THESIS
3
2006

This is to certify that the
dissertation entitled

THE USE OF INTERAURAL PARAMETERS DURING INCOHERENCE DETECTION IN REPRODUCIBLE NOISE

presented by

MATTHEW JOSEPH GOUPELL

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Physics and Astronomy

Major Professor's Signature

1 December 2005

Date

MSU is an Affirmative Action/Equal Opportunity Institution

LIBRARY
Michigan State
University

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

DATE DUE	DATE DUE	DATE DUE

2/05 p:/CIRC/DateDue.indd-p.1

THE USE OF INTERAURAL PARAMETERS DURING INCOHERENCE DETECTION IN REPRODUCIBLE NOISE

By

Matthew Joseph Goupell

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Physics and Astronomy

2005

ABSTRACT

THE USE OF INTERAURAL PARAMETERS DURING INCOHERENCE DETECTION IN REPRODUCIBLE NOISE

By

Matthew Joseph Goupell

Interaural incoherence is a measure of the dissimilarity of the signals in the left and right ears. It is important in a number of acoustical phenomenon such as a listener's sensation envelopment and apparent source width in room acoustics, speech intelligibility, and binaural release from energetic masking. Humans are incredibly sensitive to the difference between perfectly coherent and slightly incoherent signals, however the nature of this sensitivity is not well understood. The purpose of this dissertation is to understand what parameters are important to incoherence detection. Incoherence is perceived to have time-varying characteristics. It is conjectured that incoherence detection is performed by a process that takes this time dependency into account.

Left-ear-right-ear noise-pairs were generated, all with a fixed value of interaural coherence, 0.9922. The noises had a center frequency of 500 Hz, a bandwidth of 14 Hz, and a duration of 500 ms. Listeners were required to discriminate between these slightly incoherent noises and diotic noises, with a coherence of 1.0. It was found that the value of interaural incoherence itself was an inadequate predictor of discrimination. Instead, incoherence was much more readily detected for those noise-pairs with the largest fluctuations in interaural phase and level differences (as measured by the standard deviation). Noise-pairs with the same value of coherence, and geometric mean frequency of 500 Hz were also generated for bandwidths of 108 Hz and 2394 Hz. It was found that for increasing bandwidth, fluctuations in interaural differences varied less between different noise-pairs and that detection performance varied less as well. The results suggest that incoherence detection is based on the size

and the speed of interaural fluctuations and that the value of coherence itself predicts performance only in the wide-band limit where different particular noises with the same incoherence have similar fluctuations.

Noise-pairs with short durations of 100, 50, and 25 ms, and bandwidth of 14 Hz, and a coherence of 0.9922 were used to test if a short-term incoherence function is used in incoherence detection. It was found that listeners could significantly use fluctuations of phase and level to detect incoherence for all three of these short durations. Therefore, a short-term coherence function is not used to detect incoherence. For the smallest duration of 25 ms, listeners' detection cue sometimes changed from a "width" cue to a lateralization cue.

Modeling of the data was performed. Ten different binaural models were tested against detection data for 14-Hz and 108-Hz bandwidths. These models included different types of binaural processing: independent interaural phase and level differences, lateral position, and short-term cross-correlation. Several preprocessing features were incorporated in the models: compression, temporal averaging, and envelope weighting. For the 14-Hz bandwidth data, the most successful model assumed independent centers for interaural phase and interaural level processing, and this model correlated with detectability at $r = 0.87$. That model also described the data best when it was assumed that interaural phase fluctuations and interaural level fluctuations contribute approximately equally to incoherence detection. For the 108-Hz bandwidth data, detection performance varied much less among different waveforms, and the data were less able to distinguish between models.

This work is dedicated to my amazing wife Sarah and loving family.

ACKNOWLEDGEMENTS

Along my long road of schooling, I have had many mentors. I might have become a biologist, chemist, or mathematician if it had not been for my high school physics teacher. Jim Gormley opened my eyes to the world of physics and I am truly grateful to him for that.

Most memorable of my mentors were my professors at Hope College, who allowed my mind to open to the love of learning. They helped me stay on track with my studies, like Dr. Paul DeYoung. They gave me encouragement when I was down, like Dr. Graham Peaslee. They gave me higher goals to strive for, like Dr. Cathy Mader. They were good friends, like Dr. Tim Pennings.

At Michigan State University, I have received as much support as I could hope for during graduate school. My advisor, Dr. William Hartmann, has been very influential in my personal work ethic, my scholarly preparation, and my newly-acquired love for the field of psychoacoustics. If nothing else, I think he and I were the perfect match for advisor and student.

I would also like to acknowledge Dr. Steve Colburn, Dr. Nat Durlach, and Dr. Brad Rakerd for providing invaluable insights that I used to complete this work. Dr. Hayder Radha and Dr. Karim Oweiss also deserve mention for our discussions on stochastic systems and signal detection theory, topics that I have found very useful in my dissertation.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 4

Chapter 5

Chapter 6

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

KEY TO SYMBOLS

Introduction

- γ : Interaural cross-correlation function
- δ : Lag of the cross-correlation function
- x_L : Signal to left ear
- x_R : Signal to right ear
- T : Duration of signal or period

Chapter 1

Experiment 1

- x_A : Randomly generated noise A
- x_B : Randomly generated noise B
- N : Number of components
- C_n : Amplitude of Fourier component
- ω_n : Angular frequency of Fourier component
- ϕ_n : Phase of Fourier component
- f_n : Frequency of Fourier component
- x'_B : The new noise B after orthogonalization to noise A
- M : Number of samples in a channel
- A_{rms} : Root-mean-square value of noise A
- B_{rms} : Root-mean-square value of noise B
- ρ_{AB} : Correlation between noise A and B
- α : Mixing factor
- ρ : Coherence evaluated at time zero
- $s(t)$: Temporal window
- $\hat{x}$: Analytic signal
- $\Re$: Real part of analytic signal
- $\Im$: Imaginary part of analytic signal

- $\phi(t)$: Phase of analytic signal
- $E(t)$: Envelope of analytic signal
- $\Delta\Phi(t)$: Interaural phase difference
- $\Delta L(t)$: Interaural level difference
- $s_t[\Delta\Phi]$: Standard deviation over time of the interaural phase difference
- $s_t[\Delta L]$: Standard deviation over time of the interaural level difference
- $\overline{\Delta\Phi}$: Mean interaural phase difference, averaged over time
- $\overline{\Delta L}$: Mean interaural level difference, averaged over time

Experiment 5

- $s_t[E]$: Standard deviation over time of envelope
- $\overline{E}$: Mean of envelope, averaged over time

Chapter 2
Experiment 7

- $\hat{e}[f(t)]$: Exponential averaging function
- τ : Exponential averaging time constant
- T_D : Duration of exponential averaging window
- $z(t)$: Auditory image position variable
- c : Time/intensity trading ratio
- ρ_{MR} : Correlation of detection data between matched and random noise-pairs

Chapter 4
Experiment 13

- $\Psi^{\circ}_{\Delta\Phi}$: Compressed interaural phase difference
- $\Psi^{\circ}_{\Delta L}$: Compressed interaural level difference
- $w_g(t)$: Envelope weighting function
- g : Envelope weighting threshold
- $\Psi_{\Delta\Phi}$: Transformed interaural phase difference
- $\Psi_{\Delta L}$: Transformed interaural level difference

- a : Weighting parameter for independent centers models
- d_n : Model number
- $W[h, \Psi(t)]$: Threshold function
- b : Weighting parameter for lateral position models
- Ψ_z : Transformed lateral position variable
- $\gamma_{\mathrm{St}}(t)$: Short-term cross-correlation
- Ψ_{CC} : Transformed short-term cross-correlation
- $s_t[\Psi']$: Standard deviation over time of the derivative of the transformed interaural difference
- $s_t[E']$: Standard deviation over time of the derivative of the envelope

Appendix A

- N_{corr} : Number of correct responses
- N_{conf} : Number of confident responses

KEY TO ABBREVIATIONS

- 2AFC : Two-alternative forced-choice
- ASW : Apparent source width
- AVCN : Anteroventral cochlear nucleus
- *CAS* : Confidence adjusted score
- EC : Equalization-cancellation
- EE : Excitatory-excitatory
- EI : Excitatory-inhibitory
- DCN : Dorsal cochlear nucleus
- IACC : Interaural cross-correlation
- IC : Inferior colliculus
- ILD : Interaural level difference
- IPD : Interaural phase difference
- ITD : Interaural time difference
- jnd : Just-noticeable difference
- LEV : Listener envelopment
- LL : Lateral lemniscus
- LP : Lateral position units
- LSO : Lateral superior olive
- No : Noise in-phase
- NoSπ : Noise in-phase, signal out-of-phase
- Nπ : Noise out-of-phase
- MLD : Masking level difference
- MGB : Medial geniculate body
- MSO : Medial superior olive
- MTB : Medial nucleus of the trapezoidal body
- NLL : Nucleus of the lateral lemniscus

- P_c : Percent correct
- P_s : Percent selected
- pdf : Probability density function
- PVCN : Posteroventral cochlear nucleus
- RMS : Root-mean-square
- STCC : Short-term cross-correlation
- YN : Yes-no

INTRODUCTION

0.1 Road map to this work

This dissertation plans to study the ability of human listeners to detect interaural incoherence, or the difference in the signals presented to the left and right ears. Listeners are remarkably good at distinguishing slightly incoherent noise from coherent noise, when the signals in the two ears are identical. The cue used to detect interaural incoherence is thought to be much like the cue used to detect masking-level differences, where an out-of-phase tone between the two ears is more easily detectable than an in-phase tone in otherwise coherent noise. The reason for this is that the out-of-phase tone adds a small amount of incoherence to the coherent noise. There has been a vast amount of research done on masking-level differences in the last 60 years.

Incoherence detection is thought to be described wholly by the coherence function, which is an overlap integral over the left and right signals over the duration of the stimulus. If this is the case, dynamic features in the incoherent noise will be lost. However, the perception of incoherent noise has dynamic characteristics. Therefore, it seems appropriate to ask the following two questions: 1) Is cross-correlation or coherence function an adequate parameter for incoherence detection? 2) If not cross-correlation or coherence function, then what parameter is important to incoherence detection? These questions will be addressed by selecting noise-pairs that will explore the effect of bandwidth on incoherence detection. Chapter 1 will use noise-pairs with a fixed value of coherence. The results of this chapter will find that the noise-pairs with the largest standard deviation over time (fluctuations) of the interaural phase difference and interaural level difference will have incoherence that is much more readily detectable than noise-pairs with the smallest fluctuations. Since the coherence function predicts that incoherence detection should be the same in all these noises,

it will be concluded that the coherence function is not used in incoherence detection. As the bandwidth increases, the rate of fluctuations also increases and the difference between noise-pairs with the largest fluctuations and smallest fluctuations in the interaural parameters is less perceptible to listeners. In the case of wide bands, it would be possible to use the coherence function to describe the detection data, but only because of the ergodicity of the fluctuations for wide bands.

Chapter 2 will use noise-pairs with varied values of coherence, but similar distributions of fluctuations in phase and level differences. The values of coherence ranged from 0.969–998. For the average listener, incoherence would be very easy to detect in a noise-pair with a value of coherence of 0.969. For the average listener, incoherence would be almost impossible to detect in a noise-pair with a value of coherence of 0.998. There are no new results in Chapter 2, and its value is that the results of Chapter 1 are reproduced by using only the fluctuations in phase and level.

The binaural integration time is tens to hundreds of milliseconds. Therefore there may be a fundamental flaw to the results of Chapters 1 and 2, which had 500-ms noises. It may be the case that incoherence detection is done in small analysis windows much shorter than 500 ms. Chapter 3 will address this criticism by using the same method of selecting noise-pairs with large and small fluctuations as was done in Chapters 1 and 2. However, the noises will now be five to twenty times shorter in duration and only the narrow bandwidth will be used. The results of this chapter will find that listeners will be able to use interaural fluctuations of phase and level for stimuli as short as 50 ms. The results will be comparable to those in Chapters 1 and 2. For 25-ms noise-pairs, listeners will use a lateralization cue instead of a width cue to perform the task. Coherent noise is normally described as a compact sound image in the center of the head. Incoherent noise is normally described as a wide or diffuse sound image in the center of the head. A lateralization cue, for this experiment, would be a compact sound that has moved from the center of the head,

usually associated with a static interaural phase difference or static interaural level difference. It was necessary for listeners to use lateralization because it was found that the interaural differences did not change much over the 25-ms duration of the noises.

After these preliminary experiments, the results from Chapters 1–3 will motivate the direction of the modeling of the incoherence detection data. Chapter 4 will concentrate on developing and refining models that describe incoherence detection from a reference correlation of unity. The models will use different combinations of phase and level statistics and several aspects of the models will be similar to those used in describing masking-level difference data, since so much work has been done in this area. As was done in Chapter 1, Chapter 4 will use noise-pairs with a fixed value of coherence. Three types of models will be explored. The first type of model will add fluctuations of phase and fluctuations of level independently. The second type of model will combine phase and level differences and then calculate a statistic for fluctuations. In this model type, it would be possible for a phase difference to the right ear to be cancelled by a level difference to the left ear. This cancellation has been shown to occur in time-intensity trading experiments with static interaural differences. The last type of model reduces the coherence function to a short-term coherence function, that evaluates the coherence over very small time intervals. It will be shown that the incoherence detection data can be best described by a model that independently adds the standard deviation of phase to the standard deviation in level with equal weight for phase and level. Compression of the interaural difference scale, temporal averaging, and envelope weighting will also be applied to improve the model.

As was done in Chapter 2, Chapter 5 will use noise-pairs with varied values of coherence. Again, there are no new results in Chapter 5, and its value is that the results of Chapter 4 are reproduced.

Lastly, Chapter 6 will apply the modeling results of Chapters 4 and 5 to directly answer these two questions: 1) Can listeners use just phase fluctuations or just level fluctuations to detect incoherence? and 2) Do listeners use independent IPDs and ILDs as a function of time and not a single auditory image made of the combination of the two interaural parameters? It will be found that the answer to both of these questions is "yes". However, not all of the variation of the data will be explained, motivating more studies.

From the experiments in these six chapters of research, it will be concluded that incoherence detection is done by detection fluctuations in the interaural phase and level differences, contrary to other recent results. Before this dissertation, van de Par and Kohlrausch (1999) found no use for interaural parameters in incoherence detection. The reason for the discrepancy is that van de Par and Kohlrausch did not find a bandwidth dependence for interaural parameters, whereas there was one found in this dissertation. Another reason that the coherence function has been used to describe incoherence detection is possibly due to the stimuli that have been used in previous experiments. Critical bandwidth noises of about 100 Hz (for a center frequency of 500 Hz) and wide bandwidth noises of thousands of Hz have been used for experiments where individual features in the interaural parameters are shown to be too fast to consistently recognize.

