

MISLEADING AND CONFLICTING CUES IN HUMAN SOUND LOCALIZATION

By

Eric John Macaulay

A DISSERTATION

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

Physics—Doctor of Philosophy

2015

ABSTRACT

MISLEADING AND CONFLICTING CUES IN HUMAN SOUND LOCALIZATION

By

Eric John Macaulay

Human sound localization in the azimuthal plane is primarily cued by the interaural time difference (ITD) and the interaural level difference (ILD). For sine tones with frequencies greater than about 1500 Hz, the ILD is the only steady-state cue available to listeners. In free field, plane waves incident on a listener diffract around the head which results in an acoustical bright spot on the opposite side of the head as the source. This results in the ILD cue being non-monotonic with azimuth. Listeners' localization responses to stimuli in free field are highly correlated with this misleading ILD cue, as listeners err dramatically in localization of sources with large azimuths. Discrimination experiment results also confirm the misleading nature of the ILD cue. Training and feedback provide very minimal performance improvements for listeners. However, the introduction of either amplitude modulation, or narrow-band noise of the same bandwidth helped some listeners greatly, and others to a lesser extent. The non-monotonic cue remains prominent for noise with bandwidths as wide as 4 octaves when centered around 1500 kHz.

Amplitude modulation introduces an envelope interaural time difference (EITD) as a localization cue for listeners. Sinusoidally-modulated amplitude modulation signals become altered by a listener's anatomy. The EITD cue is often misleading and in conflict with the ILD cue. Nevertheless, the two cues together usually result in accurate localization. The quality of the amplitude modulation in these experiments was not degraded significantly and does not impact the weight of the EITD.

This work is dedicated to my parents.

ACKNOWLEDGMENTS

This work would not have been possible without the help of many individuals and organizations. My advisor, Dr. William M. Hartmann, has provided invaluable guidance to me during my graduate studies. I thank him for his patience and wisdom.

Dr. Brad Rakerd has been an excellent mentor and collaborator of mine. He has always been eager to share his time and insight, and is a pleasure to work with.

I thank the other members of my guidance committee for their advice and interest in my research: Dr. Alex Brown, Dr. Douglas Brungart, Dr. Jon Pumplin, and Dr. Stuart Tessmer.

Dr. Constantine Trahiotis provided an important recommendation in a discussion about accounting for in-ear distortion products for AM tones.

Mr. Zach Ryan contributed important technical work in the early phases of the non-monotonic interaural level difference experiment. Similarly, Mr. Thomas Andrews contributed to the early development of the amplitude modulation localization experiment.

Former graduate student, Dr. Peter Xinya Zhang, graciously helped to introduce the psychoacoustics laboratories to me. Former graduate student, Dr. Neil Aaronson, shared important words of advice about graduate school and research in psychoacoustics.

I have also had the pleasure of sharing office space and good discussions about research with Dr. Larisa Dunai and Dr. Tianshu Qu.

This work was supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) of the NIH under Grant No. DC 00181, and by the Air Force Office of Scientific Research (AFOSR) under Grant No. 11N002.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	x
KEY TO SYMBOLS	xx
KEY TO ABBREVIATIONS	xxii
Chapter 0 Introduction	1
Chapter 1 Non-monotonic Interaural Level Difference	3
1.1 Introduction	3
1.2 General Methods	8
1.3 KEMAR Recordings	12
1.3.1 KEMAR Results	13
1.4 Experiments, Results, and Discussion	14
1.4.1 Experiment 1: Identification Experiment	14
1.4.1.1 Identification Experiment Results	17
1.4.2 Experiment 2: Discrimination Experiment	29
1.4.2.1 Discrimination Experiment Results	30
1.4.3 Experiment 3: Identification with Training and Feedback Experiment	33
1.4.3.1 Identification with Training and Feedback Results	34
1.4.4 Experiment 4: Identification with Amplitude Modulation Experiment	43
1.4.4.1 Identification with Amplitude Modulation Results	43
1.4.5 Experiment 5: Identification with Narrow-band Noise Experiment	46
1.4.5.1 Identification with Narrow-band Noise Results	47
1.5 The effect of band-width on non-monotonicity	49
1.5.0.2 The effect of band-width on non-monotonicity results	50
1.6 Minimum Audible Angle	58
1.7 Conclusion	60
Chapter 2 Amplitude Modulation Localization	62
2.1 Introduction	62
2.1.1 Problem 1: AM Quality	68
2.1.2 Problem 2: EITD Group Delay	73
2.1.2.1 Problem 2: KEMAR Group Delay	73
2.2 Methods	77
2.2.1 Experimental Setup	77
2.2.2 Experimental Conditions and Procedure	77
2.2.3 Listeners	79

2.2.4	Computer Analysis of Signals	79
2.3	Results and Analysis	82
2.3.1	Physical Results	82
2.3.1.1	Problem 1: AM Quality Results	82
2.3.1.2	Problem 2: Envelope ITD Results	85
2.3.1.3	KEMAR Headphones Measurements	86
2.3.2	Perceptual Results	88
2.3.2.1	Interaural Cues and Responses	88
2.3.2.2	SAM and sine differences	114
2.3.2.3	Compressive Cues	120
2.3.2.4	SAM interaural cue reliability	125
2.3.3	Discussion	125
2.3.4	AM Quality (Problem 1)	133
2.3.4.1	Change in response vs. AM quality	133
2.3.4.2	The effect of AM quality on cue strength	139
2.3.4.3	Response error ratio vs. AM quality	159
2.3.4.4	Discussion	160
2.4	Conclusion	164
APPENDIX		166
BIBLIOGRAPHY		174

LIST OF TABLES

Table 1.1:	Independent ear strategy success rates, and correlations between listener responses and perfect responses.	27
Table 1.2:	The Pearson product-moment correlation coefficients for various sets of data for each listener, as well as the mean coefficients across all of the listeners. NTF1 refers to the first five passes of the experiment with no training and feedback. NTF2 refers to the second five passes of the experiment with no training and feedback. TF1 stands for the first five passes of the experiment with training and feedback. TF2 refers to the second five passes of the experiment with training and feedback. For each of these half experiments—which consist of five passes through the loudspeaker array—the mean response for each loudspeaker was used in the correlation calculations.	34
Table 1.3:	The Pearson product-moment correlation coefficients for various sets of data for each listener, as well as the mean coefficients across all of the listeners. Abbreviations are the same as those explained in the caption for Table 1.2. “Perfect” refers to a hypothetical listener who always gives the correct response to a stimulus.	40
Table 1.4:	Pearson product-moment correlation coefficients for all ten passes through the loudspeaker array. Each listener’s coefficients are shown for various sets of data as well as the mean coefficients across all of the listeners.	41
Table 1.5:	Pearson product-moment correlation coefficients for various experiments. The coefficients are calculated for each listener, and the mean coefficients are averaged over the five listeners. Equation (1.13) is used to calculate the coefficients, except that for a given listener and experiment, the data used in the correlation calculation were the mean responses of the ten passes through each loudspeaker. “Pure Tone” refers to experiment 1, “Amplitude Modulation” refers to experiment 4, and “Perfect” refers to a hypothetical listener who always gives the correct response to a stimulus.	45

Table 1.6:	Pearson product-moment correlation coefficients for various experiments. The coefficients are calculated for each listener, and the mean coefficients are averaged over the five listeners. Equation (1.13) is used to calculate the coefficients, except that for a given listener and experiment, the data used in the correlation calculation were the mean responses of the ten passes through each loudspeaker. “Noise” refers the experiment with narrow-band noise, “Pure Tone” refers to experiment 1, “Amplitude Modulation” refers to experiment 4, and “Perfect” refers to a hypothetical listener who always gives the correct response to a stimulus.	47
Table 2.1:	The percentage of amplitude modulation, m , means $> 100\%$. The means are calculated across identical interval conditions. The data are combined across all listeners and azimuths.	84
Table 2.2:	Values of m , β , and m_e calculated from recordings of KEMAR manikin (large pinna) wearing headphones (Sennheiser HD 535). The signals were 100% SAM tones with carrier frequencies of 2, 3, and 4 kHz, and a modulation frequency of 100 Hz. The monaural parameters were calculated using matched filtering.	87
Table 2.3:	Fisher’s method p-values. This meta-analysis uses a one-sided χ^2 test with $2n$ degrees of freedom on the summation over the logarithms of the t-test p-values for the loudspeakers, where n is the number of t-test p-values in the summation. Results are show for each frequency and listener over the 13 loudspeakers, for each frequency across all listeners and loudspeakers, and finally across all frequencies, listeners, and loudspeakers.	109
Table 2.4:	The correlation coefficients, r , for Figs. 2.44–2.46 and four other tests. The data include all listeners and all frequencies. For each loudspeaker, the data are the change in mean responses (AM–sine) vs. the mean independent variable. For m -left and m -right folded over 1, values of $m < 1$ are transformed to values of $m > 1$ while keeping $ m - 1 $ the same.	138
Table 2.5:	Pearson product-moments, r , for the Δ response residuals from best-fit lines vs. the AM quality metric shown by the color scale in Figs. 2.47–2.61. The correlations are calculated for each listener, frequency, and AM quality metric.	159

Table .1:	Measurements for SAM tone in the reverberation room at a modulation frequency of 100 Hz. For each angle, location and distance (d), quantities shown are the AM in the left ear (m_{left}), the AM in the right ear (m_{right}), the QFM in the left ear (β_{left}), the QFM in the right ear (β_{right}), the envelope interaural time difference (EITD) in μs , the interaural envelope coherence (E. Coh.), and the interaural level difference (ILD) in dB.	170
Table .2:	Same as Table .1 but for the reverberation room at 500 Hz.	171
Table .3:	Same is Table .1 but for the the “laboratory” at 100 Hz.	172
Table .4:	Same is Table .1 but for the the “laboratory” at 500 Hz.	173

LIST OF FIGURES

Figure 1.1:	ILD vs. azimuth for the spherical head model with a head radius of 8.75 cm. At frequencies greater than about 1000 Hz, the curves become seriously non-monotonic.	5
Figure 1.2:	The level vs. azimuth for the left and right ears relative to the level that would be at the center of the head if no head were present. The ILD is plotted vs. azimuth. They are shown for the spherical head model with a frequency of 1500 Hz and a head radius of 8.75 cm. The ILD is shown to be non-monotonic because the level at the left ear is non-monotonic. This is because the level at the left ear increases after approximately 50° , which is due to the acoustical bright spot. .	6
Figure 1.3:	Relationship between the listener's orientation and the array of loudspeakers. The 13 loudspeakers, numbered 0–12, span 90° . The angular spacing between loudspeakers is 7.5°	9
Figure 1.4:	Levels, ILD, and IPD vs. azimuth for a KEMAR manikin. KEMAR's own internal microphones and electronics were used for these recordings. The interaural graphs are polygons, where the mean \pm the standard deviation are plotted for each loudspeaker, and straight lines are drawn between the points. However, the standard deviations are so small that no polygon can be seen. There were 10 passes through the loudspeaker array. Since each loudspeaker is presented twice in a single pass, there are 20 measurements for each source number. . . .	15
Figure 1.5:	Levels, ILD, and IPD vs. azimuth for a KEMAR manikin. Probe microphones were inserted into the KEMAR's ear canals and were used to make these recordings. The interaural graphs are polygons, where the mean \pm the standard deviations are plotted for each loudspeaker, and straight lines are drawn between the points. There were 10 passes through the loudspeaker array. Since each loudspeaker is presented twice in a single pass, there are 20 measurements for each source number.	16

Figure 1.6:	Results for the identification experiment for listener B. Panel (a) shows listener responses for each trial. Response number is plotted vs. source number or azimuth. Panel (b) shows the mean level for the near (right) and far (left) ears. The levels have been set to 0 dB at 0°. Error bars show \pm the standard deviation over the 20 measurements at each angle. In panel (c), the data points show the mean response number for each source number, where the error bars show \pm the standard deviation over the 10 responses at each source number. The shaded polygon represents the mean \pm standard deviation of the ILD for each source number. The correlation coefficient (CC) between the two sets of data is 0.952 for listener B. Panel (d) shows the same listener response data as panel (c) and also shows the ITD mean and standard deviation, which is represented by the shaded polygon. The correlation coefficient (CC) between the two sets of data is 0.512. . .	18
Figure 1.7:	Same as for Fig. 1.6 except for listener M and the correlation coefficients for panels (c) and (d) are 0.986 and 0.743, respectively. . . .	19
Figure 1.8:	Same as for Fig. 1.6 except for listener N and the correlation coefficients for panels (c) and (d) are 0.945 and 0.293, respectively. . . .	20
Figure 1.9:	Same as for Fig. 1.6 except for listener E and the correlation coefficients for panels (c) and (d) are 0.991 and 0.443, respectively. . . .	21
Figure 1.10:	Same as for Fig. 1.6 except for listener X and the correlation coefficients for panels (c) and (d) are 0.920 and 0.740, respectively. . . .	22
Figure 1.11:	Speaker Response vs. ILD. Each data point represents a particular loudspeaker for a given listener. Each listener's speaker response is plotted versus the mean ILD for a given loudspeaker. The error bars show plus and minus the standard deviation for speaker response. The standard deviations for ILD are not shown, but are generally on the order of about 1 dB. The two curves are given by equation 1 where the upper curve is fit to these data with a parameter of 5.95, and the lower curve is fit to Yost's 1981 experiment with a parameter of 8.	23
Figure 1.12:	Discrimination Experiment. The predicted percent correct and actual percent correct are plotted against each other for all five listeners. The percent correct was predicted based on the data from individual listeners in the identification experiment.	31

Figure 1.13:	The percent of correct responses vs. the average difference in ILD between the speaker pairs. The difference in ILD is taken to be the right speaker's ILD minus the left speaker's ILD. The standard deviations in the Δ ILDs are typically on the order of 1 dB or less. .	32
Figure 1.14:	Loudspeaker response numbers for listener B during the training and feedback experiment. There are ten responses per source. The circles represent responses for the first five passes through the 13-loudspeaker array, and the triangles represent responses for the second five passes through the 13-loudspeaker array. The data points have been jogged slightly off of the grid lines so that they may all be visible.	35
Figure 1.15:	Same as Fig. 1.14 except for listener E.	36
Figure 1.16:	Same as Fig. 1.14 except for listener M.	37
Figure 1.17:	Same as Fig. 1.14 except for listener N.	38
Figure 1.18:	Same as Fig. 1.14 except for listener X.	39
Figure 1.19:	Each listener's mean response for each source number in the amplitude modulation experiment.	44
Figure 1.20:	Each listener's mean response for each source number for the narrow-band noise experiment.	48
Figure 1.21:	KEMAR levels for right, left and ILD with a 1500 Hz sine-tone stimulus. The recordings were made using probe microphones. Levels are shown with respect to speaker zero, and are adjusted for the relative level of the stimulus for each loudspeaker at the center of the array.	51
Figure 1.22:	KEMAR levels for right, left and ILD with a 1/3-octave-band stimulus centered logarithmically around 1500 Hz. The recordings were made using probe microphones. Levels are shown with respect to speaker zero, and are adjusted for the relative level of the stimulus for each loudspeaker at the center of the array.	52
Figure 1.23:	Same as Fig. 1.22 but with a 1-octave band.	53
Figure 1.24:	Same as Fig. 1.22 but with a 2-octave band.	54
Figure 1.25:	Same as Fig. 1.22 but with a 3-octave band.	55
Figure 1.26:	Same as Fig. 1.22 but with a 4-octave band.	56

Figure 1.27:	Average minimum audible angle from Mills, 1958. The five sets of data are for different azimuths.	59
Figure 2.1:	Spectrum of 100% amplitude modulation. The carrier frequency, ω_c , has an amplitude of C . The sidebands, at $\omega_c - \omega_m$ and $\omega_c + \omega_m$, have amplitudes of $C/2$. The phase of the lower sideband is $\phi_c - \phi_a$ and the phase of the upper sideband is $\phi_c + \phi_a$	63
Figure 2.2:	SAM signal with 100% AM. The carrier frequency is 2000 Hz and the modulation frequency is 100 Hz. No QFM is present in this signal. In this example, the signals in the two ears are identical.	69
Figure 2.3:	Left and right waveforms of arbitrary units calculated from ear canal recordings. The carrier frequency was 2000 Hz and the modulation frequency was 100 Hz. Values for the amplitude modulation, QFM, and envelope modulation fraction are shown on the right by m , β , and m_e , respectively.	71
Figure 2.4:	Left and right waveforms of arbitrary units calculated from ear canal recordings.. The carrier frequency was 3000 Hz and the modulation frequency was 100 Hz. Values for the amplitude modulation, QFM, and envelope modulation fraction are shown on the right by m , β , and m_e , respectively.	72
Figure 2.5:	Wrapped IPD (-180° to $+180^\circ$) vs. carrier frequency. Measurements were made in the anechoic room with the 1-m array. The azimuth was 60° . The signals were sine tones at 10 Hz intervals.	74
Figure 2.6:	Left (blue) and right (red) waveforms and envelopes of KEMAR recordings of SAM tones in free field. The signal was a 2240 Hz carrier with a modulation frequency of 40 Hz.	75
Figure 2.7:	Left (blue) and right (red) waveforms and envelopes of KEMAR recordings of SAM tones in free field. The signal was a 2325 Hz carrier with a modulation frequency of 40 Hz.	76
Figure 2.8:	Amplitude modulation, m , means for each loudspeaker, frequency, and listener. Histograms for the left and right ears are displayed separately. Each histogram contains a total of 195 values.	84
Figure 2.9:	QFM, β , means for each loudspeaker, frequency, and listener. Histograms for the left and right ears are displayed separately. Each histogram contains a total of 195 values.	85

Figure 2.10:	Histogram of Envelope ITD means for each loudspeaker, frequency and listener.	86
Figure 2.11:	Responses and probe-microphone measurements for listener B, 2 kHz. The source numbers span the right front quadrant. The Pearson product-moment (PPM) correlation coefficient is shown between the mean responses and the interaural cues. (a) Sine tone: Circles indicate mean responses and error bars are two standard deviations in overall length. The hatched region shows the ILD. It is centered on the mean and is two standard deviations high. (b) SAM tone: Circles indicate mean responses and error bars are two standard deviations in overall length. The hatched region shows the ILD. It is centered on the mean and is two standard deviations high. (c) SAM tone: Circles indicate mean responses and error bars are two standard deviations in overall length. The hatched region shows the EITD. It is centered on the mean and is two standard deviations high.	90
Figure 2.12:	Same as Fig. 2.11 but for listener B at 3 kHz.	91
Figure 2.13:	Same as Fig. 2.11 but for listener B at 4 kHz.	92
Figure 2.14:	Same as Fig. 2.11 but for listener C at 2 kHz.	93
Figure 2.15:	Same as Fig. 2.11 but for listener C at 3 kHz.	94
Figure 2.16:	Same as Fig. 2.11 but for listener C at 4 kHz.	95
Figure 2.17:	Same as Fig. 2.11 but for listener M at 2 kHz.	96
Figure 2.18:	Same as Fig. 2.11 but for listener M at 3 kHz.	97
Figure 2.19:	Same as Fig. 2.11 but for listener M at 4 kHz.	98
Figure 2.20:	Same as Fig. 2.11 but for listener L at 2 kHz.	99
Figure 2.21:	Same as Fig. 2.11 but for listener L at 3 kHz.	100
Figure 2.22:	Same as Fig. 2.11 but for listener L at 4 kHz.	101
Figure 2.23:	Same as Fig. 2.11 but for listener V at 2 kHz.	102
Figure 2.24:	Same as Fig. 2.11 but for listener V at 3 kHz.	103
Figure 2.25:	Same as Fig. 2.11 but for listener V at 4 kHz.	104

Figure 2.26:	p-value vs. speaker number for changes in response at 2 kHz. Each listener is plotted according to the symbols in the legend. The p-values are based on a two-tailed t-test and represent the probability that the responses in the sine tone and AM trials are statistically similar. A reference line is shown for $\alpha = 0.05$	106
Figure 2.27:	Same as Fig. 2.26 but for 3 kHz.	107
Figure 2.28:	Same as Fig. 2.26 but for 4 kHz.	108
Figure 2.29:	Pearson product-moment (PPM) correlation coefficients for perfect responses (source number) and actual responses are shown for AM in red and sine in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.	111
Figure 2.30:	Pearson product-moment (PPM) correlation coefficients for AM responses and AM ILD are in red. Correlations for sine responses and sine ILD are in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.	112
Figure 2.31:	Pearson product-moment (PPM) correlation coefficients for AM ILD and EITD are in red. Correlations for AM ILD and sine ILD are in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.	113
Figure 2.32:	Pearson product-moment (PPM) correlation coefficients for AM responses and AM ILD are in red. Correlations for AM responses and EITD are in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.	115

Figure 2.33:	Pearson product-moment (PPM) correlation coefficients for the change in response and the change in ILD are shown in red. Correlation coefficients for the change in responses and the EITD are shown in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.	116
Figure 2.34:	Change in response (AM – sine (mean)) vs. EITD for all negative EITD. All listeners and frequencies at combined. The vertical axis indicates AM responses for individual trials minus the mean sine response for the same loudspeaker. The horizontal axis indicates the EITD measured in the individual AM trials. The correlation coefficient— r —, slope, and y-intercept for the best fit line are above the plots. The color scale indicates the value of the change in ILD between the individual AM and average sine runs. A few of the Δ ILD values clip the top of the color scale at +5dB. A reference line for zero Δ response is shown.	118
Figure 2.35:	Residuals from the best fit line in Fig. 2.34 vs. Δ ILD. A line of best fit is shown as well as reference lines through the origin.	119
Figure 2.36:	FIG. 3 from Yost (1981). In this headphone experiment, listeners lateralized on-going IPD cues at 500 Hz. The different symbols indicate different listeners, and the vertical lines indicate ranges of responses for IPDs of 0° , $\pm 90^\circ$, and 180°	121
Figure 2.37:	Changes in response and compressed interaural cues for listener B. The first row is 2 kHz, the second row is 3 kHz, and the third row is 4 kHz. For each plot, red circles indicate changes in mean response (AM–sine) and error bars are two standard deviations in overall length. For plots (a), (c), and (e), blue triangles indicate the change in compressed ILD (AM–sine). For plots (b), (d), and (f), blue diamonds indicate the compressed EITD. For each plot, the maximized Pearson product-moment (PPM) correlation coefficient between the changes in response and compressed interaural cue is shown, as well as the compression exponent.	126
Figure 2.38:	Same as Fig. 2.37 but for listener C.	127
Figure 2.39:	Same as Fig. 2.37 but for listener M.	128
Figure 2.40:	Same as Fig. 2.37 but for listener L.	129

Figure 2.41:	Same as Fig. 2.37 but for listener V.	130
Figure 2.42:	Maximized Pearson product-moment (PPM) correlation coefficients for the change in AM and sine responses and the change in compressed ILD are shown in red. Correlation coefficients for the change in AM and sine responses and the compressed EITD are shown in blue. The green circles indicate the compression exponent. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.	131
Figure 2.43:	Pearson product-moment (PPM) correlation coefficients for AM tones. Correlations between ILD and source azimuth are in red. Correlations between EITD and source azimuth are in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.	132
Figure 2.44:	Change in response vs. m -left ear. All five listeners are shown here. Each point represents the mean change in response across 10 trials vs. the mean m across 20 intervals for the listener's left (far) ear. The correlation is $r = -0.1031$ and $r^2 = 0.01063$	135
Figure 2.45:	Change in response vs. m -right ear. All five listeners are shown here. Each point represents the mean change in response across 10 trials vs. the mean m across 20 intervals for the listener's right (near) ear. The correlation is $r = -0.0973$ and $r^2 = 0.00947$	136
Figure 2.46:	Change in response vs. interaural envelope coherence. All five listeners are shown here. Each point represents the mean change in response across 10 trials vs. the mean m across 20 intervals. The correlation is $r = -0.0633$ and $r^2 = 0.00440$	137

Figure 2.47:	Changes in response vs. EITD with AM quality for listener B at 2 kHz. All plots display the same data and differ in the AM quality metric shown in color. The vertical axis indicates AM responses for individual trials minus the mean sine response for the same loudspeaker. The horizontal axis indicates the EITD measured in the individual AM trials. The correlation coefficient— r —, slope, and y-intercept for the best fit line are above the plots. The color scale in plot (a) indicates the value of m in the left ear in the individual AM trials. The color scale in plot (b) indicates the value of m in the right ear in the individual AM trials. The color scale in plot (c) indicates the value of the envelope coherence in the in individual AM trials. The standard deviations, σ , of the AM quality are shown above the plots.	141
Figure 2.48:	Same as Fig. 2.47 but for listener B at 3 kHz.	142
Figure 2.49:	Same as Fig. 2.47 but for listener B at 4 kHz.	143
Figure 2.50:	Same as Fig. 2.47 but for listener C at 2 kHz.	144
Figure 2.51:	Same as Fig. 2.47 but for listener C at 3 kHz.	145
Figure 2.52:	Same as Fig. 2.47 but for listener C at 4 kHz.	146
Figure 2.53:	Same as Fig. 2.47 but for listener M at 2 kHz.	148
Figure 2.54:	Same as Fig. 2.47 but for listener M at 3 kHz.	149
Figure 2.55:	Same as Fig. 2.47 but for listener M at 4 kHz.	150
Figure 2.56:	Same as Fig. 2.47 but for listener L at 2 kHz.	151
Figure 2.57:	Same as Fig. 2.47 but for listener L at 3 kHz.	152
Figure 2.58:	Same as Fig. 2.47 but for listener L at 4 kHz.	153
Figure 2.59:	Same as Fig. 2.47 but for listener V at 2 kHz.	154
Figure 2.60:	Same as Fig. 2.47 but for listener V at 3 kHz.	155
Figure 2.61:	Same as Fig. 2.47 but for listener V at 4 kHz.	156
Figure 2.62:	Residuals from the best fit line in Fig. 2.47 vs. m -left. The correlation is $r = -0.064$ and $r^2 = 0.0041$	158

Figure 2.63:	The ratio of the error in response rms for sine tones to AM tones vs. <i>m</i> -left averaged across all speakers. The five listeners and three frequencies are all represented. The correlation is, $r = 0.1160$ and $r^2 = 0.01346$	161
Figure 2.64:	The ratio of the error in response rms for sine tones to AM tones vs. <i>m</i> -right averaged across all speakers. The five listeners and three frequencies are all represented. The correlation is, $r = -0.1459$ and $r^2 = 0.02129$	162
Figure 2.65:	The ratio of the error in response rms for sine tones to AM tones vs. the envelope coherence averaged across all speakers. The five listeners and three frequencies are all represented. The correlation is, $r = -0.0221$ and $r^2 = 0.000488$	163

KEY TO SYMBOLS

A_L : cosine intergral of the left ear

A_R : cosine intergral of the right ear

B_L : sine intergral of the left ear

A_R : sine intergral of the right ear

x_L : recording in the left ear

x_R : recording in the right ear

P_L : “power” in the left ear

P_R : “power” in the right ear

ϕ_L : phase of the left ear

ϕ_R : phase of the right ear

Ψ_{ILD} : response for a given ILD

C : amplitude of the carrier

m : amplitude modulation

ω_m : modulation frequency

ϕ_a : phase of the amplitude modulation

ϕ_c : phase of the carrier frequency

$\Delta\omega$: phase excursion

ϕ_f : phase of the frequency modulation

β : quasi frequency modulation

ω_ℓ : lower sideband

ω_u : upper sideband

A_c : cosine integral of the carrier

B_c : sine integral of the carrier

A_ℓ : cosine integral of the lower sideband

B_ℓ : sine integral of the lower sideband

A_u : cosine integral of the upper sideband

B_u : sine integral of the upper sideband

E : envelope

m_e : envelope modulation fraction

KEY TO ABBREVIATIONS

AM: amplitude modulation

ETID: envelope interaural time difference

FM: frequency modulation

ILD: interaural level difference

IPD: interaural phase difference

ITD: interaural time difference

KEMAR: Knowles Electronics Manikin for Acoustic Research

PPM: Pearson product-moment

SAM: sinusoidal amplitude modulation

QFM: quasi frequency modulation

Chapter 0

Introduction

Sound localization is a remarkable sensory capability of humans and other animals. Unlike vision (retina) or touch (skin) where localization is topical, sound localization needs to be computed. In hearing, topical coding corresponds to frequency. Compared to synaptic delays of 1 ms, the unique and remarkable computational machinery of the auditory system features interaural time difference (ITD) sensitivity at least as small as 20 μs . Sound localization as studied here is part of the topic of binaural hearing, which also includes the topic of binaural detection advantage. Binaural hearing has been studied for 100 years. It acquires new significance because of the increasing use of bilateral cochlear implants.

There are two experimental studies on human sound localization in this thesis. The non-monotonic interaural level difference study is in Chapter 1 and the amplitude modulation localization study is in Chapter 2. In general, the study of human sound localization aims to understand why listeners perceive the location of sources of sound where they do. This is done in three dimensions, and the coordinate system of choice uses the azimuthal angle in the horizontal plane of the listener's two ears, the vertical angle as measured above the horizontal plane, and the distance from the source to the listener. However, the two most important cues for sound localization, the interaural time difference (ITD) and the interaural level difference (ILD) are able to be studied by only considering sound sources that are in the horizontal plane. The studies in this thesis are each restricted to sound sources in this plane.

Additionally, these studies both take place in free field—an environment in which there are no acoustical reflections or reverberation. The advantage of this type of experiment is that the sound field can be controlled. A listener is presented with a stimulus from only one direction in space. The locations of the sound sources are also far enough away from the listener that to a good approximation the wavefronts arriving at the location of the listener are plane waves. The listener’s task is to report only the azimuthal angle—not the distance, and not the elevation.

A notable feature in both of these studies is the extensive use of recordings of the sound pressure waves inside the listener’s ear canals. These recordings are possible due to the use of small probe microphones which are inserted into the listener’s ear canals. These recordings allow for the interaural localization cues—the ILD and the ITD—to be calculated. Although the ILD and ITD are typically useful cues for listeners, due to the interaction of the incident sound waves with a listener’s anatomy, they can sometimes have misleading qualities. Chapter 1 studies a situation in which the interaural level difference cue is misleading, and Chapter 2 studies a situation in which the interaural time difference in the envelope of a sound wave is misleading, or in conflict with the the ILD.

Chapter 1

Non-monotonic Interaural Level Difference

1.1 Introduction

Human sound localization relies on two primary interaural cues. These are the interaural time difference (ITD), and the interaural level difference (ILD). The interaural time difference is the elapsed time between the arrival of a sound wave at a listener's two ears. Humans are only sensitive to fine-structure ITDs for frequencies lower than about 1500 Hz. The interaural level difference, measured in decibels (dB), is the difference in sound pressure level at a listener's two ears. This difference is caused mostly by the acoustical shadow that the head casts in the ear that is on the opposite side of a listener's head from a sound source. The ILD tends to be more useful at high frequencies, where the wavelength is small relative to the listener's head size.

Can a listener localize a pure sine tone in an anechoic environment using only interaural level difference cues? Consider the azimuthal plane for a listener and let a source in front of the listener define zero degrees and a source to the right of the listener define $+90$ degrees. If the ILD (right level minus left level) rises monotonically as a function of this angle, then it is possible that the listener can use ILD cue to localize a sine tone. However, if ILD vs. angle is a non-monotonic function, then localization via the ILD would be problematic because

more than one azimuthal location could have the same ILD.

Consider a widely-used spherical model of the head with a radius of 8.75 cm [15], and antipodal ears. Also assume that the spherical head is hard (no sound absorption). The most accessible spherical head model was proposed by Kuhn [27, 28]. Kuhn got the formula from Rschevkin [44], and Rschevkin built on the work of Morse [38] and Rayleigh [47]. This model solves the wave equation for a free-field plane wave under the boundary condition of a rigid sphere. Therefore, the solution exhibits all of the properties of diffraction and interference. Figure 1.1 shows how the ILD behaves as a function of azimuth for several frequencies. The azimuth is defined as the incident angle of the incoming plane wave where the ears are located at $\pm 90^\circ$ from the front. For frequencies greater than approximately 1000 Hz, the curves become significantly non-monotonic [29]. The complexity also increases with frequency, as more local extrema appear at higher frequencies. For the sake of simplicity, consider the 1500-Hz ILD curve. Here, there is only one extremum, which appears at approximately 50° with an ILD of about 8 dB.

Why does the 1500 Hz curve behave this way? To answer this question one needs to examine the individual levels at the left and right ears. Figure 1.2 shows the calculated individual levels of the left and right ears and the difference between the two as a function of azimuth for a sound source on the right side. The level at the right ear (the ear near the source) slowly and monotonically increases by about 3 dB over the span of 90° . The level at the far (left) ear does not vary monotonically. The level decreases until about 50° , where it then begins to increase. The increase in level is the result of the acoustical analog to the optical bright spot, or Arago spot [22]. As the source of the incident plane wave approaches the right side of the head, the sound waves diffract around the head and recombine at the left ear increasingly in phase. As a result the intensity at the left ear increases.

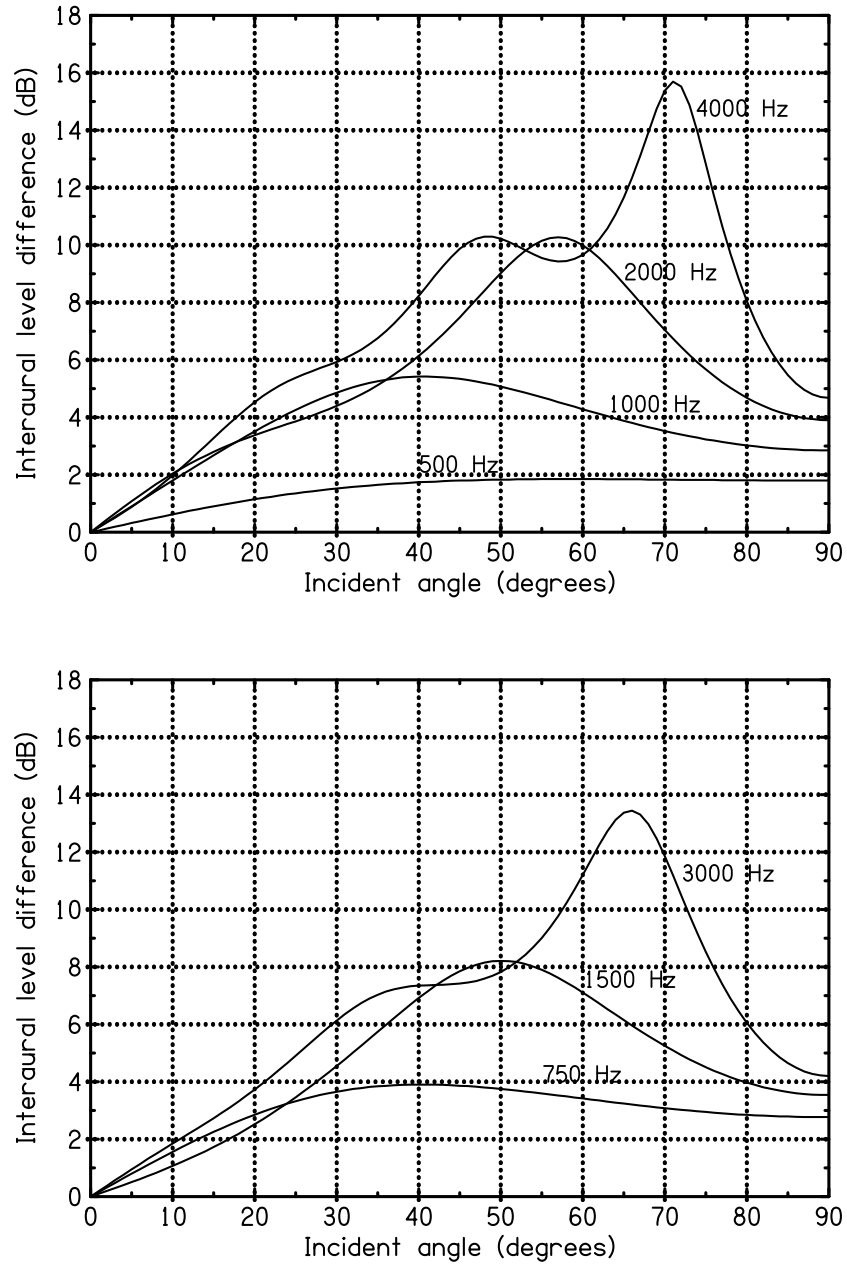


Figure 1.1 ILD vs. azimuth for the spherical head model with a head radius of 8.75 cm. At frequencies greater than about 1000 Hz, the curves become seriously non-monotonic.

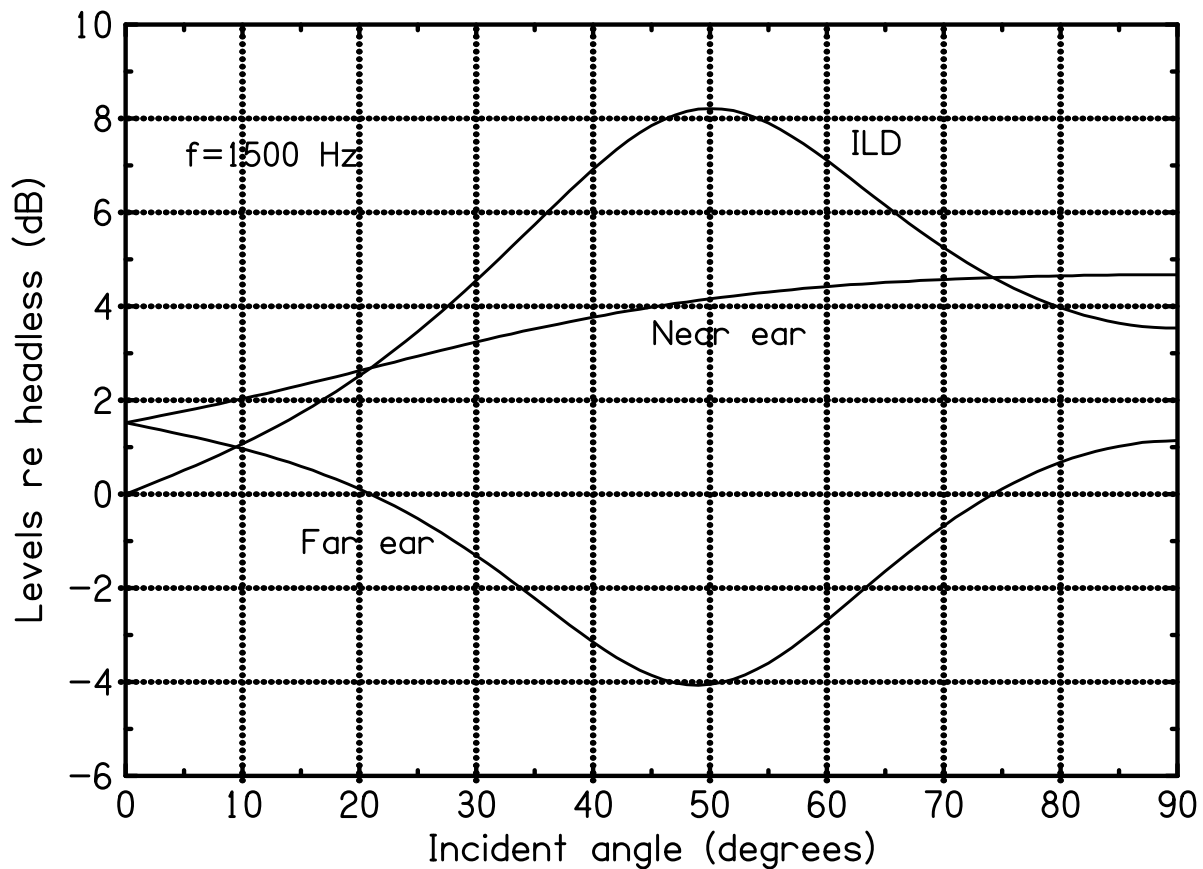


Figure 1.2 The level vs. azimuth for the left and right ears relative to the level that would be at the center of the head if no head were present. The ILD is plotted vs. azimuth. They are shown for the spherical head model with a frequency of 1500 Hz and a head radius of 8.75 cm. The ILD is shown to be non-monotonic because the level at the left ear is non-monotonic. This is because the level at the left ear increases after approximately 50° , which is due to the acoustical bright spot.

It was speculated that human heads would exhibit similar non-monotonic ILD curves. If so, do listeners rely solely on ILD cues to localize these stimuli, or can they take advantage of the individual levels at the near and far ears? If they could use individual levels at the near and far ears, then they would successfully use all of the information available to them, and not simply pay attention to the ILD. The question is: Do human listeners successfully use all of the information available to them in the individual levels, or do they simply pay attention to the ILD? A listener that can take advantage of individual levels may be able to successfully localize a sine tone in an anechoic room. A listener that only relies on ILD cues will likely become confused in regions where the slope of the ILD curve is negative, and will also likely be confused by multiple loudspeakers that share similar ILDs. It is expected that these listeners would give incorrect and/or ambiguous responses to identification experiments.

In the duplex theory of sound localization, first proposed by Lord Rayleigh [48], the interaural level difference (ILD) and the interaural time difference (ITD) both contribute to the perception of the location of a sound source, but contribute differently at low and high frequencies [32]. At low frequencies—roughly less than 500 Hz—the ILD cue is small, and the ITD cue dominates in localization [20, 49]. At these lower frequencies, the wavelength is much larger than the size of the human head, so the sound easily diffracts around the head, and both ears tend to be exposed to sounds of a very similar level. However, at 1500 Hz and higher, the ITD sensitivity disappears entirely [7, 39, 52], and only the ILD cue remains. As the wavelength becomes much smaller than the size of the human head and the IPD π -limit is passed [17], it becomes impossible to determine which wave cycle arriving in one ear corresponds to the same cycle in the other ear, as there are many cycles in between. At intermediate frequencies—roughly 500 to 1500 Hz—both cues play important roles.

1.2 General Methods

All experiments were performed in the following manner. The listener was seated in an anechoic room (Industrial Acoustics Company 107840, with dimensions 3 m wide by 4.3 m long by 2.4 m high) with a cutoff frequency of less than 100 Hz. An array of 13 loudspeakers (Radio Shack Minimus 3.5, consisting of a 2.5-in driver in a sealed box) was placed approximately in the azimuthal plane of the listener’s head, which is defined by the two ears and the nose. The 13 loudspeakers (numbered 0–12) spanned the right-front quadrant of the azimuthal plane, with speaker number zero at an angle of zero degrees in front of the listener and speaker number 12 at an angle of 90 degrees to the right of the listener. The angular spacing between the speakers was 7.5 degrees, and the grilles of the speakers were located 111.5 cm from the listener forming one-fourth of a circle, as shown in Fig 1.3. The listener sat in a chair equidistant from the loudspeakers and was instructed to face speaker zero and keep his body and head still for the duration of the experiment. A metal rod located just on top of the listener’s head served as a reference so that the listener could sense head motion and minimize it.

A Tucker Davis System II with DD1 digital-analog and analog-digital converter was used to produce all waveforms and was used to make all recordings. The sample rate was 50 kHz with 16 bits/sample.

The stimulus was a 1500-Hz sine tone and was presented by only one speaker at a time. Each tone had an overall duration of 1000 ms, where the level increased to its maximum in the first 250 ms and decreased during the last 250 ms. Raised cosine envelopes were used for the rise and fall. This long rise/fall time is sufficient to rule out the use of ITD (interaural time difference) cues from the onset of the signal [42].

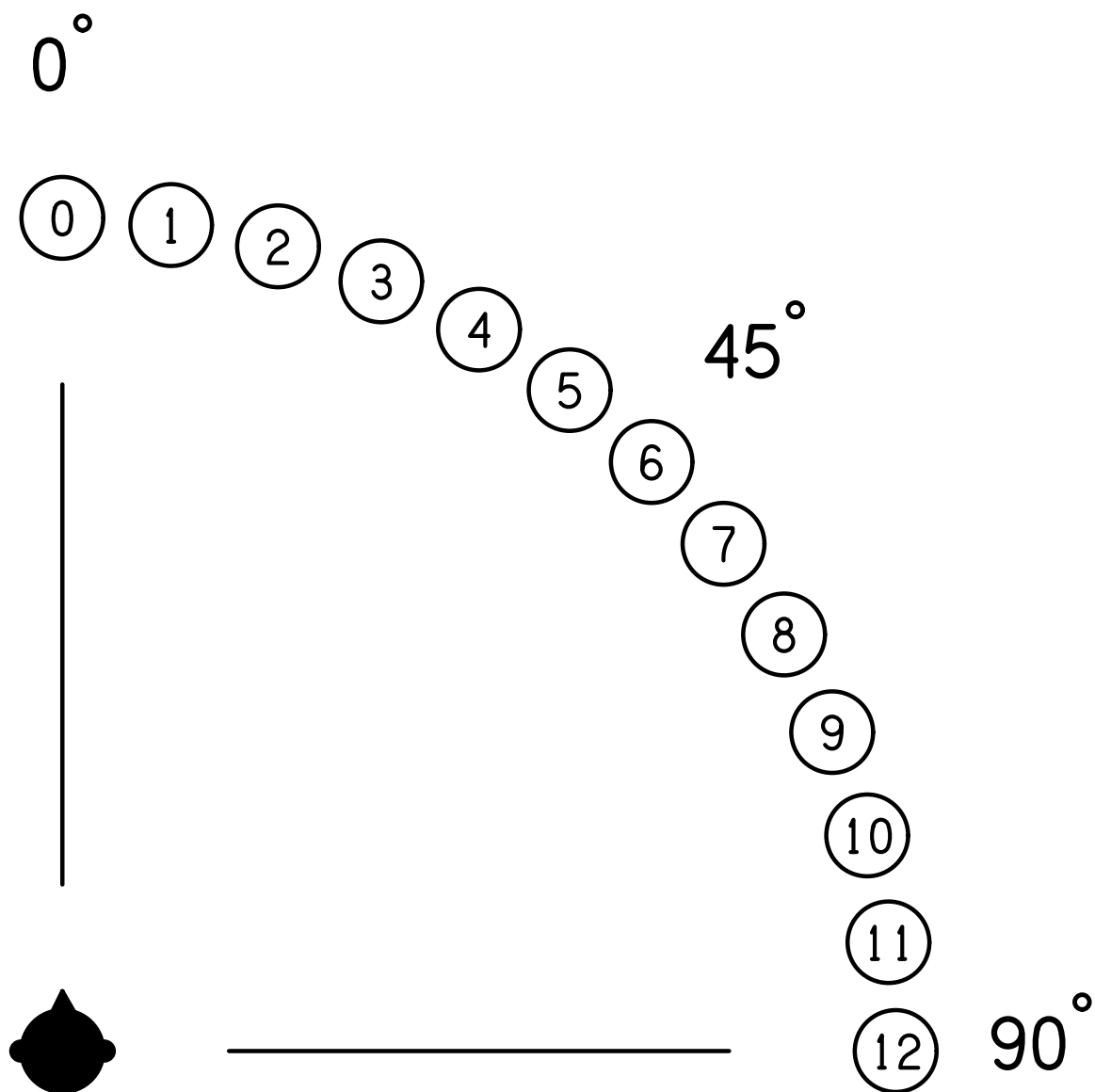


Figure 1.3 Relationship between the listener's orientation and the array of loudspeakers. The 13 loudspeakers, numbered 0–12, span 90° . The angular spacing between loudspeakers is 7.5° .

A 1500-Hz sine signal was employed in the experiment because if a higher frequency were used instead, then as shown in Fig. 1.1, the peak of the ILD curve would move closer toward 90 degrees, and therefore there would be less data past the peak and fewer opportunities to observe the non-monotonic effect. Additionally, the ILD curve at 1500 Hz tends to simply increase and then decrease, whereas at higher frequencies more ripples and complexities tend to appear in the curve. If a lower frequency were used instead there is a possibility that a listener can use ITD in the continuous waveform fine structure as a cue. It is generally believed that frequencies greater than about 1300 Hz are high enough to rule this out [52]. (As will be described later, the experimental procedure checked for localization using ITD cues).

For the entire duration of each stimulus, recordings were made inside both of the listener's ear canals using Etymotic probe microphones (ER-7C). The probe microphone signals were amplified by matched pre-amplifiers and then by a second pre-amplifier (Audio Buddy) to voltages of a few volts. Each recording was filtered in the controlling program with a very narrow sine-cosine (matched) filter—very narrow based on one-second recordings at 1500 Hz—at the same frequency as the stimulus. The matched filter produces two numbers for each ear (L and R) where the A 's refer to the cosine Fourier integrals and the B 's refer to the sine Fourier integrals in equations (1.1)–(1.4).

$$A_L = \frac{2}{T} \int_0^T x_L(t) \cos \omega t dt \quad (1.1)$$

$$B_L = \frac{2}{T} \int_0^T x_L(t) \sin \omega t dt \quad (1.2)$$

$$A_R = \frac{2}{T} \int_0^T x_R(t) \cos \omega t dt \quad (1.3)$$

$$B_R = \frac{2}{T} \int_0^T x_R(t) \sin \omega t dt \quad (1.4)$$

The duration of the recording is T , $x_L(t)$ and $x_R(t)$ are the recordings in the left and right ears, respectively, and ω is the angular frequency of the tone. The “power”, P_L and P_R , in the left and right ears, respectively, were then calculated as shown in equations (1.5) and (1.6).

$$P_L = A_L^2 + B_L^2 \quad (1.5)$$

$$P_R = A_R^2 + B_R^2 \quad (1.6)$$

Then the ILD is simply

$$\text{ILD} = 10 \log \left(\frac{P_R}{P_L} \right). \quad (1.7)$$

The phases, ϕ_L and ϕ_R , were each calculated from the A ’s and B ’s in equations (1.1)–(1.4) as

$$\phi = \begin{cases} \tan^{-1} \left(\frac{B}{A} \right) & A \geq 0 \\ \tan^{-1} \left(\frac{B}{A} \right) + \pi & A < 0, B \geq 0 \\ \tan^{-1} \left(\frac{B}{A} \right) - \pi & A < 0, B < 0. \end{cases} \quad (1.8)$$

This function is similar to the Arg and atan2 functions and restricts the angle between $(-\pi, \pi]$. The IPD, then, is just the difference between the left and right phases.

$$\text{IPD} = \phi_R - \phi_L \quad (1.9)$$

For sine tones, the ITD is

$$\text{ITD} = \frac{\text{IPD}}{\omega}. \quad (1.10)$$

Before each perception experiment, the levels of the loudspeakers were equalized. This was achieved by placing an omni-directional microphone at the center of the semi-circular array and measuring the level of each speaker for 1500-Hz sine tones. Then, compensations were made in the controlling program, such that every speaker produced the same level at the listener’s location. Overall signal levels for each run were standardized by adjusting the gain on the power amplifier and a VU meter. The absolute level (72 dBA) was chosen so as to be comfortable for the listener, yet sufficiently high to perform the tasks.

There were five male listeners used in the various experiments to follow. Four of the listeners were between the ages of 20 and 35—labeled E, N, and X—and the other listener, labeled B, was 57. All listeners had normal hearing thresholds within 15 dB according to pure-tone audiometry. One listener, M, reported a slight asymmetrical hearing loss with the left ear not as sensitive as the right ear around 2000 Hz. This was observed in the pure tone audiometry, however the listener was still within the 15 dB range.

1.3 KEMAR Recordings

As a check on the experimental methods, acoustical measurements were made using a Knowles Electronics Manikin for Acoustic Research (KEMAR) [8]. KEMAR has a head and torso with dimensions corresponding to the median human adult. It has silicone pinnae, ear canals, and acoustically realistic middle ears (Zwislocki couplers). KEMAR was placed at the center of the loudspeaker array, facing loudspeaker zero. Measurements were made with KEMAR’s internal microphones for 10 passes through the 13 loudspeakers (Etymotic ER11 with pre-amplifiers) and also with the probe microphones in KEMAR’s ear canals. The probe microphones were removed and re-inserted into KEMAR’s ears for 10 passes

through the 13 loudspeakers. This was done to simulate variation in positioning that one might expect in a human subject using probe microphones. The ten passes through the 13 loudspeakers gave 20 recordings per loudspeaker—as each speaker was played twice in one pass. KEMAR’s ILD and IPD curves were calculated for both of the recording techniques. The comparison between KEMAR’s internal microphones and the probe microphones was to show that under the parameters of this experiment, the probe microphones used with human listeners were producing consistent and accurate recordings.

1.3.1 KEMAR Results

It should be noted that in the reports below for both the KEMAR and the human subject results, all levels and phases have been set to zero where the incident angle is zero. Hence, the figures report the change in the interaural cues with azimuthal angle. The recording method did not make absolute measurements of the sound pressure level. This is because the left and right channels each had overall gains which were not necessarily the same as each other. Fortunately, the absolute sound pressure levels are not of much interest. The effective level difference depends only on how the levels change as incident angle changes, and this is what is represented in the following figures.

Figure 1.4 shows the left and right ear levels, ILD, and IPD vs. azimuth for the KEMAR when using KEMAR’s internal microphones and electronics. The ILD and IPD graphs are polygons, where for each data point, the mean \pm the standard deviation are each plotted and straight lines are drawn between data points. The standard deviations are nearly impossible to see on these plots, indicating excellent reproducibility. Figure 1.5 shows the left and right ear levels, ILD, and IPD vs. azimuth for KEMAR when using probe microphones. Once again, these graphs are polygons that show the means \pm the standard deviations. The

standard deviations here are larger than the ones on Fig. 1.4. This is due to the fact that between trials, the probe microphones were removed and re-inserted (ten total insertions). Still, the standard deviations are quite small, which indicates that changing the location of the probe tips within the ear canal does not result in a significant difference when measuring ILD and IPD.

It is notable that there is very good agreement between the two recording techniques for the KEMAR. This strongly suggests that using probe microphones on human subject is a valid way to measure ILD. Additionally, it is notable that the data generally agree with the trends suggested by the spherical head model. Both curves peak at around 60 degrees with an ILD of about 14 dB. (The advantage of studying the effect of varying microphone positions in the KEMAR is that it decouples the microphone’s movement within the ear canals from movement of a listener’s head).

1.4 Experiments, Results, and Discussion

1.4.1 Experiment 1: Identification Experiment

In experiment 1, termed the identification experiment, the listener was asked to listen to the stimulus presented from a loudspeaker, and to indicate the location of the loudspeaker that had sounded by calling out its number over an intercom. A map of the loudspeaker locations was placed in front of the listener so that the listener would not have to turn his head to find the number corresponding to some location in space. The stimulus set consisted of 1500-Hz tones presented in two one-second intervals with 250-ms raised cosine edges. There was a 1.5-second pause between the two intervals. Each listener performed 2 runs. Each run consisted of five passes through the 13 loudspeakers presented in random order, giving 65

Kemar Ears and Electronics

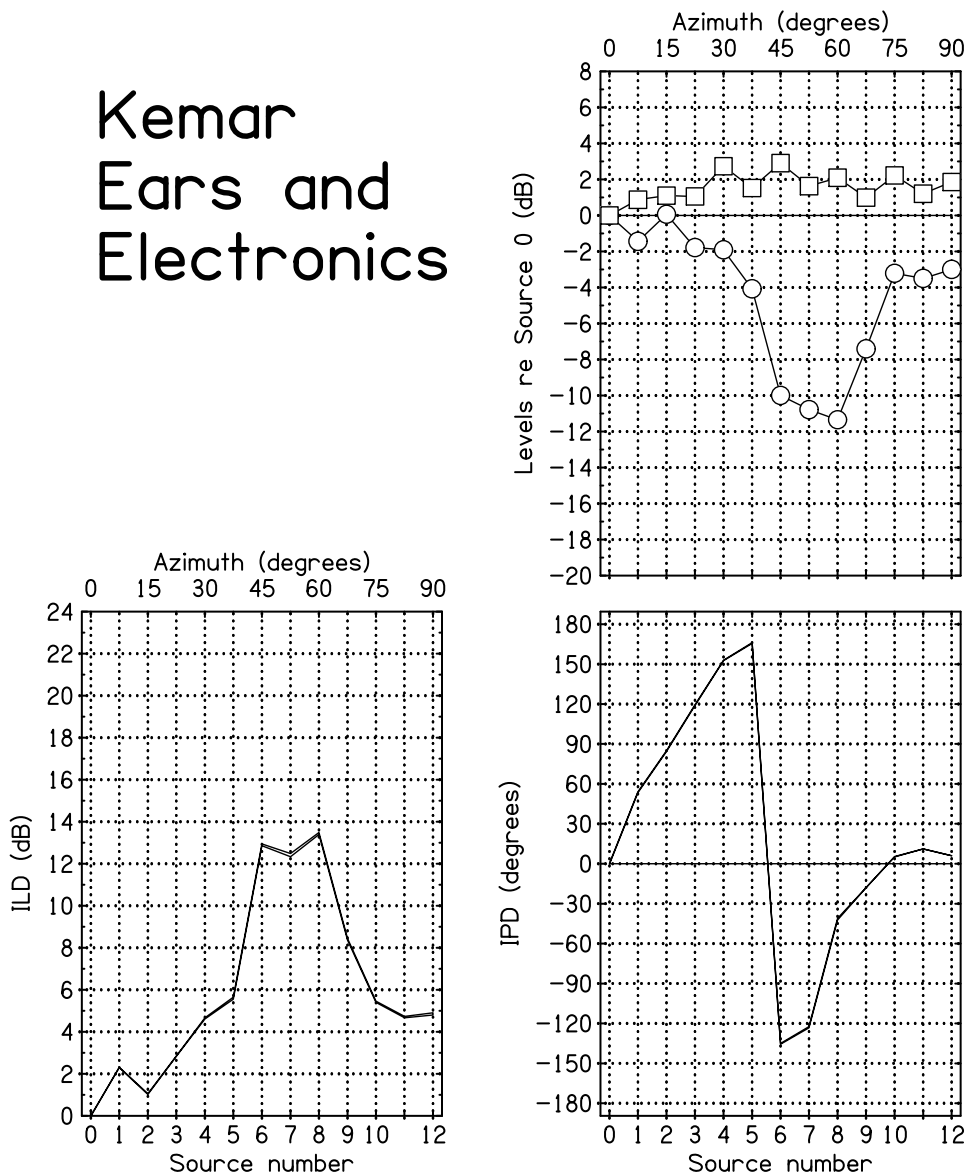


Figure 1.4 Levels, ILD, and IPD vs. azimuth for a KEMAR manikin. KEMAR's own internal microphones and electronics were used for these recordings. The interaural graphs are polygons, where the mean \pm the standard deviation are plotted for each loudspeaker, and straight lines are drawn between the points. However, the standard deviations are so small that no polygon can be seen. There were 10 passes through the loudspeaker array. Since each loudspeaker is presented twice in a single pass, there are 20 measurements for each source number.

Kemar with Probe Mics

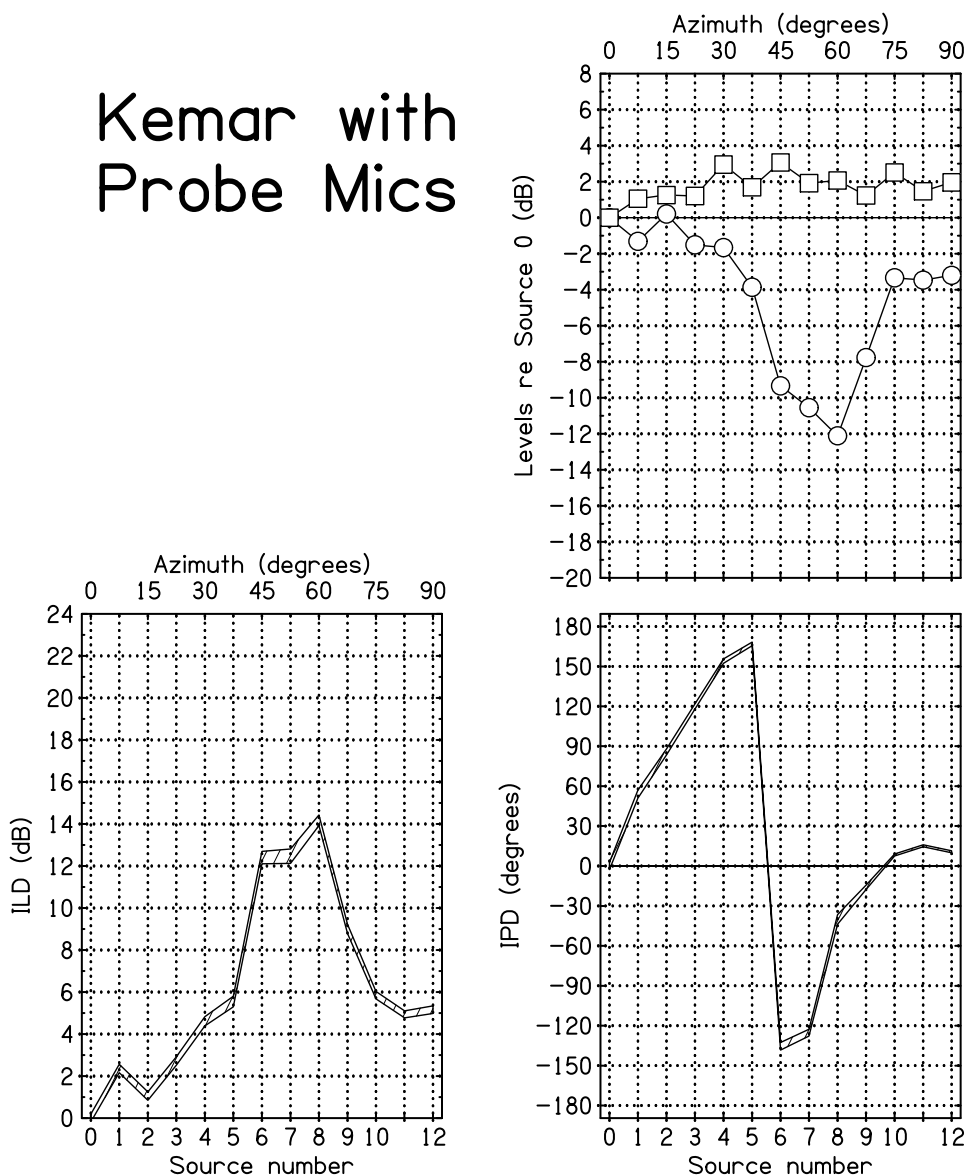


Figure 1.5 Levels, ILD, and IPD vs. azimuth for a KEMAR manikin. Probe microphones were inserted into the KEMAR's ear canals and were used to make these recordings. The interaural graphs are polygons, where the mean \pm the standard deviations are plotted for each loudspeaker, and straight lines are drawn between the points. There were 10 passes through the loudspeaker array. Since each loudspeaker is presented twice in a single pass, there are 20 measurements for each source number.

trials per run, and 130 trials per listener. Each pass of 13 loudspeakers was completed before the next pass began. The probe microphones were used to make recordings in the listener’s ears. Because each trial presented two tones, there were 20 recordings per loudspeaker for each listener. The runs lasted approximately ten to fifteen minutes each.

1.4.1.1 Identification Experiment Results

In Figs. 1.6–1.10, the physical measurements appear on panels (b), (c), and (d), whereas the listener responses appear in panels (a), (c), and (d). Panels (c) and (d) also display the correlation coefficient between the physical measurements and the listener responses. For all five listeners there is a good correlation between the responses and the measured ILD. However, there is a weaker correlation between the responses and the measured ITD. For each listener, the response curve peak tends to occur at the same angle where the ILD curve peaks. This strongly suggests that listeners are localizing based on only ILD, and are easily fooled by speakers to the right of the peak of their ILD curves.

Panel (b) of Figs. 1.6–1.10 shows the levels at each listener’s near and far ears. For listeners M and X, with increasing azimuth there is an increase of about 10 dB, and for listener E, the level becomes negative. Overall there are remarkable differences between individuals for the near ear and equally remarkable similarities for the far ear. This is not a result that one would have anticipated. One might have expected the reverse. The differences for the near ear are true individual differences, not measurement variability.

It is particularly useful to consider the dip in each listener’s far ear level. For the five listeners, this dip ranges from about -10 to -16 dB. The location of the dip is somewhat consistent, however. For listener B and for KEMAR, the dip occurs at speaker 8 (60°), and for the other four listeners, the dip occurs at speaker 7 (52.5°).

