plots of the combined data at a source distance of 5.0 m in room 10B and in the reverberant room. Each panel shows the data remaining after points with waveform coherence $\gamma_w < \psi$ and those with $\gamma_w > (1 - \psi)$ have been removed. The best fit of Equation 3.5 to the remaining data is shown as a solid line and the power parameter n is given in each panel.

	Room 10B		Reverberant Room	
1.0 m				
ψ	n	RMS Error	n	RMS Error
0	2.17 ± 0.020	0.0024	2.11 ± 0.035	0.0072
0.1	2.17 ± 0.037	0.0036	2.11 ± 0.039	0.0081
0.2	2.09 ± 0.15	0.0065	2.10 ± 0.043	0.0085
0.25	-	-	2.09 ± 0.047	0.0088
3.0 m				
0	2.09 ± 0.028	0.0061	2.07 ± 0.060	0.0088
0.1	2.09 ± 0.031	0.0068	2.07 ± 0.068	0.0097
0.2	2.08 ± 0.033	0.0070	2.07 ± 0.075	0.0099
0.25	2.08 ± 0.036	0.0072	2.09 ± 0.089	0.011
5.0 m				
0	2.07 ± 0.042	0.0079	2.07 ± 0.068	0.0092
0.1	2.07 ± 0.047	0.0088	2.07 ± 0.077	0.010
0.2	2.07 ± 0.050	0.0089	2.09 ± 0.090	0.010
0.25	2.09 ± 0.048	0.0082	2.08 ± 0.10	0.011

Table 3.8. 95% confidence intervals about the mean of the power parameter n for different source distances in room 10B and the reverberant room. The parameter n was found by nonlinear regression of Equation 3.5 to data combined across 1/3-octave bands after points with waveform coherences $\gamma_w < \psi$ or $\gamma_w > (1 - \psi)$ were removed. Thus, these values of n and the related RMS errors are determined only by data with abscissae near 0.5. Increasing values of ψ remove increasingly many points near the boundary values of γ_w . In the case of a 1.0 m source distance, most data points are grouped on the high end of the best-fit curve (see Figure 3.14), and a cutoff of $\psi = 0.25$ left too few points in the data set to perform a regression.

tion 3.5 near $\gamma_w = 1$, the points at that end of the curve would return a small RMS percent error for nearly *any* value of n . It must also be considered that with increasing ψ , the number of data points used in the regression decreases, thus increasing the width of the confidence interval. Even though the RMS percent error of the fit tends to increase with ψ , the resulting values of n should perhaps be considered to be more reliable than those found when $\psi = 0$ (Table 3.7).

Of course, there is a limit to which ψ can be increased before important and meaningful data are eliminated, but the exact value of this limit is not clear. The largest value of cutoff used here, $\psi = 0.25$, only allows points with $\gamma_w > 0.25$ and $\gamma_w < 0.75$, the middle 50% of the allowable range, to be used in the regression to Equation 3.5. This is perhaps the greatest value of ψ that would return a reliable estimate of n .

3.5 Conclusions

The relationship between envelope and waveform interaural coherences measured in real rooms agrees remarkably well with computer simulations that calculate coherences from Gaussian white noise pairs. This close agreement suggests that this relationship is robust to the acoustic environment (reverberation time, source distance, et cetera). Both within 1/3-octave bands and across bands, Equation 3.5 provides a reliable relationship between waveform and envelope coherences:

$$\gamma_e = \frac{\pi}{4} + \left(1 - \frac{\pi}{4}\right) \gamma_w^n.$$

Although the power parameter n may change across 1/3-octave bands for a given room and source distance, the changes tend to be small and unpredictable for individual measurements in any specific band. On average, n seems to decrease with increasing source distance. By combining data across bands for a given room and source distance, a reasonably good value of n can be found for any room, given the

source distance. Moreover, this value can be applied to any particular band reasonably well. Although it may at first seem that the value of n for any given room depends on the reverberant characteristics of that room, this is not supported by current observations, which instead suggest that there is little difference between rooms with vastly different reverberation times.

On average it appears that the value of n both within and across bands tends to decrease to some minimum threshold around 2.07. However, further analysis puts this into doubt. The exact value of n found by nonlinear regression of the envelope coherence versus waveform coherence data is very sensitive to small variations in the data. This sensitivity arises because the exact value of n does not have a large impact on the shape of the curves, so curves of similar shape can have quite different values of n (see Figure 3.15). This also implies that the exact value of n may not be important for describing the behavior of the relationship between envelope and waveform coherences. Instead, a “ballpark” estimate may suffice in most cases. Nevertheless, data reduction techniques provide a basis for determining the most reliable possible value of n . Specifically, results show that fitting data within a limited range of coherences between $\gamma_w = 0$ and $\gamma_w = 1$ (i.e. “trimming” the data) is effective.

Calculations of coherences resulting from trimming data seem to confirm the notion that the relationship between waveform and envelope coherence is insensitive to or independent of source distance and the reverberation characteristics of the room. Although the values of n found for source distances of 0.5 m initially seem to violate this ideal behavior, at such small source distances, all of the values of γ_w are relatively large. It has been argued that data with only large values of γ_w will tend to give an unreliable value of n using regression to Equation 3.5 because of the very small difference between curves with quite different values of n near $\gamma_w = 1$. It is suggested that a reliable equation describing the relationship

between waveform coherence γ_w and envelope coherence γ_e in any room and for any source distance is:

$$\gamma_e = \frac{\pi}{4} + \left(1 - \frac{\pi}{4}\right) \gamma_w^{2.1}. \quad (3.6)$$

The envelope coherences computed from Equation 3.6 are not exact. In order to quantify the expected deviation of envelope coherences from Equation 3.6, waveform-coherence–envelope-coherence pairs were combined across all frequencies into one large set. Three such sets were constructed — one for both of the two tested rooms, and one for the computer simulation data. The data in each set were binned by the value of their waveform coherence into bins of width 0.1, thus creating ten bins across the range of waveform coherences. Within each bin, two methods of measuring error were employed: (1) calculation of the mean absolute difference between the measured envelope coherences and envelope coherences predicted by Equation 3.6; and (2) calculation of the absolute difference from Equation 3.6 for which 95% of the data deviate less than the specified amount (i.e. the “95% bound”).

The absolute errors for the rooms turned out to be equal to or slightly less than the errors in the simulated coherences for both methods of error estimation. Thus, a conservative estimate of the absolute error in Equation 3.6 is given by the errors measured in the computer simulation of coherences. The absolute errors for both the Mean and 95% Bound methods are shown as a function of the waveform coherence, γ_w , in Figure 3.17. This figure shows third-order polynomial function fits to the absolute errors. The mean absolute error $\bar{E}$ can then be described as a function of the waveform coherence by:

$$\bar{E} = 0.011 + 0.0083\gamma_w - 0.031\gamma_w^2 + 0.013\gamma_w^3 \quad (3.7)$$

The 95% bound is approximately three times as large as the mean absolute error. A user of Equation 3.6 may expect the difference between envelope coherences calculated from Equation 3.6 and the actual envelope coherences to be approximately

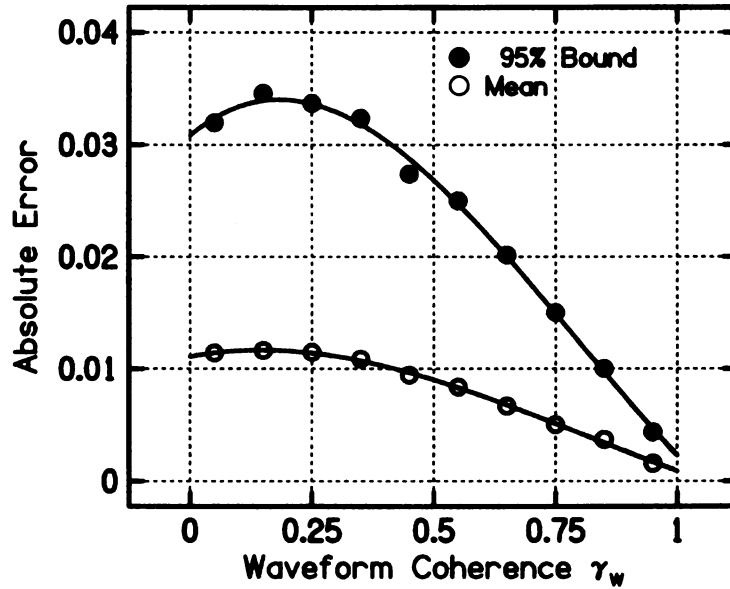


Figure 3.17. The 95% Bound curve (closed circles) gives the absolute error in envelope coherence for which 95% of the simulated coherences deviate by that amount or less from the coherence predicted by Equation 3.6. The Mean curve (open circles) gives the mean absolute error relative to Equation 3.6 for simulated envelope coherences. Both curves are third-order polynomials fitted to the absolute errors of the binned data.

$\pm \bar{E}$ on average, and may expect envelope coherences to deviate no more than $\pm 3\bar{E}$ from Equation 3.6.

It is both interesting and important to realize the remarkable agreement between the apparent coherence relationship in rooms as measured through KEMAR and that determined for simulated noises. For pairs of Gaussian noises with coherences spanning the range from 0 to 1, it was found that $n = 2.11$ fits the data well in every 1/3-octave band. This is essentially identical to the relationship between envelope and waveform coherences in rooms and suggests that the value of the power parameter $n = 2.1$ is indeed a good estimate. Furthermore, since n does not vary very much across frequency for the simulated noises, additional credence may be given to the notion that n is essentially the same for any 1/3-octave band in real rooms.

CHAPTER 4

Measurement of Binaural Properties in Rooms as a Function of Azimuth

4.1 Experiment 9: Measurements of Binaural Properties as a Function of Horizontal Angle in Anechoic Conditions

Experiment 8 showed that interaural coherences both in the waveform and the envelope may be calculated for signals recorded by KEMAR in a real room. The experiments of the present chapter measured other binaural properties, specifically interaural time differences (ITDs) and interaural level differences (ILDs). As these are perhaps the two most important binaural cues in sound source localization for sources in the horizontal plane [48], it is worth carefully measuring them and comparing the results to physical predictions. As the properties of a room may interfere with estimations of ILD and ITD in part due to reverberation, initial measurements were made in anechoic conditions. ILD and ITD vary primarily with the angle from which the sound source is located relative to the head, and so the

incident angle of the sound is the independent parameter in this experiment. At the same time, coherences in both the waveform and envelope were measured as function of angle.

The simplest model of the head approximates it as a solid, rigid sphere with antipodal ears. In this model, the head itself will have a small effect on the cross-correlation function, though its orientation (angle relative to the location of the sound source) will shift the location of the peak in time. Thus, the shape of the cross-correlation function is expected to mimic that of the auto-correlation function of a noise, $a(\tau)$, shifted in time such that its peak occurs for a lag equal to the ITD. For flat-spectrum noise, this auto-correlation function is [47]:

$$a(\tau) = \frac{2}{\Delta\omega\tau} \sin\left(\frac{\Delta\omega\tau}{2}\right) \cos(\bar{\omega}\tau) \quad (4.1)$$

In the above equation, $\Delta\omega$ is the bandwidth, and $\bar{\omega}$ is the center frequency of the noise band. The autocorrelation can be seen to be a cosine function scaled by the total noise power and enveloped by a sinc function. This gives an idea of its shape. For 1/3-octave bands of flat-spectrum noise, $\Delta\omega = 0.23\bar{\omega}$, or for octave bands of noise, $\Delta\omega = \frac{2}{3}\bar{\omega}$. The coherence derived from Equation 4.1 (the peak value) would always be 1, and the ITD (the location of the peak) would vary with the orientation of the head. Hartmann [47] cites particularly favorable cases for how well this approximation works with signals measured through KEMAR, oriented at 45° relative to the direction of incident sound, using a band of noise approximately an octave wide, from 504 Hz to 1000 Hz. In that example, the peaks of Equation 4.1 and the cross-correlation calculated from the signals at the KEMAR ears agree quite well. The coherences measured in the following experiment are expected to differ somewhat from those predicted by Equation 4.1 because flat-spectrum noise is not being used, and the geometry of the KEMAR head is more complicated than that of a simple rigid sphere as it is intended to mimic the shape of an average adult's head and torso.

A method of predicting ITDs comes from a theory of the diffraction of harmonic plane waves by a rigid sphere. Rzhevkin [78] shows that the total pressure on the surface, p_t , normalized by the incident free field plane wave pressure, p_o , is given by:

$$\frac{p_t}{p_o} = \left(\frac{1}{ka} \right)^2 \sum_{m=0}^{\infty} \frac{i^{m+1} P_m(\sin \theta) (2m+1)}{j'_m(ka) - i n'_m(ka)} \quad (4.2)$$

Here, P_m are Legendre polynomials of degree m , j_m are spherical Bessel functions of order m , and n_m are spherical Neumann functions (Bessel functions of the second kind) of order m . The parameter a is the radius of the sphere or the equivalent head radius, and $k = \frac{2\pi f}{c}$ is the acoustic wavenumber where f is the frequency of the incident wave and c is the speed of sound.

Kuhn [62] derives from this a low-frequency approximation for the interaural phase difference (IPD) between antipodal ears — ears on exactly opposite sides of the head, each 90° from the nose — on a rigid spherical head of radius a when exposed to low-frequency plane waves at an incident angle θ relative to the median sagittal plane:

$$IPD = 2 \tan^{-1} \left(\frac{3}{2} ka \sin \theta \right)$$

This approximation assumes that $(ka)^2 \ll 1$. The IPD and ITD are related simply by $ITD = \frac{IPD}{2\pi f}$, and so the ITD in this limit is:

$$ITD = \frac{2}{kc} \tan^{-1} \left(\frac{3}{2} ka \sin \theta \right) \quad (4.3)$$

From this, in the low-frequency limit where $\frac{3}{2} ka \sin \theta \ll \frac{\pi}{2}$, the first order series expansion of the arctangent function in Equation 4.3 can be taken, giving:

$$ITD \simeq \frac{3a}{c} \sin \theta \quad (4.4)$$

This approximation, with a sinusoidal shape for the ITD as it varies over the azimuthal angle of the source, describes the low-frequency ITDs quite well (Figure 4.3).

For high frequencies ($ka > 10$), Kuhn further derives Woodworth's high-frequency approximation for the ITD [94]:

$$ITD \simeq \frac{a}{c} (\sin \theta + \theta) \quad (4.5)$$

This equation suggests a more triangular shape for the ITDs at higher frequencies (which will later be demonstrated by Figure 4.4).

4.1.1 Methods

The same Mackie loudspeaker used in previous experiments was used as a sound source for MLS noise. The loudspeaker and KEMAR were placed in an anechoic chamber, 3.0 m wide by 4.3 m long by 2.4 m high, positioned 1.8 m apart. Three sound-absorbing foam wedges were placed on the seat of a sturdy cushioned chair such that they formed a ledge, atop which the loudspeaker was placed. In this way, the center of the loudspeaker was at the ear height of the KEMAR and the loudspeaker was acoustically isolated from the chair.

With the KEMAR facing the loudspeaker, only minor binaural differences were expected. This was known as the 0° condition. Other angles were tested by rotating the KEMAR on its base, leaving the loudspeaker in place. The full set of angles tested was: $\theta = \{0, \pm 15, \pm 30, \pm 45, \pm 60, \pm 75, \pm 90, \pm 120, \pm 150, 180\}$ degrees. Positive angles corresponded to clockwise rotations of the KEMAR (such that the loudspeaker was to the left), and negative angles corresponded to counterclockwise rotations (such that the loudspeaker was to the right). This convention is depicted in Figure 4.1. Correspondingly, positive ITDs were defined as those for which the left ear led the right, and vice versa. Positive ILDs were defined as those for which the sound level in the left ear was greater than those in the right ear, and vice versa for negative ILDs.

In order to align the KEMAR to specific angles relative to the loudspeaker, a

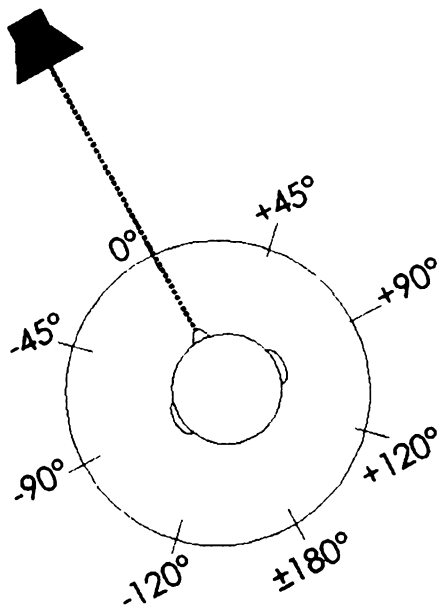


Figure 4.1. A schematic view of the alignment of the KEMAR head relative to the fixed loudspeaker and some of the angles θ , indication the direction of the KEMAR nose. When the KEMAR is rotated in place and the loudspeaker remains stationary, the angles of incidence are positive for clockwise rotations and negative for counterclockwise rotations.

protractor was mounted on the top of the KEMAR head such that the 0° mark of the protractor lay along the KEMAR median axis. The KEMAR was aligned to 0° by projecting a laser beam down the 0° mark of the protractor and adjusting the angle of the KEMAR until the laser beam struck the center of the loudspeaker. Then, the laser beam was projected along each other angle θ while keeping the KEMAR in the 0° position. Marks were made on the walls of the chamber where the laser beam landed for each angle θ . The protractor was then removed. In order to then align the KEMAR to any given angle in θ , a laser beam was projected along the KEMAR median axis and the KEMAR was rotated until the laser beam struck the appropriate mark on the wall of the chamber.

For a single trial, an order-18 MLS was generated with random initial input and played through the loudspeaker. Recordings through the left and right ears of the KEMAR were made at a sample rate of approximately 200 kHz. Thus, the signal was about 1.3 seconds long. Measurements of interaural waveform coherence, ITD, and ILD were made (Equations 2.2, 2.4, 2.5, and 2.6) in twenty-two 1/3-octave bands between 80 and 16000 Hz. Each trial was repeated five times consecutively for each angle, each time with a different MLS of the same order, and averages and standard deviations were computed across trials for each quantity measured at each angle.

The entire experimental procedure was repeated, but with the KEMAR and loudspeaker at different places in the same anechoic room. The relative separation between the KEMAR and loudspeaker remained unchanged, but their positions within the room were changed. This change in position is depicted in Figure 4.2. Such a shift of position in an anechoic environment should not change the results of the measurements made in that environment, and that is exactly what this repeat in the experiment was meant to check. Any significant differences in the results between those made at the two different positions could be attributed to imperfec-

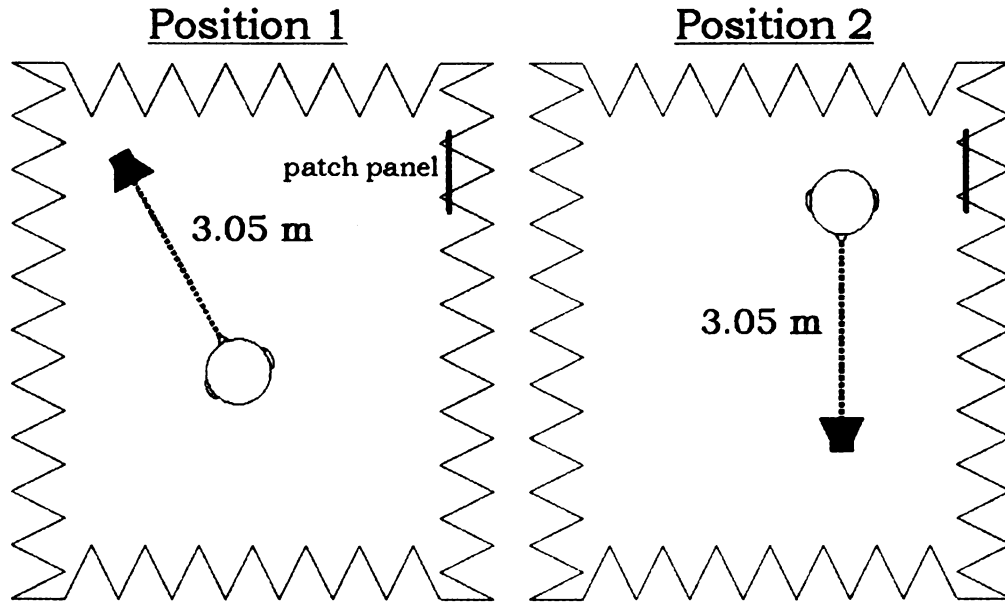


Figure 4.2. A depiction of the locations of the KEMAR and loudspeaker relative to the surroundings of the anechoic room in Positions 1 and 2. In each position, the relative separation of the KEMAR and loudspeaker remains the same, though their positions within the room change.

tions in the anechoic environment. Similarly, any anomalies that would appear in both sets of measurements could not be blamed on imperfections in the anechoic environment.

4.1.2 ITD Results

ITD was estimated as the lag at which the peak of the cross-correlation function occurred in 1/3-octave bands. Plots of ITD across angle within each frequency band are shown in Figures 4.3 and 4.4. More apparent in these figures than in their companion plots of coherence data is the effect of “period errors.” In measuring the binaural coherence and ITD, it is sometimes possible that the highest peak in the cross-correlation function occurs not at the “correct” delay, but one or more periods away (where the periodicity of the cross-correlation function in a given frequency band is understood to be close to that of a sine wave with a frequency

equal to that of the center frequency of the band). These “alias points” were noted as well by Hartmann *et al.* [51] when making similar measurements in a real room at a distance of 6 m.

In the course of this experiment, the ITD and waveform coherence of period errors were noted and recorded along with the number of periods in error. Correct peaks could be identified by locating the peak in the cross-correlation function that was nearest to the ITD measured in other surrounding bands. This, along with the expectations that period errors are unlikely in lower frequency bands (since few periods fall within the ± 1 ms limit in lag), and that ITD is expected to decrease somewhat with increasing frequency, allowed for the correct peak to be determined. The location and height of the correct peak then replaced the incorrect ITD and waveform coherence in the data set. The period errors are shown in Figures, 4.3, 4.4, 4.11, and 4.12 as filled-in counterparts to the open symbols.

Period errors tended to be more frequent at higher frequencies. At both positions, period errors most frequently occurred in the 16 kHz band. Errors of two periods occurred two times at each position — in position 1 at -120° in the 8000 Hz band and at 90° in the 12.5 kHz band, and in position 2 at 45° and 90° in the 16 kHz band. No error of more than two periods was observed within the parameter space of this experiment.

The ITDs are, as Kuhn predicts (Equation 4.3), sinusoidal in shape at low frequencies. At higher frequencies, the pattern of ITDs take on a more triangular shape, as predicted by Woodworth’s formula (Equation 4.5). Both Equation 4.3 and Equation 4.5 were fitted to the ITD data in order to arrive at an estimate of the equivalent head radius of KEMAR. These equations were fitted to the given ITD data in a manner in which the head radius a is a parameter that is allowed to vary with frequency. The result is that Kuhn’s low-frequency approximation predicts a head radius of about 8.0 cm given the ITD data in the lowest frequency bands, and

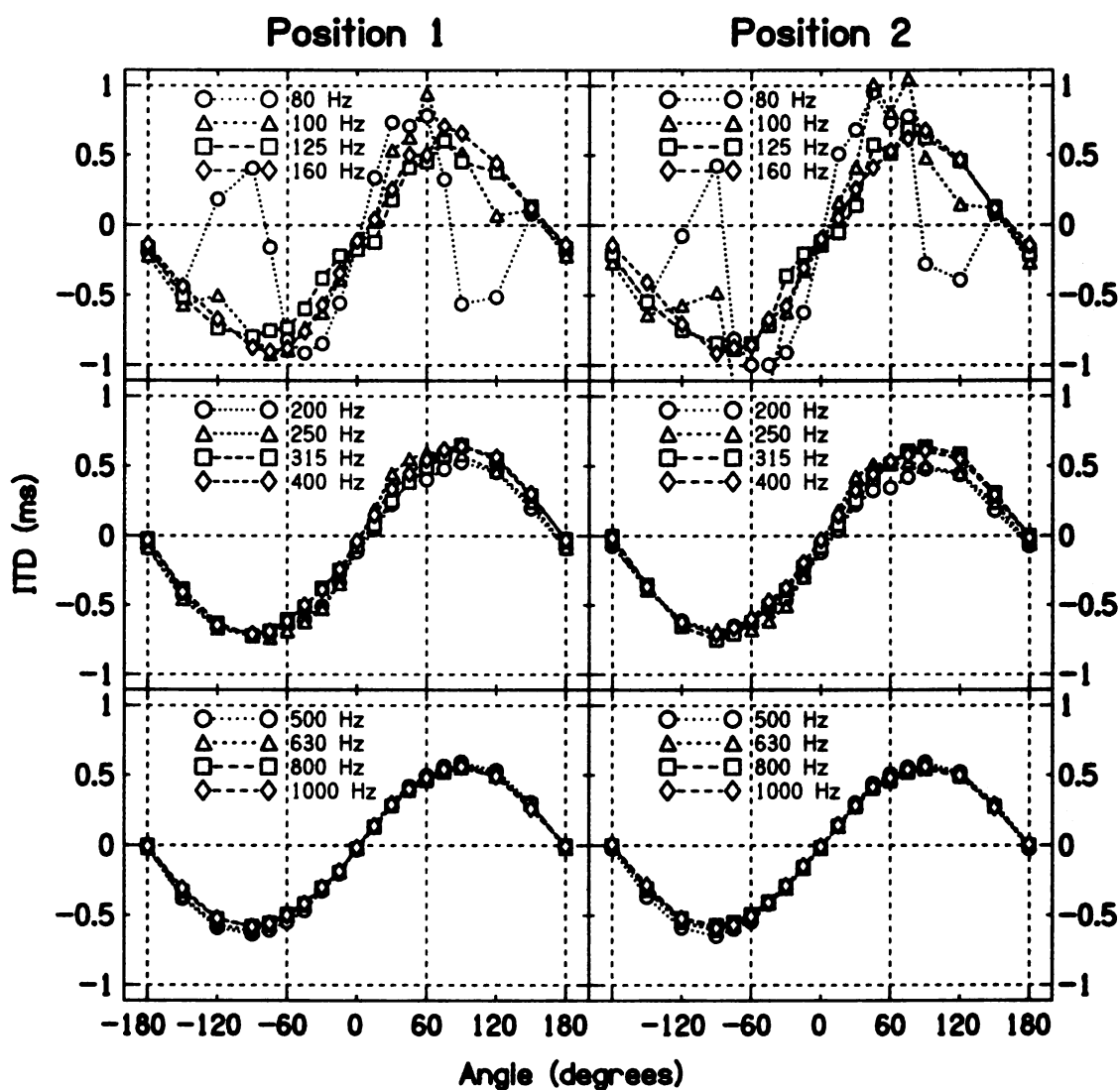


Figure 4.3. ITD in milliseconds as a function of angle in 1/3-octave bands, plotted in the same format as Figure 4.11. ITDs for the remaining higher frequency bands are shown in Figure 4.4.

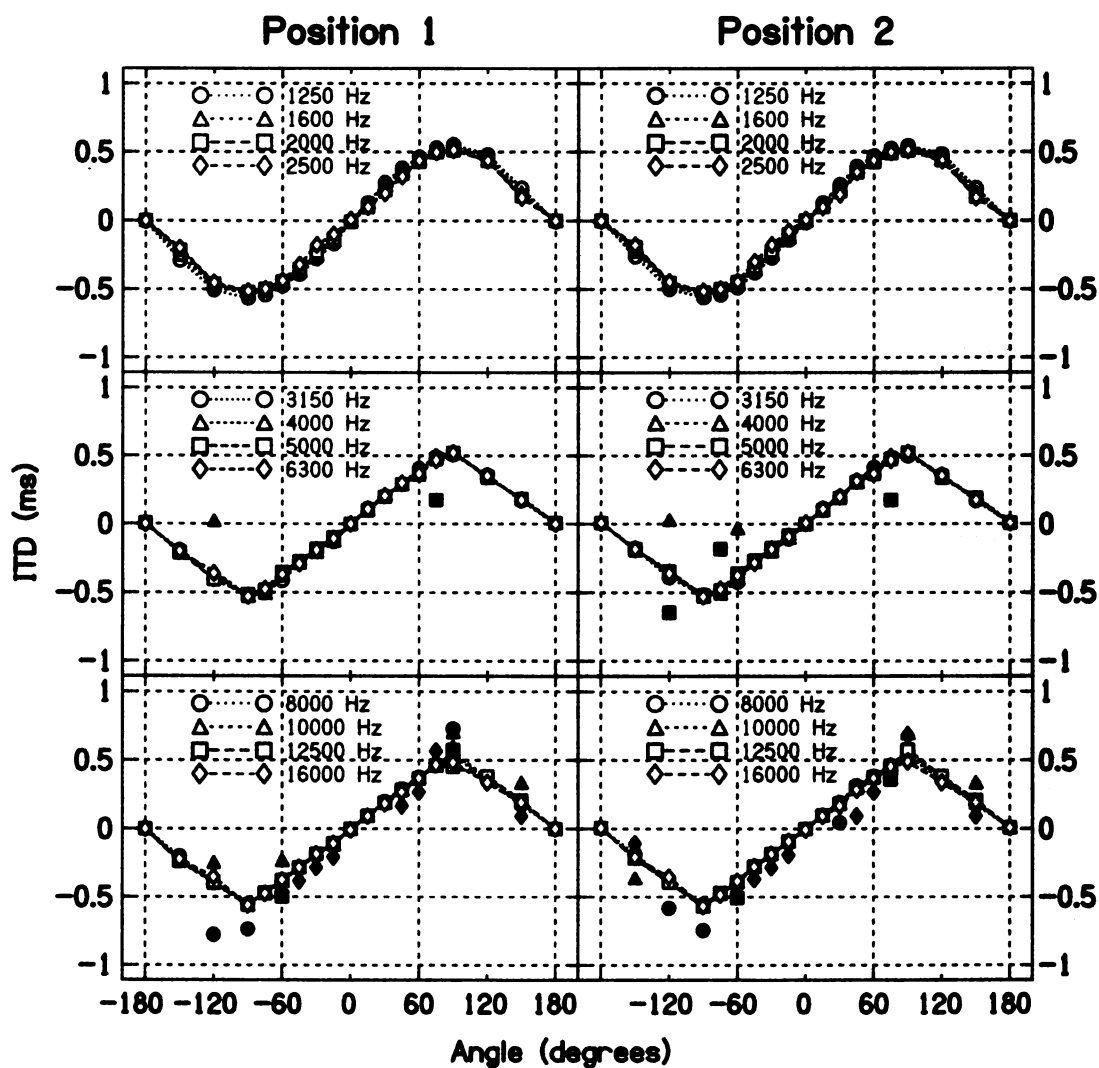


Figure 4.4. ITD in milliseconds (ms) as a function of angle in 1/3-octave bands, plotted in the same format as Figure 4.12. Closed symbols represent original data before corrections for period errors were made. ITDs for lower frequency bands are shown in Figure 4.3.

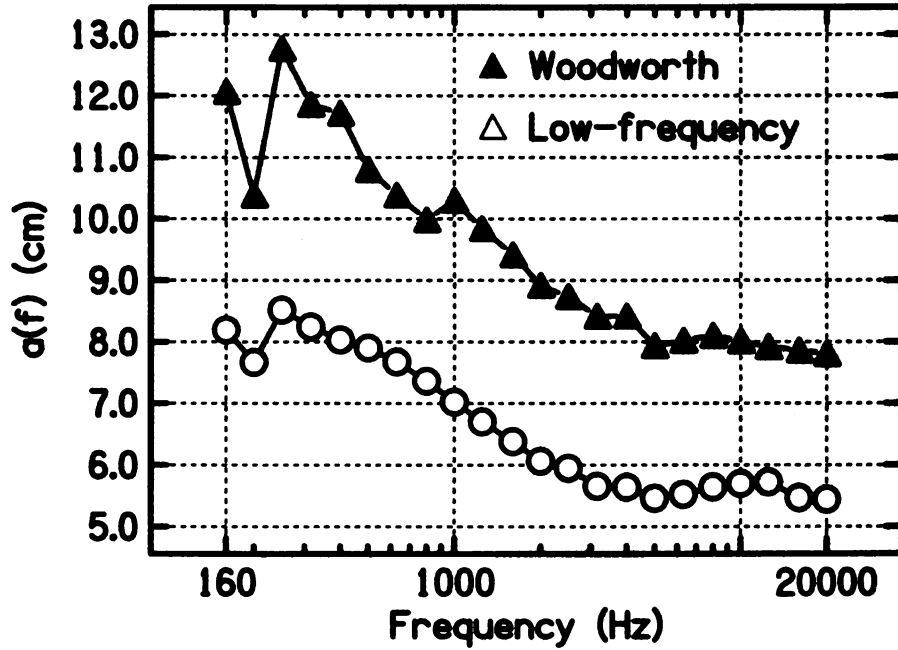


Figure 4.5. Head radii a as a function of frequency, calculated by fitting Kuhn's low-frequency approximation (Equation 4.3) and the Woodworth formula (Equation 4.5) to ITD data in 1/3-octave bands. The results of fitting the Woodworth equation are shown by closed triangles and the results of the low-frequency approximation are represented by open circles. The low-frequency approximation predicts a head radius of about 8.0 cm in the low-frequency bands, and this is in close agreement with the predictions of the Woodworth formula in high-frequency bands.

the Woodworth formula predicts a nearly equivalent head radius in the highest frequency bands. The trends are shown in Figure 4.5. These fits are least-squares regressions over the full 360° range shown in Figures 4.3 and 4.4.

Previous studies in which the Woodworth formula was shown to agree well with measured data include those of Algazi *et al.* [2], Duda *et al.* [32], and Feddersen *et al.* [37]. In those studies, the Woodworth formula was shown to correspond well to measured ITDs. However, Feddersen *et al.* took measurements with microphones on the sides of heads, just in front of the ear canals of human listeners. Duda *et al.* took measurements on opposite sides of a rigid sphere, though human ears are not truly antipodal. In neither case were measurements taken at the ex-

act location of the ear canals, nor did those measurements include the effects of pinnae or ear canals on the signals. Algazi *et al.* took measurements of human beings, but did not compare those to measurements in KEMAR or any other head simulator. The Woodworth formula itself was derived from a model of the head as a rigid sphere with antipodal ears, and so it is perhaps not surprising that the measurements of Duda *et al.* conform to it so well. It should also not be surprising, however, that some inconsistencies arise between it and the current data, which were measured through the ears of KEMAR, pinnae and ear canals included.