Two types of models that are often used to describe masking-level difference data are the equalization-cancellation model and correlation model. However, from the results that will appear in the following chapters, both of these models should only be applied to stimuli that have wide bandwidths and similar fluctuations. This is because these models are energy-based models that have no mechanism to analyze interaural parameters differences over short intervals.

The results of this dissertation have the potential to motivate several new studies in the area of incoherence detection. For example:

- More research will need to be done to understand why listeners confuse monaural and interaural fluctuations. The monaural confusion results can then be applied to false-alarm rates as measured in single-interval masking-level difference experiments in reproducible noises (Gilkey *et al.*, 1985, 1986; Isabelle and Colburn, 1991; Evilsizer *et al.*, 2002).

- It will be important to study the effect of varying the center frequency of the incoherent noises and apply the results to high frequency masking-level difference data (Bernstein and Trahiotis, 1992).

- It will be important to vary the value of the interaural coherence to change the distribution of fluctuations of phase and level. The detection of incoherence might be different enough that results in this dissertation does not apply to other values of coherence.

- It will be important to make a more sophisticated model to describe incoherence detection data. This can be done by modeling brainstem nuclei firing patterns in the superior olivary complex.

0.2 Coherence

Interaural coherence is a measure of the similarity of signals between a listener's two ears. It is derived from the interaural cross-correlation function, $\gamma(\delta)$, which is a function of the interaural time shift δ,

$$\gamma(\delta) = \frac{\int_0^T x_L(t)x_R(t+\delta)dt}{\sqrt{\int_0^T x_L^2(t)dt \int_0^T x_R^2(t)dt}}, \tag{1}$$

where x_R is the signal in the right ear, and x_L is the signal in the left. The cross-correlation is bounded by $-1 \leq \gamma \leq 1$.

With respect to perception, interest normally centers on the peak of $\gamma(\delta)$. The

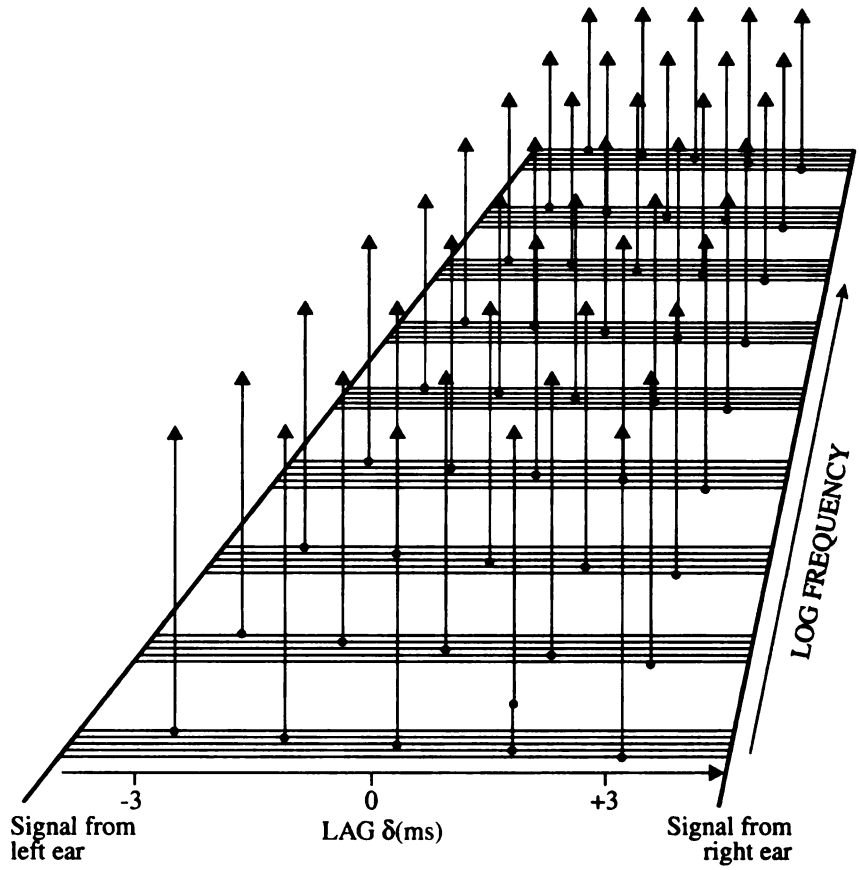

Figure 1: The Jeffress coincidence matrix. The characteristic frequency of bands (horizontal lines) increase logrithmically. The vertical arrows represent fibers attached to neurons (black dots) that fire when signals from both the left and right ears temporally coincide for different values of the interaural lag δ. (From Hartmann, 1999)

value of δ for which the peak occurs is the perceptually relevant interaural time difference (ITD) cue for the location of the sound image. This value of δ was given a place representation in the famous binaural model by Jeffress (1948). Figure 1 shows how the Jeffress coincidence model is composed of different bands with a certain characteristic frequency. Different cells fire depending on the ITD and frequency of the signals presented to the ears.

The height of the peak is thought to determine the compactness of the image. If the sounds in the two ears are identical except for an interaural delay or interaural

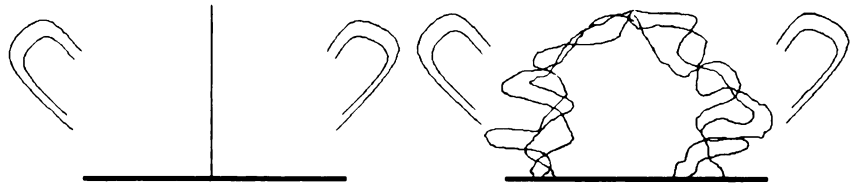

Figure 2: An artist's rendition of incoherence perception. The left figure shows a compact, coherent stimulus between the listener's ears. The right figure shows how the image is broader and more diffuse for an incoherent stimulus.

level difference, then γ has its maximum value of 1, and the image is expected to be maximally compact. If the height of the peak is less than 1, the image is broader or more diffuse (Barron, 1983; Blauert and Lindemann, 1986). Figure 2 shows a cartoon of the perception of incoherence.

The height of the cross-correlation peak is defined as the coherence. It is postulated that to be physiologically relevant, the peak must occur at a value of the interaural lag that lies in a range of perceptible ITDs ($\pm$ 763 μs for the average human head, which is determined by dividing the width of the head by the speed of sound), and this requirement may place restrictions on the peaks for which the concept of coherence is valid.

As applied in architectural acoustics, the coherence is called the IACC (interaural cross-correlation), and it refers to the height of $\gamma(\delta)$ for δ in the range $-1 \leq \delta \leq 1$ ms (Beranek, 2004). In recent years this objective architectural measure has been divided into two separate measures. One is the apparent source width (ASW), based on coherence within 80 ms of the onset of a sound (Barron and Marshall, 1981). The other is listener envelopment (LEV), determined by the coherence of later arriving sound, as measured after 80 ms (Bradley and Soulodre, 1995; Barron, 2004). Together, the ASW and LEV greatly influence the spatial impression of a sound in a room. Normally, the architectural measurements are made with microphone techniques and not with artificial heads.

Perceptual aspects of cross-correlation function and interaural coherence have been

studied by psychoacousticians, usually using bands of noise as stimuli. Noise provides an abstraction of real-world sounds that is devoid of meaningful information, and it affords many opportunities for parametric variations. Using broadband noise, Pollack and Trittipoe (1959a,b) found thresholds for changes in cross-correlation for two values of the reference correlation, namely 1.0 (i.e. No) and −1.0 (i.e. Nπ). They explored the effects of duration, sound level, frequency range, and interaural level difference.

Listeners are particularly sensitive to deviations from a reference correlation of 1.0. Using narrowband noise, Gabriel and Colburn (1981) found that listeners could easily distinguish between noise with a coherence of 1.0 and noise with a coherence of 0.99. They also reported the somewhat counterintuitive result that as the bandwidth of the noise increases, the just-noticeable difference (jnd) also increases. One might have expected the jnd to decrease instead given that a wider bandwidth generally offers the listener more information.

For a reference coherence less than unity, a listener's ability to discriminate a difference in incoherence degrades appreciably. This decrease is approximately an order of magnitude for a reference coherence of 0.0 (Gabriel and Colburn, 1981). It is possible for the signals to the two ears to become anti-correlated (when the signal in the left ear is equal to the signal in the right ear except for a difference in the sign). An anti-correlated signal places an image at both of the listener's ears, rather than at the center of the head. Boehnke *et al.* (2002) found that discrimination from a reference coherence of −1.0 was significantly worse than a reference of +1.0. It was also found that discrimination from a reference coherence of 0.0 was worse for partially anticorrelated noise when compared to partially correlated noise.

A reference coherence of 1.0 is also of interest in connection with the masking-level difference (MLD). The MLD phenomenon is the difference that occurs in the threshold to detect a tone in noise depending on the binaural masker and signal phase

configuration. Numerous studies have concluded that the threshold signal-to-noise ratio in the NoSπ (noise in-phase, signal out-of-phase by 180°) condition is essentially determined by the ability to detect the incoherence introduced by the out-of-phase signal (Wilbanks and Whitmore, 1967; Koehnke *et al.*, 1986; Durlach *et al.*, 1986; Bernstein and Trahiotis, 1992; van de Par and Kohlrausch, 1995).

The concept of coherence may be important to speech recognition, as discussed by Culling *et al.* (2001). The line of reasoning is as follows: Jain *et al.* (1991) showed that the reduction in the interaural cross-correlation function of the signal is monotonically related to the strength of the signal for signal intensities less than that of the masker. A crucial factor in identifying several speech sounds is an accurate estimation of the frequency of the first formant. It is thought that the accuracy of the estimation process is determined by relative intensities of peripherally-resolved harmonics in that frequency region. If this is the case, then the ability to discriminate different degrees of decorrelation might explain the binaural intelligibility level difference (Licklider, 1948; Carhart *et al.*, 1969a,b; Levitt and Rabiner, 1967a,b; Bronkhorst and Plomp, 1988). The binaural intelligibility level difference is the difference between using monaural and binaural information when trying to understand speech in a noise background.

0.3 Fluctuations of interaural parameters

Despite the prevalent use of the cross-correlation function to describe incoherence detection, time-dependent fluctuations seem to play a part in the perception of incoherent stimuli. This perception, which occurs when incoherent stimuli are presented with headphones, is usually described as including an auditory width or lateral displacement that varies in time. Given that the form of Eq. 1 is a "time-averaging" function that integrates over the entire duration of the stimulus, as opposed to a time-varying function, there appears to be a disconnection between the perception of incoherence and the measure of it via Eq. 1.

From the extensive work on modeling incoherence detection and the present knowledge of auditory physiology, it seems possible that the interaural parameters are more physically relevant than incoherence *per se* to the incoherence detection task. Fluctuations in interaural phase differences (IPD) and interaural level differences (ILD) have been studied for a few conditions. For example, probability distributions of IPDs and ILDs have been calculated by Zurek (1991) and Breebaart *et al.* (1999); they have been measured in rooms by Nix and Hohmann (2001).

0.4 Physiology

The pathway by which physical signals are transformed into neural signals starts with the acoustic wave impinging on the tympanic membrane (ear drum). The wave is then transferred via the ossicles (incus, malleus, and stapes) to the cochlea. The ossicles act as an impedance matching device; if the sound met the oval window directly, only 10% (or less) of the acoustic wave energy would be transferred to the inner ear or cochlea for a 1 kHz tone. Instead 65% of the acoustic wave energy is transferred (Møller, 1965). The cochlea is filled with fluid; the acoustic wave, via the middle-ear ossicles, causes pressure waves in the fluid, which, in turn, cause the basilar membrane to move. Attached to the basilar membrane are hair cells, which depolarize under mechanical stress and stimulate the generation of action potentials by the auditory nerve. Measurements of vibration patterns of the basilar membrane show that it is *tonotopically* organized or organized by increasing frequency. All of the auditory system is tonotopically organized.

From the cochlea, information travels on the auditory nerve. The cochlear nucleus is subdivided into the sections: the anteroventral nucleus (AVCN), the posteroventral nucleus (PVCN), and the dorsal nucleus (DCN). The AVCN projects to both the ipsilateral and contralateral superior olivary complexes and therefore is important to providing binaural information such as ITDs and ILDs, which are used in sound

localization. The DCN projects only contralaterally and therefore is important to monaural acoustic information such as level discrimination. The DCN processes the signals with non-monotonic firing rate-intensity functions and inhibitory sibebands. The PVCN is thought to perform both monaural and binaural processing because it has characteristics of both the AVCN and DCN. Unlike the DCN, the AVCN does little processing of the signals and projects to the superior olivary complex with myelinated fibers that keep the timing information intact.

It is well known that the superior olivary complex processes time and level difference between the ears in mammals. Specifically, Goldberg and Brown (1968, 1969) found that the medial superior olivary nucleus (MSO) in dogs analyzes time differences. This is done with excitatory-excitatory (EE) cells that fire when impulses are received at the same time. Tsuchitani and Boudreau (1966) and Tsuchitani (1977) found that the lateral superior olivary nucleus (LSO) in cats analyzes level differences. The LSO contains mostly excitatory-inhibitory (EI) cells.

Yin and Chan (1990) and Smith, Joris, and Yin (1993) showed that it is plausible for the MSO to act as the Jeffress coincidence matrix seen in Figure 1. They discovered delay lines from the cochlear nucleus to the EE cells in the MSO of cats. However, new work on the necessity of including EI cells in the MSO to accurately measure time differences has recently presented some problems to the Jeffress model (Brand *et al.*, 2001). Measurements by Goldberg and Brown (1968) showed that the MSO contains approximately 75% EE cells and 25% EI cells. However, the results of Brand *et al.* (2001) imply that a majority of MSO cells need inhibitory inputs. In either case, an additional inhibitory component to the Jeffress model appears both physiologically possible and pertinent.

Beyond the superior olivary complex, neurons project to the nucleus of the lateral lemniscus (NLL), then the inferior colliculus (IC) where auditory maps of space are made (Knudsen and Konishi, 1979), then the medial geniculate body (MGB), and

finally the auditory cortex. The processing of stimuli becomes increasingly complex and is still being researched in these areas. Given that interaural parameters are processed in the midbrain and that detection of incoherence is perceived as time-varying fluctuations, investigating the tie between incoherent stimuli and interaural differences is both psychologically and physiologically motivated.