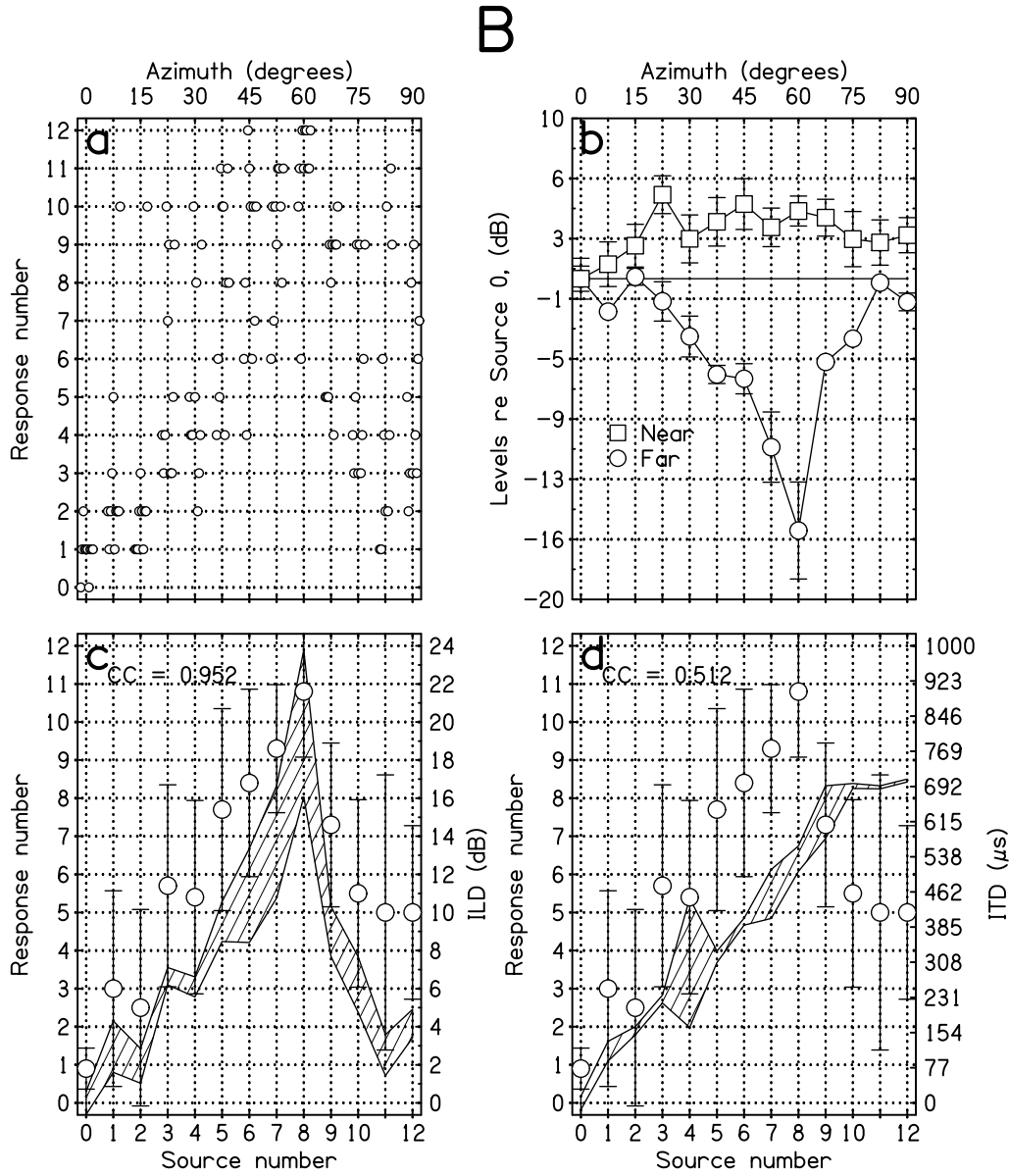


Figure 1.6 Results for the identification experiment for listener B. Panel (a) shows listener responses for each trial. Response number is plotted vs. source number or azimuth. Panel (b) shows the mean level for the near (right) and far (left) ears. The levels have been set to 0 dB at 0°. Error bars show \pm the standard deviation over the 20 measurements at each angle. In panel (c), the data points show the mean response number for each source number, where the error bars show \pm the standard deviation over the 10 responses at each source number. The shaded polygon represents the mean \pm standard deviation of the ILD for each source number. The correlation coefficient (CC) between the two sets of data is 0.952 for listener B. Panel (d) shows the same listener response data as panel (c) and also shows the ITD mean and standard deviation, which is represented by the shaded polygon. The correlation coefficient (CC) between the two sets of data is 0.512.

M

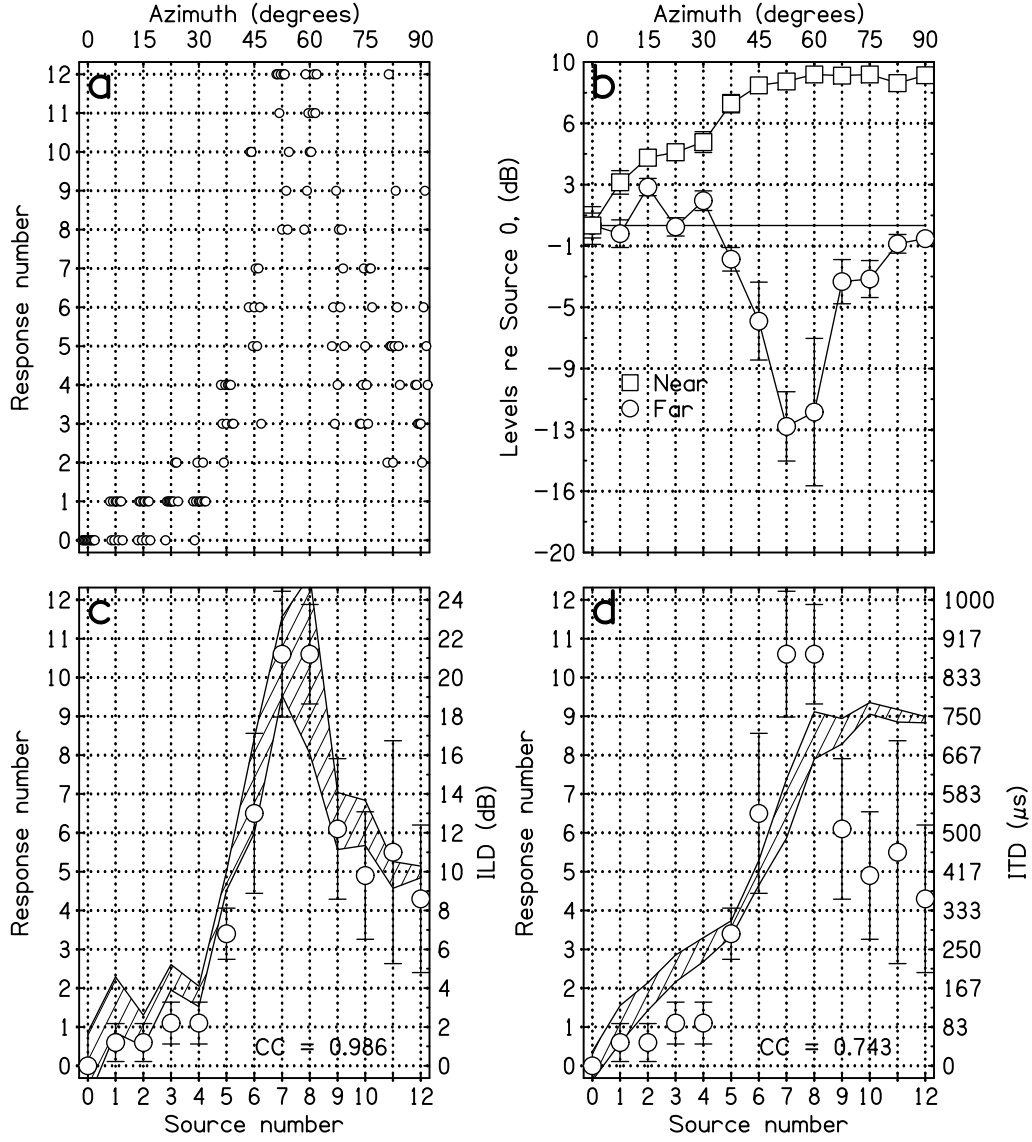


Figure 1.7 Same as for Fig. 1.6 except for listener M and the correlation coefficients for panels (c) and (d) are 0.986 and 0.743, respectively.

N

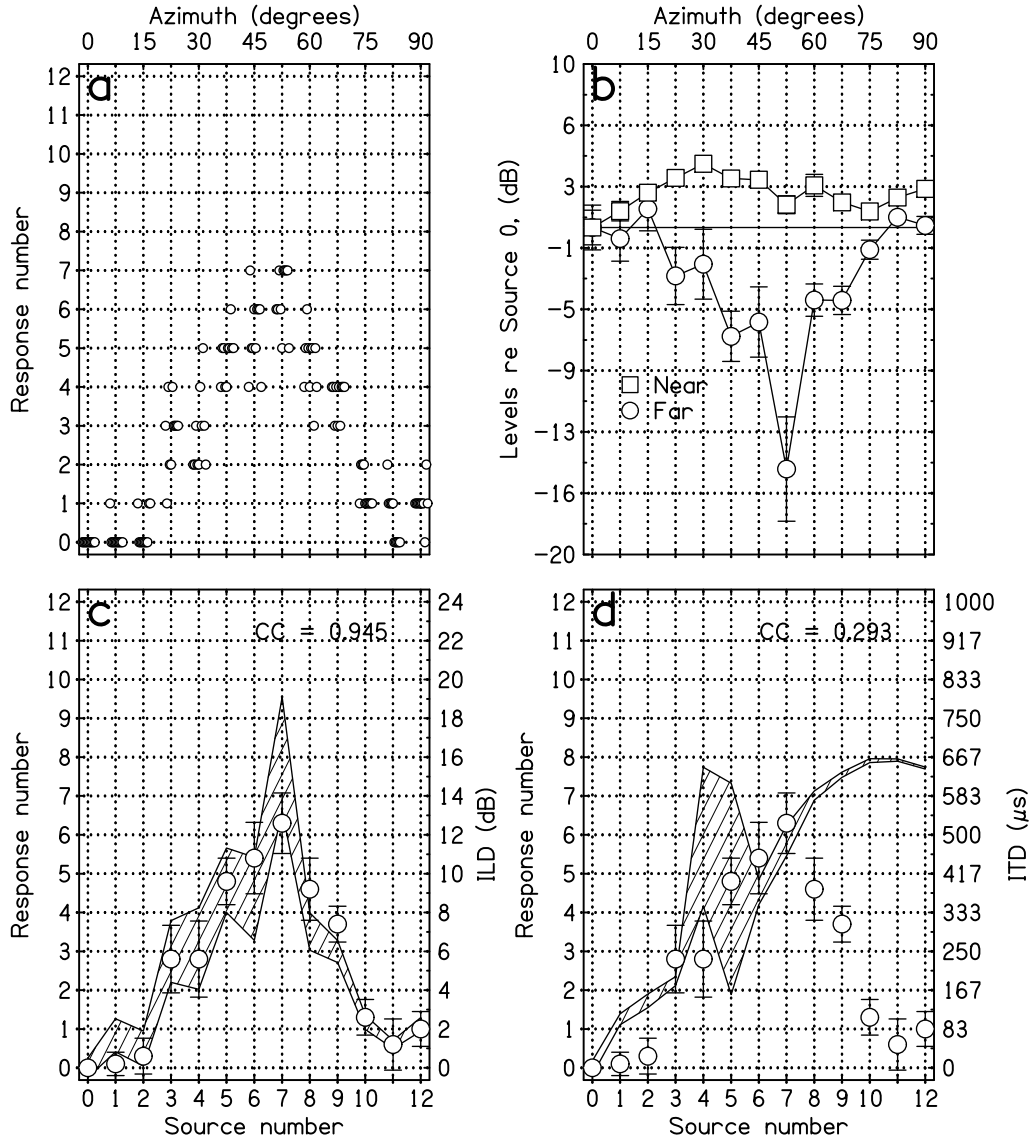


Figure 1.8 Same as for Fig. 1.6 except for listener N and the correlation coefficients for panels (c) and (d) are 0.945 and 0.293, respectively.

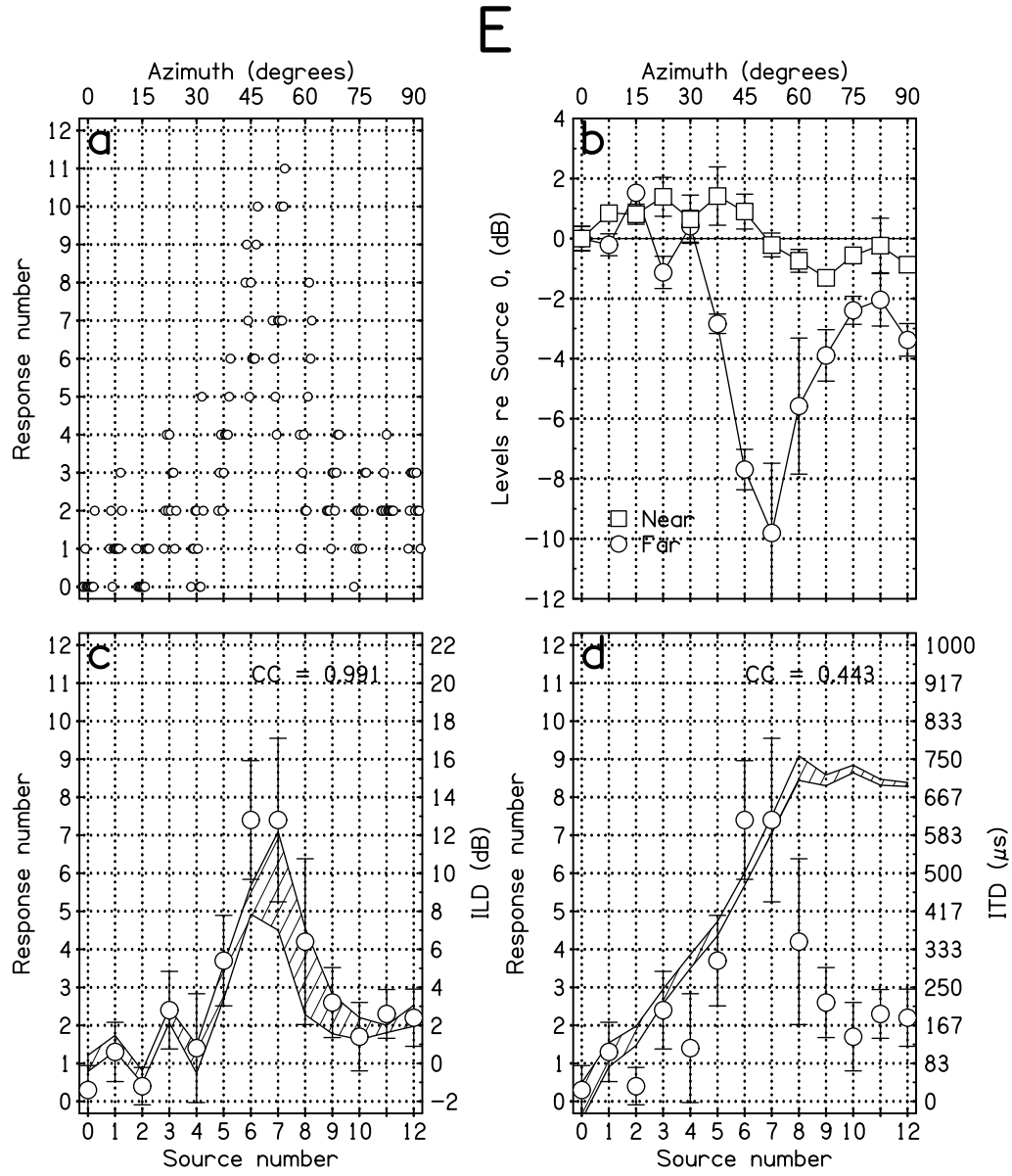


Figure 1.9 Same as for Fig. 1.6 except for listener E and the correlation coefficients for panels (c) and (d) are 0.991 and 0.443, respectively.

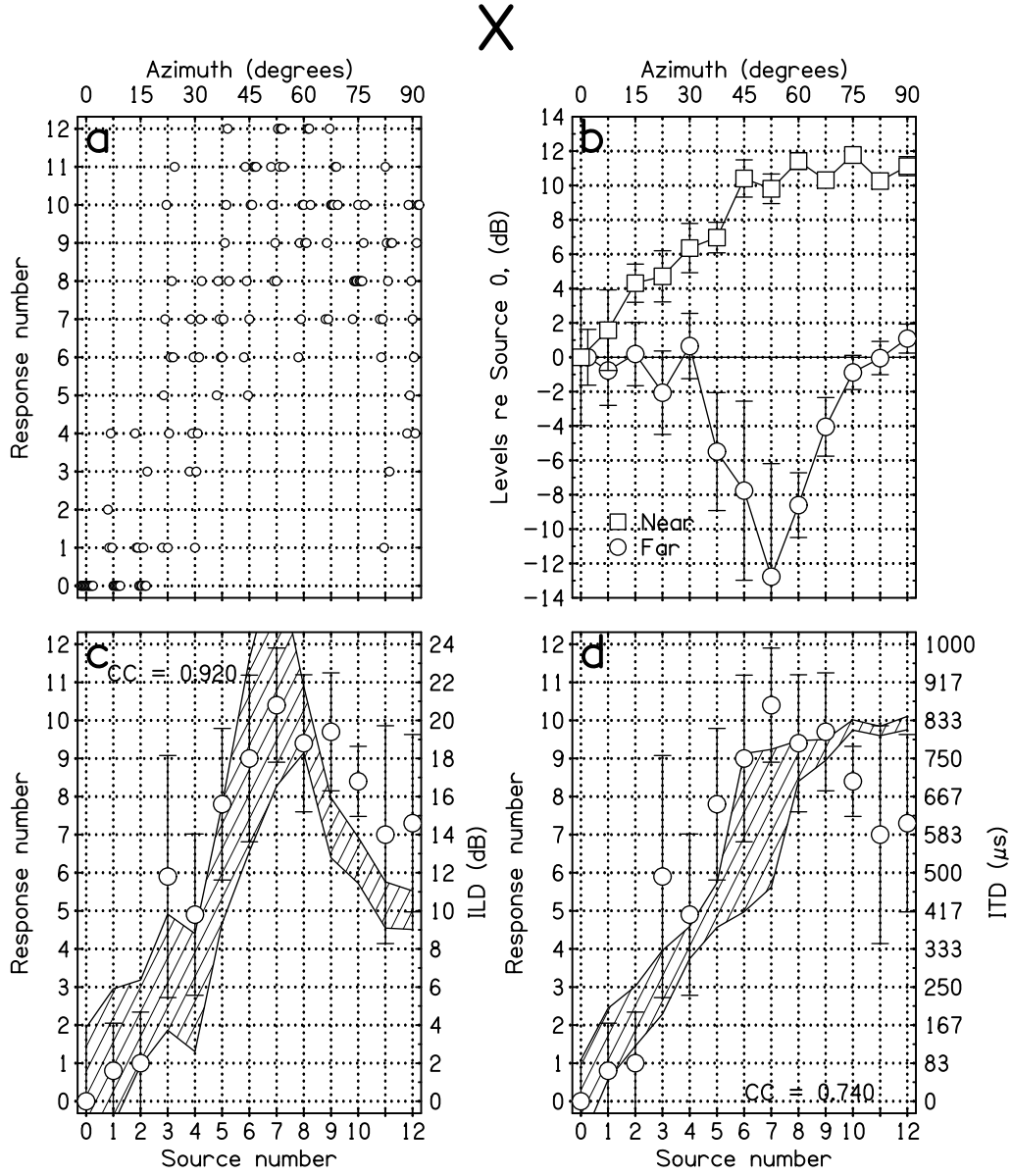


Figure 1.10 Same as for Fig. 1.6 except for listener X and the correlation coefficients for panels (c) and (d) are 0.920 and 0.740, respectively.

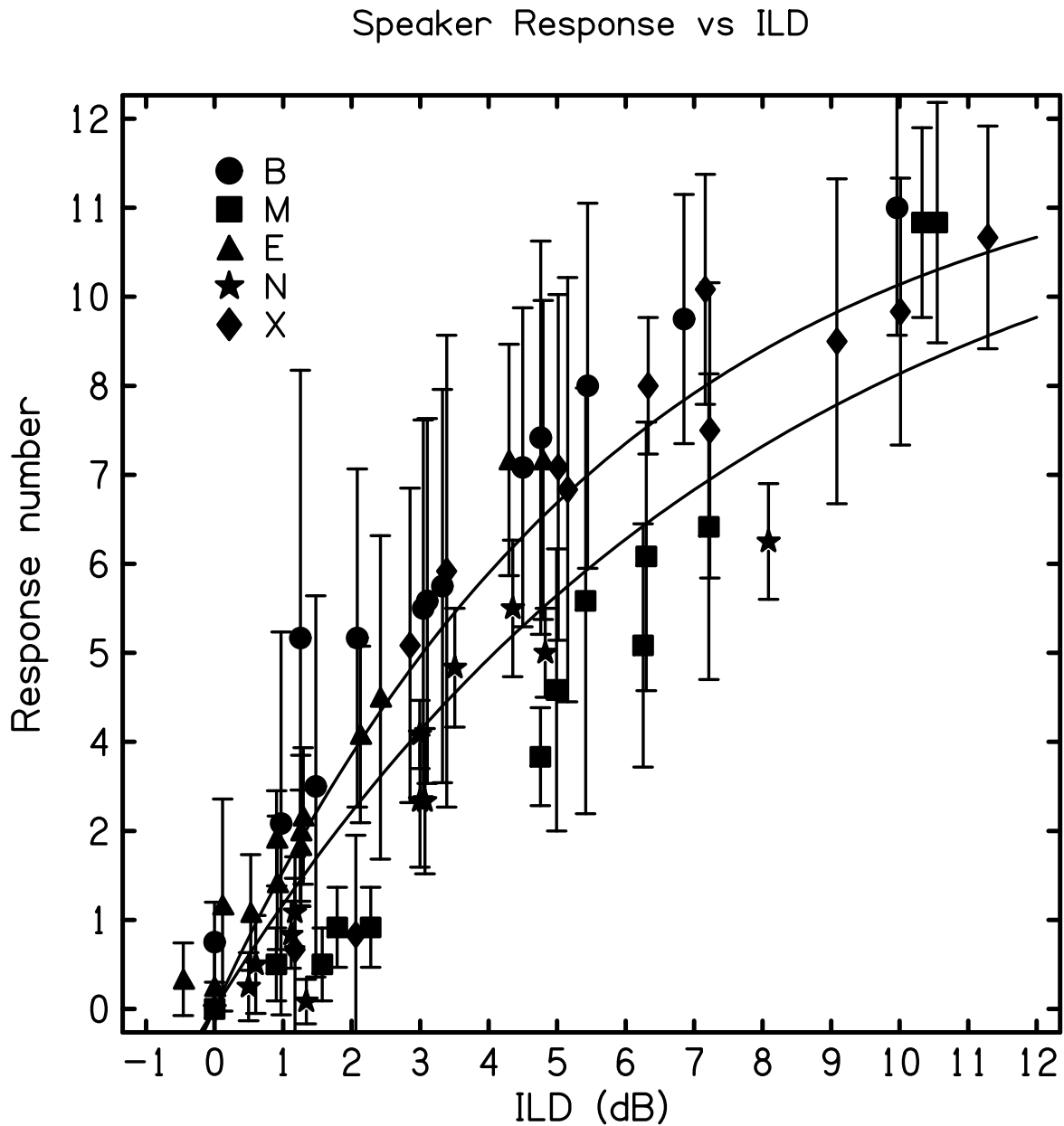


Figure 1.11 Speaker Response vs. ILD. Each data point represents a particular loudspeaker for a given listener. Each listener's speaker response is plotted versus the mean ILD for a given loudspeaker. The error bars show plus and minus the standard deviation for speaker response. The standard deviations for ILD are not shown, but are generally on the order of about 1 dB. The two curves are given by equation 1 where the upper curve is fit to these data with a parameter of 5.95, and the lower curve is fit to Yost's 1981 experiment with a parameter of 8.

As shown in panel (c) of Figs. 1.6–1.10, the peak of the listeners’ ILD curves fall into two categories. Listeners B, M and X had peaks of about 11 dB. Listener N had a peak of about 6 dB, and listener E had a peak of about 7 dB. The peak for the spherical head model was 8 dB, and for KEMAR it was 14 dB.

The variation in ILD for a given listener and source is probably due to the motion of the listeners during the trials. This is likely true because of the very small standard deviation of the KEMAR’s ILD when removing and re-inserting the probe microphones over 10 trials. Overall, though, the human subject’s ILD standard deviations were reasonably small (a few decibels), as can be seen by the shaded regions in panel (c) on Figs. 1.6–1.10. Looking at the scatter plots in panel (a) of Figs. 1.6–1.10, it is evident that the variation in responses is largely caused by variation around the mean, rather than by extreme outliers.

It is interesting that listener E and especially listener N were reluctant to give responses greater than speaker seven. This may be due to the field of view that the listeners had, as it is difficult for listeners to look directly at speakers far to the right side of their head. So although these types of listeners might have been successful in correctly identifying sources up to and including the peak of their ILD curves, they still experienced confusion about sources beyond the peak. These listeners may concede that the best that they can do is to correctly locate the peak of their ILD curve, and simply decide to not respond with sources greater than a particular speaker number. Listener N seemed to do this while not minding that all responses for sources past the peak would be incorrect. Other listeners tended to stretch out their responses, as if they assumed that the greatest ILD that they heard had to be coming from one of the speakers on the far right. These different strategies may have been the result of various amounts of practice and knowledge gained from some of the preliminary experiments.

Figure 1.11 displays the results from the identification experiment by plotting each listener's speaker response vs. the average ILD measurement for each source for that listener. Responses increase roughly with ILD, which is to be expected based on the correlation coefficients shown in panel (c) of Figs. 1.6–1.10. However, these data may better be described by a compression function. Goupell and Hartmann [14] used equation (1.11),

$$\Psi_{\text{ILD}} = 12 \operatorname{sgn}(\text{ILD})(1 - e^{-|\text{ILD}|/p}) \quad (1.11)$$

to fit Yost's 1981 lateralization experiments. The variable, Ψ_{ILD} , is the response for a given ILD. The pre-factor of 12 represents the range of possible responses (in this case 0 to 12 loudspeakers). The parameter, p , was found to be equal to 8 for Yost's data. For this current experiment, p was found to be 5.95. In Fig. 1.11, the lower curve shows, equation (1.11) with p equal to 8 (for Yost's data), and the upper curve shows p equal to 5.95. The upper curve appears to be a better fit to the eye. More importantly, though, the data do appear to fit the shape of this curve. This figure indicates that listeners are basing their localization responses on ILD.

The high correlations between listeners' responses and ILD suggest that they are relying heavily on the ILD as a means of localization. Is the ILD the only potentially useful cue that is available to listeners in this experiment? ITD in the fine structure is not available because 1500 Hz is too high [7]. The onsets and offsets are too slow to provide a cue. Changes in the spectrum are used to localize complex signals. Particularly, the pinna causes spectral peaks and notches at high frequencies which disambiguate signal sources that are in front of, above, or behind listeners [6, 9]. One could imagine that spectral cues could be useful in the azimuthal plane as well. Because of the raised cosine onset and offset in the signal, there is

some spectral splatter in the 1500 Hz tone. However, there is very little power at the lowest audible frequencies [21], and even less at frequencies above 4 kHz, where spectral cues are known to be of use [51]. Assuming that there is not any noticeable noise or distortion in the loudspeakers, the only plausible cues other than the interaural level difference are the individual levels in both ears.

Because the ILD is non-monotonic, there are ranges of azimuthal angles where it is an ambiguous cue. Could a listener with both the knowledge of and the ability to detect individual levels in both ears be able to develop a strategy to disambiguate sound sources that share the same ILD? Two possible strategies utilizing the individual levels in both ears are suggested. The first strategy begins by considering the ILD. For each angle after the peak of the ILD, there is a corresponding angle before the peak that shares the same ILD. By considering ILD alone, there is no way to determine which of these two angles is the location of the sound source. However, this is a hypothetical listener that knows that at this particular frequency the ILD is ambiguous. Through experience, this listener knows how the ILD behaves vs. azimuth for any frequency. Furthermore, this listener knows how the near and far ear levels behave vs. azimuth for any frequency. For a pair of angles with an ambiguous ILD, a listener could consult either the left or right levels to resolve the ambiguity if the difference in level for the pair of angles is greater than some just-noticeable-difference, for example about 1 dB. Either or both ears meeting this criteria would be enough for this hypothetical listener to resolve sources that are ambiguous in ILD. It does not matter if the individual levels are increasing or decreasing with the azimuth. This might make this particular strategy less realistic for a listener to employ.

Strategy 2 does not consider the ILD at all. Instead the individual levels are the only cue that is considered. Because the non-monotonicity is mostly the result of the changing

Listener	Number of speakers to the right of the peak	Strategy 1 success rate	Strategy 2 success rate	Correlation between listener responses and perfect responses
B	4	25%	50%	0.49
E	5	100%	0%	0.29
M	5	60%	80%	0.64
N	5	60%	0%	0.20
X	5	80%	80%	0.76

Table 1.1 Independent ear strategy success rates, and correlations between listener responses and perfect responses.

level in the far ear, this cue itself is ambiguous. A listener trying to resolve an ambiguity in angle caused by the non-monotonic level in the far ear may consider the level in the near ear to resolve the ambiguity. To make this strategy easier to utilize, assume that the level in the near ear must increase with azimuth. If the level does not increase with azimuth for a pair of ambiguous sources, then this strategy will fail. But, this avoids the problem of the listener memorizing a large amount of information about how the near ear changes with azimuth for any frequency. So assuming that the near ear is monotonic, then if the difference in level between these ambiguous angles is greater than the just-noticeable-difference, then the listener should be able to determine the correct angle of the source. This strategy is simpler to employ, because the ILD is ignored and only the two individual levels are analyzed. Additionally, consulting only the near ear to resolve the ambiguity is simpler because the near ear tends to increase monotonically. However, this simplicity means that strategy 2 may be less successful than strategy 1.

Using a just-noticeable-difference of 1 dB, the success rates for implementation of these two strategies were calculated. Each source to the right of the ILD was paired with an angle to the left of the peak by interpolating the cue. Then the individual levels at these angles were compared.

As shown in Table 1.1, the success rate using strategy 1 is quite high. This is perhaps because this technique assumes that the listener is too good at being able to analyze the available information. The success rate for strategy 2 gives more mixed results. This is probably largely in part because of the requirement that the level in the near ear increases with azimuth. For listener E, the near ear level tends to decrease with azimuth. This is why the success rate changed from 100% to 0%. It is interesting that the listeners with the highest correlations between their loudspeaker responses and perfect responses (Listeners M and X) tend to have the higher success rate using strategy 2. Notice that in panels (b) of Figs. 1.7 and 1.10, the far ear (squares) increases more dramatically with azimuth than for other listeners. One possibility is that these two listeners really are using the levels in their near ears to aid their localization. However, because of this increase in near-ear level, the ILD for these listeners (panels c) decreased to only about 10 dB for speaker 12, but for the other listeners the ILD for speaker 12 was about 2–4 dB. Therefore listeners M and X made smaller errors in localization at large azimuths. This alone could account for their better overall performance. Listener B has a moderate success rate using strategy 2, and a moderate correlation between his responses and perfect response. Finally listeners E and N have the lowest correlations between their responses and perfect response, and also have the lowest success rate using strategy 2. This does not necessarily mean that listeners are effectively employing a strategy like strategy 2, but it does shed some light on the possibility that listeners are attempting to use individual ear levels to localize, and the only ones that have success in reality are the ones for which this model predicts success. Either way, no listeners could be said to have been successfully localizing these stimuli; they are clearly fooled by the non-monotonic ILD.

1.4.2 Experiment 2: Discrimination Experiment

Experiment 2, termed, the discrimination experiment, was identical to the identification experiment, except that two different loudspeakers presented the two successive tones. The listener was asked to determine if the source moved from right to left or from left to right across the two intervals. The goal of the discrimination experiment was to discover whether the same perceptual process observed in the identification experiment would also be observed in the discrimination experiment; specifically, whether the discrimination results could be predicted from the identification results.

The listener's response was indicated by pressing a left button or a right button to indicate the perceived direction of motion. Recordings were made of each stimulus using the probe microphones. For each run, six pairs of speakers were chosen, and each pair was presented 10 times. The ordering of the 60 pairs was random, and the direction that the source moved for each pair was random. Each listener performed 2 runs; therefore there were 12 different speaker pairs. The 12 speaker pairs were not randomly chosen. They were deliberately chosen mainly according to expectations based on the ILD calculations for the spherical head model. The intent was to choose a variety of speaker pairs, with a range of ILD changes, and the expectation was that the listener would have a range of success rates depending on the speaker pair. Success rates were expected to be mostly correct, mostly incorrect, or ambiguous responses.

The hypothesis was that the results for the percentage of correct responses for a given speaker pair in the discrimination experiment could be predicted by using the results from the identification experiment. It was assumed that the underlying distribution of the internal representation of azimuth for each loudspeaker was normal with a mean and standard

deviation given by the mean and standard deviation of the identification experiment. The probability of responding correctly that source B was—for example—to the right of source A was the probability that a response location chosen from the identification distribution for speaker B is greater (further to the right) than the response location chosen from the identification distribution for speaker A. Or in other words, assuming that speaker B is to the right of speaker A, and a particular listener has mean responses for speakers A and B, μ_A and μ_B , respectively, and standard deviations for speakers A and B, σ_A and σ_B , respectively, then the predicted probability, P , of the listener giving a correct response in the discrimination experiment is

$$P = \sqrt{\frac{1}{2\pi\sigma^2}} \int_0^\infty e^{-\frac{[x-(\mu_B-\mu_A)]^2}{2\sigma^2}} dx \quad (1.12)$$

where $\sigma^2 = \sigma_A^2 + \sigma_B^2$. For example, when $(\mu_B - \mu_A)$ is large and positive, P will approach 100%. When $(\mu_B - \mu_A)$ is large and negative, P will approach 0%. When $(\mu_B - \mu_A)$ is zero, P is 50%.

1.4.2.1 Discrimination Experiment Results

Figure 1.12 shows that the predictions of the percentage of correct responses was reasonably successful. If the predictions were perfect, then all of the data would lie on the dashed 45° line. This was not the case, but it is important to remember that each data point only represents ten trials. If more trials were performed, it would not be unreasonable to expect that there would be some variation in the actual percent correct values. Still, as expected, most of the data lie in the lower-left and upper-right quadrants of the plot. However, there does appear to be a large clustering of data at 0% and 100%. This would indicate

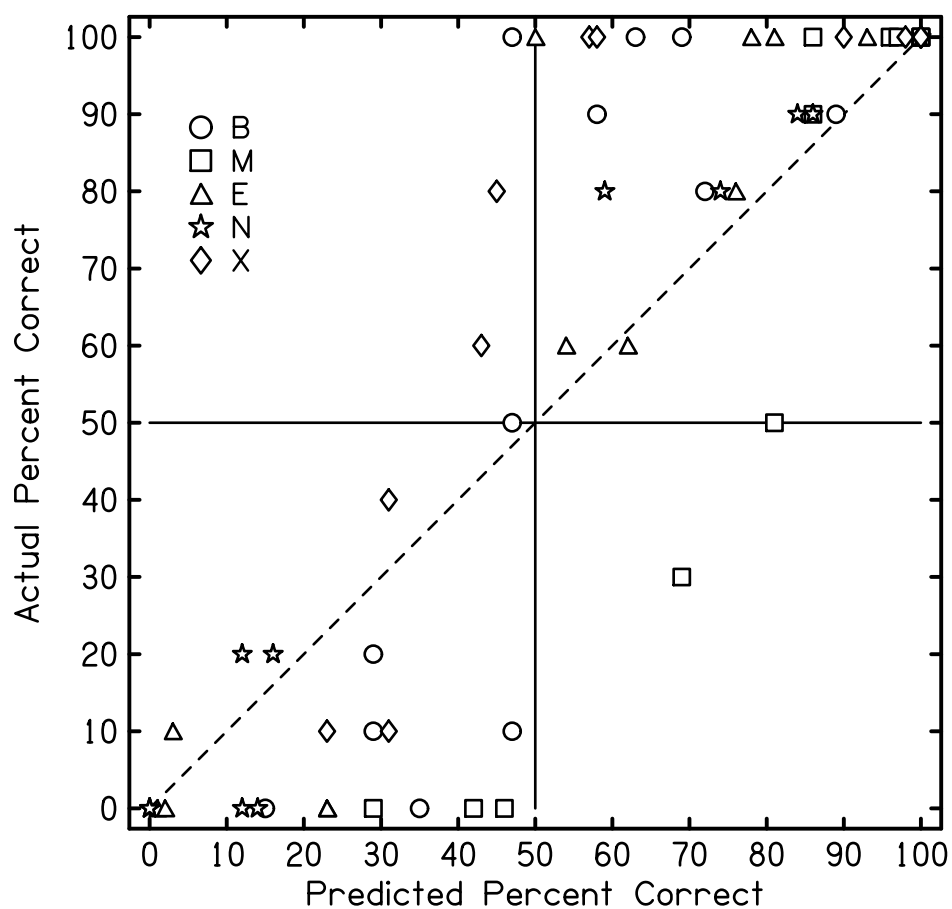


Figure 1.12 Discrimination Experiment. The predicted percent correct and actual percent correct are plotted against each other for all five listeners. The percent correct was predicted based on the data from individual listeners in the identification experiment.

that listeners were much better at discriminating the two tones than was predicted by the identification experiment. This is understandable. The discrimination task seemed to be much easier for listeners. In the discrimination task, a listener only has to listen to stimuli from two sources at a time, and make a judgment, but in the identification experiment, the listener has to create a mental map of 13 sources and try to remember how these 13 stimuli sound over the course of the entire identification experiment. This understandably leads to a large standard deviation for each speaker's response number. This would tend to cause

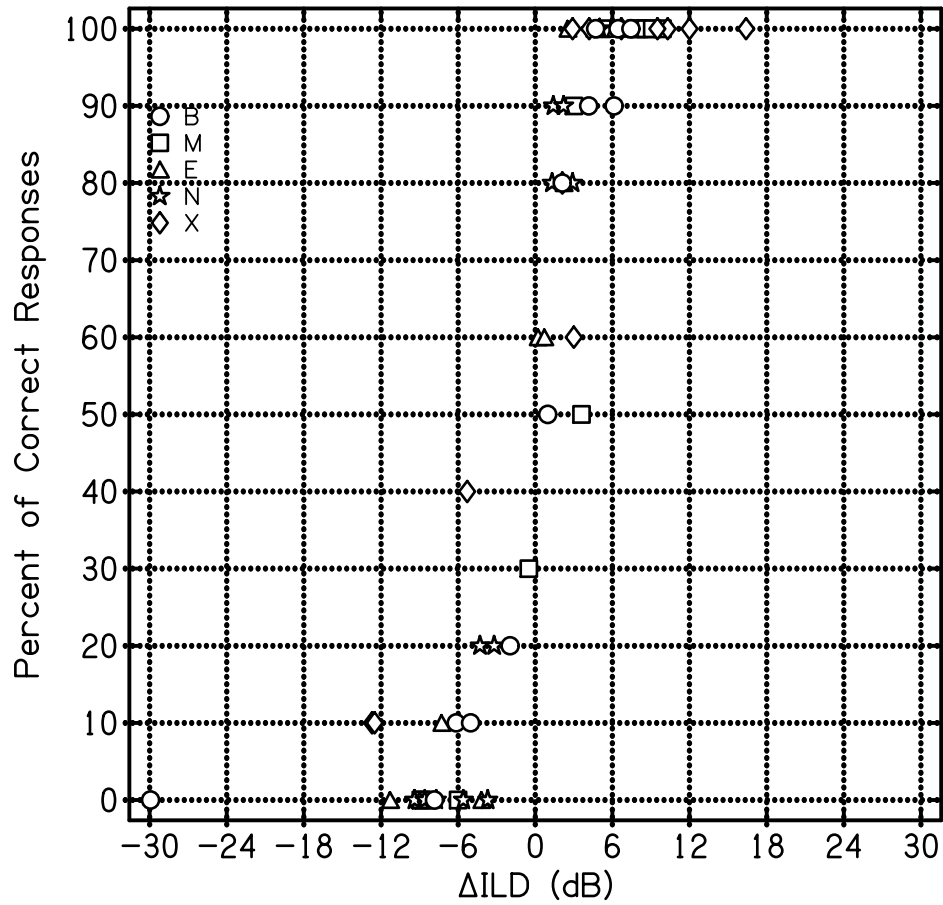


Figure 1.13 The percent of correct responses vs. the average difference in ILD between the speaker pairs. The difference in ILD is taken to be the right speaker's ILD minus the left speaker's ILD. The standard deviations in the Δ ILDs are typically on the order of 1 dB or less.

Fig. 1.12 to look as it does; with the predictions spread somewhat evenly from 0% to 100%, but with many of the actual percentage correct values at either 0% or 100%.

Figure 1.13 displays the actual percent correct vs. the average ILD that was measured between the speaker pairs. When the ILD increased from left to right, the listeners tended to perform very well. Similarly, when the ILD decreased from left to right, the listeners performed very poorly. This is what was expected. Only when the $|\Delta\text{ILD}|$ was less than about 4 dB, did the listeners tend to give ambiguous responses.

Although the identification experiment was not entirely successful in predicting the discrimination experiment results, these results agree with the identification results in that there is no evidence that the listeners were able to rely on anything other than the ILD to complete the task. For example, the listeners were unable to use the individual left and right ear levels to successfully perform the discrimination experiment.

1.4.3 Experiment 3: Identification with Training and Feedback Experiment

The motivation for this experiment was to see if listeners were able to learn to successfully localize the stimuli. From the results of the identification experiment, it is known that untrained listeners are not successful at localization. Perhaps if the listeners were given training and feedback during the experiment, their responses would improve. In the identification with training and feedback experiment, the listener performed the same task as in the identification experiment, but was presented with training before each pass through the 13 loudspeakers, and was presented with feedback during each pass. Each listener performed five runs, each with two passes through the 13 loudspeakers for a total of ten passes through

each speaker. The reason there were only two passes per run was that the training and feedback caused the runs to be longer than a typical identification run. The training before each pass consisted of playing the 13 loudspeakers in order—from the front to the right—for the listener. The listener knew that the loudspeakers were being presented in order and was instructed to make an attempt to use the training to his advantage. The nature of the feedback was to inform the listener as to whether the stimulus came from speakers zero through six, or from speakers seven through twelve. The feedback was given only after the listener gave his response, and was presented by means of two flashing lights which were located directly below speaker zero.

1.4.3.1 Identification with Training and Feedback Results

Column	1	2	3	4	5	6
x	NTF1	NTF1	NTF1	NTF2	NTF2	TF1
y	NTF2	TF1	TF2	TF1	TF2	TF2
Listener						
B	0.91	0.86	0.90	0.94	0.90	0.92
E	0.90	0.78	0.87	0.87	0.94	0.93
M	0.97	0.92	0.90	0.87	0.91	0.96
N	0.96	0.93	0.86	0.97	0.81	0.81
X	0.97	0.81	0.85	0.76	0.83	0.94
Mean	0.94	0.86	0.88	0.88	0.88	0.91

Table 1.2 The Pearson product-moment correlation coefficients for various sets of data for each listener, as well as the mean coefficients across all of the listeners. NTF1 refers to the first five passes of the experiment with no training and feedback. NTF2 refers to the second five passes of the experiment with no training and feedback. TF1 stands for the first five passes of the experiment with training and feedback. TF2 refers to the second five passes of the experiment with training and feedback. For each of these half experiments—which consist of five passes through the loudspeaker array—the mean response for each loudspeaker was used in the correlation calculations.

As seen in Figs. 1.14–1.18, the identification with training and feedback responses are

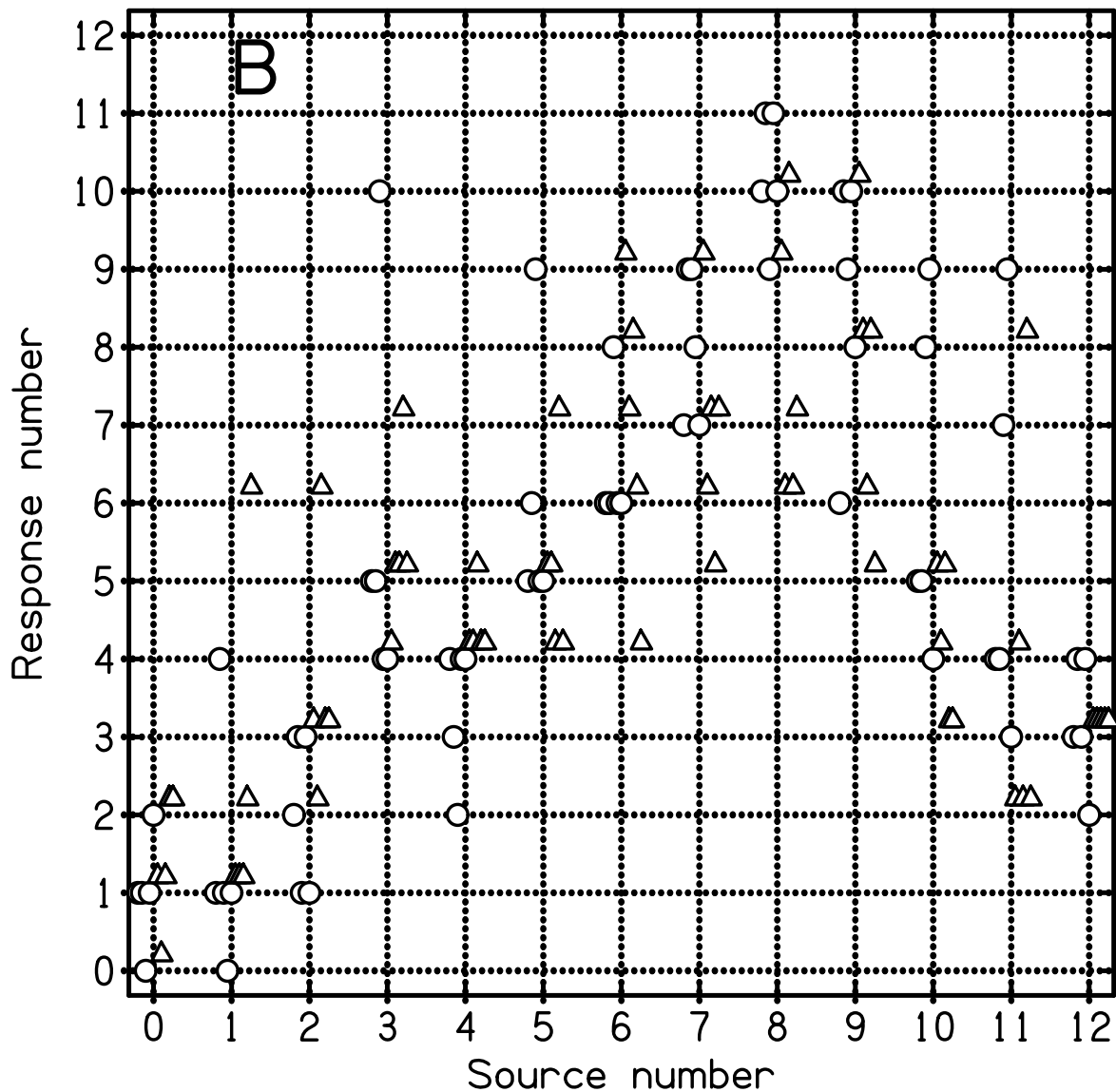


Figure 1.14 Loudspeaker response numbers for listener B during the training and feedback experiment. There are ten responses per source. The circles represent responses for the first five passes through the 13-loudspeaker array, and the triangles represent responses for the second five passes through the 13-loudspeaker array. The data points have been jogged slightly off of the grid lines so that they may all be visible.

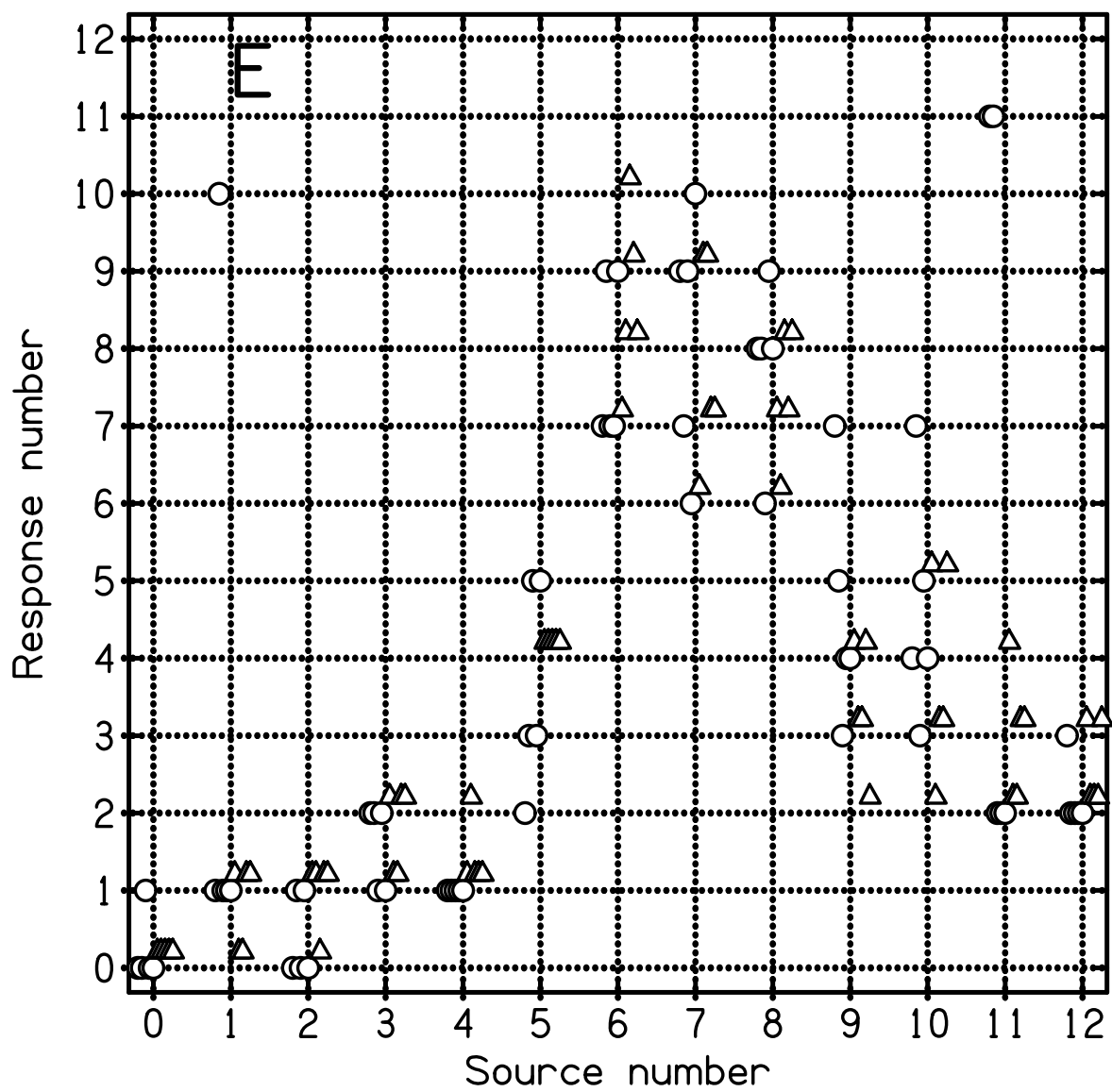


Figure 1.15 Same as Fig. 1.14 except for listener E.

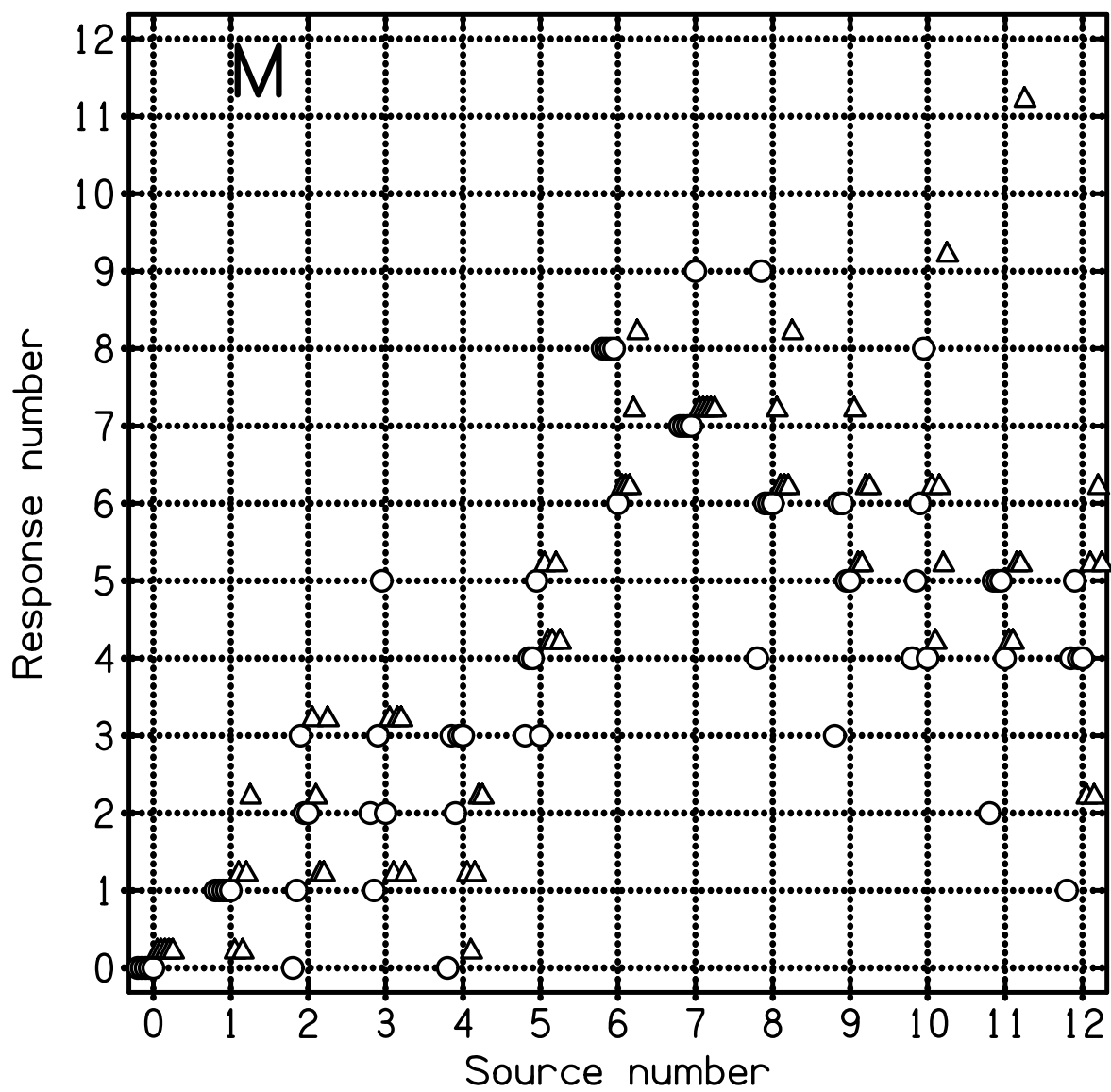


Figure 1.16 Same as Fig. 1.14 except for listener M.

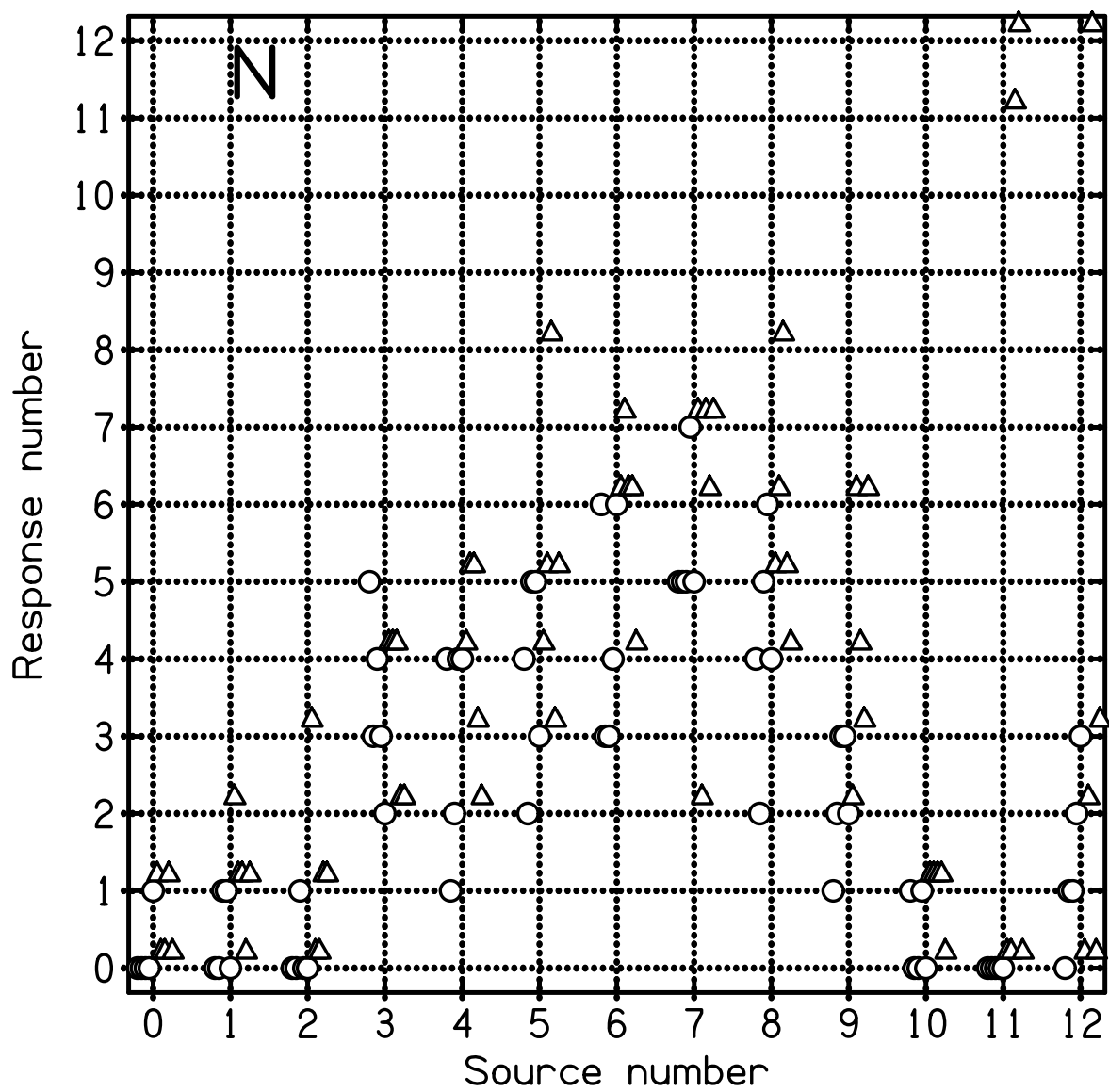


Figure 1.17 Same as Fig. 1.14 except for listener N.

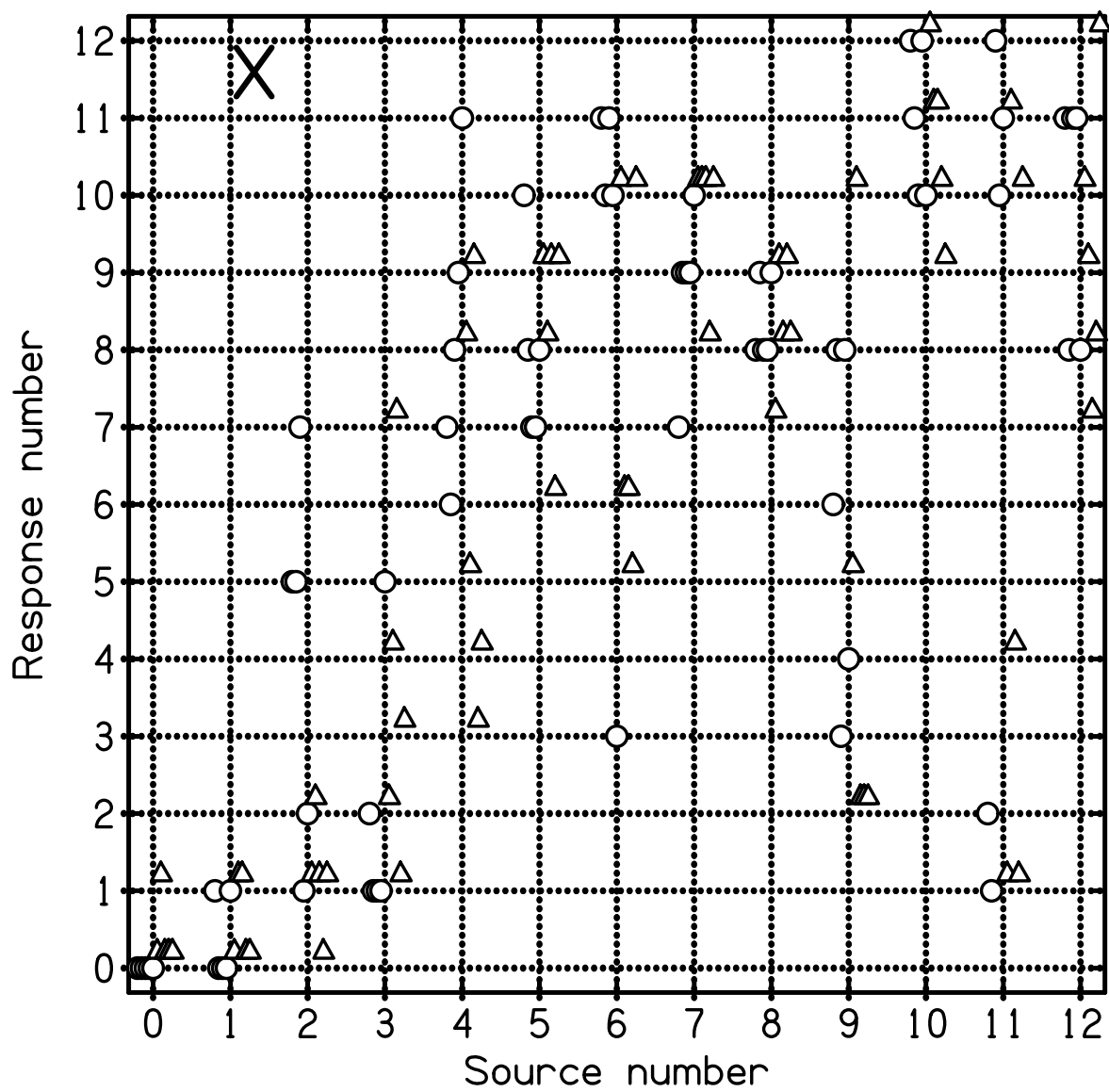


Figure 1.18 Same as Fig. 1.14 except for listener X.

Column	7	8	9	10
x	Perfect	Perfect	Perfect	Perfect
y	NTF1	NTF2	TF1	TF2
Listener				
B	0.35	0.60	0.56	0.37
E	0.31	0.27	0.53	0.44
M	0.58	0.70	0.64	0.77
N	0.28	0.12	0.07	0.45
X	0.80	0.72	0.79	0.75
Mean	0.46	0.48	0.52	0.56

Table 1.3 The Pearson product-moment correlation coefficients for various sets of data for each listener, as well as the mean coefficients across all of the listeners. Abbreviations are the same as those explained in the caption for Table 1.2. “Perfect” refers to a hypothetical listener who always gives the correct response to a stimulus.

non-monotonic for all listeners, just as are the responses without training and feedback. Subjectively, the data in these figures suggest that listeners were not successful at localizing the stimuli with the aid of training and feedback. Additionally, in looking at the differences between the circle data points and the triangle data points, it appears as though there is not a significant difference in learning between the first five passes through the array and the second five passes through the array.