Burkhard and Sachs [24] quote the ear-to-ear distance of KEMAR to be 15.2 cm, which gives a head radius of 7.6 cm. Measurements of various dimensions on and about the head were made by Algazi *et. al* [3] on 43 human subjects. The average, plus and minus one standard deviation, of the ear-to-ear head width across all 43 subjects was 14.5 ± 0.95 cm. The head depth from forehead to the back of the head was 21 ± 1.2 cm, and the circumference of the head was 57 ± 2.5 cm. These three measurements yield three different estimators of the correct equivalent head radius - 7.25 cm, 10.5 cm, and 9.1 cm respectively. Algazi *et al.* in another study compute the average optimal effective head radius of 25 different human listeners to be 8.7 cm [2]. All of these measurements indicate head radii close to, though usually larger than the effective radius of the KEMAR head.

An alternative procedure for fitting the ITD data may come from varying not the head radius, but the contribution of the linear term θ in Equation 4.5. This would be a fit of the ITD data to an equation of the form:

$$ITD = \frac{a}{c} [\sin \theta + b(f)\theta] \quad (4.6)$$

In Equation 4.6, the head radius a is fixed, and the contribution of the linear term, b , is allowed to vary with frequency. Using the estimate of the head radius from Algazi, *et. al*, the average of parameter b for frequency bands at and above 2500 Hz is found to be 0.25. The variation across frequency is shown in Figure 4.6.

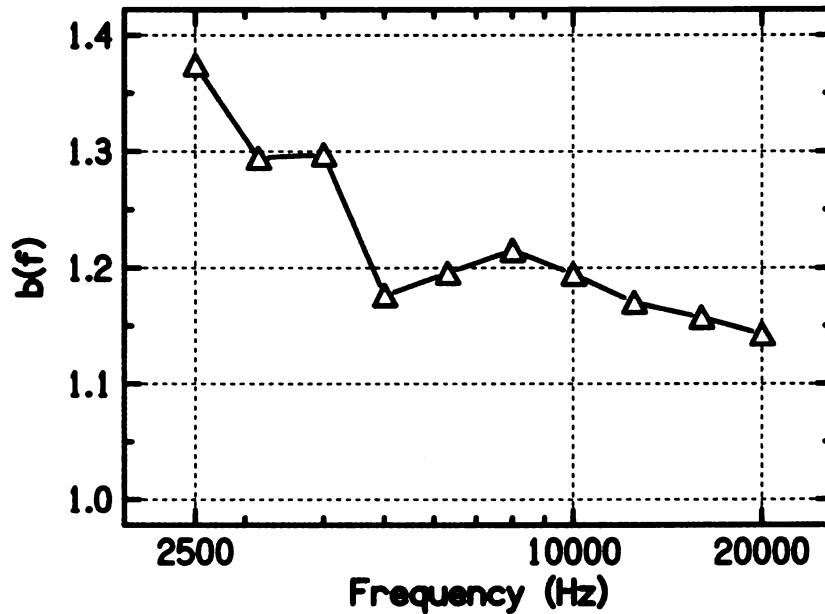


Figure 4.6. The linear parameter b as a function of frequency, calculated by fitting Equation 4.6 to ITD data for frequencies at and above 2500 Hz. This parameter describes the relative contribution of the linear term θ in Woodworth's equation (Equation 4.5) when the head radius is fixed to be 7.25 cm.

Figure 4.7 shows data plotted from one particular high-frequency band (centered at 8000 Hz), and plots of the best fitting equations. The solid line shows the best fit of Equation 4.5 to this data, which is achieved for $a = 8.10$ cm. The dashed line is the best fit of Equation 4.6, which is achieved for $b = 1.22$ when the head radius is set to $a = 7.25$ cm. Woodworth's equations (Equations 4.5 and 4.6) seem to fit the data rather well, though they perhaps do not capture the triangular shape of ITDs at high frequencies particularly well.

4.1.3 ILD Results

Plots of ILD across angles for all frequency bands at both positions are shown in Figures 4.8 and 4.9. At low frequencies, head shadow is expected to be minimal because long wavelengths are diffracted around the head. That Figure 4.8 does not reflect this at the lowest frequencies is an indication that the anechoic room is not

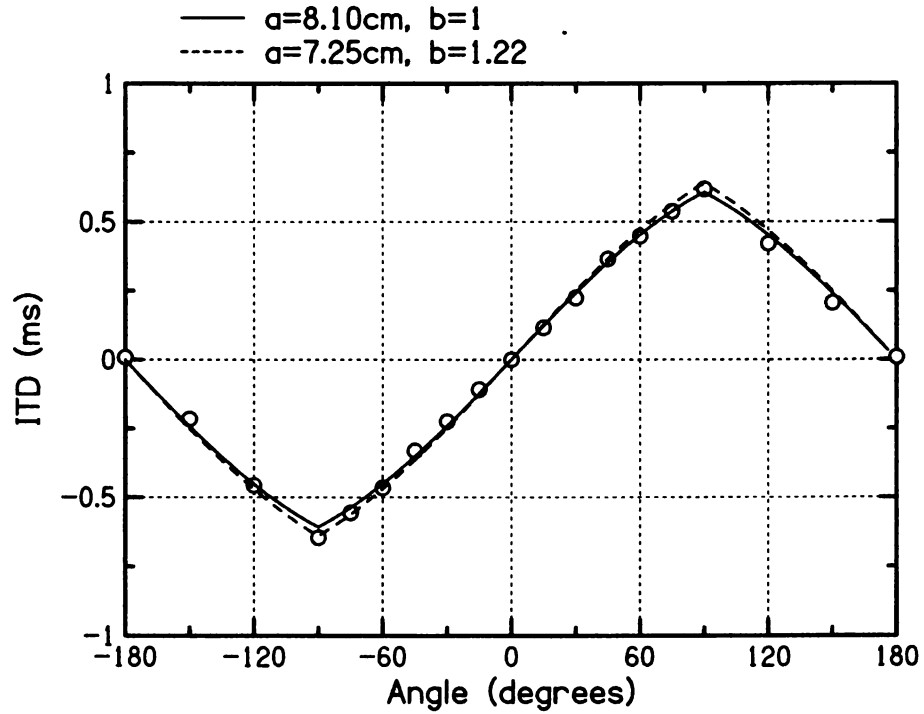


Figure 4.7. ITD data from the 8000 Hz 1/3-octave band is plotted across frequency as open circles. A fit of Equation 4.5 to this data, shown by the solid line, is made when the fitting parameter (the head radius) is $a = 8.2$ cm. The dashed line shows a fit to this data of Equation 4.6, in which the head radius is set to $a = 7.25$ cm, and the fitting parameter is found to be $b = 1.22$.

truly anechoic at those frequencies. This supposition is supported by the aberrant behavior of the ITDs at very low frequencies (see Figure 4.3). The anechoic room used for these measurements, which is lined with sound-absorbing foam wedges of approximately 0.89 m in length, is not truly anechoic at these low frequencies. The room can be expected to be anechoic for wavelengths shorter than four times the length of the absorbing wedges (3.56 m) [10], which corresponds to a frequency of approximately 96 Hz. Further discussion of binaural properties in the anechoic condition and comparisons of further results to the anechoic environment will be understood to exclude the measurements at these four lowest frequency bands (those below 160 Hz) unless otherwise noted.

At 0° and 180° , the ILD is expected to be zero. Here, the average ILD, plus and minus one standard deviation, across all frequency bands (not including the lowest three) at 0° was 0.3 ± 0.83 dB at position 1 and 0.3 ± 0.91 dB at position 2. Here, any difference in the ILD from zero at 0° , found to be roughly equal at both positions, is likely the fault of slightly unequal gains in the two recording channels. At 180° , the average ILD was found to be 1.0 ± 0.83 dB at position 1 and 1.0 ± 0.42 dB at position 2. Differences in ILD between 0° and 180° are likely due to asymmetries in the KEMAR, but these differences are explored in greater detail in a later section.

Dips in the magnitude of the ILD function were found at 90° in many of the higher frequency bands. This is the acoustical equivalent of Fresnel's "bright spot." Its existence was derived by Poisson as a result of a theory of the wave-nature of light put forth by Fresnel and later experimentally verified by Arago [53]. A rigid spherical object will diffract light in such a way that a bright spot will appear on the side of the object directly opposite of the light source. A similar effect is expected of acoustical plane waves incident on a rigid spherical diffracting object. A "bright" acoustical spot at the far ear for an angle of 90° will result in a lower

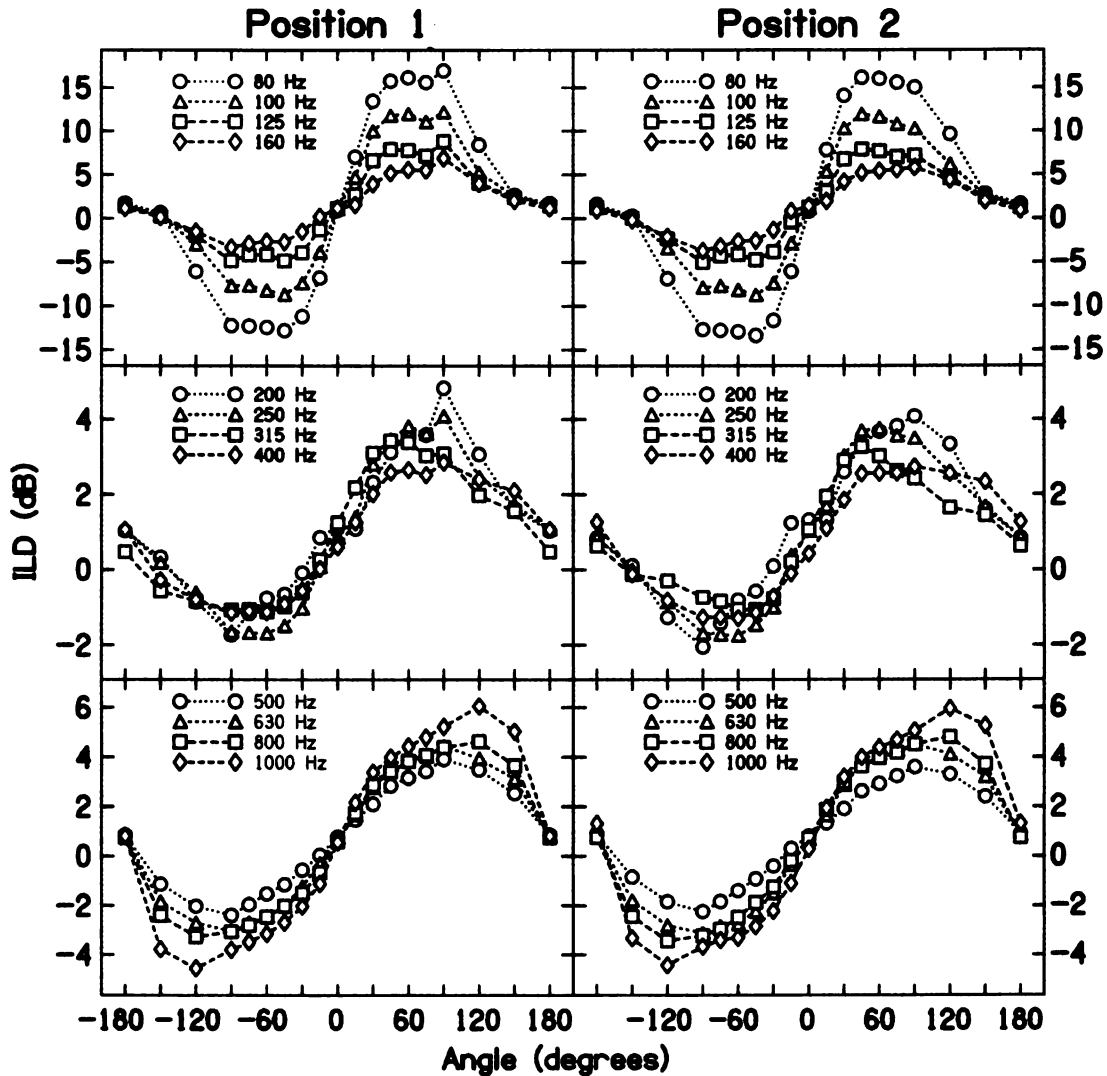


Figure 4.8. ILD in decibels (dB) as a function of angle in 1/3-octave bands, plotted in the same format as Figure 4.11. ILDs for the remaining higher frequency bands are shown in Figure 4.9. In each panel, the vertical scale changes, as shown by axes labels on the left and right sides of the figure.

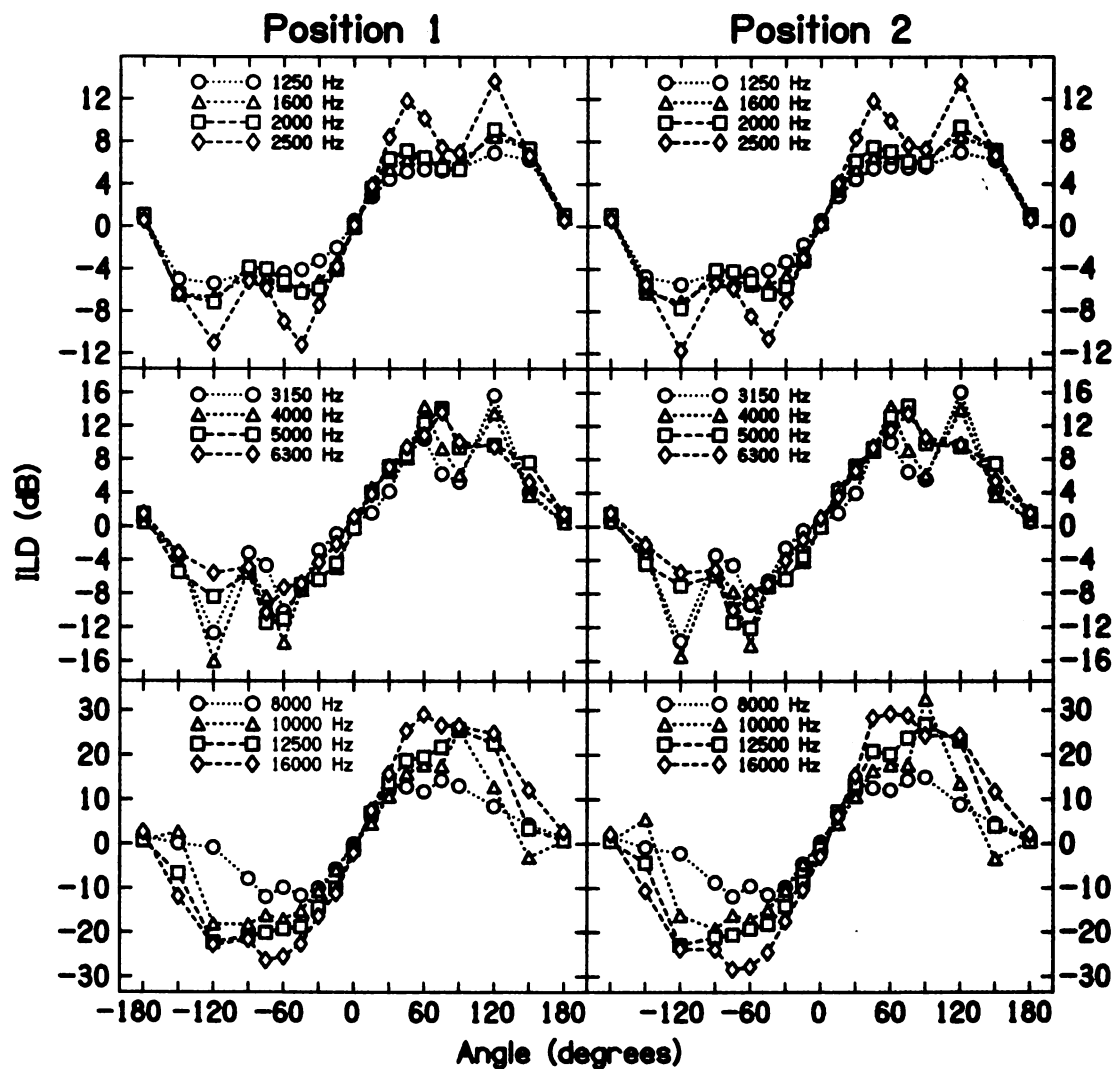


Figure 4.9. ILD in decibels (dB) as a function of angle in 1/3-octave bands, plotted in the same format as Figure 4.12. ILDs for lower frequency bands are shown in Figure 4.8. In each panel, the vertical scale changes, as shown by axes labels on the left and right sides of the figure.

ILD for wavelengths short enough to be appreciably diffracted by the head. This can be seen in frequency bands as low as 1250 Hz as a flattening of the ILD curve at 90° such that the curve looks to deviate from the sinusoidal shape seen in lower frequency bands. By 2500 Hz, there is a clear dip in the magnitude of the ILD at 90° relative to the angles around it.

4.1.4 Coherence Results

The coherences measured in the anechoic environment are plotted as a function of frequency for several values of the incident angle in Figure 4.10. Most notable about these plots is the fact that low-frequency coherences are small at certain incident angles, most notably in the range of 30° to 90°. This trend is remarkably similar when measured at each location, suggesting that it is not an effect of the room itself. A more interesting view of what is happening is given by plotting the coherence as a function of incident angle in various center frequencies.

Results for coherence in 1/3-octave bands from 80 to 1000 Hz at each position are shown in Figure 4.11, and those for bands from 1250 to 16000 Hz are shown in Figure 4.12. There was not much difference in coherence measurements made at position 1 versus those made at position 2. The greatest difference between positions was for -15°, where on average, plus and minus one standard deviation, the coherences were 0.023 ± 0.031 greater at position 2 than at position 1. Pooling all pairs of data together from position 1 and position 2, the paired differences show that the coherences are greater at position 2 than at position 1 by only 0.005 ± 0.036 (paired *t*-test, $p = 0.001$). Though the data indicate a slightly higher set of coherences at position 2 versus position 1 on average, the absolute percent difference of 0.5% is minimal. This difference is likely unimportant in the sense of comparing coherences at two different places in a room.

Of particular interest in the plots of coherence across angle is the trend for co-

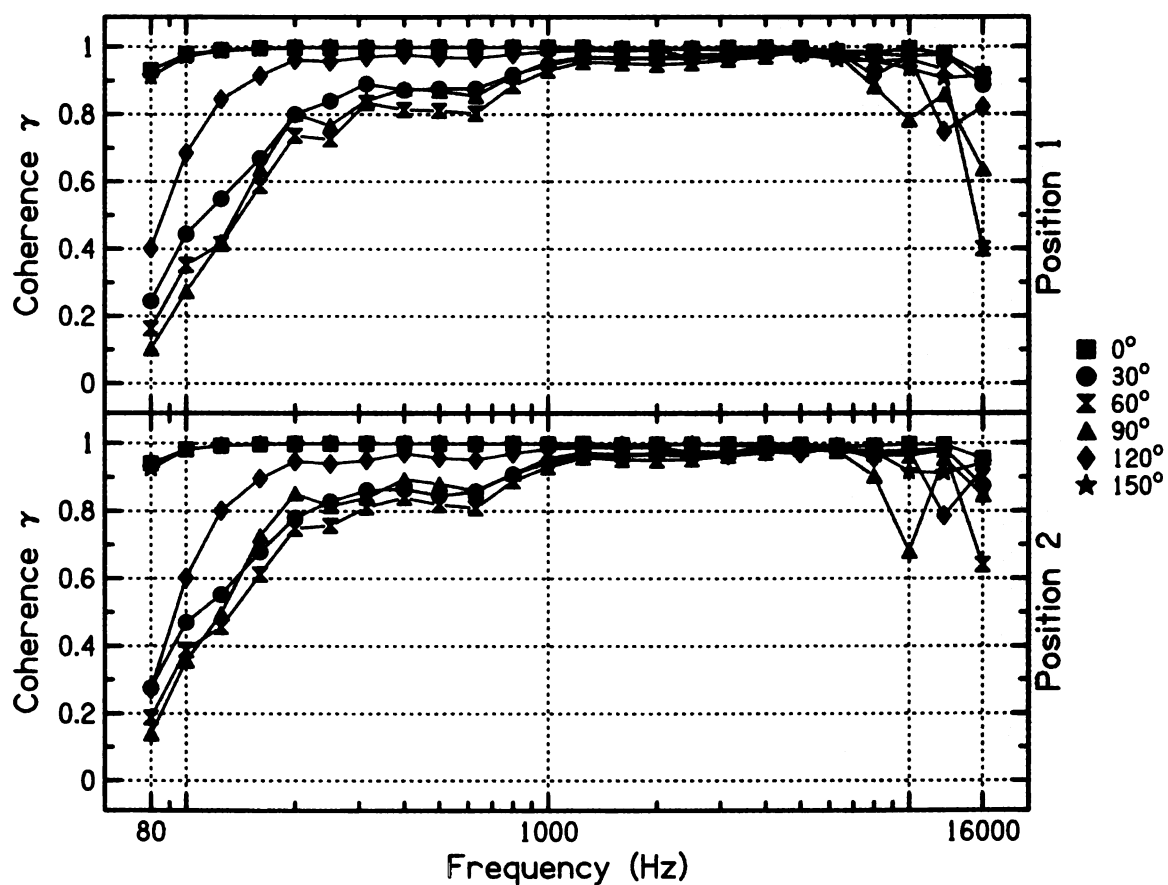


Figure 4.10. Coherence γ as a function of frequency for positive angles in 30° increments. The top panel shows measurements made at Position 1, and the bottom panel shows measurements made at Position 2 in the anechoic environment.

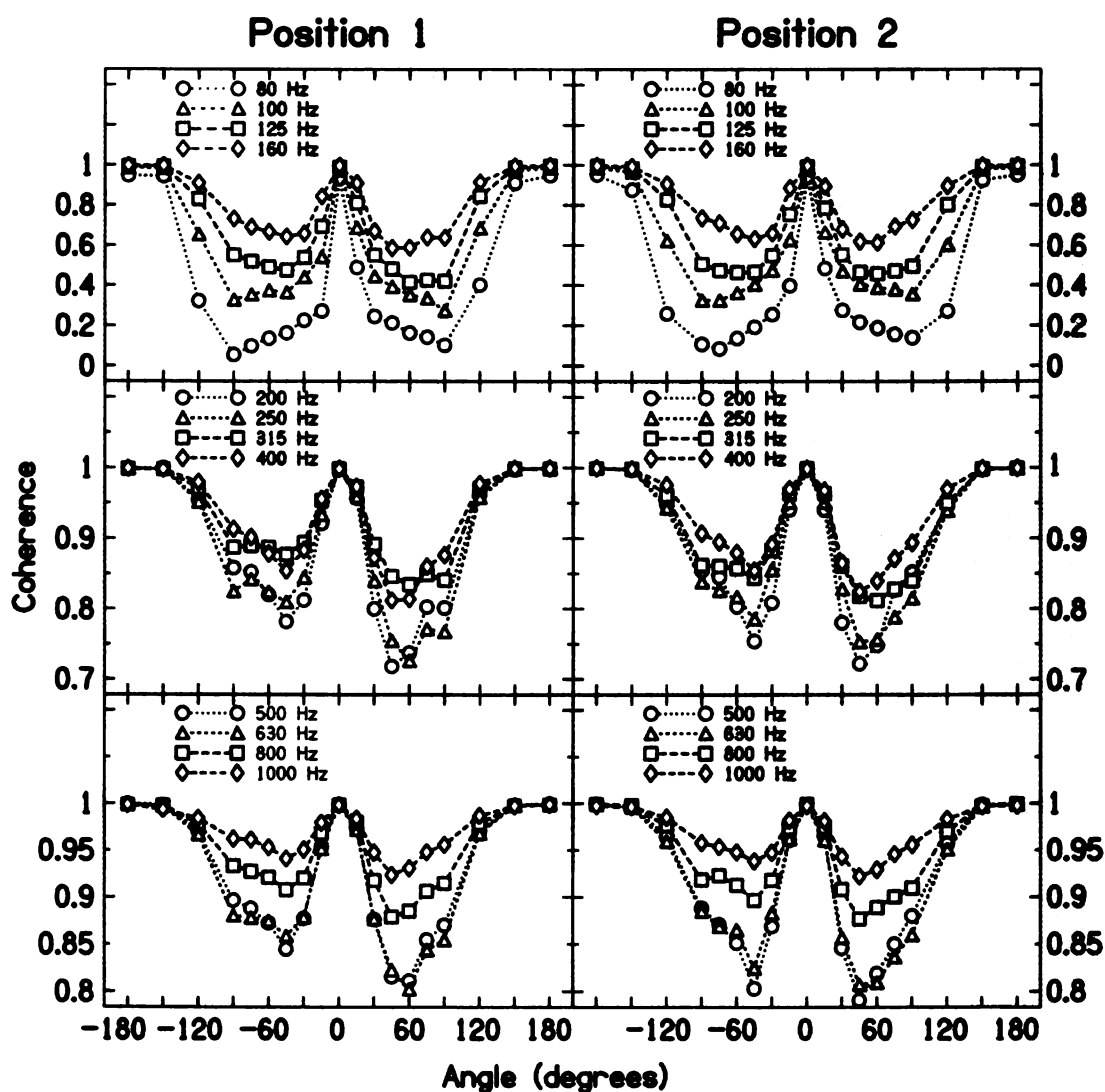


Figure 4.11. Coherences as a function of angle in 1/3-octave bands for each of two different positions in the same anechoic room. Four frequency bands are shown per plot, with results for position 1 on the left and for position 2 on the right. Shown in this plot are coherences for frequency bands from 80 to 1000 Hz. Filled-in symbols represent original data before corrections for period errors were made. Coherences for the remaining higher frequency bands are shown in Figure 4.12.

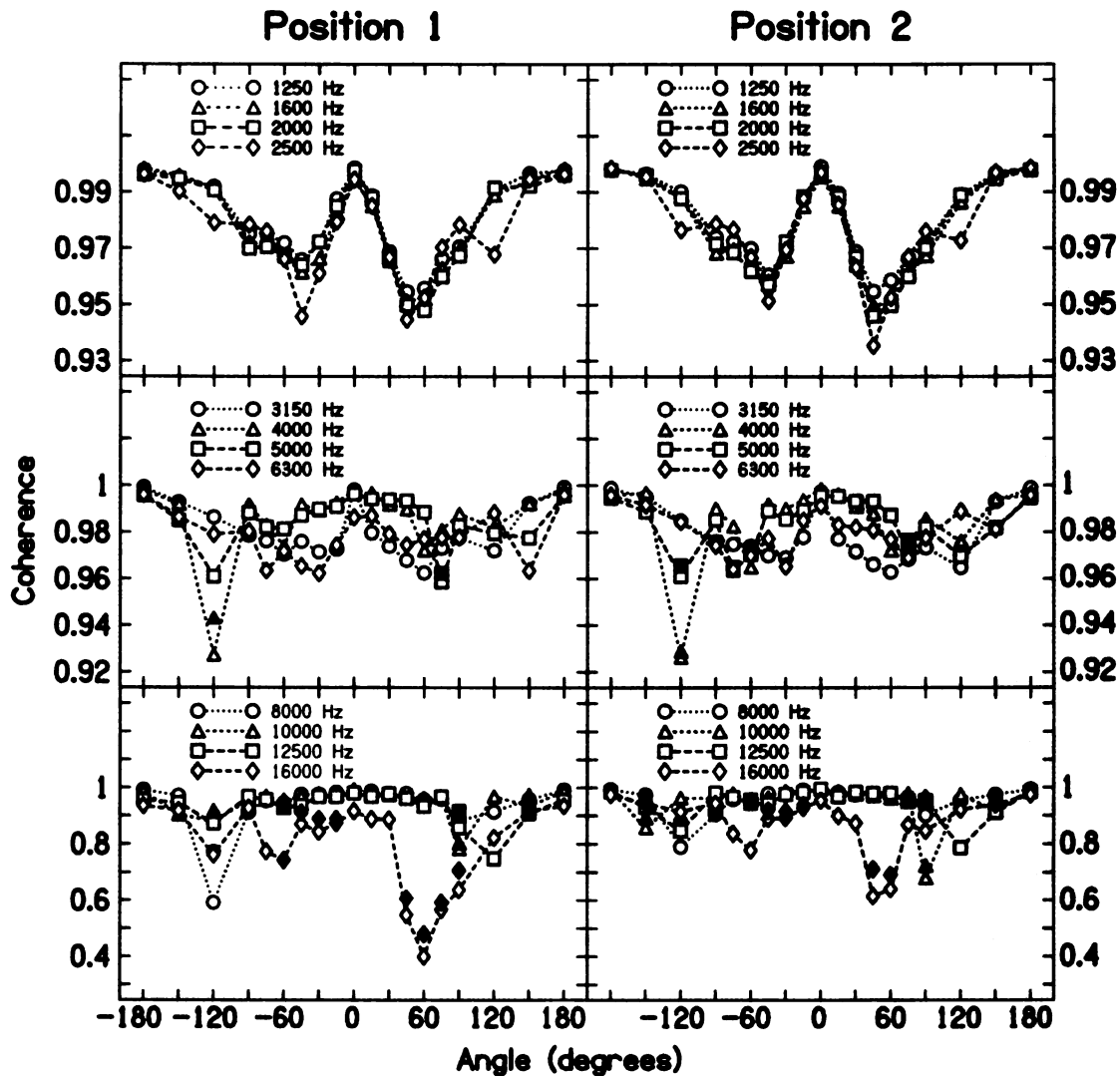


Figure 4.12. Coherences as a function of angle in 1/3-octave bands for each of two different positions in the same anechoic room. Four frequency bands are shown per plot, with results for position 1 on the left and for position 2 on the right. Shown in this plot are coherences for frequency bands from 12500 to 16000 Hz. Filled-in symbols represent original data before corrections for period errors were made. Coherences for the lower frequency bands are shown in Figure 4.11.

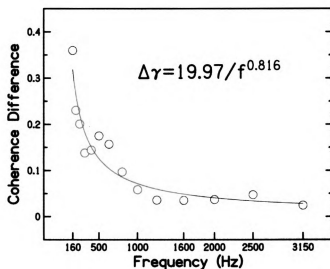


Figure 4.13. In open circles, a plot of average difference in coherence between those measured at 0° and 45° across 1/3-octave frequency bands from 160 to 3150 Hz in anechoic conditions. The trend line and its accompanying equation represent a nonlinear least squares regression to the data.

herences to dip down sharply, with minima near $\pm 45^\circ$. This unexpected effect was found consistently in the range of 200–3150 Hz. With increasing frequency, the difference between perfect coherence (1) and the coherence measured at 45° steadily decreased from an average difference of 0.360 in the 160 Hz band to 0.025 in the 3150 Hz band. The difference in waveform coherence between 0° and 45° decreases roughly as a power function in frequency of the form:

$$\Delta\gamma = \frac{19.97}{f^{0.816}} \quad (4.7)$$

A plot of these differences and the best fit line corresponding to Equation 4.7 are shown in Figure 4.13. The cause of the anomalous behavior in coherence was explored further in follow-up experiments.

Incoherence due to Signal Degradation

One possible cause for unusually low coherence is decay of the signal before it reaches one of the ears. This may be a result of diffraction by the head leading

to partially destructive interference at one ear. Incoherence caused by signal decay should correspond to an anomalous ILD. For instance, in the 4000-Hz band at -120° , there is a sharp dip in coherence as seen in Figure 4.12; this is matched by a similarly out-of-place increase in the magnitude of the ILD at that angle in that frequency band. Though less pronounced, this same coincidence can be seen for the 2500 Hz band at 120° . It is important to note that an anomalous ILD may be correlated with, but is not the cause for incoherence.

The drops in coherence for frequency bands through 3150 Hz at angles of $\pm 45^\circ$ cannot be universally explained simply by an unusual degradation of the signal in one of the ears because a corresponding increase in the magnitude of the ILD is not often seen. There are several other places at higher frequencies where this disconnect between unusually low coherence but normal ILD takes place, such as in the 5000 and 8000 Hz bands at -120° , and the 16000 band around 60° , but never in so systematic a way across a range of frequencies as for 45° . Since an unusual anomaly in the overall level of one of the signals is not the source of the drop in coherence at 45° , it must be that one signal is otherwise somehow changed relative to the other, as if by filtering.

At frequencies above 4000 Hz, filtering effects of the pinna may become important [52], as they affect the spectrum of the sound in a manner often thought to aid localization in the MSP. Asymmetrical spectral effects of the pinnae may in certain instances result in incoherence in those frequency bands affected. However, the spectral effects of the pinnae are not generally present at such low frequencies as where these systematic dips in the coherence were seen in the present study (around 45°).

Incoherence due to Spectral Properties of Signals

Signals can also be made incoherent by means of additive noise, but that is unlikely to be the cause of the dips in coherence seen in measurements made with KEMAR in the current experiments (see Figures 4.11 and 4.12). The cause for the interaural incoherence at and around $\pm 45^\circ$ may be due instead to differences within bands in either the time or the amplitude spectrum of the sounds in the left and right ears. When either ITDs or ILDs are frequency-dependent within a noise band, the interaural coherence will be imperfect (less than 1) as a result [49]. In general, interaural incoherence may be the result of a combination of frequency-dependent ILDs and frequency-dependent ITDs within bands. In this experiment, it is possible to separate out the individual contributions of these two factors.

To see the effect of the ILDs within bands alone, the signals recorded in the KEMAR ears were altered such that the IPD (and thus the ITD) between the ears was set to zero for each spectral component. To do this, the signals in the left and right ears were both Fourier transformed, and the phase spectrum of the signal in one ear was altered, component by component, to match the phase spectrum in the other ear. In so doing, the left and right ear signals then had identical phase spectra, but the ILDs (which result from differences in the amplitude spectra) were unchanged. The signals were then inverse Fourier transformed, the cross-correlation between the signals was computed, and the new “equal-ITD coherence” was found. Any incoherence found could then be attributed entirely to differences in the ILD spectra of the two signals. A similar procedure was performed in which the original signals were made to have equal amplitude spectra while maintaining their different phase spectra. Then an “equal-ILD coherence” due only to differences in ITDs between the signals could be found. These coherences were computed in the 500 Hz 1/3-octave band for five different measurements at each of three positions: $\theta = \{-45^\circ, 0^\circ, +45^\circ\}$.