0.5 Purpose of this work

This dissertation is concerned with incoherence detection starting with perfectly coherent noise as a reference. Its working hypothesis is based on the suspicion that the extreme sensitivity shown by listeners to small amounts of incoherence is not properly described by the cross-correlation function. The reason is that when a small amount of incoherence is added to an otherwise perfectly coherent noise, the image of the noise acquires lateral fluctuations that are not present when the coherence is 1.0. The hypothesis continues with the observation that, whereas coherence is a measure that is averaged over time, the fluctuations that are imagined to be the basis of coherence discrimination are dynamic. Thus, although the coherence measure may be a mathematically useful characterization of the similarity or dissimilarity of signals in the two ears, this measure may not be the most perceptually relevant characterization. Instead, it is possible that some measure that specifically considers fluctuations is better. A similar point of view with respect to the MLD has been taken by a number of authors, e.g. Jeffress *et al.* (1956). The rest of this dissertation describes experiments that test this hypothesis.

1 INCOHERENCE AND FLUCTUATIONS OF INTERAURAL PARAMETERS

1.1 EXPERIMENT 1: NARROW BANDWIDTH

The purpose of Experiment 1 was to determine whether the height of the peak of the cross-correlation function (the coherence) adequately describes incoherence detection given a reference coherence of 1.0. The experiments employed reproducible noises (to be called left-right noise-pairs, or noise-pairs), as have been used in MLD experiments by Gilkey *et al.* (1985, 1986), Isabelle and Colburn (1991), and Evilsizer *et al.* (2002). An advantage of using reproducible noises instead of randomly generated noises for incoherence detection is that reproducible-noise data should provide a more stringent and confining test of the binaural models. This is because the individual properties of the noise-pairs will be accessible to the experimenter after the experiment is performed. An advantage of using incoherence detection data instead of MLD data is that incoherence detection employs a simpler stimulus, which is spectrally homogeneous unlike a tone-in-noise situation. Certain problems may arise for MLD tasks that would not during incoherence detection; Wightman (1971) noted inconsistencies in narrowband MLD results that may have been due to off-frequency listening.

In Experiment 1 and all the other experiments in this chapter, the different reproducible left-right noise-pairs had the same value of interaural coherence. If the coherence is an adequate measure of perception, then all the reproducible noise-pairs will be equally distinguishable from perfectly coherent noise.

1.1.1 Stimuli

A collection of 100 two-channel noises with reproducible amplitudes and phases was created for Experiment 1. The process began with two waveforms, A and B, written

as a sum of cosines in the form

$$x_A(t) = \sum_{n=1}^{N} C_n^A \cos(\omega_n t + \phi_n^A)$$

and

$$x_B(t) = \sum_{n=1}^{N} C_n^B \cos(\omega_n t + \phi_n^B)$$

where the C_n's and ϕ_n's are the amplitudes and phases of the spectral components. The narrowband noises were generated with components having random phases over a frequency range of 490–510 Hz and with a frequency spacing of 2 Hz. Components between 495–505 Hz had equal amplitudes of unity. Frequencies below 495 Hz had a raised-cosine window applied to the amplitude spectrum of the form

$$C_n^A = C_n^B = \sin^2\left[\frac{\pi(f_n - 490)}{10}\right] \text{ for } 490 \leq f_n \leq 495 \tag{2}$$

to minimize any spectral edge effects in the noises. The amplitudes of components from 505–510 Hz were similarly windowed. This frequency shaping can be seen in Figure 3. The 3-dB bandwidth of the noise was therefore 14 Hz.

For each noise in a collection of 100 reproducible noise-pairs, the B noise (x_B) was orthogonalized to the A noise (x_A) by the Gram-Schmidt orthogonalization procedure, which was used in Culling *et al.* (2001), and which is described below. The orthogonalized B noise is here denoted as x'_B. This was done to ensure that the signals were uncorrelated and that the final value of the cross-correlation after mixing would be precise. For large bandwidths, the large number of random variables (the ϕ_n's are uniformly distributed) yields statistically independent noise-pairs. In the case of large noise bands, orthogonalization is probably not necessary. However, there are a small number of random variables (random phases) for the 14-Hz bandwidth noises of this experiment. Experimental observations showed the projection between randomly

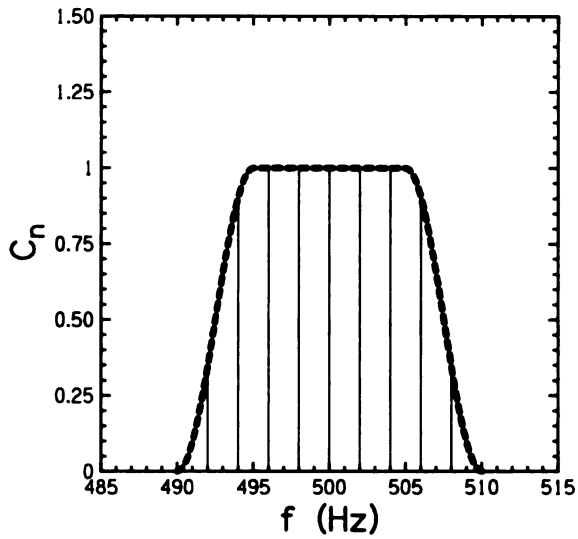

Figure 3: The frequency shaping in Eq. 2 that is applied to all stimuli in Experiment 1. This raised-cosine window was applied to remove any spectral edge effects.

generated noise-pairs can be as large as 10% for this bandwidth. Therefore, it was imperative to use the orthogonalization procedure for the 14-Hz bandwidth noise-pairs in order to be able to generate noise-pairs with the desired interaural coherence.

To orthogonalize x_B to x_A, the root-mean-square (RMS) is found for each channel:

$$A_{\text{rms}} = \sqrt{\frac{\sum_{t=1}^{M}[x_A(t)]^2}{M}} = \sqrt{\frac{\sum_{t=1}^{M}[\sum_{n=1}^{N} C_n^A \cos(\omega_n t + \phi_n^R)]^2}{M}} = \sqrt{\frac{1}{2}\sum_{n=1}^{N}(C_n^A)^2},$$

and

$$B_{\text{rms}} = \sqrt{\frac{\sum_{t=1}^{M}[x_B(t)]^2}{M}} = \sqrt{\frac{\sum_{t=1}^{M}[\sum_{n=1}^{N} C_n^B \cos(\omega_n t + \phi_n^L)]^2}{M}} = \sqrt{\frac{1}{2}\sum_{n=1}^{N}(C_n^B)^2}$$

where M is the number of samples in each sound. Then the overlap between the two channels is calculated:

$$\begin{aligned}\rho_{AB} &= \frac{\sum_{t=1}^{M} x_A(t)x_B(t)}{M A_{\text{rms}} B_{\text{rms}}} = \frac{\sum_{t=1}^{M}\sum_{n=1}^{N} C_n^A C_n^B \cos(\omega_n t + \phi_n^A)\cos(\omega_n t + \phi_n^B)}{M A_{\text{rms}} B_{\text{rms}}} \\ &= \frac{1}{2}\sum_{n=1}^{N}\frac{C_n^A C_n^B \cos(\phi_n^A - \phi_n^B)}{A_{\text{rms}} B_{\text{rms}}}.\end{aligned}$$

Next, the correlated component of x_A is subtracted from a scaled x_B ensuring the two channels have equal power,

$$x'_B(t) = \frac{A_{\text{rms}}}{B_{\text{rms}}\sqrt{1-\rho_{AB}^2}} x_B(t) - \frac{\rho_{AB}}{\sqrt{1-\rho_{AB}^2}} x_A(t) \text{ for } 1 \leq t \leq M.$$

The two perfectly uncorrelated noises were then mixed, with mixing factor α, to create the final left and right (L and R) noise-pairs to be sent to the listeners,

$$\begin{aligned} x_L &= x_A \\ x_R &= \sqrt{(1-\alpha^2)}\, x_A + \alpha\, x'_B. \end{aligned} \tag{3}$$

The coherence, as defined by Eq. 1, is then computed to yield,

$$\rho = \gamma(0) = \sqrt{1-\alpha^2}. \tag{4}$$

For values of ρ near 1, the value of interaural lag that maximizes ρ is zero. Because the mixing factor used in all of the experiments was $\alpha = 0.125$, the interaural coherence of all the noise-pairs was $\rho = 0.9922$.

After mixing, each noise was given a time interval shape with a total duration of 500 ms. A temporal window, $s(t)$, was applied such that there were cosine-squared edges with rise/fall times of 30 ms and a full-on duration of 440 ms, which can be seen in Figure 4. The application of the temporal window could change the coherence of the noise; therefore, the value of the coherence was measured after the window was applied so that $\rho = 0.9922 \pm 0.0001$. Noise-pairs that did not meet this criterion were rejected. For the 100 noise-pairs accepted in this experiment, 875 were rejected. Also, note that after orthogonalization and applying the temporal window, the noise-pairs do not necessarily have equal amplitudes.

To determine the time-dependent interaural phase difference (IPD) and interaural

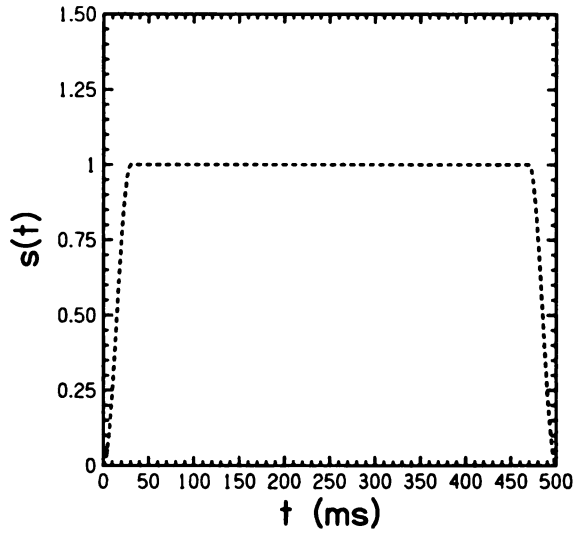

Figure 4: The temporal window, $s(t)$, that is applied to all stimuli in this chapter. All the stimuli had 30 ms rise/fall times and a full-on duration of 440 ms.

level difference (ILD), the analytic signals were found. By eliminating the negative frequencies, the analytic signal for either x_L or x_R is

$$\hat{x}(t) = s(t) \sum_{n=1}^{N} C_n \exp[i(\omega_n t + \phi_n)],$$

where $s(t)$ is the temporal window, the C_n's are the left or right amplitudes, and ϕ_n's are the left or right phases as required.

By Euler's relation, the analytic signal becomes

$$\hat{x}(t) = s(t) \sum_{n=1}^{N} C_n[\cos(\omega_n t + \phi_n) + i \sin(\omega_n t + \phi_n)] = \Re(t) + i\Im(t). \tag{5}$$

The phase and envelope of the analytic signal as a function of time are

$$\phi(t) = \arg[\Im(t), \Re(t)] \tag{6}$$

and

$$E(t) = \sqrt{\Re^2(t) + \Im^2(t)} \tag{7}$$

where the arg function is the arctangent with possible quadrant correction. The time-dependent IPD (radians) and ILD (dB) of the analytic signal are then defined as

$$\Delta\Phi(t) = \phi_R(t) - \phi_L(t). \tag{8}$$

and

$$\Delta L(t) = 20\log_{10}\left[\frac{E_R(t)}{E_L(t)}\right]. \tag{9}$$

Equation 8 yields a positive value of $\Delta\Phi(t)$ for signals that lead in the right ear. Similarly, Eq. 9 gives a positive value of $\Delta L(t)$ for a signal that has a larger level in the right ear. The interaural phase $\Delta\Phi(t)$ was required to remain in the physically relevant region of $\pm\pi$ radians at every point in time, and was corrected by adding or subtracting 2π when necessary. The ITD is proportional to the IPD for narrow bandwidths. The ITD can be calculated by dividing the IPD by the angular center frequency of $2\pi \times 500$.

1.1.2 Signal Structure Analysis

Figure 5 shows a plot of the IPD and ILD as a function of time for an arbitrarily chosen noise-pair from the collection of noise-pairs with $\rho = 0.9922$ and a bandwidth of 14 Hz. The figure shows that the IPD and ILD fluctuate as a function of time. Peaks in the IPD and ILD sometimes coincide in time and can be directed towards the same ear or opposite ears.

An analysis was done to find how often the interaural differences were in the same and different lateral directions. This was done by comparing the direction of the interaural differences for each of the 4000 samples in all the 100 noise-pairs from this experiment. The percentage of time that the interaural differences were in the same direction was calculated for the average noise-pair (averaged over 100). In addition, the individual noise-pairs with the minimum and maximum percentage of samples

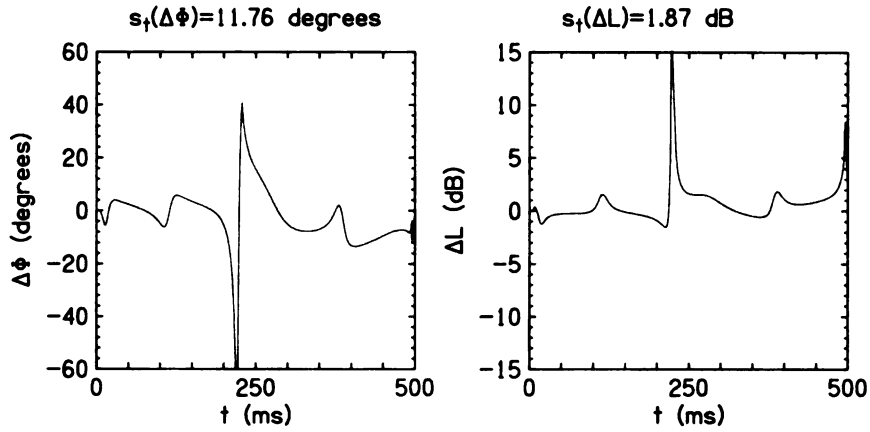

Figure 5: The IPD and ILD plotted as a function of time for an arbitrary 14-Hz bandwidth noise-pair. Above the plot are the fluctuation values of the IPD and ILD. Positive values of the IPD are defined as leading in the right ear; positive values of ILD are defined as being louder in the right ear.

that had interaural differences in the same direction were found; so were the noise-pairs that had interaural differences in opposite directions. Lastly, it was found how often the IPD was large when the ILD was large, irrespective of lateral direction. Each parameter was defined to be large when it was greater than the respective RMS value of the interaural differences over time.

Table 1: Average (over 100 noise-pairs), minimum (over 100 noise-pairs), and maximum (over 100 noise-pairs) percentage of samples where IPD and ILD are directed towards the same ear or different ears. It also shows when the IPD and ILD are large with respect to the RMS interaural difference.