Tables 1.2 and 1.3 show Pearson product-moment correlation coefficients for each listener. These coefficients were calculated with the mean speaker responses during the first and second halves of various experiments. Each coefficient, $r_{x,y}$ is then calculated using equation (1.13)

$$r_{x,y} = \frac{\sum_{i=0}^{12} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{12} (x_i - \bar{x})^2 \sum_{i=0}^{12} (y_i - \bar{y})^2}} \quad (1.13)$$

where x_i and y_i are the mean responses for speaker i for the five passes indicated in the

Column	11	12	13
x	Training & Feedback	Training & Feedback	No Training and Feedback
y	No Training & Feedback	Perfect	Perfect
Listener			
B	0.94	0.49	0.49
E	0.91	0.49	0.29
M	0.92	0.71	0.64
N	0.94	0.28	0.20
X	0.83	0.78	0.76
Mean	0.91	0.55	0.48

Table 1.4 Pearson product-moment correlation coefficients for all ten passes through the loudspeaker array. Each listener’s coefficients are shown for various sets of data as well as the mean coefficients across all of the listeners.

column headings in Tables 1.2 and 1.3. Symbols \bar{x} and \bar{y} and are the means of the x_i and y_i , respectively. In Table 1.4, the coefficients are calculated using the mean speaker responses for all ten passes through the loudspeaker array.

Overall, there is a high correlation between the first and second halves of the training and feedback experiment (column 6), which would indicate that listeners were not responding very differently during the beginning and end of this experiment. This was also the case during the experiment with no training and feedback (column 1). Listeners changed their behavior more during training and feedback more than without, but not by a lot (0.91 versus 0.94). There is, however, less correlation between experiments with and without training and feedback (columns 2-5). Values in columns 2–5 have correlations that are approximately 0.87. This indicates that listeners did change behavior slightly with training and feedback, but that the behavior change likely happened early in the experiment, rather than gradually attempting to learn as the experiment progressed. Did this small change in behavior result in better localization? To answer this question, the first and second halves of the two experiments were also compared with a hypothetical listener that always gives the correct

responses. Looking at these correlations, it appears that the listeners did improve slightly with training and feedback, but not by very much.

Table 1.4 shows the correlation coefficients that were calculated using the mean responses across all ten passes through the loudspeaker array, rather than the first or second half of each experiment. Column 11 compares the experiment with training and feedback to the experiment with no training and feedback, column 12 compares the experiment with training and feedback to a hypothetical listener who always gives correct responses, and column 13 compares the experiment without training and feedback to a hypothetical listener who always gives correct responses. Column 11's results are similar to columns 2–5 in Table 2, which also compare experiments with training and feedback to experiments without training and feedback. The mean of column 12 (0.55) is slightly larger than the mean of column 13 (0.48), which would indicate that on the average, listeners were performing slightly better with training and feedback than without. Listener E improved the most (0.29 to 0.49), however his performance was worse than the mean both with and without training and feedback. Although the correlations with perfect responses are higher with training and feedback than without, it would not be appropriate to state that any listener was successfully localizing during either of these experiments, although, the mean correlations increase in the order that one might expect them to. In summary, listeners were not able to successfully localize the stimuli with training and feedback.

1.4.4 Experiment 4: Identification with Amplitude Modulation

Experiment

Generally, it is known that listeners are unable to use the ITD in the waveform to localize for frequencies greater than about 1500 Hz [7]. For the listeners in this experiment, this claim is supported by panel (d) in Figs. 6–10. These plots show poor correlation between listeners’ responses and ITD. However, it is also known that listeners are able to localize on the basis of ITD in the envelope for modulated high-frequency tones. Modulating the amplitude of the sine tone at a low frequency adds envelope fluctuations with potentially useful ITD cues [23, 33, 46].

The amplitude modulation experiment consisted of a 1500-Hz tone modulated at 100 Hz. This tone lasted for 1000 ms with a rise/fall time of 250 ms. The listener completed two runs with five passes each for a total of ten passes through the 13 loudspeakers. Acoustical measurements in the ear canals were not made during this experiment.

1.4.4.1 Identification with Amplitude Modulation Results

Figure 1.19 and Table 1.5 show that some listeners were successful in localizing the stimuli, and others were not. On average, listeners improved from data column 3 (0.48 correlation between pure tone and perfect responses) to data column 2 (0.86 correlation between amplitude modulation and perfect responses). Listener N was the most successful localizer in this experiment. Listener N’s responses are monotonic—with the exception of speaker 12—and appear to lie on the 45° diagonal. This is in agreement with the correlation between amplitude modulation and perfect responses of 0.99 for listener N. This is a dramatic improvement over this listener’s correlation between pure tone and perfect responses of 0.20. Listener B

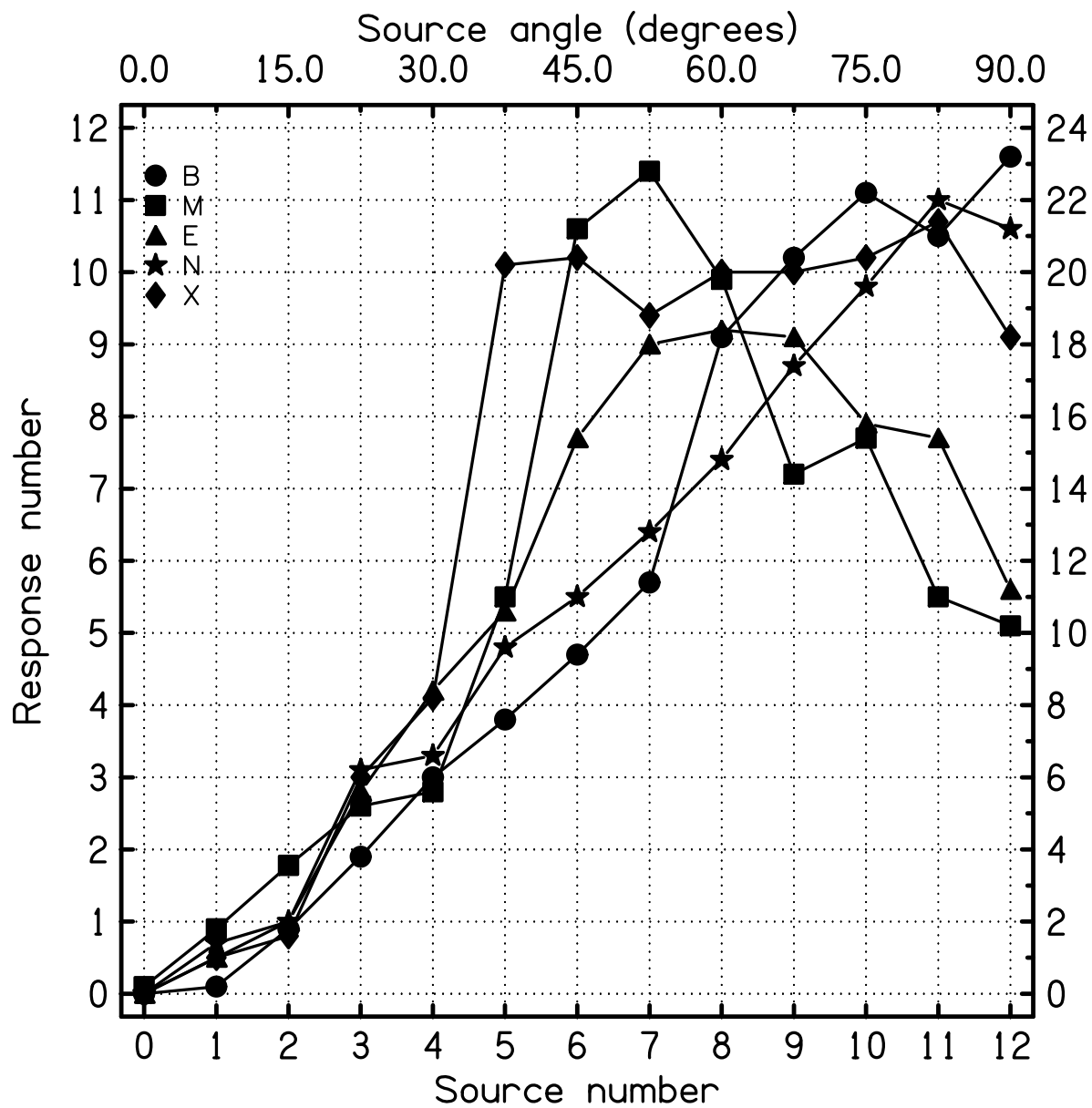


Figure 1.19 Each listener's mean response for each source number in the amplitude modulation experiment.

Column	1	2	3
x	Amplitude Modulation	Amplitude Modulation	Pure Tone
y	Pure Tone	Perfect	Perfect
Listener			
B	0.46	0.98	0.49
E	0.66	0.82	0.29
M	0.94	0.63	0.64
N	0.20	0.99	0.20
X	0.93	0.87	0.76
Mean	0.64	0.86	0.48

Table 1.5 Pearson product-moment correlation coefficients for various experiments. The coefficients are calculated for each listener, and the mean coefficients are averaged over the five listeners. Equation (1.13) is used to calculate the coefficients, except that for a given listener and experiment, the data used in the correlation calculation were the mean responses of the ten passes through each loudspeaker. “Pure Tone” refers to experiment 1, “Amplitude Modulation” refers to experiment 4, and “Perfect” refers to a hypothetical listener who always gives the correct response to a stimulus.

performed well in this experiment also with a correlation between amplitude modulation and perfect responses of 0.98. With the addition of amplitude modulation, listeners B and N could be said to be giving more weight to ITD than ILD. Listener X’s responses plateau at speaker number five, rather than rolling over as they did in the pure tone identification experiment. This would suggest that listener X was giving moderate amounts of weight to both ITD and ILD cues in this experiment. Although listener X was the best pure-tone localizer, this listener still found room for improvement with the addition of amplitude modulation. Although listener E did not perform as well as the previous three listeners (amplitude modulation and perfect correlation of 0.82), he did improve substantially from the pure tone experiment (pure tone and perfect correlation of 0.29). This listener was also likely giving moderate amounts of weight to both ITD and ILD cues. Listener M was by far the poorest performer. This was the only listener with a correlation between amplitude modulation and perfect responses which was equivalent to his correlation between pure tones and perfect

responses. It appears as though this listener gave almost no weight to the newly introduced ITD cues. Overall, these results support the idea that listeners weight ITD and ILD cues in a wide variety of different ways.

1.4.5 Experiment 5: Identification with Narrow-band Noise Experiment

In the identification with narrow-band noise experiment, it is presumed that listeners may be able to use envelope ITD cues to localize, in a similar fashion to the amplitude modulation experiment.

In the identification with narrow-band noise experiment, the signal used was a narrow band of noise from 1400 to 1600 Hz (which has the same bandwidth as the amplitude modulated signal used in experiment 4). The noise file, $x(t)$, had 201 components of equal amplitude and random phase as seen in equation (1.14), where ϕ_f is the random phase for frequency, f .

$$x(t) = \sum_{f=1400}^{1600} \sin(2\pi ft + \phi_f) \quad (1.14)$$

The file was then scaled so that the magnitude of the peak was +5 volts, and treated with a raised cosine window, giving the file a length of 1000 ms and the same rise/fall time of 250 ms that was used in the previous experiments. There were 13 different noise files used—the random phases were different in each one—and a random file was selected for each stimulus. The 13 noise files were normalized so as to have equal power. As in the regular identification experiment, the listener completed a total of ten passes through the 13 loudspeakers. This was accomplished in two runs with five passes each through

Column	1	2	3	4
x	Noise	Noise	Pure Tone	Noise
y	Pure Tone	Perfect	Perfect	Amplitude Modulation
Listener				
B	0.68	0.91	0.49	0.94
E	0.56	0.86	0.29	0.94
M	0.94	0.67	0.64	0.98
N	0.20	0.99	0.20	0.99
X	0.92	0.93	0.76	0.98
Mean	0.66	0.87	0.48	0.97

Table 1.6 Pearson product-moment correlation coefficients for various experiments. The coefficients are calculated for each listener, and the mean coefficients are averaged over the five listeners. Equation (1.13) is used to calculate the coefficients, except that for a given listener and experiment, the data used in the correlation calculation were the mean responses of the ten passes through each loudspeaker. “Noise” refers the experiment with narrow-band noise, “Pure Tone” refers to experiment 1, “Amplitude Modulation” refers to experiment 4, and “Perfect” refers to a hypothetical listener who always gives the correct response to a stimulus.

the loudspeaker array. Acoustical measurements were not made in ear canals during this experiment.

1.4.5.1 Identification with Narrow-band Noise Results

Figure 1.20 shows that some listeners were able to localize successfully while others were not. Table 1.6 shows that the average listener improved from column six (0.48 correlation between pure tone and perfect responses) to column five (0.87 correlation between narrow-band noise and perfect responses). The narrow-band noise results are similar to the amplitude modulation results. Some listeners showed improvement in their ability to successfully localize the stimuli, when compared with their results from the 1500-Hz sine-tone. In particular listener N performed quite well in this experiment just as he did in the amplitude modulation experiment. Listener X’s responses do appear to be monotonic, however a plateau begins

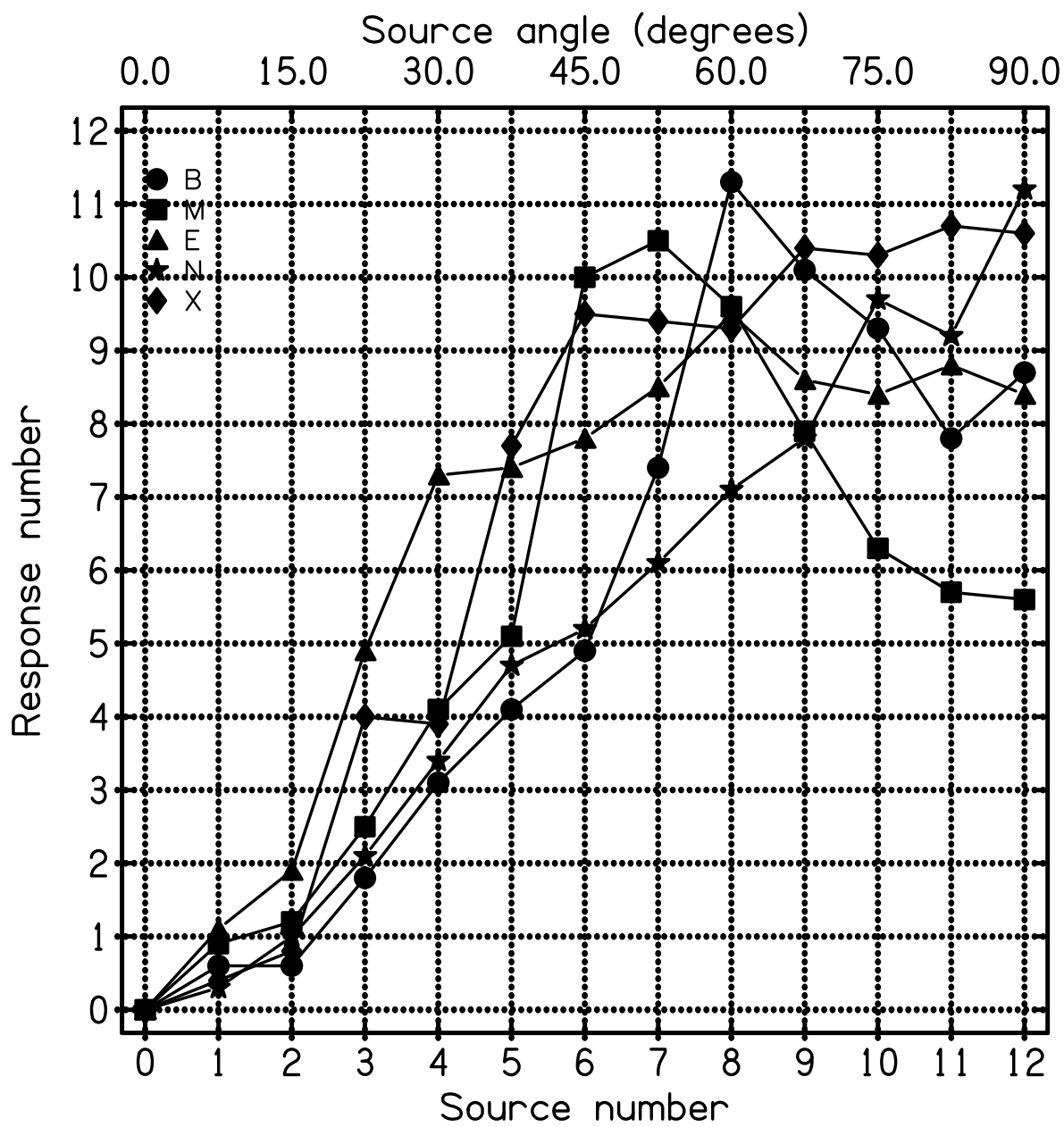


Figure 1.20 Each listener's mean response for each source number for the narrow-band noise experiment.

to form around speaker number nine. Listener E also shows a plateau around speaker six. Meanwhile, listeners B and M show little or no improvement. Overall, listeners performed very similarly to the amplitude modulation experiment (listeners B and E had the lowest correlations, 0.94, between these experiments). Generally speaking, it appears that listeners have a wide variety of results in this experiment. Just as in the amplitude modulation experiment, it seems that some listeners weight envelope ITD (EITD) cues more heavily than ILD cues, while others do the opposite. The data do not show a systematic difference between the results from the amplitude modulation experiment and the narrow-band noise experiment. The average correlation between these experiments was 0.97 (from data column 7). This is also supported by the equivalence between the average correlations between amplitude modulation and perfect responses (0.86 from data column 2), and between narrow-band noise and perfect responses (0.87 from data column 5).

1.5 The effect of band-width on non-monotonicity

Sine tones at 1500 Hz have significant non-monotonic ILDs. However, more realistic listening environments have bands of noise of various widths. It would be expected that the non-monotonic character of the ILD curve should decline as the bandwidth increases. For the spherical-head model, Fig. 1.1 shows that at different frequencies, the peak of the ILD occurs at different azimuths. Additionally, the shape of the curve, and the height of the peak changes with frequency. At lower frequencies the height of the peak is smaller, so perhaps these frequencies do not alter the overall ILD as much as at higher frequencies. The general trend as frequency increases is, the peak increases, the azimuth of the peak increases, and the curve as a whole appears to be more complicated. Perhaps the ILD that results

from a band containing frequencies of this variety will become smooth and monotonic.

The effect of bandwidth on ILD was studied using the KEMAR manikin. In addition to a 1500-Hz sine tone, the bandwidths ranged from a one-third of an octave to 4 octaves, and increased in 1/3-octave increments. The bands were logarithmically centered around 1500 Hz. The components of each stimulus were separated by 1 Hz. They had equal amplitudes and random phases. The bands of noise were played continuously on a loop, and were played in place of a sine tone using the computer program.

Probe microphones were used in the KEMAR's ear canals to make the recordings. The ILDs for each speaker were calculated from the rms voltages in the left and right probe microphones. For each band of noise, recordings were also made using the omni-directional microphone in the center of the loudspeaker array. This allowed for differences in level between different loudspeakers be compensated for so that that the levels in the left and right ears could be calculated separately as well. The levels at loudspeaker zero were subtracted from the individual ear levels at all loudspeakers, and the ILDs at loudspeaker zero were subtracted from the ILDs at all loudspeakers. Therefore, all levels and ILDs are shown relative to speaker zero, and they are shown as if each loudspeaker were to produce the same level at the center of the array. The above procedure was done in place of the loudspeaker equalization process, which was used for the perceptual experiments.

1.5.0.2 The effect of band-width on non-monotonicity results

Figures 1.21–1.26 show the progression in the levels for the right ear, left ear, and the ILD. The measurements made at the 1/3-octave increments which aren't shown here represent only small intermediate changes. Figure 1.21 shows slightly different levels than are shown in Figs. 1.4 and 1.5. For example, the peak ILD in these measurements now occurs at

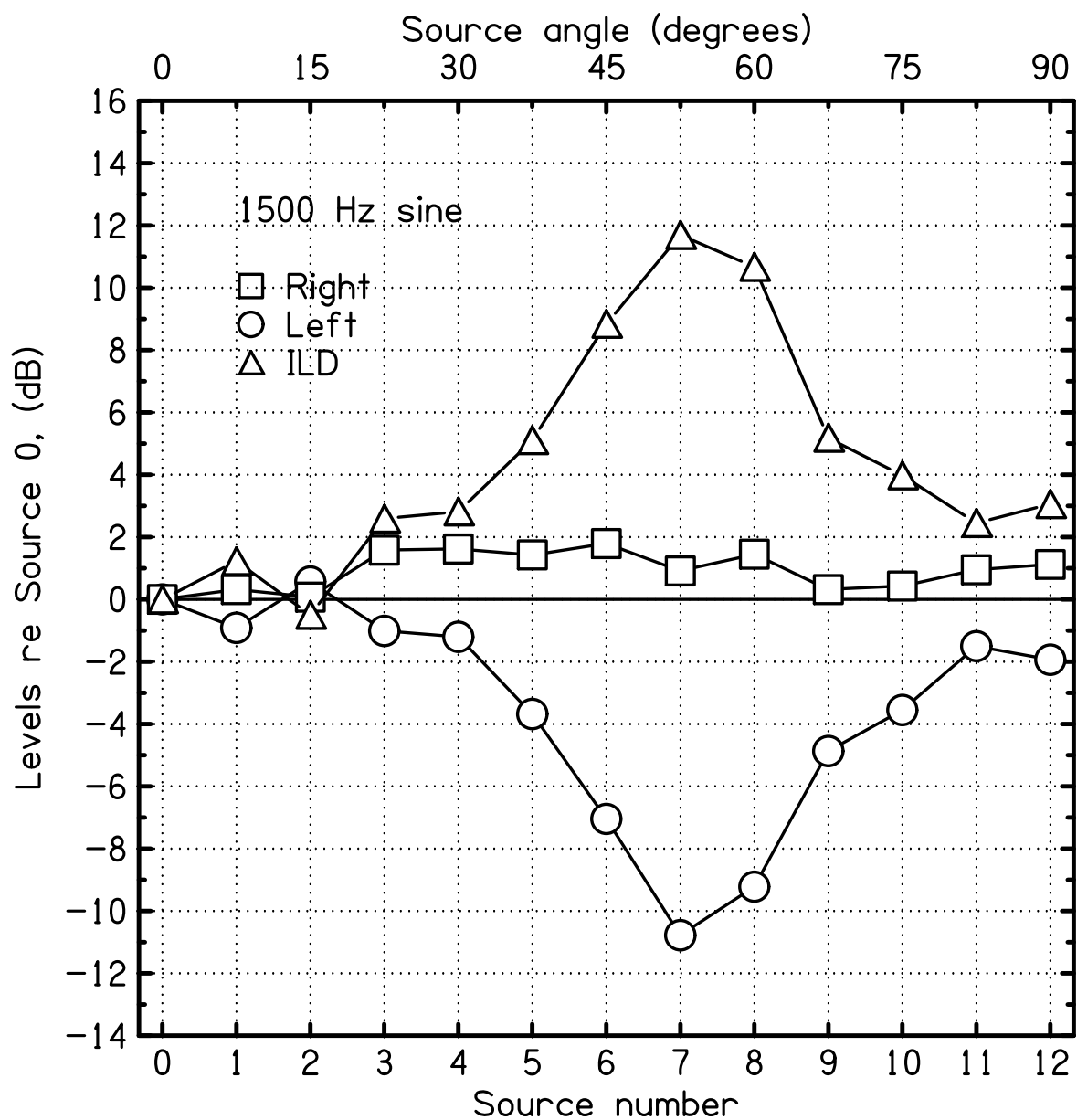


Figure 1.21 KEMAR levels for right, left and ILD with a 1500 Hz sine-tone stimulus. The recordings were made using probe microphones. Levels are shown with respect to speaker zero, and are adjusted for the relative level of the stimulus for each loudspeaker at the center of the array.

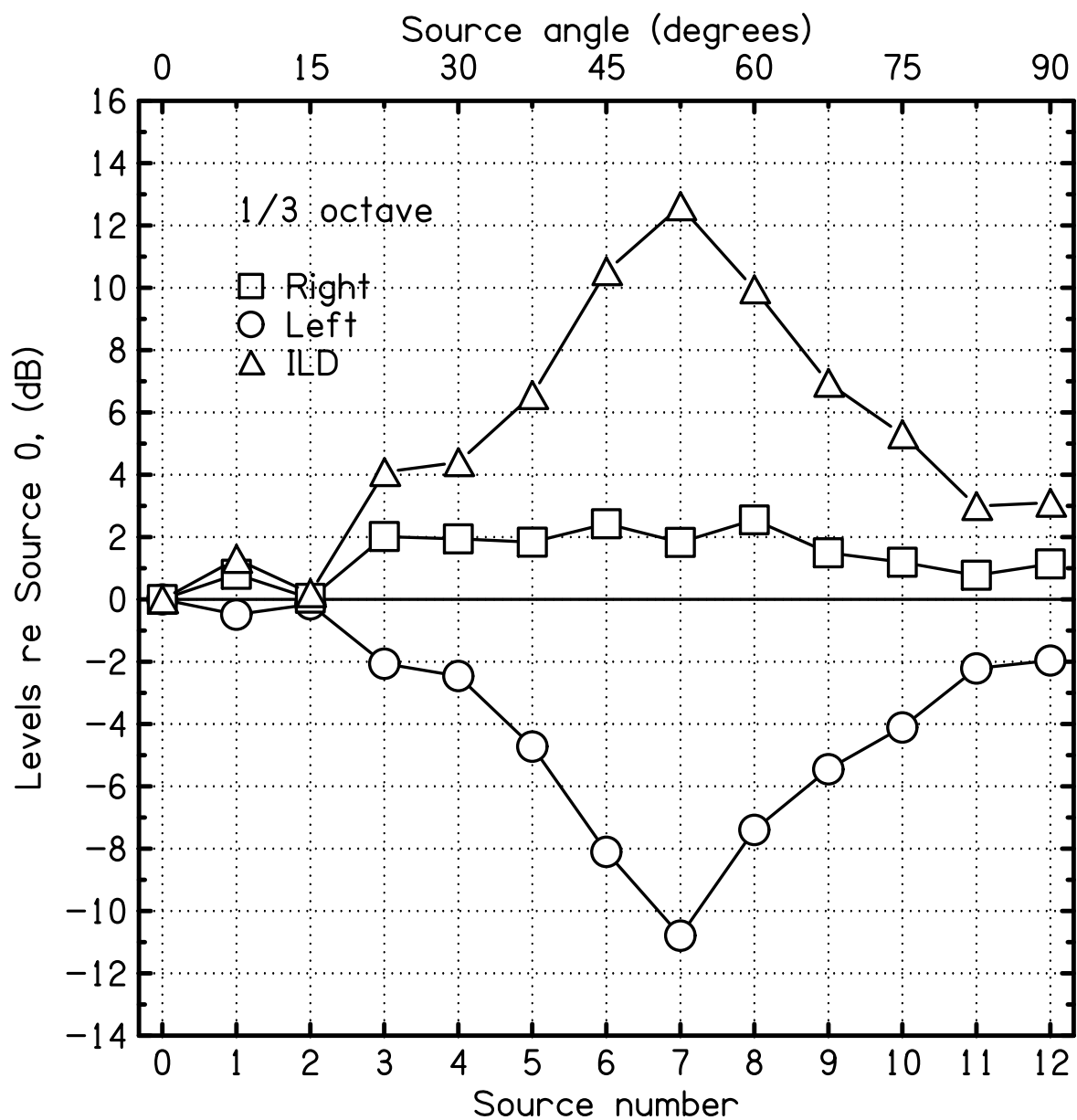


Figure 1.22 KEMAR levels for right, left and ILD with a 1/3-octave-band stimulus centered logarithmically around 1500 Hz. The recordings were made using probe microphones. Levels are shown with respect to speaker zero, and are adjusted for the relative level of the stimulus for each loudspeaker at the center of the array.

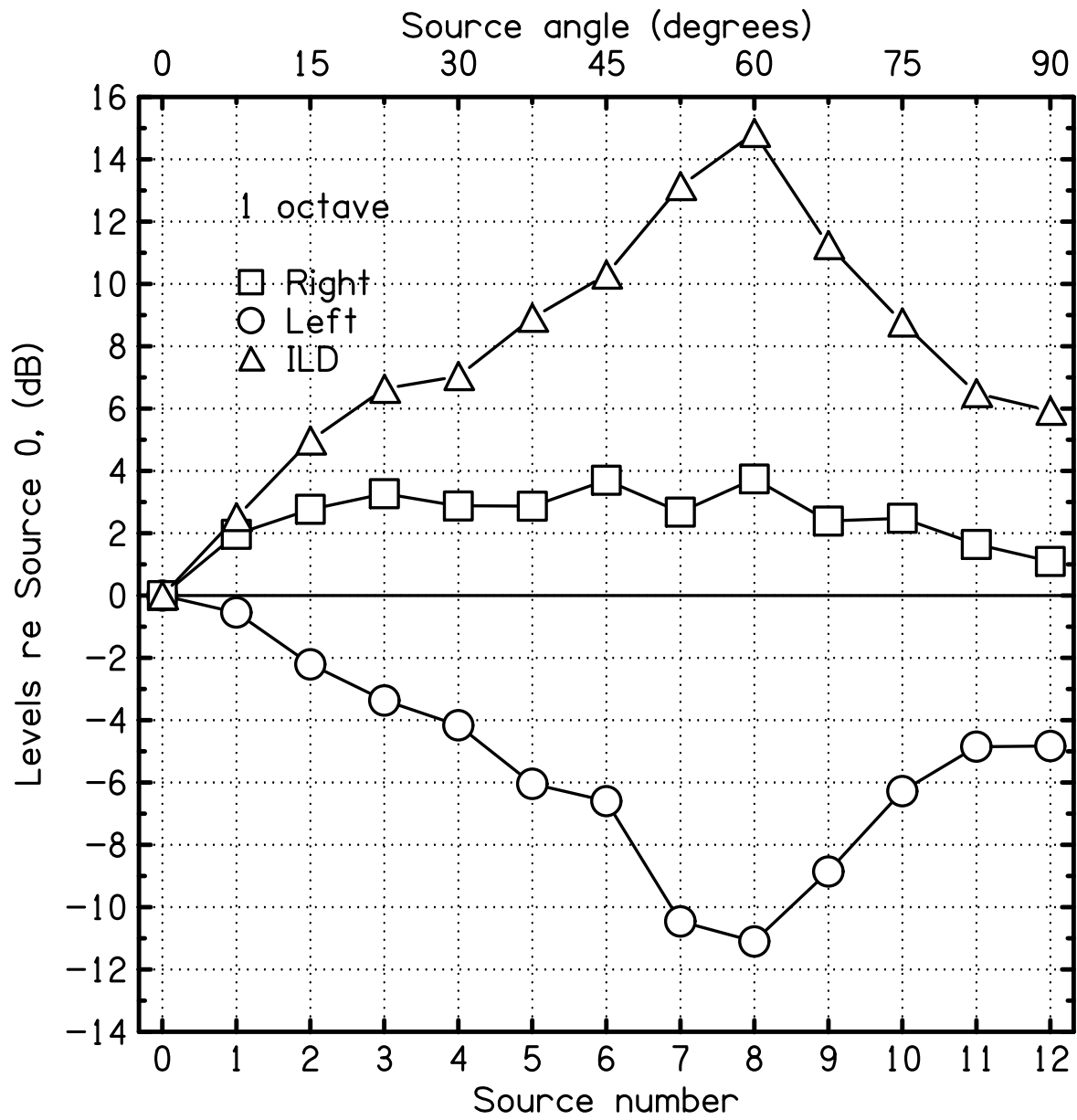


Figure 1.23 Same as Fig. 1.22 but with a 1-octave band.

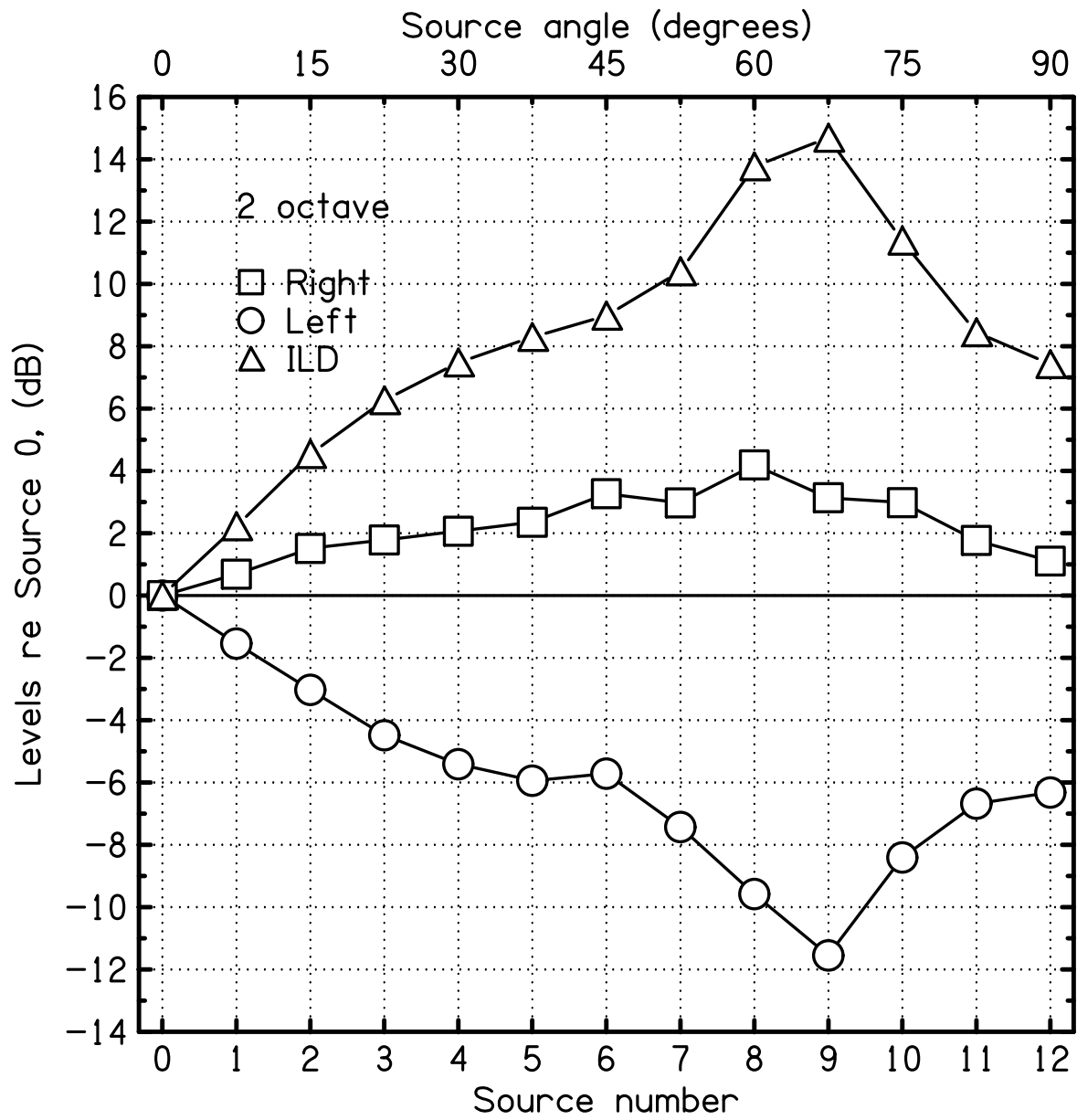


Figure 1.24 Same as Fig. 1.22 but with a 2-octave band.

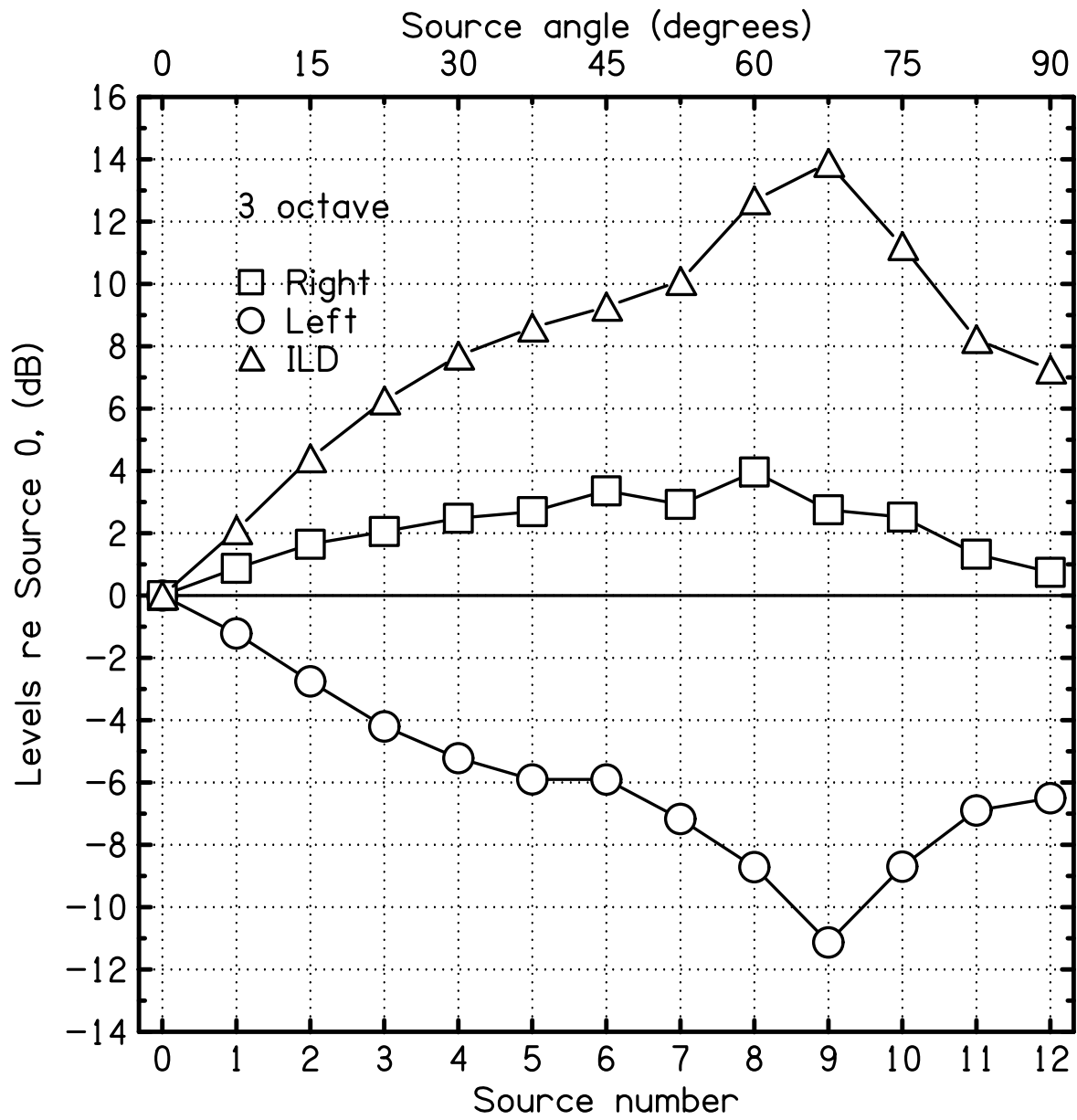


Figure 1.25 Same as Fig. 1.22 but with a 3-octave band.

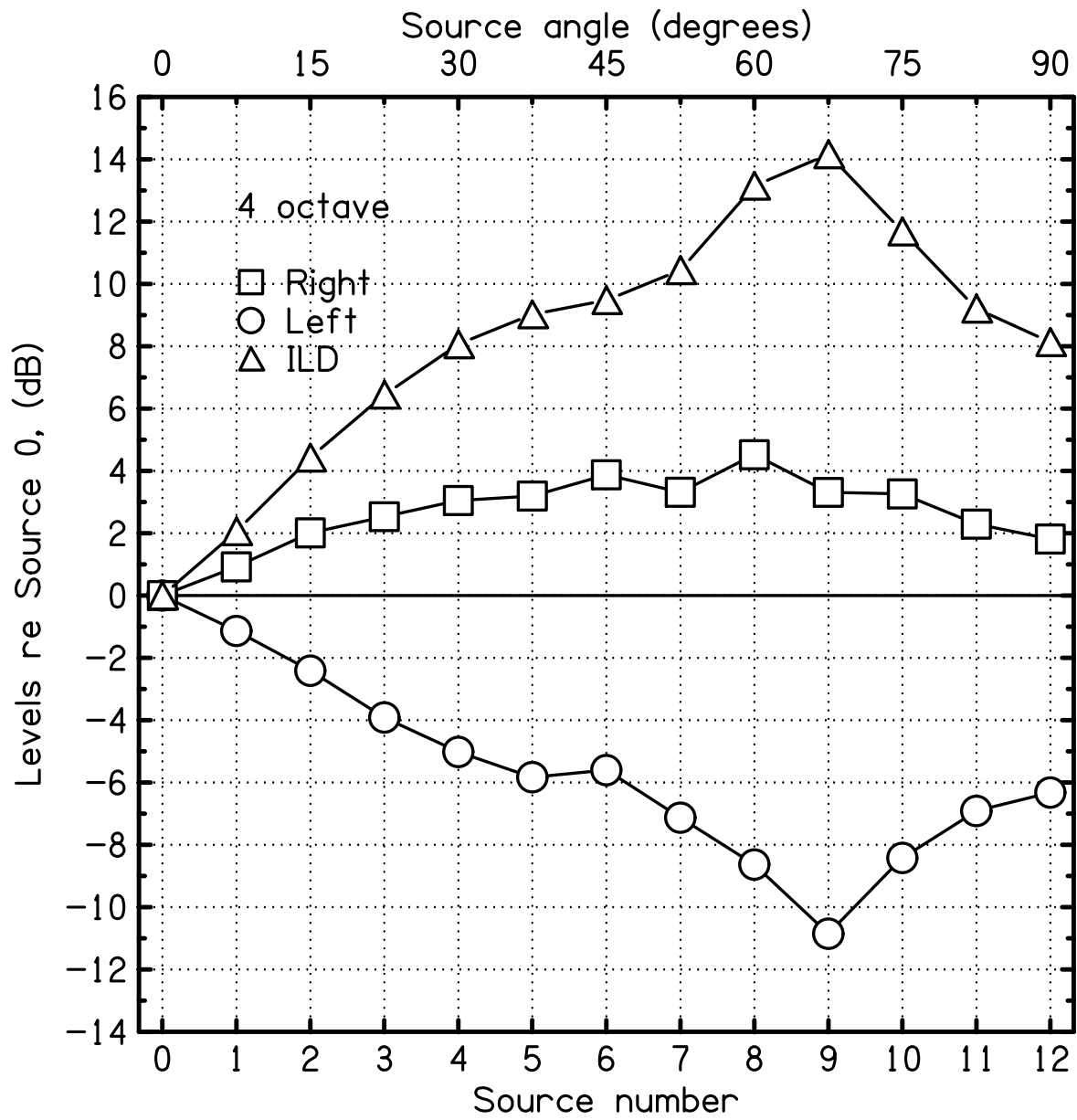


Figure 1.26 Same as Fig. 1.22 but with a 4-octave band.

loudspeaker 7 (52.5°) where the previous measurements shown in Fig. 1.4 show a peak ILD at loudspeaker 8 (60°). Additionally, the ILD in Fig. 1.21 has a maximum value of about 11 dB, where in Fig. 1.4 the maximum value is about 12 dB. Figure 1.21 also shows negative ILDs for loudspeakers 2 and 3. No negative ILD's occur in Fig. 1.4. These differences are likely caused by variations in the repositioning of the loudspeaker array and the KEMAR in the anechoic chamber. All measurements in Figs. 1.21–1.26, however, were made without any repositioning.

Differences between the 1500-Hz sine-tone ILDs in Fig. 1.21 and the one-third octave band in Fig. 1.22 are typically 1 dB or less. This indicates that the ILD of the one-third octave band is effectively equivalent to the sine tone. Figure 1.23 shows the one-octave band. The ILD peak has moved to speaker 8 (60°). The ILD also now falls off more gradually than for the sine-tone. However, a strong monotonic ILD curve is still present for the one-octave band. The two-octave band in Fig. 1.24 continues this trend. The peak has moved to a larger azimuth again. It is now at speaker nine (67.5°). This peak appears to be about as broad as the one-octave peak in Fig. 1.23.

Things change very little as the bandwidth is increased to 3 and 4 octaves in Figs. 1.25 and 1.26, respectively. The 4-octave band spans from 375 to 6000 Hz. Not only is the seriously non-monotonic ILD persisting at such a large bandwidth, it appears to be stabilizing as the bandwidth increases. Although these bands are logarithmically centered around 1500 Hz, where the non-monotonicity poses the most serious problem for listeners, these results suggest that the non-monotonic ILD will manifest in many signals beyond sine-tones or narrow bands. If listeners do perform better in localization as bandwidth increases, it is likely related to the introduction of ITDs or EITDs, rather than due to a change in the ILD.

1.6 Minimum Audible Angle

The results from this experiment may explain some results from Mills' 1958 article "On the Minimum Audible Angle" [34,35]. In this experiment, Mills measured the minimum audible angle (MAA) for sine-tones in an anechoic room. In the minimum audible angle experiment a listener heard two tones in succession. The listener then was required to indicate if the source of the sound moved to the left or the right. For each frequency and azimuth, $\Delta\theta$ was varied and the percent of correct responses was tabulated for each condition. Mills defined the minimum audible angle as being half of the angular separation between 25% correct response and 75% correct responses. This is essentially equivalent to stating that a listener must achieve 75% correct responses at a particular $\Delta\theta$, for this angle to be considered the minimum audible angle.

Figure 1.27 shows Mills' minimum audible angle experiment results. What is interesting is that for higher azimuths (60° and 75°), Mills was unable to find the minimum audible angle from about 1250 Hz to 2000 Hz. The results of this current study would suggest that this is because this is just where the ILD vs. azimuth curve has a negative slope, rather than a positive one. The listeners in this current study had ILD peaks between 55° and 60° for 1500 Hz. The listeners in Mills' experiment were likely perceiving left-moving sounds as right-moving sounds and vice versa. This would explain why Mills was able to find a minimum audible angle at 45° but not at 60° for 1500 Hz. As the frequency increased up to approximately 3000 Hz, Mills was able to measure minimum audible angle again for both 60° and 75° . Figure 1.1 shows for a spherical head model that as frequency increases, so does the peak of the ILD vs. azimuth curve. If a listener's ILD peak lies beyond 75° for 3000 Hz and higher, then it would be reasonable to expect this listener to be able to successfully

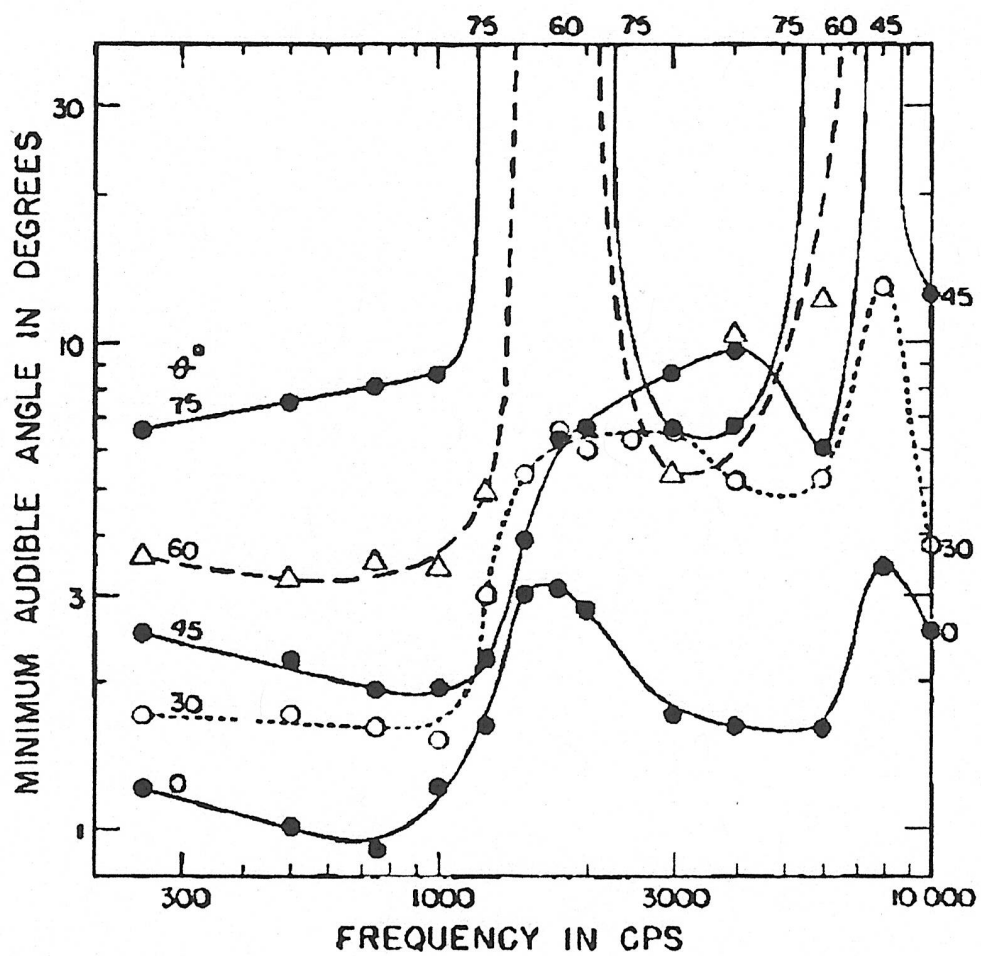


Figure 1.27 Average minimum audible angle from Mills, 1958. The five sets of data are for different azimuths.

localize the sound source.

1.7 Conclusion

Human sound localization in the horizontal plane depends on two interaural cues: the interaural time difference (ITD) and the interaural level difference (ILD). For sine tones with frequencies of 1500 Hz and higher, humans are insensitive to the ITD. The ILD is the only ongoing cue that remains. As the azimuthal angle increases to the right the level at a listener's left ear (far ear) initially decreases because the head creates an acoustical shadow. However, due to diffraction of sound waves around the head, the level in the far ear begins to increase again as the sound source angle approaches the opposite side of the head. This is known as the acoustical bright spot. Because of this, the level in the far ear is non-monotonic with azimuth. Subsequently, the level difference between the two ears is non-monotonic. This is problematic for listeners because a non-monotonic ILD is an ambiguous cue, where two, widely-separated azimuthal angles share the same ILD cue.

Listeners were not able to successfully localize 1500-Hz pure tone stimuli with azimuths greater than about 50 degrees in a free-field environment. Even after training and feedback, listeners were unable use either interaural cues or absolute level information at each ear to successfully localize. When envelope ITD cues were introduced to the signal—either with amplitude modulation or with narrow-band noise—listeners displayed a wide variety of results. One listener localized both stimuli very well and thus seemed to heavily weigh envelope ITD cues. Three other listeners also showed significant localization improvement, but were still affected by ILD cues in addition to the envelope ITD cues. One listener showed no localization improvement and thus continued to heavily weigh ILD only. The

non-monotonic ILD explains the the data from a minimum audible angle experiment by Mills [34, 35]. The minimum audible angle could not be measured at azimuthal angles where the ILD has a negative slope.

Chapter 2

Amplitude Modulation Localization

2.1 Introduction

In Chapter 1 some listeners were able to benefit from the introduction of amplitude modulation (AM). However, other listeners benefited to a lesser extent or not at all. What is responsible for this difference in performance? One possibility is that inter-listener anatomical differences lead to differences in the quality of the envelope interaural time difference (EITD) cues. Alternatively, there may be inter-listener differences in the way that the interaural cues are processed.

In sound localization, a sound source is perceived to be at some location in space. In sound lateralization, which typically occurs when listeners are using headphones, a sound source is perceived to be somewhere along the interaural axis, or the line connecting a listener's ears. Listeners wearing headphones are able to use envelope interaural time differences (EITD) to lateralize stimuli [2]. This is true even at high frequencies where no fine-structure interaural time difference cues are available [3,5,23,30,33]. Based on the results of sinusoidally amplitude-modulated (SAM)-tone lateralization experiments using headphones, it can be predicted that adding envelope fluctuations to signals in free field will have similar results. SAM tones are signals that can be lateralized based on the EITD, and are defined by equation (2.1),

$$x_0(t) = C[1 + m \cos(\omega_m t + \phi_a)] \sin(\omega_c t + \phi_c) \quad (2.1)$$

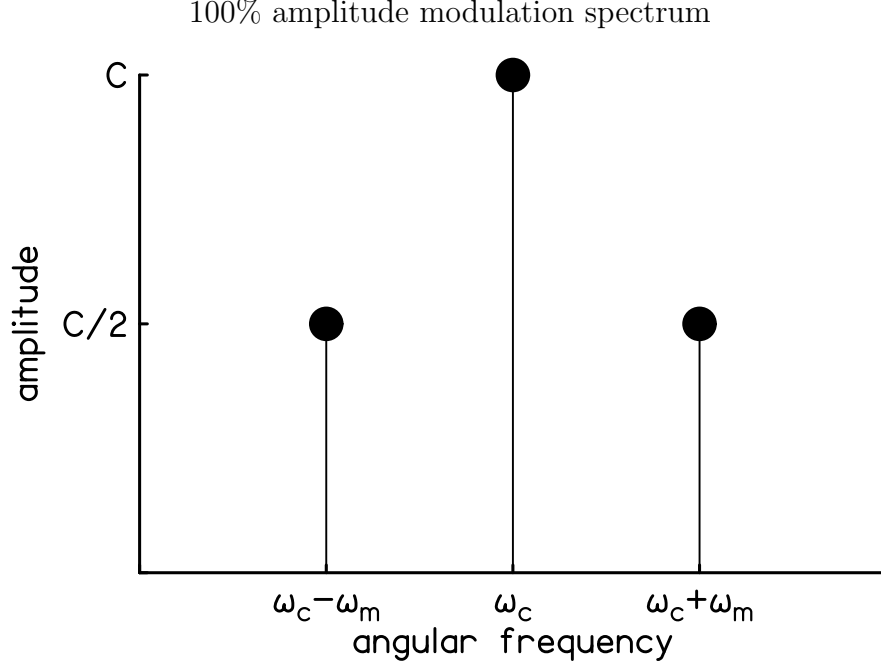


Figure 2.1 Spectrum of 100% amplitude modulation. The carrier frequency, ω_c , has an amplitude of C . The sidebands, at $\omega_c - \omega_m$ and $\omega_c + \omega_m$, have amplitudes of $C/2$. The phase of the lower sideband is $\phi_c - \phi_a$ and the phase of the upper sideband is $\phi_c + \phi_a$.

where C is the amplitude of the carrier, m is the amplitude modulation (AM) percentage, ω_m is the modulation frequency, ϕ_a is the phase of the amplitude modulation, ω_c is the carrier frequency, and ϕ_c is the phase of the carrier. In the case of 100% amplitude modulation, $m = 1$. There are three spectral components as shown in Fig. 2.1.

However, localization of SAM tones may be a different matter than lateralization. Although there have been many headphone experiments with amplitude-modulated stimuli, and a study of localizing amplitude-modulated high-frequency noise in free field [11], this is the first known free-field listening study of SAM (actual localization). Unlike headphone lateralization experiments, in SAM localization experiments, the signals are not presented directly to the listener's ears. Instead they are presented from loudspeaker sources and the sound waves must interact with the listener's anatomy. This results in a mixed modula-

tion signal in the listener's ear canals. In general, mixed modulation includes amplitude modulation and quasi-frequency modulation (QFM) which will be defined shortly.

An experiment was designed to examine the localization of 100% SAM tones in free field. Both the nature of the physical cues, and the perceptual consequences are to be examined. How often do listeners in free field encounter problems with the quality of the AM (m)? Does the degradation of AM quality effect the strength of the EITD cue? How often is the sign or magnitude of the envelope ITD cue misleading to listeners? This topic of envelope ITD localization is particularly important when it comes to cochlear implants [36]. For users of bilateral cochlear implants, the fine-structure ITD cue is unavailable. The only timing cue available to these listeners is the envelope ITD.

To obtain an expression for a signal with mixed modulation, first frequency modulation (FM) is introduced to equation (2.1) by considering a time-varying frequency in place of the carrier frequency,

$$\omega_c \rightarrow \omega(t) = \omega_c + \Delta\omega \cos(\omega_m t + \phi_f) \quad (2.2)$$

where $\Delta\omega$ is the frequency excursion and ϕ_f is the phase of the frequency modulation. The argument of the sine function of equation (2.1) is replaced by the phase,

$$\begin{aligned} \int \omega(t) dt &= \int (\omega_c + \Delta\omega \cos(\omega_m t + \phi_f)) dt \\ &= \omega_c t + \frac{\Delta\omega}{\omega_m} \sin(\omega_m t + \phi_f) + \phi_c. \end{aligned} \quad (2.3)$$

The general expression for a mixed-modulation signal is then,

$$x_{\text{MM}}(t) = C[1 + m \cos(\omega_m t + \phi_a)] \sin[\omega_c t + \beta \sin(\omega_m t + \phi_f) + \phi_c] \quad (2.4)$$

where $\beta = \frac{\Delta\omega}{\omega_m}$. However, for a linear system with a listener in free field, the signals arriving in the ear canals only contain spectral components from the three frequencies in the original spectrum. When equation (2.4) is expanded, only terms to first order in m and β should be kept. The resulting signal is

$$x(t) = C \sin(\omega_c t + \phi_c) + C m \cos(\omega_m t + \phi_a) \sin(\omega_c t + \phi_c) + C \beta \sin(\omega_m t + \phi_f) \cos(\omega_c t + \phi_c) \quad (2.5)$$

which only consists of the original spectral components ω_c , $\omega_\ell = \omega_c - \omega_m$, and $\omega_u = \omega_c + \omega_m$ where ω_ℓ is the lower sideband, and ω_u is the upper sideband. QFM with no AM occurs when $m = 0$.

To determine the nature of the waveforms at a listener's ear canals, small probe microphones (Etymotic ER-7C) were used. Small flexible probe tubes were inserted into a listener's ear canals. Because the tubes are thin—an outer-diameter of 0.95 mm—the voltages corresponding the desired signal were very small and had a noticeable amount of unwanted electrical noise. Because of this electrical noise, and also any unwanted acoustical noise, it was potentially difficult to determine the nature of the mixed-modulation signals arriving at a listeners's ear canals. Fortunately, because the frequencies of the target SAM signal were known, a matched-filtering routine can extract only the spectral components of interest.

Equation (2.5) can be written in terms of the lower, carrier, and upper frequencies as shown in equation (2.6).

$$\begin{aligned} x(t) = & A_c \cos \omega_c t + B_c \sin \omega_c t \\ & + A_\ell \cos \omega_\ell t + B_\ell \sin \omega_\ell t \\ & + A_u \cos \omega_u t + B_u \sin \omega_u t \end{aligned} \quad (2.6)$$

The digital matched-filtering routine calculates the coefficients, A_c , B_c , A_ℓ , B_ℓ , A_u , and B_u , which in the analytical realm can be obtained from Fourier transform integrals on the pure waveform, $x(t)$, as shown in equations (2.7)–(2.12).

$$A_\ell = \frac{2}{T} \int_0^T x(t) \cos \omega_\ell t \, dt \quad (2.7)$$

$$B_\ell = \frac{2}{T} \int_0^T x(t) \sin \omega_\ell t \, dt \quad (2.8)$$

$$A_c = \frac{2}{T} \int_0^T x(t) \cos \omega_c t \, dt \quad (2.9)$$

$$B_c = \frac{2}{T} \int_0^T x(t) \sin \omega_c t \, dt \quad (2.10)$$

$$A_u = \frac{2}{T} \int_0^T x(t) \cos \omega_u t \, dt \quad (2.11)$$

$$B_u = \frac{2}{T} \int_0^T x(t) \sin \omega_u t \, dt \quad (2.12)$$

The relationships between these coefficients and C , m , β , ϕ_c , ϕ_a , and ϕ_f are as follows in equations (2.13)–(2.24).

$$C = \sqrt{A_c^2 + B_c^2} \quad (2.13)$$

$$\tan \phi_c = \frac{A_c}{B_c} \quad (2.14)$$

$$\tan \phi_a = \frac{(A_u - A_\ell) \cos \phi_c - (B_u - B_\ell) \sin \phi_c}{(B_u + B_\ell) \cos \phi_c + (A_u + A_\ell) \sin \phi_c} \quad (2.15)$$

$$\tan \phi_f = \frac{(A_u + A_\ell) \cos \phi_c - (B_u + B_\ell) \sin \phi_c}{(B_u - B_\ell) \cos \phi_c + (A_u - A_\ell) \sin \phi_c} \quad (2.16)$$

The solutions for m and β are as follows.

$$m = \frac{2}{C} \frac{B_u \sin(\phi_c + \phi_f) - A_u \cos(\phi_c + \phi_f)}{\sin(\phi_f - \phi_a)} \quad (2.17)$$

$$\beta = \frac{2}{C} \frac{A_u \cos(\phi_c + \phi_a) - B_u \sin(\phi_c + \phi_a)}{\sin(\phi_f - \phi_a)} \quad (2.18)$$

In the case where $\sin(\phi_f - \phi_a) = 0$, m and β become

$$m = \frac{A_u}{C \sin(\phi_c + \phi_a)} + \frac{A_\ell}{C \sin(\phi_c - \phi_a)} \quad (2.19)$$

and

$$\beta = \frac{A_u}{C \sin(\phi_c + \phi_a)} + \frac{A_\ell}{C \sin(\phi_c - \phi_f)}. \quad (2.20)$$

When both $\sin(\phi_f - \phi_a) = 0$ and $\sin(\phi_c + \phi_a) = 0$, the solution for m is

$$m = \frac{B_u}{C \cos(\phi_c + \phi_a)} - \frac{B_\ell}{C \cos(\phi_c - \phi_a)}. \quad (2.21)$$

When both $\sin(\phi_f - \phi_a) = 0$ and $\sin(\phi_c - \phi_a) = 0$, the solution for m is

$$m = \frac{B_u}{C \cos(\phi_c - \phi_a)} + \frac{B_\ell}{C \cos(\phi_c + \phi_a)}. \quad (2.22)$$

When both $\sin(\phi_f - \phi_a) = 0$ and $\sin(\phi_c + \phi_f) = 0$, the solution for β is

$$\beta = \frac{B_u}{C \cos(\phi_c + \phi_f)} - \frac{B_\ell}{C \cos(\phi_c - \phi_f)}. \quad (2.23)$$

When both $\sin(\phi_f - \phi_a) = 0$ and $\sin(\phi_c - \phi_f) = 0$, the solution for β is

$$\beta = \frac{B_u}{C \cos(\phi_c - \phi_f)} - \frac{B_\ell}{C \cos(\phi_c + \phi_f)}. \quad (2.24)$$

It is the interaural time difference in the envelope of the signals that is of perceptual

importance at high frequencies [23]. The envelope, $E(t)$, can be obtained by calculating the Hilbert transform of the signal, \hat{x} , and adding it to the original signal in quadrature as shown in equation (2.25) [16].

$$E^2(t) = x^2(t) + \hat{x}^2(t) \quad (2.25)$$

Then, from the mixed-modulation signal from equation (2.6), the envelope is given by

$$\begin{aligned} E^2(t) = & [A_c + (A_u + A_\ell) \cos \omega_m t + (B_u - B_\ell) \sin \omega_m t]^2 \\ & + [B_c + (B_u + B_\ell) \cos \omega_m t + (A_\ell - A_u) \sin \omega_m t]^2. \end{aligned} \quad (2.26)$$

2.1.1 Problem 1: AM Quality

Figure 2.2 shows an SAM tone with 100% AM, $m = 1$. However, due to anatomical filtering in free field, a listener will not generally hear a 100% SAM tone with $m = 1$. There will generally be some QFM present in the signal. A listener should experience over-modulation ($m > 1$) or under-modulation ($m < 1$). To what extent there is over- or under-modulation will depend on an individual's anatomy, the sound source angle, the frequency, and the modulation frequency. The depth of modulation—particularly if there is under-modulation—may have an effect on a listener's ability to utilize the EITD cue to localize the source of sound. Furthermore, there is no reason to expect that the mixed modulation parameters, including m , will be the same in a listener's left and right ears. In this situation, the left and right signals may become too dissimilar for the envelope ITDs to be of use [37].