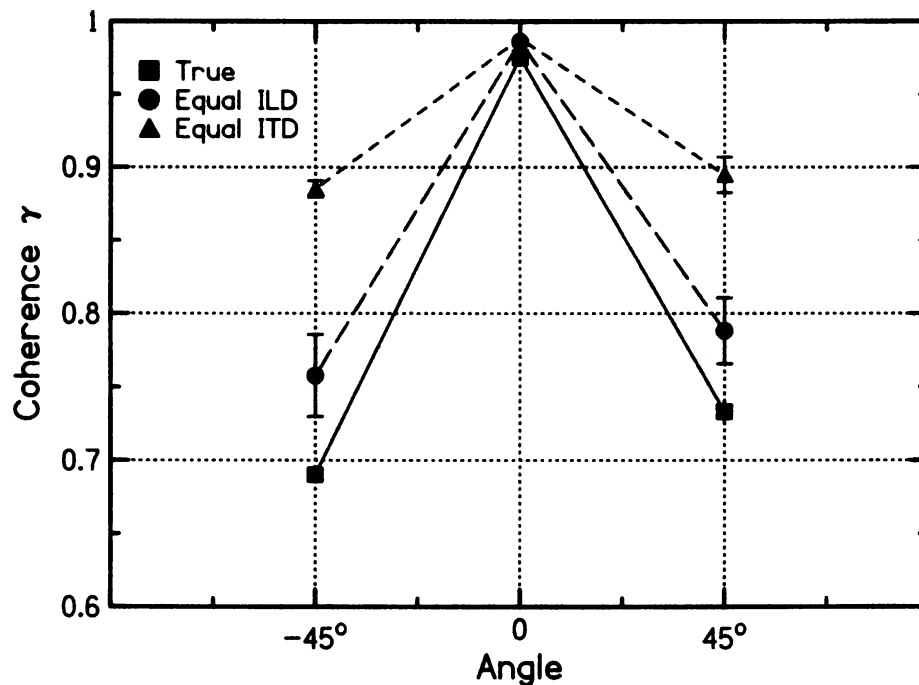


Figure 4.14. Coherences in a 500 Hz 1/3-octave band measured in KEMAR for sound at incident angles of $\theta = \{-45^\circ, 0^\circ, +45^\circ\}$. Symbols represent averages over five measurements and error bars extend one standard deviation in each direction. Square symbols represent the “true” interaural coherence, measured between unaltered signals recorded in the left and right ears. Circles represent coherences when the ILD was made to be constant within the frequency band, and thus are representative of the effect of varying ITDs within the band. Triangles represent coherences when the ITD was made to be constant within the band, and thus are representative of the effect of varying ILDs within the band.

The results of the calculations investigating the role of frequency-dependent ILDs and ITDs in the 500 Hz band were shown in Figure 4.14. When the ITDs within the 500 Hz band are set to be equal across the entire band (triangle symbols), the coherence at $\pm 45^\circ$ increased substantially compared to the true coherence. However, when the ILDs (but not the ITDs) were set to be equal across the entire band (circles), the coherence was much closer to the true coherence and thus much more incoherent. Thus, the difference in phase spectra within a 1/3-octave band is the primary source of incoherence.

The ITD spectrum within a band is computed by first finding the phase spectra

of the signals from each ear filtered into that band, taking the difference between the phase spectra, and dividing each component of the phase difference by its angular frequency. The ITD spectra for the bands with center frequencies of 500 Hz and 10 kHz are illustrated by Figure 4.15. In a high frequency band such as 10 kHz, where the coherence across angles is relatively flat, the ITD spectrum -45° , 0° , and $+45^\circ$ looks qualitatively flat, indicating a consistent ITD across the entire band. This is also true in the 500 Hz band at 0° , though there is more jaggedness than in the 10 kHz band. However, at $\pm 45^\circ$ in the 500 Hz band, a great deal of variation in the ITDs across the band is quite evident. Curiously, this only occurs when the loudspeaker is in the anterior half of the horizontal plane - a symmetric effect in the posterior half of the horizontal plane (which would occur at $\pm 135^\circ$) is not observed. This implies that such an effect could not be predicted by a simple theory of acoustical scattering from a rigid spherical head in which the interaural axis is exactly perpendicular to the MSP (antipodal).

Incoherence Due To Dispersion

The observed behavior in the interaural coherences near 45° incident sound direction has been attributed mainly to the phase spectrum of the signals measured in the ears of KEMAR. The phase spectrum can affect coherence if the ITDs are frequency-dependent within the 1/3-octave bands in which they are calculated. Over narrow bands, systematic dispersion, such as that found in the spherical head model, can be approximated as being linear. Here the effect of a linear dependence of the ITD on frequency is investigated. In this section, a mathematical framework is constructed for estimating the effect of such a linear dependence on interaural coherence and the results are compared to the measurements made earlier in this experiment.

Start by considering the signals in the left and right ears, $x_L(t)$ and $x_R(t)$ re-

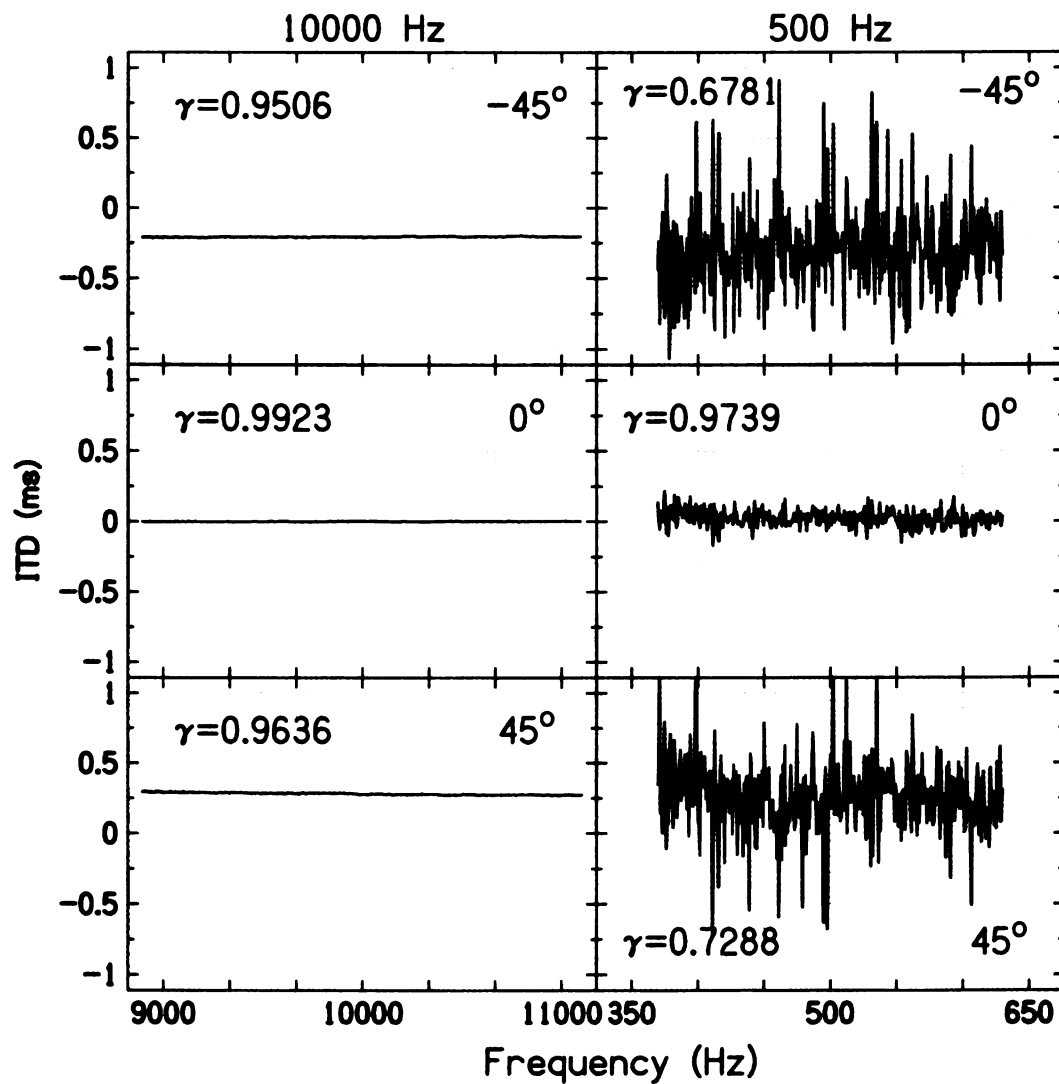


Figure 4.15. The ITD spectra in the 10000 Hz-centered band (left plots) and the 500 Hz-centered band (right plots). The three sets of plots show the variation in the ITDs within the indicated 1/3-octave frequency band when the KEMAR is oriented -45° , 0° , and $+45^\circ$ relative to the direction of the sound source. At $\pm 45^\circ$, the large variability of the ITDs in the 500-Hz band may lead to significant incoherence.

spectively. With an appropriately fine frequency spacing, any such signal can be represented as a sum of cosines with a phase and amplitude for each frequency component.

$$x_L(t) = \sum_{n=1}^{\infty} A_{L,n} \cos(\omega_n t + \phi_{L,n})$$

$$x_R(t) = \sum_{m=1}^{\infty} A_{R,m} \cos(\omega_m t + \phi_{R,m})$$

The cross-correlation of these signals is given by Equation 2.3. If the signals are long, then the frequency spacing is quite close and we can replace the numerator of Equation 2.3 by an integral over time.

$$c_{x_L x_R}(\tau) = \frac{\int_0^{\infty} dt \sum_{n=1}^N \sum_{m=1}^N A_{L,n} A_{R,m} \cos(\omega_n t + \phi_{L,n}) \cos(\omega_m (t + \tau) + \phi_{R,m})}{N \sqrt{P_L P_R}}$$

The integral that appears then in the numerator is greatly simplified by the mutual orthogonality of the cosine functions for different frequencies. Taking this into account, with some manipulation, this can be simplified to an integrated form:

$$c_{x_L x_R}(\tau) = \frac{\sum_{n=1}^{\infty} A_{L,n} A_{R,n} \cos(\omega_n \tau + \phi_{R,n} - \phi_{L,n})}{\sqrt{P_L P_R}} \quad (4.8)$$

Now, if it is assumed that the ILD is constant and the 1/3-octave band is flat, then the amplitudes A in the numerator of Equation 4.8 cancel with the calculations of power in the denominator. Of course, this experiment uses gammatone-shaped bands, but this calculation will suffice for the purpose of exploring the effect of dispersion on coherence in a slightly simpler case. The remaining cross-correlation equation is simply:

$$c_{x_L x_R}(\tau) = \sum_{n=1}^{\infty} \cos(\omega_n \tau + \phi_{R,n} - \phi_{L,n})$$

Finally, the analytic cross-correlation function can be easily calculated if the sum over frequencies is turned into an integral over the bandwidth. Having filtered the signals into a frequency band with center frequency ω_0 and bandwidth $\Delta\omega$;

and defining $\delta = \Delta\omega/2$ such that δ is half the bandwidth, we can write the cross-correlation in an analytic form:

$$c_{x_L x_R}(\tau) = \frac{1}{2\delta} \int_{\omega_0 - \delta}^{\omega_0 + \delta} \cos(\omega\tau + \phi_R - \phi_L) d\omega \quad (4.9)$$

Now consider the possible behavior of the phase difference in Equation 4.9. For ITDs that are constant over a frequency band, the relationship between the IPD, $\phi_L - \phi_R$, and ITD, T , is simply:

$$\phi_L - \phi_R = \omega T \quad (4.10)$$

As it is, this leads to a cross-correlation function that appears to be of the form of a cosine function enveloped by a sinc function, centered with a peak at the ITD, T :

$$c_{x_L x_R}(\tau) = \frac{\sin[\delta(T - \tau)]}{\delta(T - \tau)} \cos[\omega_0(T - \tau)] \quad (4.11)$$

This cross-correlation function is shown in Figure 4.16 for a center frequency of 1000 Hz, a 1/3-octave bandwidth, and an ITD of 0.27 ms.

Now, assume that the ITD is not constant, but varies linearly across the frequency band. Let the ITD at the center frequency $\omega = \omega_0$ be $T = T_0$. The ITD varies linearly from $T = T_0 - \zeta$ to $T = T_0 + \zeta$ over the frequency range $\omega = [\omega_0 - \delta, \omega_0 + \delta]$. This dispersion relation is depicted in Figure 4.17. The ITD can be written in this case as:

$$T = T_0 + \zeta(\omega - \omega_0)/\delta$$

The ITD at the center frequency of the band is T_0 . The parameter ζ characterizes the dispersion, and equals the difference between the ITD and T_0 at $\omega = \omega_0 \pm \delta$. Using Equations 4.9 and 4.10, the cross-correlation function becomes:

$$c_{x_L x_R}(\tau) = \frac{1}{2\delta} \int_{\omega_0 - \delta}^{\omega_0 + \delta} \cos \left[\omega(\tau - T_0) - \zeta \frac{\omega - \omega_0}{\delta} \omega \right] d\omega \quad (4.12)$$

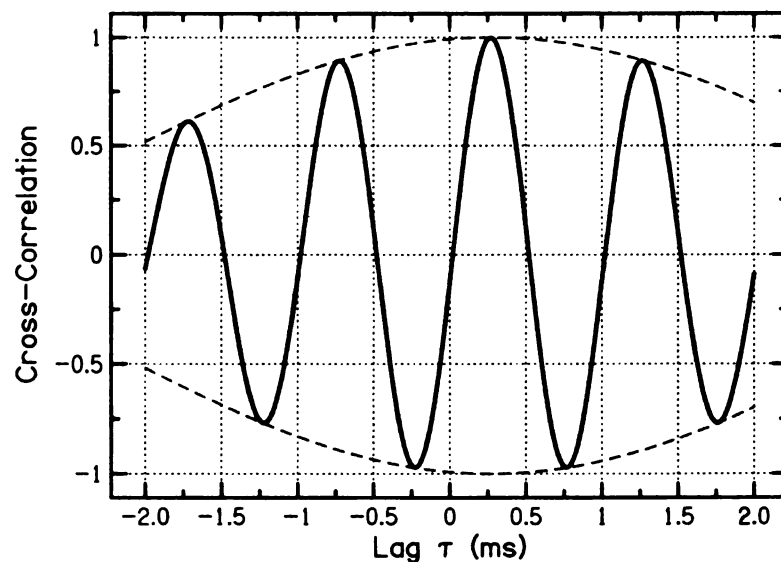


Figure 4.16. A plot of the cross-correlation function as described by Equation 4.11 with a center frequency of 1000 Hz, a 1/3-octave bandwidth, and an ITD of 0.27 ms. The dashed line shows the sinc function enveloping the cosine.

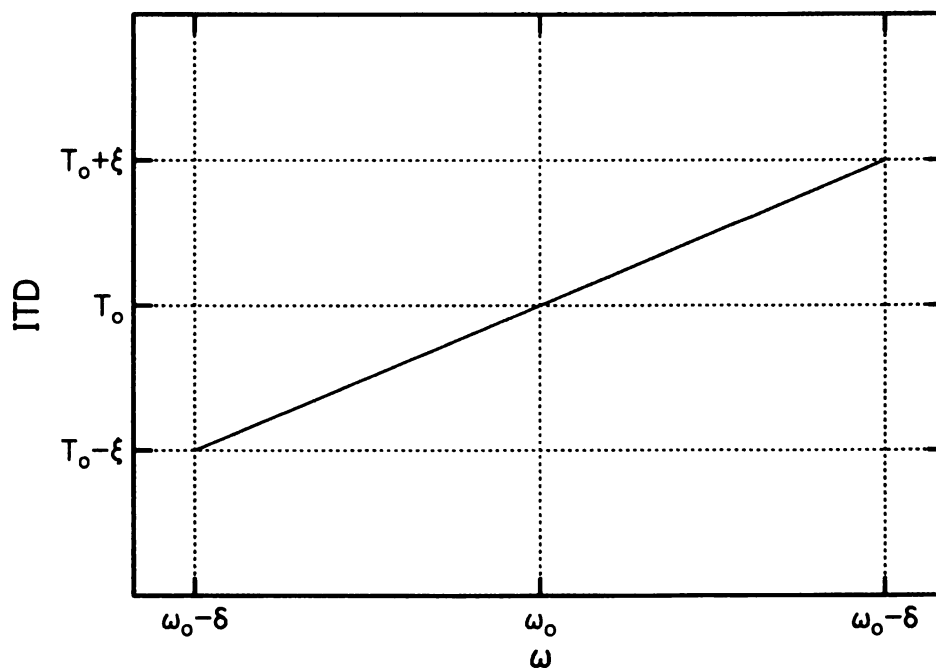


Figure 4.17. A linearly varying ITD over some finite frequency band with center frequency ω_0 and bandwidth 2δ . The parameter ξ describes the maximum difference between the ITD at any given frequency and the average ITD, T_0 .

A solution to this integral is tenable with some manipulation. Grouping terms in orders of ω , the argument of the cosine function is a quadratic. Completing the square gives:

$$\cos \left[\omega(\tau - T_0) - \xi \frac{\omega - \omega_0}{\delta} \omega \right] = \cos \left[\frac{\xi}{\delta} (\omega - \nu)^2 - \frac{\xi \nu^2}{\delta} \right]$$

where $\nu = \frac{\delta}{2\xi} \left(\tau - T_0 + \frac{\xi \omega_0}{\delta} \right)$ is used for brevity. Trigonometric expansion of the difference term in the cosine function leads to:

$$\begin{aligned} c_{x_L x_R}(\tau) = & \frac{1}{2\delta} \int_{\omega_0 - \delta}^{\omega_0 + \delta} \cos \left[\frac{\xi}{\delta} (\omega - \nu)^2 \right] \cos \left(\frac{\xi \nu^2}{\delta} \right) d\omega + \\ & + \frac{1}{2\delta} \int_{\omega_0 - \delta}^{\omega_0 + \delta} \sin \left[\frac{\xi}{\delta} (\omega - \nu)^2 \right] \sin \left(\frac{\xi \nu^2}{\delta} \right) d\omega \end{aligned} \quad (4.13)$$

Equation 4.13 involves integrals of the form of Fresnel cosine and Fresnel sine integrals. These are, as defined by Abramowitz and Stegun [1]:

$$C(z) = \int_0^z \cos \left(\frac{\pi}{2} u^2 \right) du \quad (4.14)$$

$$S(z) = \int_0^z \sin \left(\frac{\pi}{2} u^2 \right) du \quad (4.15)$$

The series expansion of the Fresnel integrals is:

$$C(z) = \sum_{n=0}^{\infty} \frac{(-1)^n (\pi/2)^{2n}}{(2n)!(4n+1)} z^{4n+1} \quad (4.16)$$

$$S(z) = \sum_{n=0}^{\infty} \frac{(-1)^n (\pi/2)^{2n+1}}{(2n+1)!(4n+3)} z^{4n+3} \quad (4.17)$$

The Fresnel Integrals are plotted in Figure 4.18. They have the limits:

$$\lim_{z \rightarrow \pm\infty} \{C(z), S(z)\} = \pm \frac{1}{2}$$

Using these gives an equation for the cross-correlation function in terms of the

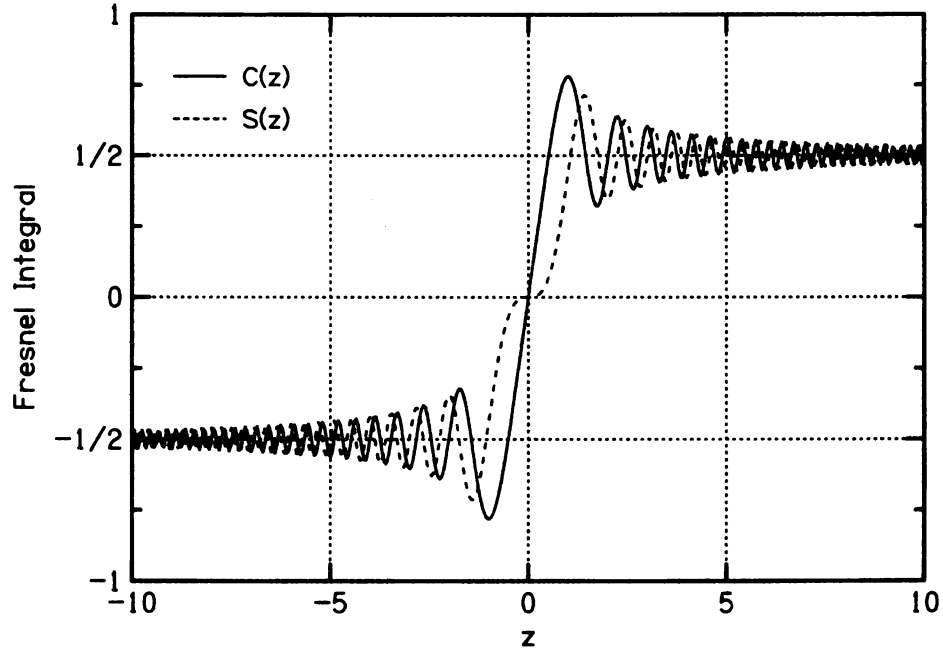


Figure 4.18. Plots of the Fresnel integrals $C(z)$ (solid line) and $S(z)$ (dashed line).

Fresnel integrals:

$$\begin{aligned}
 c_{x_L x_R}(\tau) = & \sqrt{\frac{\pi}{8\delta\xi}} \left\{ \cos \left[\frac{\xi v^2}{\delta} \right] \left(C \left[\sqrt{\frac{2\xi}{\pi\delta}} (\omega_o + \delta - \nu) \right] - \right. \right. \\
 & \left. \left. - C \left[\sqrt{\frac{2\xi}{\pi\delta}} (\omega_o - \delta - \nu) \right] \right) + \right. \\
 & \left. + \sin \left[\frac{\xi v^2}{\delta} \right] \left(S \left[\sqrt{\frac{2\xi}{\pi\delta}} (\omega_o + \delta - \nu) \right] - \right. \right. \\
 & \left. \left. - S \left[\sqrt{\frac{2\xi}{\pi\delta}} (\omega_o - \delta - \nu) \right] \right) \right\} \quad (4.18)
 \end{aligned}$$

Finding the ITD and waveform coherence from Equation 4.18 requires finding its first derivative with respect to the delay τ , $\frac{dc}{d\tau} = \frac{dc}{d\nu} \frac{d\nu}{d\tau}$, setting it equal to zero, and solving for τ . Unfortunately, the result does not lead to an algebraic function of τ . Further progress with Equation 4.18 can be made with numerical methods to solve for the estimated ITD, $\hat{T}$ (the value of τ which maximizes Equation 4.18) and the interaural waveform coherence, γ (the maximum value of Equation 4.18).

Note that Equation 4.18 involves the ITD, T_o . The actual value of ITD measured

by this method will not necessarily be equal to T_0 due to the dispersion, but will be shifted by some amount $\Delta T = \hat{T} - T_0$. Numerical simulation of the cross-correlation requires definition of a center frequency ω_0 , a bandwidth 2δ , the ITD T_0 , and the dispersion parameter ξ . For a few important center frequencies of a 1/3-octave bandwidth $2\delta = 2^{1/3}\omega_0$ and an ITD equal to that measured in KEMAR at 45° , the dependence of the ITD shift and coherence on ξ for values of ξ up to 0.1 ms is shown in Figure 4.19. This simulation is performed via numerical computation of Equation 4.18.

In frequency bands up to and including the 3150-Hz band, the trend for ΔT to increase and for γ to decrease with increasing ξ follows a common form. Simple equations may be fitted to these trends:

$$\Delta T = a \xi \quad (4.19)$$

$$\gamma = 1 - b \xi^m \quad (4.20)$$

The best fit values for the parameters a , b , and m are shown in Figure 4.19 for the frequency bands plotted there. The trend in these parameters as a function of frequency is depicted in Figure 4.20. The tendency for a to decrease with increasing frequency indicates that the ITD shift ΔT decreases with increasing frequency for any given dispersion parameter ξ , but the magnitude of this shift is rather small. The power parameter m in the coherence function generally decreases with increasing frequency, though over a very small range. The scaling parameter b , however, increases by two orders of magnitude between 160 and 3150 Hz, suggesting that linear dispersion can have a significant effect on coherence, especially at higher frequencies.

At frequencies higher than about 3150 Hz, the behavior of the ITD shift and coherence seems to diverge from Equations 4.19 and 4.20. In fact, since the functions in Equation 4.18 rely on the product of the dispersion parameter ξ and the

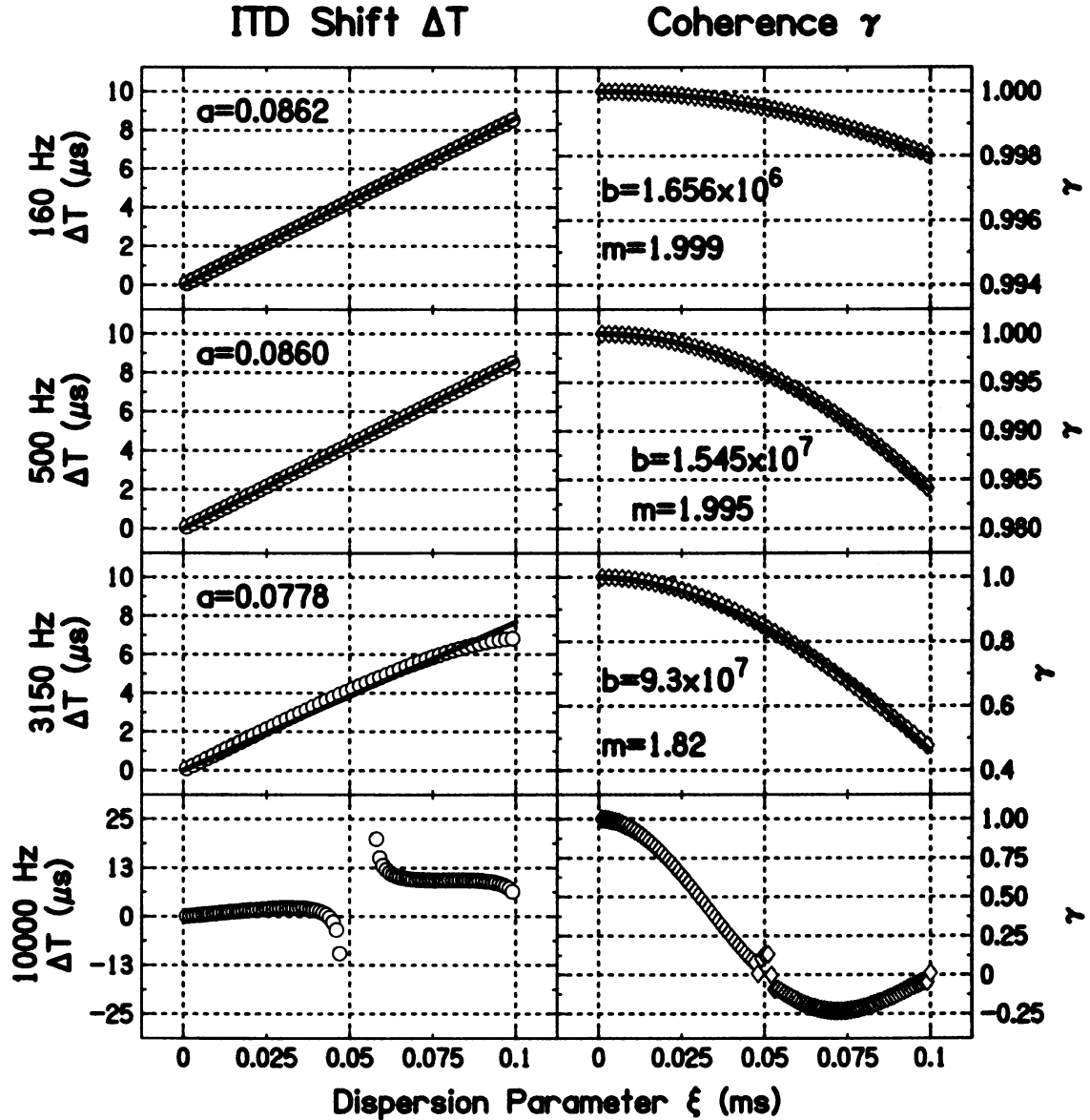


Figure 4.19. The ITD shift ΔT and the coherence γ for various bands with center frequencies indicated on the vertical axis and flat 1/3-octave bandwidths as a function of the dispersion parameter ξ . Data points for various values of ξ are shown in open symbols (circles for ΔT and diamonds for γ), and best fitting curves to Equations 4.19 and 4.20 are plotted in solid lines. As either ξ or center frequency increases, these equations do not hold, and so no such fit lines were plotted for the highest frequency band. In principle, only one plot is necessary because γ is a function of $\delta \xi$ (see Equation 4.23).

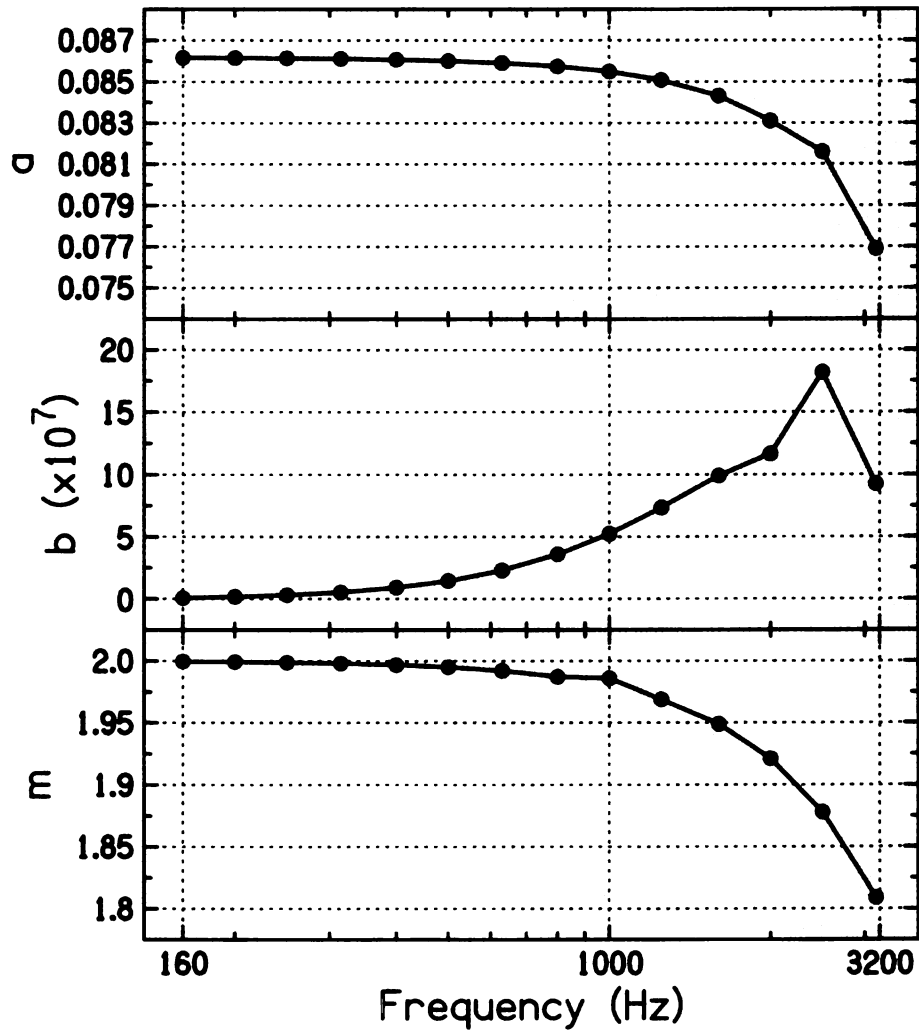


Figure 4.20. Best fit parameters a , b , and m across band center frequency. These parameters fit ITD shift data and coherence data to equations of the form of Equations 4.19 and 4.20.

center frequency ω_o (a dimensionless factor $\omega_o\zeta$), Equations 4.19 and 4.20 will fail to describe the behavior of ΔT and of γ as either ω_o or ζ become large.

Consider a calculation of the coherence, $\gamma = c_{x_L x_R}(\tau = \hat{T})$. First, the half-bandwidth δ depends on how wide the bands are chosen to be, but in general we may write:

$$\delta = \beta \frac{\omega_o}{2}$$

The parameter β determines the bandwidth. For instance, $\beta = 2^{1/3} - 1$ would give 1/3-octave bandwidths. For small $\omega_o\zeta$, a simplification to Equation 4.18 may be made by noting that at low frequencies, $\hat{T} - T_o \simeq a\zeta$ (Equation 4.19). Then the parameter ν becomes:

$$\begin{aligned} \nu &= \frac{\delta}{2\zeta} \left(a\zeta + \frac{\omega_o\zeta}{\delta} \right) \\ &= \frac{\delta}{2} \left(a + \frac{2}{\beta} \right) \\ &= \frac{\omega_o}{2} \left(\frac{a\beta}{2} + 1 \right) \end{aligned}$$

The argument of the trigonometric functions in Equation 4.18 is then:

$$\frac{\zeta\nu^2}{\delta} = \frac{\omega_o\zeta}{2\beta} \left(\frac{a\beta}{4} + \frac{1}{2} \right)^2$$

If the frequency is small, then the function $\sin(\zeta\nu^2/\delta)$ will be much smaller than $\cos(\zeta\nu^2/\delta)$ since the product $\omega_o\zeta$ is small. In order for the sin function to be less than 5% of the cosine function, it is required that the factor $\omega_o\zeta$ must be smaller than some limit that depends on a and β . For 1/3-octave bandwidths and using $a = 0.086$, this limit is $\omega_o\zeta < 0.0254$. For a dispersion parameter of $\zeta = 0.1$ ms, this would allow a maximum center frequency of only 40.4 Hz. It seems that this is an unnecessarily strict condition to place on the usage of this approximation since a glance at Figure 4.19 suggests good agreement with this approximation through at least the 500-Hz band.