	Average	Minimum	Maximum
Same direction	46.3%	19.4%	72.9%
Opposite direction	48.8%	20.1%	74.8%
Large in both	43.0%	20.8%	82.0%

Table 1 shows that for the 100 noise-pairs in this experiment the IPD and ILD can be either in the same direction (both towards the right or both towards the left) or in the opposite direction (one to the right, one to the left). The IPD and ILD are in the same direction and opposite directions about 50% of the time for the

average noise-pair. The total of 46.3% and 48.8% do not add up to 100% because either the IPD or ILD is zero 4.9% percent of the duration for a typical noise-pair. However, there are improbable noise-pairs where the interaural differences go in the same direction as little as 19.4% of the duration and as much as 72.9%. This is similar to the percentages for the interaural differences going in different direction.

Possibly the most interesting part of this analysis is whether a large peak in one interaural difference implies a large peak in the other interaural difference at the same sample number. Figure 5 shows a large peak (greater than the RMS value) in IPD going to the left ear at 225 ms and large peak in ILD going to the right at the same time. On average, only 43.0% of the large peaks in IPD or ILD correspond to large peaks in both IPD and ILD. The most extreme stimuli have peaks that coincide as little as 20.8% of the time and as much as 82.0% of the time. Therefore, there is a great amount of variability in individual noise-pairs with respect to fluctuations of IPD and ILD depending on the phase relationship.

1.1.3 Signals

Next, a quantitative measure of interaural fluctuations is defined. The IPD fluctuation over time is defined as

$$s_t[\Delta\Phi] = \sqrt{\frac{1}{M}\sum_{t=1}^{M}[\Delta\Phi(t) - \overline{\Delta\Phi}]^2} \tag{10}$$

where M is the number of samples in the 500 ms noise (M = 4000 samples), $\overline{\Delta\Phi}$ is the mean IPD computed over time, and the t subscript indicates that this standard deviation value is computed over time. The ILD fluctuation over time is defined as

$$s_t[\Delta L] = \sqrt{\frac{1}{M}\sum_{t=1}^{M}[\Delta L(t) - \overline{\Delta L}]^2}. \tag{11}$$

Figure 5 shows the IPD and ILD changing as a function of time, with values of $s_t[\Delta\Phi]$

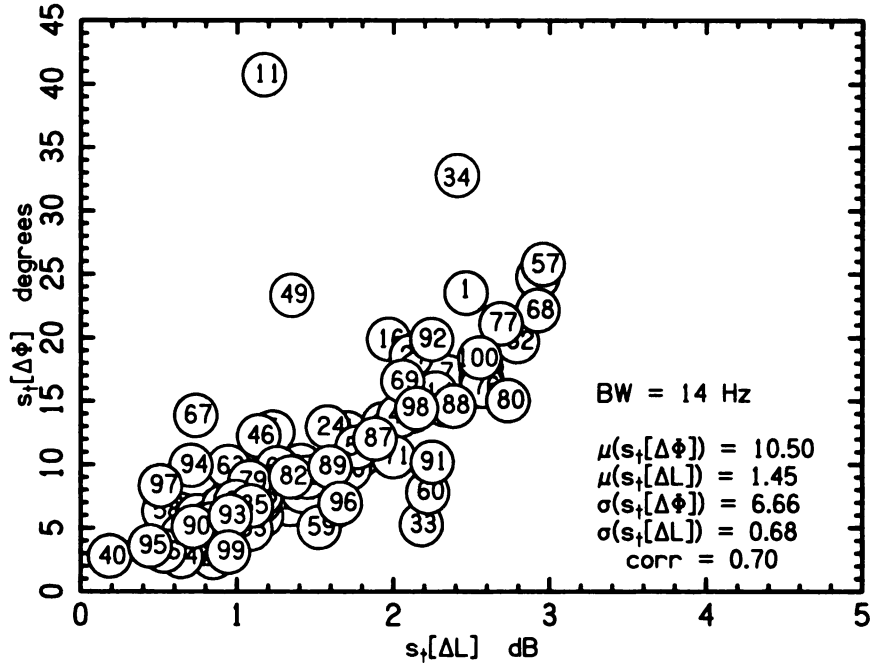

Figure 6: Fluctuations of IPD versus fluctuations of ILD for the collection of 100 reproducible noise-pairs having a 14-Hz bandwidth, as used in Experiment 1. Each noise-pair is labeled by a serial number indicating only the order of creation. The means, standard deviations, and IPD-ILD correlation of the distributions are reported.

and $s_t[\Delta L]$ above each panel. For the entire collection of 100 14-Hz bandwidth noise-pairs, Figure 6 shows $s_t[\Delta\Phi]$ plotted against $s_t[\Delta L]$. The mean, standard deviation, and correlation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ are reported on the figure. This figure shows that $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ are strongly correlated. However, there are still some improbable noise-pairs, which were predicted to occur from the signal structure analysis of the previous section. For example, noise-pair #11 has the largest IPD fluctuations but only average ILD fluctuations. (It is interesting to see that the correlation between phase fluctuations and level fluctuations across different noise-pairs is fairly strong, 0.70. However, as shown in Appendix B, a value of 0.70 is somewhat low relative to a value computed for 5000 noise-pairs and for this bandwidth.)

Numerical studies were done to compare the equal-amplitude noise-pairs used in this experiment to Rayleigh distributed noise-pairs. It was found that there was no

difference in the probability density functions of the IPD and ILD over time. It was also found that using Rayleigh distributed noise-pairs did not affect the distributions of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$.

To perform Experiment 1, the five noise-pairs with the greatest fluctuations of IPD and the five noise-pairs with the smallest fluctuations of IPD were selected to form a *phase set* of ten reproducible noise-pairs. Similarly, the ten noise-pairs with the greatest and smallest fluctuations of ILD were selected to form a *level set.* The convention in this dissertation is to plot detection data of phase sets with circles and level sets with boxes.

The two channels of noise were computed and downloaded by a Tucker-Davis AP2 array processor (System II) and converted from a digital signal by 16-bit DACs (DD1). The buffer size was 4000 samples per channel and the sample rate was 8 ksps. The noise was lowpass filtered with a corner frequency of 4 kHz and a −115 dB/octave roll off. The noises were presented at 70 dB ± 3 dB with levels determined by programmable attenuators (PA4) operating in parallel on the two channels. The levels of the two channels were equal and were randomized over a range of ±3 dB for each of the three intervals within a trial in order to discourage the listener from trying to use overall level cues to perform the task.

1.1.4 Procedure

Listeners were tested individually, seated in a double-walled sound attenuating room, and using Sennheiser HD414 headphones. Six runs were devoted to listening to a set of ten reproducible-noise pairs. A noise-pair could be presented either incoherently - the dichotic presentation of x_L and x_R - or it could be presented coherently - the diotic presentation of x_L. A run consisted of 60 trials, where each of the ten reproducible noise-pairs in a set was presented incoherently a total of six times. Thus, listeners heard an individual noise-pair incoherently a total of 36 times (six runs times six

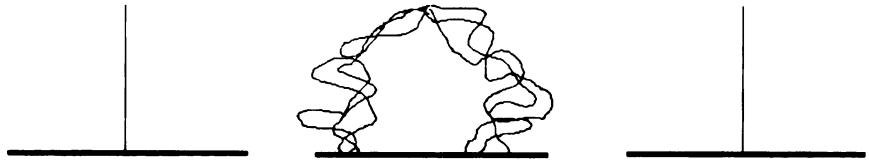

Figure 7: A pictorial representation of the three-interval sequence used in this experiment. The first interval was always coherent, represented by the straight line on the far left. The second interval can be either incoherent or coherent. In this figure the second interval is incoherent, represented by a fuzzyness or increased width between the ears. The last interval needs to be opposite from the second interval. Therefore, for this example, it is a coherent, compact image.

presentations per run).

On each trial the listener heard a three-interval sequence, as illustrated in Figure 7. The first interval was the standard interval, which was always a coherent noise. The second interval was randomly chosen to be either incoherent or coherent. The third interval was the opposite of the second (e.g. if the second interval was coherent, the third interval was incoherent). The two coherent presentations were randomly selected from the remaining nine reproducible noises in the set except that they were required to be different from the incoherent "odd" interval and to be different from one another. The inter-interval duration was 150 ms.

Listeners were instructed to "choose the interval that was different from the other two." Initial training with noise-pairs with a small values of coherence (i.e. $\rho = 0.95$) were used so that the difference between incoherent and coherent stimuli would be obvious. Feedback was given to the listeners so that the task could be properly learned. The value of coherence for the noise-pairs was increased to a final value of 0.9922 as listeners became better at identifying incoherent noise-pairs. The training was finished when listeners could identify the incoherent noise-pair approximately 75% of the time for noise-pairs with a bandwidth of 136 Hz.

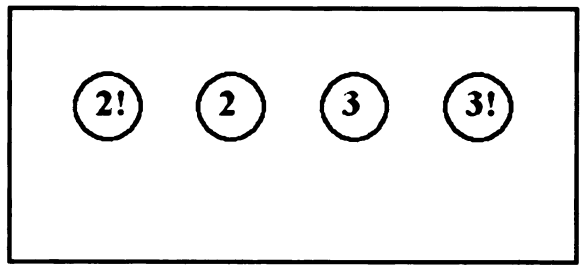

Figure 8: The response box used in the experiments.

1.1.5 Data collection

Listeners used a four-button response-box to make decisions as seen in Figure 8. Four buttons were used so that the listeners could respond to the correct interval with a confidence estimate. The buttons from left to right were 2!, 2, 3, and 3!, representing confident second interval, second interval, third interval, and confident third interval respectively. Listeners were instructed to use a confident response only if there was no uncertainty as to which interval was incoherent. If a run included more than one incorrect confident response, the run was terminated immediately and the listener was obliged to replace this run. There was no time limit for a response. After a decision was made by the listener, the next trial began following an inter-trial duration of 900 ms.

There were several reasons to introduce the confidence measure in this experiment. The first was that for a given coherence, it proved to be noticeably easier to detect incoherence in some stimuli. Thus, extra weighting was wanted for identifying obviously incoherent sounds. The second reason was that it was necessary to use the same waveforms, with a coherence of $\rho = 0.9922$, for all the listeners, and it was further desired to use the same coherence for waveforms with different bandwidths. However, some listeners were better at the task than others, as shown by the percentage of correct (P_c) responses, and some bandwidths led to better performance than others. Consequently there was a need to increase the "dynamic range" of the experiment to

prevent ceiling effects for the most successful listeners and easiest bandwidths.

The data collection procedure kept track of both the percentage of correct responses, which ignored the confidence estimate (e.g. responses of 2 and 2! were not treated differently), and a confidence adjusted score (CAS). The CAS is defined as the number of times the listener responded correctly plus the number of times that the listener was confident about the correct response. Since an individual noise-pair was heard 36 times, it was possible for a listener to get a score of 72 if the listener was able to respond correctly and confidently for all 36 presentations. In comparison with P_c, the use of CAS improved inter-listener correlation, moved p-values of t-tests to greater significance, and improved the agreement that was achievable by models of binaural processing. Further justification of this technique is given in Appendix A.

1.1.6 Listeners

The experiments in this chapter employed four male listeners, D, M, P, and W. Listeners D, M, and P were between the ages of 20–30 and had normal hearing according to standard audiometric tests and histories. Listener W was 64 and had a mild bilateral hearing loss, but only at frequencies four octaves above those used in this experiment. Listener M is the author.

1.1.7 Results

Figure 9 shows the selected P_c values and Figure 10 shows the selected CAS values for the phase set in Experiment 1. The five smallest $s_t[\Delta\Phi]$ noise-pairs are to the left of the dashed line. The five largest $s_t[\Delta\Phi]$ noise-pairs are to the right of the dashed line. The dashed line thus represents a gap of 90 unused noise-pairs. All four listeners show, in general, a greater P_c and CAS for the largest $s_t[\Delta\Phi]$ noise-pairs compared to the smallest.

The statistical test that will be used to test the hypothesis that the five noise-pairs

with the largest fluctuations in phase have detection scores that are indistinguishable from the five noise-pairs with the smallest fluctuations in phase is the two-sample t-test. If there is a difference between the largest and the smallest fluctuation noise-pairs, it will be considered significant when there is at least a 95% level of confidence (the p-value is 0.05 or less). It will be considered more significant if there is at least a 98% level of confidence (the p-value is 0.02 or less). Significant differences would mean that the value of the coherence function does not describe the detection data from the phase and level sets. The t-test will be used on both the phase and the level sets that are selected in the first three chapters of this dissertation.

In a two-sample t-test, three of four differences were significant at the 0.05 level for the P_c data. All four differences were significant at the 0.02 level for the *CAS* data. The individual p-values are shown in Table 2. Figures 9 and 10 also show that listeners tend to agree upon the difficultly of detecting incoherence in individual noise-pairs. Table 3 shows the correlations between listeners; all of them are equal to or greater than 0.391 for the P_c data and are equal to or greater than 0.769 for the *CAS* data.

Table 2: The p-values from a one-tailed t-test for 14-Hz bandwidth data. The test compared the five noise-pairs with the largest fluctuations to the five noise-pairs with the smallest fluctuations. The p-values with at least a 95% level of confidence are in bold.