For a pure SAM tone from (2.1), the depth of modulation is determined only by m . However, in mixed modulation the QFM, β , can also affect the overall depth of modulation and may be perceptually important. So it is useful to consider another measure of the depth of modulation. The envelope modulation fraction, m_e in equation (2.27), is the same as

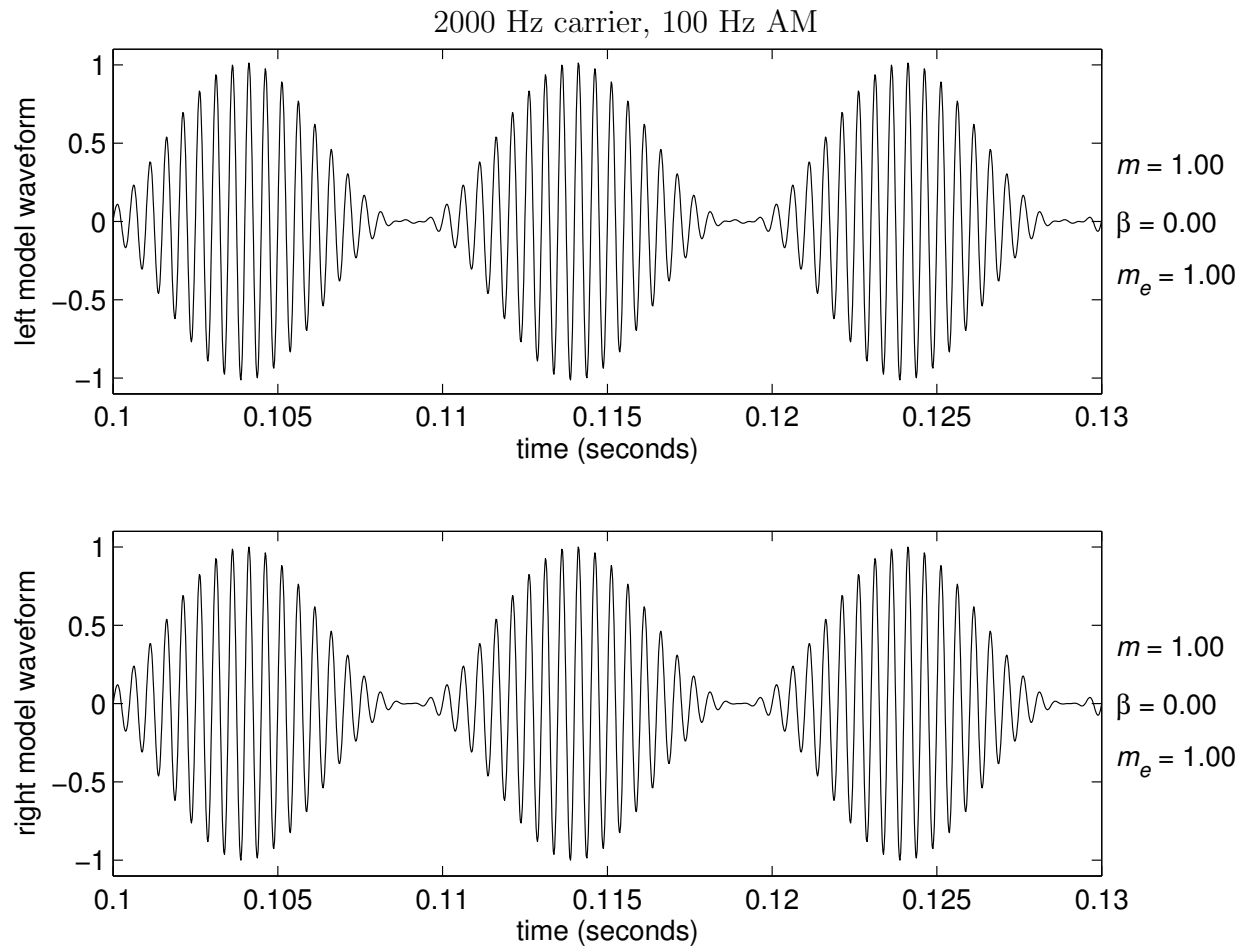


Figure 2.2 SAM signal with 100% AM. The carrier frequency is 2000 Hz and the modulation frequency is 100 Hz. No QFM is present in this signal. In this example, the signals in the two ears are identical.

m for a pure SAM tone of arbitrary modulation depth. It is the ratio of the difference between the maximum and minimum values of the envelope and twice the amplitude of the carrier frequency. The envelope is never negative, but can be zero for cases of 100% AM or over-modulation. Over-modulation results in a secondary peak in the envelope between the primary peaks. Although this peak may be perceptually relevant, it is not explicitly included in the calculation of equation (2.27).

$$m_e = \frac{\max(E) - \min(E)}{2C} \quad (2.27)$$

The degradation of AM quality could potentially pose a problem to listeners for two reasons. First, if the signal is under-modulated, the EITD cue will likely be weaker and become non-existent as the modulation approaches zero. Second, the shape of the envelope may not be the same in a listener's left and right ears. How might the binaural system process and compare the time difference between two envelopes if they have different shapes? Figure 2.3 shows an example of the waveform for a 2000-Hz carrier signal. In this interval, the signal in the listener's left ear is under-modulated, while the signal in the listener's right ear is over modulated. Figure 2.4 shows another example of a waveform with a 3000-Hz carrier signal. In this interval, an asymmetrical modulation pattern is seen in both ears. The left ear signal contains a large amount of QFM, with $\beta = 0.56$. These mixed modulation waveform examples—which differ substantially from SAM signals listeners heard in headphone experiments—are not necessarily representative of a randomly selected recording from this experiment, but do illustrate the some of the more unusual binaural signals a listener will encounter when listening to SAM tones in free field.

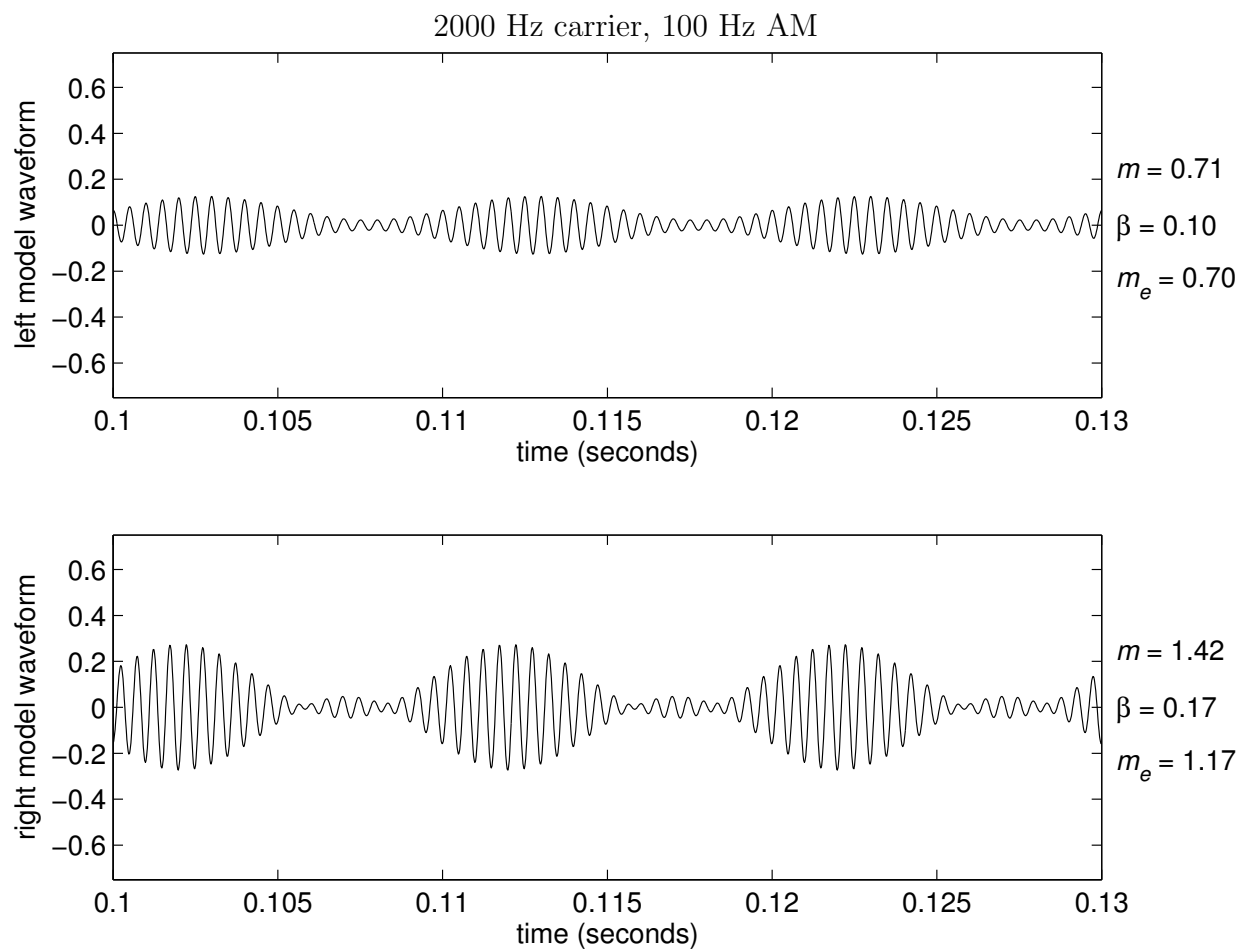


Figure 2.3 Left and right waveforms of arbitrary units calculated from ear canal recordings. The carrier frequency was 2000 Hz and the modulation frequency was 100 Hz. Values for the amplitude modulation, QFM, and envelope modulation fraction are shown on the right by m , β , and m_e , respectively.

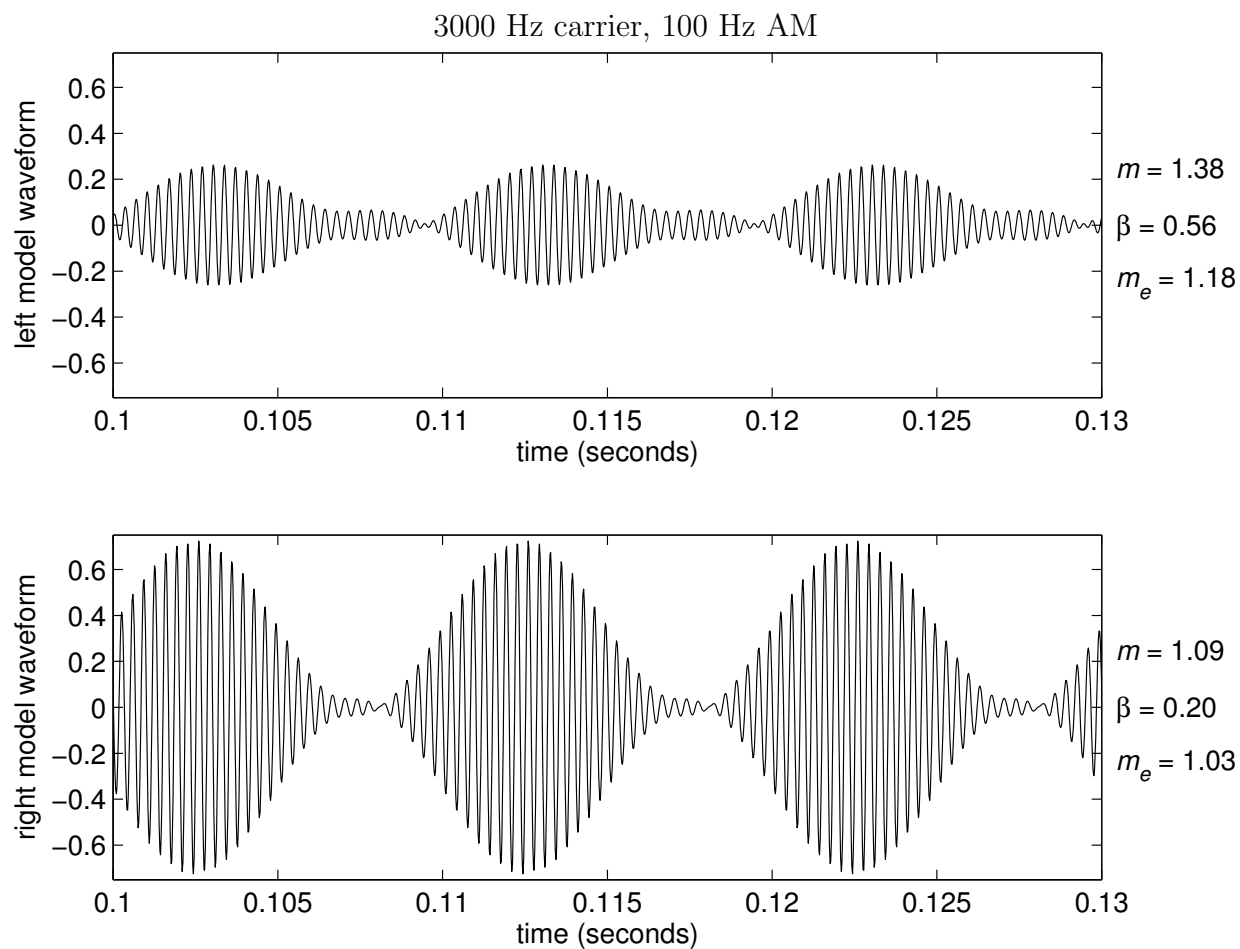


Figure 2.4 Left and right waveforms of arbitrary units calculated from ear canal recordings.. The carrier frequency was 3000 Hz and the modulation frequency was 100 Hz. Values for the amplitude modulation, QFM, and envelope modulation fraction are shown on the right by m , β , and m_e , respectively.

2.1.2 Problem 2: EITD Group Delay

The fine-structure ITD is the phase delay of the interaural phase difference, shown by the simple relationship in equation (2.28).

$$\text{ITD} = \frac{1}{\omega} \text{IPD} \quad (2.28)$$

The ITD and IPD relationship is linear and monotonic, so the fine-structure cue is always consistent in sign and magnitude with the IPD and ITD. The envelope ITD cue, however, is the group delay of the IPD [26] as shown in equation (2.29)

$$\text{EITD} = \frac{d}{d\omega} \text{IPD} \quad (2.29)$$

The envelope ITD cue depends on the slope of the IPD with respect to frequency. There is no reason to expect a priori that this function will be linear or monotonic for a human head in free field. Therefore, the EITD cue may sometimes be incorrect in sign for a given azimuth. This would also mean that the EITD cue may not increase monotonically with azimuth and may mislead listeners.

2.1.2.1 Problem 2: KEMAR Group Delay

Measurements were made in an anechoic room with the KEMAR manikin described in Chapter 1. Figure 2.5 shows the wrapped IPDs measured over a frequency range of 1800–2800 Hz with a sound source at an azimuth of +60 degrees. On the whole, the IPD increases with frequency, but the increase is non-monotonic. There are ranges where the slope is very steep, such as line A, and ranges where the slope is negative, such as line B. The EITD of

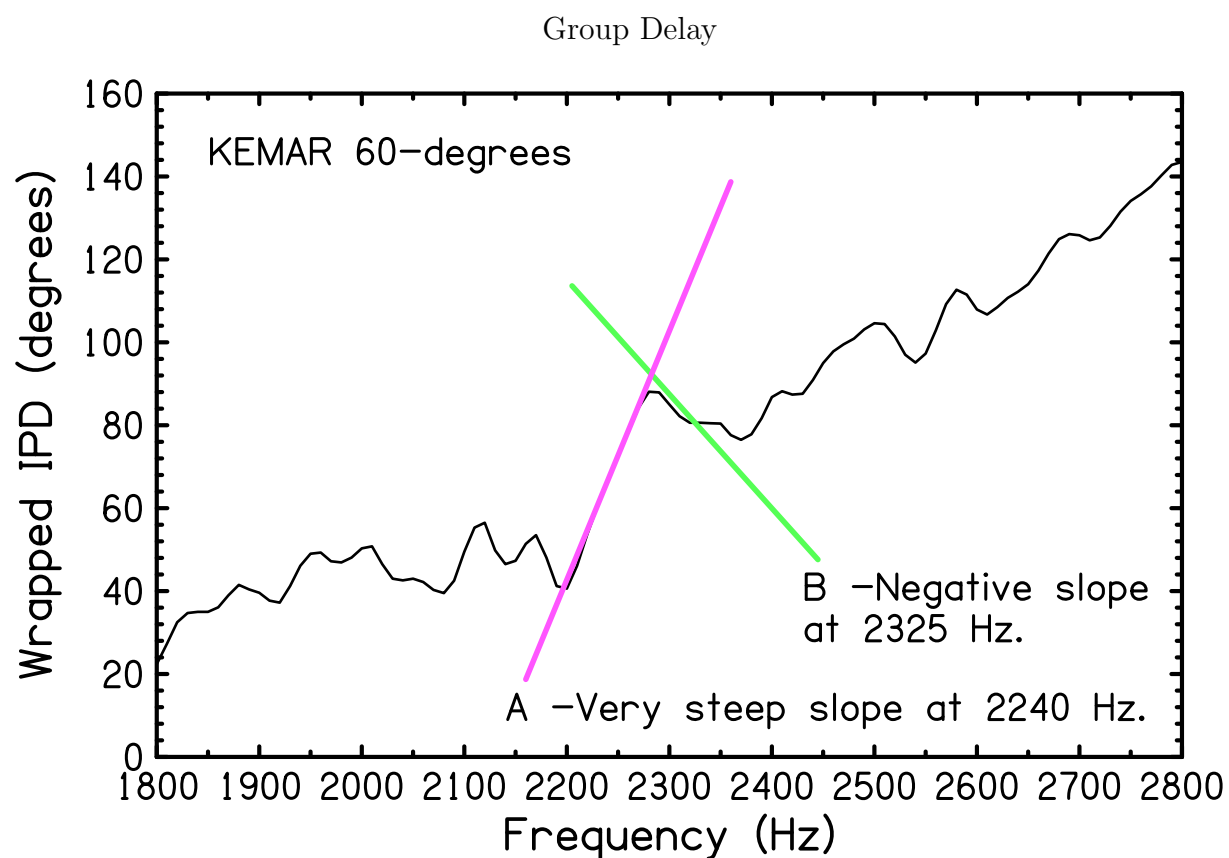


Figure 2.5 Wrapped IPD (-180° to $+180^\circ$) vs. carrier frequency. Measurements were made in the anechoic room with the 1-m array. The azimuth was 60° . The signals were sine tones at 10 Hz intervals.

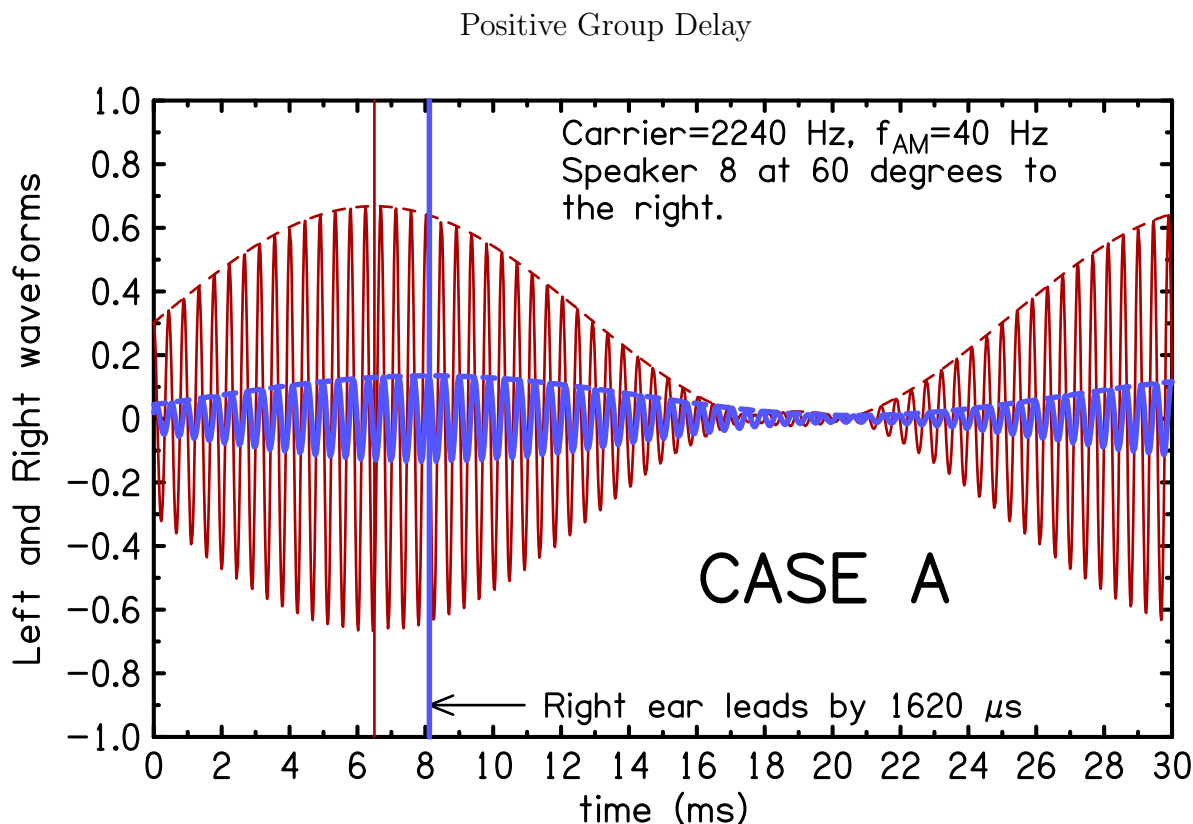


Figure 2.6 Left (blue) and right (red) waveforms and envelopes of KEMAR recordings of SAM tones in free field. The signal was a 2240 Hz carrier with a modulation frequency of 40 Hz.

an SAM tone is the group delay, or the slope of the IPD.

To investigate regions A and B, SAM tones with a modulation frequency, $f_m = 40$ Hz, were presented from an azimuth of 60° . Figure 2.6 shows the left and right waveforms for the SAM tone with a carrier frequency of $f_c = 2240$ Hz, corresponding to line A. The envelope of the right ear (red) is leading the envelope of the left ear (blue), as is expected. Figure 2.7, shows the left and right waveforms for the SAM tone with a carrier frequency of $f_c = 2325$ Hz, corresponding to line B. Here, the envelope of the left ear (blue) is leading the envelope of the right ear (red), even though the the source is on the right. These measurements show that for SAM tones, the EITD can be a misleading cue in free field.

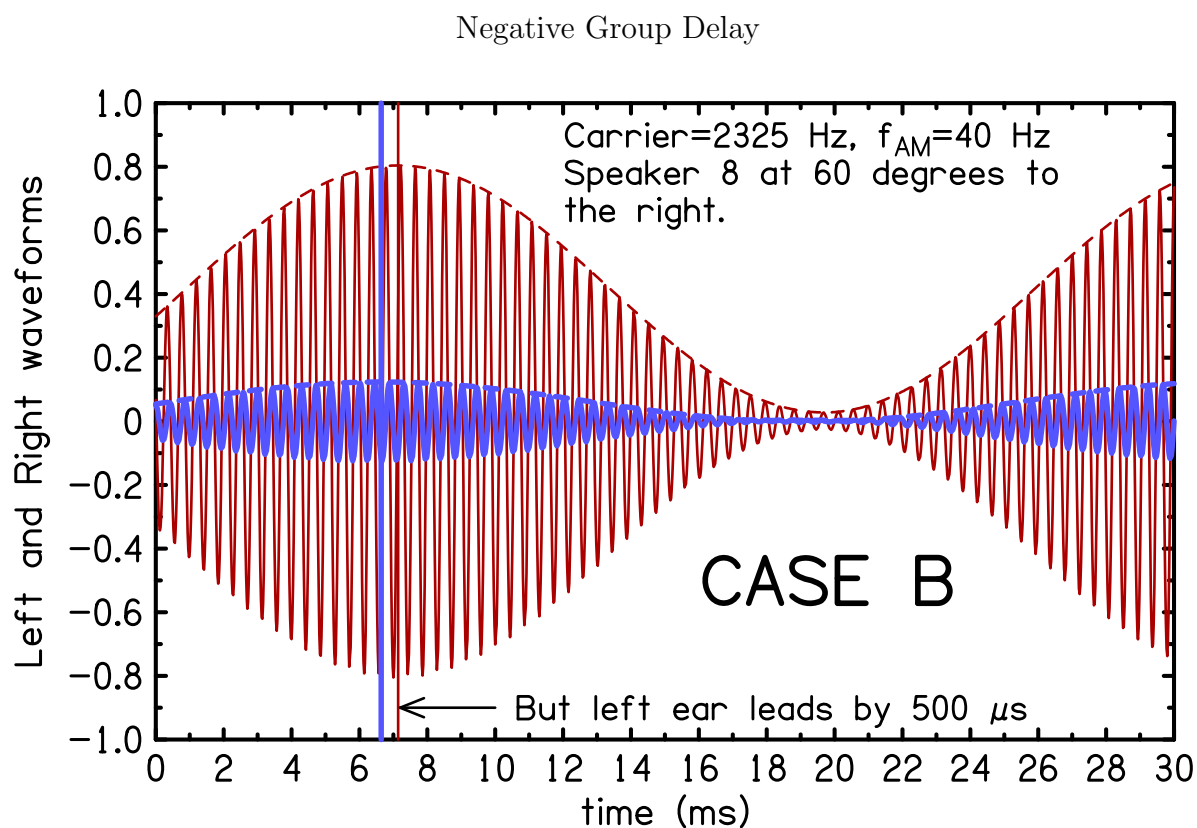


Figure 2.7 Left (blue) and right (red) waveforms and envelopes of KEMAR recordings of SAM tones in free field. The signal was a 2325 Hz carrier with a modulation frequency of 40 Hz.

2.2 Methods

2.2.1 Experimental Setup

The experimental setup was the same as in Chapter 1 except for the following. The construction of the array was different. The grills of the speakers were located 197 cm from the listener, but otherwise had the same angular spacing as shown in Fig. 1.3. For the new array, the loudspeakers were mounted on four pieces of wooden flooring with velcro. The pieces of wooden flooring were supported by microphone stands. The previous array height was slightly above the azimuthal plane for a typical listener, but the new array was lowered so that the loudspeaker cones were in the azimuthal plane for the listeners. This new array was designed to reduce scattering off of the mounting hardware. Additionally, because the array was larger, there should have been less scattering off of neighboring loudspeakers. Finally, the increased distance between the sources and the listener resulted in a better realization of the far-field limit than the old 111.5-cm array [41, 45].

Directly behind the listener was a masking loudspeaker (Boston Acoustics A40). The loudspeaker faced up toward the listener's head. A masker signal, described below, was produced by a CD player.

2.2.2 Experimental Conditions and Procedure

There were six experimental conditions for each listener, defined by different signals that were presented. These consisted of sine tones at 2 kHz, 3 kHz, and 4 kHz, and sinusoidally amplitude modulated tones with the same carrier frequencies and a modulation rate of 100 Hz (100% modulation). The modulation frequency of 100 Hz is near the region around 128 Hz shown to be most sensitive [4, 5, 10]. A critical band is a range of frequencies that is small

enough that auditory processing is not independent for different frequencies. The 100-Hz modulation frequency was small enough to ensure that all of the spectral components were in the same critical band for each of the three carrier frequencies [13,24,25]. Therefore the SAM tones were perceived as single acoustical entities. The stimuli had a total duration of 1000 ms. There was a 250-ms linear rise/fall at the beginnings and ends on the signals. The target stimuli had an average level of 65 dBA as measured at the location of the listener. In each trial the level roved randomly by +2, +1, 0, -1, or -2 dB. This was done to ensure that the listener could not use any level differences between individual loudspeakers or non-linear distortion products to identify the source.

The masker signal was intended to mask difference-tone distortion products at 100 and 200 Hz, which might be an alternative localization cue. It consisted of frequencies with an even number of hertz from 50 to 250 Hz. The components had equal amplitude and random phase. This 0.5-s period signal was played continuously throughout the course of a run with a level of 50 dBC as measured at the location of the listener's head. To measure the power of the masker tone in the listeners' ear canals, each run began with a silent trial where recordings were made while only the masker was on.

At the beginning of each run, a calibration sine tone from speaker zero was turned on while the experimenter viewed the probe microphone signals on an oscilloscope. The experimenter instructed the listener to adjust his or her head to ensure that the fine-structure IPD was as close to zero as possible. This was done under the constraint that the listener felt confident that he or she was facing speaker zero. Additionally, the experimenter adjusted the gains on the preamplifier to ensure that the ILD was within 1 dB of zero.

A run consisted of 5 random passes through the 13-loudspeaker array. In each trial, there were 2 intervals presented by the same loudspeaker. There was a pause of about 1 second

between the 2 intervals. The listener was then asked to verbally respond with a loudspeaker number. The diagram in Fig. 1.3 was located just below speaker zero so that the listener could judge speaker numbers without turning his or her head. Listeners were allowed to respond with negative speaker numbers if the source was perceived to be on their left, and with source numbers greater than 12 (or less than -12) if the source was perceived to be behind them.

Each listener completed 2 runs for each stimuli, which resulted in 10 trials and 20 binaural recordings for each stimuli/loudspeaker combination.

2.2.3 Listeners

There were 5 listeners. Listener B was a male aged 59. Listeners C, M, and L were males aged 20–25. Listener V was a female aged 19. Listener B was the only listener that also participated in the experiments from Chapter 1. Listeners M, L, and V had normal hearing thresholds within 15 dB of audiometric zero out to 8 kHz. Listener B had a mild hearing loss typical of males his age, but normal thresholds at the frequencies of these experiments. Listener C had normal hearing thresholds except for about 20 dB of hearing loss in his left ear between 1.5 and 4 kHz.

2.2.4 Computer Analysis of Signals

The analysis of the recordings was limited to the half second from 256 ms to 756 ms, and at a sample rate of 50 kHz, contained 25,000 samples per channel. This eliminated the 250-ms rise/fall times at the beginnings and ends of the signals, and accounted the 6-ms lag time for the sound to travel from the loudspeakers to the listener.

The raw recordings, x_{raw} , contained electrical noise from the pre-amplifiers, acoustical noise—including the continuous noise of the masker—and distortion. To eliminate the effects of the noise and distortion, the signals were digitally filtered. Because the masker had a period of 500 ms, its effect was totally negated by the 500-ms matched-filtering. Using the discrete-time equivalents of equations (2.7)–(2.12), six matched-filtering coefficients were obtained for both ears for each raw recording, x_{raw} . These coefficients were then used to calculate the model waveform, x_{model} , using equation (2.6). The model waveform power, P_{model} , was calculated for each ear with equation (2.30).

$$P_{\text{model}} = A_{\ell}^2 + B_{\ell}^2 + A_c^2 + B_c^2 + A_u^2 + B_u^2 \quad (2.30)$$

The power of the noise plus distortion in the raw recordings was defined as

$$P_{N+D, \text{ raw}} = \frac{1}{25000} \sum_{t=1}^{25000} (x_{\text{raw}}[t] - x_{\text{model}}[t])^2. \quad (2.31)$$

Since the raw recordings also contained noise from the masker and some pre-amplifier electrical noise, the power of the masker recordings, P_{masker} , was subtracted from the power of the noise plus distortion in the raw recordings. This left the noise and distortion power only due to loudspeaker distortion or other sources of sound in the room, P_{N+D} .

$$P_{N+D} = \frac{1}{25000} \sum_{t=1}^{25000} (x_{\text{raw}}[t] - x_{\text{model}}[t])^2 - P_{\text{masker}} \quad (2.32)$$

Finally, the percentage of noise plus distortion was calculated as

$$100(N + D) = \frac{P_{N+D}}{P_{\text{model}}}. \quad (2.33)$$

Recordings and responses for trials where the noise plus distortion was greater than or equal to 10% were discarded from further analysis.

In addition to the model waveform and the model envelope, C , m , β , ϕ_c , ϕ_a , and ϕ_f were calculated using the coefficients and equations (2.13)–(2.24) and m_e using (2.27).

The model ILDs were calculated using model powers from equation (2.30).

$$\text{ILD} = 10 \log \frac{P_{\text{model, right}}}{P_{\text{model, left}}} \quad (2.34)$$

The value of the ILD for zero degrees azimuth was subtracted from the ILD for every azimuth. The ILDs as recorded had a small offset because of the unknown difference in gain between the left and right channels.

The model envelopes were calculated using the matched-filtering coefficients and equation (2.26). Envelope ITDs were calculated using a cross correlation, $\gamma(\tau)$, as a function of lag time, τ , between the left and right model envelopes, E_ℓ and E_r , respectively.

$$\gamma[\tau] = \begin{cases} \frac{\sum_{t=1}^{25000-\tau} E_\ell[t+\tau] E_r[t]}{\sqrt{\sum_{t=1}^{25000} E_\ell^2[t] \sum_{t=1}^{25000} E_r^2[t]}} & \tau \geq 0 \\ \frac{\sum_{t=1}^{25000+\tau} E_\ell[t] E_r[t+\tau]}{\sqrt{\sum_{t=1}^{25000} E_\ell^2[t] \sum_{t=1}^{25000} E_r^2[t]}} & \tau < 0 \end{cases} \quad (2.35)$$

Given that the modulation frequency is $f_m = \omega_m/2\pi = 100$ Hz, and that the carrier frequencies are multiples of 100 Hz, the period of the signals and envelopes should be $T = 1/f_m$.

Neglecting the small end-effects of the finite cross-correlation calculation, $\gamma[\tau]$ must also have a period of T . Therefore, the maximum of $\gamma[\tau]$ was searched for over a range of $-T/2 \leq \tau \leq T/2$. The value of $\gamma[\tau]$ at the peak is the envelope coherence, and the corresponding indexed time, τ , is the envelope ITD.

2.3 Results and Analysis

A small number of trials were removed from the analysis due to noise plus distortion percentages that were greater than or equal to 10% in either ear of either interval of the trial. In all, 9 trials out of a grand total of 3,900 were removed. The means and standard deviations of the responses were calculated for each loudspeaker across the two runs for a particular listener, frequency, and waveform (usually 10 responses). Similarly, the means and standard deviations of all physical quantities from the ear canal recordings were calculated for each loudspeaker across the two runs of a particular listener, frequency, and waveform (usually 20 intervals). In cases where a listener responded with a source location from behind, the recorded responses were flipped to the front symmetrically across speaker 12 at 90° azimuth.

2.3.1 Physical Results

2.3.1.1 Problem 1: AM Quality Results

The amplitude modulation, m , varied somewhat depending on listener and frequency. Most of the values remained close to unity. The left ear, which was on the opposite side of the loudspeaker array, experienced more variation than the right ear. The variation also tended to be greatest when the left ear was in the acoustical dark spot. This would correspond to the azimuth where the level in the left ear was the smallest, and corresponds well to

the location of the peak of the ILD. Near this azimuth, the intensity changes greatly with azimuth, so it is likely that the three frequency components of the SAM tone experienced the greatest amount of disparity in diffraction here as well. Therefore m deviated from 1 the most in these cases.

Figure 2.8 is a histogram of the amplitude modulation, m , for all loudspeakers, listeners, and frequencies. Although the modal value for m is near 1 for these distributions, there are large number of cases with a notable deviation from 1, although never much lower than about 0.75 or higher than 1.75. The left ear exhibits more over-modulation, including one outlier near 3. The bulk of the left (far) ear distribution is wider than the right (near). This is likely because the magnitudes and phases of the three frequency components become more scrambled when diffracting around the listener’s head.

Table 2.1 indicates how frequently over-modulation occurred in each ear for the three carrier frequencies and across all frequencies. Both ears experienced some degree over-modulation the majority of the time. At lower carrier frequencies, the over-modulation occurred more frequently. This may be understood due to the fact that the 100-Hz spacing between the frequency components was a smaller ratio of the carrier frequency at higher frequencies. Therefore, assuming constant-Q diffraction, the three components would have been diffracted in a more similar manner at higher carrier frequency, and the original waveform would have been better preserved.

Figure 2.9 is a histogram of the QFM, β , for all listeners and all frequencies. The histograms for both ears peak around 0.1 to 0.2, and there are not many instances where β exceeds 0.4. However, the QFM in the left (far) ear frequently deviates drastically from zero—the value of β in the signal played by the loudspeakers. Given that the left ear is on the far side of the head for each loudspeaker except the front loudspeaker, this is not

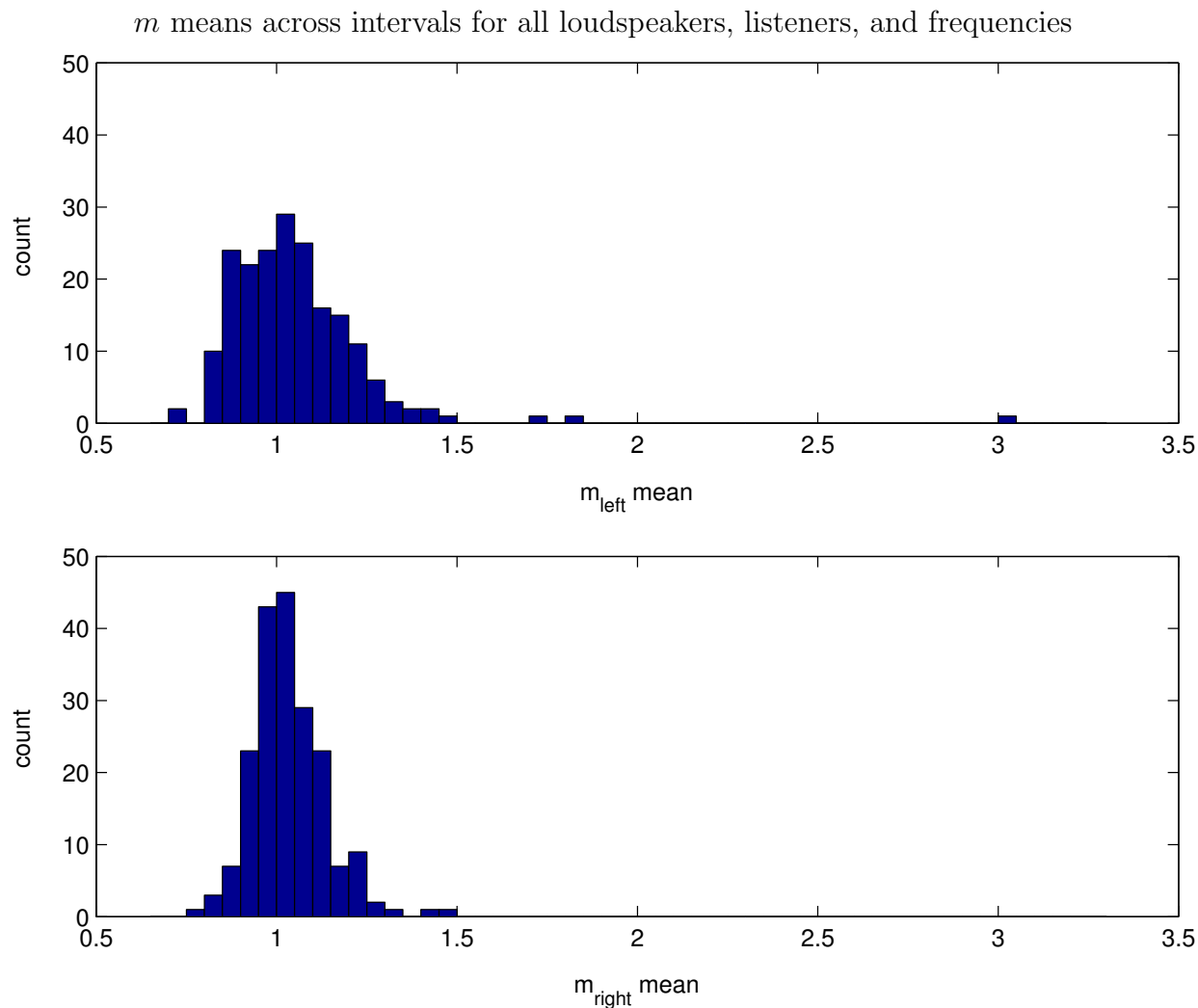


Figure 2.8 Amplitude modulation, m , means for each loudspeaker, frequency, and listener. Histograms for the left and right ears are displayed separately. Each histogram contains a total of 195 values.

Percent of m means $> 100\%$		
carrier frequency	left ear	right ear
2 kHz	71%	86%
3 kHz	69%	66%
4 kHz	23%	38%
all	57%	60%

Table 2.1 The percentage of amplitude modulation, m , means $> 100\%$. The means are calculated across identical interval conditions. The data are combined across all listeners and azimuths.

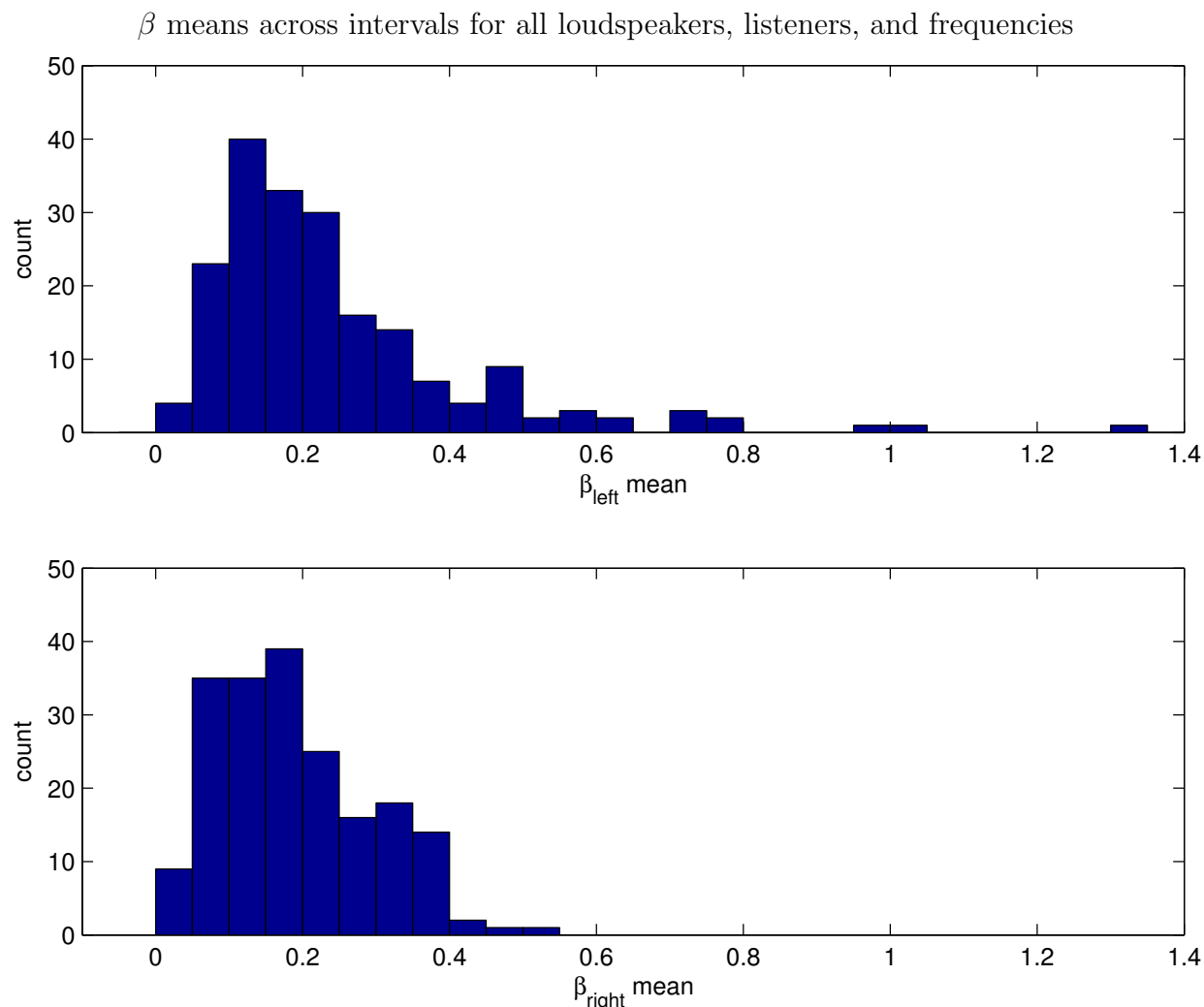


Figure 2.9 QFM, β , means for each loudspeaker, frequency, and listener. Histograms for the left and right ears are displayed separately. Each histogram contains a total of 195 values.

surprising. Although the effects of different azimuths, carrier frequencies, and listeners are not shown in this histogram, what's apparent is that in SAM free-field listening, some degree of QFM is always produced in the ear canals.

2.3.1.2 Problem 2: Envelope ITD Results

The distribution of envelope ITDs in this experiment is shown in Fig. 2.10. The sign of the EITD is substantially positive with a modal value of about $600 \mu s$, which is consistent with

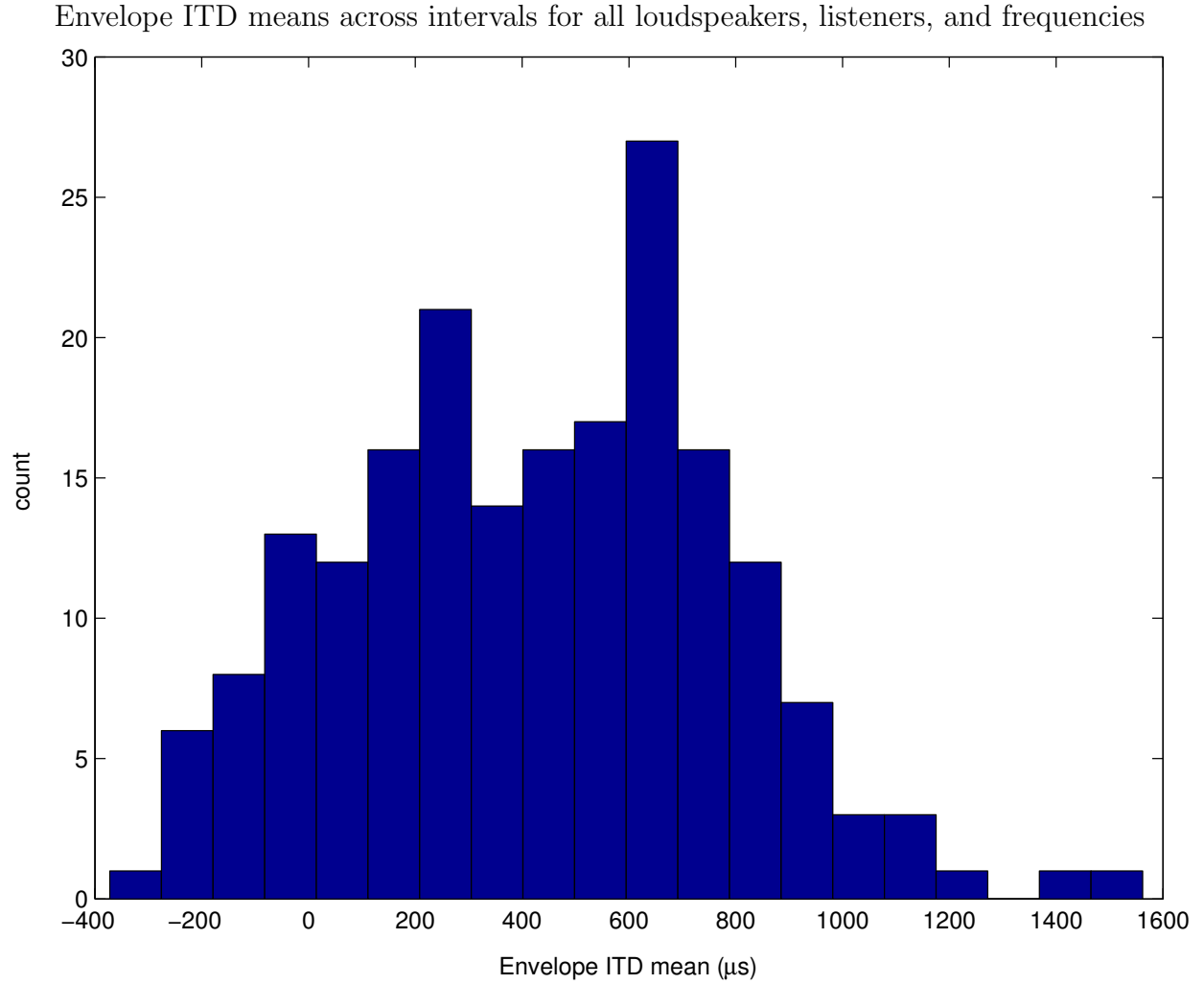


Figure 2.10 Histogram of Envelope ITD means for each loudspeaker, frequency and listener.

the source azimuths. There are, however, a number of negative EITDs, which occurred 14% of the time. These cues have the wrong sign and should therefore be misleading cues to listeners.

2.3.1.3 KEMAR Headphones Measurements

Since there has been a great deal of headphone-based AM research, it is useful to know to what extent the problem of AM quality exists when AM tones are presented over headphones. While the degradation of AM quality in free field appears to be attributable to diffraction

KEMAR SAM Headphone Recordings				
carrier frequency	ear	m	β	m_e
2 kHz	left	1.0156	0.1629	1.0057
	right	1.0072	0.1566	1.0012
3 kHz	left	0.9998	0.0925	0.9974
	right	1.0130	0.1071	1.0053
4 kHz	left	0.9851	0.0627	0.9762
	right	0.9939	0.0439	0.9882

Table 2.2 Values of m , β , and m_e calculated from recordings of KEMAR manikin (large pinna) wearing headphones (Sennheiser HD 535). The signals were 100% SAM tones with carrier frequencies of 2, 3, and 4 kHz, and a modulation frequency of 100 Hz. The monaural parameters were calculated using matched filtering.

around the head, it may be that some degradation could also occur while listening over headphones. For over-ear headphones, this would likely be due to effects from the pinna. In order to determine the magnitude of this effect, recordings were made of SAM tones with the KEMAR manikin’s own ears (large pinna) while wearing over-ear headphones (Sennheiser HD 535). The 2, 3, and 4 kHz SAM tones with 100 Hz AM were presented to both ears.

The results of the KEMAR headphones measurements are shown in Table 2.2. All of the values of m are within 2% of the signal value of 1.0000. Similarly, all of the values of m_e are within 3% of the signal value of 1.0000. By contrast, the QFM, β are quite a bit larger than the signal value of 0. The deviation from 0 appears to decrease as the carrier frequency increases. This could be understood by considering that as the modulation frequency remains constant, the sidebands move closer to the carrier frequency on a logarithmic scale. Therefore the magnitudes and phases of the frequency components are changed less in relation to each other at higher frequencies. While the QFM values measured from the KEMAR wearing headphones are larger than expected, they are not nearly as large as some of the values found in free-field for human listeners (Fig. 2.9).

2.3.2 Perceptual Results

2.3.2.1 Interaural Cues and Responses

Listener responses when localizing the SAM tones and probe microphone measurements made at the same time are shown in Figs. 2.11–2.25. The non-monotonic ILD [31] is observable in the sine and AM trials for each listener and frequency. As predicted by the spherical head model (Chapter 1), as the frequency increases the ILD tends to peak at larger azimuths, and the peak tends to grow as well. In the sine tone trials—(a) plots—listeners are clearly misled by this cue. This is especially evident at large source angles where the ILD is small. The responses for the sine tone are highly correlated with the ILD for all listeners and all frequencies as seen in the (a) plots. This is easily explained by the fact that the ILD is the only available interaural cue available to listeners for these trials.

The ILDs in the AM trials—(b) plots—are similar to the sine-tone trials, but differ somewhat. There are two possible sources of these differences. The first is that the listener orientation can change slightly between trials. It should also be noted that the variability in ILD within plots (a) and (b) may be attributable to changes in listener orientation between the two trials for a given condition. Additionally, a listener may have made small movements during a single run. The second possible source for ILD differences between sine and AM runs is that the signals are different. Although the SAM tone’s carrier frequency is the same as the sine tone, the two sidebands at ± 100 Hz will behave slightly differently. Despite the small differences in ILD, one can see that the trend is that responses do not correlate as well with the AM ILD as they do with the sine ILD. The correlation is still strong, however. A summary of the correlations between response and ILD for AM and sine tones is shown in Fig. 2.30.

The lower correlations in the AM trials may be caused by the availability of an alternative ITD cue—specifically the envelope ITD (EITD). The (c) plots show the EITD along with the same responses from the (b) plots. Like the ILD, the EITD is non-monotonic. The EITD is negative in some cases, although never with a very large magnitude. This was addressed above (Problem 2: the group delay of the IPD being negative in some cases). This tends to occur at smaller azimuthal angles, and the trend of the EITD is to rise with azimuth.

At smaller azimuths, the EITD is usually not very large. At the larger azimuths the EITD tends to reach larger positive values. Larger azimuths are also where the ILD has decreased significantly, whereas the EITD remains relatively large. When comparing responses between the sine tone trials and the AM trials, one can see that the biggest changes tend to occur in these situations. Despite the relatively low correlations between AM responses and EITD, the addition of AM has in some cases improved listeners' most serious errors. The EITD happens to be the most useful exactly where the ILD cue fails. There are two clear examples of this. In Fig. 2.11, at 2 kHz listener B shows improvement for source numbers 11 and 12 where the ILD decreases sharply and the EITD is at its largest. The mean responses for these loudspeakers increase by about 30 degrees. Similarly, in Fig. 2.16 at 4 kHz listener C shows improvement for source numbers 10, 11, and 12.

Listener B, 2 kHz

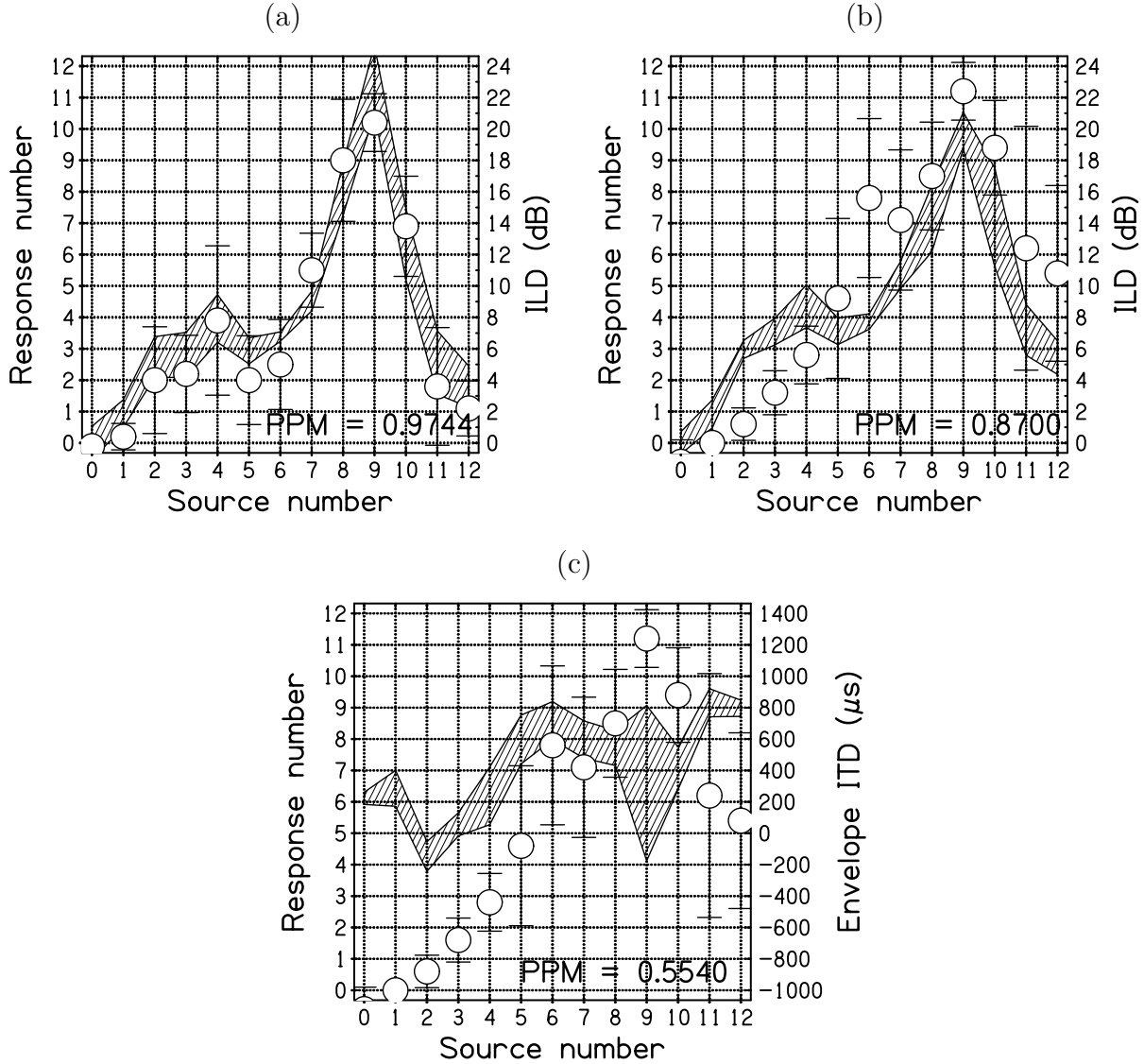


Figure 2.11 Responses and probe-microphone measurements for listener B, 2 kHz. The source numbers span the right front quadrant. The Pearson product-moment (PPM) correlation coefficient is shown between the mean responses and the interaural cues. (a) Sine tone: Circles indicate mean responses and error bars are two standard deviations in overall length. The hatched region shows the ILD. It is centered on the mean and is two standard deviations high. (b) SAM tone: Circles indicate mean responses and error bars are two standard deviations in overall length. The hatched region shows the ILD. It is centered on the mean and is two standard deviations high. (c) SAM tone: Circles indicate mean responses and error bars are two standard deviations in overall length. The hatched region shows the EITD. It is centered on the mean and is two standard deviations high.

Listener B, 3 kHz

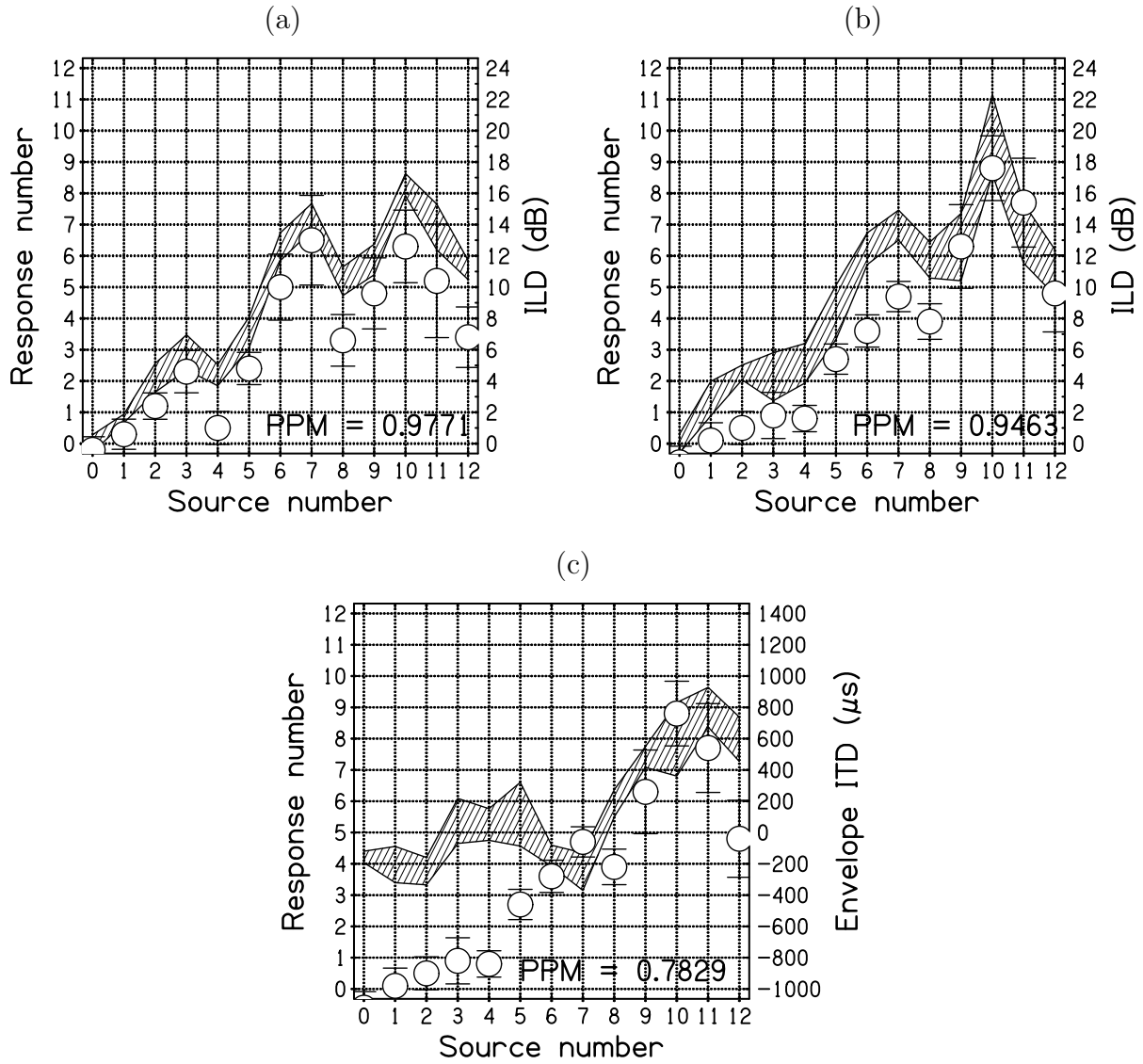


Figure 2.12 Same as Fig. 2.11 but for listener B at 3 kHz.

Listener B, 4 kHz

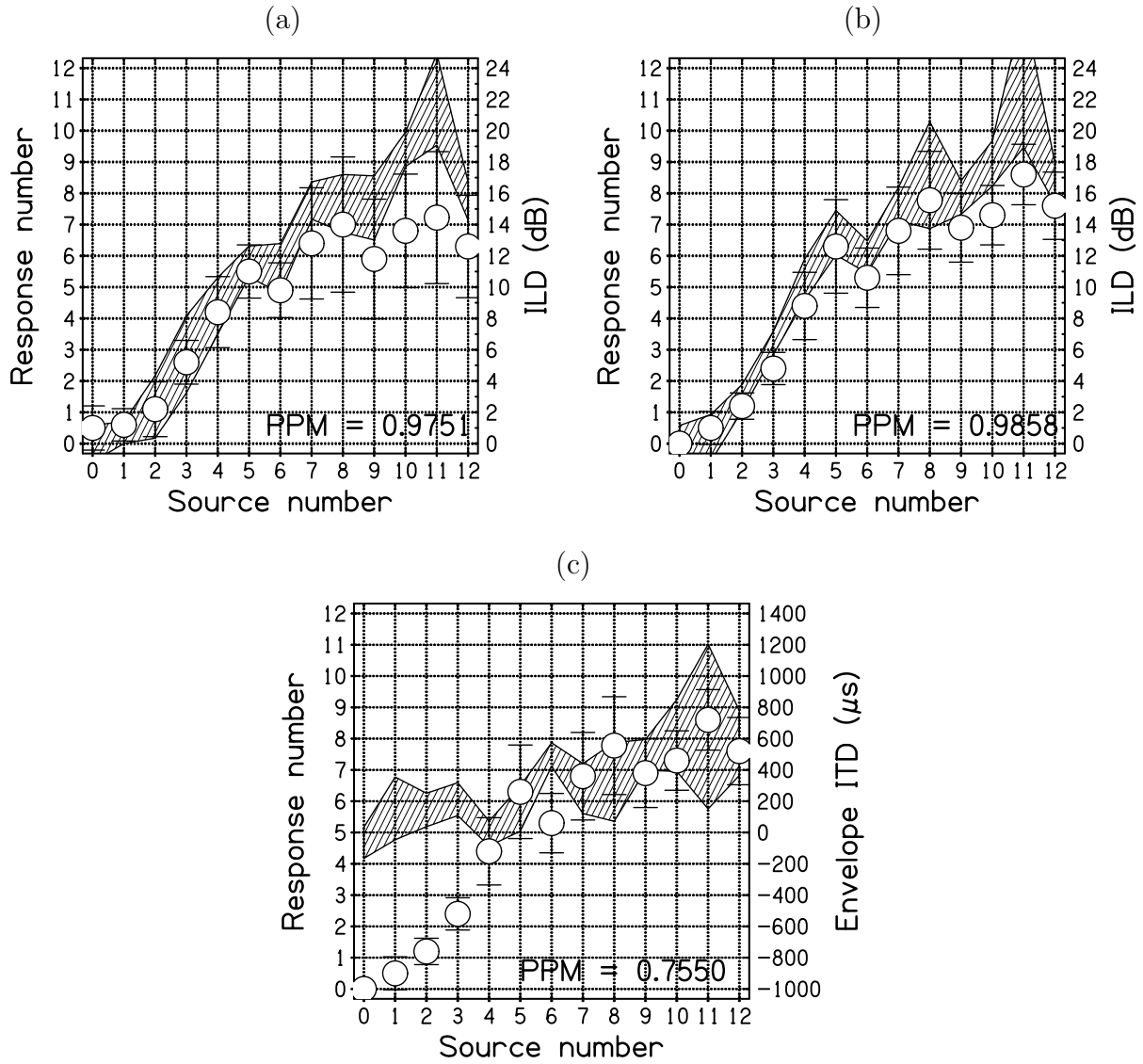


Figure 2.13 Same as Fig. 2.11 but for listener B at 4 kHz.