Assuming small $\omega_o \zeta$, the terms involving the sine function are ignored. Now, the remaining Fresnel integrals may be series expanded to first order via Equation 4.16 ($C(z) \simeq z$). After some elementary simplification, what is left is a low-frequency approximation for γ :

$$\gamma = \cos \left[\frac{\omega_o \zeta}{2\beta} \left(\frac{a\beta}{4} + \frac{1}{2} \right)^2 \right]. \quad (4.21)$$

Expanding this as a series for small values of the product $\omega_o \zeta$ yields:

$$\gamma \simeq 1 - \frac{\omega_o^2 \zeta^2}{8\beta^2} \left(\frac{a\beta}{4} + \frac{1}{2} \right)^4. \quad (4.22)$$

This equation has the same form as Equation 4.20 with $m = 2.0$, the approximate value which best fits frequency bands up to about 1000 Hz (see Figure 4.20). The coefficients on ζ^2 are reasonably close in this approximation to those calculated by fitting Equation 4.20. For instance, given a center frequency of 160 Hz, $\beta = 0.26$ (a 1/3-octave band), and $a = 0.0862$, the coherence becomes:

$$\gamma = 1 - 1.954 \times 10^6 \zeta^2$$

The coefficient of ζ^2 has units of inverse-seconds squared. This is close to the value $b = 1.656 \times 10^6$ that was found from fitting Equation 4.20 to the data generated from the complete solution, Equation 4.18. Alternately, this may be expressed as a function of the bandwidth δ times ζ , in which case the multiplicative coefficient is unitless:

$$\gamma \simeq 1 - \frac{1}{2} (\delta \zeta)^2 \left(\frac{a}{4} + \frac{1}{2\beta} \right)^4 \quad (4.23)$$

Plugging in the aforementioned values of β and a , this becomes:

$$\gamma = 1 - 7.15 (\delta \zeta)^2$$

A derivation of the ITD for small $\omega_o \zeta$ involves setting the derivative of Equation 4.13 or Equation 4.18 with respect to τ equal to zero and solving for τ (or

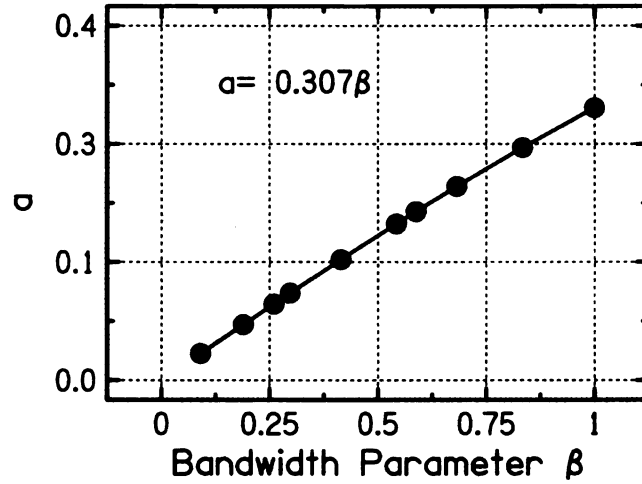


Figure 4.21. The parameter a relating ΔT to the dispersion parameter ζ as a function of the bandwidth parameter β . An approximate relationship between a and β is given by Equation 4.24 to be $a = \beta/3$.

$\tau - T_0$). Efforts to this extent have thus far proved ineffective. If simplifications are made to the extent that ΔT is linear in ζ , the coefficients are largely inaccurate compared to the values that led to Equation 4.20 and Figure 4.20. The difficulty likely lies in the oscillatory nature of the functions involved. Without delving into the details of the derivation, the relationship between ΔT and ζ for small values of $\omega_0 \zeta$ best fits a relationship of the form:

$$\Delta T = \frac{\beta}{3} \zeta \quad (4.24)$$

For a 1/3-octave bandwidth, $\frac{\beta}{3} = 0.0867$. This is very close to the value found when using the complete solution (Equation 4.18), $a \simeq 0.086$. Returning to the computer model using Equation 4.18, it is possible to find the coefficient a from Equation 4.19 as a function of β . This relationship is shown in Figure 4.21. Though the exact solution gives $a = 0.307\beta$ instead of $a = \beta/3$ (predicted in the limit of small $\omega_0 \zeta$), the two are quite close. Thus, useful equations have been found to predict the behavior of the coherence and the measured ITD in the case of linear ITD dispersion. The telltale sign of the presence of linear dispersion is not in the

ITD, where even if the average ITD T_0 is known. The difference between it and the calculated ITD τ is very small according to Equation 4.24, even for a calculation within broad bandwidths. Coherence within bands is a far more obvious indicator of linear dispersion. For even a moderate dispersion, the resulting incoherence is noticeable, and becomes more extensive with increasing frequency. Linear dispersion cannot describe the behavior of the interaural coherence measured in KEMAR at angles near 45° , however, unless the dispersion parameter ζ decreases with increasing frequency. Certainly it is possible that ζ is a function of both frequency and incident angle, but it is unlikely to be the only factor in the incoherence seen in KEMAR. Even then, the linear dispersion observed in the ITD spectrum of measurements made in KEMAR is rather small. In the 500 Hz band, for instance, the linear dispersion is approximately 0.1 ms, which would result in a coherence of approximately 0.985, far larger than the coherence actually observed. The noisiness of the ITD spectrum is also a likely contributor (see Figure 4.15), and may in fact be the dominant factor in determining coherence.

4.1.5 Incoherence due to Pseudorandom Dispersion

A glance at Figure 4.15 suggests a noisy ITD spectrum at lower frequencies, which may contribute to incoherence along with the apparent linear (or approximately linear) dispersion present in the ITD spectrum. Thus, a new model is proposed for the cross-correlation function. In this model, the ITD is a Gaussian random variable with mean T_0 equal to the average ITD across the band, and standard deviation σ .

$$\text{ITD} \sim \mathcal{N}[T_0, \sigma]$$

Alternatively, the ITD may be written as:

$$\text{ITD} = T_0 + n$$

where n is normally distributed about zero with standard deviation σ .

Beginning with Equation 4.9, the IPD can be written as:

$$\phi_R - \phi_L = \omega (T_o + n)$$

And the cross-correlation function then becomes:

$$c_{x_L x_R}(\tau) = \frac{1}{2\delta} \int_{\omega_o - \delta}^{\omega_o + \delta} \cos[\omega(\tau - T_o) + \omega n] d\omega \quad (4.25)$$

A computer simulation of Equation 4.25 is evaluated for values of the standard deviation on the range $\sigma \in [0, 50 \mu\text{s}]$ in $1 \mu\text{s}$ steps. This is accomplished by computing a Riemann sum. The random additional phase ωn has a different random value for each frequency component drawn from $\mathcal{N}[0, \sigma]$ added to it. These random added phases are held constant as the cross-correlation function is calculated for different lags τ . Lags are computed between -2.0 ms and 2.0 ms in steps of $10 \mu\text{s}$. For each value of σ , the maximum value of the cross-correlation function (the coherence γ) and the difference in measured versus mean ITD ($\tau - T_o$) is recorded. For each value of σ , this process was repeated ten times and average coherences and ITDs are computed.

The results of the computation of cross-correlations with pseudorandom phases are depicted for a few frequencies in Figure 4.22. The coherences behave in a way simply related to the center frequency of the band and the noise variance σ^2 :

$$\gamma = \text{Exp}\left[-\frac{1}{2}\omega_o^2\sigma^2\right] \quad (4.26)$$

In fact, this result can be derived from Equation 4.25. First, the cosine function is expanded to give:

$$c_{x_L x_R}(\tau) = \frac{1}{2\delta} \int_{\omega_o - \delta}^{\omega_o + \delta} \{\cos[\omega(\tau - T_o)] \cos[\omega n] - \sin[\omega(\tau - T_o)] \sin[\omega n]\} d\omega$$

Now, the coherence occurs for $\gamma = c_{x_L x_R}(\tau = T_o)$. Thus, the sine terms vanish

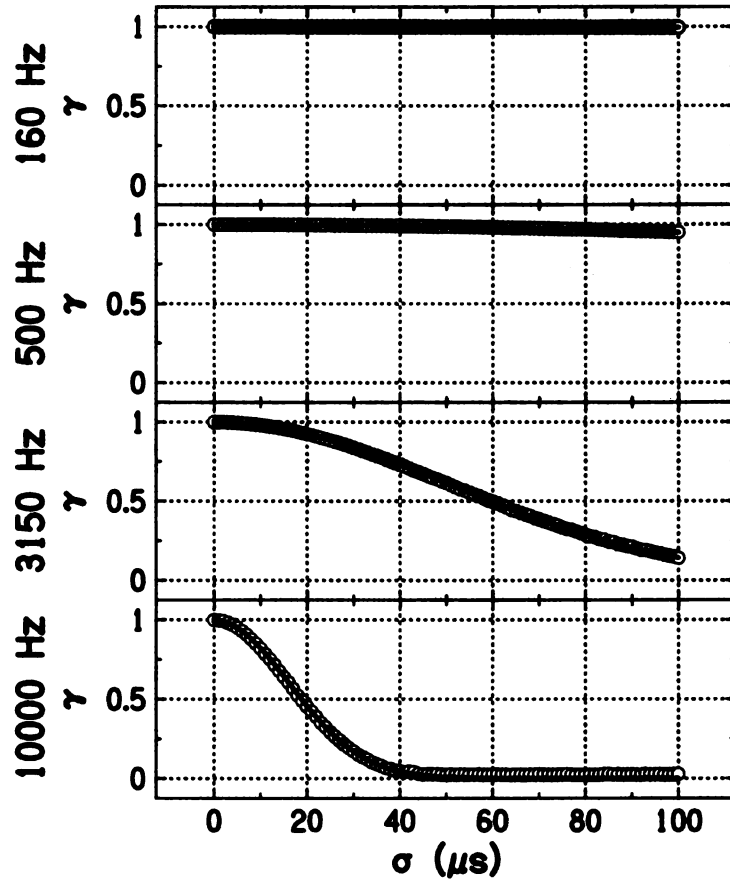


Figure 4.22. Coherence in various 1/3-octave bands as a function of the standard deviation of the noise added to the ITD of the signals. Best fitting curves corresponding to Equation 4.26 are drawn with the data.

entirely, and what is left is:

$$\gamma = \frac{1}{2\delta} \int_{\omega_0-\delta}^{\omega_0+\delta} \cos [\omega n] d\omega$$

In this form, the integral is impossible to evaluate, but statistical methods allow an evaluation of the ensemble average of the coherence $\langle \gamma \rangle$. In fact, this is what the results of the numerical calculation of Equation 4.25 represent.

$$\langle \gamma \rangle = \frac{1}{2\delta} \int_{\omega_0-\delta}^{\omega_0+\delta} \langle \cos [\omega n] \rangle d\omega \quad (4.27)$$

The ensemble average of $\cos [\omega n]$ is simple if we note that the product ωn is, for any given value of ω , a normally-distributed random variable with mean 0 and standard deviation $\omega\sigma$. If no value of ω carries any more or less weight than any other value, as is true here, then ω can be replaced in the ensemble average with the average value of ω . Thus, the ensemble average can be equivalently written as:

$$\langle \cos [\omega n] \rangle = \langle \cos [\omega_0 n] \rangle$$

The ensemble average of the cosine function can be written as an integral involving the probability density function (PDF) $f(n)$ of the normally distributed random variable n :

$$\langle \cos [\omega_0 n] \rangle = \int_{-\infty}^{\infty} f(n) \cos (\omega_0 n) dn \quad (4.28)$$

This can be compared to the characteristic function of a normally distributed random variable evaluated at ω_0 , $\chi_n(\omega_0)$, with mean zero:

$$\begin{aligned} \chi_n(\omega_0) &= E [e^{i\omega_0 n}] \\ &= \int_{-\infty}^{\infty} f(n) e^{i\omega_0 n} dn \end{aligned} \quad (4.29)$$

Now since the PDF of a normally-distributed random variable with mean 0 is a purely even function, the odd (sine) components of the exponential term vanish in

the integration, leaving only the even (cosine) terms. This leaves:

$$\chi_n(\omega_o) = \int_{-\infty}^{\infty} f(n) \cos(\omega_o n) dn \quad (4.30)$$

Fortunately, this is exactly the integral of Equation 4.28, and so the ensemble average coherence $\langle \gamma \rangle$ is exactly equal to this characteristic function. The characteristic function of a normally-distributed random variable is well understood. For a normally-distributed random variable with mean 0 and standard deviation σ , the characteristic function evaluated at x is:

$$\chi_n(x) = \text{Exp} \left[-\frac{1}{2} \sigma^2 x^2 \right]$$

The ensemble average coherence is therefore:

$$\langle \gamma \rangle = \text{Exp} \left[-\frac{1}{2} \omega_o^2 \sigma^2 \right] \quad (4.31)$$

This is exactly the form of the coherence found by fitting curves to the numerical data (Equation 4.26).

It should be noted now that σ as it has been defined here is the standard deviation in the noise inherent in the phase *difference* between two signals, and not the noise inherent in the phase of either signal independently. However, if the noise in the phase of each signal is independent and identically distributed, then the variance in each signal, σ_s , is:

$$\sigma_s^2 = \sigma^2 / 2$$

Then the coherence may be related to the variance in the noise inherent to the ITD spectrum of each signal by:

$$\gamma = \text{Exp} \left[-\omega_o^2 \sigma_s^2 \right]$$

The incoherence noticed in KEMAR for incident angles near 45° is likely a combination of linear dispersion and a noisy ITD spectrum. It is apparent that even a modest standard deviation in the ITD spectrum can lead to significant incoherence.

Thus, it is likely the variance, and not the slope, of the ITD spectra that lead to significant incoherence. Consider for instance the 500-Hz band shown in Figure 4.15. At 45° , the standard deviation in the ITD is $\sigma = 0.27$ ms. Plugging this into Equation 4.31 gives $\langle \gamma \rangle = 0.7061$, which is very close to the coherence arrived at via the cross-correlation function, $\gamma = 0.7288$.

4.1.6 Comparison of KEMAR Ears

It is important to note the possibility of absolute differences in the behavior or response of the left and right ears of the KEMAR. In the previous results of this experiment, such differences were not noticeable because those measurements were only concerned with how those binaural differences changed as the angle between the direction of incident sound and the facing direction of the KEMAR changed. In later studies concerning room acoustics, however, several measurements will be taken using a head-on (0°) geometry.

At an angle of 0° , an ideal model of the head where the ears are an equal angular displacement from the center of the face gives zero ITD and zero ILD. In an ideal anechoic environment where there is no random incidence of reflected sound on the ears, coherence is expected to be 1 in all 1/3-octave bands. The same results are expected for an angle of 180° since the ears would again be equally separated from the sound source. However, asymmetry between the ears may lead to deviations from these ideals.

Differences in measurements at 0° and 180° may be due either to external or internal effects. Internal effects are those such as asymmetries in the location of the KEMAR ears relative to the center of the face, and differences in the height, shape, or canal length of the ears, differences in the response of the electronic components at the ears of the KEMAR or in other electronic components involved in the transmission and processing of the received sound. External effects are mainly those

of the auditory environment itself such as reflections and reverberation. To identify and separate external and internal effects, measurements of coherence, ITD, and ILD in 1/3-octave bands taken at both 0° and 180° were compared. Theoretically, both the ITD and ILD should be zero when both ears are equidistant from the sound source in an anechoic environment. Any measured values that are different from zero may be attributed to either internal or external effects. If such a difference occurs but changes sign between 0° and 180° , then this difference is likely an external effect, while those that keep the same sign under a 180° rotation are likely internal.

ITDs in 1/3-octave bands from 80 to 16,000 Hz at 0° and 180° for both positions 1 and 2 are shown in Figure 4.23. Measurements at 0° are shown in circles, and measurements at 180° are shown in triangles. Open symbols are for measurements taken at position 1 and closed symbols are for measurements taken at position 2. Error bars smaller than the size of the points are invisible. The ITD is nearly zero, as expected, everywhere except at frequencies at or below 250 Hz, where the ITD is negative at both 0° and 180° . Errors in ITD at low frequencies may be due to the breadth of the peak in the cross-correlation functions at low frequencies (see Figure 2.2), leading to errors in estimation of the location of the peak. This also accounts for the variance seen in ITD measurements at low frequencies, whereas there is almost no variation in ITD estimations in higher frequency bands.

Since the ITDs at low frequencies are negative for measurements taken at both 0° and 180° at either position, this suggests an internal effect. One possibility is that a capacitor in the preamplifier is responsible for causing a phase shift in low-frequency inputs, though it is unusual for such a thing to happen for frequencies as high as 200 Hz, as is observed here. If internal effects are indeed responsible, then there may also be an effect noticeable in the measurement of coherence across frequency bands. Such a plot is shown in Figure 4.24. Another possible explanation

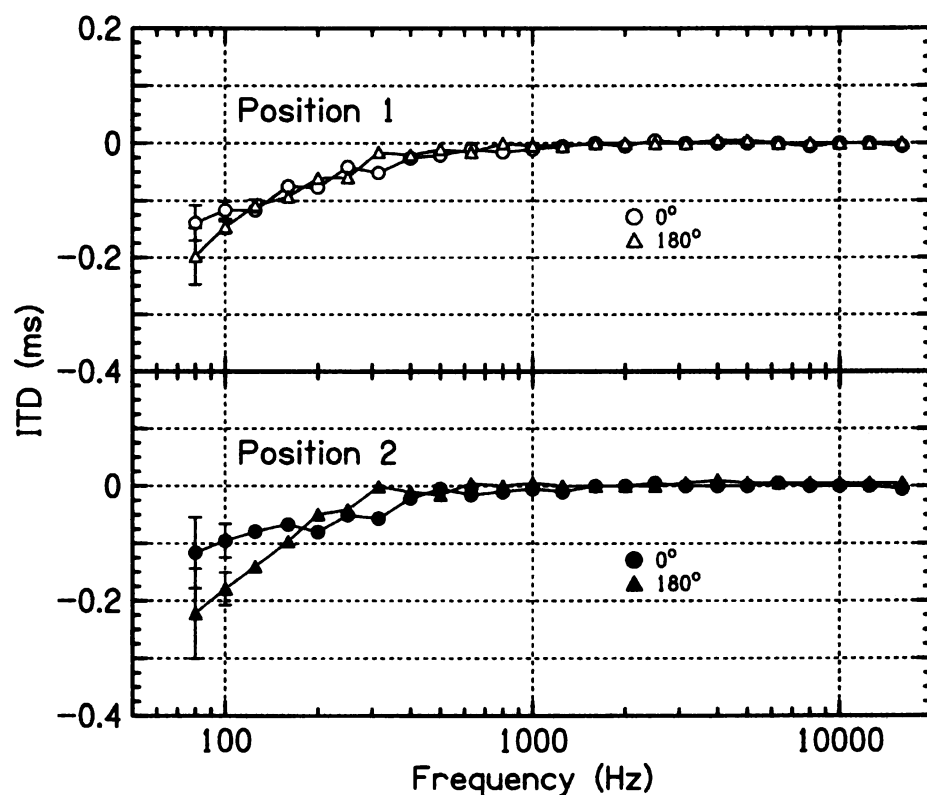


Figure 4.23. Measurements of the ITD taken at 0° (circles) and 180° (triangles) in an anechoic environment in 24 $1/3$ -octave bands from 80 to 16000 Hz. Measurements were taken at two different positions in the anechoic room, each with the same separation between KEMAR and loudspeaker. Measurements taken at position 1 are indicated by open symbols and are shown in the top panel. Measurements taken at position 2 are indicated by filled symbols and are shown in the bottom panel.

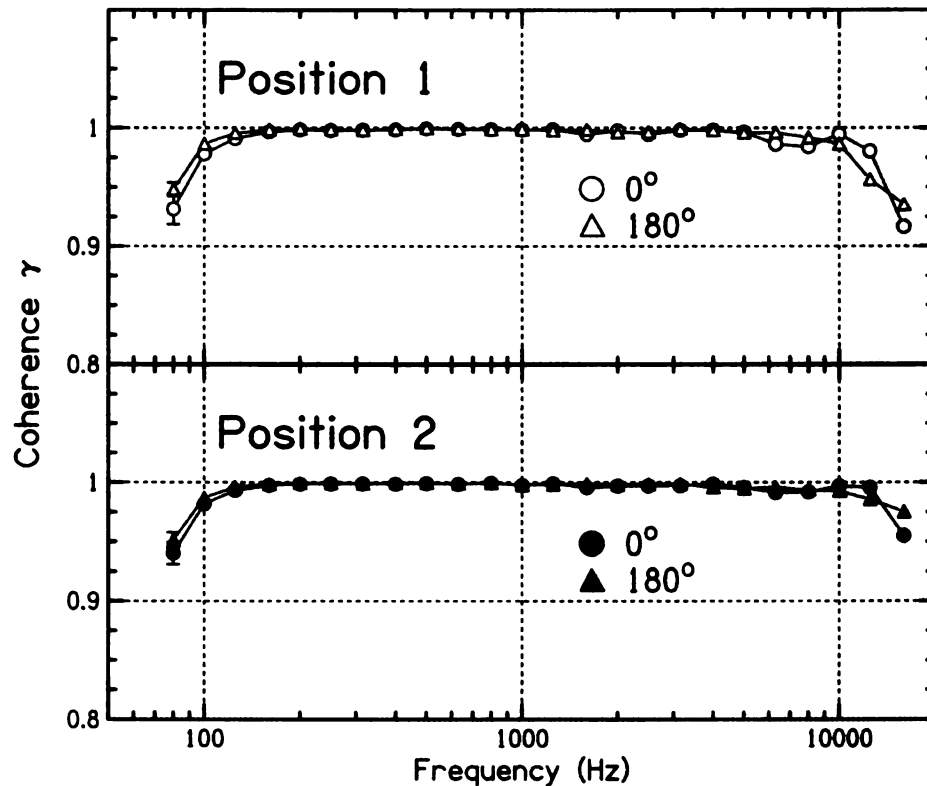


Figure 4.24. Measurements of the interaural coherence taken at 0° (circles) and 180° (triangles) in an anechoic environment in 24 $1/3$ -octave bands from 80 to 16000 Hz. Measurements were taken at two different positions in the anechoic room, each with the same separation between KEMAR and loudspeaker. Measurements taken at position 1 are shown in the top panel, and measurements taken at position 2 are shown in the bottom.

lies in the previously noted imperfection of the anechoic environment, resulting in some reflections at low frequencies. However, external effects are expected to change sign between 0° and 180° , and no such sign change is noticed here.

In Figure 4.24, a drop in coherence can be seen for frequencies at or below 125 Hz at both positions 1 and 2. This is likely due to the imperfection of the anechoic environment at low frequencies. The deviation in measured coherences from the expected value of 1 is more limited in the range of frequencies over which it occurs than the imperfections seen in ITD. Thus, it seems that the deviations in ITD at frequencies below 125 Hz is likely a combination of room (external) effects

and internal effects. If so, then a similar phenomenon is expected when ITDs are measured in other locations due to the internal effect. It should be noted, however, that the front-back differences suggest that ITDs agree with ideal predictions at all but the lowest frequency band of 80 Hz. Deviations in coherence at high frequencies in the anechoic room are likely an external effect caused by reflections off of objects in the anechoic room which are not padded. At 10000 Hz, the wavelength of sounds in air is about 3.5 cm, and even small reflective objects may produce some scattering.

Measurements of the ILD at 0° and 180° are shown in Figure 4.25 in a manner similar to that of Figure 4.24. First, it should be noted that most ILDs are above 0 dB, usually near 1 dB. This consistent difference is likely due to a mismatch in gain between the left and right channels of the preamplifier (an internal effect). This mismatch seems to have been consistent over measurements made at both positions 1 and 2, as evidenced by good agreement of measurements made at 0° and at 180° at both positions.

There are significant ILDs in the measurements. These are especially noticeable at high frequencies, but the magnitude of the ILDs observed are relatively small compared to the range of naturally-occurring ILDs at these frequencies (see for instance Figure 4.9). In the 16,000 Hz center frequency, an ILD of about -2 dB is observed for an angle of 0° and an ILD of more than 2 dB at 180° . This indicates excellent symmetry about the setup angle at high frequencies, where sensitivity to small errors in orientation is expected to be greatest. Given the sign change, this seems to be an external effect of the room and may point to another frequency region where the behavior of the anechoic room is not ideal. Support for this notion comes from the high frequency coherences (see Figure 4.24), which drop below 1 at high frequencies.

Some reverberation in the room may occur for high frequency bands, where

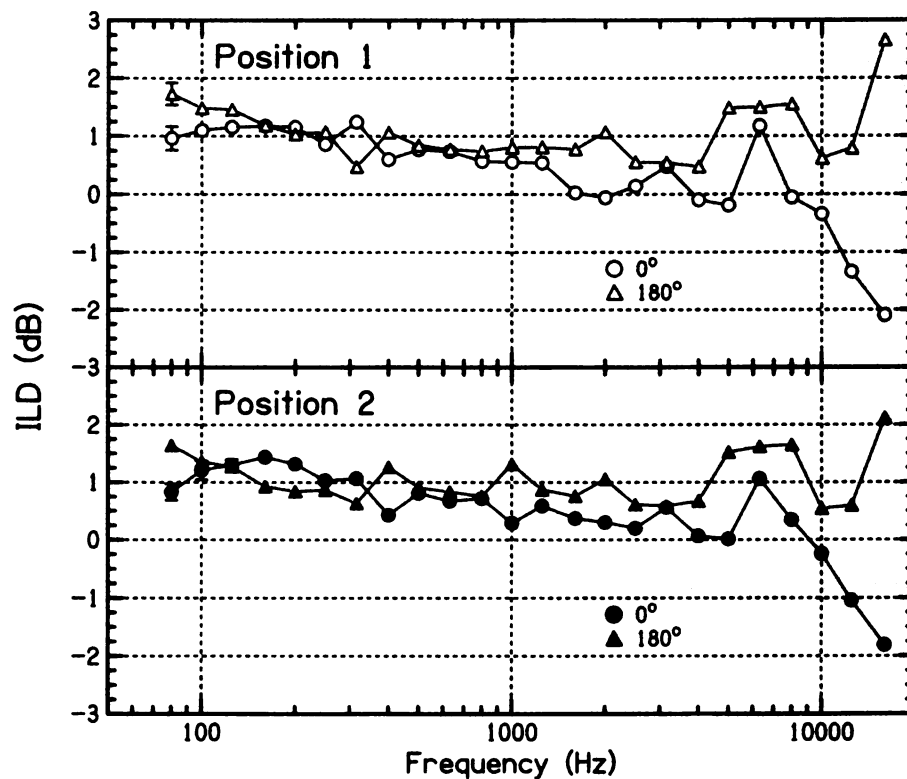


Figure 4.25. Measurements of the ILD taken at 0° (circles) and 180° (triangles) in an anechoic environment in twenty-four 1/3-octave bands from 80 to 16000 Hz. Measurements were taken at two different positions in the anechoic room, each with the same separation between KEMAR and loudspeaker. Measurements taken at position 1 are indicated by open symbols of either type and are located in the top panel. Measurements taken at position 2 are indicated by filled symbols and are located in the bottom panel.

reflections off of even small or thin non-absorbing objects in the room may occur. This would then lead to a decrease in interaural coherence and an ILD. A patch panel near one corner of the room, though curtained with sound-absorbing foam, is a likely source of such reflections. However, it seems unusual for the effect to persist in the same manner when the KEMAR and Mackie were moved from one position to another. Instead, this effect at high frequencies might result from asymmetries in the pinnae of the left and right ear.

The observed interaural differences are unexpected in the ideal model of a rigid, spherical head with symmetrically located ears. At 0° and 180° , the sound source is located in the MSP. Though measurable, these binaural cues in the median sagittal plane are small, and are not thought to be particularly useful in aiding sound localization [70].

4.2 Further Experiments on the Coherence at 45°

Thus far, the degraded coherences measured in KEMAR near angles of $\pm 45^\circ$ have gone largely unexplored. In this section, several experiments are presented which attempt to determine the cause for these incoherences and their possible effect on human listeners. Experiments are also presented which attempted to measure this effect in listeners.

4.2.1 Variations on Methods of Coherence Measurement in KEMAR

A set of further measurements were made with the KEMAR in anechoic conditions to further investigate the nature of the drops in coherence at $\pm 45^\circ$. Under test conditions identical to those previously used in Experiment 9, several more

measurements were made at angles of 45° , 0° , and -45° . For each set of measurements, the experimental setup was varied in one of the following ways:

- A. KEMAR wearing external microphones near ears
- B. KEMAR with both pinna removed
- C. KEMAR with both pinna removed and skull smoothed
- D. KEMAR only left pinna removed
- E. KEMAR only right pinna removed
- F. JVC foam dummy head
- G. KEMAR as normal for this experiment, with both pinnae
- H. KEMAR without torso

Methods

In variation A, KEMAR was fitted with two small microphones placed about 1 cm in front of the KEMAR ear canals on each side of its head. Recordings were made through those microphones instead of through the KEMAR ears. This effectively eliminated the effects of the pinna, ear canal, and internal electronics of the KEMAR itself.

In variation B, recordings were made through the KEMAR ears, but both pinnae were removed. This left a recess on each side of the KEMAR head in which was the KEMAR internal microphones. This eliminated the contribution of both the KEMAR pinnae and ear canals but not the internal recording equipment.

In variation C, the recesses left on each side of the KEMAR head when the pinnae were removed were filled by pieces of cardboard. The cardboard pieces each had a hole cut into them such that sound could get through to the KEMAR microphones. In this way, the KEMAR was fitted with ear canals while still lacking pinnae. This effectively eliminated the indentations on each side of the KEMAR head, leaving a smooth surface with a hole for the ear canal.

In variations D and E, only one pinna was removed at a time. Nothing was inserted to take the place of the missing pinna. Measurements at all three angles of

interest were made with the only left pinna missing, and again with only the right pinna missing.

In variation F, the KEMAR was not used at all, but instead a JVC foam dummy head designed to make binaural recordings was used. The JVC head was fitted with a headphone-like apparatus with microphones on the outside where the ear canals of the dummy head would be. The JVC head was attached to a microphone stand and its height off the ground was adjusted so that the height of the JVC microphones was the same as those of the KEMAR ears.

In variation H, the KEMAR head was removed from its torso and situated on a microphone stand. The height of the stand was adjusted such that the KEMAR ears were at the same height as they would have been with the torso still attached. Both pinnae were in place, and recordings were made through the KEMAR ears. In this way, the effect of the torso could be measured.

Results

The results of the coherence measurements made at 45° , 0° , and -45° for all seven variations of the experimental setup are shown in Figure 4.26. At an angle of 0° , the coherence does not differ much across different situations, except for some minor differences at very low and very high frequencies. At angles of $\pm 45^\circ$, however, there is a clear effect in certain variations.

At -45° , coherence is significantly worse at frequencies up to about 1250 Hz in variations D (only right pinna present) and G (both pinna present). At 45° , coherence is degraded for variations E (only left pinna present) and G. No other variation changes the coherence so significantly at $\pm 45^\circ$ compared to 0° as do variations D, E, and G.

In variations D and E, the coherence is only degraded when the pinna that is present is near to the sound source. For instance, when the KEMAR is turned

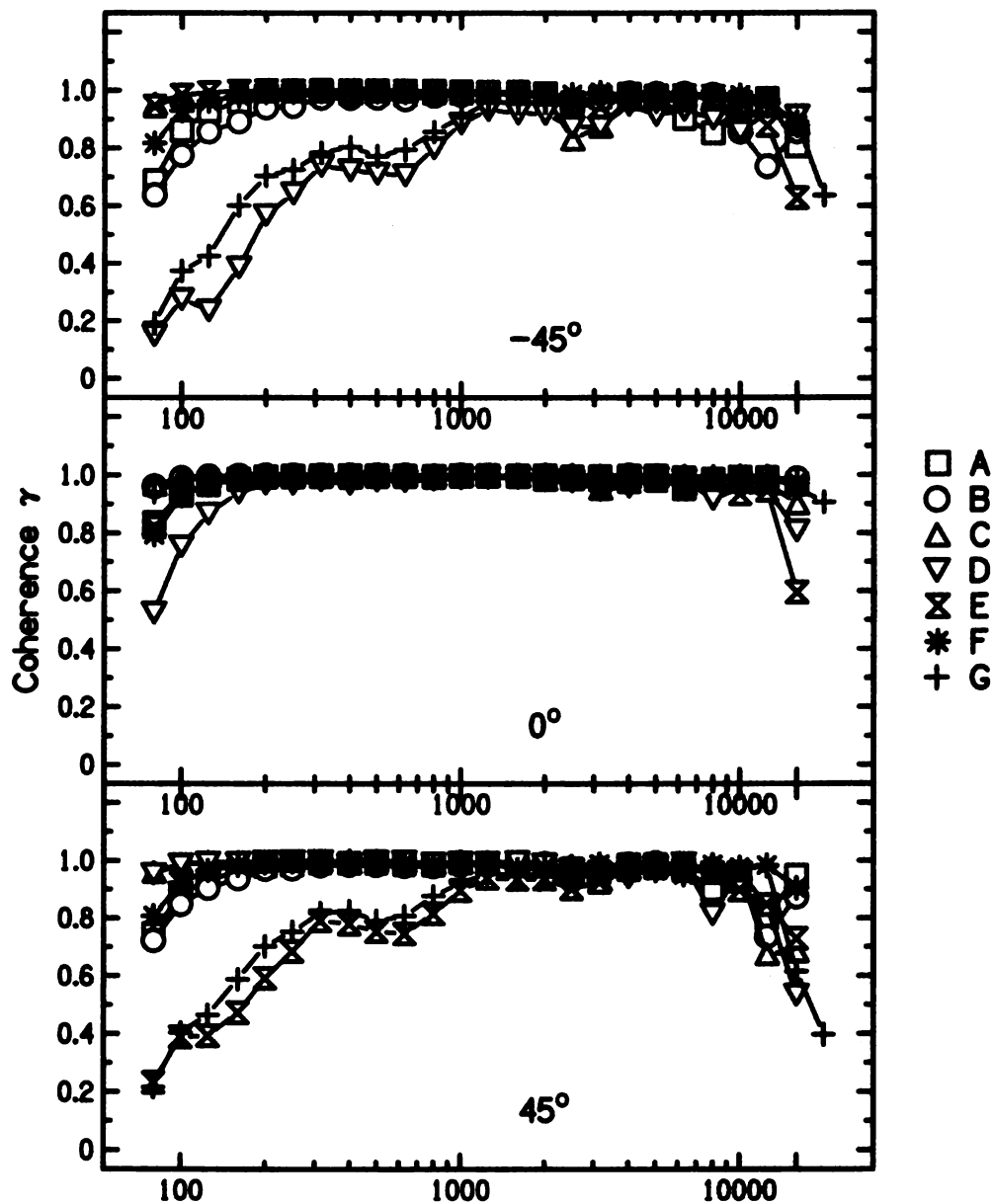


Figure 4.26. Coherences measured in 1/3-octave bands in an anechoic environment using several variations of the experimental apparatus to record the signals in both ears. Each variation is labeled with a different letter and is plotted separately with different symbols. The variations are A: measurements through headphone-mounted microphones placed on KEMAR; B: recordings through KEMAR ears but with pinnae removed; C: as variation B, but with cardboard "ear canals;" D: as variation B, but with only the left pinna removed; E: as variation B, but with only the right pinna removed; F: recordings through microphones mounted on a JVC foam dummy head; G: recordings made through KEMAR normally with both pinnae present.

to -45° , the loudspeaker is to its right, and the right ear is closer to the sound source than the left ear. If the right pinna is present, then the coherence is less than it would be if the right pinna were absent, regardless of the presence of the left pinna. A similar point can be made about the left pinna in the 45° configuration. The unequivocal result is that the drops in coherence only occur when the pinna of the ear nearer the sound source is present.