P_c	Phase	Level	*CAS*	Phase	Level
Listener	Set	Set	Listener	Set	Set
D	0.095	**0.006**	D	**<0.001**	**<0.001**
M	**0.021**	**0.047**	M	**0.002**	**0.002**
P	**0.025**	**0.007**	P	**0.011**	**<0.001**
W	**0.015**	**0.014**	W	**0.001**	**<0.001**

Figure 11 shows the P_c values and Figure 12 shows the *CAS* values for the level set for Experiment 1. The figures show results similar to Figure 9 and 10 for the phase

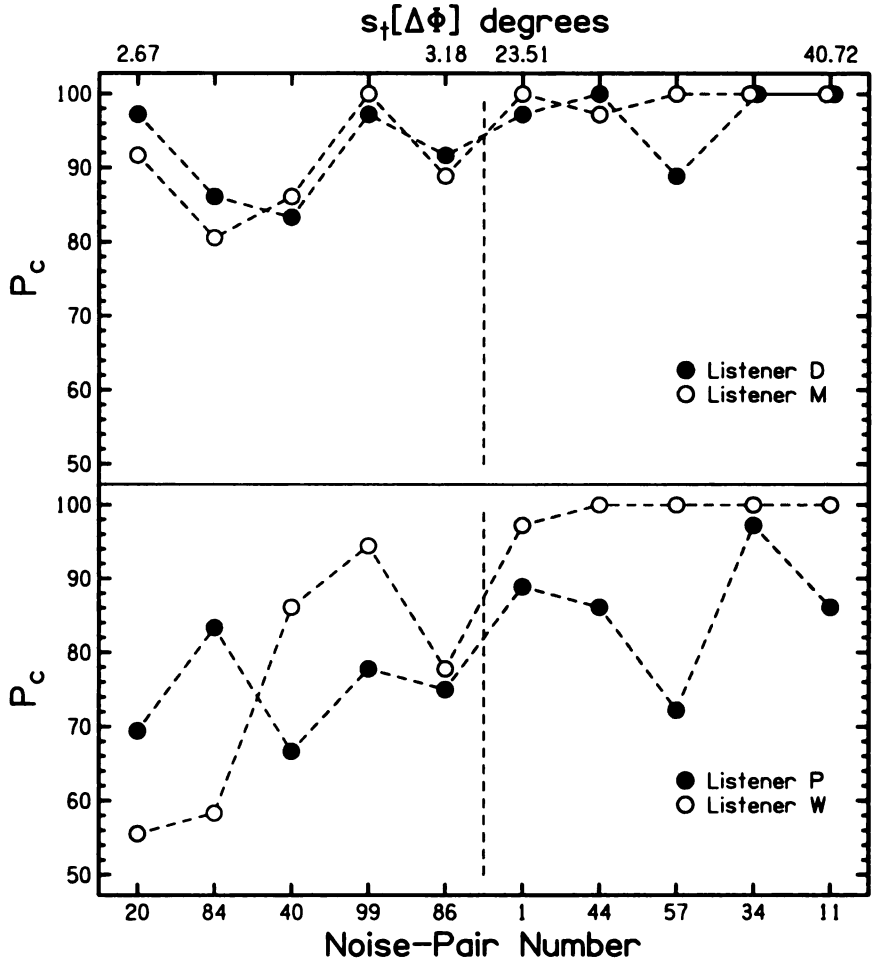

Figure 9: The percent (P_c) for five listeners for the *phase set* of Experiment 1, the 14-Hz noise-pairs. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta\Phi]$ in the collection of 100 noise-pairs. The noise-pairs are rank ordered by increasing $s_t[\Delta\Phi]$ along the horizontal axis. The vertical dashed line represents 90 unused reproducible-noise pairs. The P_c values are higher for noise-pairs with the largest $s_t[\Delta\Phi]$ than for noise-pairs with the smallest $s_t[\Delta\Phi]$ for all listeners. The plots of P_c scores vs. noise-pair serial number show a large measure of agreement among listeners.

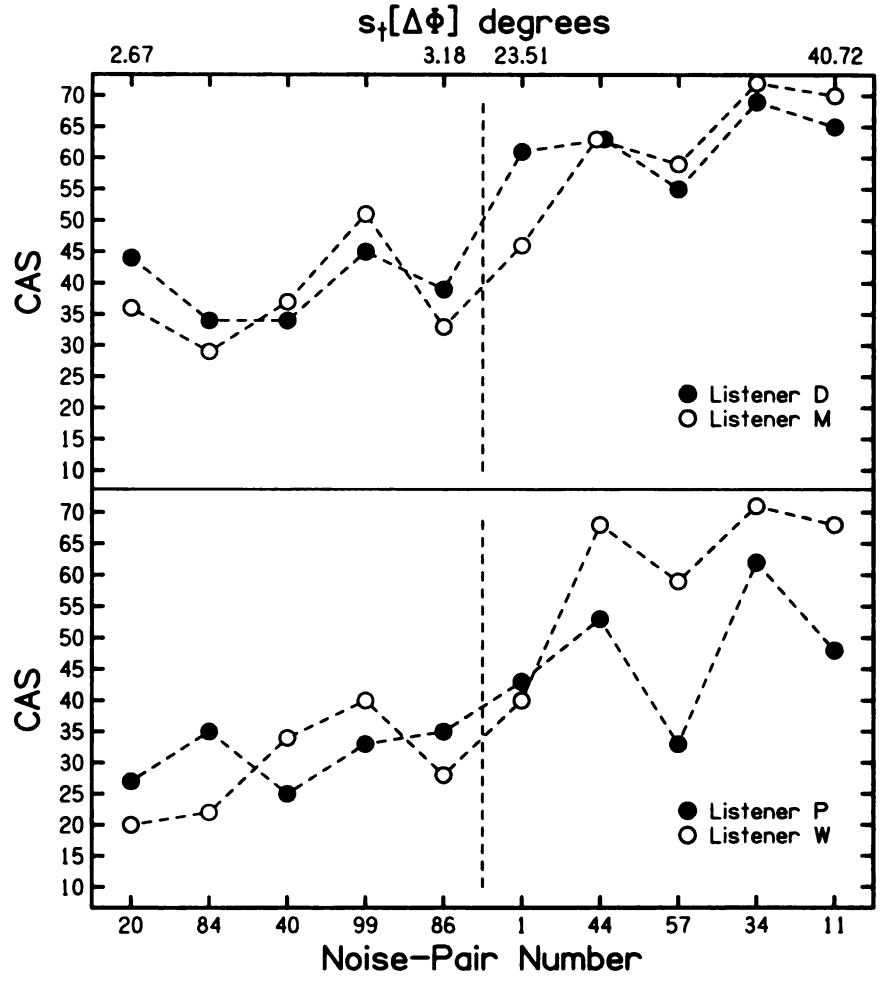

Figure 10: The confidence adjusted scores (CAS) for five listeners for the *phase set* of Experiment 1, the 14-Hz noise-pairs. The CAS values are higher for noise-pairs with the largest $s_t[\Delta\Phi]$ than for noise-pairs with the smallest $s_t[\Delta\Phi]$ for all listeners. The plots of CAS scores vs. noise-pair serial number show a large measure of agreement among listeners.

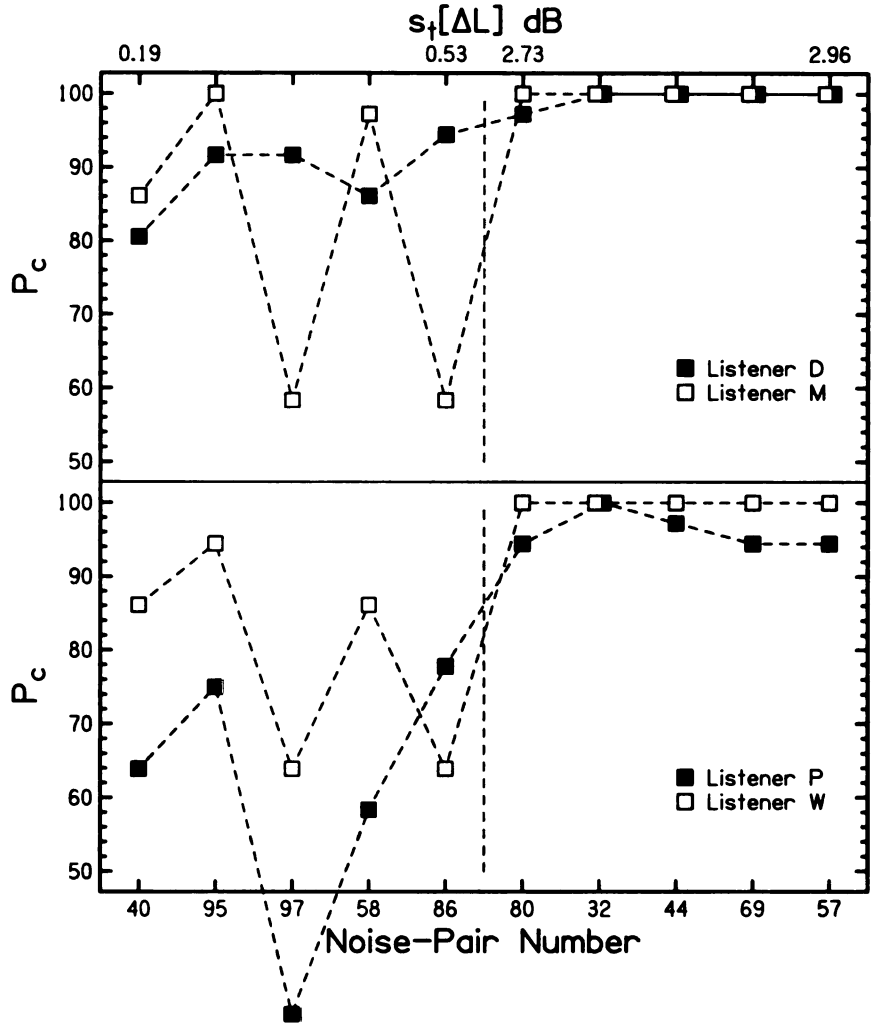

Figure 11: The percent correct (P_c) for five listeners for the *level set* of Experiment 1, the 14-Hz noise-pairs. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta L]$ in the collection of 100 noise-pairs.

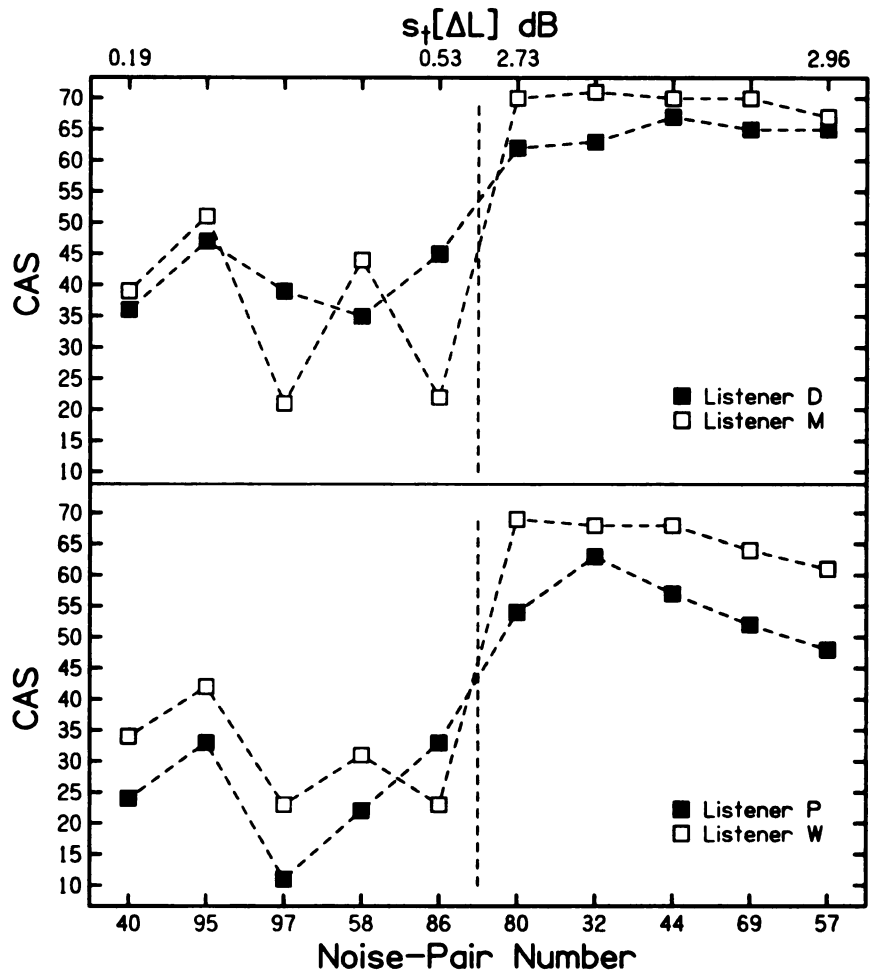

Figure 12: The confidence adjusted scores (CAS) for five listeners for the *level set* of Experiment 1, the 14-Hz noise-pairs. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta L]$ in the collection of 100 noise-pairs.

Table 3: The inter-listener correlations for 14-Hz bandwidth IPD data. It can be seen that correlations between listeners are near 1.0 for the *CAS* data. It can also be seen that using the *CAS* data increases the average correlation compared to the P_c data.

P_c	D	M	P	W
D	1	0.732	0.614	0.391
M	–	1	0.421	0.796
P	–	–	1	0.429
W	–	–	–	1
Average	0.564			

CAS	D	M	P	W
D	1	0.902	0.846	0.868
M	–	1	0.769	0.970
P	–	–	1	0.787
W	–	–	–	1
Average	0.857			

set. As shown by Table 2, all the p-values are significant at the 0.05 level for the P_c data and at the 0.02 level for the *CAS* data. Table 4 shows correlations between the individual listeners are all greater than or equal to 0.291 for the P_c data and increase to greater than or equal to 0.855 for the *CAS* data.

1.1.8 Discussion

Experiment 1 shows that the peak of the cross-correlation function is not an adequate predictor of the detectability of incoherence for narrowband noise. Instead, the fluctuations in interaural phase and interaural level clearly play a role in incoherence detection. This can explicitly be seen in the p-values of the t-tests for the phase and level sets chosen in this experiment.

The large inter-listener correlations of this experiment indicates that listeners agree in detail about the kinds of fluctuations that are easy or hard to detect. This correlation can be seen graphically in Figures 9–12 or numerically in Tables 3 and 4.

Table 4: The inter-listener correlations for 14-Hz bandwidth ILD data. Correlations are comparable to those for the IPD data. Once again, inter-listener correlations are higher for the *CAS* data compared to the P_c data.

P_c	D	M	P	W
D	1	0.291	0.707	0.447
M	–	1	0.662	0.967
P	–	–	1	0.774
W	–	–	–	1
Average	0.641			

CAS	D	M	P	W
D	1	0.855	0.931	0.925
M	–	1	0.885	0.974
P	–	–	1	0.933
W	–	–	–	1
Average	0.917			

It general, p-values moved to greater significance and inter-listener correlations increased when using the *CAS* values over the P_c values in this experiment. Although it was probably not necessary to introduce the *CAS* to show conclusive results that the cross-correlation function is not an adequate predictor of the detectability of incoherence for narrowband noise, it was predicted that the *CAS* would be necessary for larger bandwidths.

Lastly, it should be reported that some noise-pairs led to values of percent correct that were well below chance. Through informal listening experiments, it was found that these noise-pairs had very little roughness or action in the envelope of their waveforms. It seems possible that when the fluctuations of IPD and ILD are barely detectable, listeners may sometimes mistake roughness or action for incoherence. They would then be reluctant to say that a particularly smooth sounding noise-pair is incoherent, which could lead to a P_c less than 50%. Figure 13 shows the left and right channels for three sample noise-pairs. It appears that the envelope for

noise-pair B is smoother (more gradually sloped) or has less action than the other two noise-pairs. Listeners detected the incoherence in noise-pair B with a P_c of only 27%. This can be compared to noise-pair A with $P_c = 96\%$ and noise-pair C with $P_c = 71\%$. This idea of using monaural cues during incoherence detection will be more thoroughly explored in Experiment 5.