Listener C, 2 kHz

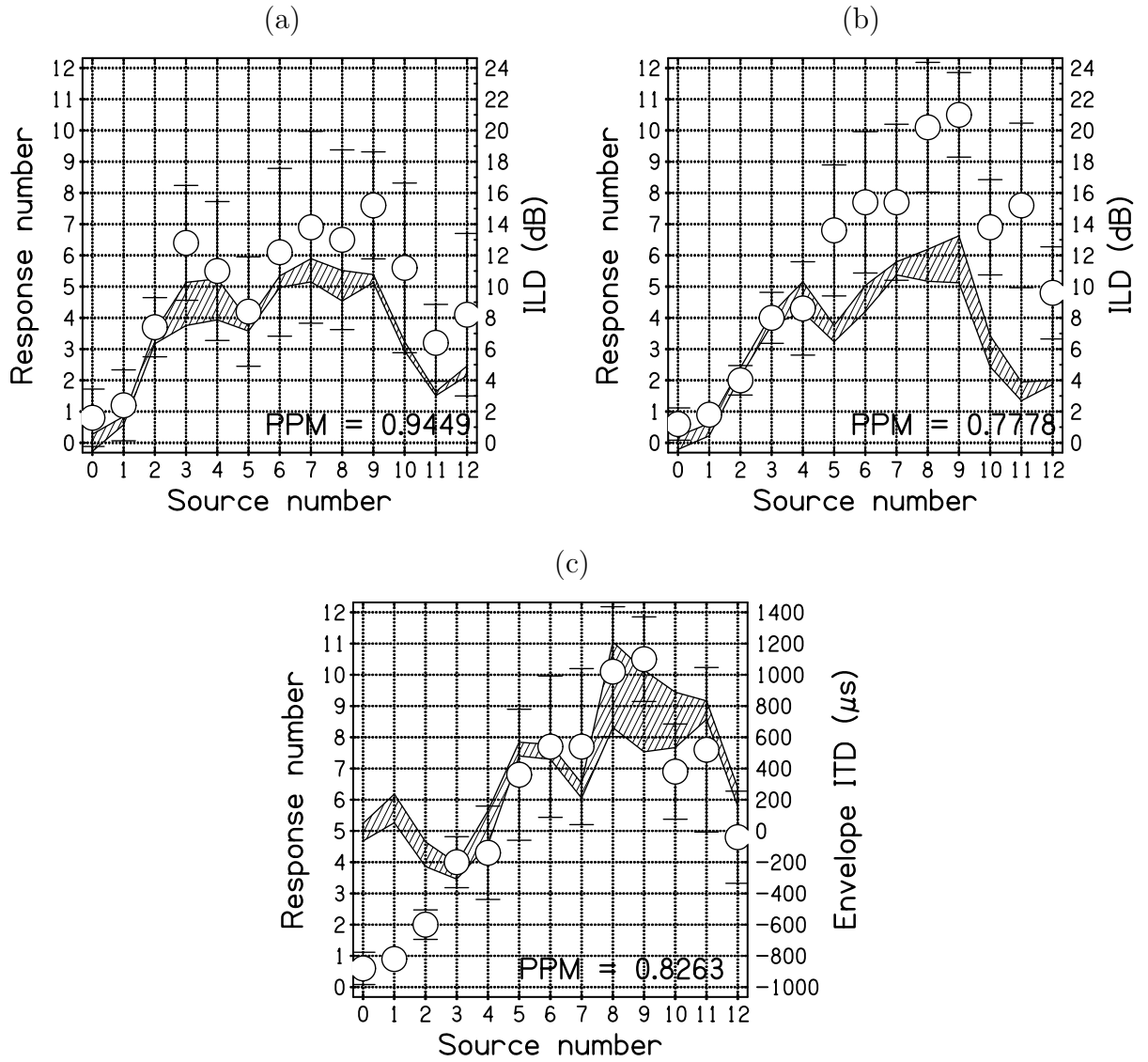


Figure 2.14 Same as Fig. 2.11 but for listener C at 2 kHz.

Listener C, 3 kHz

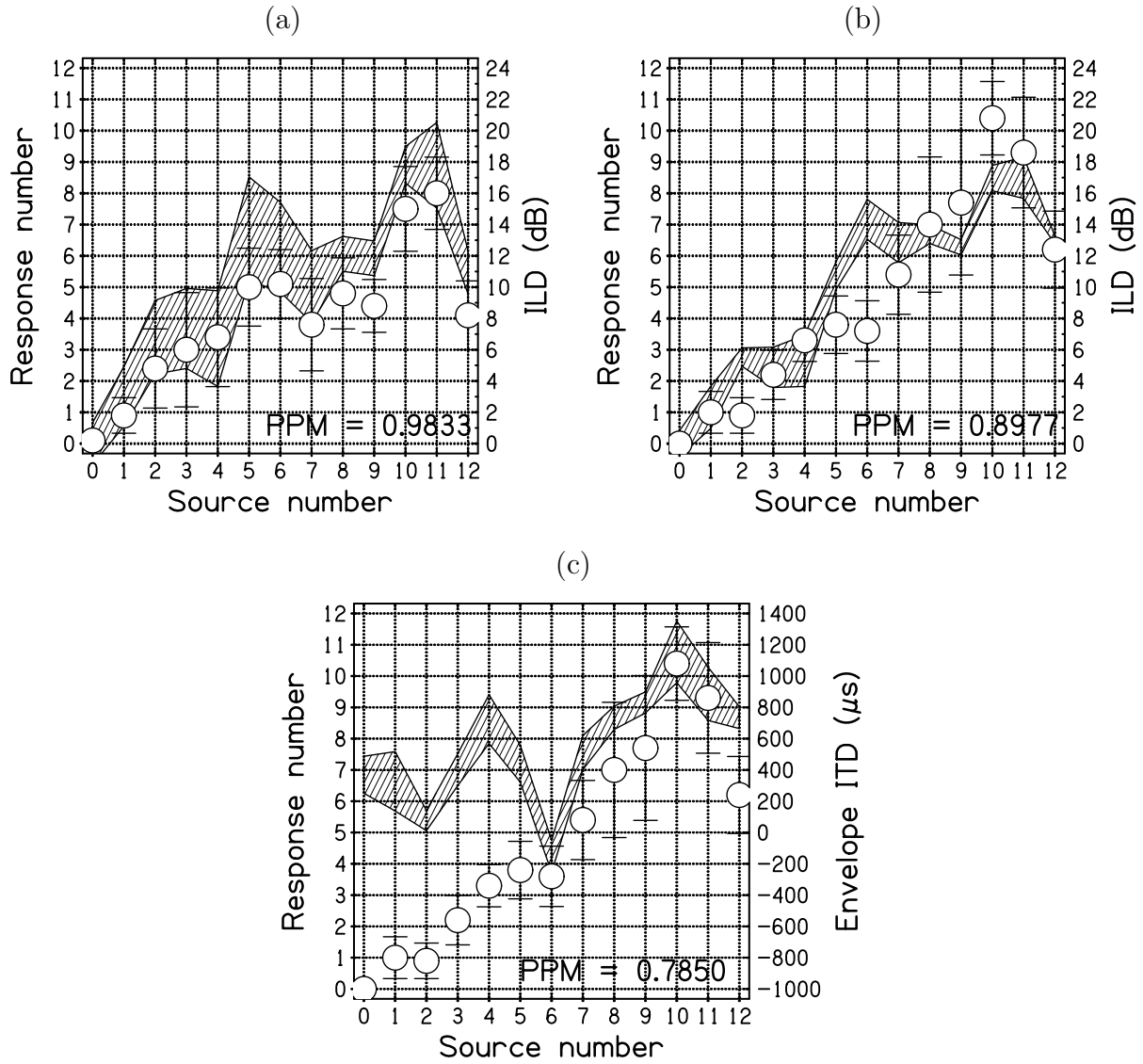


Figure 2.15 Same as Fig. 2.11 but for listener C at 3 kHz.

Listener C, 4 kHz

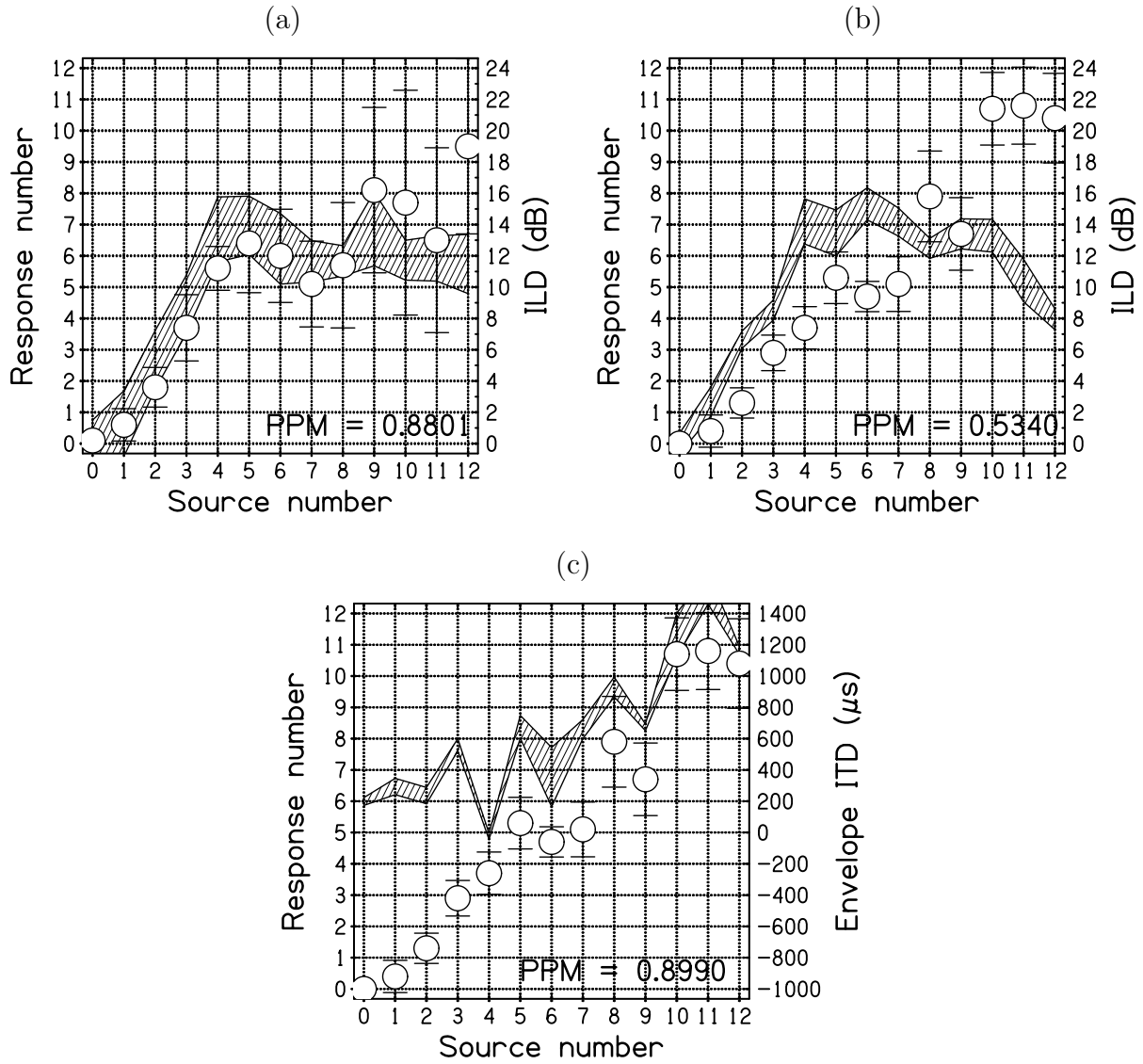


Figure 2.16 Same as Fig. 2.11 but for listener C at 4 kHz.

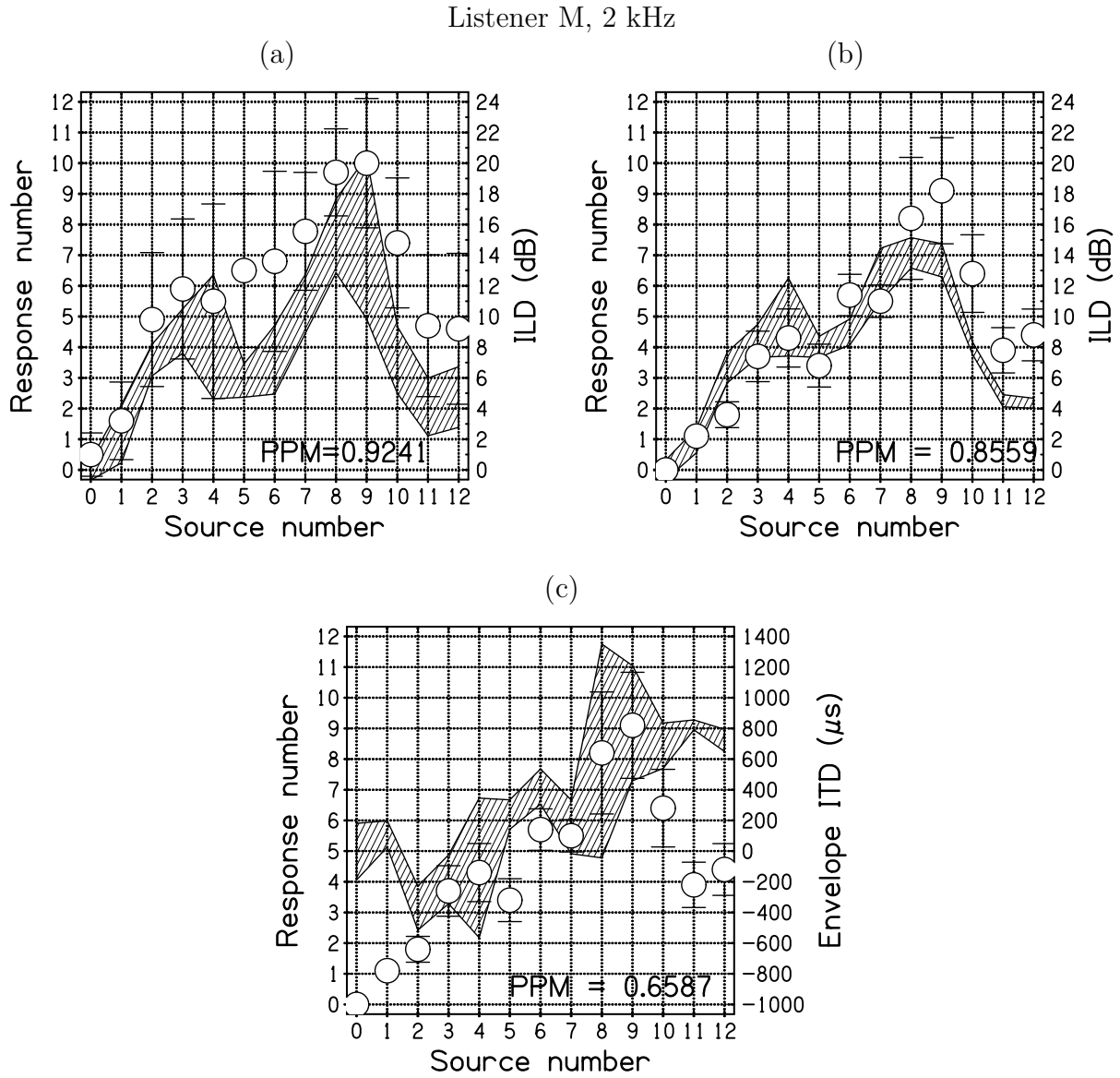


Figure 2.17 Same as Fig. 2.11 but for listener M at 2 kHz.

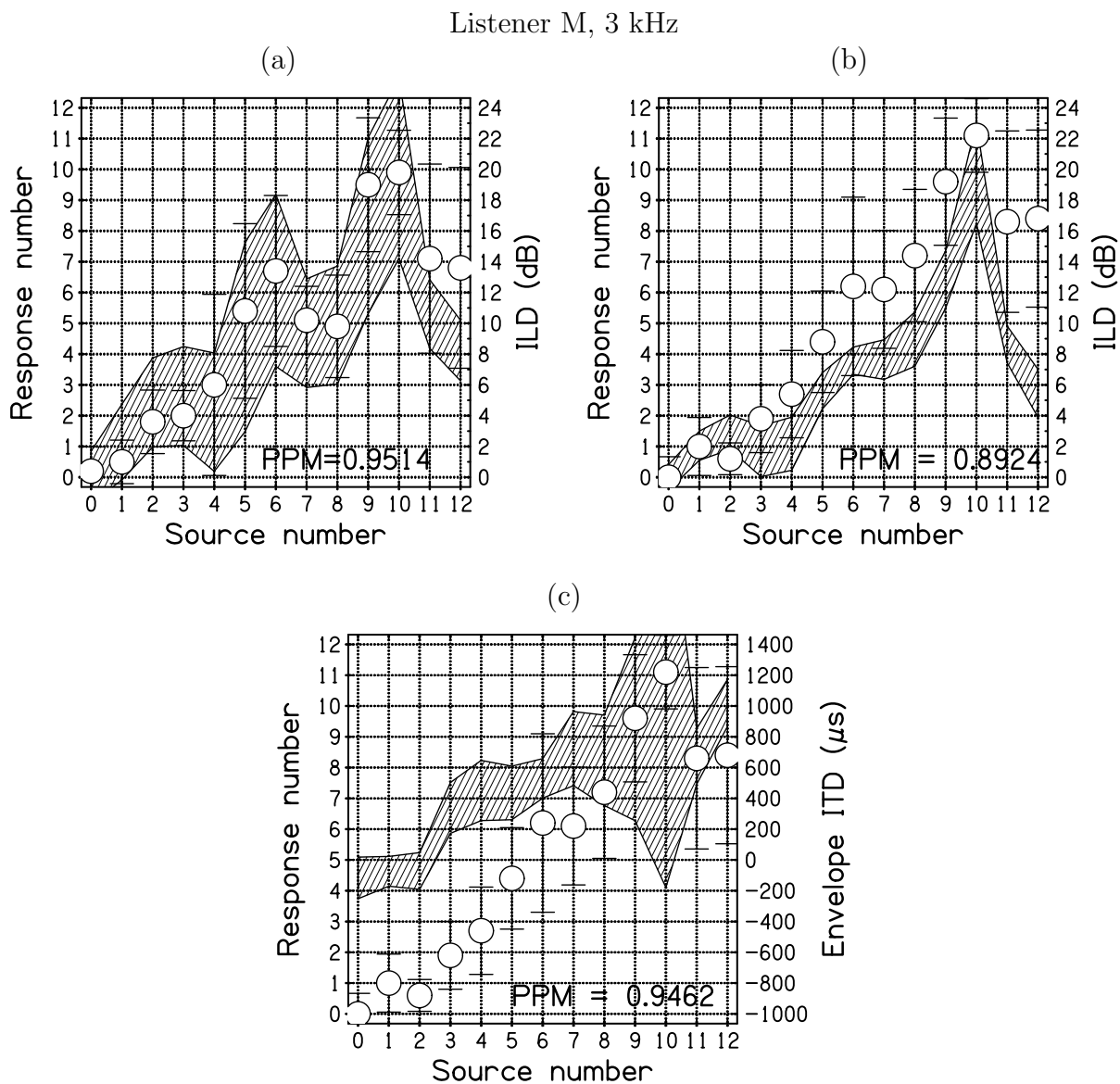


Figure 2.18 Same as Fig. 2.11 but for listener M at 3 kHz.

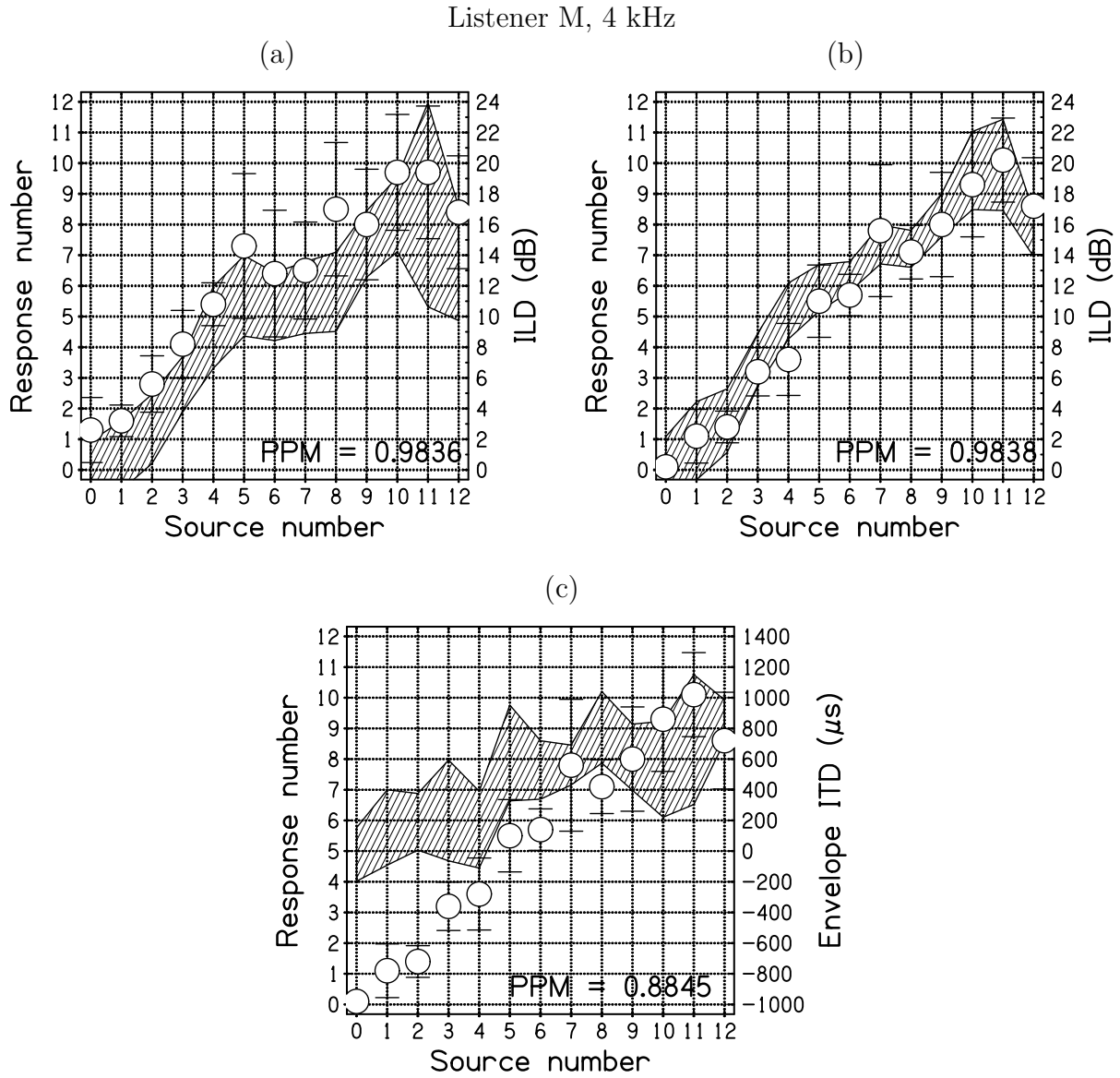


Figure 2.19 Same as Fig. 2.11 but for listener M at 4 kHz.

Listener L, 2 kHz

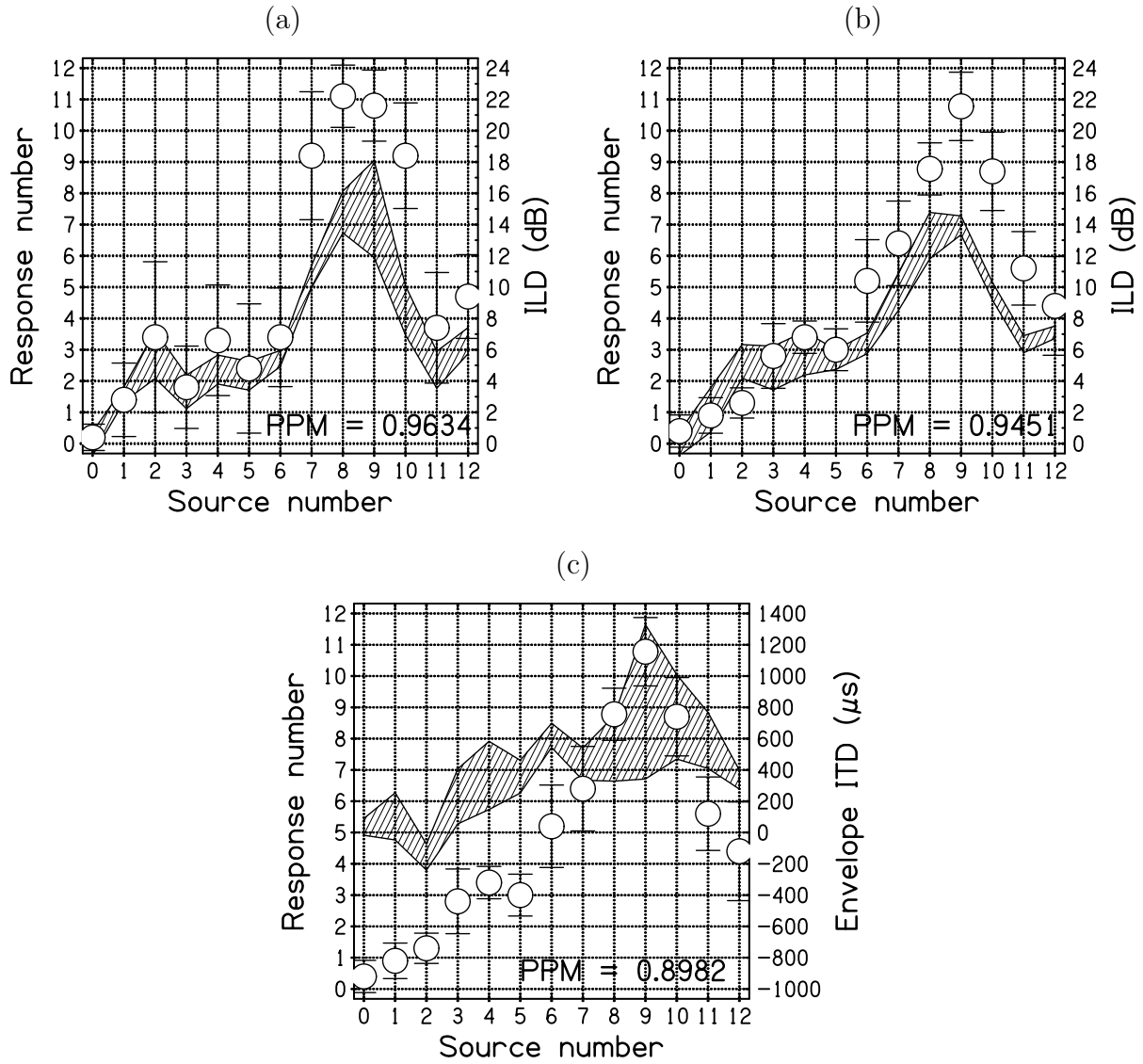


Figure 2.20 Same as Fig. 2.11 but for listener L at 2 kHz.

Listener L, 3 kHz

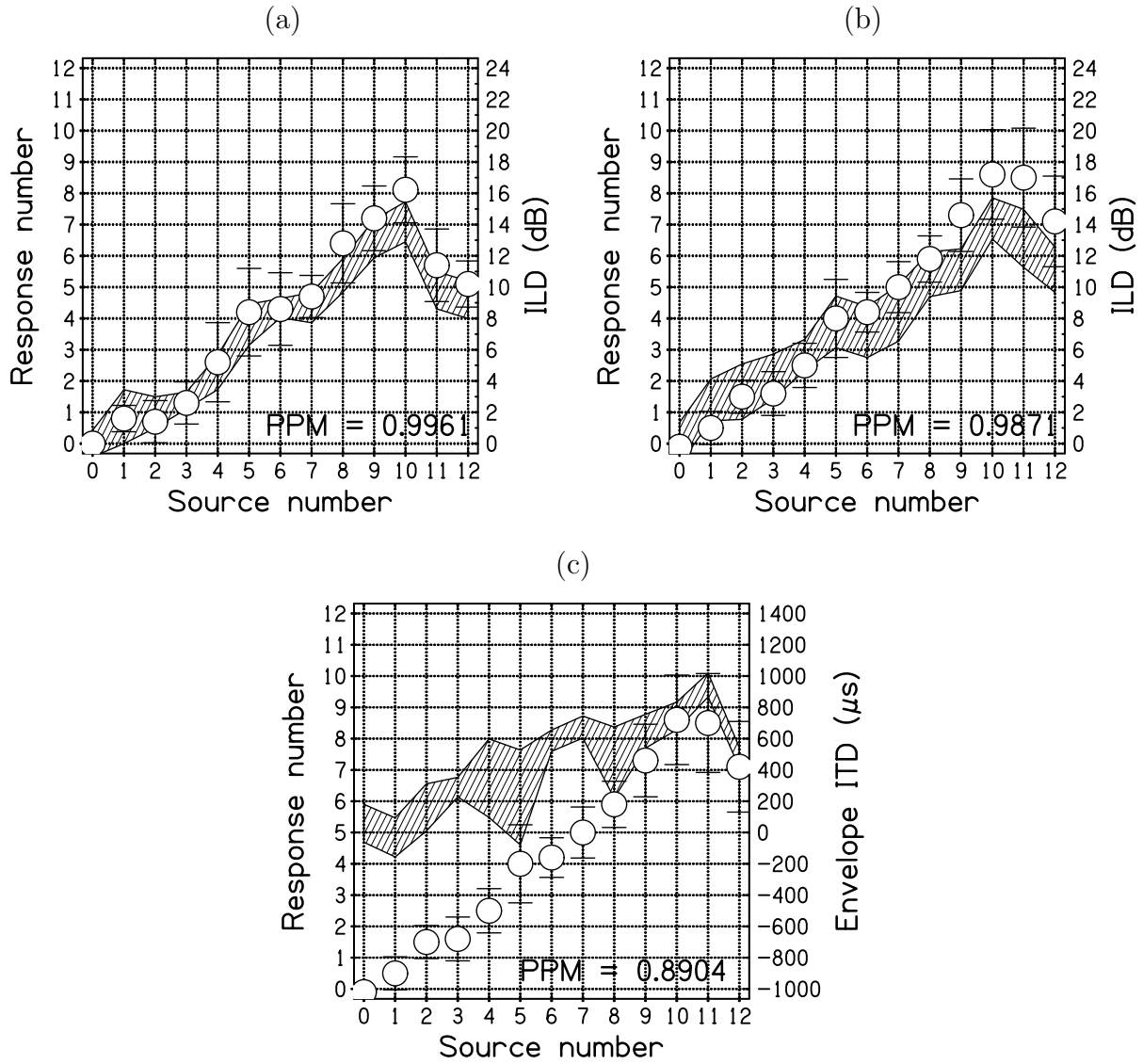


Figure 2.21 Same as Fig. 2.11 but for listener L at 3 kHz.

Listener L, 4 kHz

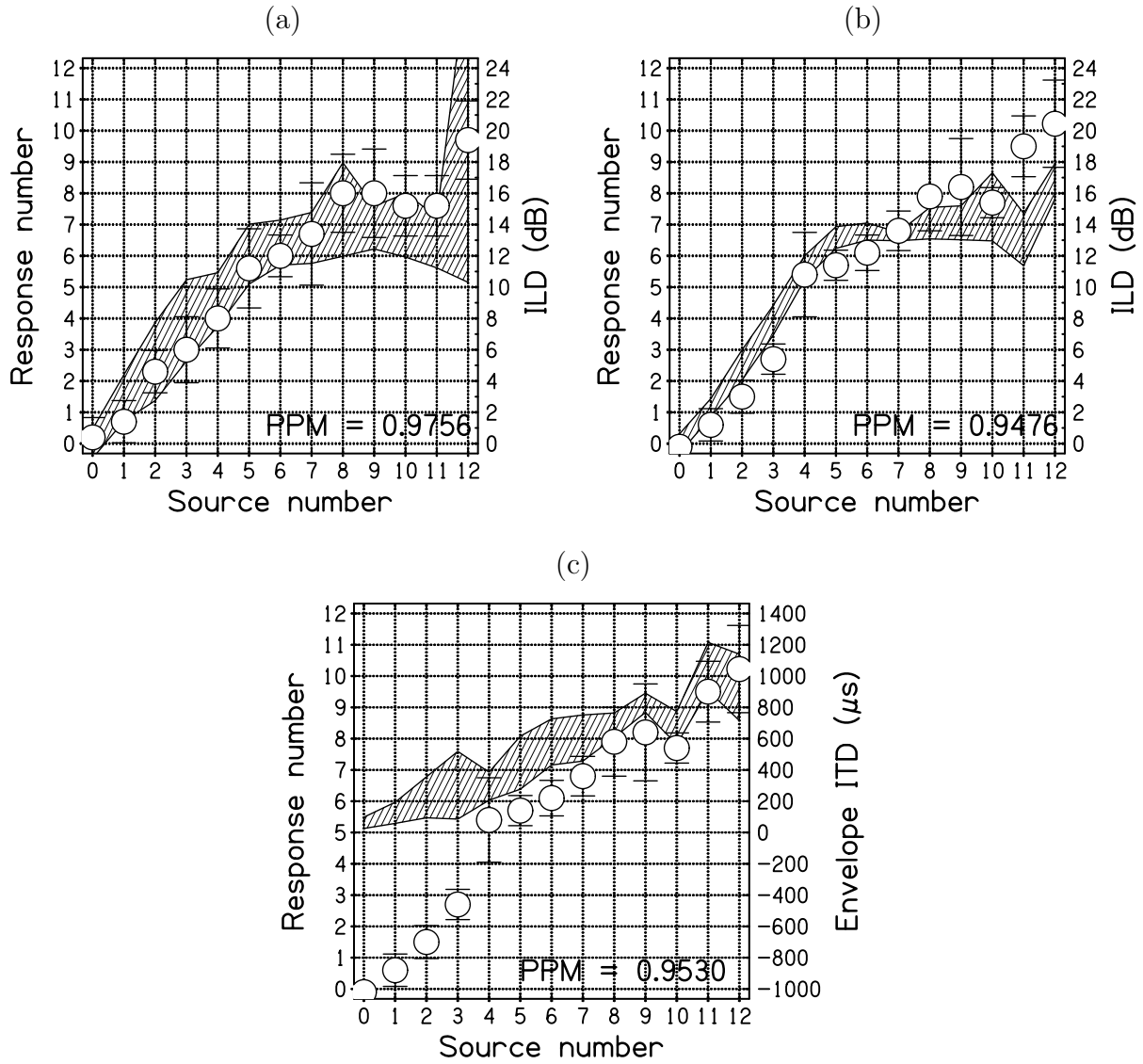


Figure 2.22 Same as Fig. 2.11 but for listener L at 4 kHz.

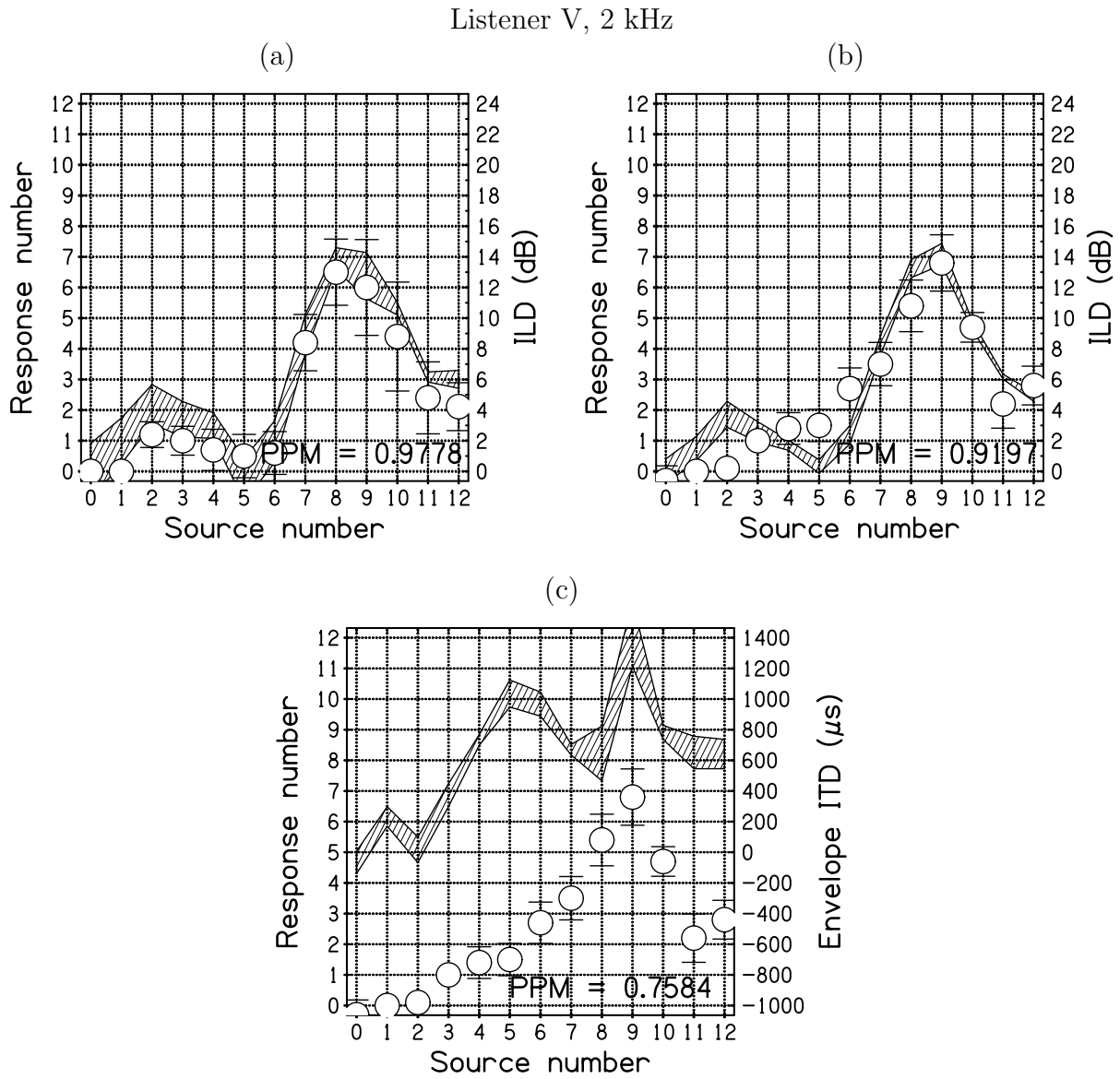


Figure 2.23 Same as Fig. 2.11 but for listener V at 2 kHz.

Listener V, 3 kHz

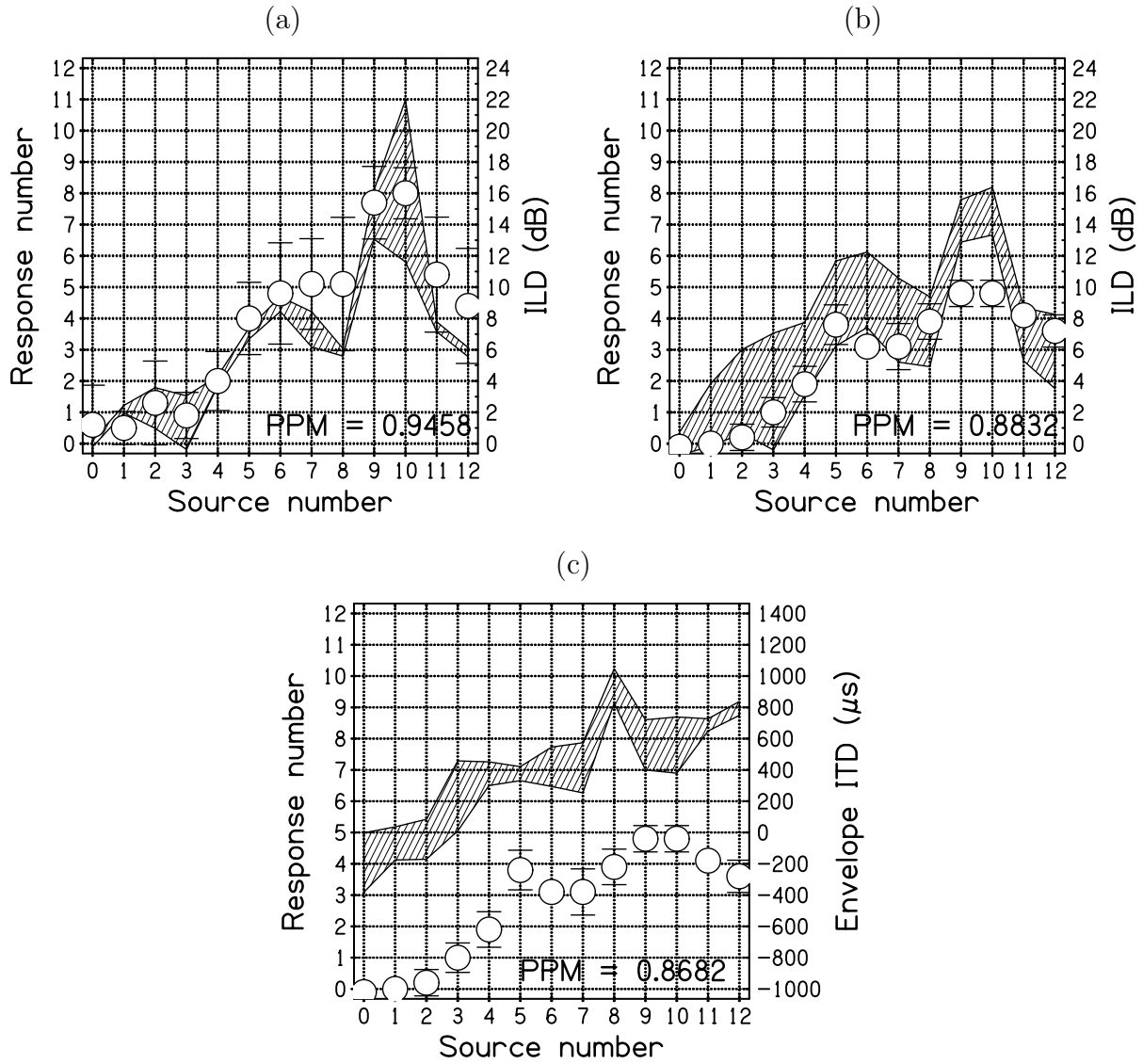


Figure 2.24 Same as Fig. 2.11 but for listener V at 3 kHz.

Listener V, 4 kHz

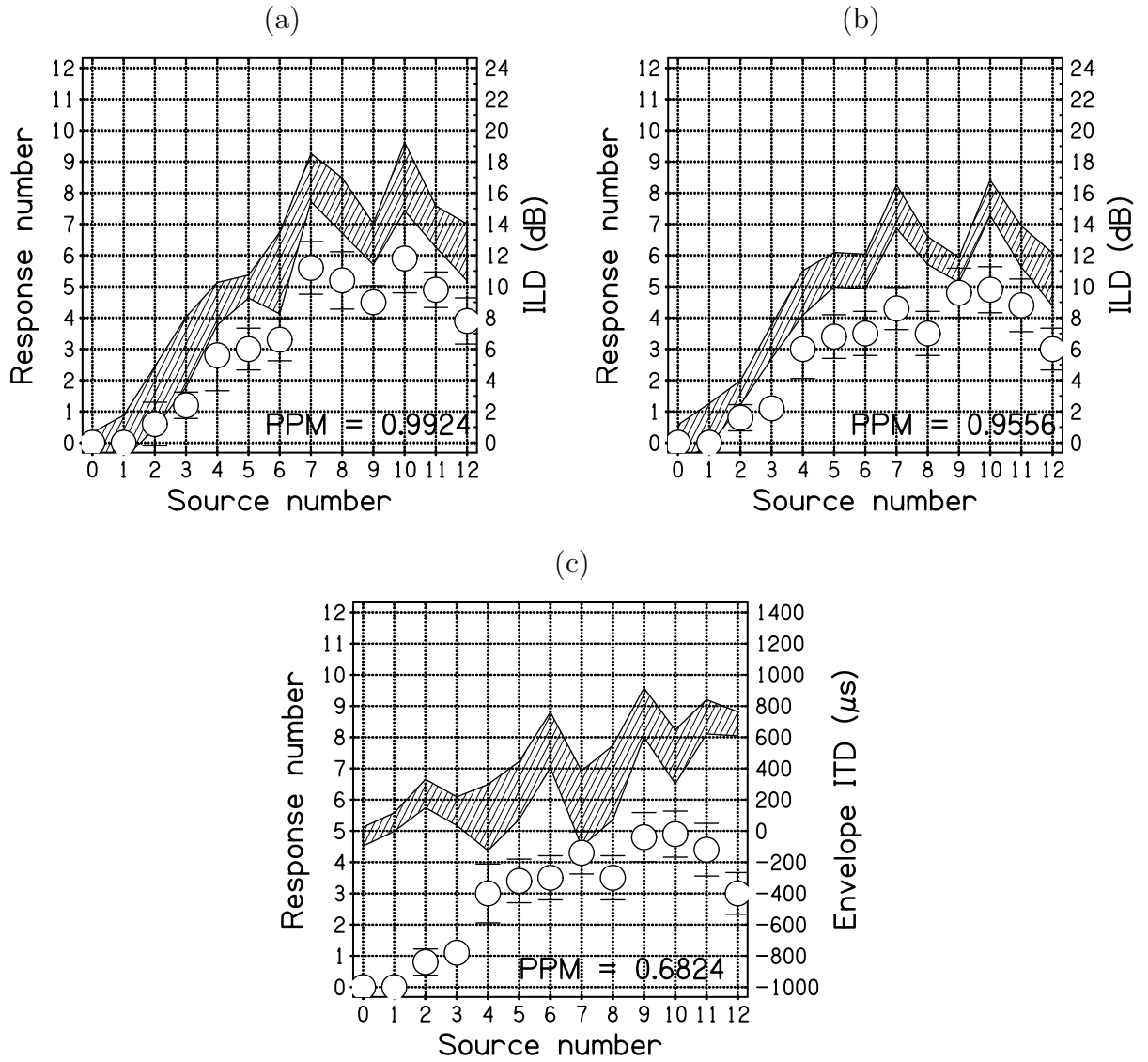


Figure 2.25 Same as Fig. 2.11 but for listener V at 4 kHz.

Results of a statistical analysis of the change in responses associated with the presence of AM is shown in Figs. 2.26–2.28. The p-values are calculated from a two-tailed t-test with the null hypothesis that responses to SAM are the same as responses to sine tones. The p-values confirm that there are many cases where the AM trials produced significantly different responses ($p < 0.05$). This is not necessarily due to the EITD cue, but may also be influenced by a change in ILD. The ILD is expected to change little due to the addition of AM itself. Most of the change is probably random and caused by the listener’s position. With the exception of when p-values very close to 1.0, cases with larger p-values do not necessarily indicate that the responses were statistically indistinguishable. This could simply be the result of an inconsistent listener response. The p-values should only be expected to be small in situations where listener responses are consistent and either the ILD has significantly changed and/or the EITD cue is significantly conflicting with the ILD. Because a large number of t-tests were performed, it is likely that there are instances in which the null hypothesis was incorrectly rejected. Therefore, these p-values should be considered as a whole, rather than placing emphasis on any one in particular.

The statistical significance of the changes in response vary greatly on a loudspeaker-by-loudspeaker basis. Fisher’s method [12] is a meta-analysis of the p-values based on a one-sided χ^2 test and shows the significance of the changes in response across all loudspeakers. The null-hypothesis is that the mean responses for all loudspeakers are the same in the sine and AM runs. The alternative hypothesis is that the mean response for at least one of the loudspeakers is different for the sine and AM runs. In other words, the alternative hypothesis is that listeners gave different responses with AM than with sine tones. At a significance level of $\alpha = 0.05$, all of the Fisher’s method p-values reject the null hypothesis except for listener B at 4 kHz. Again, although it is clear that listeners gave different response in the

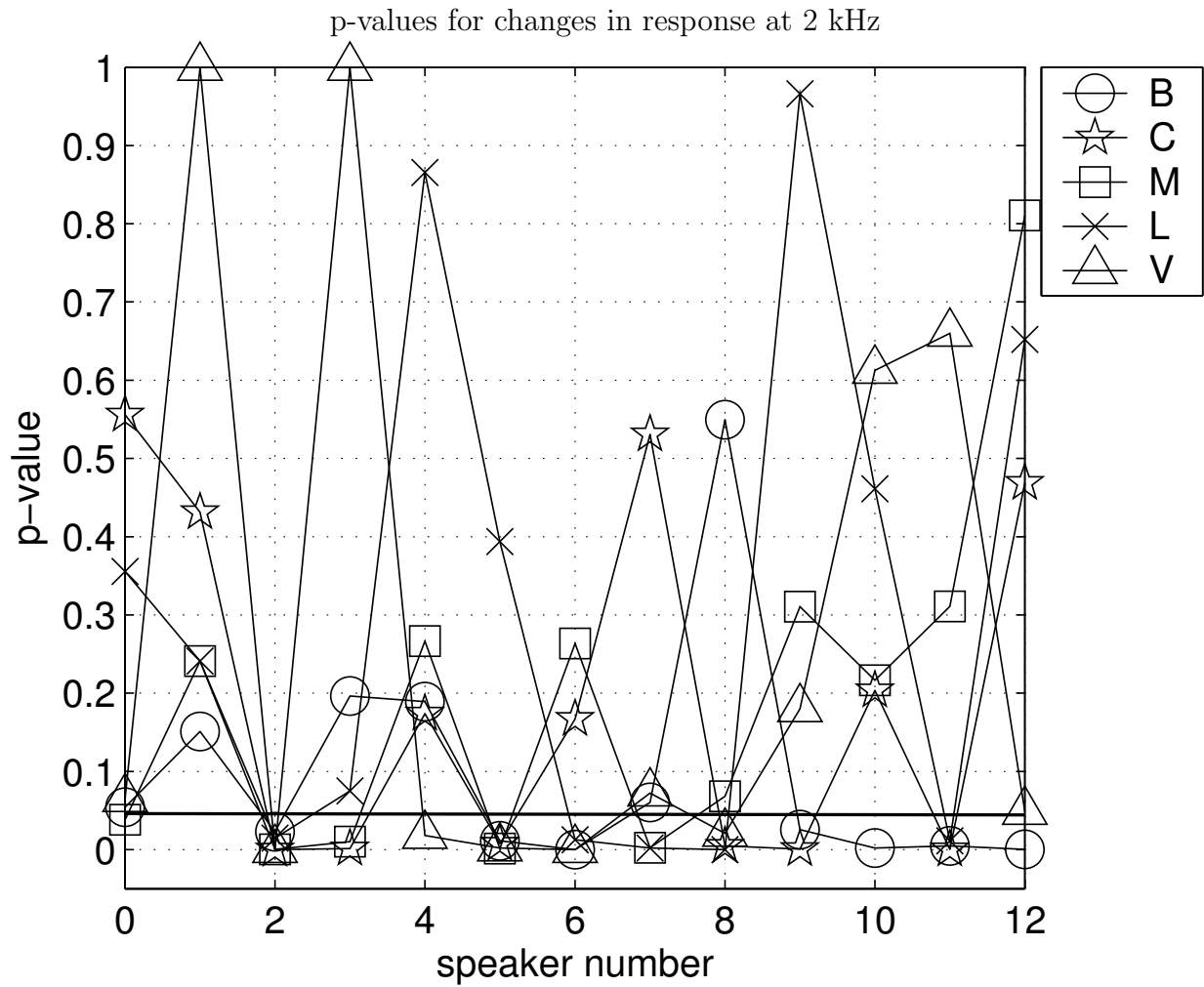


Figure 2.26 p-value vs. speaker number for changes in response at 2 kHz. Each listener is plotted according to the symbols in the legend. The p-values are based on a two-tailed t-test and represent the probability that the responses in the sine tone and AM trials are statistically similar. A reference line is shown for $\alpha = 0.05$.

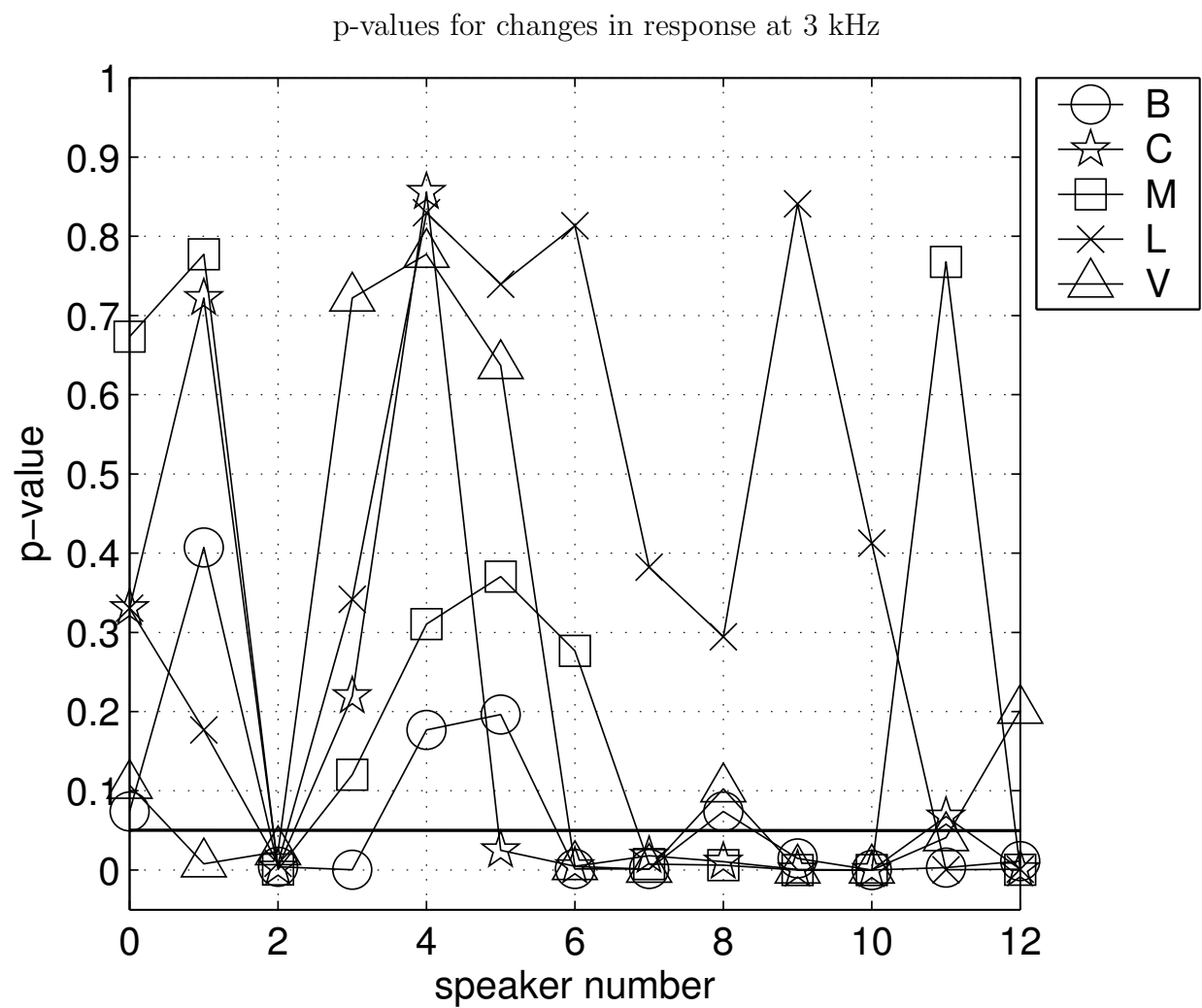


Figure 2.27 Same as Fig. 2.26 but for 3 kHz.

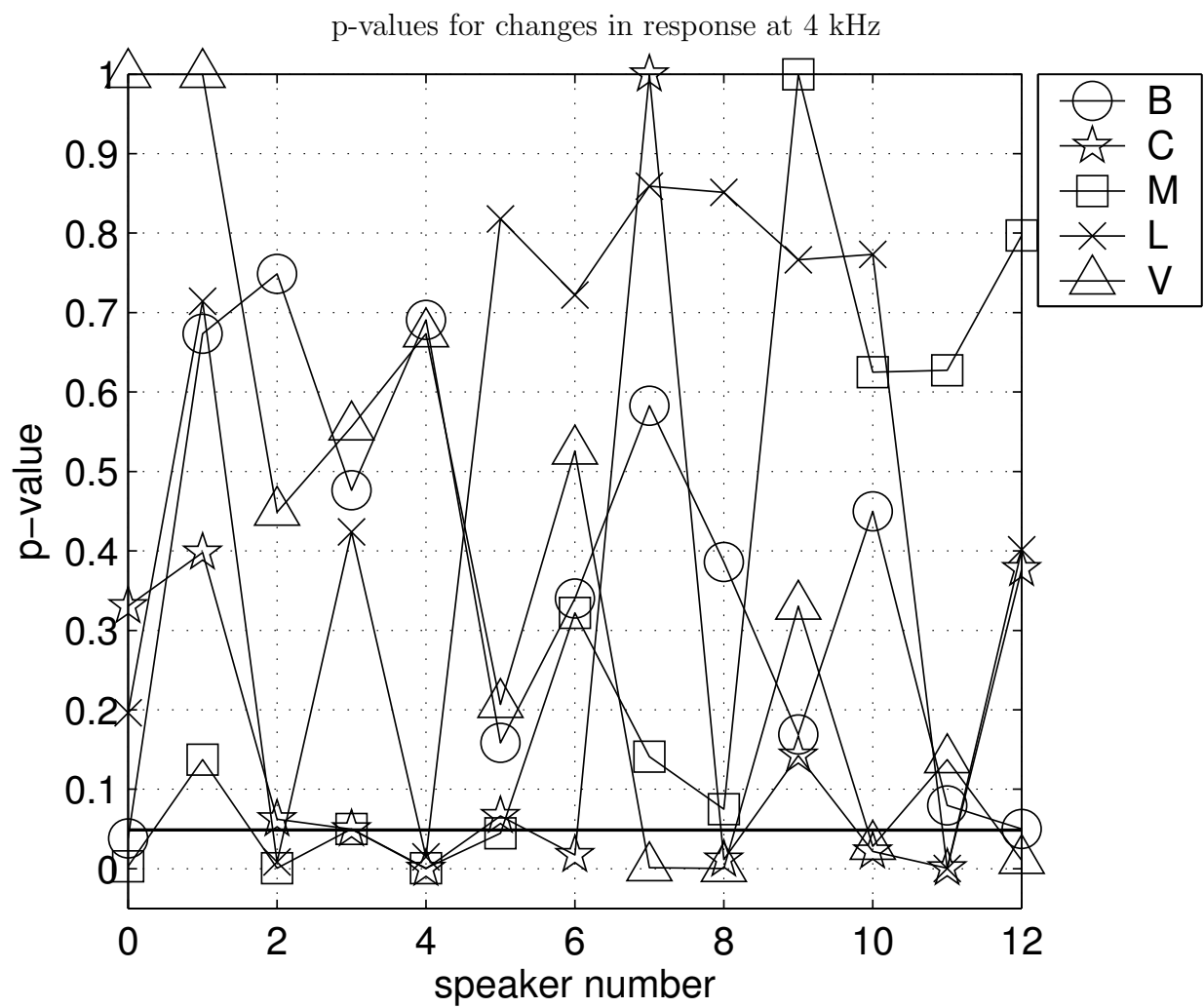


Figure 2.28 Same as Fig. 2.26 but for 4 kHz.

Fisher's method p-values			
Listener	2 kHz	3 kHz	4 kHz
B	$4 * 10^{-12}$	$4 * 10^{-14}$	0.1084
C	$8 * 10^{-11}$	$8 * 10^{-12}$	$5 * 10^{-9}$
M	$3 * 10^{-7}$	$2 * 10^{-13}$	$4 * 10^{-6}$
L	$2 * 10^{-6}$	$7 * 10^{-4}$	0.0151
V	$7 * 10^{-11}$	$6 * 10^{-16}$	$1 * 10^{-4}$
All listeners	$5 * 10^{-39}$	$1 * 10^{-49}$	$1 * 10^{-15}$
All listeners/frequencies	$1 * 10^{-97}$		

Table 2.3 Fisher's method p-values. This meta-analysis uses a one-sided χ^2 test with $2n$ degrees of freedom on the summation over the logarithms of the t-test p-values for the loudspeakers, where n is the number of t-test p-values in the summation. Results are shown for each frequency and listener over the 13 loudspeakers, for each frequency across all listeners and loudspeakers, and finally across all frequencies, listeners, and loudspeakers.

AM runs, this cannot necessarily be fully attributed to the presence of the EITD, as the ILD is known to have changed as well, even if only slightly.

A measure of the accuracy of the listeners' responses is shown in Fig. 2.29. This plot shows the correlations between listener response means and the actual source azimuth. It does not involve the interaural cues. The correlations represent how well the listeners localized the source.

In every case, except for listener V at 4 kHz, the AM correlations are higher than the sine correlations. Although some of the improvements are modest, the introduction of amplitude modulation virtually never resulted in poorer listener performance at any of these frequencies. For both the sine tones and SAM tones, performance tends to increase as frequency increases. This can be attributed to the fact the the peak of the ILD curve advances to larger azimuths as the frequency increases, with the exception of listeners C and V at 4 kHz. Therefore, fewer loudspeakers in the array produce misleading ILDs. This also results in the correlation between ILD and source azimuth increasing with frequency, except for listeners C and L at

4 kHz (see Fig. 2.43). In other words, the ILD cue tends to be more reliable as the frequency increases. Looking back at Fig. 2.29, the AM correlations tend to not increase as much as the sine correlations. This may be explained by the EITD solving the problem of the misleading non-monotonic ILD cue. It is effective at all of these frequencies. This can also explain why the improvement in performance increases as the frequency decreases (Fig. 2.29).

Figure 2.30 shows correlations between listener responses and the ILD for sine and AM runs. Most correlations are quite high. This is expected to be the case with the sine tones. With the exception of listener C, the AM correlations are also large. The only cases where the AM correlation is larger than the sine are listener B and M at 4 kHz. The AM correlations are expected to be lower since responses are also influenced by the EITD. However, it is striking how small the correlation differences are. There are two possible reasons for this. The first is that the EITD may not have had a strong effect on listener responses. The second is that the EITD and the AM ILD themselves were highly correlated. In this case, the cues would be redundant and correlations with response would be high for each of them.

The red bars in Fig. 2.31 show this correlation. One can see that cases where the AM ILD and EITD correlation is smaller corresponds to situations in Fig. 2.30 with larger disparities between the two correlations. For example, in Fig. 2.30 listener C has fairly low correlations between response and AM ILD (red bars), while listener L has very high correlations. Figure 2.31 shows that listener C has low correlations between the two cues, whereas for listener L they are high. This shows that even though the correlations between response and AM ILD is high, that the ILD is not necessarily solely causing the responses. The EITD must be relevant too.

The blue bars in Fig. 2.31 show the similarity between the ILD in the sine and AM runs. As stated previously these cues are indeed very similar.