Results of Removing the Torso

The effect of removing the torso can be seen by comparing variations G and H as shown in Figure 4.27. The presence of the torso had a statistically significant effect on the coherence (in a one-sided, paired t -test, $t = 5.37$, $df = 37$, $p < 0.001$), but it is clearly not the dominant effect. The mean improvement in coherence, plus and minus one standard deviation, at $\pm 45^\circ$ when the torso is removed is 0.03 ± 0.033 . Even if only the frequency bands below 1 kHz are considered, the mean improvement in coherence when the torso is taken away is 0.05 ± 0.035 . The effects of the shoulders and torso have been noted to have an effect on the measurement of head-related transfer functions (HRTFs) at frequencies as low as 1 kHz [25]. Low-frequency effects of the shoulders and torso have been previously noted to provide important elevation cues [5] (as do the pinnae). However, the effect observed here would point to a low-frequency effect of the torso in the horizontal plane, albeit a small one.

The degradation of interaural coherence at $\pm 45^\circ$ seems to be primarily a low-frequency pinna effect. The effect on the ITDs within frequency bands shown in Figure 4.15 is then due to the pinna nearer the sound source, and the opposite pinna has little or no effect on the coherence. However, no explanation has yet been given for why angles around $\pm 45^\circ$ are so special in this respect, and why a similar effect does not occur at 135° . In the geometry of the pinnae, the angle of

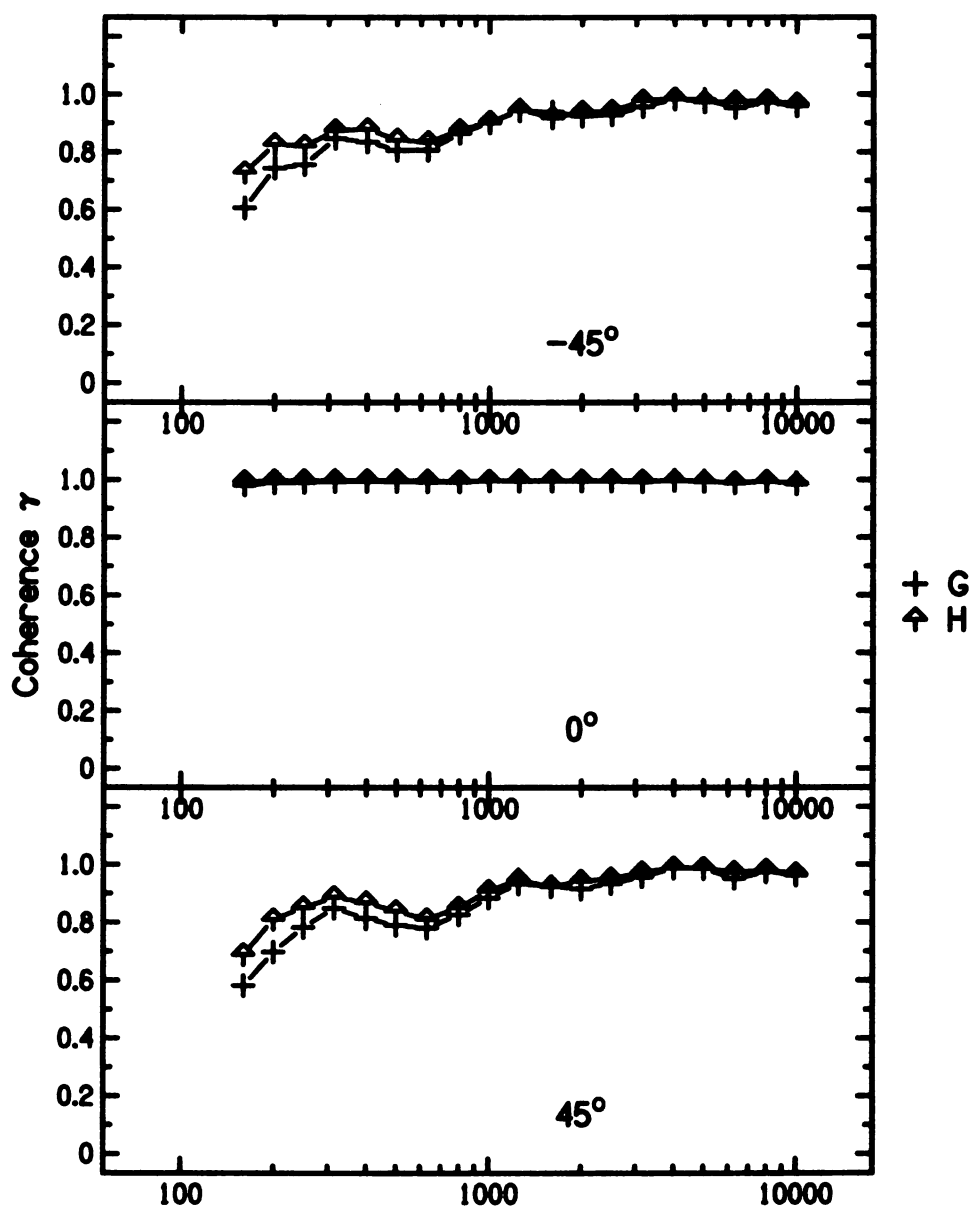


Figure 4.27. Coherences measured in 1/3-octave bands in an anechoic environment normally through KEMAR with pinnae in place (variation G, + symbols), and through a KEMAR head, detached from the KEMAR torso (variation H, up-arrow symbols).

45° does not seem to have any special place. The angle between the pinnae and the side of the head, for instance, is on average 24.1° for the average adult, and 22° for KEMAR [24].

The main effect of the pinnae is traditionally thought to be to aid in sound localization in the median sagittal plane. This is accomplished by means of the comb-filtering that occurs when reflection of sounds off the pinna interfere with the direct sound incident on the pinna [7]. These effects are thought to be particularly important above about 6 kHz (corresponding to a wavelength of about 5.7 cm), and this makes sense given that the length of the KEMAR ear is reported to be 5.89 cm by Burkhard and Sachs [24]. At a low frequency of, say, 200 Hz, the wavelength of the sound (1.72 m) is 29 times the length of the pinna, and so scattering by the pinna at low frequencies must be very small. At even lower frequencies, any scattering by the pinnae must be even smaller. An effect is observed, however, which is greater at lower frequencies and yet is attributed to the pinnae. It seems implausible then that simple acoustical scattering by the pinnae is responsible for the degraded coherences observed at certain angles of incidence.

4.2.2 Localizability of Noise Bands at 45° by Human Listeners

Poor interaural coherence potentially leads to poor localization of sounds via ITD cues. However, coherence is generally more important in the ITD localization of higher frequency noises (where ILD cues are not present), and becomes less important for lower frequencies. Constan [29] reports measurements of the threshold coherences necessary to correctly localize shifts in ITD of high-pass filtered noises. He observes that higher frequency cutoffs require greater coherences (nearer to 1) in order for a particular magnitude of ITD to be detectable. The drops in coherence measured previously in this experiment for KEMAR recordings at a 45° angle of incidence are most dramatic at low frequencies, and therefore may not be as im-

Angle	γ	ITD (ms)
-45°	0.6865	-0.2816
45°	0.7303	0.2400

Table 4.1. Coherence and ITD of 500-Hz noise bands recorded through KEMAR at incident angles of $\pm 45^\circ$ in an anechoic environment.

portant a factor in human localization of sounds in those frequency bands. The depth of the dips in coherence suggest, however, that even low-frequency localization tasks may become difficult for sounds with those coherences. This experiment measured the effect of noises with coherences as measured through KEMAR on the localizability of sounds by human listeners.

Methods

To determine the significance of the drops in coherence at 45° , the KEMAR recordings made at $\pm 45^\circ$ were used in a discrimination task designed to determine the ability of real listeners to localize coherent and incoherent noise bands. Two sets of recordings were made through KEMAR at -45° and $+45^\circ$. These noises were filtered into a 1/3-octave band centered at 500 Hz. The coherences of the noises in this frequency band as well as the measured ITDs are shown in Table 4.1.

An “incoherent” noise presentation was made to listeners by presenting them with the signals exactly as measured through the KEMAR ears for $\pm 45^\circ$ incident angles. This was referred to as a “standard *incoherent* noise pair.” A coherent noise presentation was made by presenting the same signal (that recorded in either the left or the right channel) to both ears, and thus the signals in this presentation had an interaural coherence of exactly 1. In the case of coherent noise presentations, the

signals were circularly shifted so as to match the ITD, and scaled so as to match the ILD of the incoherent presentations at the same angle. This was then referred to as a “standard *coherent* noise pair.”

The standard coherent and standard incoherent noises were perceptually shifted in space by altering the ITD between the left and right ear signals. Since ITD is thought to be the dominant cue for binaural sound localization in the region of 500 Hz, only the ITD was altered. Also, the relative magnitude of ILDs was small in the 500 Hz band, so a small shift in incident angle, which corresponds to a significant change in ITD, corresponds as well to a very small change in ILD. Thus, it was unnecessary to alter the ILDs.

The standard noises were subjected to a set of delays (in number of samples) of $\{0, \pm 3, \pm 6, \pm 9, \pm 12, \pm 15\}$. These noises were recorded at and were to be played back at a sample rate of 195,312.5 Hz, so that every three samples of shift represents a change in ITD of $15.4 \mu\text{s}$ and a corresponding change in perceived angle of, coincidentally, approximately 3° . Noises were shifted such that negative delays corresponded to a perceptual shift to the left and positive delays corresponded to a shift to the right. In this way, a set of shifted noises was created.

Listeners were presented with a task in which two noises were presented, one after the other, with a 500 ms pause between them. The first noise of the stimulus was a standard noise and the second was a shifted noise. Listeners were asked to indicate whether they heard the second noise to the left or to the right of the first. The full delay set was presented four times in random order, giving 44 noise pairs. This constituted a single run. In any given run, the noises presented to the listener were either all coherent or all incoherent.

All noise pairs in a single run corresponded to only one angle, and were all coherent noises or incoherent noises. Thus, there were four main conditions: 45° coherent, 45° incoherent, -45° coherent, and -45° incoherent. Listeners were given

a training period of several runs of both coherent and incoherent noises at each angle until their performance at both tasks no longer improved. This process of learning how to do the task was especially important in the case of incoherent noises. Listeners were then put through three runs of each condition for which their discrimination responses were recorded.

Four listeners, all of them males between the ages of 21 and 27, participated as subjects. All four had normal hearing, and were inexperienced in listening experiments of this type. Stimuli were presented at a level of 70 dB SPL through Etymotic ER-7 in-ear headphones. The headphones were placed in the ear canals of the listeners, so that the only pinna effects present were those of the KEMAR.

The percentage of times a listener responded that they heard the second noise to the right of the first was calculated for each value of the delay. A response rate at or above 75% was taken to be the threshold for correct discrimination of positive delays, and a response rate at or below 25% was taken to be the threshold for negative delays. By finding the smallest delay for which threshold discrimination was reached, a just-noticeable difference (JND) in shift was measured for each listener in each condition. Larger JNDs indicate greater difficulty in localizing sounds.

Results

The results for all four listeners are shown in Figures 4.28 and 4.29. The JNDs for shifts to the left are the delays at which curves crossed the 75% line, and the JND for shifts to the right are the delays at which curves cross the 25% line. Open circles represent incoherent noise, and filled-in circles represent coherent noise.

Listeners unanimously described the coherent noise bands as compact and easy to localize, whereas the incoherent noises were quite diffuse and difficult to localize. Listeners were able to perform well in the coherent localization task almost immediately, but required several runs with the incoherent noise before they learned

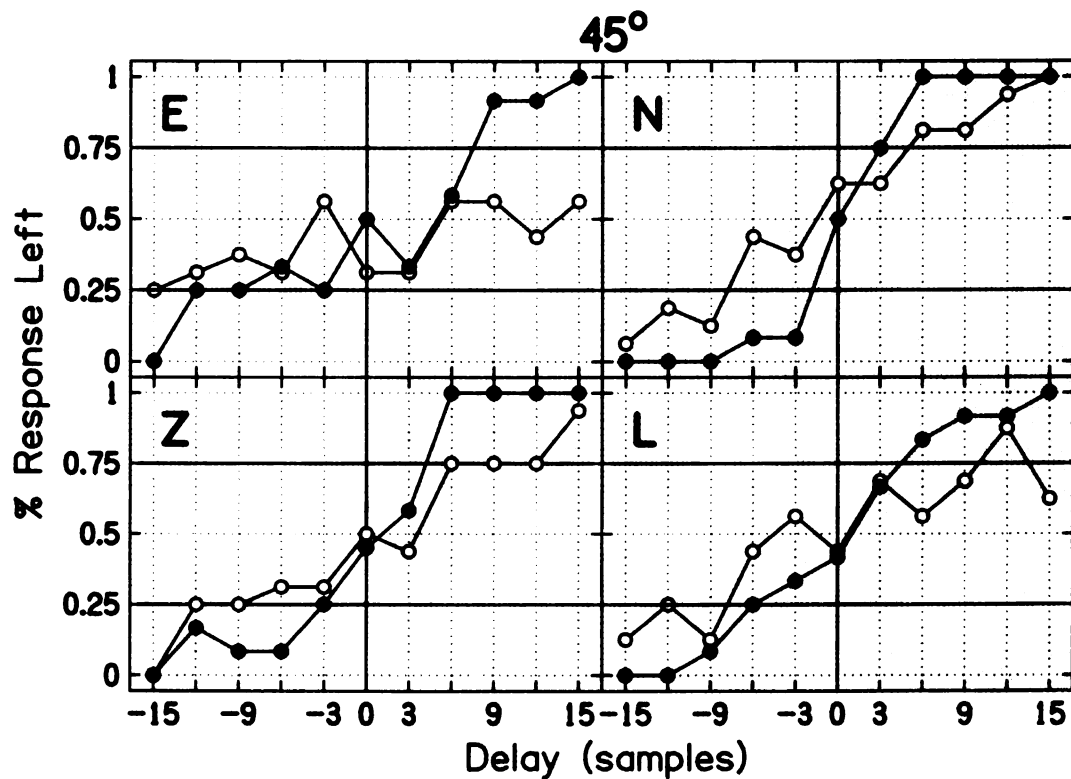


Figure 4.28. Percentage of "right" responses as a function of delay for noises at an angle of 45° for four different listeners. Responses corresponding to coherent noises are plotted with filled circles. Responses corresponding to incoherent noises are plotted with open circles. Thresholds for discrimination are plotted as solid horizontal grid lines. A shift of one sample corresponds to a perceptual shift of approximately one degree.

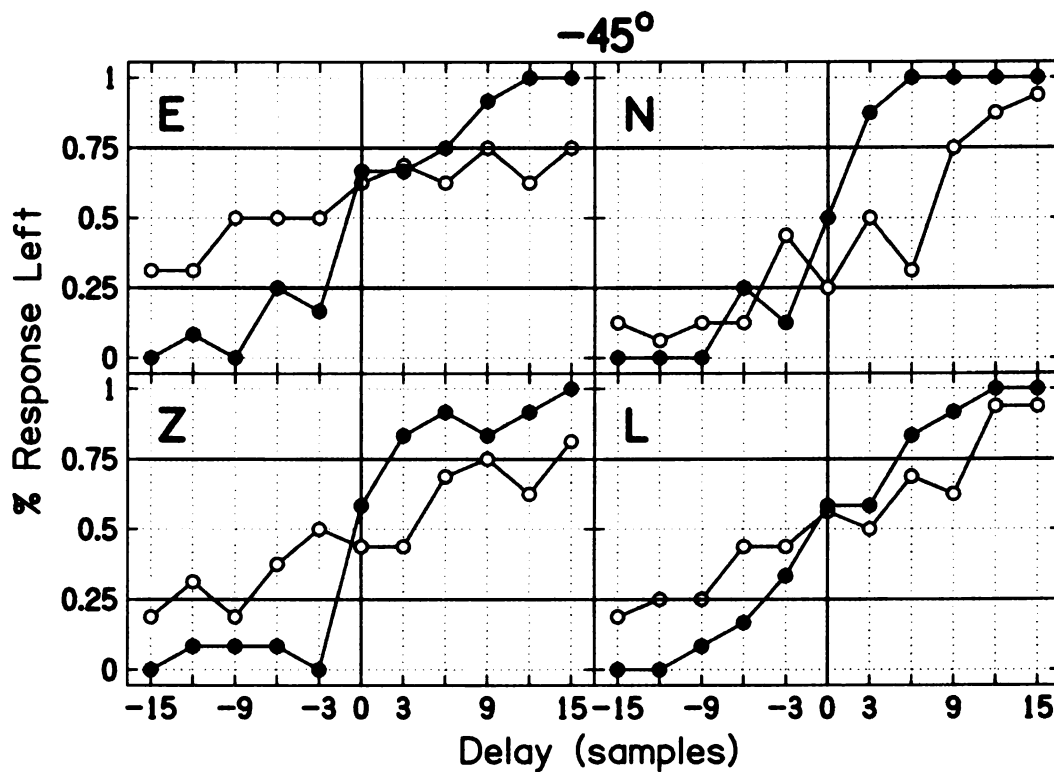


Figure 4.29. Percentage of "right" responses as a function of delay for noises at an angle of -45° for four different listeners. Responses corresponding to coherent noises are plotted with filled circles. Responses corresponding to incoherent noises are plotted with open circles. Thresholds for discrimination are plotted as solid horizontal grid lines. A shift of one sample corresponds to a perceptual shift of approximately one degree.

to localize the incoherent noises at all. Following training, listeners described various methods that they had developed in the learning process to help them localize the incoherent noises, such as listening to the “center” of the noises or focusing on only the highest frequency parts of the noise. No listeners reported that such special techniques were necessary for the localization of the coherent noise.

The JNDs for each listener in each condition are summarized in Table 4.2. The left side of the table contains the JNDs for sounds moving to the left, and the right side of the table shows JNDs for sounds moving to the right. In all cases, the JND for coherent noises is less than that for incoherent noises. In two cases, for incoherent noises moving to the left at 45° and moving to the right at -45° , listener E did not reach the threshold of discrimination within a shift of 15° . The average difference between the magnitude of the JNDs of incoherent and coherent noises in the same conditions across listeners was 5° , or an ITD of approximately $26 \mu\text{s}$. The JNDs for incoherent noises in each condition were significantly greater in magnitude than their coherent counterparts (in a paired, one-sided t -test, $t = -6.88$, $df = 13$, $p < 0.001$).

4.2.3 Measurements in KEMAR and Human Listeners with Probe Microphones

A different method of measuring the interaural coherence in KEMAR, which could be repeated in real listeners, would be to make recordings of signals through in-ear probe microphones. These microphones receive sound at the ends of long plastic tubes which can be snaked into the ear canal of the subject. Using such a technique on KEMAR would serve to confirm the drops in coherence seen in Experiment 9 for angles near $\pm 45^\circ$. More importantly, using such a technique on real listeners would determine whether or not this phenomenon was specific to KEMAR, or was actually present in human listeners.

Listener	Left JND				Right JND			
	Coherent		Incoherent		Coherent		Incoherent	
	45°	-45°	45°	-45°	45°	-45°	45°	-45°
E	7°	6°	> 15°	9°	-3°	-3°	-15°	< -15°
N	3°	2°	5°	9°	-2°	-2°	-8°	-5°
Z	4°	2°	6°	9°	-3°	-2°	-9°	-8°
L	5°	5°	10°	10°	-6°	-5°	-8°	-9°
Avg.	4.8°	3.8°	> 9.0°	9.3°	-3.5°	-3.0°	-10°	< -9.3°

Table 4.2. Just-noticeable differences (JNDs) for coherent and incoherent noises moving to the left or to the right. JNDs for sounds moving to the left are on the left side of the table and JNDs for sounds moving to the right are on the right side of the table. One degree on the table corresponds to an ITD of $5.1 \mu\text{s}$. Incoherent sounds at 45° had an interaural coherence of 0.7303 and those at -45° had a coherence of 0.6865. These JNDs tend to be significantly larger for the incoherent noises than for coherent noises, indicating greater difficulty of listeners to localize incoherent noises under otherwise identical conditions.

Methods

Three listeners from the experiment on the localization of noise bands recorded through KEMAR participated in this experiment (E, N, and Z). Listeners were seated on a rotating stool in the anechoic room used in previous experiments 3.05 m from the loudspeaker. Small paper markers on the walls of the room at approximately eye level indicated where the listeners should fixate in order to face in the -45° and +45° directions, respectively.

Listeners were fitted with a velcro headband. Attached to the headband near the ears were the connectors for ER-7C probe microphones. A thin plastic tube was snaked from the connectors into the ear canal of each listener. Sound recorded at the ends of the tubes were fed to the external preamplifier. No special care was taken to ensure that the position of the microphones was identical in each listener, except to make certain that the end of the tubes was well inside the ear canal. Though this may be expected to lead to variance across listeners, the point of this experiment was not to make identical measurements in different listeners, but to

look for a large effect on coherence in human listeners at $\pm 45^\circ$ when compared to 0° , and so the present setup and methodology was sufficient.

Measurements were made through the probe microphones for listeners facing each of three positions: -45° , 0° , and $+45^\circ$. At each position, listeners were exposed to three different MLS noises of order 18 at a level of 85 dB SPL. Recordings were made at a sample rate of 195,312.5 Hz. These conditions were identical to those used to test KEMAR in previous experiments in the anechoic environment.

Listeners were instructed to sit upright, fixate directly on the paper marker on the wall, and to not move in between noise bursts. Once three noises were measured at one position, listeners were instructed to gently rotate themselves on the stool until they were facing the next position, and another three noises were presented and recorded. Listeners were instructed to begin by looking at the -45° marker, then directly at the center of the loudspeaker (0°), then at the $+45^\circ$ marker. No special measures were taken to ensure that the orientation of the listeners' heads was exact beyond what has been described already. However, the drop in coherence observed in KEMAR was reasonably broad, and so it should not be necessary for a listener to be oriented such that their interaural axis is exactly 45° to the incident sound. An identical set of measurements was made using the probe microphones in the KEMAR.

Results

The interaural coherences were measured in 1/3-octave bands between 200 and 10,000 Hz. Measurements in frequency bands lower than 200 Hz were unreliable because of the poor response of the probe microphones at those frequencies. The results of these measurements for each listener are shown in Figure 4.30. The average coherence measured for each listener in each band is plotted, with error bars spanning one standard deviation in each direction.

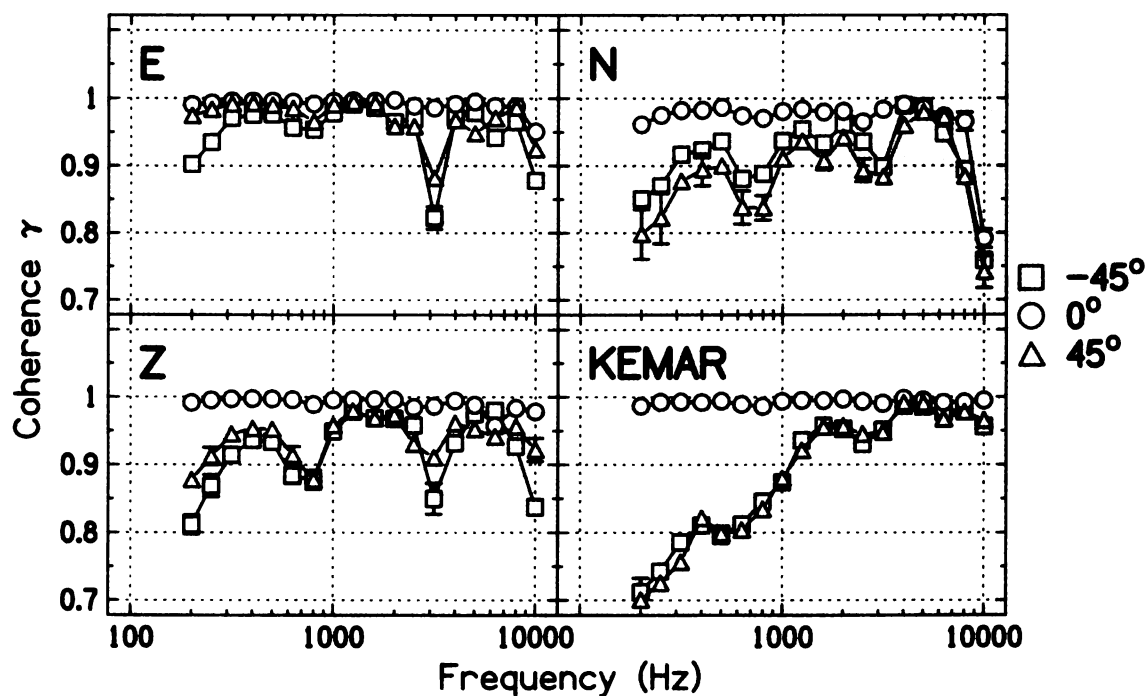


Figure 4.30. Coherences measured in three human listeners and a KEMAR at incident angles of sound of $\pm 45^\circ$ and 0° . Averages and standard deviations are calculated over three measurements made at each angle. Error bars span one standard deviation in each direction. Error bars smaller than the size of the data points are invisible.

The results show that coherences as measured in human listeners are also somewhat degraded at angles of $\pm 45^\circ$ compared to coherences at 0° , but not as much as in KEMAR. There is significant variation in the coherences measured at $\pm 45^\circ$ across listeners, but this is not surprising given that the probe microphone tubes were not placed in any precise way in the ear canals of the listeners. To reiterate, the precise numbers are not as important as the observation of the overall effect in this study.

Some greater care was taken with probe microphone measurements in KEMAR than was taken in human listeners. In the case of KEMAR, the probe microphones were inserted as far in as possible, until they nearly touched the transducer “ear canals.” In this way, a measurement as similar as possible to those measured through KEMAR itself was made. A plot of the coherence across 1/3-octave bands at 45° measured both through KEMAR and with probe microphones inserted into KEMAR is shown in Figure 4.31. The probe microphone measurements yield almost identical results to those made through the KEMAR electronics, and the two have a Pearson product-moment correlation coefficient of $r = 0.986$. Measurements in KEMAR with probe microphones thus confirm measurements made through the KEMAR ears.

The degradation of coherence around $\pm 45^\circ$ in human listeners is evident in Figure 4.30, but to far less an extent than in measurements made in KEMAR. Comparing coherences for the bands from 200 to 1000 Hz, where the differences between KEMAR and human listeners is most evident, the coherences measured in KEMAR were significantly less than those measured in any of the listeners (in several one-sided paired t -tests between listener and KEMAR data, $t < -7.53$, $df = 7$, $p < 0.001$). In the 500 Hz band at $+45^\circ$, for instance, the average coherence across all three trials, plus and minus one standard deviation, measured in KEMAR was 0.799 ± 0.077 . In the same band, the average coherence across all three human lis-

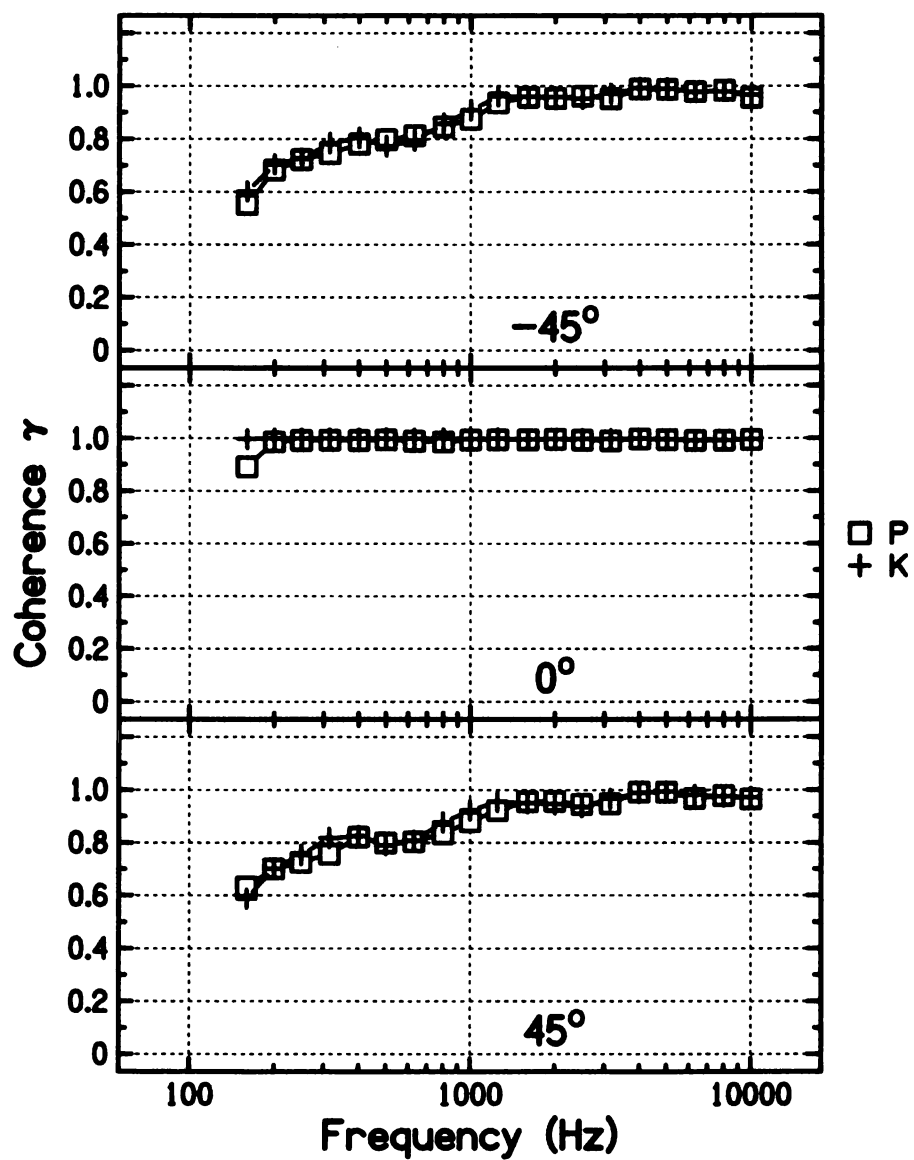


Figure 4.31. Coherences measured at 45° from recordings made through KEMAR Zwislocki couplers and electronics (+ symbols) and through ER-7 probe microphones (open squares).

teners was 0.94 ± 0.056 . Though the drop in coherence measured here for human listeners is much smaller than that measured in KEMAR, such coherence drops should be nonetheless detectable [29]. This may still imply that there is an effect on the ability of listeners to localize sounds incident at or around $\pm 45^\circ$.

4.2.4 Localization of noises with coherence as measured in human listeners

The sounds presented to listeners in previous parts of this experiment were noises filtered into 1/3-octave bands. Previous experiments regarding the detectability of incoherence indicate that human listeners can be quite sensitive to even a modest amount of incoherence. At a center frequency of 500 Hz and a bandwidth of 100 Hz, listeners can discriminate perfectly coherent noises from noises with a coherence as high as, on average, 0.99 [29]. The 1/3-octave bandwidth about 500 Hz as used in this experiment is approximately 115 Hz. Gabriel and Colburn [42] note that discrimination of interaural incoherence compared to a reference condition of coherence 1 is best at or below bandwidths of 115 Hz.

Human listeners were shown to experience a far less dramatic drop in coherence for incident angles near $\pm 45^\circ$ than those measured in KEMAR. However, the average coherence measured in listeners at $\pm 45^\circ$ was 0.94, easily within the threshold of incoherence detection at the 500 Hz band [29, 42]. Previously, listeners were presented signals as recorded through KEMAR filtered into a 1/3-octave band centered at 500 Hz, which had an average interaural coherence of 0.71. It is likely that noises with coherences equal to those actually measured in listeners are far easier to localize than those previously presented to listeners.

In this experiment, noises were constructed in a 500 Hz band identical to that used in previous experiments but with an interaural coherence equal to that measured in human listeners. These noises were presented to listeners in a manner

Listener	Coherence γ		
	-45°	0°	$+45^\circ$
E	0.9763	0.9960	0.9899
N	0.9363	0.9868	0.9000
Z	0.9337	0.9972	0.9524
KEMAR	0.7971	0.9942	0.7988

Table 4.3. Interaural coherences in the 500-Hz band from signals measured with probe microphones in KEMAR and human listeners E, N, and Z.

identical to that of the previous JND localization experiment. In this way, the ability of listeners to localize noises with coherences as measured in their own ears was explored and compared to previous results for the localization of perfectly coherent noises.

Methods

A localizability experiment was performed in exactly the same manner as the previous localizability experiment for human listeners. Stimuli were made from MLS signals, altered to have the same phase and amplitude spectra as those signals measured in each listener's ears using probe microphones. The coherences of the noises presented to each listener were then exactly the same as what was previously measured. These coherences in the 500 Hz band are summarized for each listener and for KEMAR in Table 4.3.

Results

The results for all four listeners are shown in Figures 4.32 and 4.33. The JNDs for shifts to the left are the delays at which curves cross the 75% line, and the JND for shifts to the right are the delays at which curves cross the 25% line. Open circles represent incoherent noise, and filled-in circles represent coherent noise.