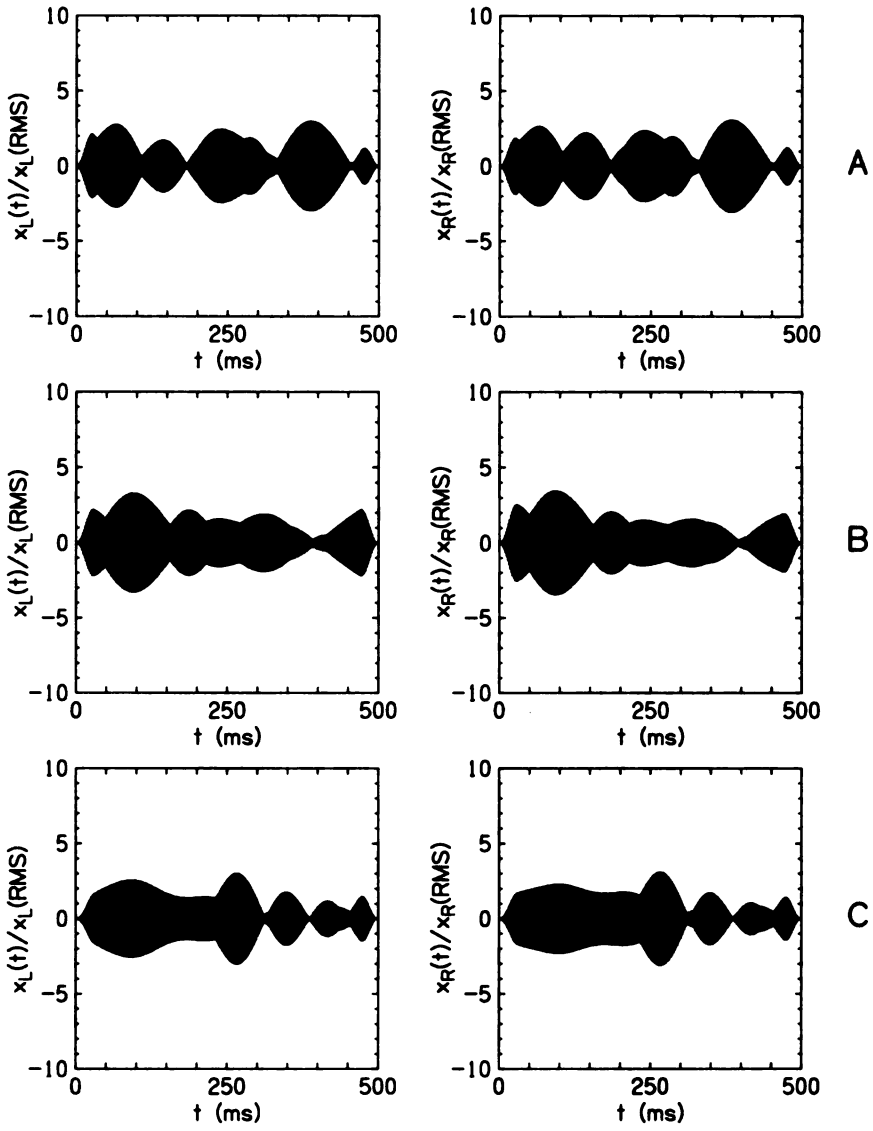

Figure 13: The left and right channels for three sample noise-pairs A, B, and C with ρ = 0.9922 and a bandwidth of 14 Hz. Each channel was normalized by the RMS value of the channel. It appears that B is smoother or has less action in it compared to the other noise-pairs, which led to the hypothesis that listeners may be using monaural cues during incoherence detection for noise-pairs with small interaural fluctuations.

1.2 EXPERIMENT 2: CRITICAL BANDWIDTH

Experiment 2 was identical to Experiment 1 except that the bandwidth was increased from 14 Hz to 108 Hz, near a critical bandwidth at 500 Hz. The value of the coherence remained fixed at 0.9922.

1.2.1 Method

As for Experiment 1, the geometric mean frequency was 500 Hz and the spectral spacing was 2 Hz. The bandwidth was increased to 108 Hz. As for Experiment 1, the spectral edges were 5 Hz wide. Therefore, components between 444 and 449 Hz and between 555 and 560 Hz were given a raised-cosine edge; components between 449 and 555 Hz had unity amplitude. The variation of the interaural parameters as a function of time can be seen in Figure 14. Notice that the fluctuations are approximately eight times faster when compared with Figure 5 because the bandwidth is approximately eight times larger. Note that with a larger bandwidth of 108 Hz, the conversion of IPD to ITD is not necessarily as straightforward as in the 14-Hz bandwidth case. However, given an IPD, the value of the ITDs is uncertain by only approximately 10% for this bandwidth.

Another collection of 100 noise-pairs was used in this experiment. A phase set and level set were chosen from the collection, as in Experiment 1. Figure 15 shows $s_t[\Delta\Phi]$ versus $s_t[\Delta L]$ for the first collection of 108-Hz bandwidth noise-pairs. The black dots represent the noise-pairs with 14-Hz bandwidth from Experiment 1 for comparison. The means of the distributions of fluctuations increased slightly when compared to the the 14-Hz bandwidth waveforms, but the standard deviations over the noise-pair ensemble of the phase fluctuations decreased from 6.66 degrees to 3.72 degrees. The standard deviation of the level fluctuations decreased from 0.68 dB to 0.40 dB. The correlation between $s_t[\Delta\Phi]$ and $s_t[\Delta L]$, evaluated over the ensembles of 100 waveforms, was essentially the same, 0.73 compared to 0.70. Appendix B shows

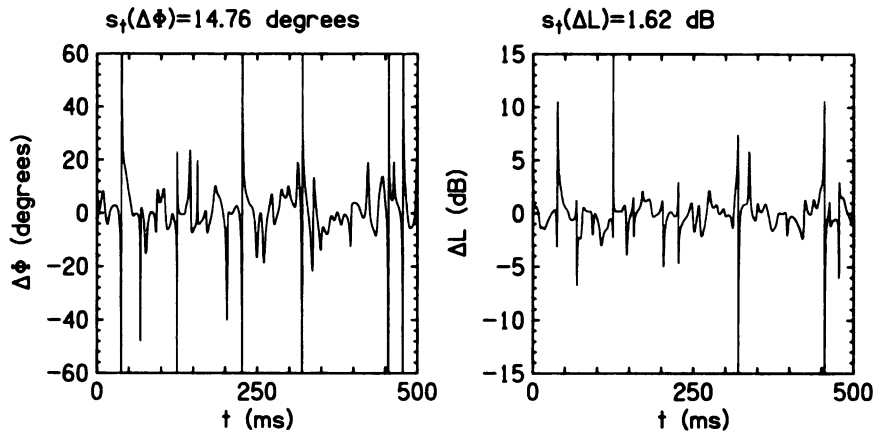

Figure 14: The IPD and ILD plotted as a function of time for an arbitrarily chosen 108-Hz bandwidth noise-pair. Above the plot are the fluctuation values of the IPD and ILD. Note that when compared with Figure 5, the fluctuations are about eight times faster.

that a value of 0.73 is in line with expectation based on statistics for 5000 noise-pairs.

1.2.2 Results

Figures 16 and 17 show the P_c and *CAS* values for the phase set for the 108-Hz noise-pairs of Experiment 2. A comparison between the phase set of this experiment and the phase set of Experiment 1 show that listeners are not nearly as good at distinguishing incoherence from coherence at this larger bandwidth. Similar results are seen for the level set in Figures 18 and 19.

The Figures 16 and 18 show Listeners D and M near the ceiling of P_c for many of the ten noise-pairs and always above 75%. However, Listeners D and M do not reach the ceiling for *CAS* as seen in Figure 17 and 19. Listeners P and W are not so dramatically close to ceiling for P_c or *CAS*.

Table 5 shows that only one of the four t-tests from the phase sets P_c values led to differences that were significant at the 0.05 level between the results for the five largest fluctuations and the results for the five smallest fluctuations. This increased to three of four significant p-values when using the *CAS* values. One of the four

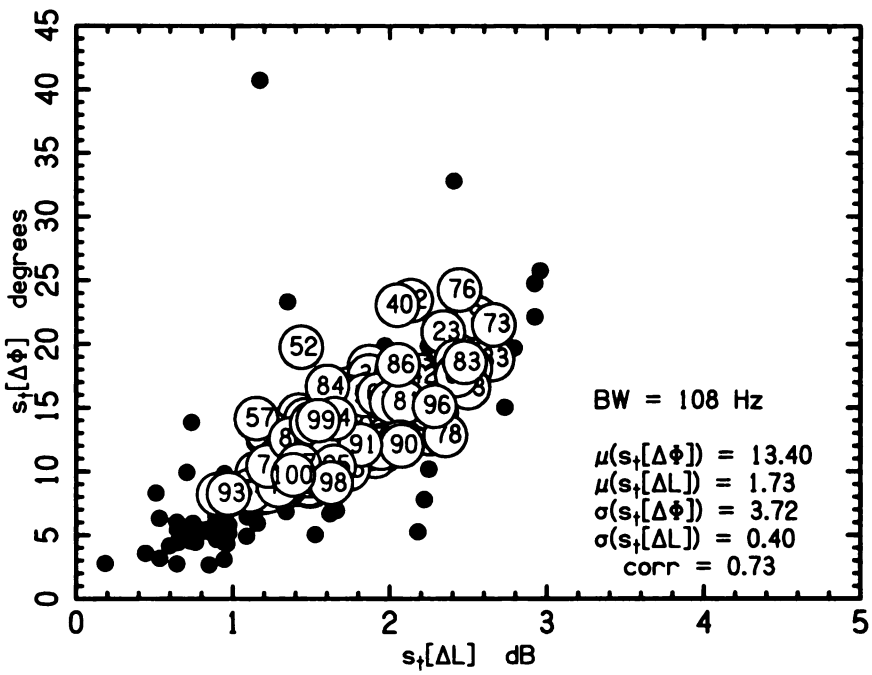

Figure 15: The fluctuations of IPD versus fluctuations of ILD for the collection of 100 reproducible noise-pairs having a 108-Hz bandwidth, as used in Experiments 2. Each noise-pair is labeled by a serial number indicating only the order of creation. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. The black dots represent the 14-Hz bandwidth noise-pairs from the collection in Experiment 1.

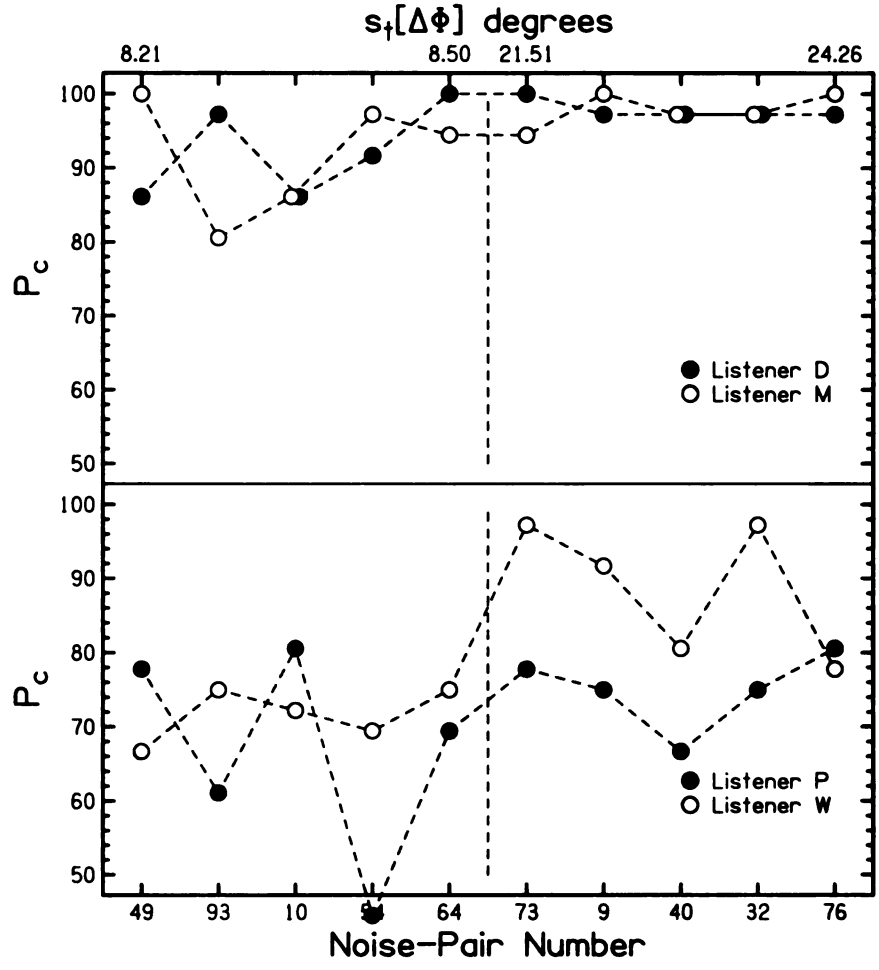

Figure 16: The P_c results for the phase set of Experiment 2 with a 108-Hz bandwidth. The phase set was constructed by the same method as was used in Experiment 1. When compared to the 14-Hz bandwidth phase set in Figure 9, there is a much less drastic difference between noise-pairs with the largest and smallest fluctuations of IPD.

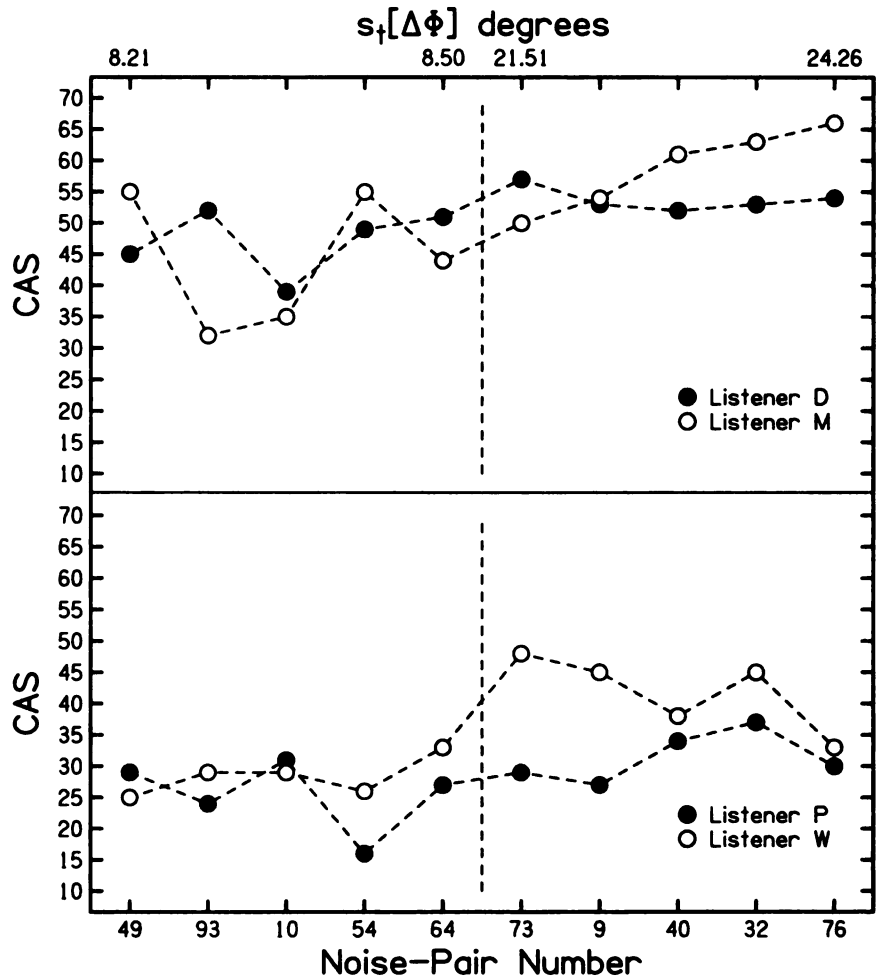

Figure 17: The *CAS* results for the phase set of Experiment 2 with a 108-Hz bandwidth. The phase set was constructed by the same method as was used in Experiment 1. When compared to the 14-Hz bandwidth phase set in Figure 10, there is a much less drastic difference between noise-pairs with the largest and smallest fluctuations of IPD.