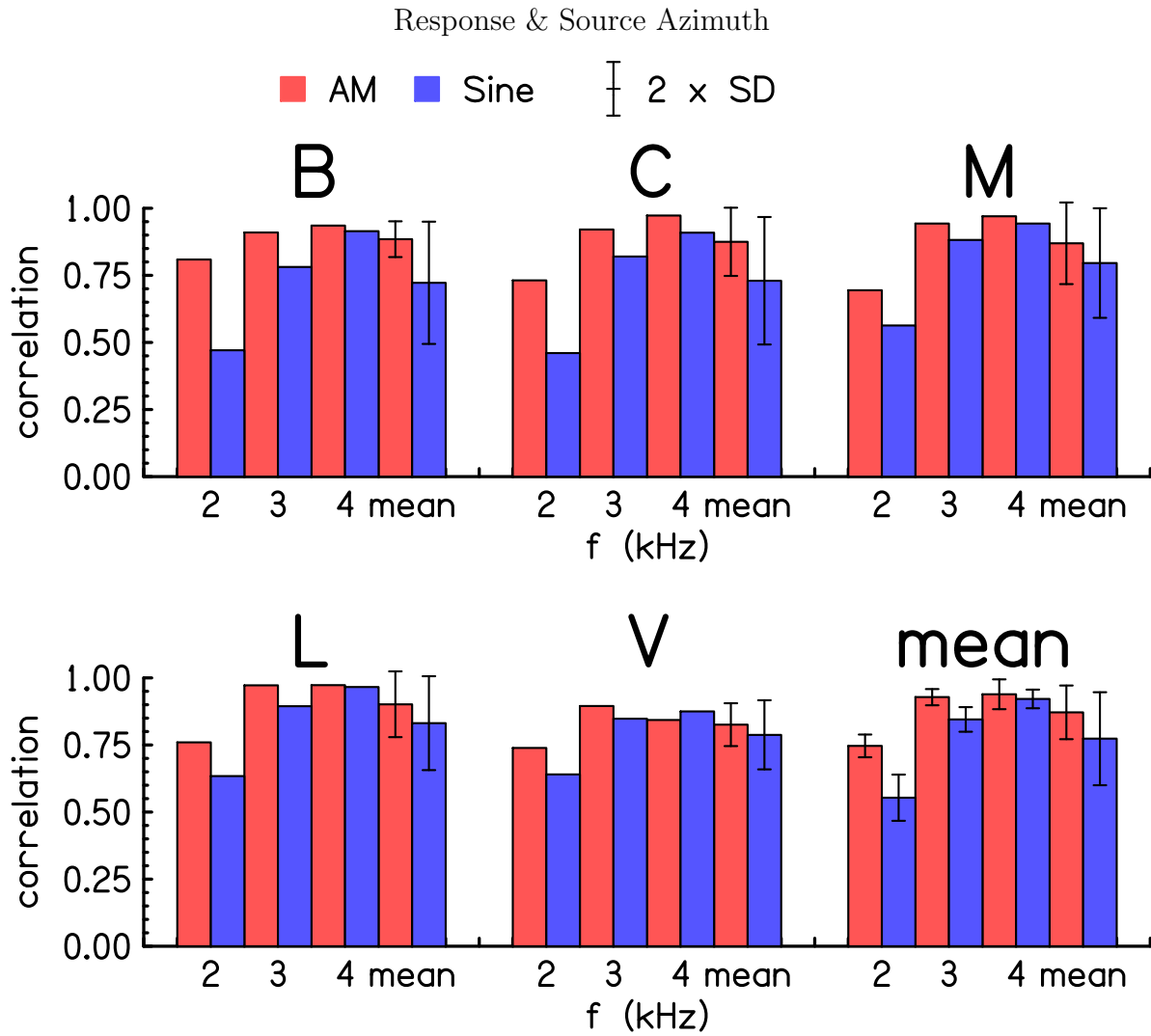


Figure 2.29 Pearson product-moment (PPM) correlation coefficients for perfect responses (source number) and actual responses are shown for AM in red and sine in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.

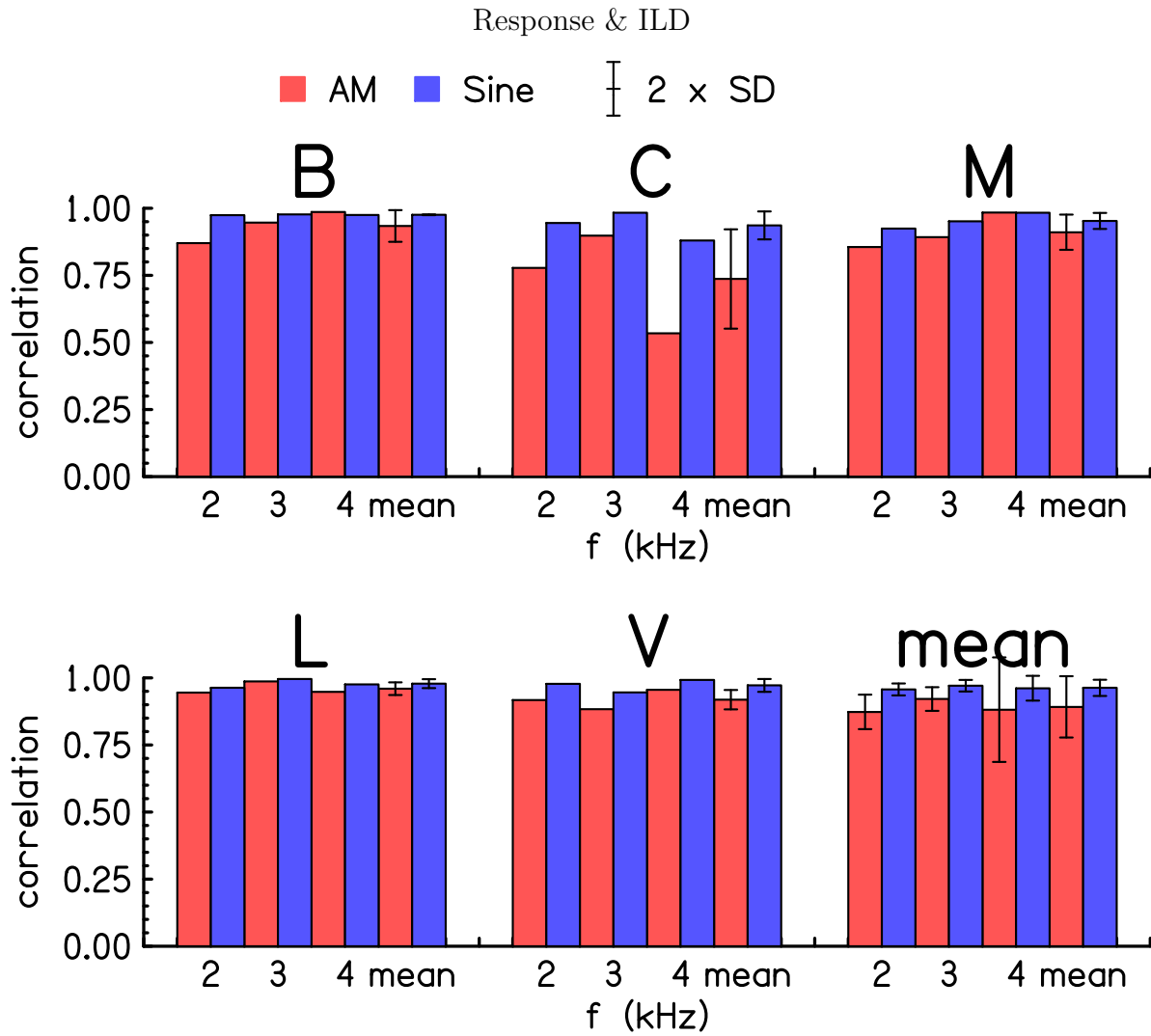


Figure 2.30 Pearson product-moment (PPM) correlation coefficients for AM responses and AM ILD are in red. Correlations for sine responses and sine ILD are in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.

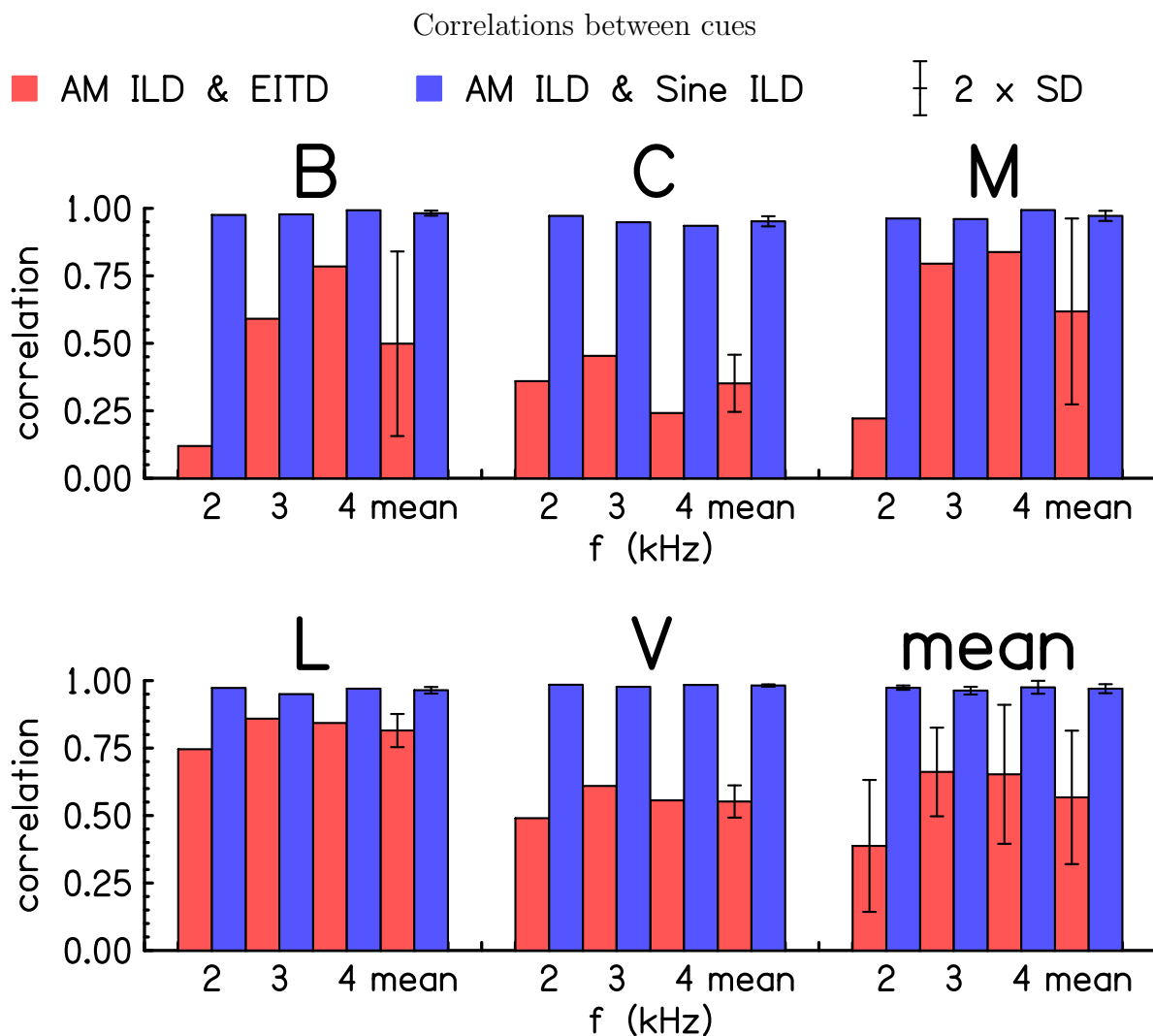


Figure 2.31 Pearson product-moment (PPM) correlation coefficients for AM ILD and EITD are in red. Correlations for AM ILD and sine ILD are in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.

To determine which cue may be playing a larger role in SAM localization one can compare the correlations between response and the two cues, as in Fig. 2.32. These plots are a summary of the correlations shown in Figs. 2.11–2.25. Although not true in every case, the trend is for the ILD cue to have a higher correlation with responses than the EITD.

2.3.2.2 SAM and sine differences

Even though the ILD seems to have more weight than the EITD in determining listener responses, one would expect that the EITD is the cue responsible for changing listener responses with the introduction of amplitude modulation. This is because the ILD has changed very little, whereas the EITD is a new, systematic cue. To test this, the correlations between the change in listener response (AM–sine) and the change in cue were investigated. These results are shown in Fig. 2.33.

While listener B and C appear to be strongly influenced by the introduction of the EITD cue, due to the large correlations between change in response and EITD, the other three listeners do not. Because of this, and since the change in ILD is essentially random, one might expect that listeners B and C have improved the most with the introduction of AM. Figure 2.29 supports this hypothesis.

Figure 2.34 shows the effect of negative EITDs on listener responses. The AM response for individual trials minus the mean sine-tone response for the same loudspeaker is plotted vs. EITD for all listeners and frequencies in instances where the EITD is negative. A line of best fit is shown for these data ($r = 0.23066$, $r^2 = 0.053204$). Most of the changes in response are negative, indicating that the conflicting EITD cue is most often causing listeners to localize sound sources closer to the midline. The change in ILD (AM – sine) is indicated by the color of the data points. The negative changes in ILD appear to correspond

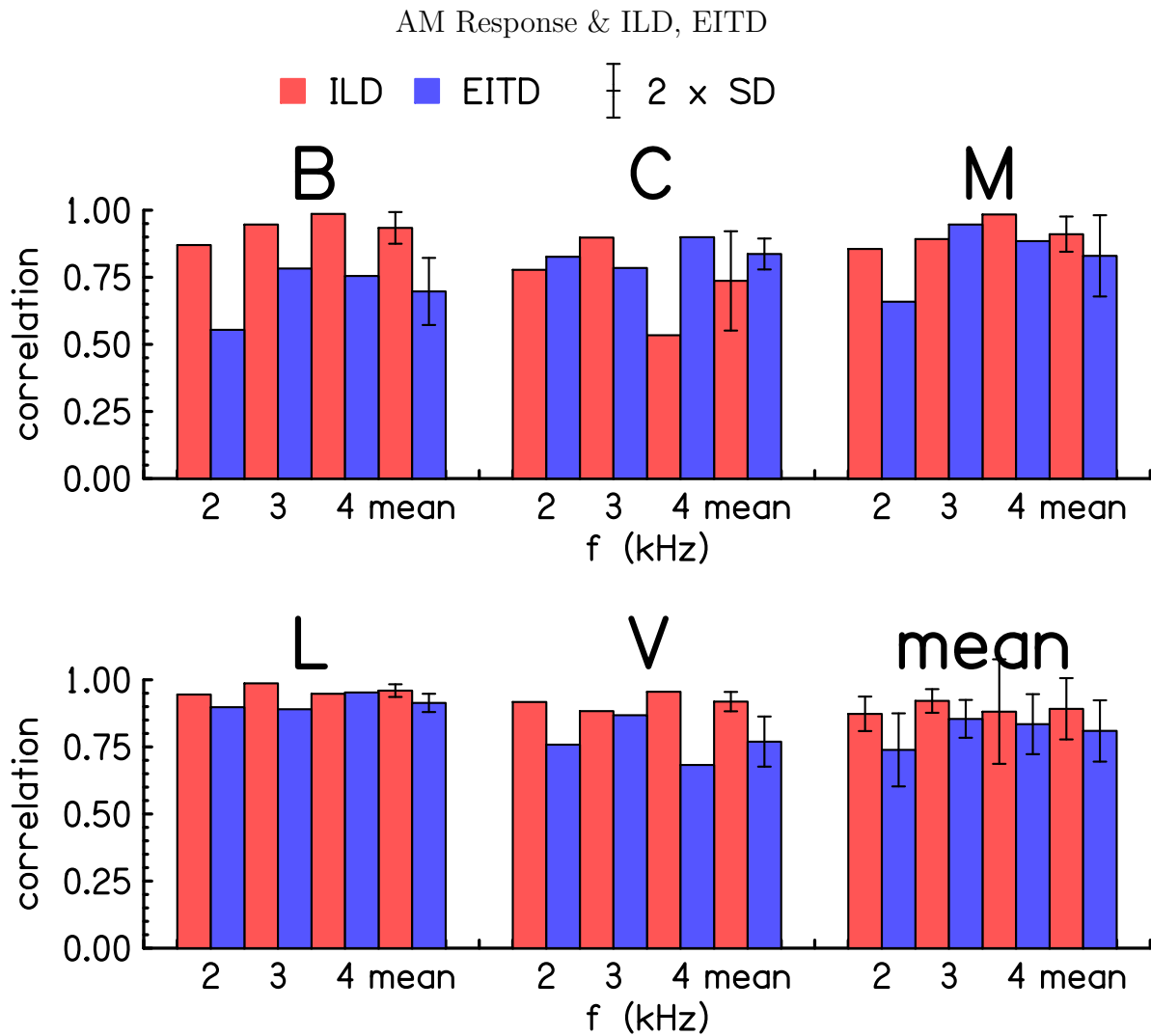


Figure 2.32 Pearson product-moment (PPM) correlation coefficients for AM responses and AM ILD are in red. Correlations for AM responses and EITD are in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.

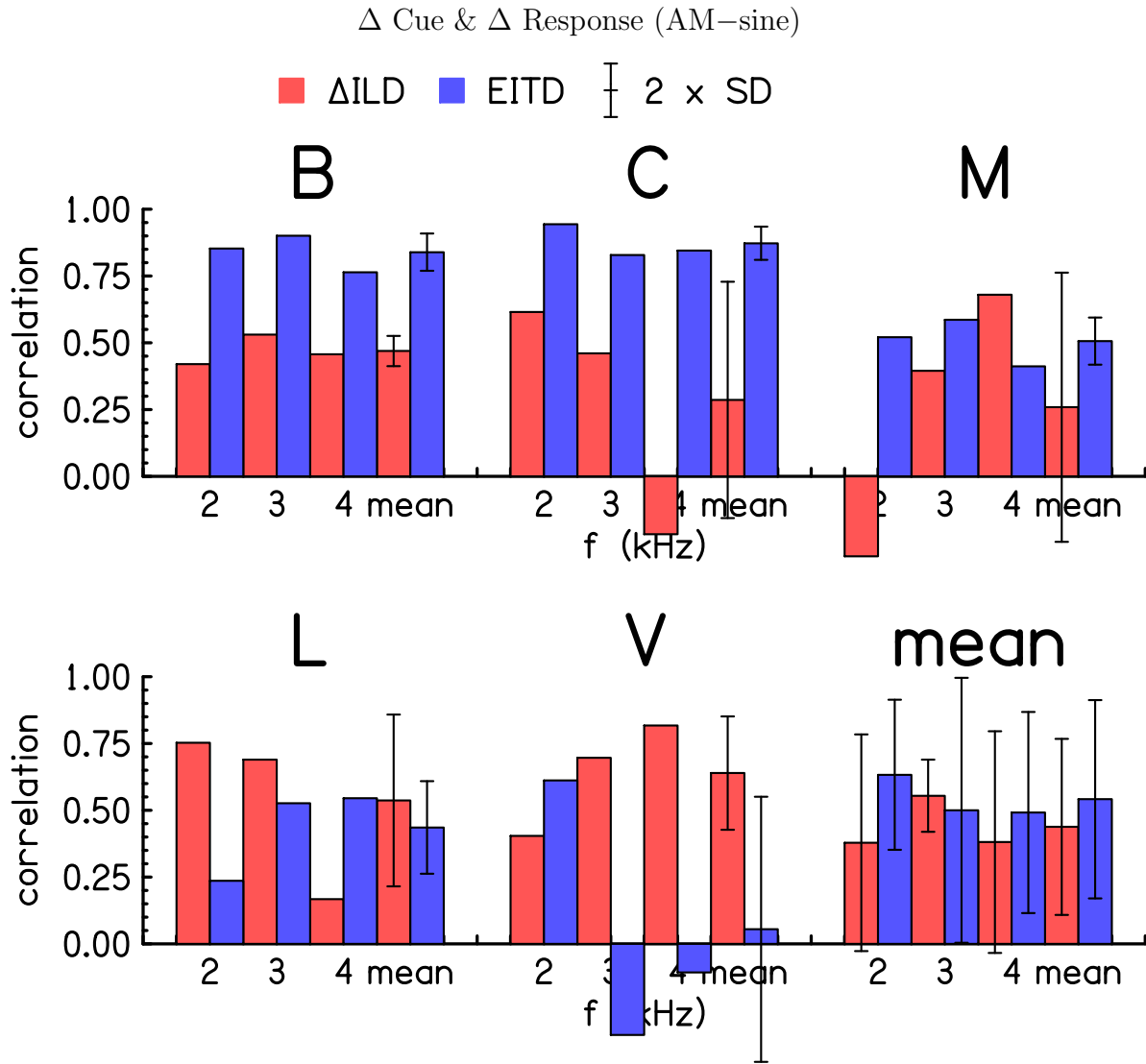


Figure 2.33 Pearson product-moment (PPM) correlation coefficients for the change in response and the change in ILD are shown in red. Correlation coefficients for the change in responses and the EITD are shown in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.

to smaller changes in response, and positive changes in ILD appear to correspond to larger changes in response. Therefore, the change in ILD may account for the cases where the change in response was positive, even though the EITD was negative. Figure 2.35 examines the residuals, or the vertical deviation from the line of best fit in Fig. 2.34, vs. the Δ ILD. Although there is some scatter, presumably due to the inconsistency of listener responses ($r = 0.31731$, $r^2 = 0.10069$) the line of best fit shows that there is a trend that as the Δ ILD increases, so do the residuals from Fig. 2.34.

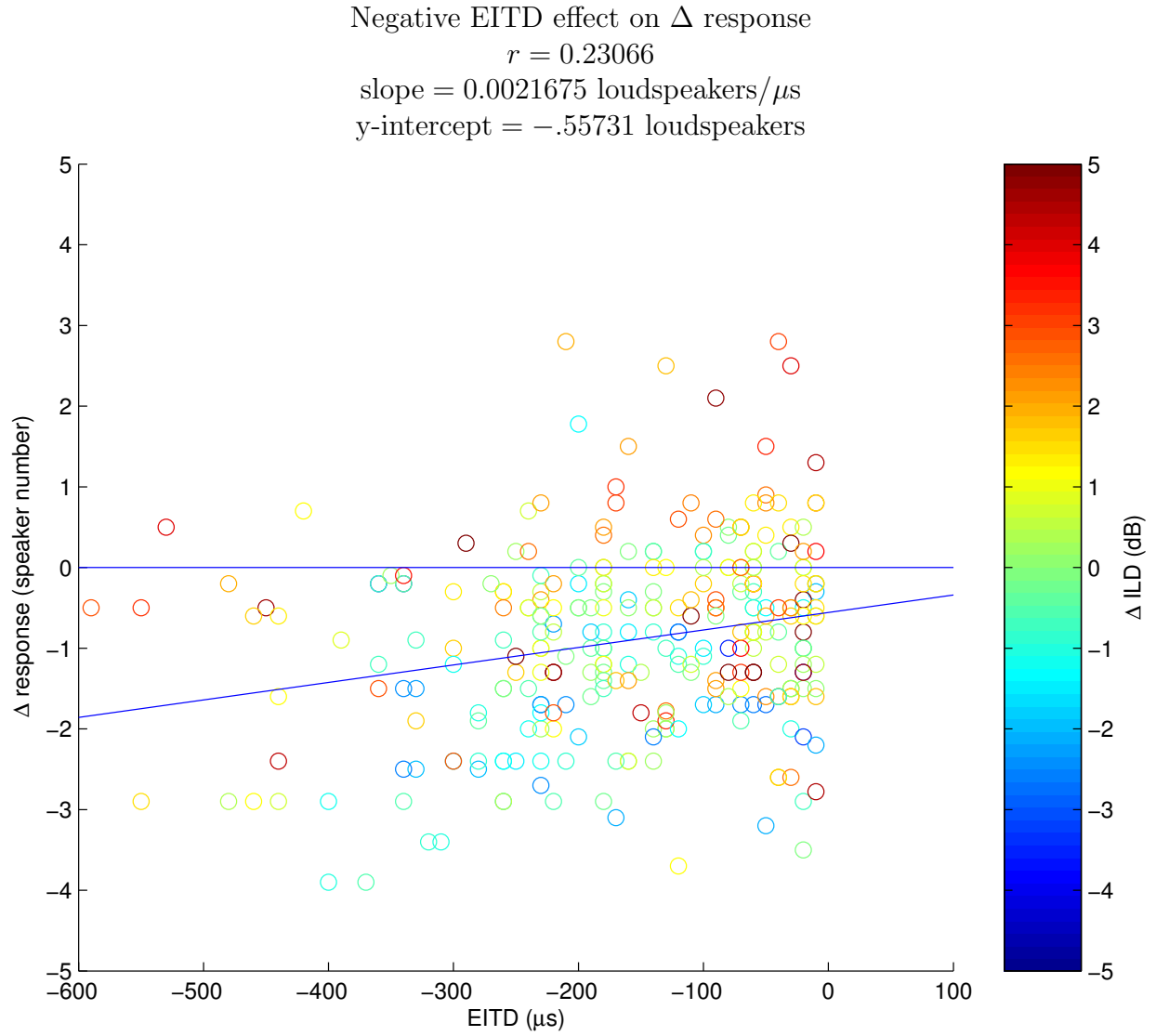


Figure 2.34 Change in response (AM – sine (mean)) vs. EITD for all negative EITD. All listeners and frequencies at combined. The vertical axis indicates AM responses for individual trials minus the mean sine response for the same loudspeaker. The horizontal axis indicates the EITD measured in the individual AM trials. The correlation coefficient— r —, slope, and y-intercept for the best fit line are above the plots. The color scale indicates the value of the change in ILD between the individual AM and average sine runs. A few of the Δ ILD values clip the top of the color scale at +5dB. A reference line for zero Δ response is shown.

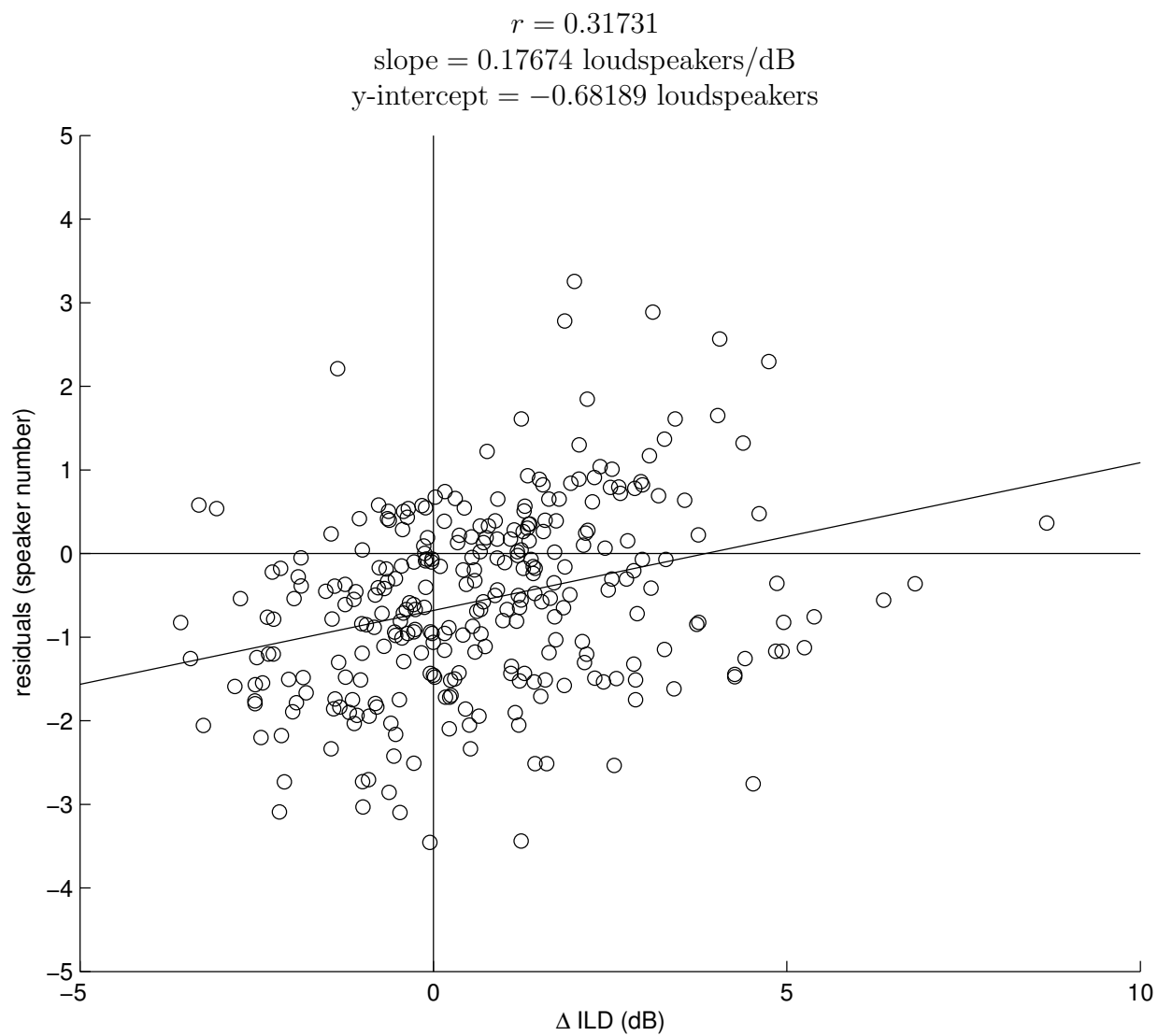


Figure 2.35 Residuals from the best fit line in Fig. 2.34 vs. Δ ILD. A line of best fit is shown as well as reference lines through the origin.

2.3.2.3 Compressive Cues

Figure 2.36 shows a plot from an IPD lateralization experiment by Yost (1981) [50]. For interaural cues that are small in magnitude, responses can be well-approximated to be linear. However, in this plot, one can see that after about $\pm 90^\circ$ IPD, the lateralization response judgements begin to saturate. This is also true of ILDs [50] and EITDs [3]. Saturation should also occur in localization tasks. Perceptual responses to interaural cues are compressive functions, of some kind or another.

An analysis that takes this compressive nature into account may better illustrate the effect of the EITD vs. the Δ ILD on differences in listeners' changes in responses. The previous analysis of the correlation between the change in response and the change in interaural cue was repeated using compressed cues. The amount of compression was variable between no compression and extreme compression. For each listener, frequency, and interaural cue, a particular amount of compression was chosen such that the correlation between the change in response and cue was maximized. Although neither of the compression extremes represent what may be considered to be correct, this analysis will shed light on how meaningful each of the cues were in changing listener responses. The cues are given the best possible chance to correlate with responses. Also, this served as a check to determine if this technique yields reasonable amounts of compression.

The EITD was transformed as

$$\text{EITD} \rightarrow \text{sgn}(\text{EITD})|\text{EITD}|^{x_1} \quad (2.36)$$

where the exponent, x_1 , was selected between 0.01–1.00 in 0.01 increments in order to max-

IPD lateralization data from Yost, W. A. (1981)

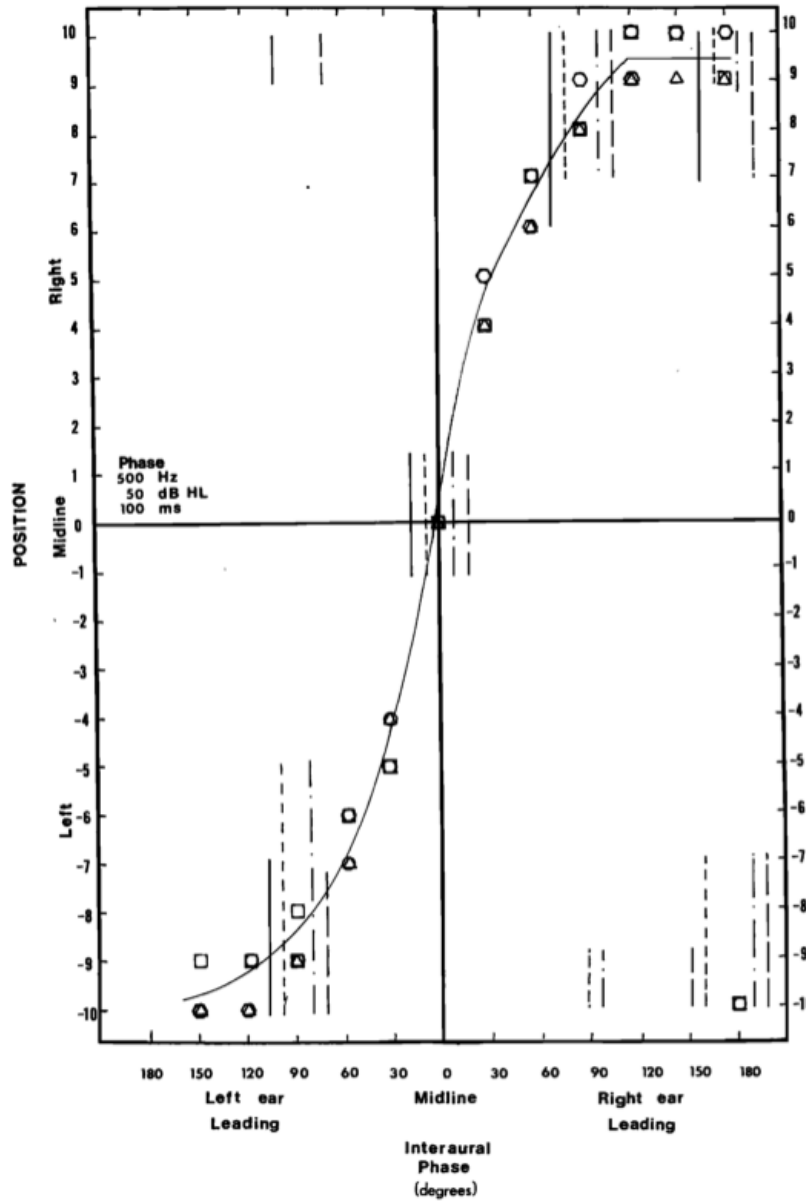


FIG. 3. Same as Fig. 2. Frequency: 500 Hz.

Figure 2.36 FIG. 3 from Yost (1981). In this headphone experiment, listeners lateralized on-going IPD cues at 500 Hz. The different symbols indicate different listeners, and the vertical lines indicate ranges of responses for IPDs of 0° , $\pm 90^\circ$, and 180° .

imize the correlation. Similarly, the change in ILD was transformed as

$$\Delta\text{ILD} \rightarrow \text{sgn}(\text{ILD}_{\text{AM}})|\text{ILD}_{\text{AM}}|^{x_2} - \text{sgn}(\text{ILD}_{\text{sine}})|\text{ILD}_{\text{sine}}|^{x_2} \quad (2.37)$$

where another exponent, x_2 , was selected between 0.01–1.00 in 0.01 increments in order to maximize the correlation. Because the correlations should not be negative, the maximization selected the most positive correlation, rather than the correlation with the largest magnitude. The correlations and the exponents, x_1 and x_2 , were then reported for each cue.

Figures 2.37–2.41 show the changes in response and compressed cues. For listener B, shown in Fig. 2.37, the EITD correlations (labeled PPM) exceed the ILD correlations at each frequency. One can see why, speaker by speaker. In panel (a), the the exponent is 0.01, yielding a correlation of only 0.5757. This amount of compression is unreasonable. However, if a totally flat change in ILD yielded the best correlation with the change in response, this change in ILD was apparently not influencing the change in listener B’s response at 2 kHz. Listener B’s changes in response are influenced to a large degree only by the EITD cue.

For listener C, shown in Fig. 2.38, the results are similar to those of listener B. The EITD correlations are fairly large and greater than the corresponding correlations for the change in compressed ILD. On a speaker-by-speaker basis, it is convincing that this listener’s changes in response were predominately affected by the introduction of the EITD cue. The exponents for the EITD cue are reasonable. For the ILD, the exponents are much smaller, especially at 3 and 4 kHz where it is 0.01.

For listener M, shown in Fig. 2.39, the effect of the EITD is less convincing than for listeners B and C. At 2 kHz, the EITD correlation beats the ILD correlation. But the EITD correlation is not convincing by itself. Only for speakers 0–4 due the changes in response

appear to correlate well with the EITD. However the ILD correlation is negative, which is totally unreasonable. Notice though that the responses did not change drastically. Because of this, a large correlation is unlikely to appear anywhere. Again, at 3 kHz, the EITD correlation beats the ILD correlation, but is not convincing by itself. But at both 2 and 3 kHz, the EITDs appear to be large enough that one would expect a larger change in response. This listener seems to be somewhat insensitive to the introduction of EITD and is largely basing his response in the existing ILD (which has not changed drastically). At 4 kHz, the ILD correlation beat the EITD correlation, but neither of them are convincingly large.

For listener L, shown in Fig. 2.40, the EITD does not have a convincing effect on the changes in response. At 2 kHz, the change in ILD correlates fairly well with the change in response. The EITD correlation is quite a lot smaller, and is maximized when the EITD is totally compressed. At 3 kHz, neither the change in ILD nor the EITD are strongly correlated with the change in response. In fact the responses did not change much at all at 3 kHz. This makes it unlikely that there would be a high correlation for either cue. The same is true at 4 kHz. At both 3 and 4 kHz the EITD cue appears as if it should be extremely helpful, as it tends to rise roughly linearly. But the response change is mostly flat. This may be because listener L is insensitive to the EITD, and prefers the ILD cue. Or possibly, this listener's sine-tone responses were already quite good and could not be improved upon. Looking at the mean responses for 3 and 4 kHz in Figs. 2.21 and 2.22, respectively, shows that there is room for this listener to improve. At 3 kHz, the mean response did not exceed speaker 9 and at 4 kHz the mean response slightly exceeded speaker 10. Or perhaps this listener's inexperience in psychoacoustical experiments contributed to an apparent lack of EITD sensitivity.

For listener V, shown in Fig. 2.41, the changes in response are not very drastic. At 2 kHz, neither correlations are very strong. At 3 and 4 kHz, the EITD correlations are negative, which essentially indicates that the listener's response was not affected by the EITD. Like listener L, listener V's responses had room for improvement, as seen in Figs. 2.23–2.25.

Figure 2.42 is a summary of the correlations from Figs. 2.11–2.25. It is similar to Fig. 2.33, except that the correlations have been maximized by the compressed-cues analysis. Unfortunately, the compressed-cues analysis did not result in drastically different correlations than the uncompressed-cues analysis. Naturally, the largest discrepancies appear when there is a lot of compression. The compression exponents tended to be very near one or zero. Although the exponents near zero yielded the largest correlations, they are totally unrealistic, and indicate that change in response and cue have weak correlation. Notice that the cases with very small compression exponents occur for the Δ ILD for listeners B and C and for the EITD for listeners L and V. This would indicate that in these instances, the listeners' changes in response are being dictated almost entirely by the opposite cues. This statement agrees with the general trend of the correlations for these listeners.

The differences between individuals here has a couple of possible explanations. The first possibility is that there are individual differences in the relative sensitivity to the ILD and EITD cues. Listeners B and C appear to be strongly influenced by the EITD. But listeners M, L, and V do not, to varying degrees. Different listeners may have different auditory weighting for different cues. However, there is another possibility. Consider the fact that these five listeners did not receive the same signals in their ear canals. This would be due mostly to differences in the external anatomy of the different listeners. It could be that listeners M, L, and V did not get any information from the EITD that could have been of any use. If the ILD was already quite large and produced a response with a large azimuth,

then there is not much room for improvement by an additional large EITD cue. The EITD cues that listeners M, L, and V received may not have conflicted very much with the ILD that was already present.

2.3.2.4 SAM interaural cue reliability

A way to understand the differences in performance change shown in Fig. 2.29 is to look at the reliability of the cues. Figure 2.29 shows the greatest improvements for listeners B and C. Did these listeners have substantially more reliable EITD cues than ILD cues? Figure 2.43 shows the correlation between the two AM cues and the azimuth. In principle, the larger these correlations are, the more reliable they are for localization.

If cue reliability is the cause of individual differences, then listeners B and C should show different cue reliability than the other listeners—especially listeners L and V. However, listeners B and C don't show more reliable EITD cues than the other listeners.

An argument could be made that listeners B and C have less reliable ILD cues than the other listeners. This could support the hypothesis that these listeners had more to gain by listening to the EITD cue than the other listeners. This may then explain why listeners B and C showed more correlation with the EITD and their change in responses in Figs. 2.33 and 2.42. Furthermore, Fig. 2.31 indicates that listeners B and C had more cases of dissimilar cues, than for listeners L.

2.3.3 Discussion

It remains somewhat unclear so far as to why there are individual differences in the correlations between listeners' changes in response and the cues shown in Figs. 2.33 and 2.42. But there is not overwhelming evidence to suggest that the cues themselves are responsible

Compressed Cues, Listener B

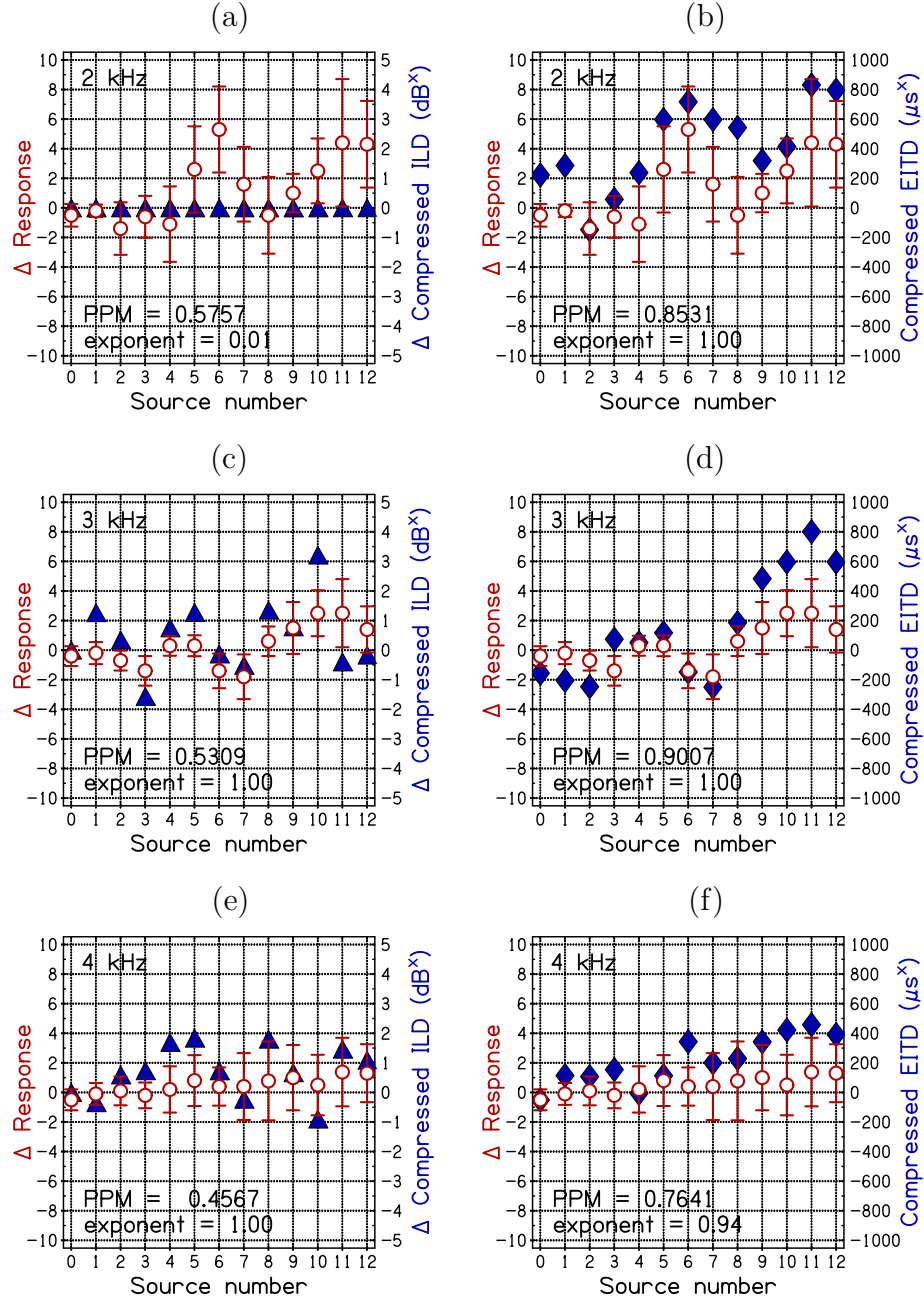


Figure 2.37 Changes in response and compressed interaural cues for listener B. The first row is 2 kHz, the second row is 3 kHz, and the third row is 4 kHz. For each plot, red circles indicate changes in mean response (AM–sine) and error bars are two standard deviations in overall length. For plots (a), (c), and (e), blue triangles indicate the change in compressed ILD (AM–sine). For plots (b), (d), and (f), blue diamonds indicate the compressed EITD. For each plot, the maximized Pearson product-moment (PPM) correlation coefficient between the changes in response and compressed interaural cue is shown, as well as the compression exponent.

Compressed Cues, Listener C

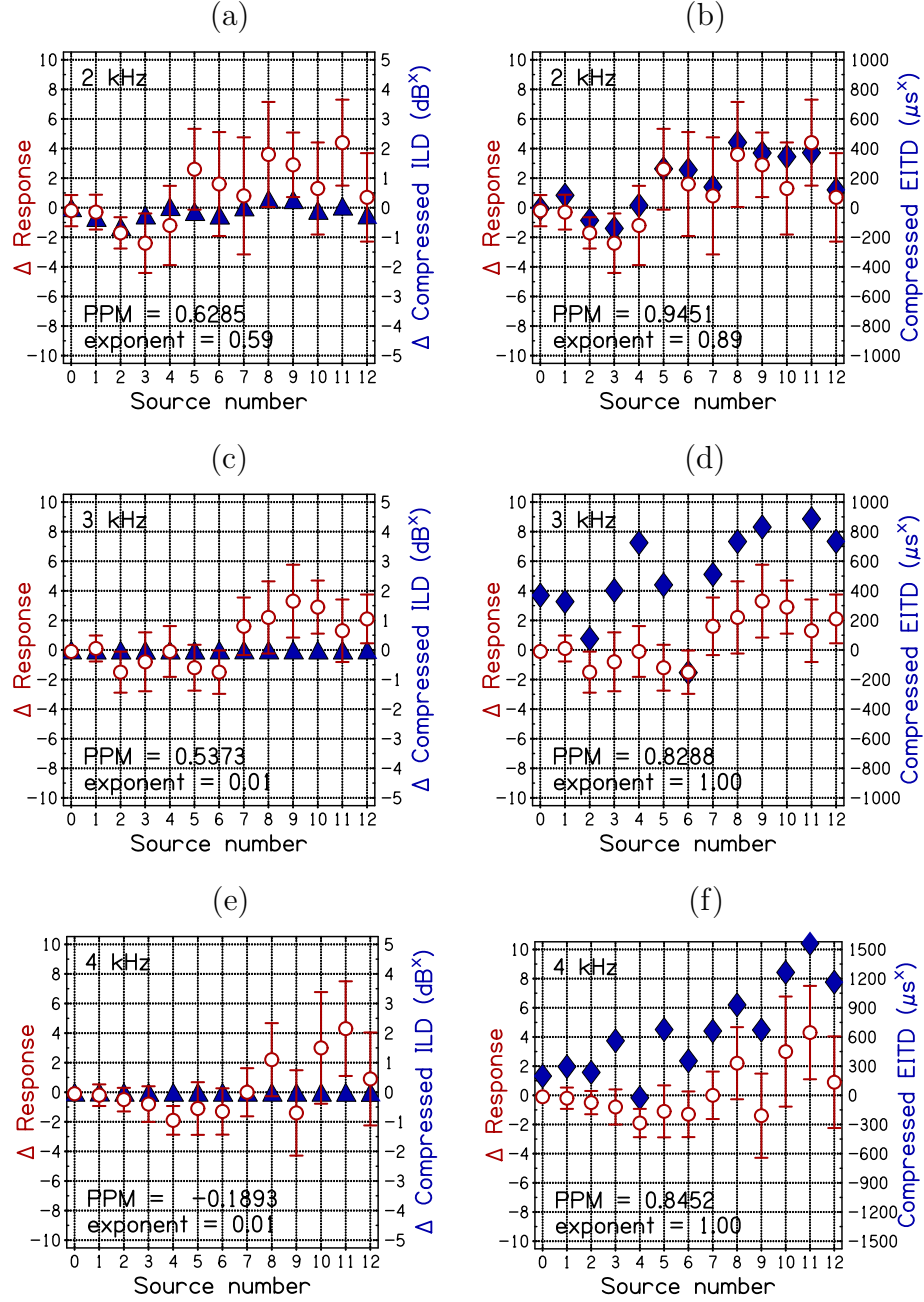


Figure 2.38 Same as Fig. 2.37 but for listener C.

Compressed Cues, Listener M

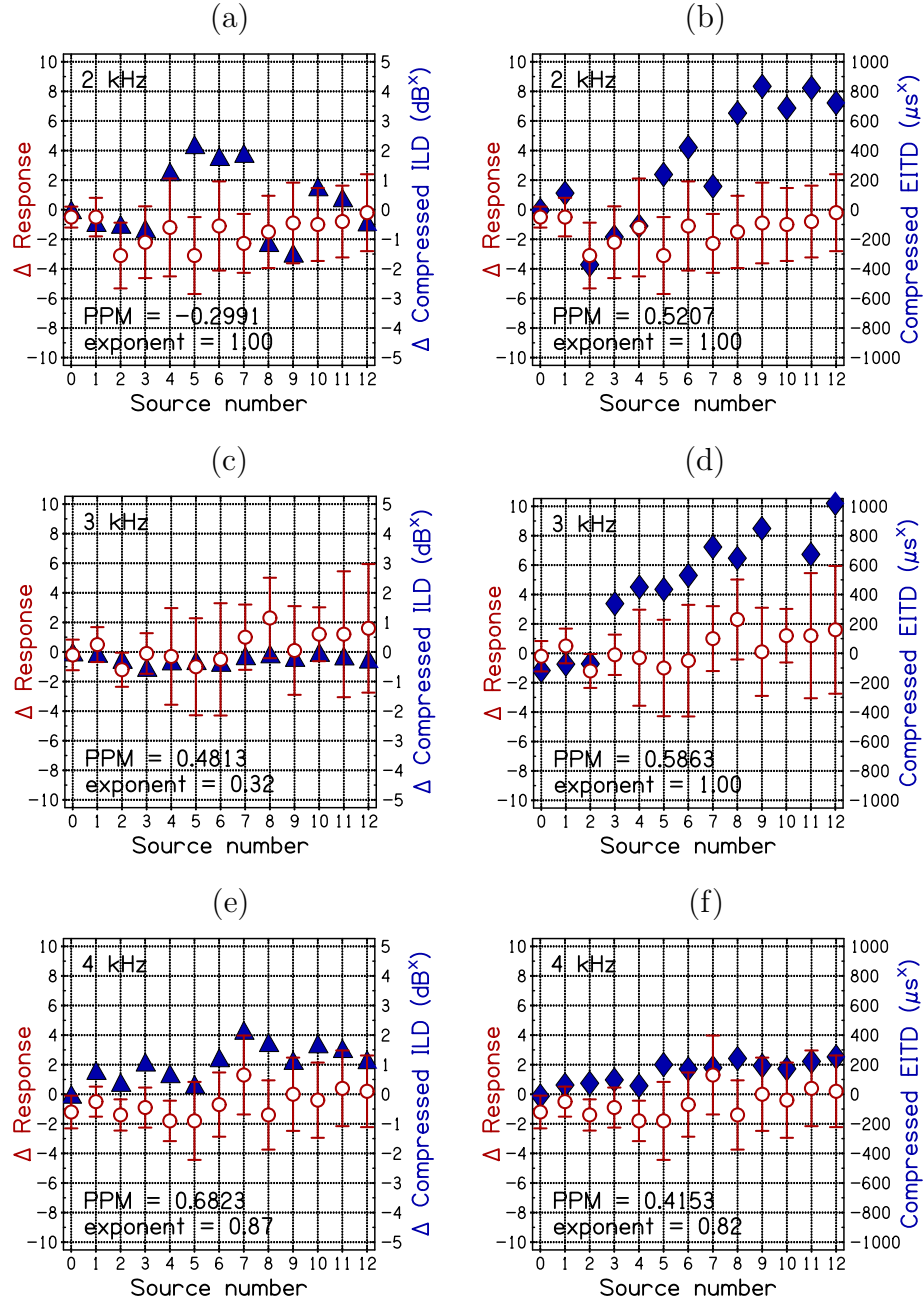


Figure 2.39 Same as Fig. 2.37 but for listener M.

Compressed Cues, Listener L

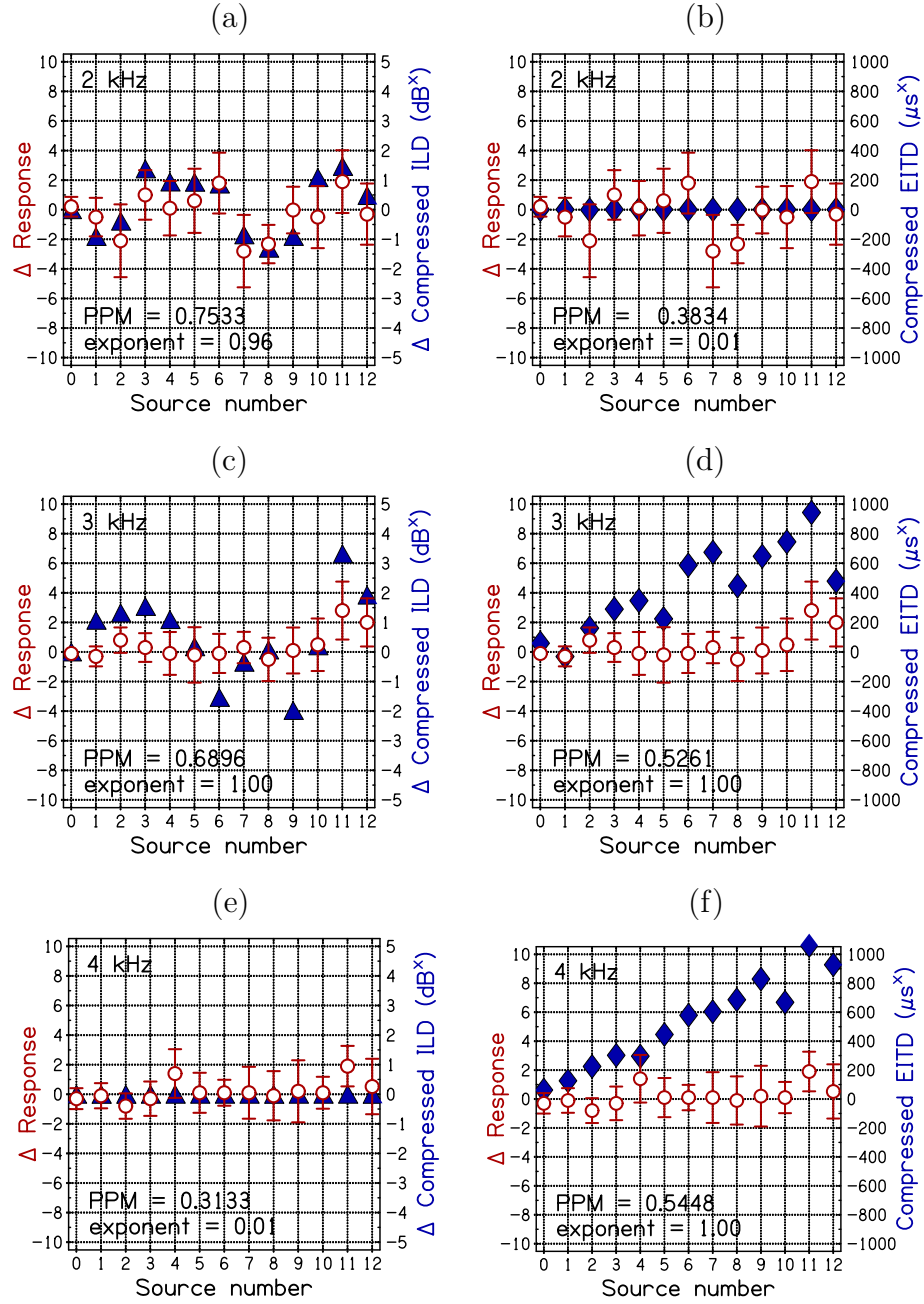


Figure 2.40 Same as Fig. 2.37 but for listener L.

Compressed Cues, Listener V

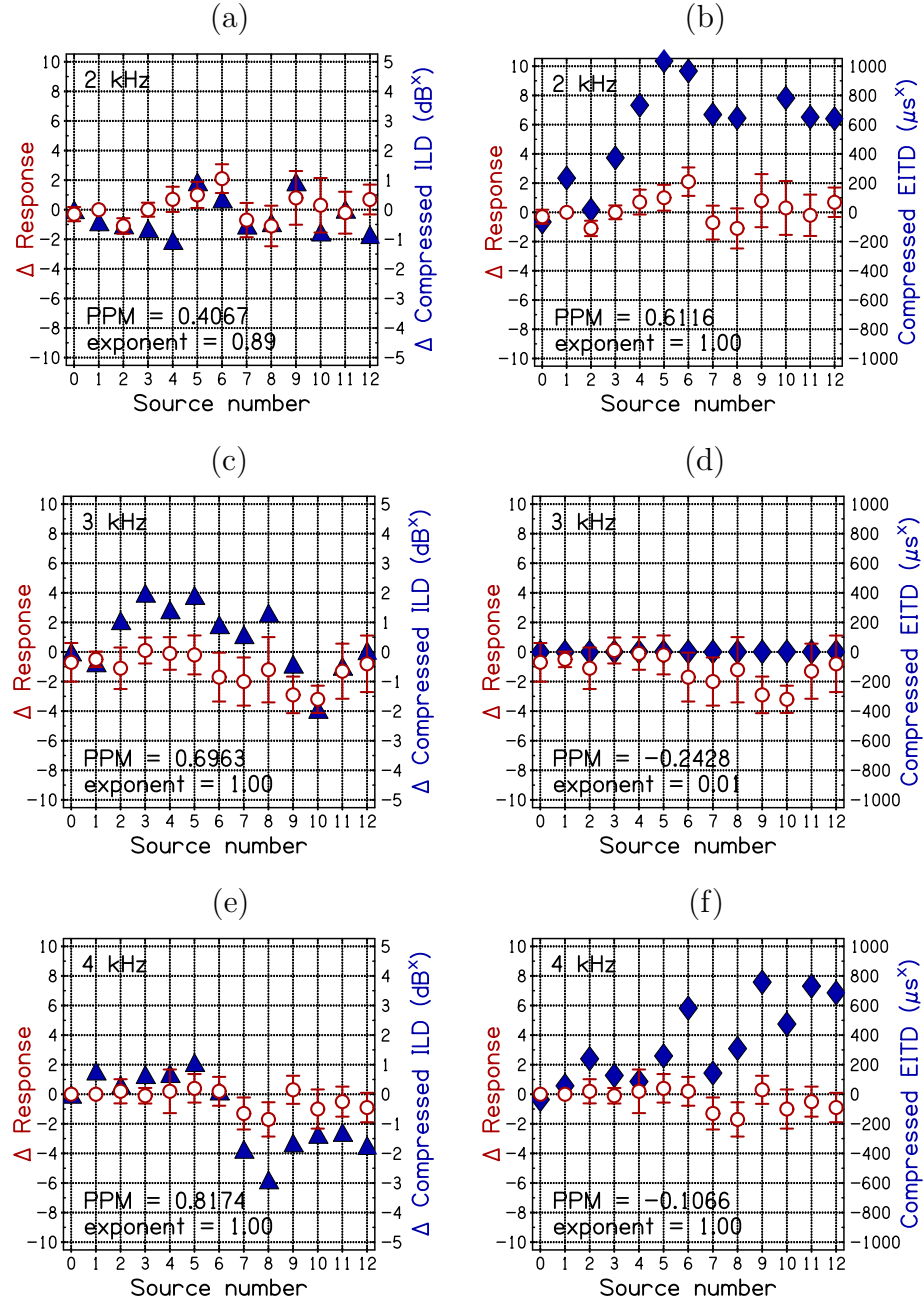


Figure 2.41 Same as Fig. 2.37 but for listener V.

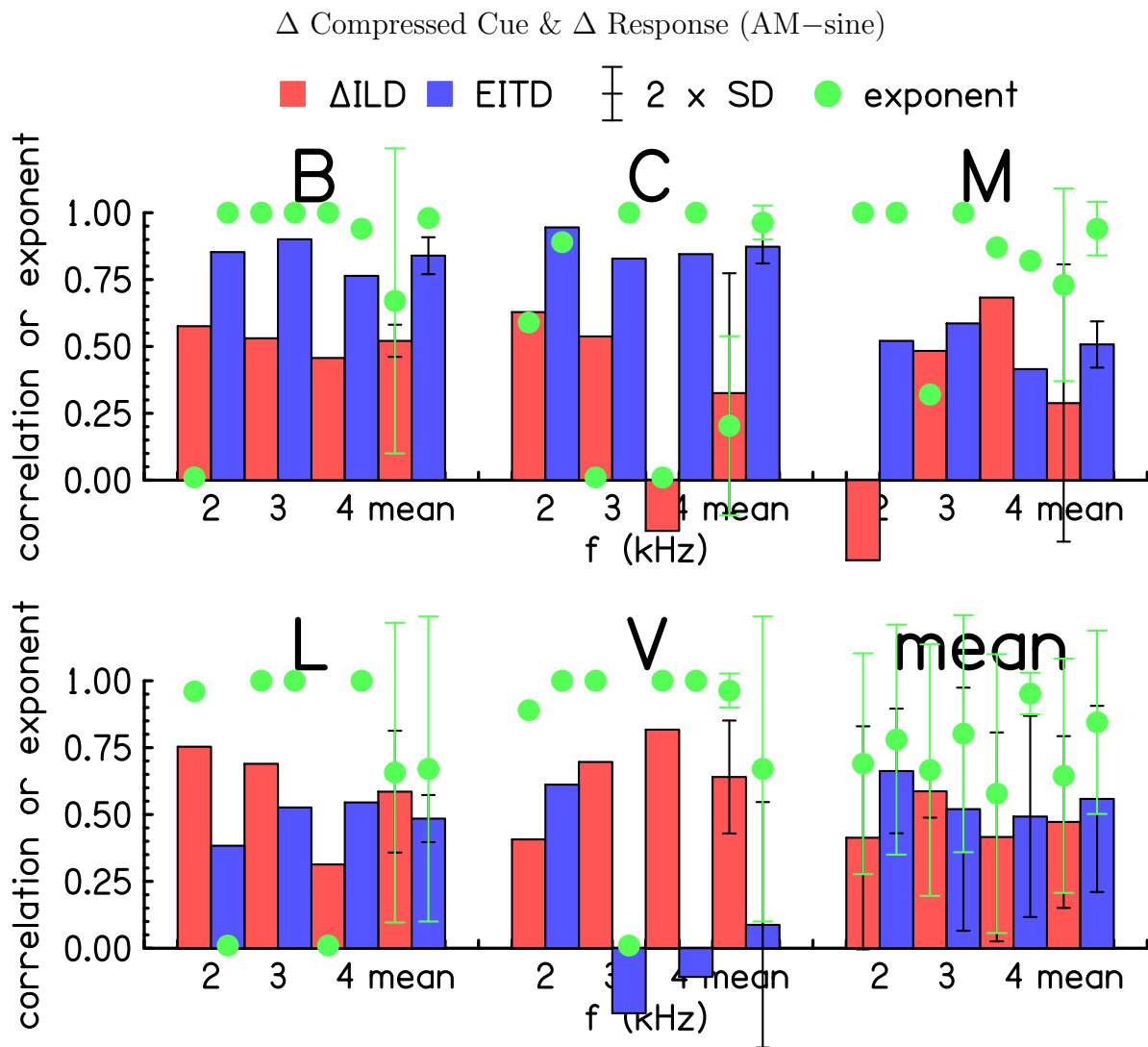


Figure 2.42 Maximized Pearson product-moment (PPM) correlation coefficients for the change in AM and sine responses and the change in compressed ILD are shown in red. Correlation coefficients for the change in AM and sine responses and the compressed EITD are shown in blue. The green circles indicate the compression exponent. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.