All listeners described this task as being quite easy compared to the task where

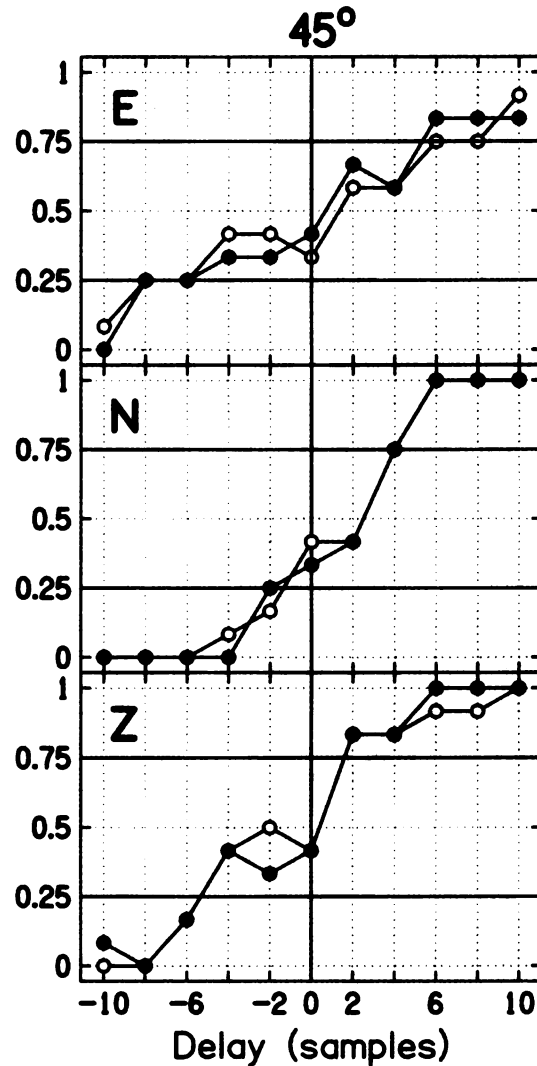


Figure 4.32. Percentage of "right" responses as a function of delay for noises at an angle of 45° for four different listeners. Noises presented to each listener had interaural coherences equal to that measured in that particular listener's ears in the 500 Hz 1/3-octave band. Responses corresponding to coherent noises are plotted with filled circles. Responses corresponding to incoherent noises are plotted with open circles. Thresholds for discrimination are plotted as solid horizontal grid lines. A shift of one sample corresponds to a perceptual shift of approximately one degree.

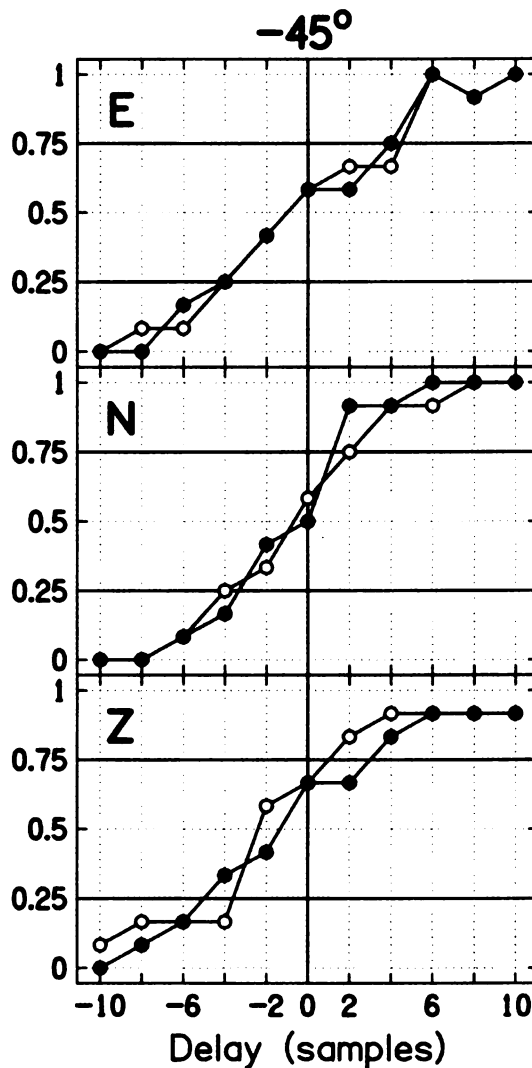


Figure 4.33. Percentage of “right” responses as a function of delay for noises at an angle of -45° for four different listeners. Noises presented to each listener had interaural coherences equal to that measured in that particular listener’s ears in the 500 Hz 1/3-octave band. Responses corresponding to coherent noises are plotted with filled circles. Responses corresponding to incoherent noises are plotted with open circles. Thresholds for discrimination are plotted as solid horizontal grid lines. A shift of one sample corresponds to a perceptual shift of approximately one degree.

Listener	Left JND				Right JND			
	Coherent		Incoherent		Coherent		Incoherent	
	45°	-45°	45°	-45°	45°	-45°	45°	-45°
E	5°	4°	6°	4°	-6°	-4°	-6°	-4°
N	4°	2°	4°	2°	-2°	-4°	-2°	-4°
Z	2°	3°	2°	1°	-5°	-5°	-5°	-4°
Avg	3.7°	3.7°	4.0°	2.3°	-4.3°	-4.3°	-4.3°	-4.0°

Table 4.4. Just-noticeable differences (JNDs) for coherent and incoherent noises moving to the left or to the right. JNDs for sounds moving to the left are on the left side of the table and JNDs for sounds moving to the right are on the right side of the table. Incoherent sounds at 45° and -45° had coherences for each listener identical to that measured in that listener with probe microphones. The specific coherences for each listener can be found in Table 4.3. There is no significant difference in JNDs for the coherent and incoherent noises.

the noises had coherences as measured in KEMAR. The JNDs for each listener in each condition are summarized in Table 4.4. The left side of the table contains the JNDs for sounds moving to the left, and the right side of the table shows JNDs for sounds moving to the right. There was no significant difference between JNDs for coherent and incoherent noises in this case (in a two-sided, paired t -test, $t = 0.00$, $df = 11$, $p = 1.000$).

That the JNDs for each listener are not different when listeners are presented with coherent noise versus when the noise has characteristics as measured in the listeners' ears may indicate a familiarity of each listener with their own acoustics. However, it is just as likely that the coherences measured in listeners' ears were so high that, for a 500 Hz center frequency, noises were just as easily localizable as if the coherences had been equal to 1, even though listeners reported being able to detect the presence of incoherence. In any case, though the effect on interaural coherence with sound source location seen in KEMAR is present in human listeners, the magnitude of the effect is much smaller and likely does not adversely affect horizontal localization.

4.3 Experiment 10: Measurements of Binaural Parameters and Room Characteristics in a Highly Reverberant Environment

Measurements made on KEMAR in anechoic conditions form a baseline for making further measurements in a real room. The patterns and data collected in anechoic conditions reflect the effects of the geometry of the KEMAR head, binaural asymmetries, and other such parameters relating to the experimental equipment. Thus they provide a basis for measurements made in a reverberant environment such that the effects of the environment and the effects of the equipment might be separated.

In the previous experiment, binaural parameters were measured for a KEMAR in an anechoic environment. While such measurements provide a useful set of baseline measurements, most environments in which listening takes place are not nearly anechoic. Listeners experience the effect of their auditory environment, such as reflections off of walls, and cope with these effects to varying degrees of success. For instance, perception of reflections of impulsive sounds within rooms is largely suppressed by the auditory system. Listeners learn to ignore these reflections and do not perceive them as separate auditory images (echoes) unless the delay in arrival time between the direct sound and a reflection is longer than about 10 ms for clicks, and 50 ms for speech [46]. The current experiment examines the influence of rooms on binaural parameters and how this relates to the reverberant properties of the room.

4.3.1 Methods

A KEMAR was placed in a reverberant chamber with dimensions 7.67 m by 6.35 m by 3.58 m high (the same reverberation room as used by Hartmann, *et al.* [51]). A MLS signal was played using the same equipment as in previous experiments, with the loudspeaker situated 3.05 m from the KEMAR, at approximately ear level. Measurements were taken through the KEMAR of five different MLS stimuli, all of order 19. Such long signals were necessary since it was expected that the room would, in some bands, have reverberation times on the order of two to three seconds. However, this also greatly increased the processing time required and so the sampling rate was reduced to 97,656.25 Hz. This resulted in signals 5.37 s long. This sampling rate was sufficient for measuring frequency bands through 10 kHz. The Schroeder frequency of this room (Equation 2.8) was expected to be between 200 Hz and 300 Hz.

To eliminate the influence of standing waves in the room, measurements were taken at six different positions in the reverberant room. Each position placed the loudspeaker and KEMAR in a different place in the room, but both were always separated by 3.05 m with the loudspeaker and KEMAR “facing” each other. Five measurements, each with a different MLS of order 19, were taken at each position. In this way, measurements of the waveform coherence γ , ITD, and ILD were taken, as well as measurements of RT_{60} in twenty-two 1/3-octave bands from 80 Hz to 10000 Hz.

A unique IR may be calculated from the signal in each ear of the KEMAR for each MLS, and a RT_{60} in each 1/3-octave band may be calculated for each IR. Thus, two estimates of RT_{60} were made in each 1/3-octave band for each MLS played. At each of the six positions tested there were then 10 measurements of RT_{60} per center frequency, leading to a total of 60 measurements of RT_{60} in each 1/3-octave band. Averages and standard deviations for RT_{60} in each band were calculated

across these 60 measurements.

4.3.2 Coherence Results

Coherences in each 1/3-octave band were averaged across all six positions in the reverberant room for which data were taken. The average coherences in each band are shown in Figure 4.34. Error bars extend to plus and minus one standard deviation. As done for measurements in the anechoic environment, coherences were corrected for period errors (peaks in coherence were expected to occur for lags near 0 ms). The differences in coherence between the correct peak and the coherence in the case of a period error were, on average (plus and minus one standard deviation), -0.009 ± 0.0097 . Because the differences between the heights of the accepted peaks and the period error peaks of the cross correlation functions are so small, the expected peak is selected rather than the period error when measuring ITD and coherence. It is reasonable to expect that the auditory system is able to perform a similar operation by comparing the ITD in each critical band to that measured in lower critical bands to ensure it has found the “correct” peak.

The shape of this graph is quite similar to that of measurements made by Hartmann, *et al.* [51] in the same room. For frequencies up to about 5000 Hz, the pattern of coherences is in fact extremely similar. At frequencies above 5000 Hz, measurements made here and those made by Hartmann, *et al.* are quite different, with coherences measured here generally higher on average by 0.184. It may be argued that measurements made in the present study are more accurate at higher frequencies because the sampling rate used in the present study (about 100 kHz) was twice that used by Hartmann, *et al.* (50 kHz). Higher sampling rates yield better temporal resolution in the cross-correlation function, and thus more accurate measurements of coherence.

The variances in the data measured by Hartmann, *et al.* were significantly

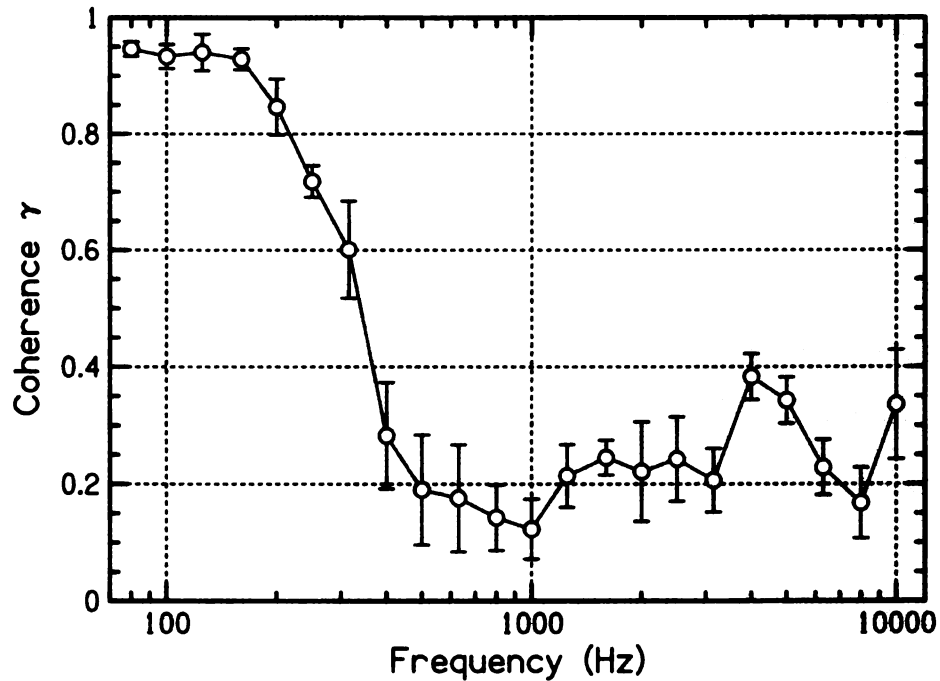


Figure 4.34. Average coherences across all measurements in 1/3-octave bands made in a reverberant room at a distance of 3.05 m from the sound source. Error bars extend one standard deviation in each direction.

greater than the variances measured here (in two-sample, one-sided, t -test, $t = 1.57$, $df = 36$, $p = 0.063$). Greater variances in the earlier study were likely due to the types of noises used. Hartmann, *et al.* made use of a sequence of equal-amplitude random phase noises, each filtered into one of several 1/3-octave bands until 20 noises were created with center frequencies ranging from 142 Hz to 9000 Hz. The noises were played in five-second bursts, one after the other, and separate recordings were made for each noise band. In the present study, a single noise with a flat frequency amplitude spectrum (i.e. a MLS) was used and filtered *a posteriori* into 1/3-octave bands. This simultaneous measurement of all frequency bands at once, as well as the similarity in the properties of different MLS of the same order, likely led to the smaller variance within frequency bands seen in the present study.

Unlike measurements made in the anechoic environment (Figures 4.11

and 4.12), there is noticeable variation in the measurements of coherence across trials in the reverberant space. The largest standard deviations tend to occur near a center frequency of 500 Hz, where the standard deviation in coherence is ± 0.094 . Coherences tended to be larger and more consistent for bands centered at 250 Hz and below than for those above 250 Hz. Standard deviations for coherences in these low frequency bands were significantly smaller than standard deviations in higher frequency bands (one-sided two-sample t-test, $t = -5.16$, $df = 16$, $p < 0.001$). The reasons for this will be further elaborated in the following section.

The cause for very low coherences in certain bands is likely variance in the ITD spectrum within bands, much as it was for recordings made through KEMAR ears in anechoic conditions for incident sound source angles near $\pm 45^\circ$. To see this, consider the ITD spectrum for one particular trial in which the waveform coherence in the 1000-Hz band was found to be only $\gamma = 0.2453$. This is plotted in Figure 4.35. In that band, the mean ITD across individual components within the band was found to be 12.7 ms, very close to the expected value of 0, and the standard deviation was 251 ms. Plugging this standard deviation into Equation 4.31 gives a coherence of $\langle \gamma \rangle = 0.2883$, which is very close to the coherence found via the peak cross-correlation function. Thus, it seems that the variance in the ITD spectrum is responsible for poor coherence in rooms.

4.3.3 Reverberation Times

Reverberation times measured in the reverberant room are shown in Figure 4.36. Average reverberation times in each 1/3-octave band are calculated for all measurements of RT_{60} in each band, including those from each ear of the KEMAR. Error bars span one standard deviation in each direction. Reverberation times tended to increase with increasing frequency until peaking at 1.25 kHz with a re-

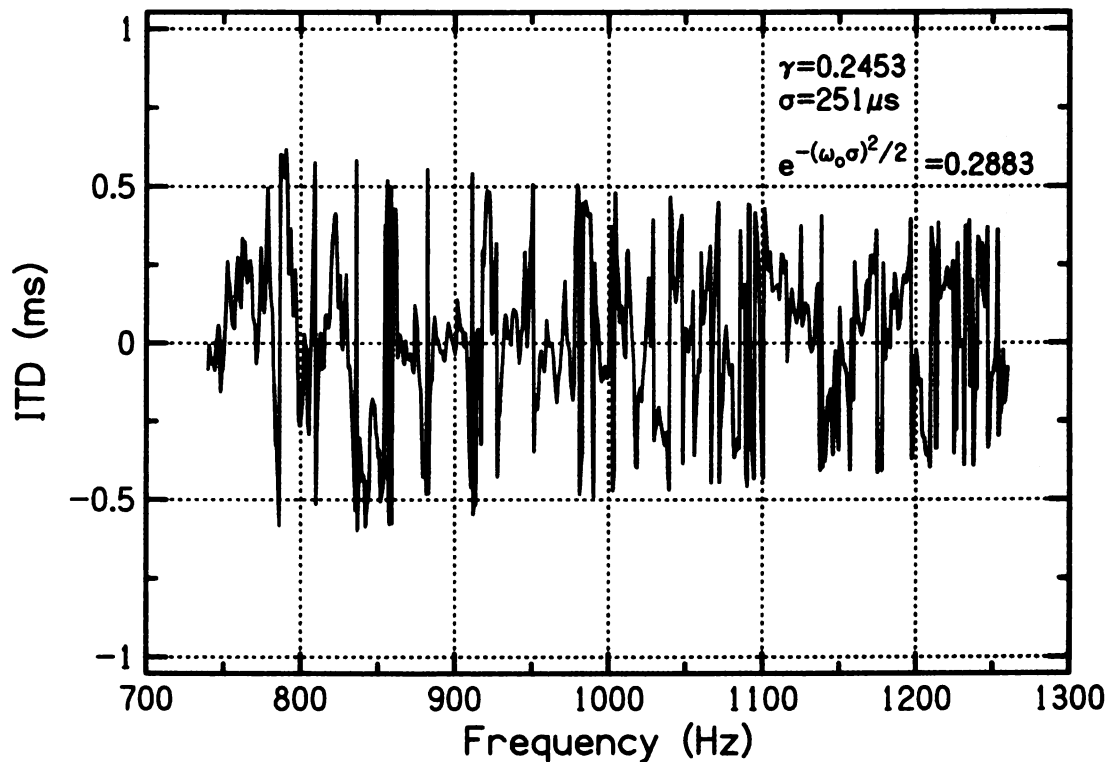


Figure 4.35. The ITD spectrum of the 1/3-octave band centered at 1000 Hz, recorded through KEMAR in the highly reverberant environment. The interaural waveform coherence within this band is $\gamma = 0.2453$. The standard deviation in ITD across the band is $\sigma = 251 \text{ ms}$, which gives a coherence of $\langle \gamma \rangle = 0.2883$ when Equation 4.31 is employed.

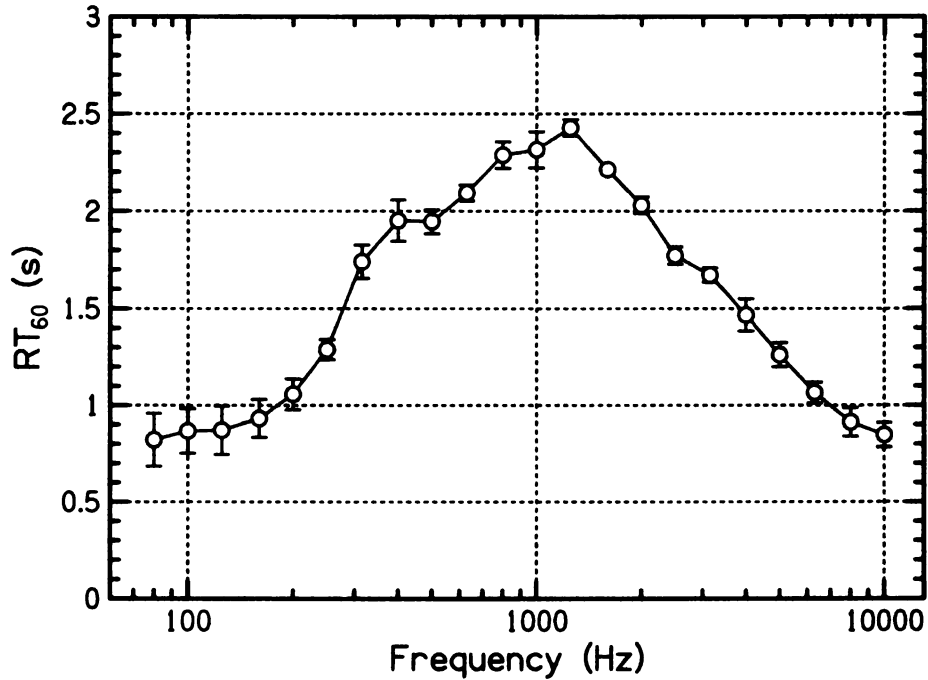


Figure 4.36. 60-dB reverberation times, averaged across all measurements from both ears of a KEMAR in 1/3-octave bands, made in a reverberant room. Error bars extend one standard deviation in each direction and indicate excellent agreement across different positions in the room and between the two ears.

reverberation time of 2.43 ± 0.042 s. For frequencies above 1.25 kHz, RT_{60} decreased monotonically. These reverberation times agree quite well with those reported for the same room by Hartman, *et al.* [51].

In general, measurements of RT_{60} were quite consistent across different positions in the room, as evidenced by the small size of the error bars. Standard deviations were significantly larger for center frequencies near or below the Schroeder frequency of the room (one-sided two-sample t-test, $t = 2.86$, $df = 7$, $p = 0.012$). This tendency in the size of standard deviations is exactly opposite that for coherence measurements.

Presumably, the prominence of individual standing waves in the room (thought to occur for frequencies below the Schroeder frequency of the room) leads to uncertainty in RT_{60} measurements across positions in the room, as different positions

in the room will occur at different points along those standing waves. At the same time, each ear of the KEMAR is equally affected by standing waves common to both ears. This leads to a larger coherence and a smaller variation in coherence across measurements at low frequencies.

It should also be noted that the difference in RT_{60} between ears was not statistically significant (one-way ANOVA, $F(1, 262) < 0.01$, $p = 0.951$). This gives some assurance that the reverberation times heard by each ear were nearly the same in each 1/3-octave band for each measurement. Thus, the practice of combining measurements of RT_{60} from both ears seems to be well justified.

4.3.4 ITD Results

The measured ITD, based on the temporal location of cross-correlation peaks corrected for period errors, is shown in Figure 4.37 in open circles. ITDs were measured in each 1/3-octave band five times at each of six positions for 30 total measurements in each band. Each measurement was individually corrected for period errors. Averages were computed across all 30 measurements in each band and standard deviations were computed across averages at each position. Error bars span one standard deviation in each direction.

The only average ITDs that did not fall within one standard deviation of 0 ms were those at 125 Hz and below, where the ITDs were less than 0 ms. This is similar to the pattern of ITDs at low frequencies found in an anechoic environment (Figure 4.23). This is an expected result given the negative ITDs seen in anechoic conditions for this range of frequencies (Figure 4.23).

Period errors were understandably frequent in the reverberant room. At each position, six or seven period errors were noticed in various 1/3-octave bands. The lowest frequency band for which period errors were found was at 1 kHz. Curiously, no period errors were found at any of the six positions tested or in any trial

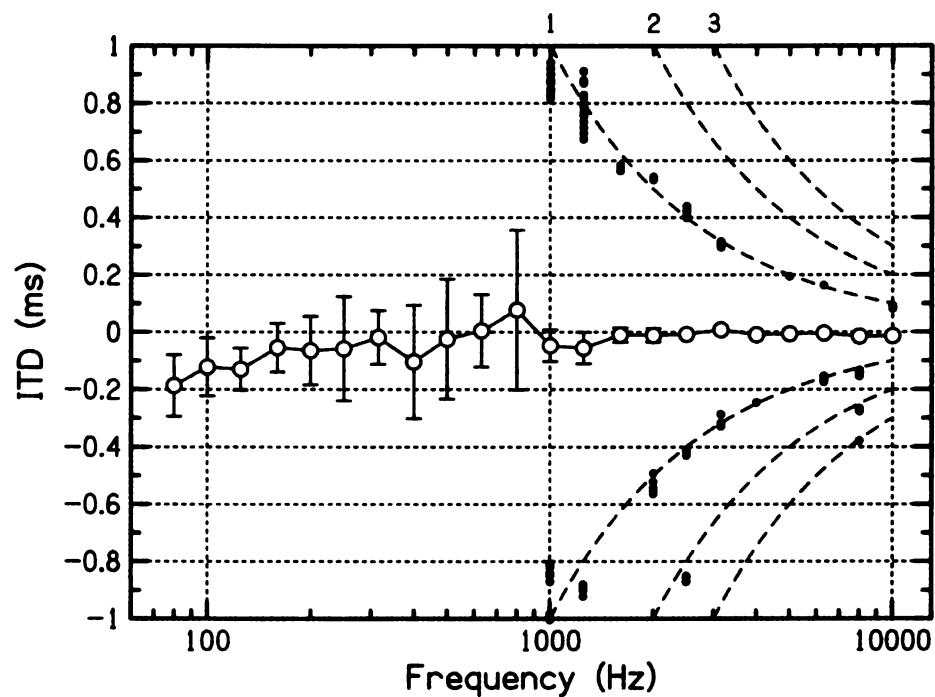


Figure 4.37. ITDs measured between two ears of a KEMAR in a reverberant room. Data points indicate average measured values, with error bars spanning one standard deviation in each direction. Averages are calculated across 30 measurements, made at six different positions in the reverberant room, and standard deviations are calculated across average values from each position. At each position, the KEMAR and sound source were separated by 3.05 m with the KEMAR facing the sound source. Small dots indicate the ITD of period errors that occurred in at least one trial. Dashed lines indicate the theoretical ITD for one, two and three period errors respectively as a function of frequency.

for the band with center frequency of 4 kHz, and there is also a local peak in the coherence at this center frequency (see Figure 4.34). Each occurrence of a period error is plotted in Figure 4.37 as a small dot. Dashed lines indicate a multiple of the theoretical period (in ms) as a function of center frequency, and thus trace curves along which period errors should occur. The observed period errors conform well to the dashed lines, indicating good agreement with prediction of where period errors should fall.

These measurements are similar to those reported by Hartmann, *et al.* [51]. In that paper, a similar figure is presented for ITD measurements taken in a room of similar volume, but shorter RT_{60} than the reverberation room presented here in connection with Figure 4.37. Both in that paper and in the observations made in the present study, wide variations in ITD were observed in frequency bands up to about 1.25 kHz. Above that frequency, ITD cues become consistent across all trials, and are nearly zero (the expected ITD in any band when the incident sound is at 0°). Though ITDs measured in most bands are within one standard deviation of 0 ms, the variances seem to become suddenly quite small once a single period error comes within the physiological range for ITDs of ± 1 ms. The variance in ITD at low frequencies may lead to a difficulty for listeners in localizing low-frequency sounds in reverberant environments, since ITD cues are thought to be quite important in low-frequency localization.

Frequency bands below 1 kHz are not subject to period errors because the calculation of the cross-correlation function that leads to a measurement of coherence and ITD is limited here to a maximum lag of ± 1 ms. Period errors at these frequencies would fall outside of this range, and thus the measured ITD and coherence in those bands occurs only at the peak in the cross-correlation function within the ± 1 ms range. It is interesting then to repeat the calculation of ITD using the same method and the same recorded signals, but with a maximum lag of ± 2 ms. In

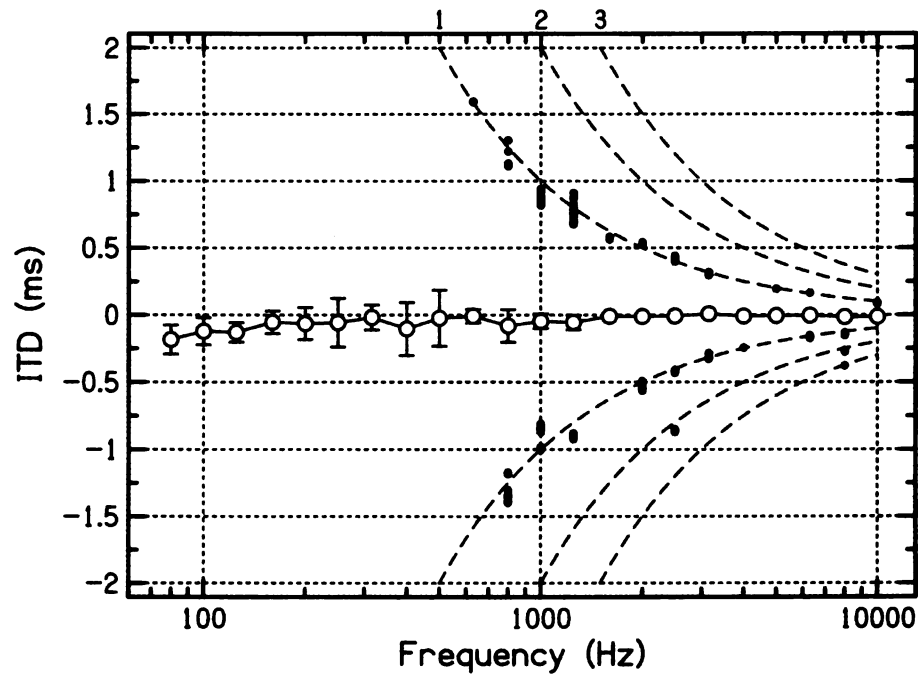


Figure 4.38. ITDs measured between two ears of a KEMAR in a reverberant room, presented in a manner identical to that of Figure 4.37. ITDs are calculated by finding the lag of the peak in a cross-correlation function between the signals recorded in the left and right ears, with a maximum lag of ± 2 ms.

some cases, the largest peaks may occur outside the ± 1 ms range. If this is so, then period errors may be accounted for in the 500–800 Hz bands as well. The result of performing this calculation with a maximum delay of ± 2 ms is shown in Figure 4.38. Period errors were found in both 630 and 800-Hz bands. The result of correcting these period errors is a lower variance in the ITD measurements in those bands.

It may also be that the peaks measured in low-frequency bands in rooms may be large and outside the maximum lag of the cross-correlation function such that the peaks captured here are actually the period errors. In other words, ITDs greater than 1 ms may occur naturally in rooms. Certainly the human auditory system is capable of localizing sounds with ITDs well outside of the “physiological limit” of ± 1 ms [79, 18], and such ITDs may occur in rooms where reflections occur.

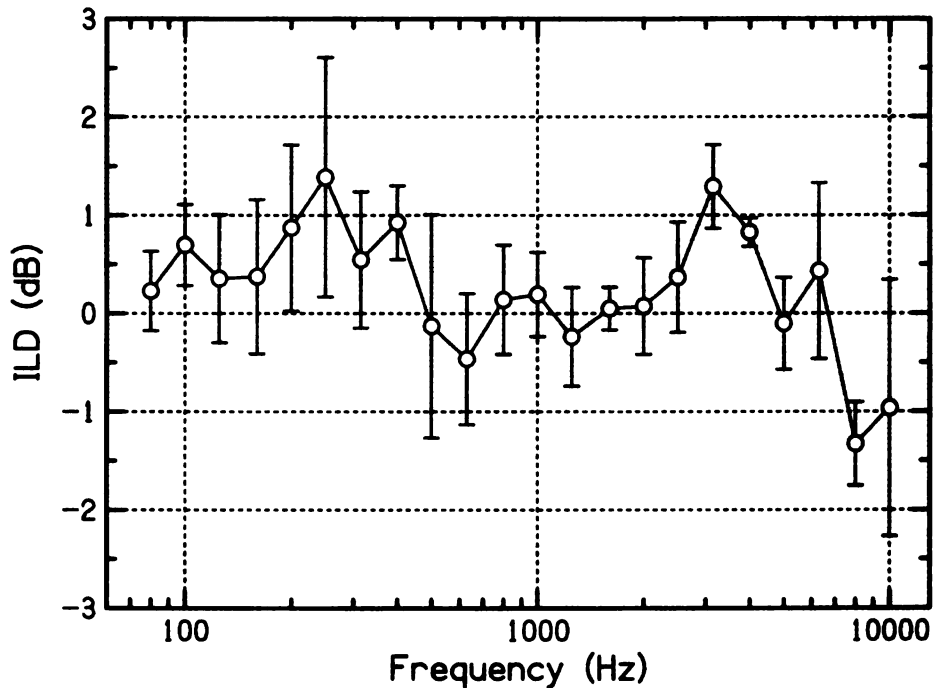


Figure 4.39. ILDs measured between two ears of a KEMAR in a reverberant room. Data points indicate average measured values, with error bars spanning one standard deviation in each direction. Averages are calculated across 30 measurements, made at five different positions in the reverberant room, and standard deviations are calculated across average values from each position. At each position, the KEMAR and sound source were separated by 3.05 m with the KEMAR facing the sound source.

4.3.5 ILD Results

The 1/3-octave ILDs measured in the reverberant room are shown in Figure 4.39. As with measurements of ITD in this room, averages are calculated across five measurements at each of six positions in the room, for a total of 30 measurements. Standard deviations were taken across averages found at each of six positions in the room. Error bars span one standard deviation in each direction.

Certain broad features of the ILD across center frequencies are similar to those seen in Figure 4.25 for ILDs measured in anechoic conditions. In low frequencies through about 400 Hz, there is an average positive ILD of about 0.5 dB, likely corresponding to an imbalance in gain between the left and right channels at the

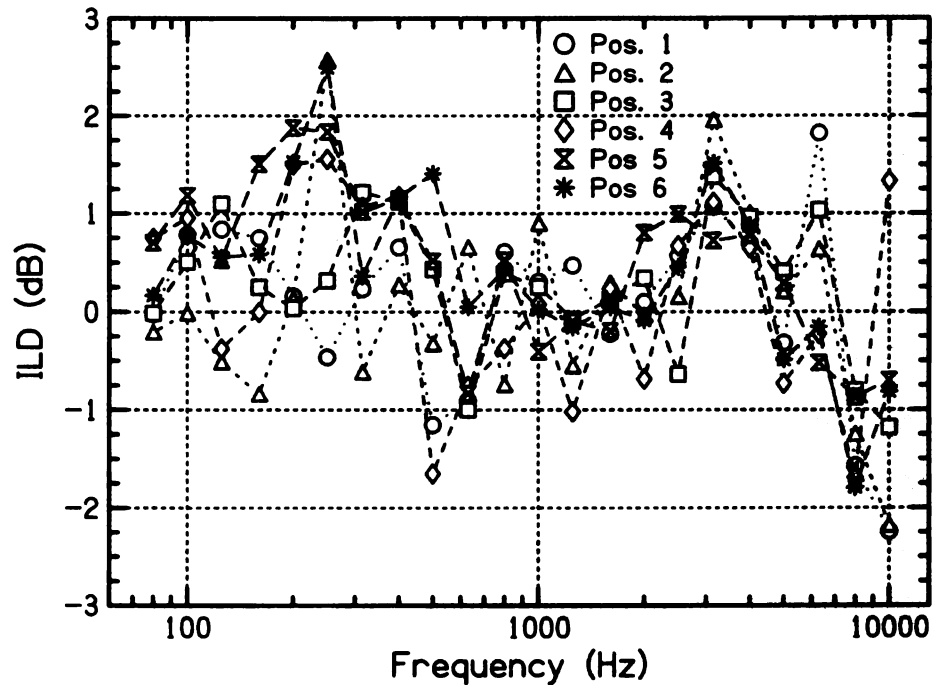


Figure 4.40. ILDs measured in a reverberant room at six different positions. The data points for each position are labeled with different symbols. Error bars extending one standard deviation in each direction are all smaller than the size of the data points, and so are not shown.

output of the preamplifier. The dip in ILD at and above 8 kHz is similar to the anechoic condition where the KEMAR was oriented at 0° relative to the direction of incident sound, and so this effect is likely internal in nature.