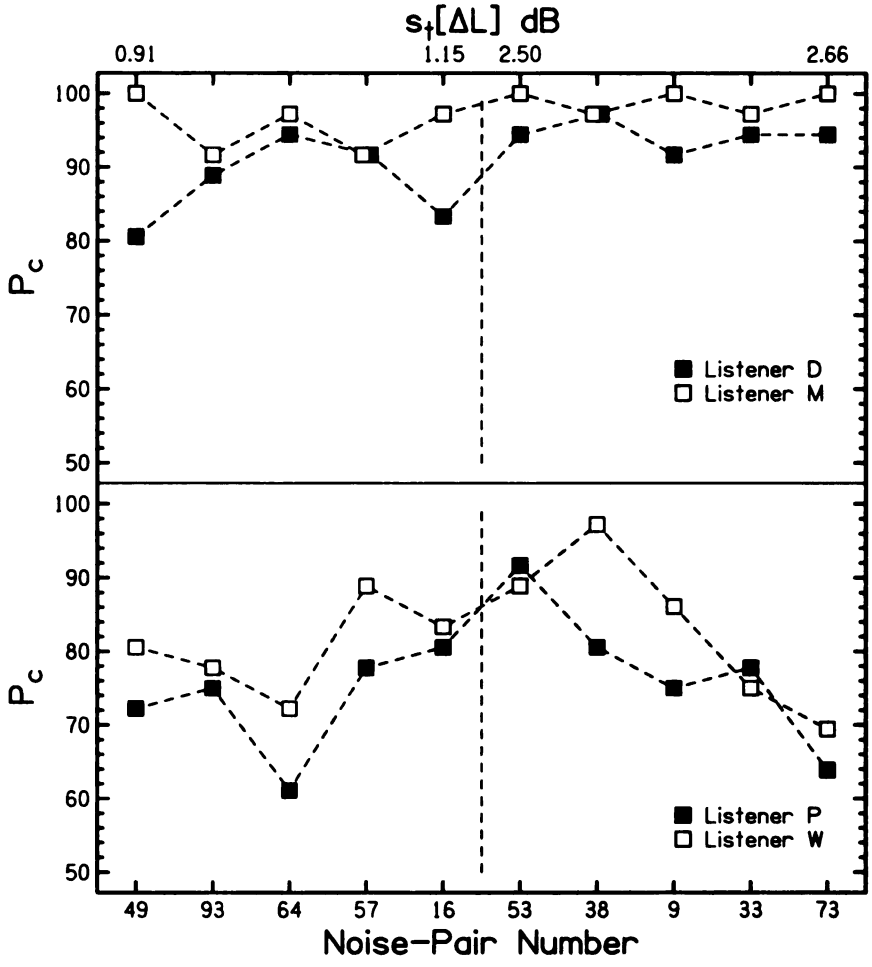

Figure 18: The P_c results for the level set of Experiment 2 with a 108-Hz bandwidth. The phase set was constructed by the same method as was used in Experiment 1. When compared to the 14-Hz bandwidth phase set in Figure 11, there is a much less drastic difference between noise-pairs with the largest and smallest fluctuations of ILD.

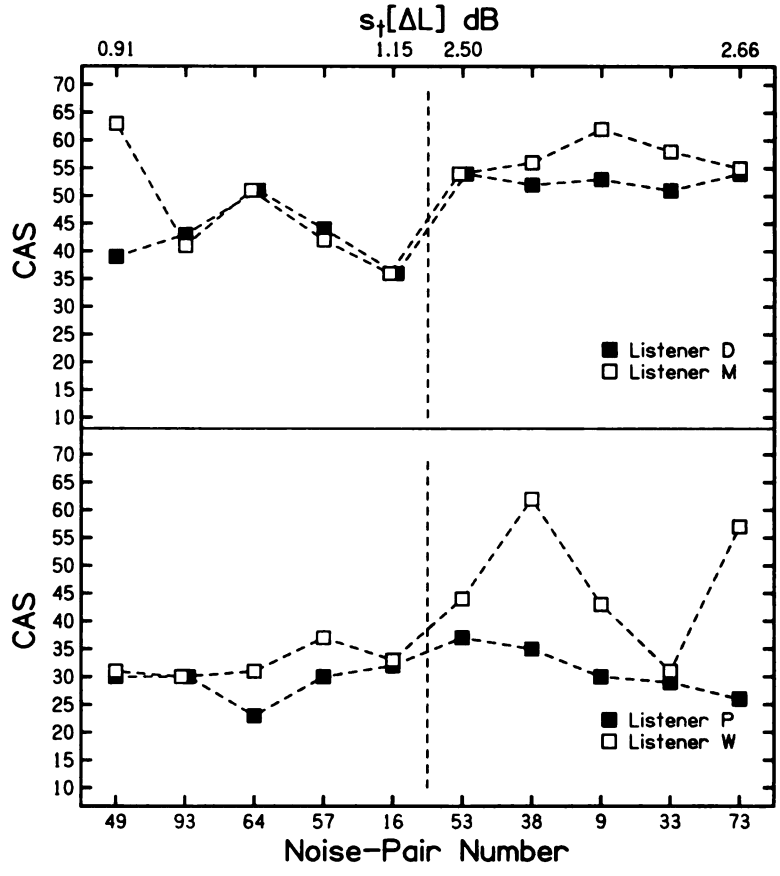

Figure 19: The *CAS* results for the level set of Experiment 2 with a 108-Hz bandwidth. The phase set was constructed by the same method as was used in Experiment 1. When compared to the 14-Hz bandwidth phase set in Figure 12, there is a much less drastic difference between noise-pairs with the largest and smallest fluctuations of IPD.

t-tests from the level sets for the P_c values led to differences that were significant at the 0.05 level. This increased to three of four significant p-values when using the *CAS* values.

Table 5: The p-values from a one-tailed t-test for 108-Hz bandwidth data.

P_c	Phase	Level	*CAS*	Phase	Level
Listener	Set	Set	Listener	Set	Set
D	0.061	**0.029**	D	**0.024**	**0.007**
M	0.085	0.060	M	**0.019**	**0.047**
P	0.248	0.226	P	0.187	0.187
W	**0.005**	0.069	W	**0.002**	**0.026**

Table 6: Inter-listener correlations for 108-Hz bandwidth phase set.

P_c	D	M	P	W
D	1	0.071	-0.044	0.634
M	–	1	0.148	0.202
P	–	–	1	0.337
W	–	–	–	1
Average	0.225			

CAS	D	M	P	W
D	1	0.422	0.053	0.668
M	–	1	0.312	0.328
P	–	–	1	0.492
W	–	–	–	1
Average	0.379			

Further, as shown in Tables 6 and 7, the correlation between listeners was smaller for the wider bandwidth for all listener pairs. The correlation between listeners was 0.225 on average for the phase set for the P_c data. This increased to 0.379 for the *CAS* data. Likewise, there was an increase in inter-listener correlation for the level set, from 0.117 to 0.278, when using *CAS* data over the P_c data.

Table 7: Inter-listener correlations for 108-Hz bandwidth level set.

P_c	D	M	P	W
D	1	0.000	0.007	0.093
M	–	1	–0.031	–0.093
P	–	–	1	0.724
W	–	–	–	1
Average	0.117			

CAS	D	M	P	W
D	1	0.543	–0.039	0.583
M	–	1	–0.026	0.298
P	–	–	1	0.309
W	–	–	–	1
Average	0.278			

1.2.3 Discussion

The standard deviation of the fluctuations of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ computed across the 100 different waveforms of the ensemble decreased by more than 40% when the bandwidth was increased from 14 Hz to 108 Hz. Since these fluctuations were found to correlate with incoherence detection from Experiment 1, it was not surprising to find that there was less variation in the listeners' ability to detect incoherence at the wider bandwidth.

The bandwidth of 108 Hz may be of special interest because this bandwidth approximately corresponds to a critical bandwidth at 500 Hz, and critical band noise has often been used in binaural experiments. For instance, Koehnke *et al.* (1986) used a noise with a bandwidth of 114 Hz centered on 500 Hz, and Evilsizer *et al.* (2002) used a bandwidth of 100 Hz. It was found that the tests comparing detection performance (CAS) with the size of the phase and level fluctuations led to a significant difference at the 0.05 level on six of eight of the t-tests. This can be compared to all of the tests being significant at the 0.02 level in Experiment 1. The P_c results had

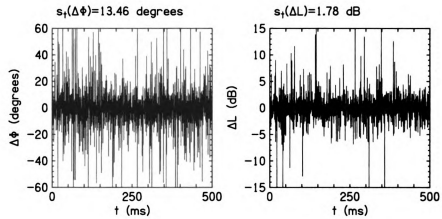

Figure 20: The IPD and ILD plotted as a function of time for an arbitrary 2396-Hz bandwidth noise-pair. Above the plot are the fluctuation values of the IPD and ILD.

only two significant p-values at the 0.05 level compared to seven significant p-values in Experiment 1. It was expected that if the bandwidth were further increased the variation in fluctuations would continue to decrease and the incoherence detection performance would be approximately the same for all the waveforms in the ensemble. In that case, the ability to detect incoherence would be only a function of the incoherence measure itself. That expectation led to Experiment 3.

1.3 EXPERIMENT 3: WIDE BANDWIDTH

1.3.1 Method

Experiment 3 was identical to Experiments 1 and 2 except that the bandwidth was increased to 2394 Hz. The geometric mean frequency of 500 Hz and the coherence of 0.9922 remained the same. Spectral components between 105 and 2495 Hz had unity amplitude and components in the ranges 100–105 and 2495–2500 Hz were shaped by the raised-cosine window. The variation of the interaural parameters can be seen in Figure 20.

Figure 21 shows the distribution of interaural phase and level fluctuations for one-hundred 2394-Hz bandwidth noises. The corresponding values for Experiments 1 and

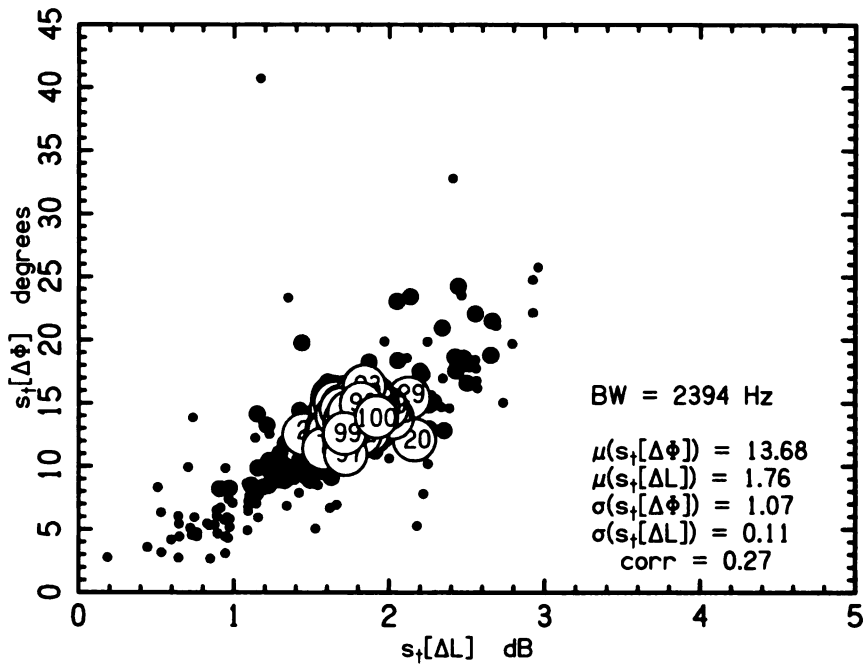

Figure 21: Fluctuations of IPD versus fluctuations of ILD for the collection of 100 reproducible noise-pairs having a 2394-Hz bandwidth, as used in Experiments 3. Each noise-pair is labeled by a serial number indicating only the order of creation. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. The small black dots represent the 14-Hz bandwidth noise-pairs from the collection in Experiment 1. The large black dots represent the 108-Hz bandwidth noise-pairs from the collection in Experiment 2.

2 are shown by the small black dots. The mean of the distribution remained about the same as in Experiments 2 in that the ensemble average phase fluctuation was about 13 degrees and the level fluctuation was about 1.7 dB. However, the standard deviations of the phase and level fluctuations over the collection of 100 noise-pairs decreased dramatically when the bandwidth was increased to 2394 Hz, as shown by the σ values in Figure 21, respectively 1.07 degrees and 0.11 dB. The correlation between level and phase fluctuations decreased to 0.27.