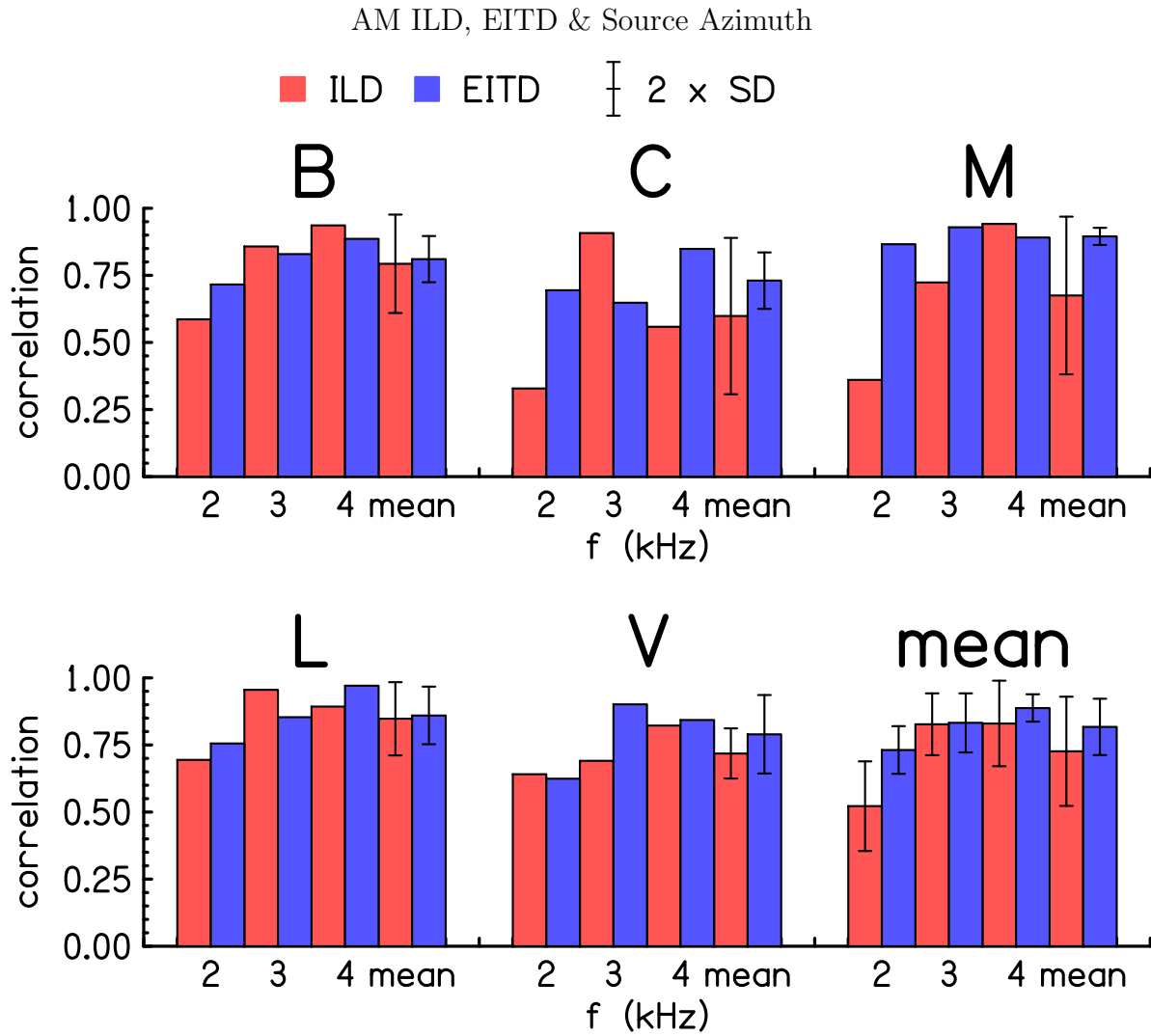


Figure 2.43 Pearson product-moment (PPM) correlation coefficients for AM tones. Correlations between ILD and source azimuth are in red. Correlations between EITD and source azimuth are in blue. For each listener, the three frequencies and the mean and standard deviation across frequencies are shown. In the lower right is the mean and standard deviation across all listeners for each frequency and the mean and standard deviation across all listeners and frequencies. The error bars are two standard deviations in overall length.

for the individual differences. It is entirely possible that listeners are processing ILD and EITD differently.

What is shown by Fig. 2.29 is that adding amplitude modulation nearly never hurts listeners, and is sometimes beneficial. Even though the reliability of the EITD cue itself is often flawed, because the ILD cue suffers from non-monotonicity, and because the EITD cue tends to be strongest after the peak of the ILD curve, it is easily understandable how AM is almost always beneficial.

2.3.4 AM Quality (Problem 1)

The key to understanding individual differences may lie in analyzing how AM quality (problem 1) affected listener responses [1]. Perhaps listeners B and C had higher AM quality, and were therefore better able to utilize the EITD cue than the other listeners. AM quality was analyzed in several ways: First, the effect of m and envelope coherence on the change in response for loudspeakers 8–12. Then the effect m and envelope coherence on the strength of EITD. Finally, the effect of m and envelope coherence on the response error ratio.

2.3.4.1 Change in response vs. AM quality

Because the EITD should be most useful at large azimuths (past the peak of the ILD), this analysis will be restricted to that region. These angles correspond to about 60° to 90° , or speakers 8–12. Here the EITD is large, so if its quality is reduced, a small percentage change in listener localization response should be easier to measure than at smaller angles where there are fewer speakers near the midline for a listener to respond to. Also at these large azimuths, the EITD and ILD begin to contradict each other, so the role of the EITD on response should be more pronounced.

Figures 2.44 and 2.45 show how response changed (AM – sine) for the left (far) ear and right (near) ear, respectively. The left ear has a much larger spread in m than the right ear, so the AM quality’s effect on it should stronger (if at all). However there seems to be very little, if any, systematic trend in these data for either ear as the correlations for the left and right ears are $r = -0.1031$ and $r = -0.0973$, respectively. This might be because all of the m values here are too high to hinder the ability to use EITD. The lowest value for the left ear is about 0.75, and for the right ear is not much less than 1. Although over-modulation is said to be AM of lower quality than the SAM signal, it is possible that in some circumstances it is at least as useful to listeners as 100% AM. However, this experiment was not designed to rigorously cover a particular range of m values. Rather it is a representation of what a listener in free field would encounter when listening to SAM tones at various azimuths. Although m is certainly affected in free field for both ears, it may be that no trend can be seen because m simply is always too high. In fact, Nuetzel and Hafter (1981) [40] showed that for 4-kHz SAM tones modulated at 150 and 300 Hz, the just-noticeable-difference (JND) in the EITD does not vary greatly with m when m is greater than about 0.5. The JND increases suddenly as m decreases below about 0.4. The value of m is greater than 0.5 for speakers 8–12, as seen in Figs. 2.44 and 2.45, and this remains true across all loudspeakers as seen in the histogram from Fig. 2.8.

Another hypothesis is that the interaural envelope coherence—the peak of equation 2.35—should affect listener responses. Since envelopes in the left and right ear must be compared during the processing of EITD, one might expect that if the envelopes are too dissimilar, the EITD would not be able to be detected as well. Figure 2.46 shows how response changed (AM – sine) vs. the envelope coherence for speakers 8–12. Surprisingly, the envelope coherence is always pretty high—usually higher than 0.99. The lowest envelope

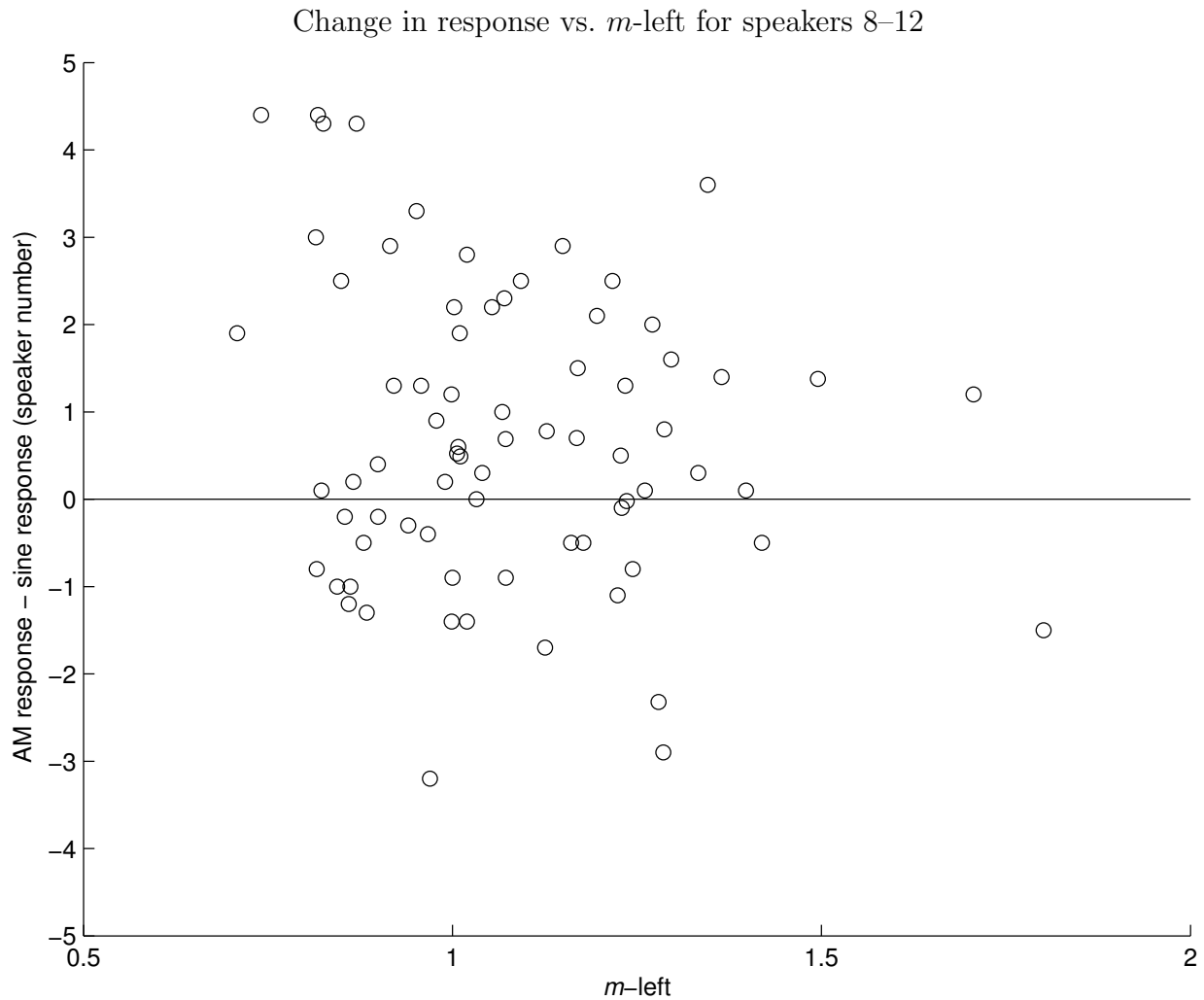


Figure 2.44 Change in response vs. m -left ear. All five listeners are shown here. Each point represents the mean change in response across 10 trials vs. the mean m across 20 intervals for the listener's left (far) ear. The correlation is $r = -0.1031$ and $r^2 = 0.01063$.

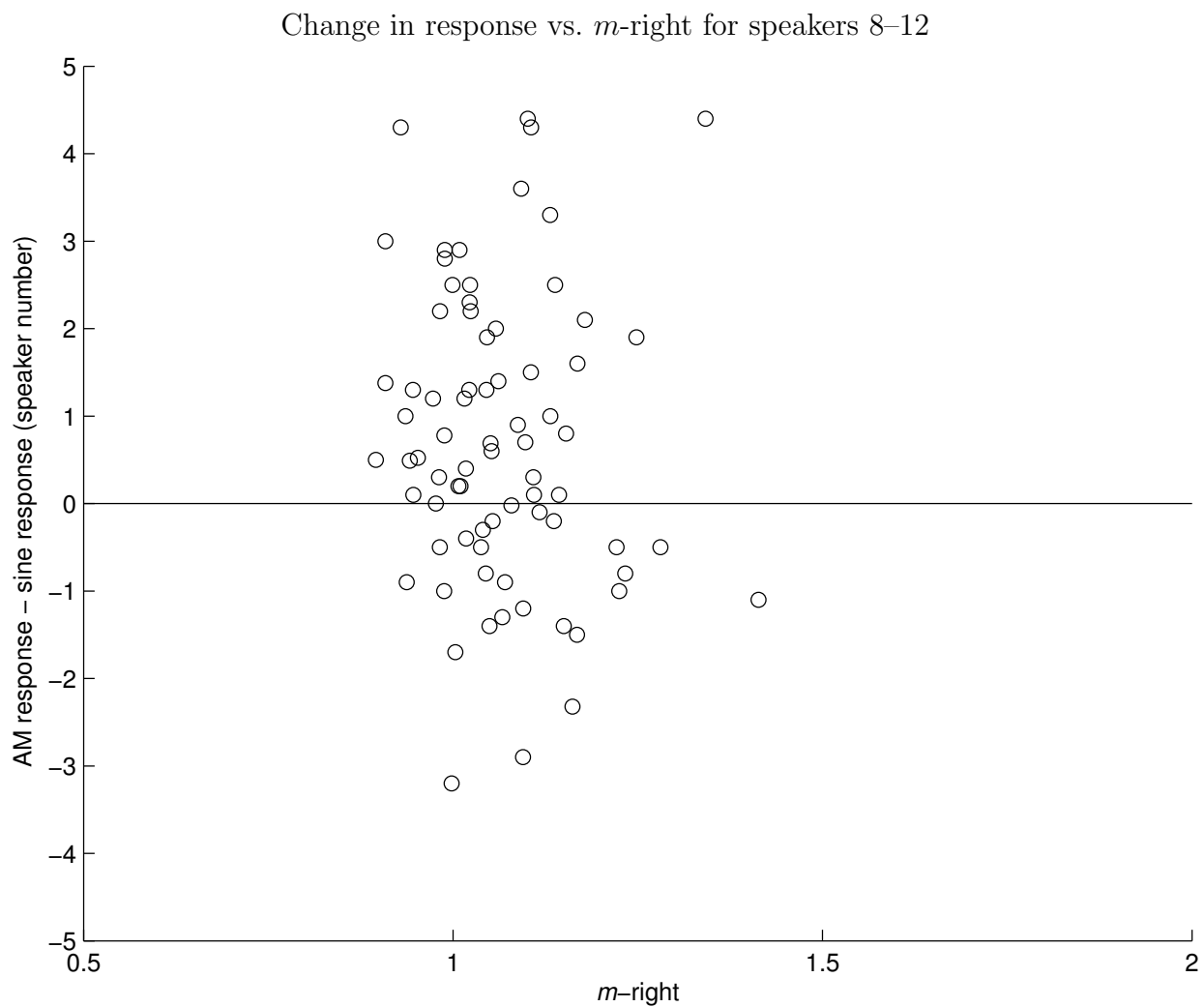


Figure 2.45 Change in response vs. m -right ear. All five listeners are shown here. Each point represents the mean change in response across 10 trials vs. the mean m across 20 intervals for the listener's right (near) ear. The correlation is $r = -0.0973$ and $r^2 = 0.00947$.

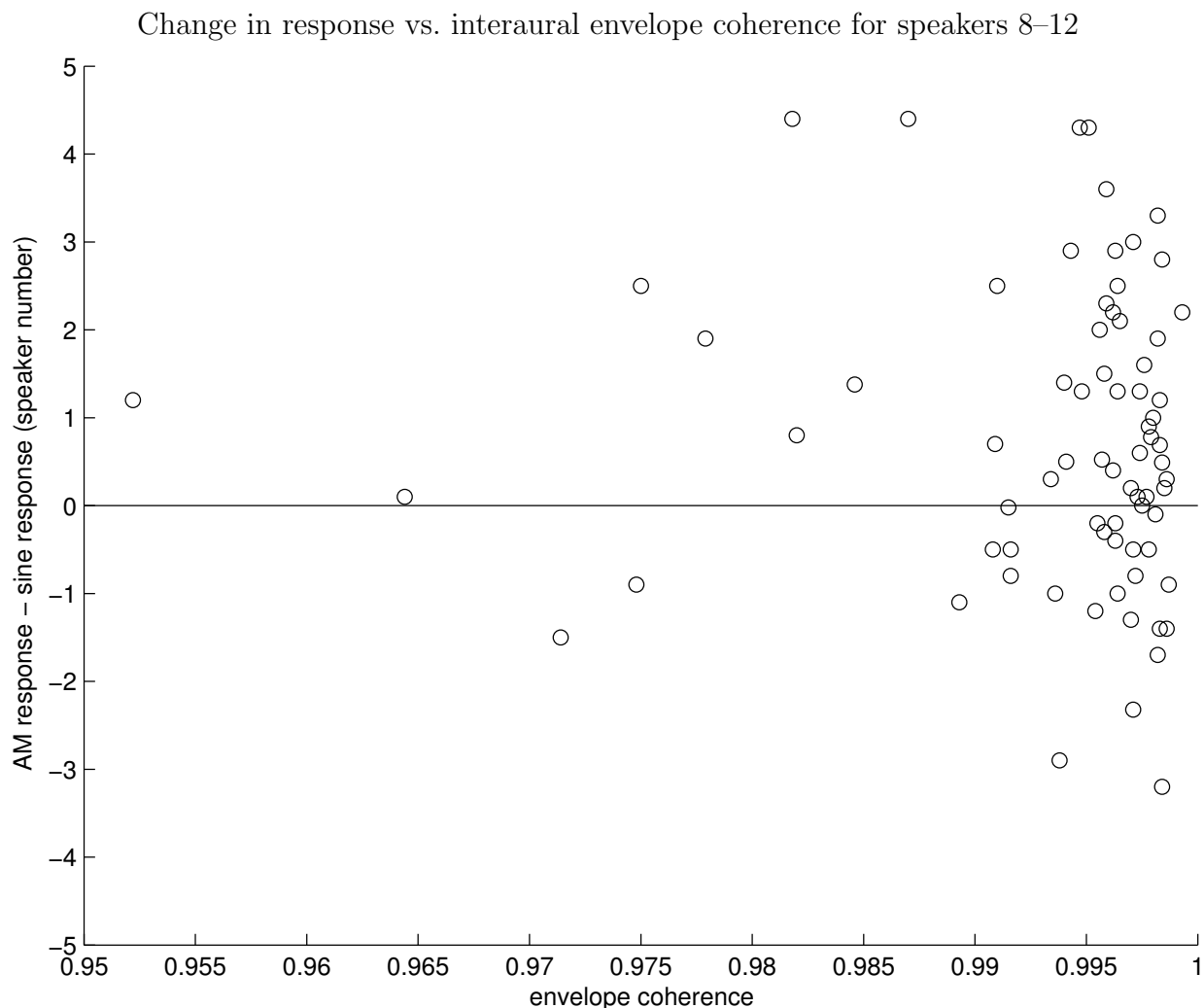


Figure 2.46 Change in response vs. interaural envelope coherence. All five listeners are shown here. Each point represents the mean change in response across 10 trials vs. the mean m across 20 intervals. The correlation is $r = -0.0633$ and $r^2 = 0.00440$.

coherence is no less than 0.95. This is despite the presence of unusual pairs of waveforms such as those in Figs. 2.3 and 2.4. The correlation is $r = -0.0633$. Unfortunately, just as was the case with m , it may be that this particular experiment did not result in an envelope coherence which was low enough to show a systematic effect on the change in response.

In an effort to find some effect, several variations on this analysis were conducted. Table 2.4 shows a summary of these correlations. Another hypothesis is that $m = 1$ is the ideal

signal, and that any deviation from this will cause problems for a listener. So, all $m < 1$ were folded over 1 such that $m_{\text{folded}} = 1 + (1 - m)$. This was done for both ears. In both cases the correlations became even weaker, and no pattern was noticeable on the scatter plots. The next hypothesis was that the smallest m (left or right) is what matters. This produced the strongest correlation of all, $r = -0.2435$. However, the expectation is that as the smallest m increases, that the EITD cue should be more effective at increasing listener responses for speakers 8–12. The negative correlation here opposes this hypothesis. Yet since this correlation is still quite low, this hypothesis should not be discounted. Finally, the m (left or right) which was furthest from 1 was analyzed. This, too, resulted in a weak correlation.

r for speakers 8–12	
independent variable	r
m -left ear	-0.1031
m -right ear	-0.0973
envelope coherence	-0.0633
m -left, folded over 1	0.0043
m -right, folded over 1	-0.0387
smallest m (left or right)	-0.2435
m , furthest from 1 (left or right)	-0.0917

Table 2.4 The correlation coefficients, r , for Figs. 2.44–2.46 and four other tests. The data include all listeners and all frequencies. For each loudspeaker, the data are the change in mean responses (AM–sine) vs. the mean independent variable. For m -left and m -right folded over 1, values of $m < 1$ are transformed to values of $m > 1$ while keeping $|m - 1|$ the same.

2.3.4.2 The effect of AM quality on cue strength

The previous section did not take into account the value of the EITD. It also only looked at the mean for each loudspeaker, and grouped all listeners and carrier frequencies together. It may be possible to find cases where m or the envelope coherence affects the strength of the EITD on a more individualized basis. In Figs. 2.47–2.58 the change in response (individual trial minus mean sine response) is plotted vs. the EITD. The color of each data point represents either the m for the left or right ears, or the envelope coherence. The goal is to observe situations where data points deviate from the best fit lines, and to look for systematic effects. Most of the effects should coincide with the largest values of EITD. Let the cue-weight hypothesis be that trials with lower quality AM will tend to fall below the line of best fit, and trials with higher quality AM will tend to fall above the line of best fit. Additionally, this effect should increase with EITD. For panels (a) and (b) the cue-weight hypothesis is that under-modulated tones will tend to fall below the best fit line, and 100%-modulated or over-modulated tones will fall on or perhaps above the line of best fit. For panels (c), the cue-weight hypothesis is that lower coherence stimuli will tend to fall below the line of best fit, and higher coherence stimuli will tend to fall on or above the line of best fit.

In Fig. 2.47 the change in response increases suddenly around $750\ \mu\text{s}$. The left ear (far ear), though, shows that these happened to be some of the most under-modulated tones in these runs. The responses increased dramatically despite being somewhat under-modulated in the far ear. This does not mean that the under-modulation caused this, but it does oppose the cue-weight hypothesis in this case. The right ear shows a mixture of modulation values at high EITD—some less than zero and some greater. The envelope coherence is large

in most cases and no obvious trend is visible in this plot.

In Fig. 2.48 the slope of the best fit line has decreased, compared for Fig. 2.47. This indicates that the EITD cue was not as impactful at 3 kHz as it was at 2 kHz. This is likely because the ILD cue is more reliable at 3 than 2 kHz, so the sine responses are somewhat more saturated and there is less room for improvement. For the left ear (panel a) there does appear to be some consistency in how m is distributed on the plot. Since there are ten trials per loudspeaker, it is not surprising to see that there are similar groupings of EITD and m . But there does not appear to be a systematic distribution of over-modulated tones above the best fit line and under-modulated tones below the line. The right ear (panel b) shows m is usually closer to unity than the left ear. Again, though, the cue-weight hypothesis is not supported here. The envelope coherence (panel c) is always very high and no effect on the change in response can be seen. For listener C, the slope in Figs. 2.50–2.52 also decreases with increasing carrier frequency. An inspection of these plots shows no support for the cue-weight hypothesis.

In Fig. 2.49, the slope of the best fit line continues to decrease compared to the the lower carrier frequencies in Figs. 2.47 and 2.48. Once again the left ear has a wider range of m than the left. No clear trend is noticeable for m -left ear, m -right ear, or the envelope coherence.

For listener M, shown in Figs. 2.53–2.55, the slopes are smaller than for listeners B and C, which is consistent with the fact that listener M was less influenced by the EITD. Even so, it may be interesting to see if m or the envelope coherence played a roll in the change in response. There are some trends in how m is distributed along the horizontal axis—for example, Fig. 2.17, panel (a). But again, this relates m to the values of the EITD itself and therefore roughly to the azimuth. What is not seen is that for over-modulated tones, changes

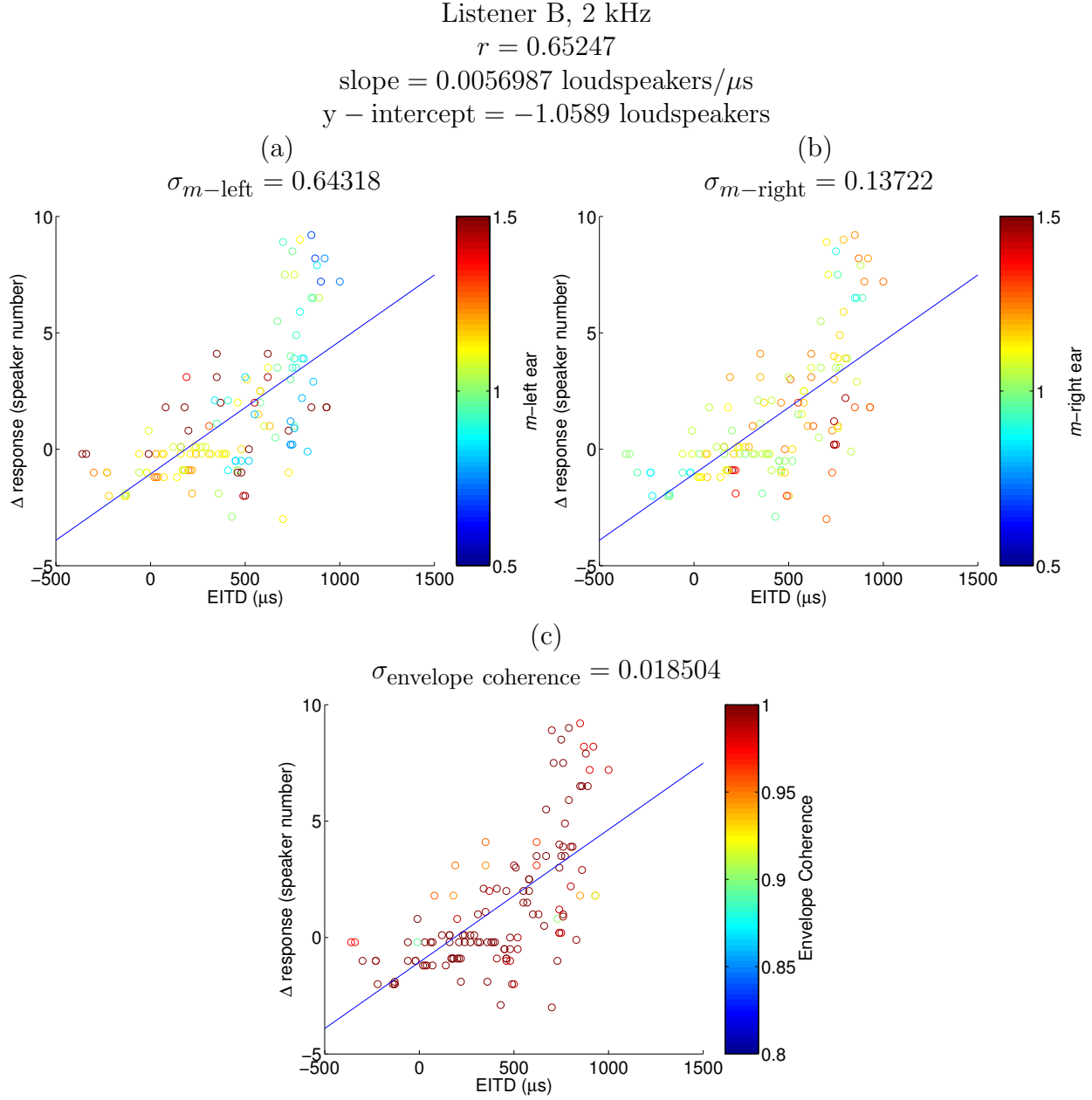


Figure 2.47 Changes in response vs. EITD with AM quality for listener B at 2 kHz. All plots display the same data and differ in the AM quality metric shown in color. The vertical axis indicates AM responses for individual trials minus the mean sine response for the same loudspeaker. The horizontal axis indicates the EITD measured in the individual AM trials. The correlation coefficient— r —, slope, and y-intercept for the best fit line are above the plots. The color scale in plot (a) indicates the value of m in the left ear in the individual AM trials. The color scale in plot (b) indicates the value of m in the right ear in the individual AM trials. The color scale in plot (c) indicates the value of the envelope coherence in the individual AM trials. The standard deviations, σ , of the AM quality are shown above the plots.

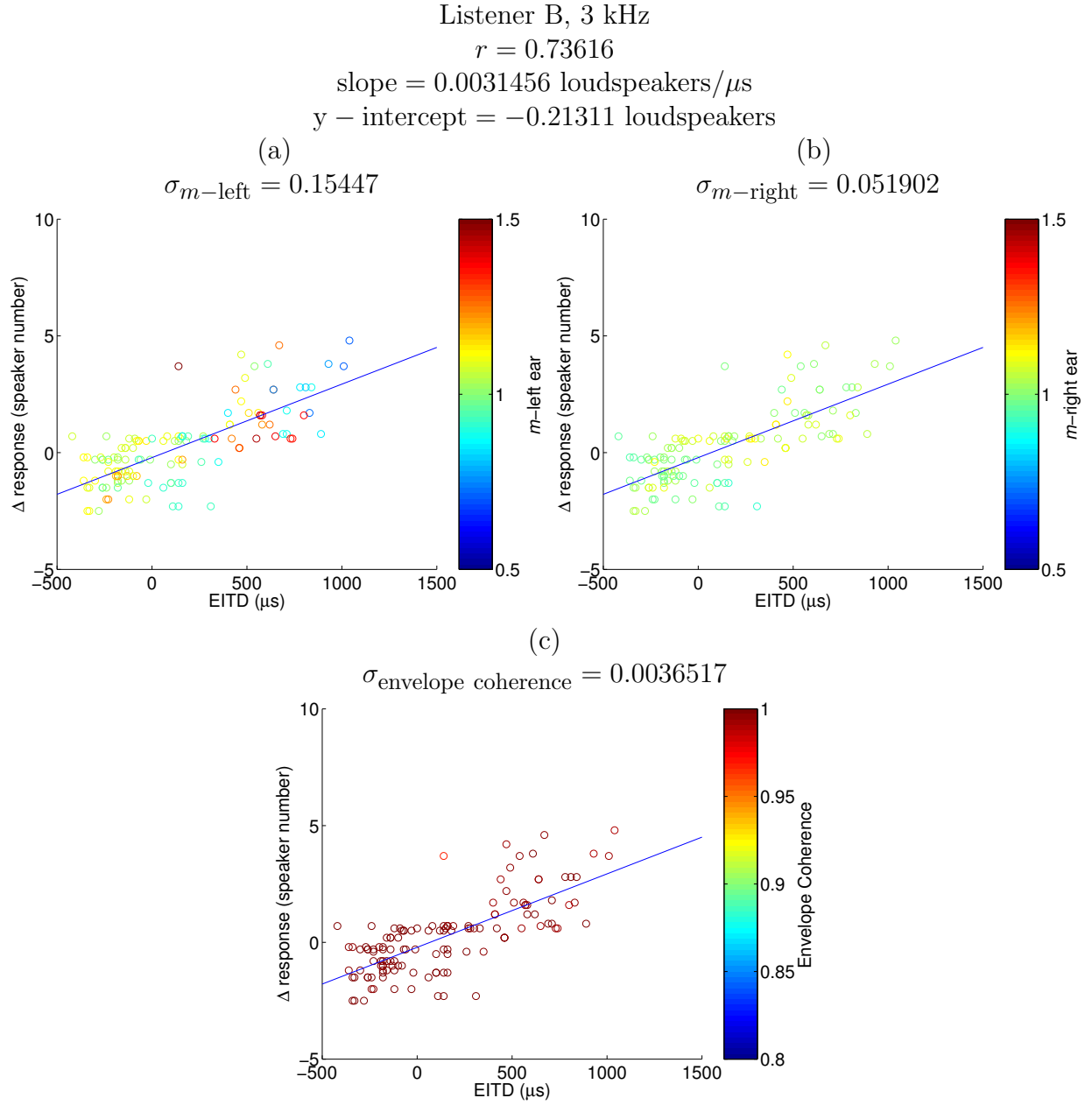


Figure 2.48 Same as Fig. 2.47 but for listener B at 3 kHz.

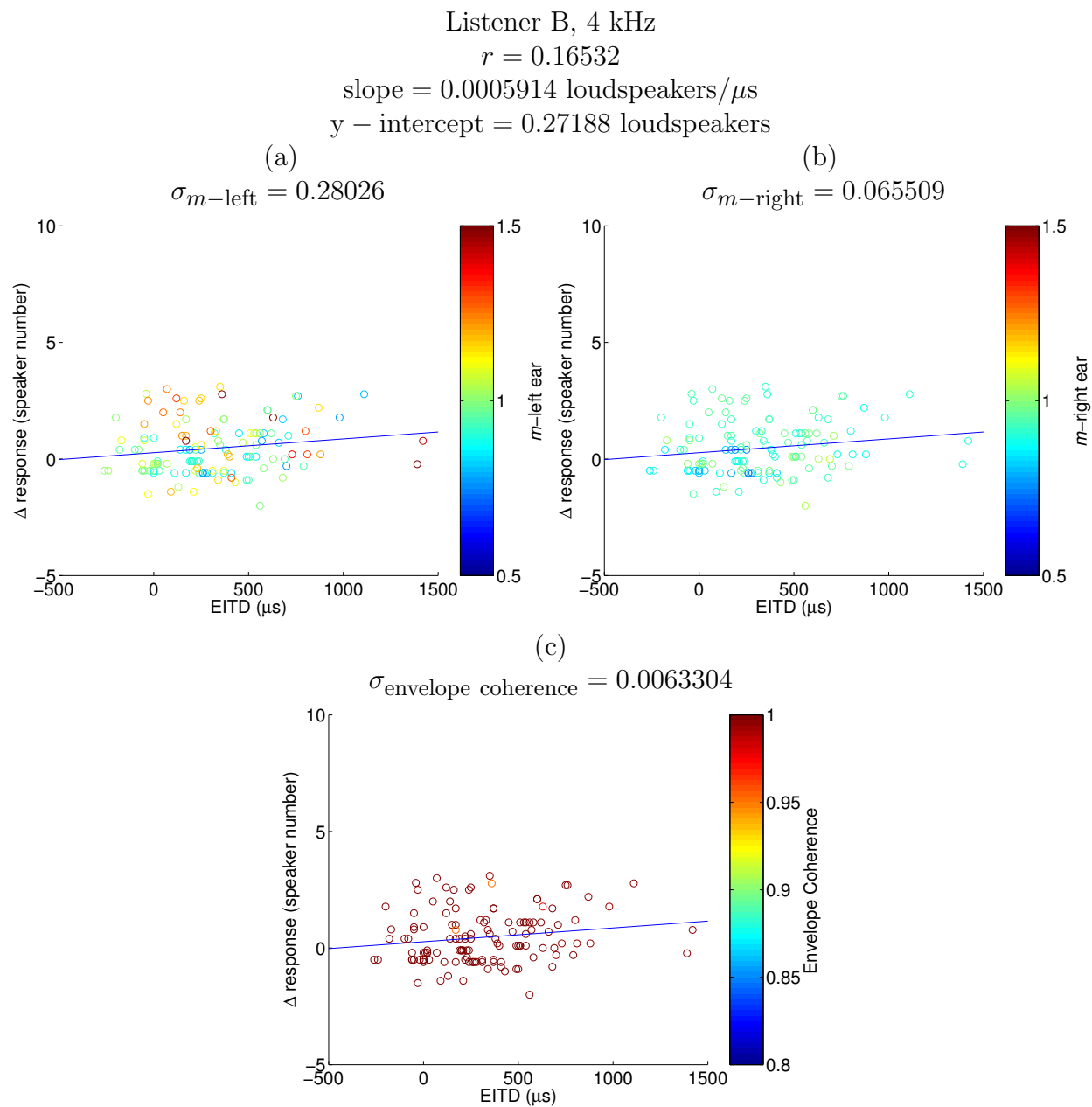


Figure 2.49 Same as Fig. 2.47 but for listener B at 4 kHz.

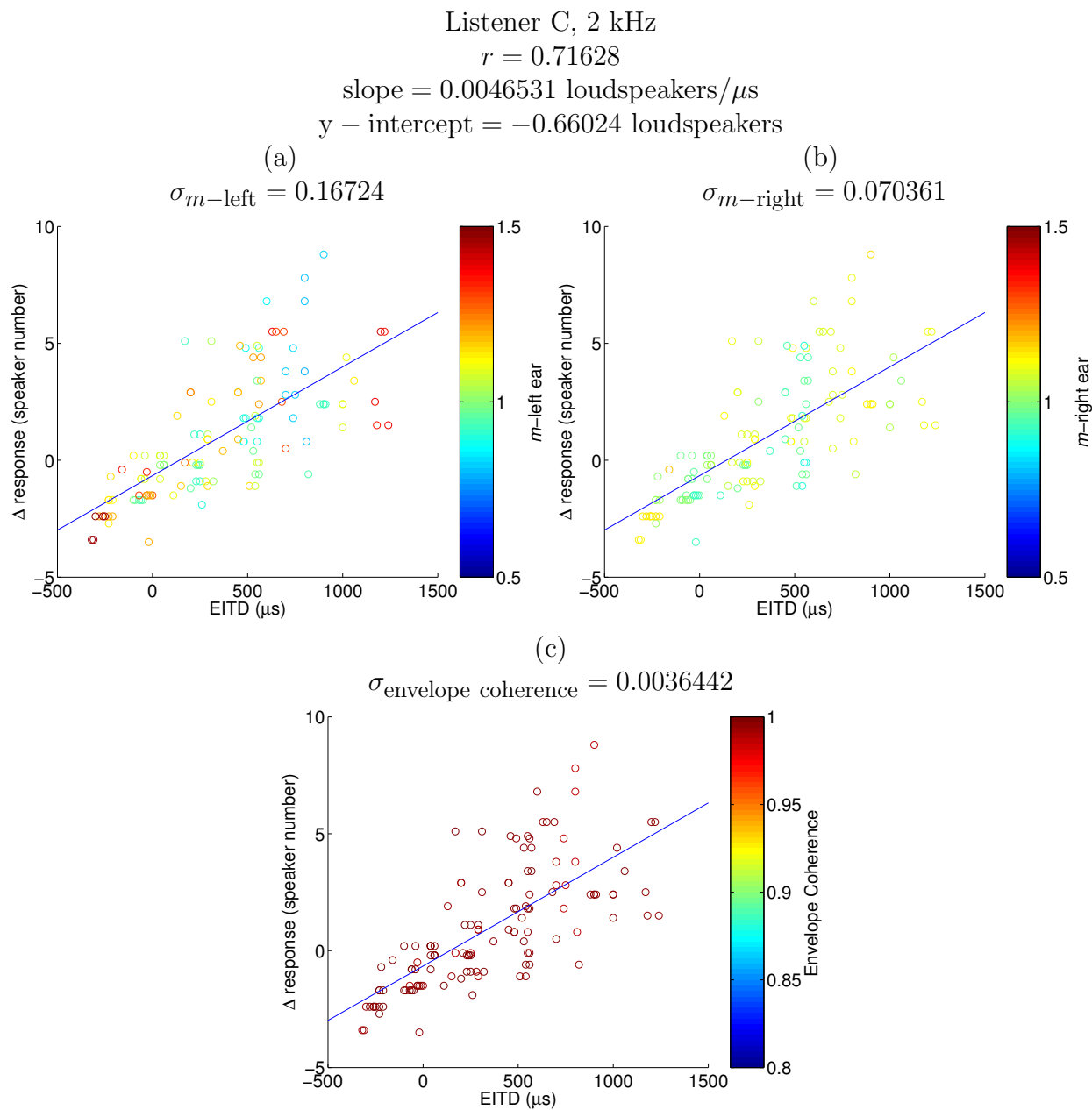


Figure 2.50 Same as Fig. 2.47 but for listener C at 2 kHz.

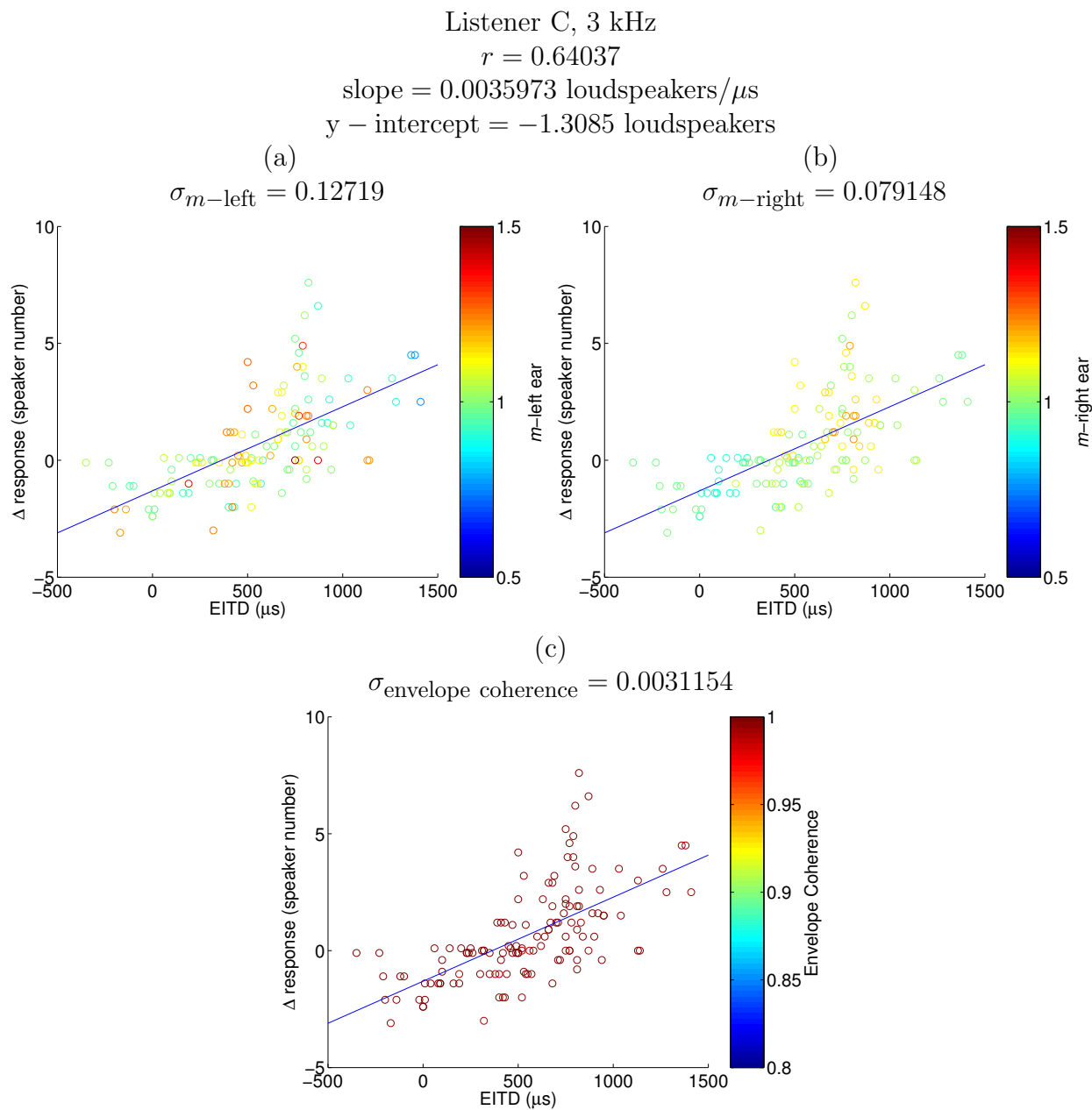


Figure 2.51 Same as Fig. 2.47 but for listener C at 3 kHz.

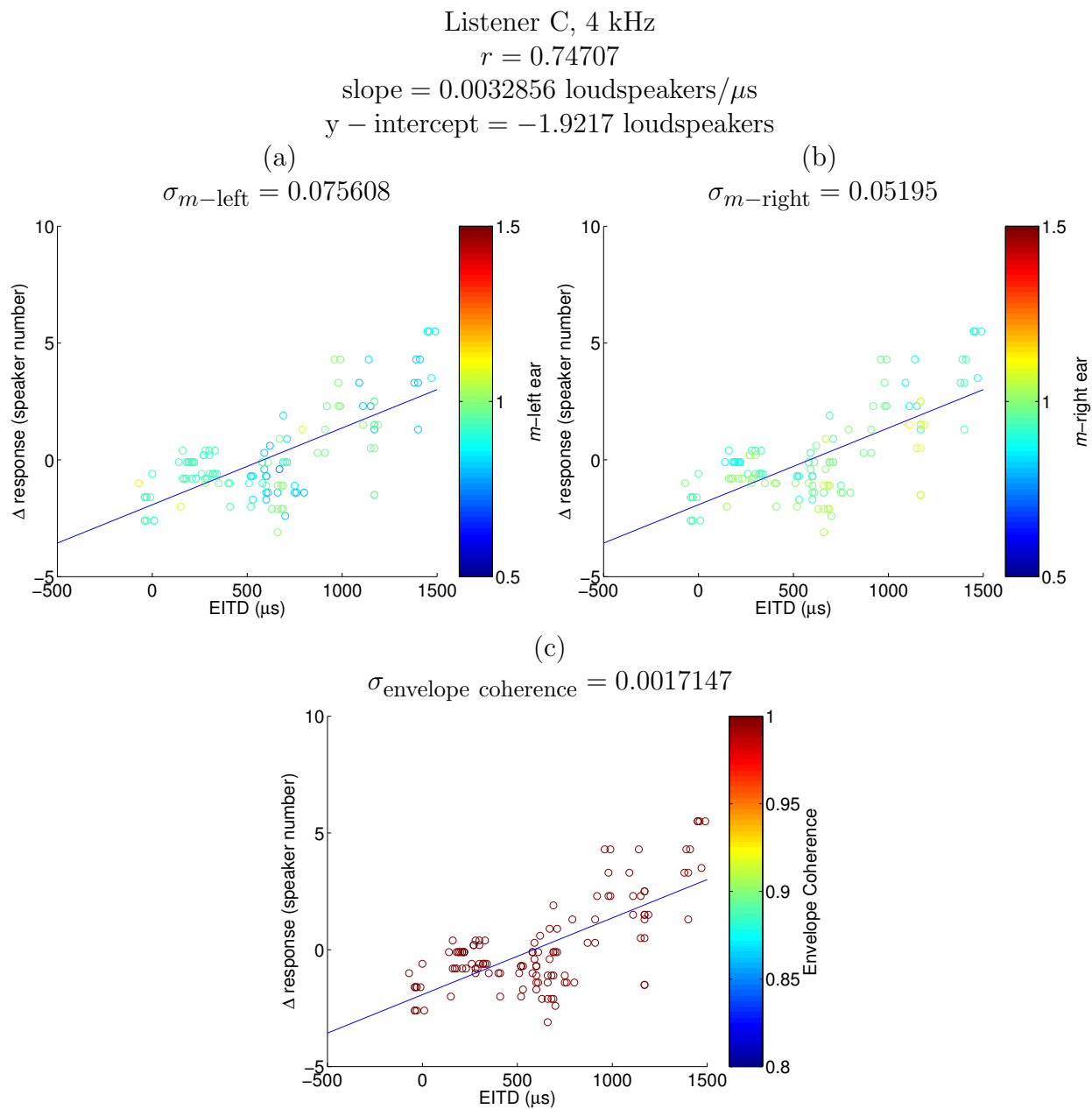


Figure 2.52 Same as Fig. 2.47 but for listener C at 4 kHz.

in response grow above the line of best fit with increasing EITD, and for under-modulated tones, change in response grow below the line of best fit with increasing EITD. Similarly, there is no trend seen with the envelope coherence. For listeners L, shown in Figs 2.56–2.58, the slopes are small like for listener M. Unfortunately, the cue-weight-hypothesis cannot be confirmed by inspecting these plots.

For listener V, shown in Figs. 2.59–2.61, the scatter plots are the most unusual. The 2 kHz plots shown in Fig. 2.59 are the only ones with a reasonably large slope, meaning that the EITD is changing this listener’s responses. At 3 kHz the slope is negative. One can see why perhaps the slope was negative in the 3 kHz case. Figure 2.24 shows the ILD for the sine tones in panel (a) and the AM tones in panel (b). In this instance, the AM ILD is noticeably smaller and appears to be driving many of the changes in response to be negative. That’s not to say that the EITD had been totally ignored. The correlation between ILD and response was 0.9458 for the sine tones and 0.8832 for the AM tones, indicating that another variable was in play. For 4 kHz, shown in Fig. 2.61, the slope is very flat. This is not as easy to explain by comparing the plots in Fig. 2.25. However, as is the case for 3 kHz, the correlation between ILD and response is lower for the AM tones in panel (b) than for the sine tones in panel (a). So the EITD is surely not being ignored by listener V, but it cannot be shown to be weighted by m in either ear nor by the envelope coherence.

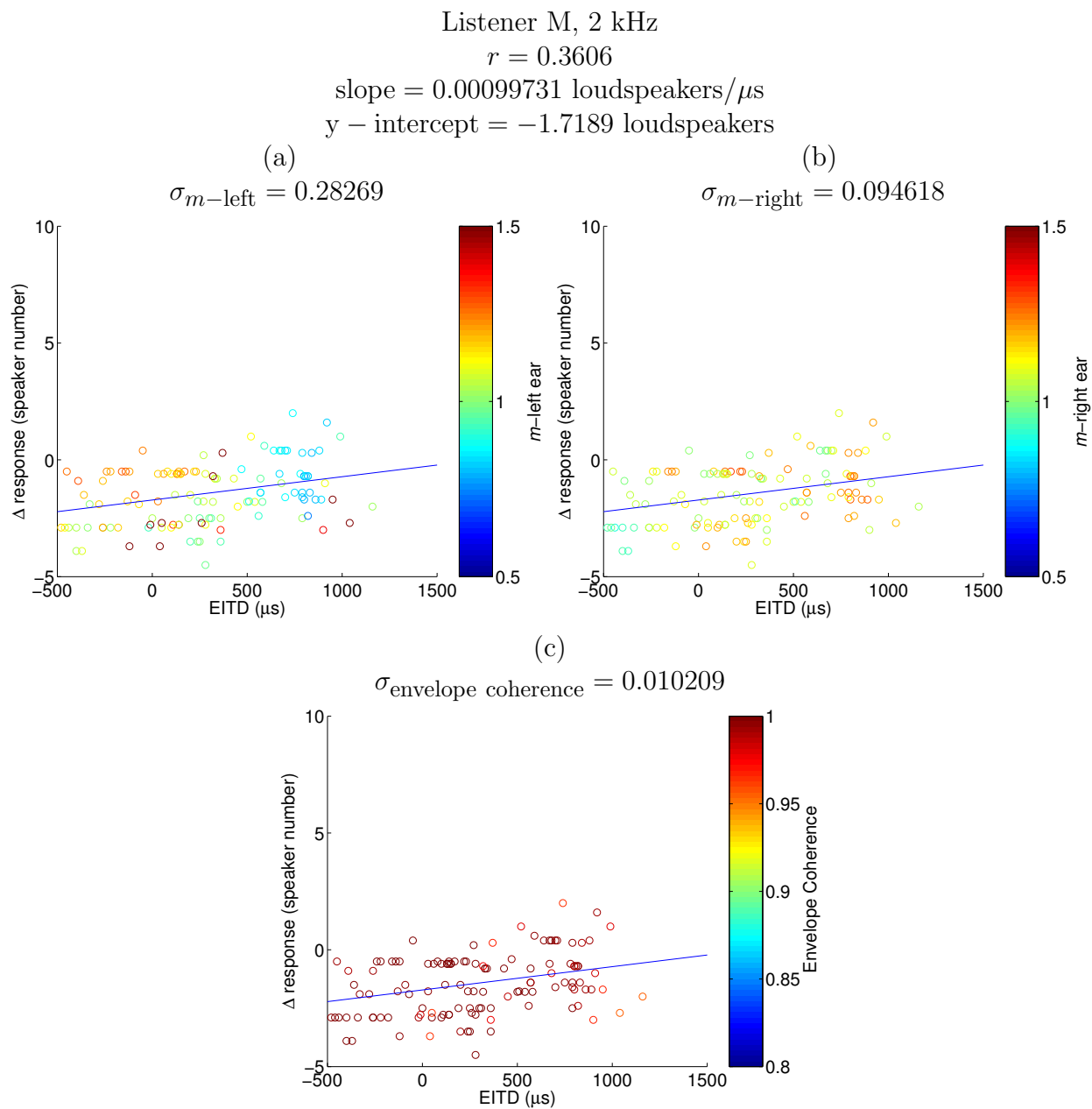


Figure 2.53 Same as Fig. 2.47 but for listener M at 2 kHz.

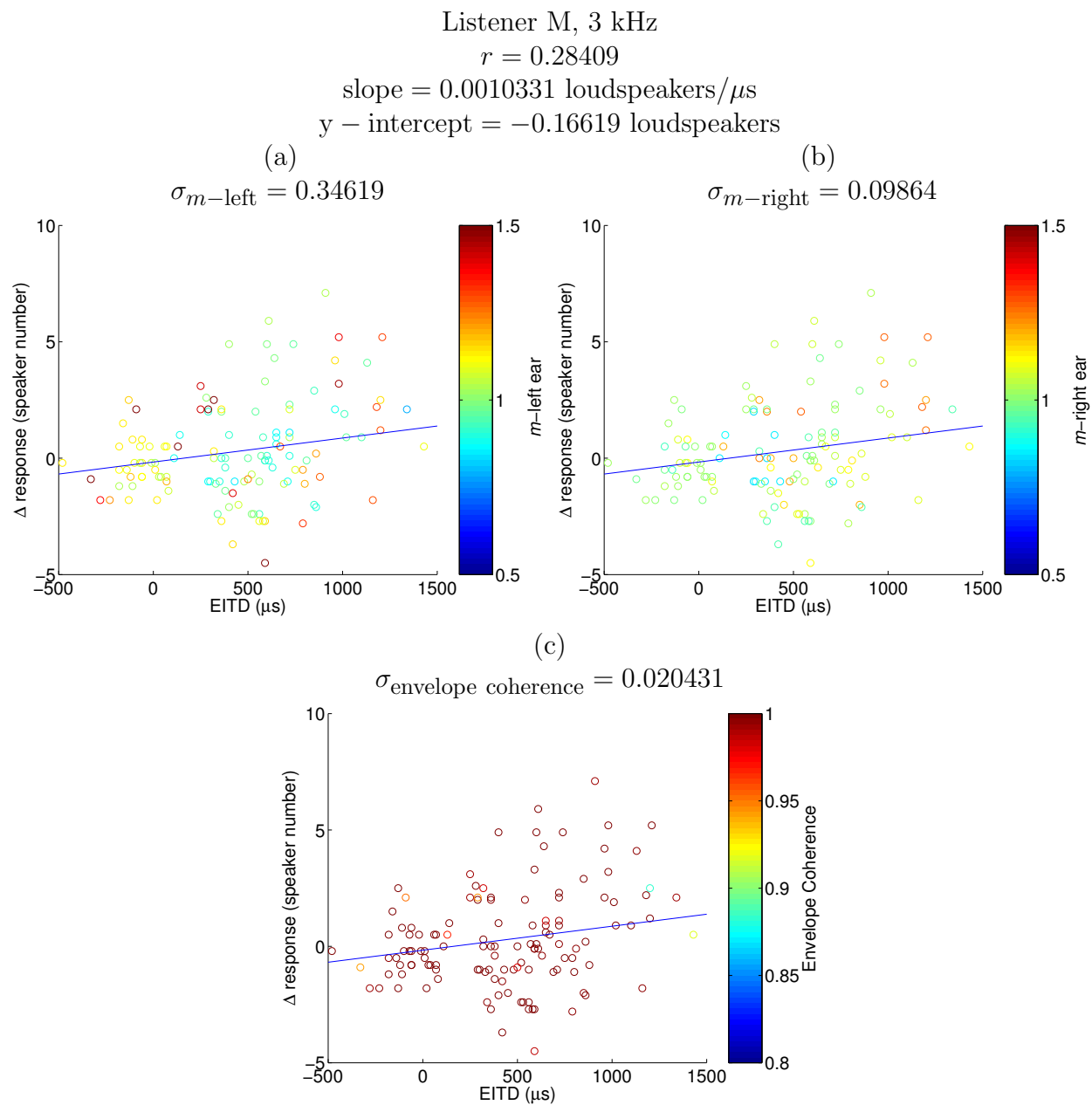


Figure 2.54 Same as Fig. 2.47 but for listener M at 3 kHz.

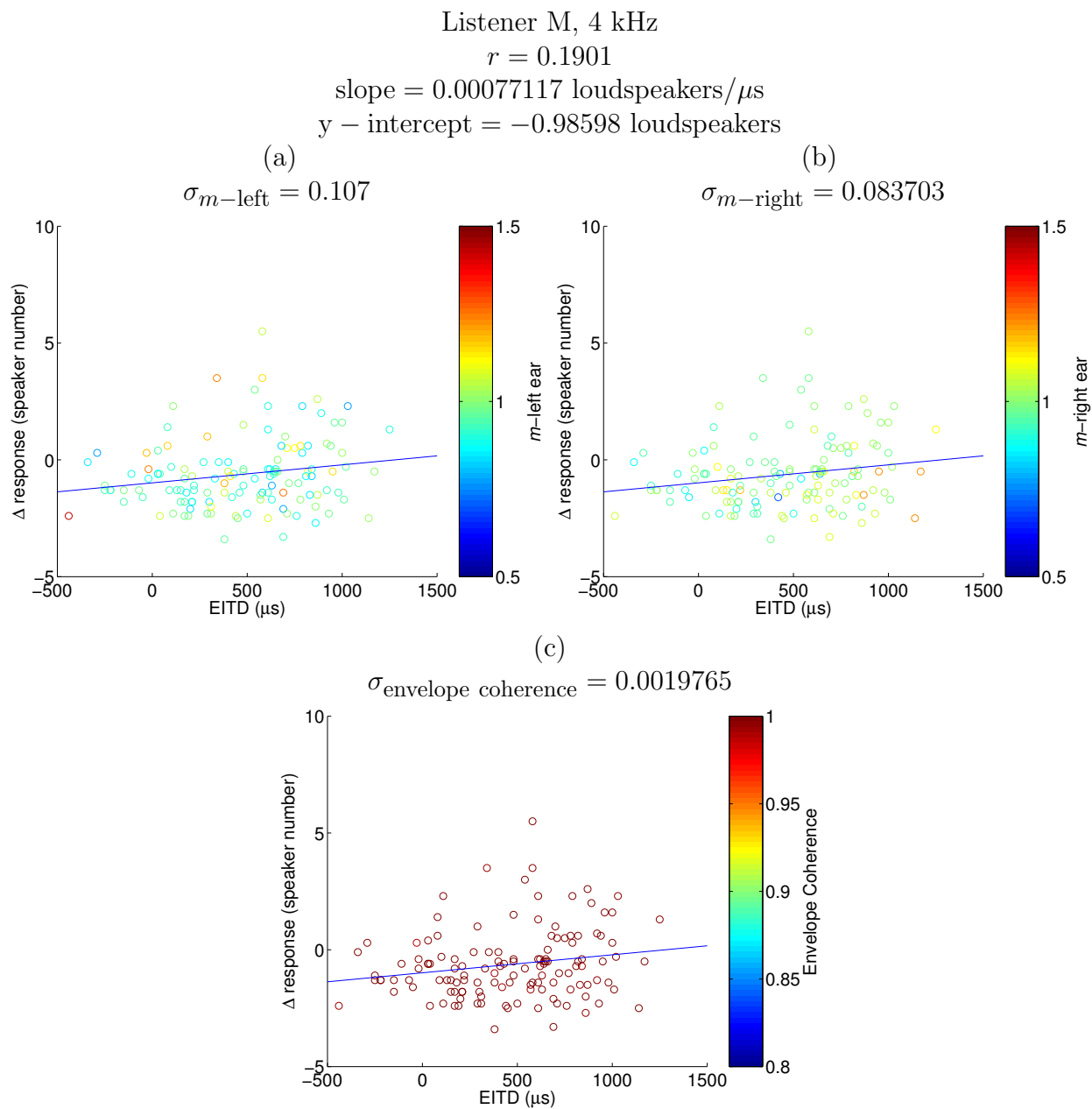


Figure 2.55 Same as Fig. 2.47 but for listener M at 4 kHz.

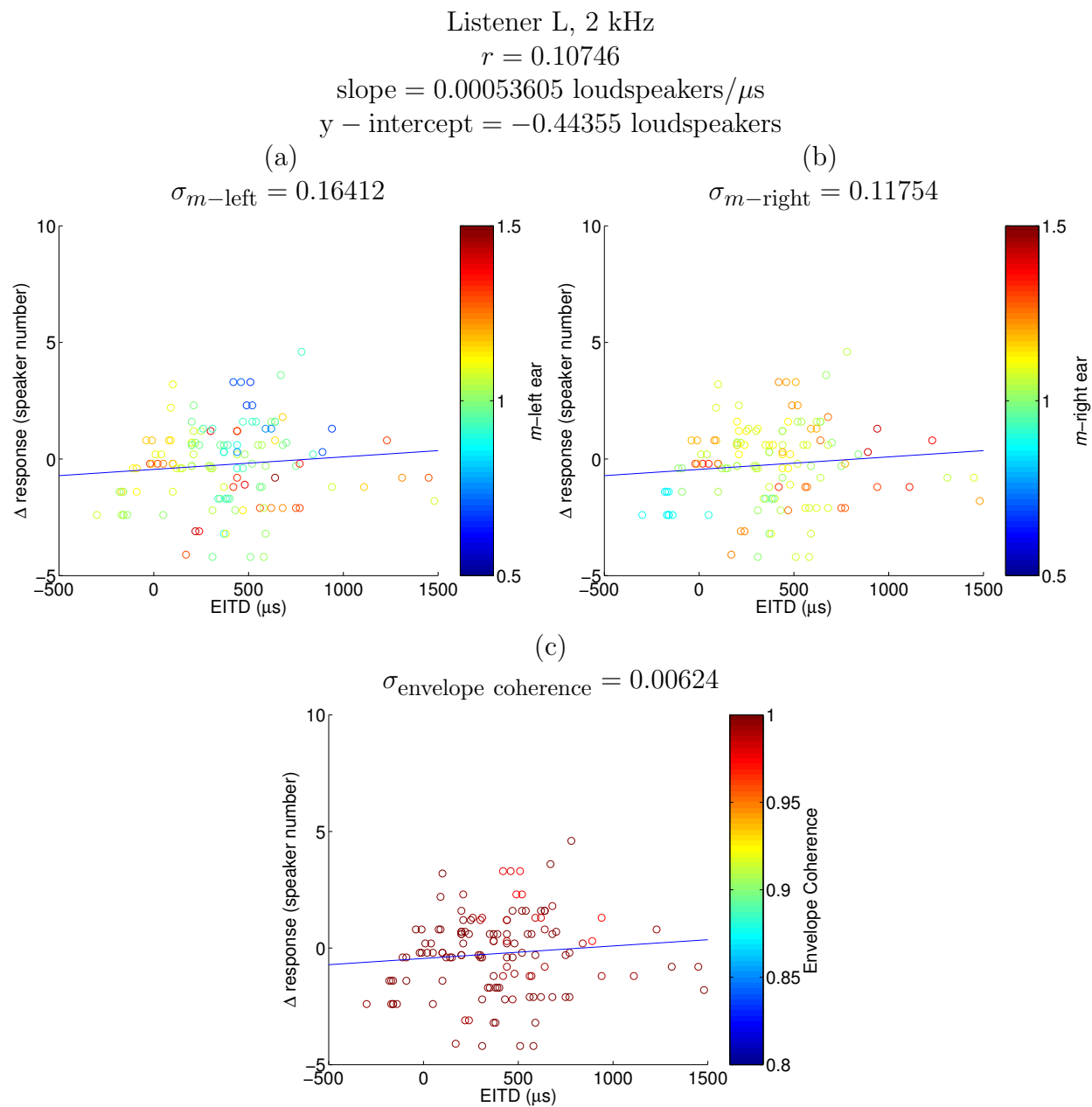


Figure 2.56 Same as Fig. 2.47 but for listener L at 2 kHz.