The variance in ILD is quite noticeable across different positions in the room, as shown in Figure 4.39. However, ILDs across all five trials (each trial with a different MLS of order 19) at each position were very consistent within each 1/3-octave band. Figure 4.40 shows the average ILDs measured at each position in each band. Averages and standard deviations were calculated across all five trials in each band for a given position. Error bars extend one standard deviation in each direction, but in almost all cases are smaller than the size of the data points, and are thus invisible. There is very little variance across trials at each position, but there is relatively little agreement across positions.

Despite the agreement in ILD across trials at a single position, the variation in ILDs across different positions in the room makes the ILD a poor binaural cue for localization in this condition. Ihlefeld and Shinn-Cunningham note that the mean ILD, calculated over several short-time measurements at a fixed location in a reverberant environment, is unreliable for determining sound source laterality, though some information on source location can be gathered by examining how the short-term ILDs vary over time [57]. Even if a listener were trained to associate certain ILD patterns with specific sound source localizations while standing at one position in the room however, those patterns of ILD would significantly change when the listener moves to another position in the room, invalidating the listeners previous associations. This is not surprising, since the sound energy is expected to be diffuse (independent of sound source location) in a reverberant space once it is some minimal distance from the listener (in the room used in the current study, this distance was found to be about 3 m). However, it seems that significant ITD cues are still available (Figure 4.37) if period errors can be ignored. At very high frequencies, as the room becomes more anechoic, ILD cues again become useful to those who are able to hear them [50].

4.4 Experiment 11: Measurements of Binaural Parameters and Room Characteristics in a Normal Room

Thus far, binaural parameters have been measured in both anechoic conditions (Experiment 9) and highly reverberant conditions (Experiment 10). It is then left to measure binaural parameters and room properties in a normal room, one with properties more like what is typically experienced by most people on a day-to-day

basis. It is expected that the effects of reverberation as seen in the highly reverberant environment will be seen to a far lesser extent in such a room. Similarly, a departure from the ideal conditions of the anechoic environment is expected to lead to binaural coherences less than those seen in that environment, though greater than those found in the highly reverberant environment.

A simple theory of reverberation in rooms is that a listener hears a sound composed of a “direct” and a “reverberant” component. Direct sound comes to a listener’s ears from the sound source via a straight line path between the two and diminishes in intensity as the inverse-square of the distance between the source and listener. This is a good description of listening in an anechoic environment, where the only sound that reaches the listener is that which comes directly from the sound source.

The reverberant component of the sound heard by a listener in a room comes from the sound reflected off the various reflecting surfaces (walls, floor, etc.). Since there are many paths for reverberant sound to take when traveling from the source to the listener, the timing and level of reverberant sounds will be chaotic, and will be different in each ear. The reverberant component of the sound interferes with the direct sound and reduces the binaural coherence γ of the overall sound. This is easily seen in the coherences measured in the reverberant environment (Figure 4.34), where the coherence is quite low for most frequencies.

A simple model of binaural coherence in rooms assumes that the direct sound contributes to coherence, while the reverberant sound, which is assumed to have random direction of incidence and random phase over time, subtracts from coherence. The “total” sound is simply the sum of the direct and the reverberant sound. Let x represent a signal, and subscripts of D and R represent the direct and reverberant components respectively. When necessary, subscripts r and l will be added to indicate the right and left ear respectively. The cross-correlation function

between signals arriving at the left and right ear may then be written as:

$$c(\tau) = \frac{\frac{1}{T} \int_0^T dt (x_{l,D}(t) + x_{l,R}(t))(x_{r,D}(t + \tau) + x_{r,R}(t + \tau))}{\sqrt{P_l P_r}} \quad (4.32)$$

The denominator contains P_l and P_r , which are the total powers of the signals in the left and right ear respectively:

$$P = \frac{1}{T} \int_0^T (x_D(t) + x_R(t))^2 dt \quad (4.33)$$

Now, if it is assumed that the reverberant components are uncorrelated with either the direct sounds or each other, then the integral in the numerator becomes simply:

$$c(\tau) = \frac{\frac{1}{T} \int_0^T dt x_{l,D}(t) x_{r,D}(t + \tau)}{\sqrt{P_l P_r}} \quad (4.34)$$

If the direct signal is the same in each ear except for a delay, then the peak of the cross-correlation function will occur at the value of τ that is equal to that delay. Such a delay makes $x_{l,D}$ and $x_{r,D}$ the same signal, and the integral in the numerator becomes identical to the power in the direct signal, P_D . It is especially expected that the direct signal levels will be the same in each ear for a source at 0° azimuth. The coherence γ is then:

$$\gamma = \frac{P_D}{\sqrt{P_l P_r}} \quad (4.35)$$

Framing this in terms of intensity, it is noted that the ratio of powers equals the ratio of intensities. Let the intensity of the direct sound be I_D , the intensity of the reverberant sound be I_R , and the intensity of the total sound be I_T . The coherence γ should then be calculable from [16]:

$$\gamma = I_D / I_T \quad (4.36)$$

Where $I_T = \sqrt{I_{T_r} I_{T_l}}$. The total intensity in 1/3-octave bands can be easily measured using techniques already described. The direct intensity can be taken to be the intensity in 1/3-octave bands as measured in anechoic conditions, which was

measured as part of Experiment 9. This experiment tested this model of coherence and examined possible further interactions between room properties and binaural parameters.

4.4.1 Methods

Measurements were made in much the same manner as the experiment in the reverberant room (Experiment 10). The room was a lab, 6.5 m by 7.5 m and 4.5 m high. This room was the same room referred to as “10B” by Hartmann *et al.* [51], and it will be referred to similarly in this work. Measurements were taken at six different positions in the room. At each position, the loudspeaker and KEMAR were separated by 3.05 m and faced each other such that the angle between the incident sound and the center of the interaural axis was 0° . At each position, five consecutive measurements were made, each with a different MLS of order 19. Given reverberation times measured in this room by Hartmann, *et al.*, the Schroeder frequency of the room (Equation 2.8) was expected to be approximately 120 Hz.

4.4.2 Coherence Results

Coherences in 1/3-octave bands were calculated and averaged as in Experiment 10. The results are shown in Figure 4.41. Results indicate coherences slightly higher than those found by Hartman, *et al.* [51], as it was for coherences measured in the reverberant environment (Experiment 10, Figure 4.34). A dip in coherence near 800 Hz, and a local minimum in the range of 6.3 – 8 kHz with a slight increase at the highest frequencies are features common to both sets of measurements.

As in the reverberant environment, variance across measurements was notably smaller for the lowest frequencies measured. At low frequencies, the pattern of coherences looks quite similar to those in the reverberant environment — coherence

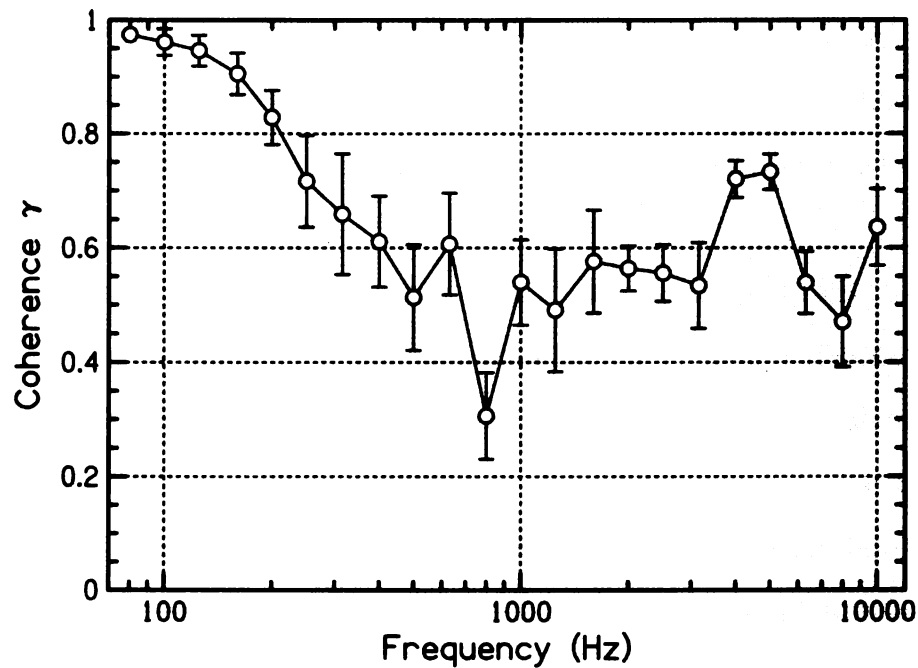


Figure 4.41. Average coherences across all measurements in 1/3-octave bands made in room 10B. Error bars extend one standard deviation in each direction.

drops to 0.72 by 250 Hz in both rooms. However, at higher frequencies, the similarities between the reverberant room and room 10B quickly disappear — at 400 Hz, the coherence in the reverberant room has dropped to 0.28 while the coherence in room 10B drops to only 0.61.

At frequencies above 250 Hz the similarities between the patterns of coherence in the reverberant room and in room 10B seem largely qualitative, but some important quantitative similarities can be observed as well. In both rooms there is a noticeable rise in coherence relative to surrounding frequency bands in the 4 and 5 kHz bands, and there is a slight rise in coherence again at the highest frequency band, 10 kHz. The magnitude of the increase in coherence from 3.15 to 4 kHz is about the same in both environments (0.187 in room 10B and 0.177 in the reverberant room). Similarly, the increase in coherence from 8 to 10 kHz in both environments is roughly the same (0.165 in room 10B and 0.168 in the reverberant room). This increase is similar in behavior to the theoretical results of Lindevald

and Benade in their study of two-ear correlation in rooms [65]. It is likely then that this effect is beneficial in situations where the coherence is not otherwise very high. If the coherence is already near 1, then this effect may increase the coherence but by an unnoticeable amount.

The reason for the rise in coherence at 4 kHz and again at 10 kHz is not fully understood. The wavelength of sound waves between 4 and 5 kHz ranges from 6.8 to 8.5 cm. This corresponds to the approximate expected equivalent head radius [24, 3]. It may be hypothesized that sounds with wavelengths that are integer fractions or multiples of the head diameter benefit coherence, but it should then be true that a frequency near 2 kHz would have the same effect. However, no such rise in coherence is seen around 2 kHz in either the reverberant room or in room 10B, nor is it seen at 8 kHz. The wavelengths of sound that yield a consistent rise in coherence relative to the coherences in the surrounding frequency bands seem related to the equivalent radius of the head. The rise in coherence at 10 kHz can be attributed to the fact that both rooms become less reverberant at higher frequencies.

To compare measured coherences to the model proposed in Equation 4.36, the ratio of total signal power measured in the anechoic room to the total signal power measured in room 10B was calculated and plotted versus the measured coherences in room 10B. The ratio of power should be equal to the ratio of intensities since it is expected that the effective receiving area of the microphones in the KEMAR head are the same in both conditions. A graph of calculated values of I_D/I_T versus measured coherence γ are shown in Figure 4.42. The expected linear trend line has the equation:

$$\gamma = -0.423 \frac{I_D}{I_T} + 0.872$$

The associated Pearson product-moment correlation coefficient is $r = -0.317$ indicating that there is very little tendency for γ and I_D/I_T to increase or decrease

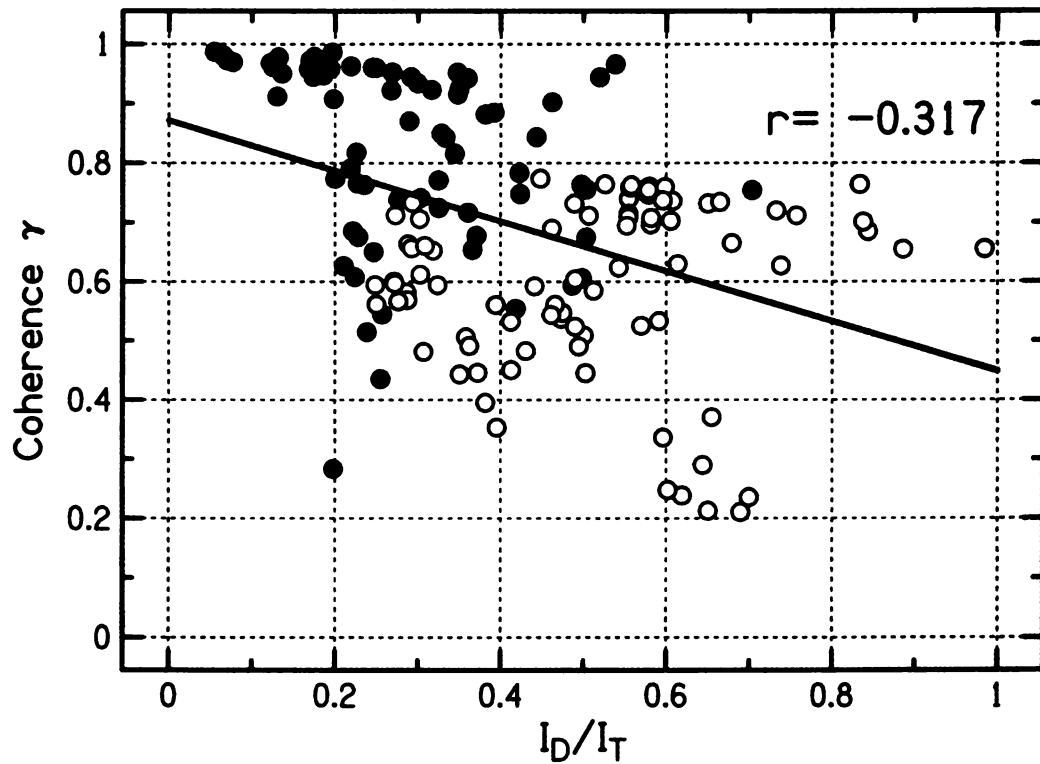


Figure 4.42. A plot of waveform coherence γ versus the ratio of direct to total sound intensity in room 10B. Direct sound intensities were measured in an anechoic environment and total intensities were measured in room 10B. The location of each point is found from the coherence and intensities measured in one of nineteen 1/3-octave bands with ISO center frequencies between 160 Hz and 10 kHz. Points from bands with center frequencies less than or equal to 1 kHz are marked as solid circles, and points from bands above 1 kHz are marked as open circles. This plot is generated from data pooled across all frequency bands and measurements taken at several locations in the room. The linear trend line is shown and the associated Pearson product-moment correlation coefficient r is noted.

together. The coefficient of determination for this fit was $R^2 = 0.174$, indicating a poor goodness of fit of the data to a linear model. It seems then that this model of coherence in rooms fails in a real room situation.

In Figure 4.42, data points from bands with center frequencies equal to or below 1 kHz are plotted as solid circles. Points from bands above 1 kHz are plotted as open circles. Most points from the low frequency group of bands exhibit low direct-to-total intensity ratios, while the direct-to-total intensity ratio tends to be larger for higher frequency bands. This is because the room tends to become more anechoic at higher frequencies and is an expected behavior.

The simple model given by Equation 4.36 supposes that reverberant sound degrades coherence from what is expected for the direct sound. Though this is surely the case, since noises added to otherwise coherent signals will decrease the coherence of those signals, the model that results from this simple starting presumption, Equation 4.36, does not adequately describe the behavior of sounds in a real room, where the “noise” is reverberant energy. This is likely because the reverberant sound itself is correlated nontrivially with the direct sound. In a real room, the reverberant sound is the sum of many attenuated and delayed copies of the direct sound with random angles of incidence. This complicates the relationship between direct and reverberant sound in how they affect interaural coherence.

4.4.3 Reverberation Times

Reverberation times measured in room 10B are shown in Figure 4.43. Averages and standard deviations in each 1/3-octave band were calculated as they were for the reverberant room. Average reverberation times were highest for low frequency bands, with an average RT_{60} , plus and minus one standard deviation, of 0.86 ± 0.022 s for bands between 80 and 315 Hz. Between the 400 and 2,500 Hz bands, the reverberation time is also fairly consistent at 0.75 ± 0.019 s. At frequen-

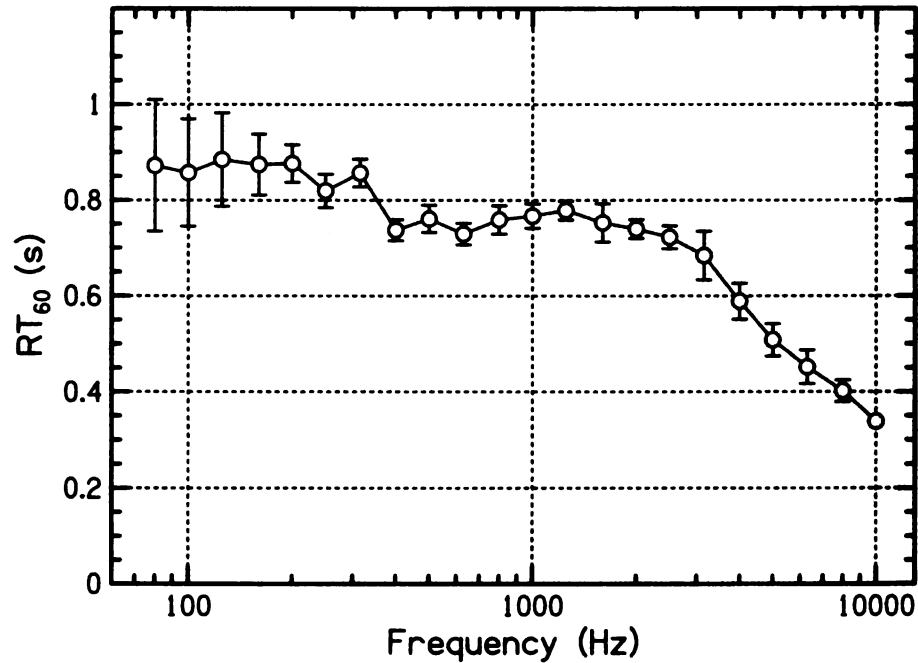


Figure 4.43. 60-dB reverberation times, averaged across all measurements from both ears of a KEMAR in 1/3-octave bands, made in room 10B. Error bars extend one standard deviation in each direction. Small error bars in most bands indicate excellent agreement across the two ears and across different positions in the room. Errors at low frequencies are similar to those seen in Figure 4.36 for the reverberant room.

cies above 2,500 Hz, the RT_{60} in room 10B gradually decreases at the rate of about 0.2 s/octave.

Similar to the reverberation times seen in the reverberant environment (Figure 4.36), the greatest variances in RT_{60} in room 10B are seen at low frequencies. In fact, the magnitude of the variances in RT_{60} in the first five bands (from 80 to 200 Hz) has an average difference from the variances in those bands as measured in the reverberant environment of only 0.021 (two-sided paired t -test, $t = 2.50$, $p = 0.067$). As noted in Experiment 10, however, it is not surprising to expect RT_{60} to vary across positions in a room for low-frequency sounds where standing waves are expected to have a significant effect on the sound field. It is also not surprising to find the variances across positions to be roughly equal in both rooms at low

frequencies since the values of RT_{60} at these frequencies are similar in both rooms. Also similar to results in the reverberant environment, the frequencies that give the greatest variance in RT_{60} in room 10B give the smallest variance in coherence (Figure 4.41).

4.4.4 ITD Results

The ITDs measured in room 10B, corrected for period errors, are shown in Figure 4.44. Results were calculated in a manner identical to that of Experiment 10 for the reverberant environment. As seen in the reverberant room, the lowest frequency bands were more than one standard deviation below 0, confirming the trend of low-frequency ITDs to be slightly negative. Again, this is likely an internal effect.

Similar to ITDs measured in the reverberant room, there is a noticeable variance in the ITDs across positions in room 10B at frequencies up to 800 Hz. This corresponds exactly to the range of frequencies for which a single period error falls outside the physiological range for ITDs of ± 1 ms. However, the variances in room 10B are significantly smaller than those in the reverberant environment (in a one-sided, two-sample t -test, $t = -2.23$, $df = 15$, $p = 0.021$). There were 22 occurrences of period errors in room 10B, compared to 89 total period errors observed for the same number of measurements over the same number of positions in the reverberant room. In a manner similar to that of calculations of ITD in the reverberant room, the ITDs were recalculated for room 10B with a maximum lag in the cross-correlation function of ± 2 ms. The results of this calculation are presented in Figure 4.45 and show reduced variance in bands in which new period errors were detected (the 500, 630, and 800-Hz bands).

In both the highly reverberant environment and in room 10B, variances in ITD measurements are relatively large across trials in frequency bands at or below

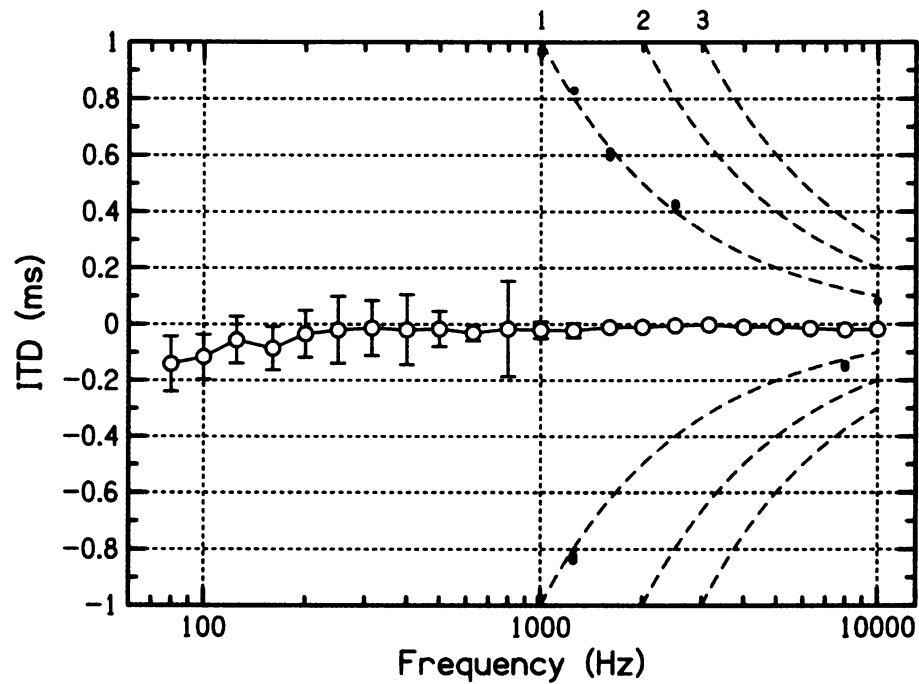


Figure 4.44. ITDs measured between two ears of a KEMAR in room 10B. Data is presented in a manner like that of Figure 4.37. Averages are indicated by open circles, with error bars spanning one standard deviation in each direction. Small dots indicate the ITD of period errors that occurred in at least one trial. Dashed lines indicate the theoretical ITD for one, two and three period errors respectively as a function of frequency.

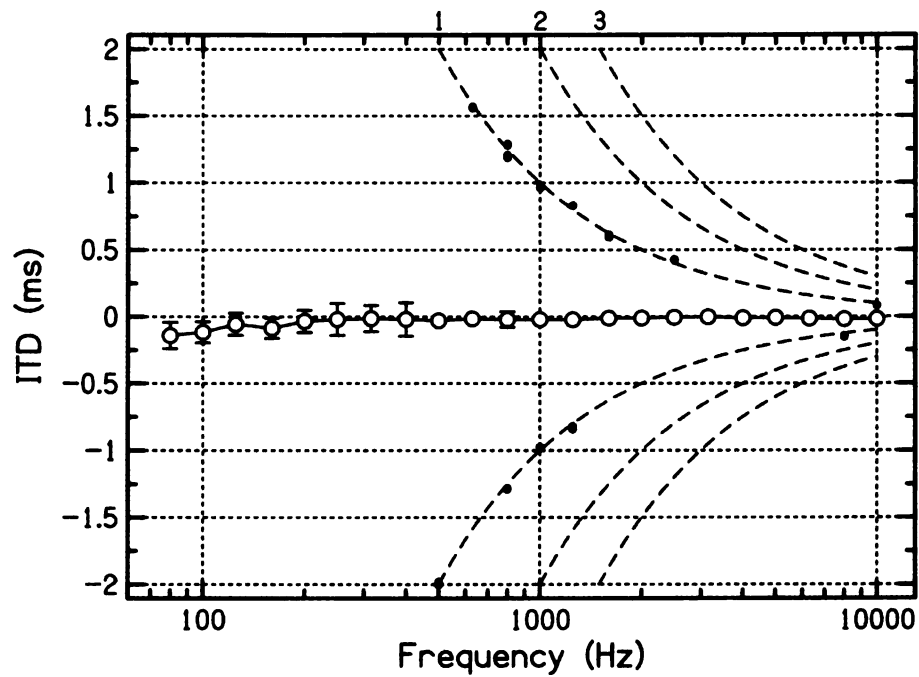


Figure 4.45. ITDs measured between two ears of a KEMAR in room 10B, presented in a manner identical to that of Figure 4.44. ITDs are calculated by finding the lag of the peak in a cross-correlation function between the signals recorded in the left and right ears, with a maximum lag of ± 2 ms.

800 Hz. In both room 10B and the reverberant environment, this is likely a bandwidth effect, since 1/3-octave bandwidths are narrower at lower frequencies. To show this, the ITD was calculated after filtering the signals from the left and right ears into a band centered at 800 Hz with various bandwidths. The estimate of the ITD and the variance of that estimate are plotted as a function of bandwidth in Figure 4.46. As expected, the variances in the estimate of ITD decrease steadily as the bandwidth is increased. It may also be noted that the ITDs, though never very large (0.045 ms for signals filtered into a 1/6-octave band around 800 Hz), approach the expected value of 0 ms as the bandwidth increases. For higher center frequencies, where 1/3-octave bandwidths are wider and reverberation becomes weaker, ITD estimates made in this way are bound to exhibit smaller variance across consecutive measurements.

If auditory critical bands are consistently about 1/3-octave wide, then this could lead to difficulty in estimating ITDs at low frequencies. The variance across position in room 10B of ITDs at low frequencies would likely lead to difficulty in localization of low-frequency sounds. Since ITDs are important for low-frequency sound localization, variation in the ITD across different positions in a room might lead to confusion, limiting the usefulness of this binaural cue. However, it is reasonable to assume that the auditory system makes several “measurements” of ITD, assuming the sound being measured is long enough to allow multiple measurements. Then the average ITD across measurements might be used as an estimate for the ITD within each low-frequency band. The average ITD within a band could then be compared to the ITDs in other low-frequency bands. This would reduce the importance of the within-band variance in ITD estimation.

In many cases in which period errors appeared at a given position, the errors were not consistent across all five trials, and so seemed to appear sporadically. Though there are cases where the same period error is observed on all five trials,

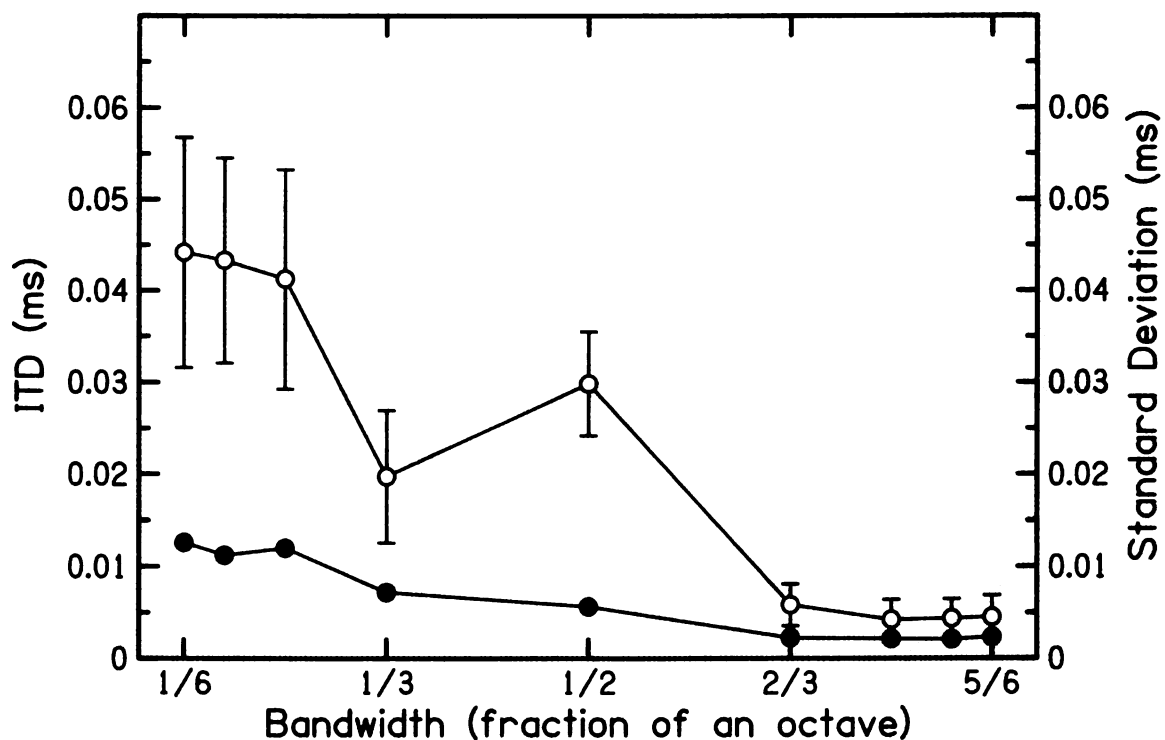


Figure 4.46. Average ITDs (open circles) plus and minus one standard deviation, and sizes of the standard deviations (closed symbols) measured in room 10B. ITDs are measured in a 800 Hz band with a bandwidth as indicated on the abscissa as a fraction of an octave. ITDs approach the expected value for sounds incident at 0° of 0 ms and the variances across measurements decrease as the bandwidth is increased.

such errors are usually isolated to a certain frequency band. Assuming the auditory system makes many measurements of ITD in 1/3-octave bands, period errors could easily be detected and corrected for in a room like 10B. In cases where a period error appears sporadically within a band, the errant measurement of ITD is easily ruled out by comparing it to other ITD estimates within the same critical band (majority rules). In the case of a consistent period error within some band, the period error would be detectable as an ITD that is much different than the ITDs measured in surrounding critical bands. Such discrimination would become increasingly difficult at high frequencies, where the absolute difference in ITD between the true ITD and a period error becomes small. This method of discrimination depends on period errors occurring inconsistently within a critical band, or at least not occurring in many surrounding bands. Such behavior was observed in room 10B and could be expected in similar auditory environments. In a highly reverberant environment, like that measured in Experiment 10, period errors are so pervasive that this model for error correction of period errors might fail or become far less reliable. This would certainly lead to difficulty or confusion in sound localization as is expected in a highly reverberant environment.

4.4.5 ILD Results

The ILDs measured for room 10B in 1/3-octave bands are shown in Figure 4.47. Methods of calculation were identical to those used to calculate ILDs in the reverberant room (Experiment 10). Average ILD values were calculated across all 30 trials (five at each position), and error bars extend one standard deviation in each direction.

The variance in ILD, like that seen in the reverberant room (Figure 4.39) is almost entirely across different positions in the room, and is very small across trials measured at any particular position. The sizes of the variances are not significantly

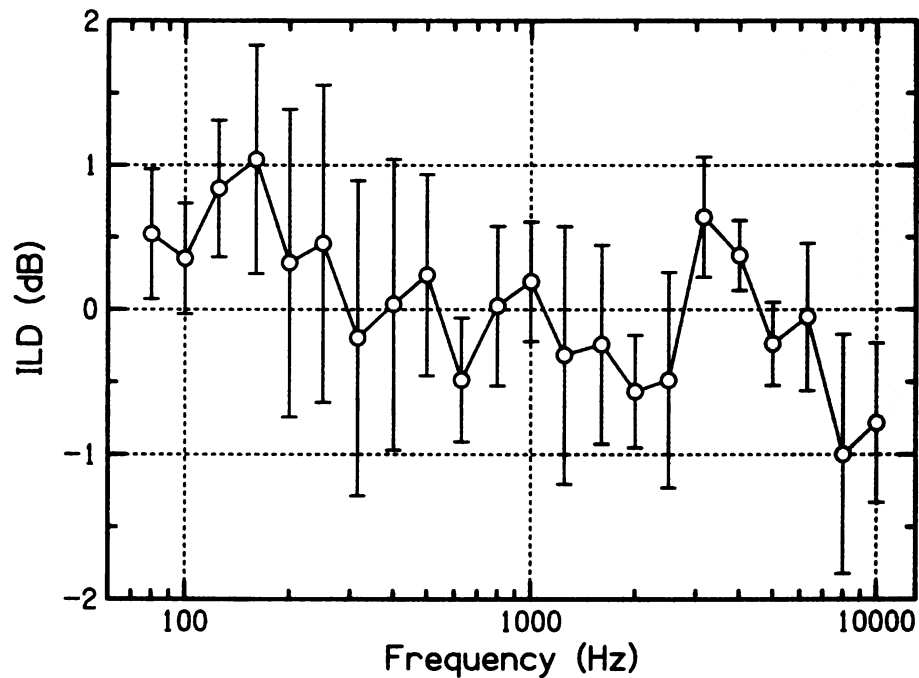


Figure 4.47. ILDs measured between two ears of a KEMAR in room 10B. The presentation of data is the same as that used in Figure 4.39. Data points indicate average measured values, with error bars spanning one standard deviation in each direction. This represents the average of the points shown in Figure 4.48, with error bars extending one standard deviation in each direction. The trend in ILDs is similar to that found in the reverberant environment.

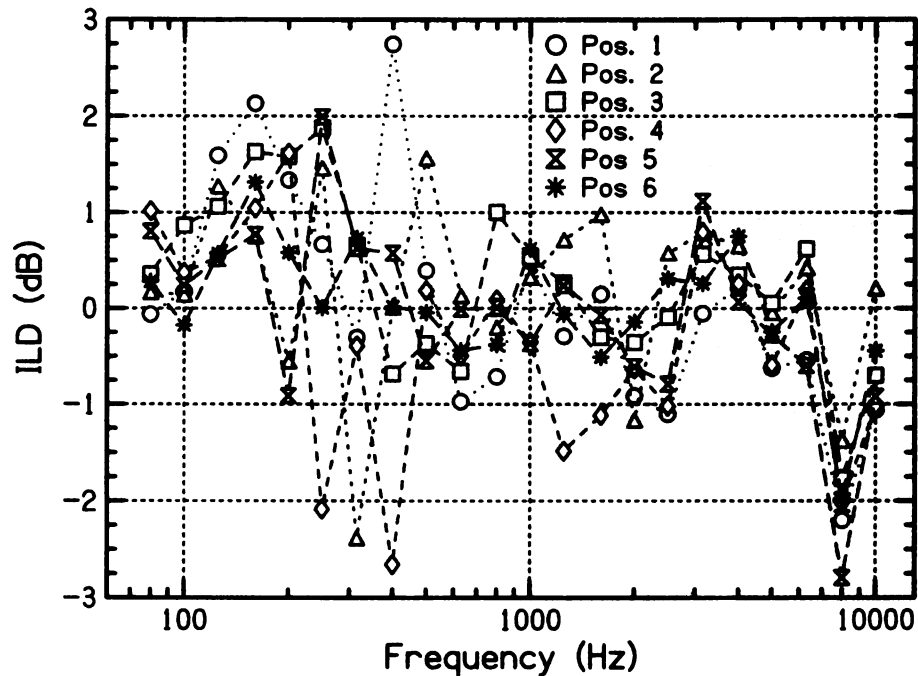


Figure 4.48. ILDs measured in room 10B at six different positions. This figure is akin to Figure 4.40 for ILDs measured in the reverberant room. The data points for each position are labeled with different symbols. Error bars smaller than the size of the data points are not shown.

different in room 10B from what was found in the reverberant room (two-sided two-sample t -test, $t = 0.68$, $df = 40$, $p = 0.499$).

The trend in ILDs across frequency is similar for both room 10B and the reverberant room. Certain features, such as a dip in ILD at 8 and 10 kHz, a local peak at 3.15 kHz, and slightly positive ILDs in the lowest frequency bands, are common to both the measurements made in room 10B and those made in the reverberant room. In fact, the two sets of measurements have a modest Pearson product-moment correlation coefficient of $r = 0.70$. Also, as for the reverberant room, the pattern of ILDs was quite different across positions in the room, as shown in Figure 4.48.

The correlation between ILDs measured in two very different environments — room 10B and the reverberation room — may indicate an effect that contributes to the overall pattern of ILDs only in the presence of significant reverberant en-

ergy. For instance, the dip in ILD at 8 kHz, which appeared at all positions in both rooms, seems to suggest a systematic internal effect, but such a dip does not appear in the ILD at 8 kHz measured in the anechoic environment (Figure 4.25). The same can be said for certain other features of the pattern of ILDs such as the peak in ILD at the 3.15-kHz band. These features are absent in the ILDs measured in anechoic conditions, while other broad features, such as the tendencies for ILDs to be slightly positive at low frequencies, are evident in both reverberant and anechoic conditions.

4.5 Summary and Conclusions

In the measurement of several binaural properties in KEMAR as a function of the incident angle of the sound, several deviations from theoretical expectations were noticed. In Experiment 9, ITD results agreed rather well with the theoretical framework set forth by Rzhevkin [78] and Kuhn [62] (Equations 4.2, 4.4, and 4.5). The general shape of the curves as a function of incident angle — sinusoidal at low frequencies, becoming increasingly triangular with increasing frequency — obeys these equations, and attempts to fit actual ITD data to these equations, yields head radii that decrease as a function of frequency from 8.74 cm at 2.5 kHz to 7.8 cm at 20 kHz.

Results of coherence measurements as a function of incident angle in the anechoic room were surprising. The measured coherence was found to drop for incident angles at and around $\pm 45^\circ$. The extent of this effect was found to be greatest at low frequencies, and decreased until it was no longer noticeable at and above frequencies of about 3.15 Hz. This effect was found to be due to the pinna of the ear nearest the sound source. The pinna introduce both jitter and linear dispersion in the ITD spectrum, though it is predominantly the latter which leads to incoher-

ence. The interaural coherence due to a noise in the ITD spectrum with mean zero and variance σ^2 is given by Equation 4.26 to be:

$$\gamma = \text{Exp} \left[-\frac{1}{2} \omega_o^2 \sigma^2 \right]$$

This incoherence is also observed in measurements taken in the ear canals of human listeners, though the degree of incoherence is significantly less than in KEMAR. Human listeners seem unperturbed by their own inherent coherence properties. In lateralization experiments, listeners showed no difficulty in detecting the direction of shift (to the left or to the right) of sounds that mimicked the coherence characteristics measured in their own ears. When given similar task using noise recorded in KEMAR, the task became quite difficult, but listeners were still able to make some use of ITD cues to detect the direction of the shift. This suggests a familiarity with one's own coherence characteristics. Since the extent of the incoherence seems to depend on the incident angle of the sound source, it may also provide a cue for localization.

In rooms, measurements made at several locations with a "head-on" arrangement (0°) show good consistency across frequency bands and positions with respect to ITD estimation. This is accomplished by locating the lag of a peak in the cross-correlation function of the signals in the left and right ears. The cross-correlation function is limited in calculation to some maximum lag, τ_m . Selection of this maximum lag has consequences for the resulting estimation of the ITD because of the existence of period errors. Period errors can only be observed in frequency bands where $1/f_c < \tau_m$, where f_c is the center frequency of the band. In those bands, the ITD estimation can be corrected when period errors occur. In bands for which $1/f_c > \tau_m$, period errors can not be detected. This leads to a larger degree of variance in the resulting average ITD measurements within those bands.

The question of how coherences come to be so low in certain frequency bands

in rooms seems to be answered by the variance in the ITD spectra. Thus, incoherence found for certain incident angles of sound in anechoic conditions, which is due to the pinnae, is similar in nature to incoherence measured in rooms for sound sources straight ahead of the listener. The expected coherence in both situations obeys Equation 4.31. Linear dispersion in the ITD has a negligible effect on the coherence. The amount of variance in the ITD spectrum required to cause a noticeable drop in coherence is greater at lower frequencies (since the coherence is a function of $\omega_0\sigma$) and this may be in part the cause for the very high coherences observed in the lowest frequency bands even in reverberant conditions.

APPENDIX A

Test for Interaural Differences

The methods used in the first chapter of this dissertation had the goal of eliminating interaural differences by confining the sources to the front-back dimension. However, interaural differences can not be completely eliminated, no matter how accurately the experimental system is aligned. Nevertheless, it is thought that the small interaural differences that exist in a geometry such as is used in the experiments on informational masking presented here have no perceptual importance [70].

To test for the perceptual effect of interaural differences, which might arise because the listener is inadvertently misaligned or because of individual anatomical asymmetry, a test experiment was run in which misalignments were deliberately introduced such that the ears were not equidistant from each loudspeaker.

The test effectively used a dramatic misalignment, equivalent to a 5° rotation of the listener. Accordingly, one ear was effectively $65 \mu\text{s}$ closer to the front loudspeaker and $65 \mu\text{s}$ farther away from the back loudspeaker, for a total difference in arrival time of $130 \mu\text{s}$. The same difference, but with opposite sign, occurred in the other ear.

Seven listeners, including B and N from the previous experiments, were pre-

sented with CRM stimuli of the same type used in Experiment 1 (target plus two speech distracters) through Sennheiser HD 414 headphones. The task was identical to that of previous experiments. The distracters were passed through delay-and-add filters before they were combined with the target speech. Two experimental scenarios were tested — one in which the delay between the maskers was $870\ \mu\text{s}$ in the left ear and $1130\ \mu\text{s}$ in the right, the “different delays” scenario; and a second condition in which the delay between the maskers was the same in both ears $1000\ \mu\text{s}$, the “same delays” scenario.

A reference delay of $1000\ \mu\text{s}$ was chosen because that delay is a typical value beyond the range at which release from energetic masking occurs (Experiment 2). For $1000\ \text{ms}$ there is substantial release (but not the greatest release) from informational masking. These delay scenarios were tested at SNRs of $-4\ \text{dB}$, $0\ \text{dB}$, and $+4\ \text{dB}$, for a total of $2 \times 3 = 6$ conditions.

Differences in performance between the scenarios were examined within and across listeners for each SNR. The most relevant SNR is $0\ \text{dB}$, at which Experiment 1 was performed. At $0\ \text{dB}$, the average listener performed marginally better with different delays than with same delays. The mean improvement in percent correct, plus and minus one standard deviation, was $7 \pm 16\%$. This was not a significant improvement (one-sided Wilcoxon Signed Rank test, $N = 5$, $W^+ = 3.0$, $p = 0.140$). There were also differences in performance across listeners — three listeners performed slightly better with different delays, two performed slightly better with same delays, and two showed no difference in performance. For lower SNR, there was a greater improvement in performance under the different-delays condition, but the difference in performance did not reach the 0.05 level of significance ($N = 5$, $W^+ = 1.5$, $p = 0.069$). For higher SNR, the differences in performance between the same-delay and different-delays conditions were entirely negligible ($N = 6$, $W^+ = 11.0$, $p = 0.583$). Note here that a worst-case scenario

(a 5° rotation of the listener) has been assumed. In reality, the error in alignment was almost certainly smaller. Thus, ruling out differences in the delay-and-add filtering between the two ears as a possible source of unmasking seems justified.

APPENDIX B

Separated-Source Presentation with Speech Distracters

The positive delay conditions of Experiment 1, particularly the conditions for which $\tau > 1$ ms, can be expected to elicit a precedence effect shift in the *perceived* location of the distracters. The precedence effect should shift the *perceived* location of the distracters from front to back. The masking release caused by this perceptual shift is hypothesized to be comparable to the release that would be obtained if there were an actual *physical* shift in the location of the distracters. Several experiments have been performed which have measured the release from both energetic masking and informational masking when the distracters are moved from the front, where they were collocated with the speech target, to the back [40, 56, 75, 99].

To test this hypothesis using the geometry of the previous experiments and the CRM stimuli, the experiment of Freyman *et al.* [40] was repeated. Three conditions were tested: (1) The Front-Only baseline, as in Experiment 1, with SNR of 0 dB. (2) The Front-Only baseline, as in Experiment 1, with SNR of 4 dB. (3) A separated-source test with the target presented from the front loudspeaker and the speech distracters presented from the back loudspeaker at a SNR of 0 dB. The first two

conditions were the FO and FO+4 dB conditions. The third condition was new.

Four listeners were tested, including listeners B and N from Experiment 1 and two inexperienced listeners, A and E. Listeners sat quietly and faced the front loudspeaker without moving their heads throughout each run. Listeners were positioned such that their ears were half way between the front and back loudspeakers. Each run consisted of five practice trials followed by 30 trials for which data were collected and scored as in Experiment 1. Each listener went through five runs of each condition.

The results of the experiment are presented in Table B.1. Performance in the separated-source condition was slightly worse than that in the FO+4 dB condition, indicating a release from masking somewhat less than 4 dB as a result of spatially separating the target and distracters. In detail, the average percent correct performance across listeners in the separated condition was $75 \pm 4.8\%$, an increase over the FO baseline performance of $38 \pm 7.4\%$. This increase in performance in the F-B condition corresponds to an average release of 3.5 dB. This is comparable in magnitude to the release seen in Experiment 1 between the FO baseline performance and the best performance at any other delay for the average listener (this occurred at $\tau = 2$ ms, where the average release was 3.5 dB).

The release from masking seen in this separated-source experiment is larger than that seen at almost all the delays in Experiment 1. However, the comparison is not quite fair because the observed release in Experiment 1 is reduced by the fact that there is twice as much masker power in Experiment 1.

In the end, the hypothesis that the effect of a perceptual shift should be comparable to the effect of an actual physical shift in the location of the distracters is supported by the comparison of the separated-source experiment and Experiment 1. However, the symmetry of the results in Experiment 1, whereby performance was similar for both positive and negative delays, proves that the release from masking

Listener	(a) FO	(b) FO+4 dB	(c) F-B
A	26 ± 8.3%	71 ± 5.0%	75 ± 8.4%
B	34 ± 8.6%	82 ± 4.5%	66 ± 7.8%
E	22 ± 9.0%	75 ± 3.8%	58 ± 6.9%
N	38 ± 9.0%	73 ± 4.7%	75 ± 8.3%
Avg	30 ± 7.4%	75 ± 4.8%	68 ± 7.4%

Table B.1. Separated-source Experiment: Percentage of correct responses (plus or minus one standard deviation) for each listener by condition. (a) Front-Only base-line condition. (b) Front-Only condition with +4 dB target level relative to a single distracter. (c) Target in front and distracters in back. Averages and standard deviations for each listener are calculated across five runs. Averages and standard deviations across listeners are shown in the bottom row.

caused by the perceived relocation or delocation of the distracters involves more than just the localization precedence effect.

APPENDIX C

Wiener filtering

One common method of reducing noise from a signal is by Wiener filtering. As opposed to traditional noise filters which attempt to remove specific parts of the spectrum of a noisy signal attributed to the noise alone, the Wiener filter takes a statistical approach to filtering, and requires some knowledge of the spectral properties of the signal and additive noise. Both processes are assumed to be linear and stationary. Though a causal Wiener filter must be physically realizable, in the following experiments, Wiener filtering will only be performed *a posteriori*, and so a non-causal filter is utilized. The Wiener filter uses a minimum mean square error criterion filter design.

Excellent introductions to the Wiener filter can be found in several texts [85, 89]. For a signal, $x(t)$, corrupted by additive stationary white noise, $n(t)$, let $y(t)$ be the observed signal such that $y(t) = x(t) + n(t)$. The output of a filter $w(t)$ on that signal and noise combination is:

$$\hat{x}(t) = w(t) \otimes [x(t) + n(t)]$$

The symbol $\otimes$ represents the convolution operation. The estimator $\hat{x}(t)$ will be related to $y(t)$ by a linear combination of a range of known values of $y(t)$ between

times t_0 and t_f via the filter w by:

$$\hat{x}(t) = \int_{t_0}^{t_f} w(t - \varepsilon) y(\varepsilon) d\varepsilon$$

The orthogonality principle states that for linear mean-square error estimation, the error in the estimate, $\tilde{x}$, must be orthogonal to the observations, y . Noting that $\tilde{x}(t) = x(t) - \hat{x}(t)$, we then have as an expectation in time:

$$E \left\{ \left[x(t) - \int_{t_0}^{t_f} w(t - \varepsilon) y(\varepsilon) d\varepsilon \right] y(\tau) \right\} = 0 \quad t_0 \leq \tau \leq t_f$$

Expanding the argument of the expectation operation gives:

$$E \left\{ x(t) y(\tau) - \int_{t_0}^{t_f} w(t - \varepsilon) y(\varepsilon) y(\tau) d\varepsilon \right\} = 0 \quad t_0 \leq \tau \leq t_f$$

This can be written in terms of cross correlation functions. Let $R_{pq}(\alpha - \beta) = E[p(\alpha)q(\beta)]$ be the cross correlation between two jointly wide-sense stationary functions, p and q . Then we can write the above equation as:

$$R_{xy}(t - \tau) - \int_{t_0}^{t_f} w(t - \varepsilon) R_y(\varepsilon - \tau) d\varepsilon = 0 \quad t_0 \leq \tau \leq t_f \quad (C.1)$$

This is often referred to as the Wiener-Hopf integral equation [89].

Let $W(s)$ be the Laplace transform of $w(t)$, and $\Phi_{xy}(s)$ be the Laplace transform of $R_{xy}(t)$. If the period of observation is very long, then the observations y and the desired estimate $\hat{x}$ are jointly stationary as long as the filter is time-invariant. The solution to the Wiener-Hopf equation for a noncausal filter in Laplace space is then:

$$W(s) = \frac{\Phi_{xy}(s)}{\Phi_y(s)} \quad (C.2)$$

In the case where the noise is uncorrelated with the input signal x , Φ_{xy} becomes simply Φ_{xx} and Φ_y becomes $\Phi_{xx} + \Phi_{nn}$. Thus, Equation C.2 can be reduced to one involving only autocorrelations. If the noise is assumed to be white then its autocorrelation function is simply a delta function (1 in Laplace space). Now, something must be known about the autocorrelation of the input signal, x . This

is far less stringent a requirement than having to know the entire signal, however, and an approximate autocorrelation will still afford the Wiener filter quite a lot of filtering ability.

The mean square error of the estimation, after utilizing the orthogonality principle, is:

$$E [\tilde{x}^2] = E \{ [x(t) - \hat{x}(t)] x(t) \}$$

Again, in terms of cross correlation functions, this error becomes:

$$E [\tilde{x}^2] = R_x(0) - \int_{t_0}^{t_f} w(t - \epsilon) R_{yx}(t - \epsilon) d\epsilon \quad (C.3)$$

In the system in which it was utilized in these experiments, the Wiener filter was used to reduce the noise in the IR. Equation C.2 shows that the Wiener filter requires *a priori* knowledge of the autocorrelations of both the noise and the signal. The corrupting noise is usually assumed to be white, and its autocorrelation is equal to its variance. The noise can be effectively isolated by examining the tail of the measured IR, where the IR is expected to be zero. Any deviation from zero in the IR at very high times, such as the last quarter of the IR, is then assumed to be due entirely to the noise. The Wiener filter was given the power of the signal in the tail of the IR measured by the RTN and then operated on the entire IR to arrive at a noise-reduced IR, h_w .

APPENDIX D

Guide to Acronyms

The following acronyms and symbols are commonly used in this dissertation, and so a friendly guide is provided for reference.

γ : interaural coherence

γ_e : interaural envelope coherence

ADD: added, delayed distracter

ASW: auditory source width

CC: cross-correlation

df : degrees of freedom

FFT: fast Fourier transform

FIR: finite impulse response

FO: front-only

HP: horizontal plane

HRTF: head-related transfer function

IIDC: integrated impulse decay curve

ILD: interaural level difference

IPD: interaural phase difference

IR: impulse response

ISO: International Organization for Standardization

ITD: interaural time difference

JND: just-noticeable difference

JVC: Victor Company of Japan

KEMAR: Knowles Electronics Manikin for Acoustic Research

LFSR: linear feedback shift register

LTI: linear, time-invariant

MLS: maximum length sequence

MSP: median sagittal plane

PDF: probability density function

RT_{60} : 60-dB reverberation time

RTN: random telegraph noise

SNR: signal-to-noise ratio

SPL: sound pressure level TF: transfer function

BIBLIOGRAPHY

- [1] A. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth Dover printing, tenth GPO printing edition, 1964.
- [2] V. R. Algazi, C. Avendano, and R. O. Duda. Estimation of a spherical-head model from anthropometry. *J. Audio Eng. Soc.*, 49, 2001.
- [3] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF database. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 2001.
- [4] Y. Ando. *Architectural Acoustics - Blending Sound Sources, Sound Fields, and Listeners*. Springer-Verlag, New York, 1998.
- [5] C. Avendano, V. R. Algazi, and R. O. Duda. A head-and-torso model for low-frequency binaural elevation effects. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 1999.
- [6] R. L. Balakrishnan, U. andd Freyman. Speech detection in spatial and non-spatial speech maskers. *J. Acoust. Soc. Am.*, 123:2680–2691, 2008.
- [7] D. W. Batteau. The role of the pinna in human localization. *Proc. Royal Soc. London. Series B, Biological Sciences*, 168:158–180, 1967.
- [8] S. Bedard, B. Champagne, and A. Stephenne. Effects of room reverberation on time-delay estimation performance. IEEE International Conference on Acoustics, Speech, and Signal Processing, Apr. 1994.
- [9] L. L. Beranek. *Acoustics*. American Institute of Physics, New York, 1986.
- [10] L. L. Beranek and H. P. Sleeper Jr. The design and construction of anechoic sound chambers. *J. Acoust. Soc. Am.*, 18:140–150, 1946.
- [11] L. Bernstein. Personal communication, Sept. 2007.

- [12] L. R. Bernstein and C. Trahiotis. Why do transposed stimuli enhance binaural processing?: Interaural envelope correlation vs envelope normalized fourth moment.
- [13] L. R. Bernstein and C. Trahiotis. 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.
- [14] L. R. Bernstein, S. van de Par, and C. Trahiotis. The normalized interaural correlation: Accounting for NoS π thresholds obtained with Gaussian and “low-noise” masking noise. *J. Acoust. Soc. Am.*, 106:870–976, 1999.
- [15] J. Blauert. Sound localization in the median plane. *Acustica*, 22:205–213, 1969/70.
- [16] J. Blauert. *Spatial Hearing*. The MIT Press, Combridge, Massachusetts, 1983.
- [17] J. Blauert and W. Lindemann. Spatial mapping of intracranial auditory events for various degrees of interaural coherence. *J. Acoust. Soc. Am.*, 79:806–813, 1986.
- [18] H. C. Blodgett, W. A. Wilbanks, and L. A. Jeffress. Effect of large interaural time differences upon the judgment of sidedness. *J. Acoust. Soc. Am.*, 28:639–643, 1956.
- [19] R. S. Bolia, W. T. Nelson, M. A. Ericson, and B. D. Simpson. A speech corpus for multitalker communications research. *J. Acoust. Soc. Am.*, 107:1065–1066, 2000.
- [20] J. Borish and J. B. Angell. An efficient algorithm for measuring the impulse response using pseudorandom noise. *J. Audio Eng. Soc.*, 31:478–487, 1983.
- [21] A. W. Bronkhorst. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acust.*, 86:117–128, 2000.
- [22] D. Brungart, B. Simpson, T. Darwin, C. Arbogast, and G. J. Kidd. Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task. *J. Acoust. Soc. Am.*, 117:292–304, 2005.
- [23] D. Brungart, B. Simpson, and R. L. Freyman. Precedence-based speech segregation in a virtual auditory environment. *J. Acoust. Soc. Am.*, 118:3241–3251, 2005.
- [24] M. D. Burkhard and R. M. Sachs. Anthropometric manikin for acoustic research. *J. Acoust. Soc. Am.*, 58:214–222, 1975.

- [25] C. I. Cheng and G. H. Wakefield. Spatial frequency response surfaces (SFRS's): An alternative visualization and interpolation technique for head-related transfer functions (HRTF's). *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999.
- [26] E. C. Cherry. Some experiments on the recognition of speech with one and two ears. *J. Acoust. Soc. Am.*, 25:975–979, 1953.
- [27] I. Chun, B. Rafaely, and P. Joseph. Experimental investigation of spatial correlation in broadband reverberant sound fields. *J. Acous. Soc. Am.*, 113:1995–1998, 2003.
- [28] M. Cohn and A. Lempel. On fast M-sequence transforms. *IEEE Trans. Info. Theory*, 23:135–137, 1977.
- [29] Z. A. Constan. *All Things Coherence*. PhD thesis, Michigan State University, 2002.
- [30] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Jr. Thompson. Measurement of correlation coefficients in reverberant sound fields. *J. Acoust. Soc. Am.*, 27:1072–1077, 1955.
- [31] J. F. Culling, H. S. Colburn, and M. Spurchise. Interaural correlation sensitivity. *J. Acoust. Soc. Am.*, 110:1020–1029, 2001.
- [32] R. O. Duda and W. L. Martens. Range dependence on the response of a spherical head model. *J. Acoust. Soc. Am.*, 104:3048–3058, 1998.
- [33] C. Dunn and M. Hawksford. Distortion immunity of mls-derived impulse response measurements. *J. Audio Eng. Soc.*, 41:314–335, 1993.
- [34] N. I. Durlach, K. J. Gabriel, H. S. Colburn, and C. Trahiotis. Interaural correlation discrimination: II. Relation to binaural unmasking. *J. Acoust. Soc. Am.*, 79:1548–1557, 1986.
- [35] N. I. Durlach, C. R. Mason, G. J. Kidd, T. L. Arbogast, H. S. Colburn, and B. G. Shinn-Cunningham. Note on informational masking (I). *J. Acoust. Soc. Am.*, 113:2984–2987, 2003.
- [36] A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. 108th Convention of the Audio Engineering Society, 2000.
- [37] W. E. Feddersen, T. T. Sandel, D. C. Teas, and L. A. Jeffress. Localization of high-frequency tones. *J. Acoust. Soc. Am.*, 29:988–991, 1957.

- [38] R. L. Freyman, U. Balakrishnan, and K. S. Helfer. Spatial release from informational masking in speech recognition. *J. Acoust. Soc. Am.*, 109:2112–2122, 2001.
- [39] R. L. Freyman, U. Balakrishnan, and K. S. Helfer. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J. Acoust. Soc. Am.*, 115:2246–2256, 2004.
- [40] R. L. Freyman, K. S. Helfer, and U. Balakrishnan. Spatial and spectral factors in release from informational masking in speech recognition. *Acta Acustica*, 91:537–545, 2005.
- [41] R. L. Freyman, K. S. Helfer, D. D. McCall, and R. K. Clifton. The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am.*, 106:3578–3588, 1999.
- [42] K. J. Gabriel and H. S. Colburn. Interaural correlation discrimination I: Bandwidth and level dependence. *J. Acoust. Soc. Am.*, 69:1394–1401, 1981.
- [43] B. R. Glasberg and B. C. J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.*, 47:102–138, 1990.
- [44] S. W. Golomb. *Shift register sequences*. Holden-Day, San Francisco, 1967.
- [45] D. W. Grantham and F. L. Wightman. Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. *J. Acoust. Soc. Am.*, 65:1509–1517, 1979.
- [46] H. Haas. On the influence of a single echo on the intelligibility of speech. *Acustica*, 1:49–58, 1951.
- [47] W. M. Hartmann. *Signals, Sound, and Sensation*. Springer-Verlag, New York, 1998.
- [48] W. M. Hartmann. How we localize sound. *Physics Today*, pages 24–29, Nov. 1999.
- [49] W. M. Hartmann. The cross-correlation function, Feb. 2007.
- [50] W. M. Hartmann and B. Rakerd. Localization of sound in reverberant spaces. *J. Acoust. Soc. Am.*, 105:1149, 1999.
- [51] W. M. Hartmann, B. Rakerd, and A. Koller. Binaural coherence in rooms. *Acta Acust.*, 91:451–462, 2005.
- [52] J. Hebrank and D. Wright. Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.*, 56:1829–1834, 1974.

- [53] E. Hecht. *Optics*. Addison Wesley, Reading, Massachusetts, 4th edition, 2001.
- [54] G. B. Henning. Detectability of interaural delay in high-frequency complex waveforms. *J. Acoust. Soc. Am.*, 55:84–90, 1974.
- [55] T. Hidaka, L. L. Beranek, and T. Okano. Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls. *J. Acoust. Soc. Am.*, 98:988–1007, 1995.
- [56] I. J. Hirsh. Relation between localization and intelligibility. *J. Acoust. Soc. Am.*, 22:196–200, 1950.
- [57] A. Ihlefeld and B. G. Shinn-Cunningham. Effect of source location and listener location on ild cues in a reverberant room. *J. Acoust. Soc. Am.*, 115:2598, 2004.
- [58] F. Jacobsen and T. Roisin. The coherence of reverberant sound fields. *J. Acoust. Soc. Am.*, 108:204–210, 2000.
- [59] L. A. Jeffress, H. C. Blodgett, and B. H. Deatherage. Effect of interaural correlation on the precision of centering a noise. *J. Acoust. Soc. Am.*, 34:1122–1123, 1962.
- [60] C. H. Keller and T. T. Takahashi. Binaural cross-correlation predicts the responses of neurons in the owl's auditory space map under conditions simulating summing localization. *J. Neurosci.*, 16:4300–4309, 1996.
- [61] G. J. Kidd, C. R. Mason, P. S. Deliwala, W. S. Woods, and H. S. Colburn. Reducing informational masking by sound segregation. *J. Acoust. Soc. Am.*, 95:3475–3480, 1994.
- [62] G. F. Kuhn. Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.*, 62:157–167, 1977.
- [63] M. Kuster. Spatial correlation and coherence in reverberant acoustic fields: Extension to microphones with arbitrary first-order directivity. *J. Acoust. Soc. Am.*, 123:154–162, 2008.
- [64] M. R. Leek, M.E. Brown, and M. F. Dorman. Informational masking and auditory attention. *Percept. Psychophys.*, 50:205–214, 1991.
- [65] I. M. Lindevald and A. H. Benade. Two-ear correlation in the statistical sound fields of rooms. *J. Acoust. Soc. Am.*, 80:661–664, 1986.
- [66] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *J. Acoust. Soc. Am.*, 106:1633–1654, 1999.

- [67] R. Y. Litovsky, B. Rakerd, T. C. T. Yin, and W. M. Hartmann. Psychophysical and physiological evidence for a precedence effect in the median sagittal plane. *J. Neurophysiol.*, 77:2223–2226, 1997.
- [68] R. W. Marsh. *Table of irreducible polynomials over GF(2) through degree 19*. Office of Technical Services, U.S. Dept. of Commerce, Washington, D.C., 1957.
- [69] D. McFadden and E. G. Pasanen. Binaural detection at high frequencies with time-delayed waveforms. *J. Acoust. Soc. Am.*, 63:1120–1131, 1978.
- [70] J. C. Middlebrooks and D. M. Green. Sound localization by human listeners. *Ann. Rev. Psych.*, 42:135–159, 1991.
- [71] B. C. J. Moore and B. R. Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.*, 74:750–753, 1983.
- [72] N. Moray. Attention in dichotic listening: Affective cues and the influence of instructions. *Quart. J. Exp. Psych.*, 11:56–60, 1959.
- [73] M. Morimoto, H. Fujimori, , and Z. Maekawa. Discrimination between auditory source width and envelopment. *J. Acoust. Soc. Japan*, 46:449–457, 1990.
- [74] T. Okano, L. L. Beranek, and T. Hidaka. Relations among interaural cross-correlation coefficient ($iacc_e$), lateral fraction (lf_e), and apparent source width (asw) in concert halls. *J. Acoust. Soc. Am.*, 104:255–265, 1998.
- [75] R. Plomp. Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). *Acustica*, 34:200–211, 1976.
- [76] B. Rakerd, N. L. Aaronson, and W. M. Hartmann. Release from speech-on-speech masking by repeated presentation of the masker. *J. Acoust. Soc. Am.*, 119:1597–1605, 2006.
- [77] S. K. Roffler and R. A. Butler. Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am.*, 48:1255–1259, 1968.
- [78] S. N. Rzhevkin. *The Theory of Sound*. Pergamon Press, Oxford, 1963.
- [79] K. Saberi, Y. Takahashi, R. Egnor, H. Farahbod, J. Mazer, and M. Konishi. Detection of large interaural delays and its implication for models of binaural interaction. *J. Assoc. Res. Otolaryng.*, Online:80–88, 2001.
- [80] D. V. Sarwate and M. B. Pursley. Crosscorrelation properties of pseudorandom and related sequences. *Proc. IEEE*, 68:593–619, 1980.

- [81] M. R. Schroeder. New method of measuring reverberation time. *J. Acoust. Soc. Am.*, 37:409–412, 1965.
- [82] M. R. Schroeder. The “Schroeder frequency” revisited. *J. Acoust. Soc. Am.*, 99:3240–3241, 1996.
- [83] M. R. Schroeder. Integrated-impulse method measuring sound decay without using impulses. *J. Acoust. Soc. Am.*, 66:497–500, 1997.
- [84] T. M. Shackleton, R. H. Arnott, and A. R. Palmer. Sensitivity to interaural correlation of single neurons in the inferior colliculus of guinea pigs. *J. Assoc. Res. Otolaryngology*, 6:244–259, 2005.
- [85] M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan. *Introduction to statistical signal processing with application*. Prentice-Hall, Inc., New Jersey, 1996.
- [86] C. Trahiotis, L. R. Bernstein, R. M. Stern, and T. N. Buell. Interaural correlation as the basis of a working model of binaural processing: An introduction. In A. N. Popper and R. R. Fay, editors, *Sound Source Localization*, pages 238–271. Springer, New York, 2005.
- [87] S. van de Par and A. Kohlrausch. Analytical expressions for the envelope correlation of narrow-band stimuli used in CMR and BMLD research. *J. Acoust. Soc. Am.*, 103:3605–3620, 1998.
- [88] S. van de Par, C. Trahiotis, and L. R. Bernstein. A consideration of the normalization that is typically included in correlation-based models of binaural detection. *J. Acoust. Soc. Am.*, 109:830–833, 2001.
- [89] H. L. van Trees. *Detection, estimation, and modulation theory*. John Wiley & Sons, Inc., New York, 2001. Part I.
- [90] M. Vorländer and H. Bietz. Comparison of methods for measuring reverberation time. *Acustica*, 80:205–215, 1994.
- [91] H. Wallach, E. B. Newman, and M. R. Rosenzweig. The precedence effect in sound localization. *Am. J. Psychol.*, 57:315–336, 1949.
- [92] C. S. Watson, W. J. Kelly, and H. W. Wroton. Factors in the discrimination of tonal patterns: II. selective attention and learning under various levels of stimulus uncertainty. *J. Acoust. Soc. Am.*, 60:1176–1185, 1976.
- [93] F. L. Wightman and D. J. Kistler. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, 105:2841–2853, 1999.

- [94] R. S. Woodworth and H. Scholberg. *Experimental Psychology*. Holt, Rinehart, and Winston, New York, 3rd edition, 1962.
- [95] N. Xiang, J. N. Daigle, and M. Kleiner. Simultaneous acoustical channel measurement via maximal-length-related sequences. *J. Acoust. Soc. Am.*, 117:1889–1894, 2005.
- [96] N. Xiang and M. R. Schroeder. Reciprocal maximum-length sequence pairs for acoustical dual source measurements. *J. Acoust. Soc. Am.*, 113:2754–2761, 2003.
- [97] W. A. Yost. The cocktail party problem: Forty years later. In R. H. Gilkey and T. R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 329–347. Erlbaum, Hillsdale, NJ, 1997.
- [98] W. A. Yost and E. R. Hafter. Lateralization. In W. A. Yost and G. Gourevitch, editors, *Directional Hearing*, page 62. Springer, New York, 1987.
- [99] P. M. Zurek. Binaural advantages and directional effects in speech intelligibility. In G. A. Studebaker and I. Hockberg, editors, *Acoustical Factors Affecting Hearing Aid Performance*. Allyn and Bacon, Needham Heights, MA, 2nd edition, 1993.
- [100] J. Zwislöcki and R. S. Feldman. Just noticeable differences in dichotic phase. *J. Acoust. Soc. Am.*, 28:860–864, 1956.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02956 7504