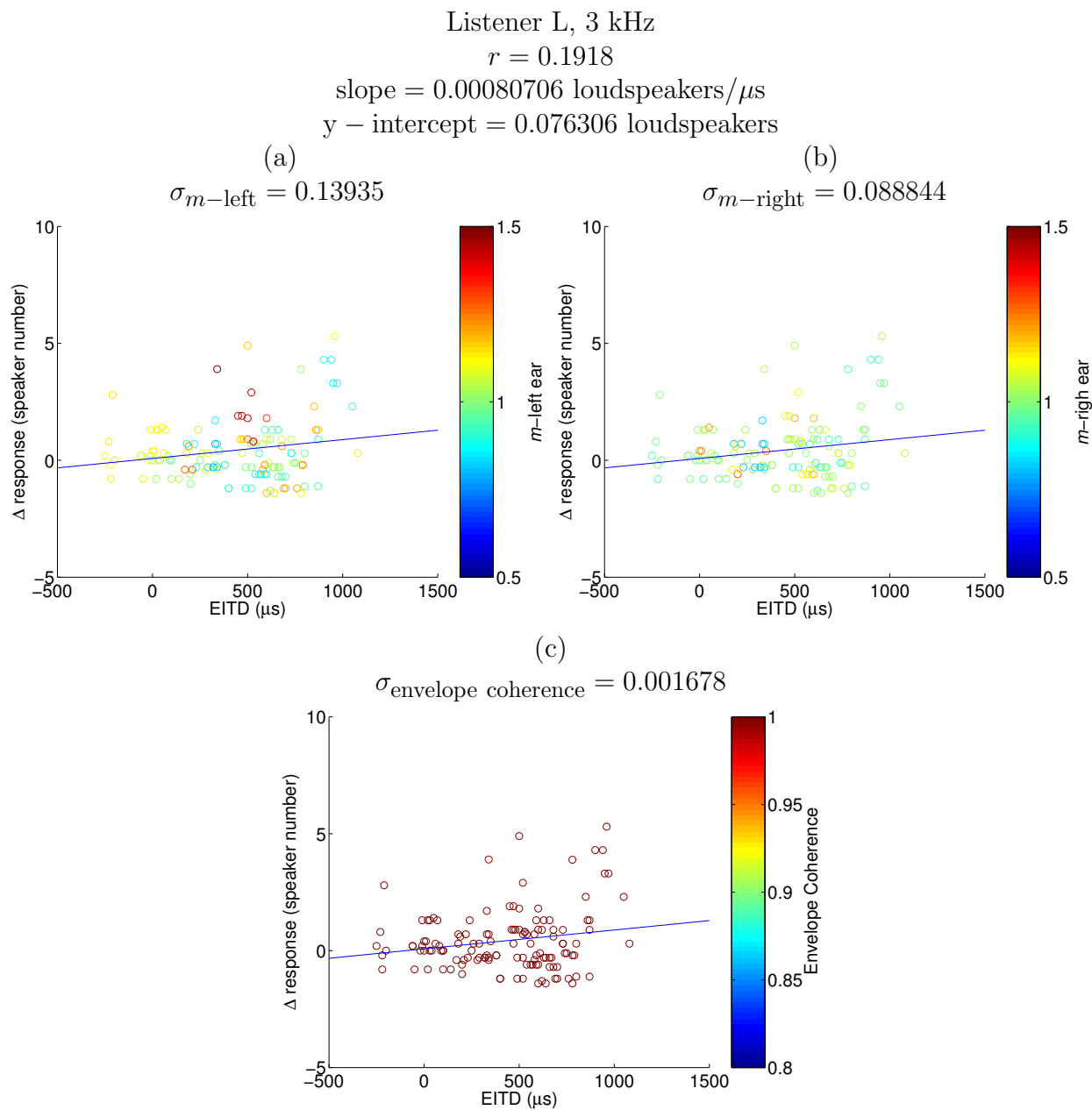


Figure 2.57 Same as Fig. 2.47 but for listener L at 3 kHz.

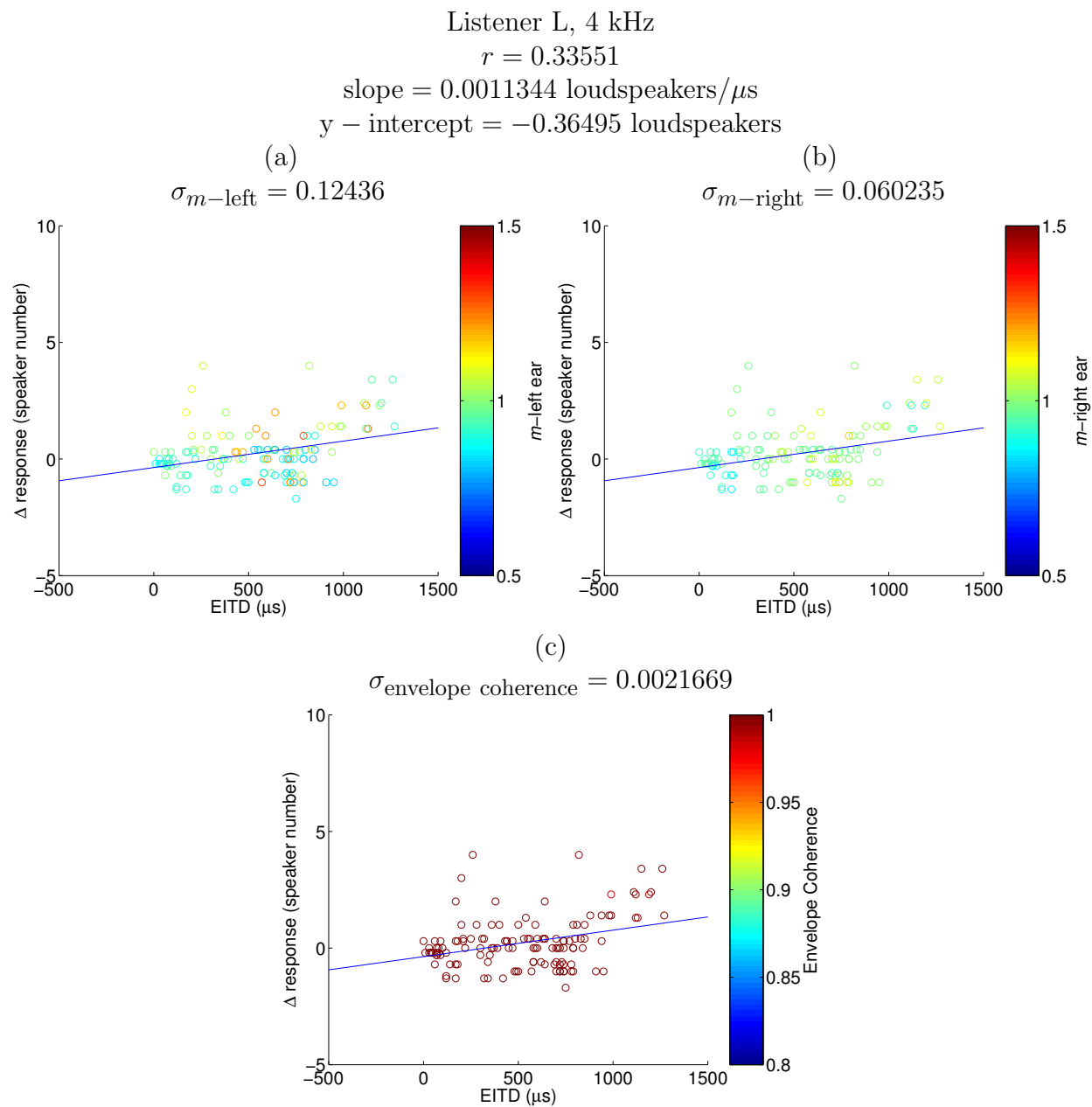


Figure 2.58 Same as Fig. 2.47 but for listener L at 4 kHz.

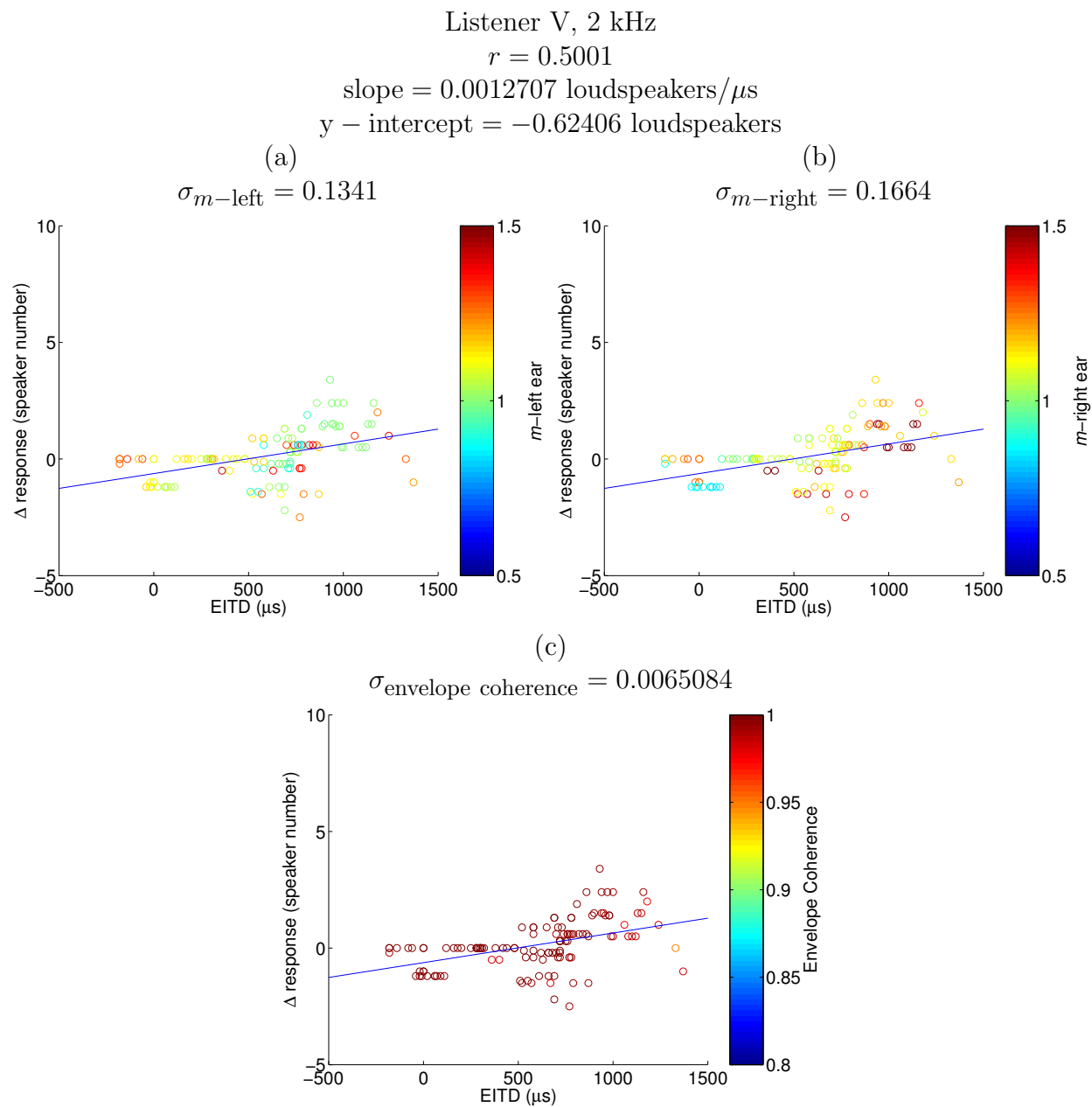


Figure 2.59 Same as Fig. 2.47 but for listener V at 2 kHz.

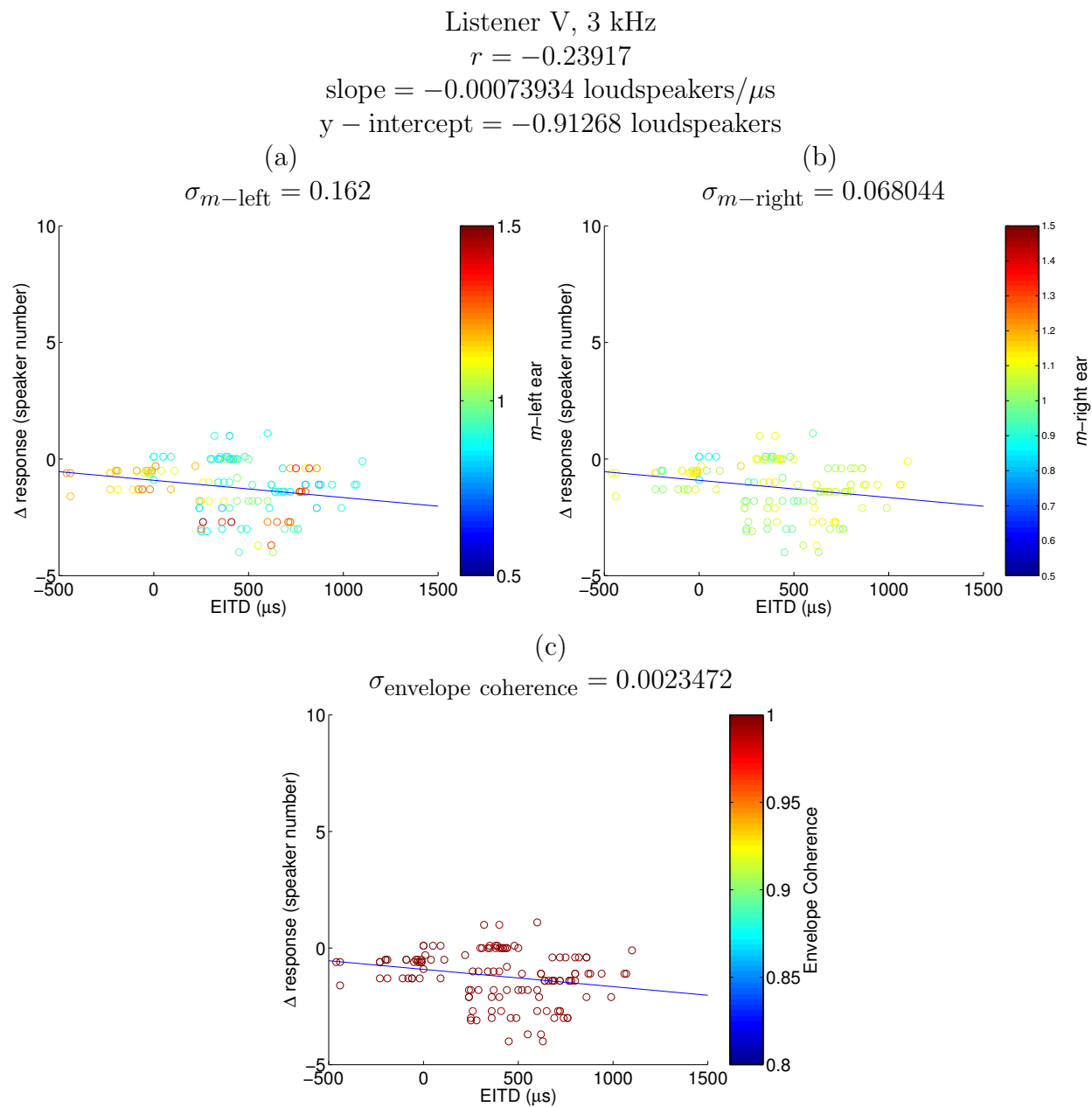


Figure 2.60 Same as Fig. 2.47 but for listener V at 3 kHz.

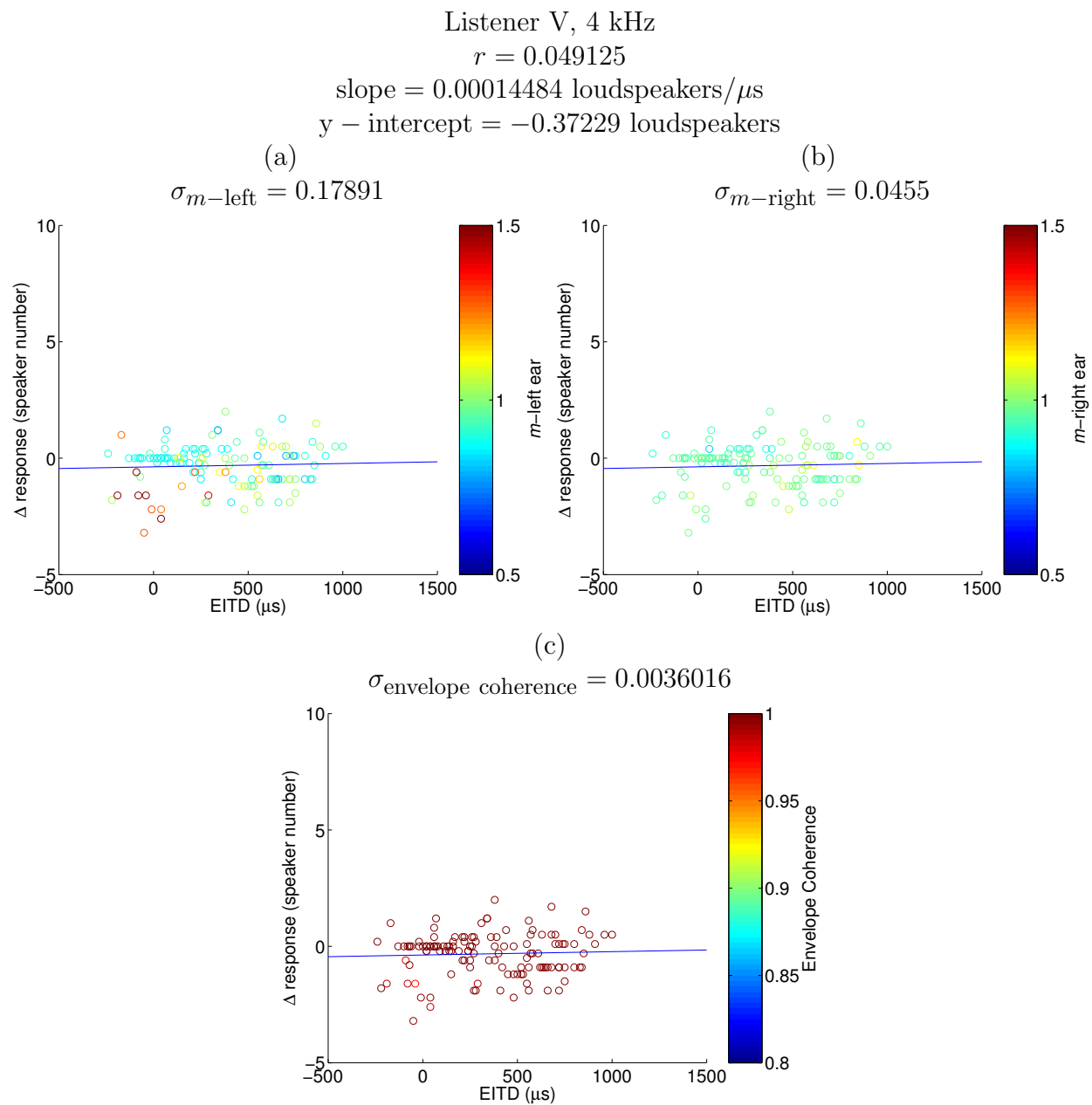


Figure 2.61 Same as Fig. 2.47 but for listener V at 4 kHz.

To investigate further how m in either ear or the envelope coherence may affect the strength of the EITD cue, an analysis of the vertical deviation (residuals) from the best fit lines in Figs. 2.47–2.61 was conducted. An example for listener B at 2 kHz is shown in Fig. 2.62. In this case, again there is little evidence that this range of m -values is affecting the change in response, as the correlation is $r = -0.064$. Table 2.5 is a summary of the correlations for each listener and frequency for m -left, m -right, and the envelope coherence. If anything, the anticipation is that the correlations will be positive, because most dramatically, the EITD cue should be pulling listeners' responses to the right for azimuths past the peak of the ILD curves. However, there are about as many negative correlations as there are positive, none of which are very large. The largest correlation is for listener V at 4 kHz for the m in the left (far) ear. This would indicate that the EITD tended to decrease this listener's responses, and that perhaps when the far ear was under-modulated, the EITD had less weight than when the modulation was higher. Even in this case, however, the statistics are not very convincing.

A possible explanation for this is that the values of m were never really that low. It may be that for these particular AM signals in free-field that m never becomes degraded to the extent that the EITD's weight becomes significantly compromised. One might expect that if the modulation frequency were larger, and therefore the spectrum were wider, that there would be more variation in the values of m , for the far ear in particular. Similarly, in a room, with many reflections, one may expect that the amplitudes and phases as measured in a listener's ear canals would be randomized to a greater extent, and that this would result in smaller values of m for both ears [37].

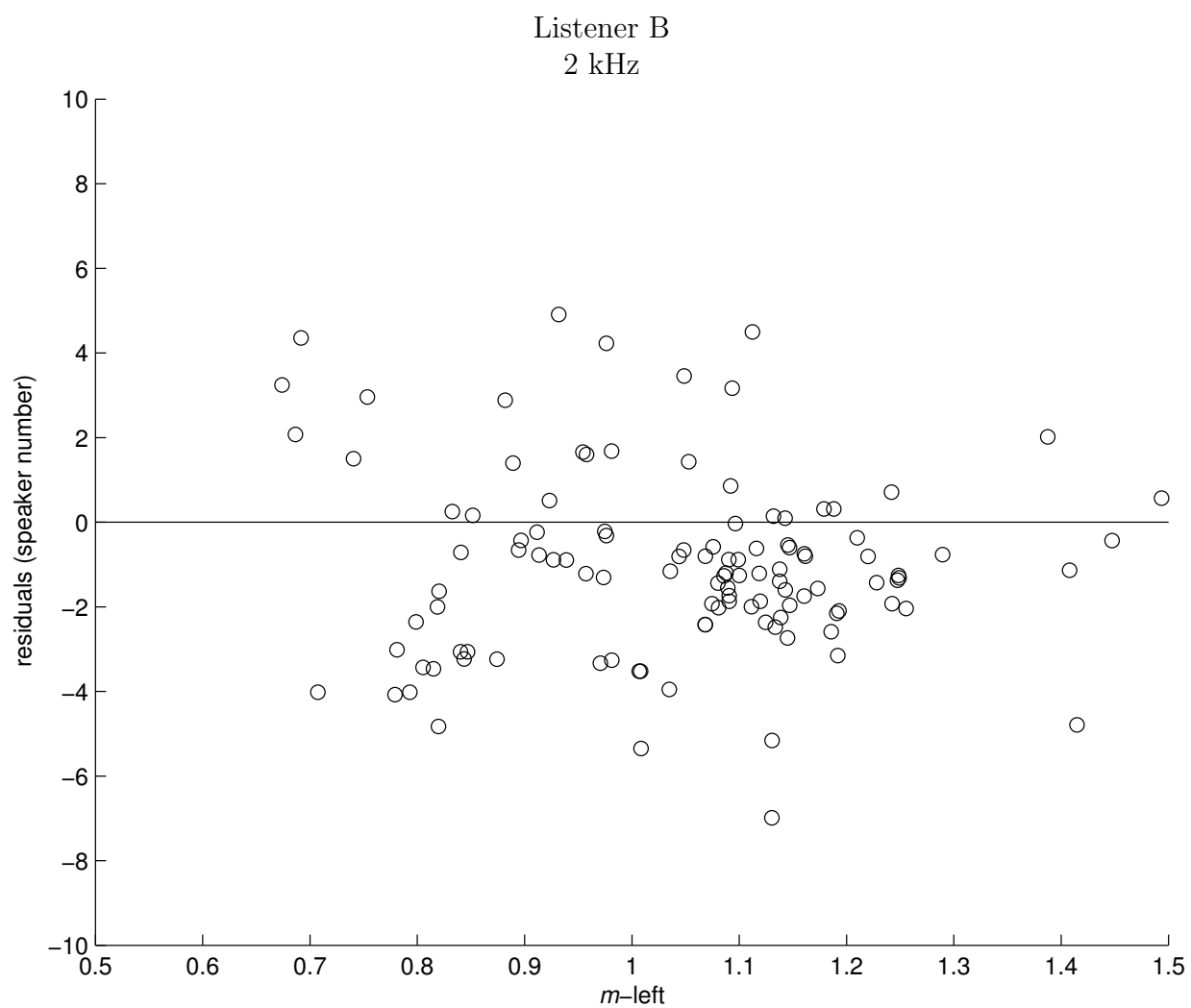


Figure 2.62 Residuals from the best fit line in Fig. 2.47 vs. $m\text{-left}$. The correlation is $r = -0.064$ and $r^2 = 0.0041$.

r for Δ response residuals vs. AM quality			
Listener and frequency (kHz)	m -left	m -right	envelope coherence
Listener B			
2	-0.064318	-0.22606	-0.062472
3	0.040750	0.10755	-0.32536
4	0.24596	0.13275	-0.23071
Listener C			
2	-0.11526	0.099286	-0.18409
3	-0.027794	0.19056	0.24153
4	0.021158	-0.45753	0.20442
Listener M			
2	0.011531	0.013131	0.041615
3	0.028100	0.15645	-0.0064281
4	0.095850	-0.098704	-0.12751
Listener L			
2	-0.35295	0.15010	-0.25949
3	0.22814	-0.011681	-0.17502
4	0.27120	-0.011603	-0.13470
Listener V			
2	-0.12742	-0.054720	0.16777
3	-0.22956	0.010630	0.023282
4	-0.41962	-0.21445	0.26906

Table 2.5 Pearson product-moments, r , for the Δ response residuals from best-fit lines vs. the AM quality metric shown by the color scale in Figs. 2.47–2.61. The correlations are calculated for each listener, frequency, and AM quality metric.

2.3.4.3 Response error ratio vs. AM quality

Despite the lack of evidence that the quality of the AM was affecting the weight of the EITD, it may be possible to determine if the quality of the AM resulted in more accurate responses. To examine this the rms of the response error was examined for the mean responses. This was calculated for each listener and each frequency. The ratio of rms error for the sine tones and the AM tones was then plotted vs. m for each ear and the envelope coherence averaged across speakers. Five listeners and three frequencies means that there are only 15 data points on each plot.

Figure 2.63 shows the rms ratio vs. m -left. Larger rms ratios correspond to improved

listener performance with AM. The correlation is positive, but not strong ($r = 0.1160$). If the data point in the upper left is ignored (listener C at 4 kHz), the correlation would be stronger ($r = 0.4829$), but it is hard to justify doing so. Figure 2.64 shows the same rms error ratios plotted vs. the m in the right ear. Here the correlation is negative, but not strong ($r = -0.1459$). Although it is known that AM did help listeners as seen in Fig. 2.29, this analysis does not show that the depth of modulation was important, at least for these stimuli.

One might think that the similarity between the envelopes in the two ears would impact the accuracy of listener responses. Figure 2.65 shows the same rms error ratios plotted vs. the envelope coherence. Here the correlation is very small ($r = -0.0221$). No trend in response accuracy can be seen as with increasing envelope coherence. The envelope coherences here happen to be pretty large, and may be above a threshold necessary to give strong weight to the EITD cue.

2.3.4.4 Discussion

The slopes of the best fit lines in Figs. 2.47–2.61 show that the EITD appears to become less effective as the carrier frequency increases, but this very likely occurs because the ILD becomes more effective. The quality of the AM cannot be shown to impact listener responses in this experiment. However it cannot be discounted either. The sample size was not large. There may have been a small effect hidden in the noise of the data. Such is the problem with an experiment in free field, as the experimenter does not directly control the nature of the signals in the listener’s ear canals. The ILD, EITD, and quality of AM are dictated by the interaction of the sound wave incident on the listener’s anatomy. A controlled headphone experiment with a wider distribution of cues and AM quality may yield different results.

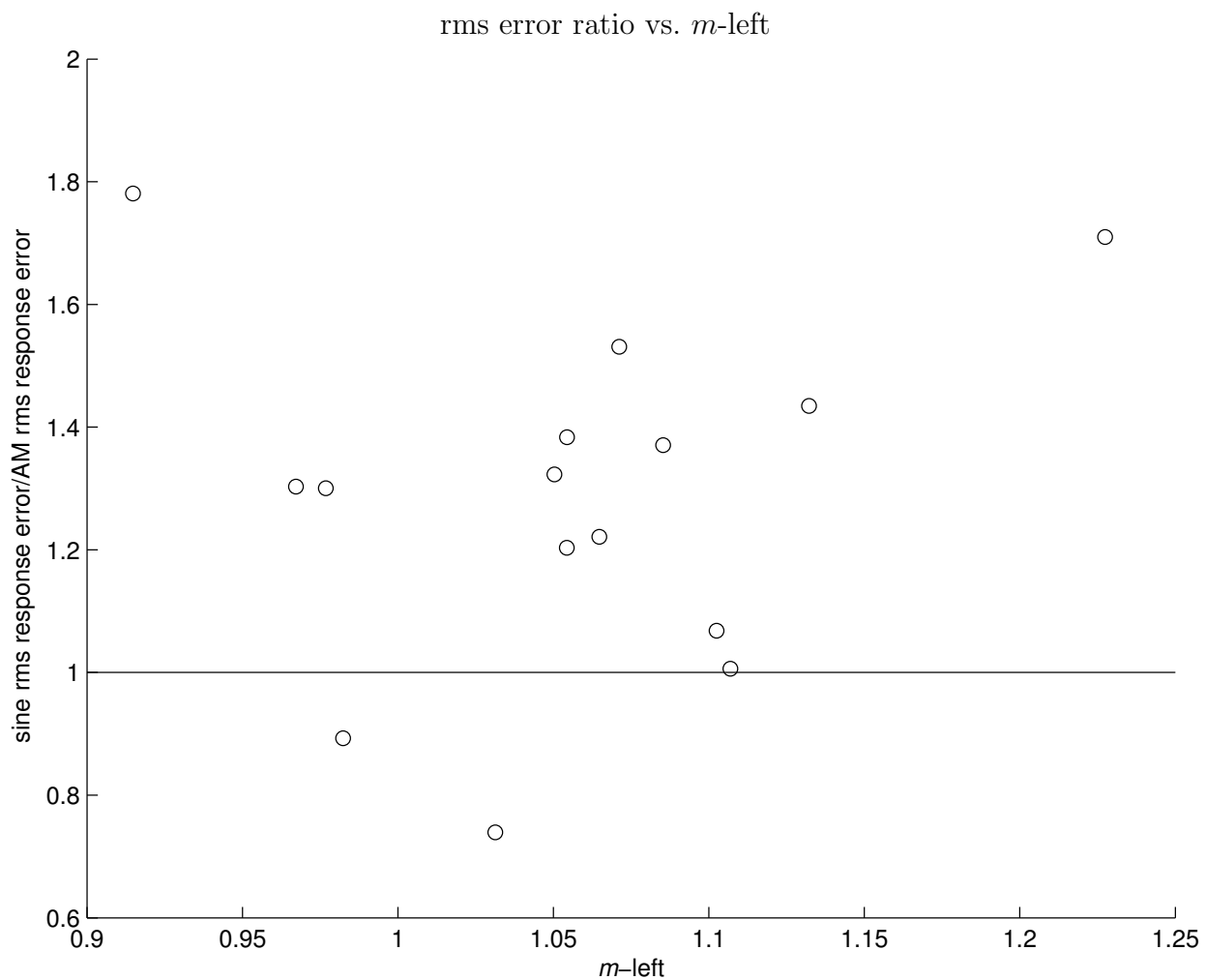


Figure 2.63 The ratio of the error in response rms for sine tones to AM tones vs. m -left averaged across all speakers. The five listeners and three frequencies are all represented. The correlation is, $r = 0.1160$ and $r^2 = 0.01346$.

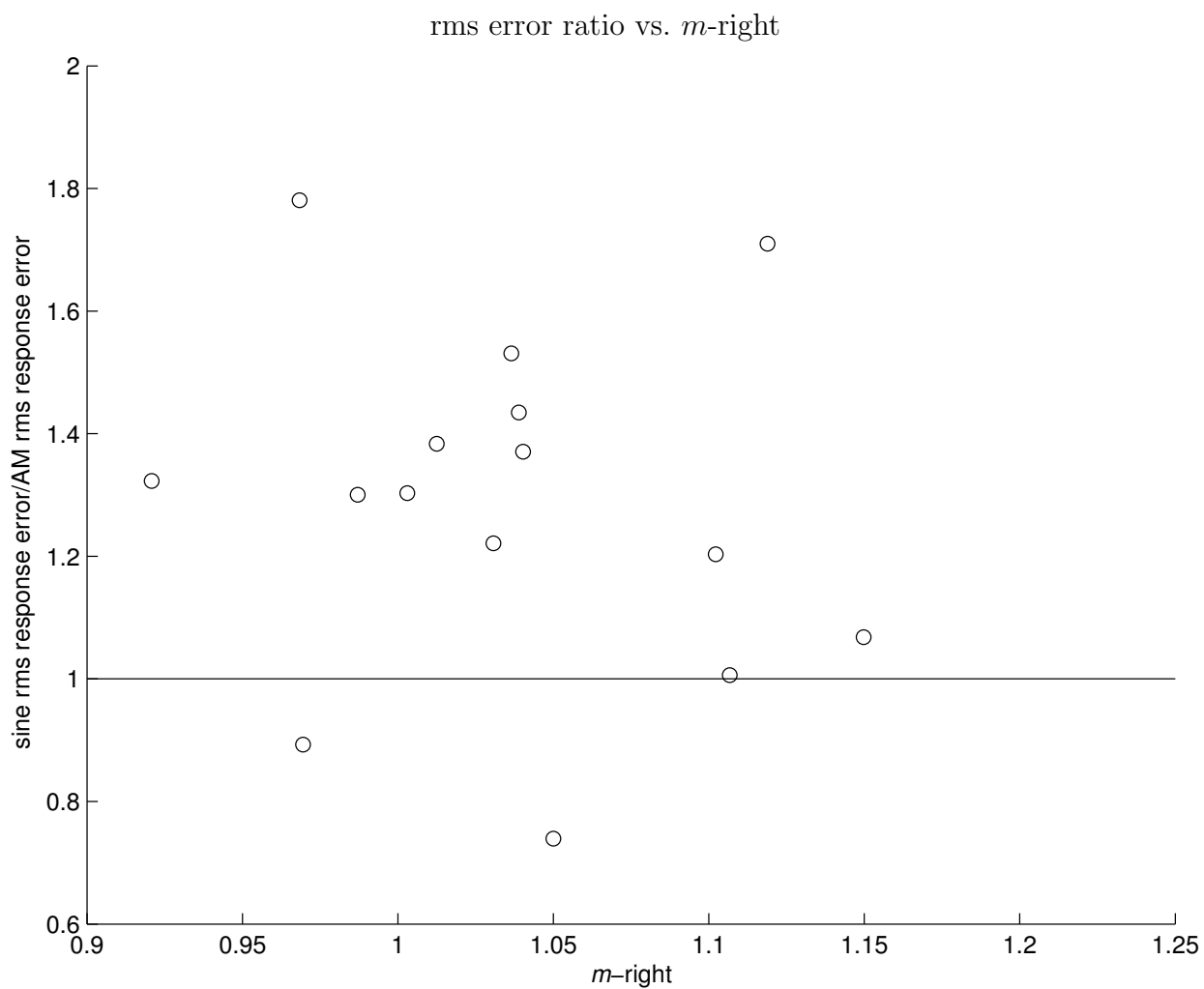


Figure 2.64 The ratio of the error in response rms for sine tones to AM tones vs. m -right averaged across all speakers. The five listeners and three frequencies are all represented. The correlation is, $r = -0.1459$ and $r^2 = 0.02129$.

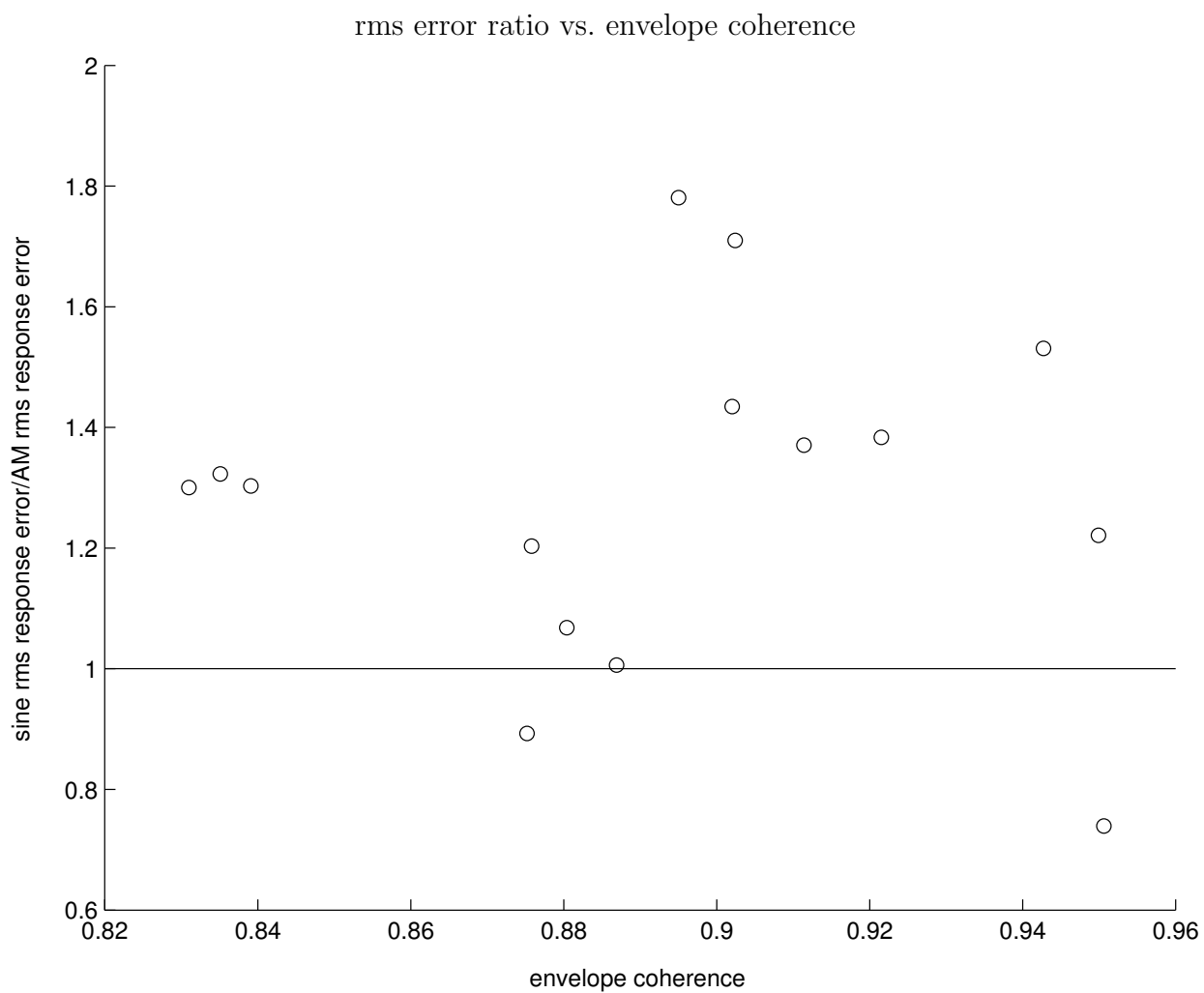


Figure 2.65 The ratio of the error in response rms for sine tones to AM tones vs. the envelope coherence averaged across all speakers. The five listeners and three frequencies are all represented. The correlation is, $r = -0.0221$ and $r^2 = 0.000488$.

2.4 Conclusion

An important cue for localizing sound sources in the horizontal plane is the interaural time difference (ITD). The ITD is the difference in the time of arrival of a sound wave between the two ears. When a sound source is located directly in front of a listener (0° azimuth), the ears are located the same distance to the source, so the ITD is zero. If a sound source is located to the right of a listener at some positive angle, then the listener's right ear is closer to the sound source than the listener's left ear, so the sound wave will arrive at the right ear before the left ear.

For frequencies lower than 1450 Hz, humans are able to use the ITD in the fine structure of a sound wave [7]. At higher frequencies, humans do not have fine-structure ITD sensitivity. However, the interaural time difference in the envelope (EITD) of a sound wave can be used as a localization cue. Psychoacoustical experiments have previously studied the EITD cue for listeners wearing headphones. A typical signal used in headphone experiments that study the EITD is a sinusoidally amplitude modulated (SAM) sine tone.

SAM localization in free field differs from lateralization experiments with headphones. As the incident sound wave interacts with a listener's anatomy, the amplitude and phase spectrum of the original SAM signal is altered. Most notably, the group delay is such that the EITD cue becomes somewhat unreliable to listeners. It often has the opposite sign as the azimuth, and does not increase monotonically.

There were inter-listener differences in performance in localizing SAM tones. A possible reason for this is that there are individual differences in the quality of the cues. Another possibility is that there are individual differences in the processing of the interaural cues. There was not any evidence found linking the AM quality to listener performance. Therefore,

this suggests that there may be individual differences in cue processing.

Nevertheless, the introduction of AM almost always helped with localization accuracy and virtually never hindered it. This is likely due to the interaural level difference (ILD) cue itself being an unreliable cue due to the acoustical bright spot. The EITD cue, although flawed, makes up for the flaws in the ILD cue, especially at large azimuths.

APPENDIX

Appendix

SAM tones in rooms

Methods

In order to understand the effect that rooms have on the interaural properties on SAM tones, a KEMAR manikin was used to make binaural recordings of SAM tones in rooms. Two rooms were used: a reverberation room and a room called the “laboratory” [18, 19, 43]. The reverberation room (Industrial Acoustics Company) was 25 by 21 ft with a 12-ft ceiling (7.7 by 6.4 by 3.6 m). The reverberation room was empty except for a few objects with minimal sound absorption. The laboratory was 28 by 24 ft with a 15-ft ceiling (8.5 by 7.3 by 4.6 m). The room had concrete-block walls, a concrete ceiling, and vinyl tile. To dampen the usually lively room, there was a canvas tarpaulin on the floor with dimensions of 16 by 10 ft. Additionally, scattered on the floor were about two dozen sound absorbing acoustical cones.

In the reverberation room, the KEMAR was positioned in three separate locations. At each location the KEMAR was rotated to four different angles: 0, 30, 60, and 90 degrees. At positive angles the KEMAR was rotated such that the loudspeaker was on the right side. At each angle the loudspeaker was placed at 3 distances, d : 50, 100, and 200 cm. Location 1 was 9 ft 2 in into the room along a line 10 ft 9 in in from the right-hand wall. Location 2 was 11 ft 8 in into the room along a line 7 ft 0.5 in in from the left-hand wall. Location 3 was 50 cm into the room along a line 7 ft 0.5 in in from the left-hand wall. For each location, the

KEMAR was facing away from the door when rotated to zero degrees and the loudspeaker was facing the door.

In the “laboratory” two locations near the middle of the room were used. In location 1 the KEMAR was facing away from the door for the zero degree rotation, and in location 2 the KEMAR was facing away from the door. The same four angles and three distances were used.

There were two SAM signals produced by a Tucker Davis System II with DD1 digital-analog and analog-digital converter. The sample rate was 50 kHz with 16 bits/sample. Both signals had a 4-kHz carrier and were 5 seconds long with hard onsets and offsets. The amplitude modulation had rates of either 100 Hz or 500 Hz and had a depth of 100%. A Radio Shack Minimus 3.5, consisting of a 2.5-in driver in a sealed box was used to generate the tones at a level of 64 dBA. Binaural recordings were made at the same time as the 5-second presentation of the tones. The signals from the KEMAR pre-amplifiers were further amplified with an Audio Buddy pre-amplifier. The gains of the left and right pre-amplifier were calibrated to be within 1 dB of each other.

Summary of Data

From each recording the following were calculated: the ILD, the EITD, the interaural envelope coherence, the AM in each ear, m_{left} and m_{right} , and the QFM in each ear, β_{left} and β_{right} . For the reverberation room, these quantities are presented in Table .1 for 100 Hz and Table .2 for 500 Hz. For the “laboratory” the quantities are presented in Table .3 for 100 Hz and Table .4 for 500 Hz.

Compared to the values of m in Fig. 2.8 for 100-Hz AM in free field, the values of m

in Table .1 for the reverberation room show more variation. This is also the case but to a lesser degree for m in the “laboratory” (Table .3). In free field, m was never found to be much smaller than about 0.75. In the reverberation room and the “laboratory” there are instances with much smaller values which may have important perceptual consequences on the strength of the EITD cue [40]. The 500-Hz modulation frequency data in Tables .2 and .4 also contain considerable variation in m with more variation in the reverberation room.

The QFM (β) also showed more variation in rooms than in free field (Fig. 2.9). In free field, β rarely exceeded a value of 1. In rooms, there are several instances of this, including one very large value at 500 Hz in the “laboratory” of 20.6270. Because β is very sensitive to changes in phase, it is not surprising that room reflections would lead to this type of distribution.

The EITD cue is inconsistent with azimuthal angle in rooms. Changes in the distance to the source, and changes in the location in the room create an EITD that appears to be mostly random. The EITD is often negative, even though the azimuthal angles are located between 0 and 90 degrees.

The interaural envelope coherence remained large in rooms. A computer simulation of $n = 1000$ random pairs of envelopes with Rayleigh-distributed amplitudes, and uniformly-distributed phases showed that the interaural envelope coherence does not deviate from unity by much. The mean value was $\mu = 0.9979$, the standard deviation was $\sigma = 0.0012$, and the standard error was $\frac{\sigma}{\sqrt{1000}} = 0.000039165$.

The ILD appeared to be a more reliable cue in rooms than the EITD. There were fewer misleading (negative) cues. The ILD appeared to increase with azimuthal angle more reliably than the EITD.

	location 1			location 2			location 3		
d (cm)	50	100	200	50	100	200	50	100	200
0°									
m_{left}	1.5445	0.8756	0.4745	1.4099	0.4198	1.7936	1.3937	0.5885	0.1531
m_{right}	0.7930	0.6594	0.6875	0.7951	0.4993	0.3572	0.8753	0.3950	0.7184
β_{left}	0.4768	0.4580	0.8928	0.5495	0.5423	0.3604	0.0717	0.5822	0.2245
β_{right}	0.1213	0.3015	0.2776	0.1236	0.1907	0.9751	0.2598	0.4194	0.8649
EITD	260	-520	1560	-340	1080	-20	-180	-440	2400
E. Coh.	0.9850	0.9971	0.9745	0.9943	0.9932	0.9217	0.9864	0.9938	0.9565
ILD	4.7	0.3	-0.6	3.0	4.3	5.9	2.4	0.6	-1.4
30°									
m_{left}	0.9720	0.3107	0.2546	0.7301	2.1296	1.8191	0.6172	0.3607	0.4048
m_{right}	1.1057	0.7871	0.6929	0.9049	1.5508	0.3378	0.9741	0.7965	0.9068
β_{left}	0.7020	1.2994	0.1239	0.1976	2.5768	1.7730	0.1594	0.0959	0.6591
β_{right}	0.0398	0.3262	0.1428	0.0623	0.4187	0.3310	0.1626	0.2097	0.1582
EITD	0	-4000	3320	200	2960	2160	140	-140	-5000
E. Coh.	0.9863	0.9182	0.9549	0.9936	0.9296	0.9594	0.9846	0.9681	0.9307
ILD	15.3	11.6	4.4	10.9	9.6	11.9	9.4	7.9	0.8
60°									
m_{left}	0.6119	0.2347	1.2575	0.3284	0.8312	0.3034	0.5120	3.8221	1.1431
m_{right}	1.1285	1.0267	0.8148	0.9464	1.8968	0.5433	0.8261	1.0279	0.5721
β_{left}	0.1917	0.4958	0.9659	0.1997	0.4159	0.3903	0.0463	3.3003	0.4881
β_{right}	0.1511	0.3505	0.5763	0.0322	1.0464	0.5704	0.2553	0.2946	0.4132
EITD	140	460	-2280	580	-20	200	880	-340	1260
E. Coh.	0.9730	0.9208	0.9942	0.9326	0.9586	0.9895	0.9852	0.8871	0.9797
ILD	10.2	9.7	2.6	10.8	2.9	-0.2	9.5	12.4	7.1
90°									
m_{left}	0.4338	0.2410	2.1282	0.7818	2.0527	0.3172	0.6372	0.3197	0.6171
m_{right}	1.2117	0.4886	0.4477	1.2439	2.2506	0.1616	0.9738	0.9701	0.7463
β_{left}	0.1921	0.3995	1.1502	0.3582	0.4298	0.2957	2.5129	0.5764	0.3911
β_{right}	0.1828	0.3078	0.9657	0.0525	1.0517	0.4135	0.1975	0.5098	0.5591
EITD	1560	-520	-580	720	1960	-3980	3700	3840	-2100
E. Coh.	0.9402	0.9850	0.9732	0.9800	0.9916	0.9842	0.8470	0.9527	0.9981
ILD	11.0	6.6	9.6	7.0	1.5	-1.9	16.9	-2.1	0.7

Table .1 Measurements for SAM tone in the reverberation room at a modulation frequency of 100 Hz. For each angle, location and distance (d), quantities shown are the AM in the left ear (m_{left}), the AM in the right ear (m_{right}), the QFM in the left ear (β_{left}), the QFM in the right ear (β_{right}), the envelope interaural time difference (EITD) in μs , the interaural envelope coherence (E. Coh.), and the interaural level difference (ILD) in dB.

	location 1			location 2			location 3		
d (cm)	50	100	200	50	100	200	50	100	200
0°									
m_{left}	0.8379	0.6355	0.5402	1.5785	0.60422	1.8565	1.0388	0.5980	0.3559
m_{right}	0.8589	0.7111	1.0643	0.5676	0.8413	0.5275	0.5028	0.7415	1.8287
β_{left}	0.1425	0.5565	0.2925	0.4589	1.2760	1.0355	0.2725	0.8793	1.2703
β_{right}	0.1019	0.5293	0.3770	0.2711	0.3469	1.2142	0.4989	0.6458	0.9776
EITD	-20	40	-140	-20	-320	-800	120	-140	-900
E. Coh.	0.9997	0.9931	0.9719	0.9752	0.9587	0.9649	0.9548	0.9928	0.9333
ILD	7.2	1.0	1.4	2.5	2.8	5.8	2.8	1.2	-1.2
30°									
m_{left}	1.4466	0.9902	0.4199	0.5205	1.0475	0.9736	0.2044	0.2446	0.5554
m_{right}	1.0008	0.8964	0.4651	0.9489	2.1246	0.1898	0.6593	0.5381	0.6729
β_{left}	1.5811	0.0486	0.2518	0.3208	3.3604	1.1770	0.1835	0.2292	0.5745
β_{right}	0.0415	0.3027	0.3982	0.2645	0.6631	0.5310	0.1594	0.5813	0.6689
EITD	260	300	-100	180	-620	-320	100	20	260
E. Coh.	0.9653	0.9949	0.9994	0.9775	0.9098	0.9575	0.9579	0.9897	0.9941
ILD	12.4	13.8	3.9	11.2	10.9	14.4	9.2	7.7	0.2
60°									
m_{left}	0.5608	0.9908	0.6755	0.4783	0.5317	0.6571	0.6676	2.0886	1.6293
m_{right}	1.1370	0.8289	0.3514	0.8255	1.1904	0.7945	0.6645	0.8727	0.9892
β_{left}	0.6541	0.8601	0.3369	0.3402	0.8234	0.3010	0.1342	1.2437	0.5079
β_{right}	0.1720	0.5388	0.7662	0.4163	2.0632	0.7148	0.2146	0.6813	0.5951
EITD	540	720	740	660	80	420	520	-380	880
E. Coh.	0.9465	0.9951	0.9628	0.9852	0.9591	0.9923	0.9987	0.9757	0.9843
ILD	9.7	7.0	3.6	10.2	3.0	0.4	8.7	14.1	7.0
90°									
m_{left}	0.6543	1.1516	1.2218	0.7902	0.9061	1.0965	3.2058	0.3902	1.2611
m_{right}	1.4227	0.5725	0.1086	0.9213	2.9357	0.6965	0.8293	1.5403	0.9417
β_{left}	0.5657	0.7351	3.8372	0.5685	1.5851	1.2577	1.7952	0.5668	0.6452
β_{right}	0.3893	0.7280	0.1497	0.5919	1.7889	0.3479	0.5439	0.9004	0.3338
EITD	-500	400	-740	280	40	40	-800	560	-20
E. Coh.	0.9734	0.9687	0.9456	0.9907	0.9822	0.9913	0.9772	0.9569	0.9986
ILD	10.7	5.1	3.8	6.2	3.4	-4.5	14.7	0.4	-1.2

Table .2 Same as Table .1 but for the reverberation room at 500 Hz.

	location 1			location 2		
d (cm)	50	100	200	50	100	200
0°						
m_{left}	0.8100	0.7885	0.8104	1.0872	0.6181	0.8844
m_{right}	0.9330	0.9409	1.4727	0.9144	0.8108	0.7636
β_{left}	0.0156	0.1622	0.5390	0.2282	0.1078	0.6571
β_{right}	0.1079	0.3242	0.7879	0.2436	0.2999	0.6011
EITD	180	240	120	-260	340	-400
E. Coh.	0.9979	0.9987	0.9809	0.9966	0.9966	0.9983
ILD	0.6	3.2	0.8	2.0	-1.6	-0.5
30°						
m_{left}	1.4911	0.7394	1.666	0.8435	1.5139	0.1340
m_{right}	0.9759	1.0049	0.9603	0.8370	1.1159	0.5985
β_{left}	0.2945	0.8047	0.9547	0.0475	1.2610	0.4465
β_{right}	0.1039	0.0846	0.2474	0.0759	0.0886	0.0963
EITD	-120	2560	-1600	760	1820	0
E. Coh.	0.9912	0.9700	0.9895	0.9981	0.9606	0.9514
ILD	13.9	12.6	11.3	11.8	12.2	9.1
60°						
m_{left}	0.6251	0.8951	0.5914	1.1320	1.2511	0.3153
m_{right}	0.9190	1.0828	1.0762	0.9178	1.1978	0.8949
β_{left}	0.3132	0.7596	0.3808	0.5171	0.4671	1.3520
β_{right}	0.0836	0.3412	0.0697	0.1208	0.2779	0.4185
EITD	600	-500	220	780	-80	-4020
E. Coh.	0.9809	0.9836	0.9656	0.9900	0.9980	0.9180
ILD	16.3	9.2	5.3	13.8	11.4	17.0
90°						
m_{left}	0.4067	1.1268	0.9762	0.2972	0.1390	0.8198
m_{right}	0.9998	1.2441	4.0648	0.9250	0.9999	3.3393
β_{left}	0.4139	0.8395	1.6093	1.6407	0.1923	0.3819
β_{right}	0.1164	0.3821	7.7200	0.2586	0.2372	1.7341
EITD	2920	1640	2820	-1060	1340	2820
E. Coh.	0.9364	0.9872	0.9360	0.8775	0.8851	0.9177
ILD	19.2	9.3	0.6	14.9	5.1	-2.4

Table .3 Same is Table .1 but for the the “laboratory” at 100 Hz.

	location 1			location 2		
d (cm)	50	100	200	50	100	200
0°						
m_{left}	0.9875	1.4277	1.3406	1.1955	0.7131	0.4907
m_{right}	0.9491	1.0009	1.2350	0.9099	0.9230	0.4349
β_{left}	0.2922	0.1002	0.7831	0.2526	0.1540	1.3852
β_{right}	0.2477	0.4749	0.7430	0.3113	0.4237	1.2396
EITD	0	40	80	-40	60	-60
E. Coh.	0.9996	0.9906	0.9891	0.9946	0.9973	0.9847
ILD	0.3	1.4	-1.9	1.8	-1.4	-0.3
30°						
m_{left}	1.2242	0.3694	1.8766	0.7852	1.0914	0.5982
m_{right}	1.0469	0.9813	0.6639	0.9128	1.0379	0.4634
β_{left}	0.7052	1.4314	1.7679	0.2565	1.1859	0.8868
β_{right}	0.1375	0.1325	0.4823	0.0973	0.4175	0.3340
EITD	160	-540	560	140	120	840
E. Coh.	0.9959	0.8899	0.9902	0.9968	0.9752	0.9907
ILD	14.4	11.9	9.5	12.2	13.8	7.9
60°						
m_{left}	1.0428	1.2232	0.4876	0.7511	0.6937	2.9010
m_{right}	1.0330	0.9984	1.4846	1.0274	1.1181	0.4452
β_{left}	0.3208	0.6330	0.5706	0.7551	1.4402	1.3217
β_{right}	0.2377	0.5541	0.6594	0.3272	1.0080	0.2584
EITD	560	580	660	580	480	-480
E. Coh.	0.9981	0.9988	0.9532	0.9778	0.9621	0.9492
ILD	15.3	9.0	5.3	15.1	11.2	11.8
90°						
m_{left}	1.2183	1.364	1.4309	1.0319	0.8004	1.0140
m_{right}	0.9477	1.0277	6.2806	0.9665	1.3526	5.0144
β_{left}	0.3307	1.6183	1.1942	0.4966	0.0670	1.2269
β_{right}	0.3815	1.0628	20.6270	0.5411	0.9354	2.7867
EITD	740	800	460	940	-40	-840
E. Coh.	0.9925	0.9781	0.9025	0.9931	0.9906	0.9464
ILD	17.6	7.1	7.0	16.9	6.0	-0.3

Table .4 Same is Table .1 but for the the “laboratory” at 500 Hz.

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] G. Andéol, E. A. Macpherson, and A. T. Sabin. Sound localization in noise and sensitivity to spectral shape. *Hear. Res.*, 304:20 – 27, 2013.
- [2] L. R. Bernstein and C. Trahiotis. Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities. *J. Acoust. Soc. Am.*, 77(5), 1985.
- [3] L. R. Bernstein and C. Trahiotis. Lateralization of sinusoidally amplitude-modulated tones: Effects of spectral locus and temporal variation. *J. Acoust. Soc. Am.*, 78(2):514–523, 1985.
- [4] L. R. Bernstein and C. Trahiotis. Enhancing sensitivity to interaural delays at high frequencies by using “transposed stimuli”. *J. Acoust. Soc. Am.*, 112(3):1026–1036, 2002.
- [5] L. R. Bernstein and C. Trahiotis. How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli. *J. Acoust. Soc. Am.*, 125(5):3234–3242, 2009.
- [6] J. Blauert. Sound localization in the median plane. *Acustica*, 22:205 – 213, 1969-70.
- [7] A. Brughera, L. Dunai, and W. M. Hartmann. Human interaural time difference thresholds for sine tones: The high-frequency limit. *J. Acoust. Soc. Am.*, 133(5), 2013.
- [8] M. D. Burkhard and R. M. Sachs. Anthropometric manikin for acoustic research. *J. Acoust. Soc. Am.*, 58(1), 1975.
- [9] R. A. Butler, S. K. Roffler, and R. F. Naunton. The role of stimulus frequency in the localization of sound in space. *J. Aud. Res.*, 7:169 – 180, 1967.
- [10] M. Dietz, L. R. Bernstein, C. Trahiotis, S. D. Ewert, and V. Hohmann. The effect of overall level on sensitivity to interaural differences of time and level at high frequencies. *J. Acoust. Soc. Am.*, 134(1):494–502, 2013.
- [11] G. Eberle, K. I. McAnally, R. L. Martin, and P. Flanagan. Localization of amplitude-modulated high-frequency noise. *J. Acoust. Soc. Am.*, 107(6), 2000.

- [12] R. Fisher. *Statistical Methods For Research Workers*. Cosmo study guides. Cosmo Publications, 1925.
- [13] B. R. Glasberg and B. C. Moore. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.*, 47(1):103–138, 1990.
- [14] M. J. Goupell and W. M. Hartmann. Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models. *J. Acoust. Soc. Am.*, 122(2), 2007.
- [15] R. V. L. Hartley and T. C. Fry. The binaural location of pure tones. *Phys. Rev.*, 18:431–442, Dec 1921.
- [16] W. M. Hartmann. *Signals, Sound, and Sensation*. Modern Acoustics and Signal Processing. Springer-Verlag, New York (American Institute of Physics Press), 1998.
- [17] W. M. Hartmann and E. J. Macaulay. Anatomical limits on interaural time differences: An ecological perspective. *Frontiers in Neuroscience*, 8(34), 2014.
- [18] W. M. Hartmann and B. Rakerd. Localization of sound in rooms IV: The Franssen effect. *J. Acoust. Soc. Am.*, 86(4), 1989.
- [19] W. M. Hartmann, B. Rakerd, and A. Koller. Binaural coherence in rooms. *Acta acustica united with acustica*, 91(3):451–462, 2005.
- [20] W. M. Hartmann and A. Wittenberg. On the externalization of sound images. *J. Acoust. Soc. Am.*, 99(6), 1996.
- [21] W. M. Hartmann and E. M. Wolf. Matching the waveform and the temporal window in the creation of experimental signals. *J. Acoust. Soc. Am.*, 126(5), 2009.
- [22] E. Hecht. Optics. Pearson Education, page 496. Addison-Wesley, 4th edition, 2002.
- [23] G. B. Henning. Detectability of interaural delay in high-frequency complex waveforms. *J. Acoust. Soc. Am.*, 55(1), 1974.
- [24] G. B. Henning. Some observations on the lateralization of complex waveforms. *J. Acoust. Soc. Am.*, 68(2), 1980.
- [25] G. B. Henning. Lateralization of low-frequency transients. *Hear. Res.*, 9(2):153 – 172, 1983.

- [26] G. B. Henning and J. Ashton. The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay. *Hear. Res.*, 4(2):185 – 194, 1981.
- [27] G. F. Kuhn. Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.*, 62(1), 1977.
- [28] G. F. Kuhn. The pressure transformation from a diffuse sound field to the external ear and to the body and head surface. *J. Acoust. Soc. Am.*, 65(4), 1979.
- [29] G. F. Kuhn. Physical acoustics and measurements pertaining to directional hearing. In W. Yost and G. Gourevitch, editors, *Directional Hearing*, Proceedings in life sciences, pages 3–25. Springer-Verlag, 1987.
- [30] D. M. Leakey, B. M. Sayers, and C. Cherry. Binaural fusion of low- and high-frequency sounds. *J. Acoust. Soc. Am.*, 30(3):222–222, 1958.
- [31] E. J. Macaulay, W. M. Hartmann, and B. Rakerd. The acoustical bright spot and mislocalization of tones by human listeners. *J. Acoust. Soc. Am.*, 127(3), 2010.
- [32] E. A. Macpherson and J. C. Middlebrooks. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *J. Acoust. Soc. Am.*, 111(5), 2002.
- [33] D. McFadden and E. G. Pasanen. Lateralization at high frequencies based on interaural time differences. *J. Acoust. Soc. Am.*, 59(3), 1976.
- [34] A. W. Mills. On the minimum audible angle. *J. Acoust. Soc. Am.*, 30(4), 1958.
- [35] A. W. Mills. Auditory localization. In J. Tobias, editor, *Foundations of Modern Auditory Theory*, volume 2, pages 301–345. Academic Press, 1970.
- [36] S. M. Misurelli and R. Y. Litovsky. Spatial release from masking in children with normal hearing and with bilateral cochlear implants: Effect of interferer asymmetry. *J. Acoust. Soc. Am.*, 132(1), 2012.
- [37] J. J. M. Monaghan, K. Krumbholz, and B. U. Seeber. Factors affecting the use of envelope interaural time differences in reverberation. *J. Acoust. Soc. Am.*, 133(4), 2013.
- [38] P. M. C. Morse. *Vibration and sound*. International series in physics. McGraw-Hill book company, inc., 1936.

- [39] J. O. Nordmark. Binaural time discrimination. *J. Acoust. Soc. Am.*, 60(4), 1976.
- [40] J. M. Nuetzel and E. R. Hafter. Discrimination of interaural delays in complex waveforms: Spectral effects. *J. Acoust. Soc. Am.*, 69(4), 1981.
- [41] M. Otani, T. Hirahara, and S. Ise. Numerical study on source-distance dependency of head-related transfer functions. *J. Acoust. Soc. Am.*, 125(5), 2009.
- [42] B. Rakerd and W. M. Hartmann. Localization of sound in rooms, III: Onset and duration effects. *J. Acoust. Soc. Am.*, 80(6):1695–1706, 1986.
- [43] B. Rakerd and W. M. Hartmann. Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise. *J. Acoust. Soc. Am.*, 128(5), 2010.
- [44] S. N. Rschewkin. *A Course of Lectures on the Theory of Sound*. Pergamonn Press, 1963.
- [45] B. G. Shinn-Cunningham, S. Santarelli, and N. Kopco. Tori of confusion: Binaural localization cues for sources within reach of a listener. *J. Acoust. Soc. Am.*, 107(3):1627–1636, 2000.
- [46] M. A. Stellmack, N. F. Viemeister, and A. J. Byrne. Discrimination of interaural phase differences in the envelopes of sinusoidally amplitude-modulated 4-khz tones as a function of modulation depth. *J. Acoust. Soc. Am.*, 118(1), 2005.
- [47] J. W. Strutt. *The Theory of Sound V. 2*. Macmillan, 1896.
- [48] J. W. Strutt. XII. On our perception of sound direction. *Philosophical Magazine Series 6*, 13(74):214–232, 1907.
- [49] F. L. Wightman and D. J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, 91(3), 1992.
- [50] W. A. Yost. Lateral position of sinusoids presented with interaural intensive and temporal differences. *J. Acoust. Soc. Am.*, 70(2), 1981.
- [51] P. X. Zhang and W. M. Hartmann. On the ability of human listeners to distinguish between front and back. *Hear. Res.*, 260(12):30 – 46, 2010.
- [52] J. Zwislocki and R. S. Feldman. Just noticeable differences in dichotic phase. *J. Acoust. Soc. Am.*, 28(5), 1956.