1.3.2 Results

The results of the incoherence detection experiments, expressed as P_c and CAS values, are shown in Figures 22–25 for the phase and level sets. Listeners D and M, the

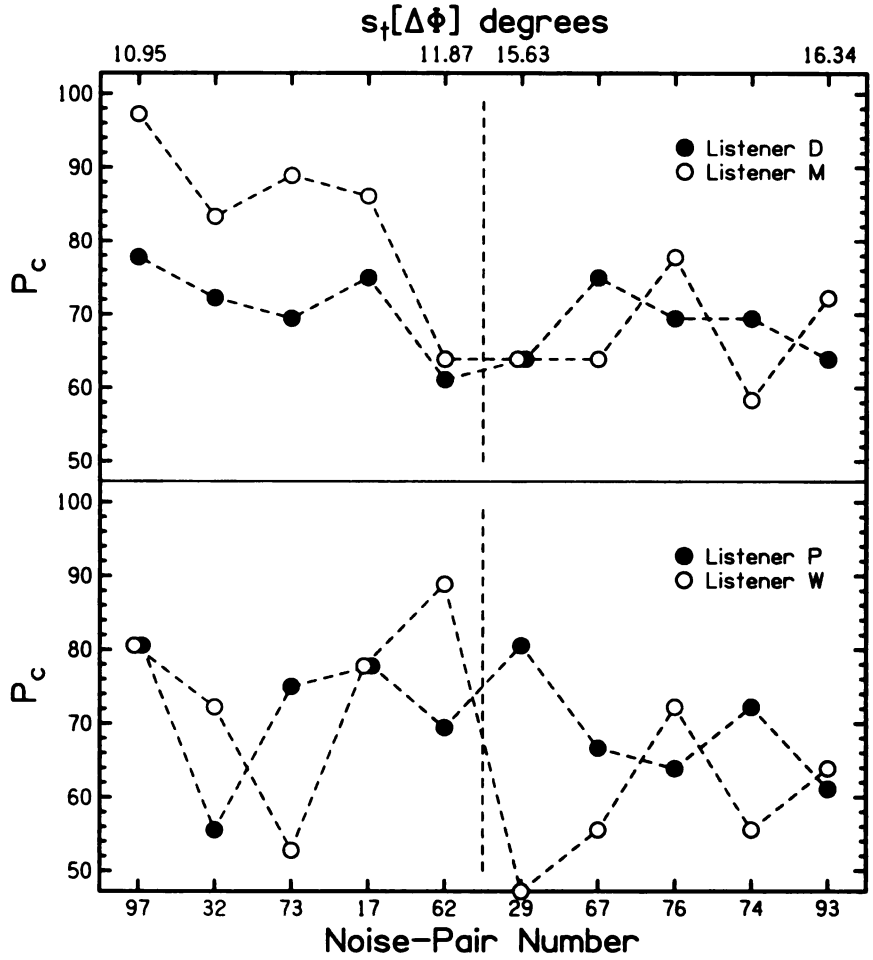

Figure 22: The P_c data for the 2394-Hz phase set.

listeners with consistently the highest values of P_c in the previous experiments, are near the ceiling for only a few noise-pairs with a 2394-Hz bandwidth.

With four listeners and two sets there were eight possible significance tests, and Table 8 shows that only one of them led to a significant difference at the 0.05 level for P_c; none for *CAS*. Further, Table 9 shows inter-listener correlations comparable to those in Experiment 2. However, Table 10 shows an overall negative inter-listener correlation, mainly because of Listener W.

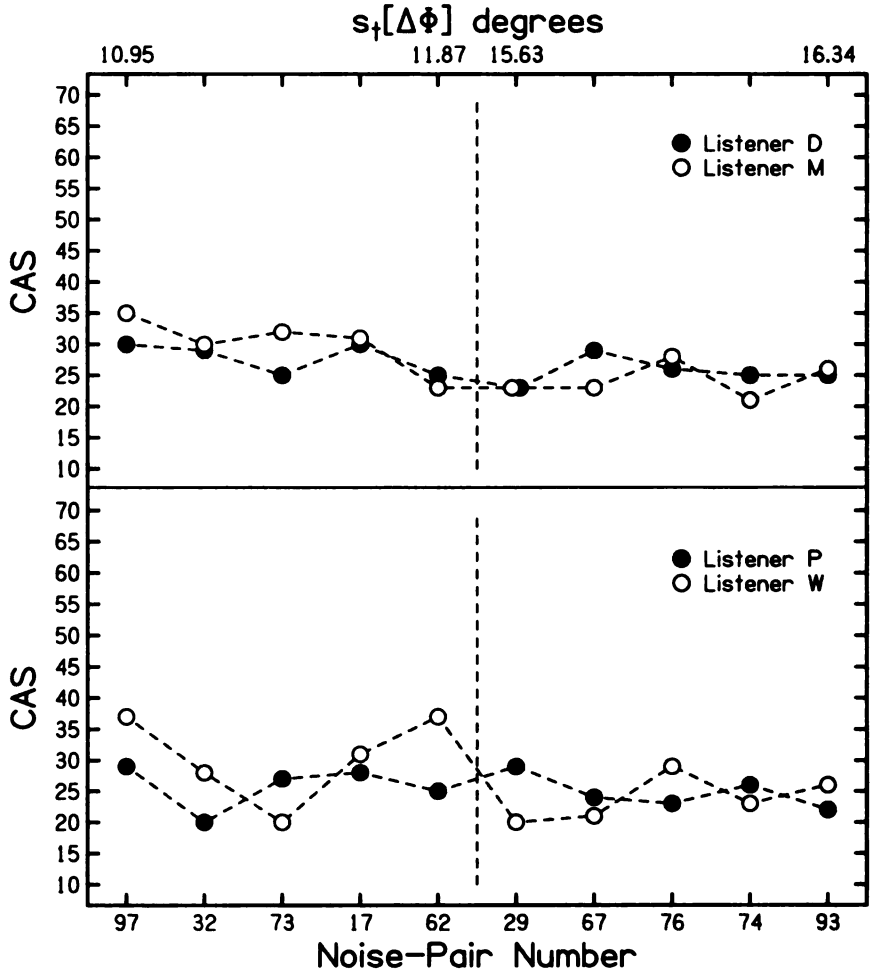

Figure 23: The *CAS* data for the 2394-Hz phase set.

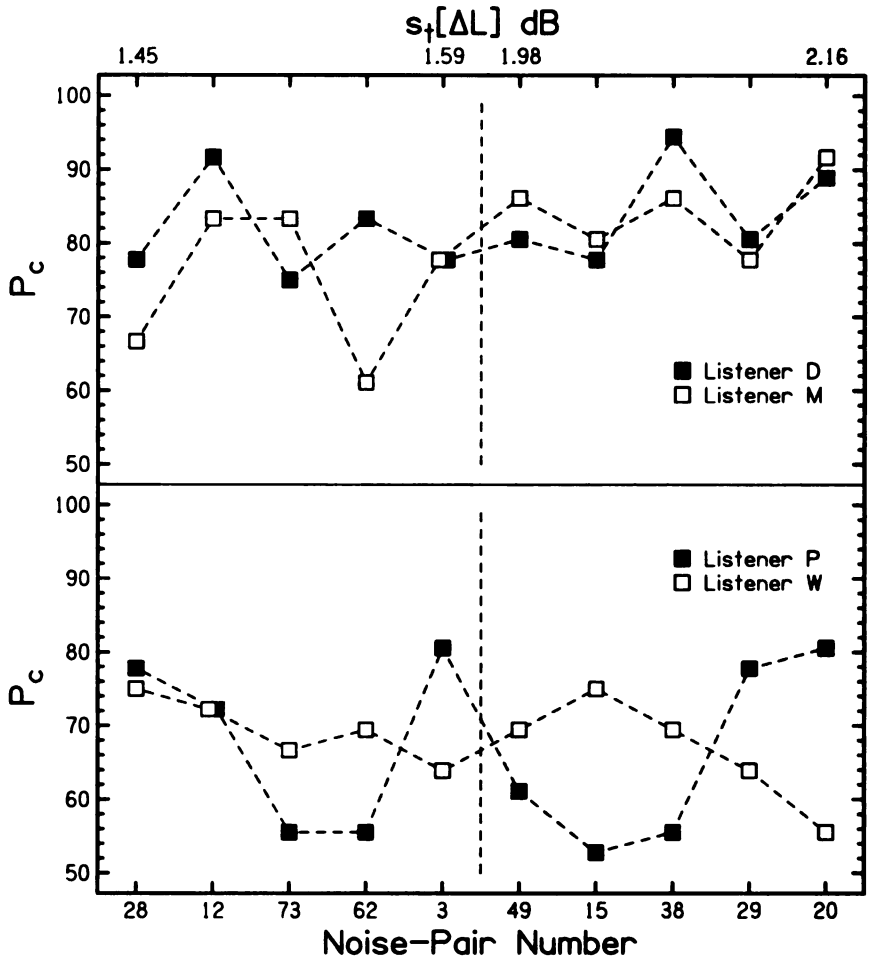

Figure 24: The P_c data for the 2394-Hz level set.

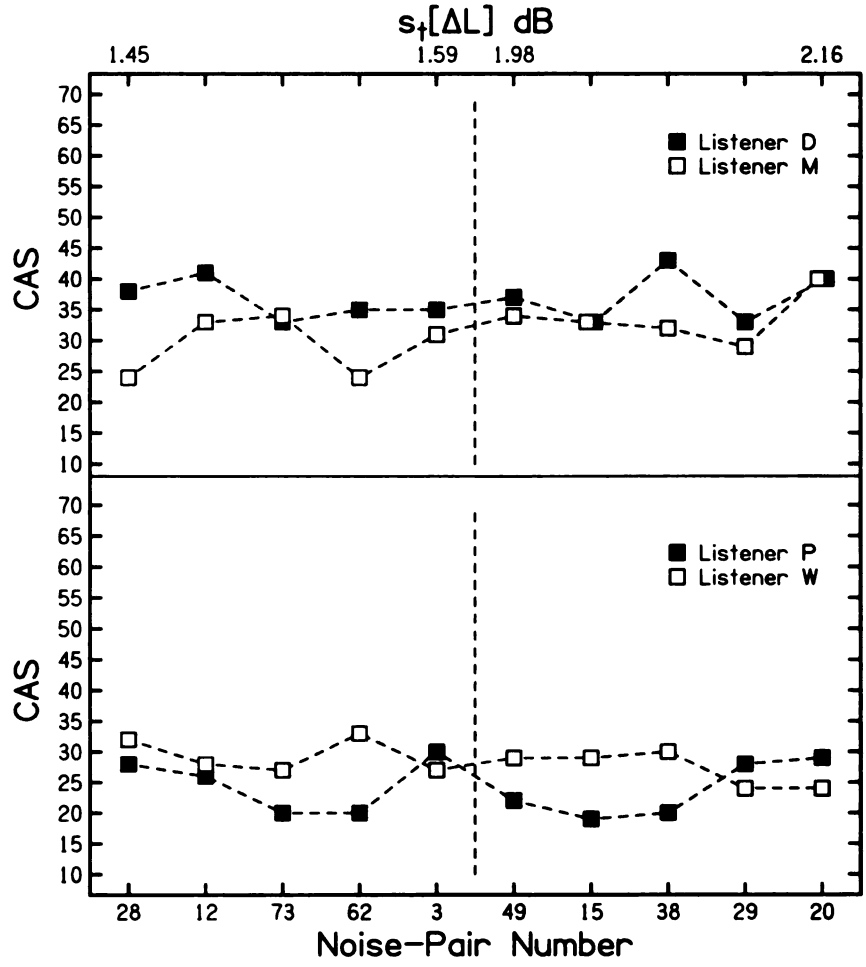

Figure 25: The *CAS* data for the 2394-Hz level set.

Table 8: The p-values from a one-tailed t-test for 108-Hz bandwidth data.

P_c	Phase	Level
Listener	Set	Set
D	0.772	0.230
M	0.981	**0.049**
P	0.683	0.633
W	0.964	0.753

CAS	Phase	Level
Listener	Set	Set
D	0.907	0.375
M	0.981	0.080
P	0.683	0.654
W	0.947	0.866

Table 9: Inter-listener correlations for 2394-Hz bandwidth phase set.

P_c	D	M	P	W
D	1	0.585	0.161	0.103
M	–	1	0.165	0.363
P	–	–	1	–0.125
W	–	–	–	1
Average	0.208			

CAS	D	M	P	W
D	1	0.572	–0.059	0.426
M	–	1	0.165	0.377
P	–	–	1	0.031
W	–	–	–	1
Average	0.252			

1.3.3 Discussion

Experiments 1, 2, and 3 clearly show trends in the nature of interaural incoherence detection. For narrow bands, detection depends on the details of interaural fluctuations, and it is not possible to predict detection performance if one knows only the value of coherence. As the bandwidth increases, the coherence becomes a better predictor of detection performance. The wideband limit, where the coherence statistic itself becomes adequate to predict performance for all stimuli, was apparently reached for the bandwidth of 2394 Hz because no systematic differences appeared between those

Table 10: Inter-listener correlations for 2394-Hz bandwidth level set.

P_c	D	M	P	W
D	1	0.364	0.024	–0.153
M	–	1	0.042	–0.439
P	–	–	1	–0.474
W	–	–	–	1
Average	–0.106			

CAS	D	M	P	W
D	1	0.253	0.136	0.130
M	–	1	0.043	–0.683
P	–	–	1	–0.488
W	–	–	–	1
Average	–0.105			

noise-pairs with the largest fluctuations and those noise-pairs with the smallest.

The above interpretation, however, is not flawless. The noise-pairs for Experiment 3 were chosen on the basis of fluctuations computed for the entire band of noise, as presented to the listeners. However, it is possible that listeners do not perform the incoherence detection experiment by listening to the entire band. Instead, they may listen to a narrower portion of the band - perhaps one critical band width wide. If so, then the choices of noise-pairs were inappropriate for the band actually used, and further, one would have no way of knowing *a priori* which band to examine when selecting stimuli based on large and small fluctuations. If this were true, then the reasoning by which it was concluded that the wideband limit is reached in Experiment 3 would be circular reasoning. One piece of evidence in favor of this view is found in the MLD experiments by van de Par and Kohlrausch (1999) which suggested that listeners detect an $S\pi$ sine in noise by listening to a critical band around the tone.

It is believed that the above criticism is incorrect and that in a wide band incoherence detection task - perhaps unlike a wideband MLD task - listeners do not attend to only a portion of the band. The evidence for that belief is that the listeners

performed much less well for the wide band in Experiment 3 compared to the 108-Hz band of Experiment 2. Comparing *CAS* values in Figures 17 and 23 (phase sets) and comparing values in Figures 19 and 25 (level sets) show that the most successful listeners, D and M, had higher scores even for the most difficult (lowest values of P_c) noise-pairs at 108 Hz than for *any* noise-pairs with wide bands. Listeners P and W also scored consistently better at 108-Hz for the level sets. If listeners were able to take advantage of the slower and potentially larger fluctuations in a critical-band portion of the wide band one would have expected that some listener would have scored well for some one of the noise-pairs, contrary to the results of Experiment 3. Consequently it is believed that Experiment 3 reached a wideband limit wherein the relevant interaural fluctuations are completely characterized by the coherence value.

1.4 EXPERIMENT 4: THE ROLE OF BANDWIDTH

Experiments 1, 2, and 3 demonstrated that, as the bandwidth increases, two effects occur. First, the ranges of fluctuations in IPD and ILD become narrower. Second, the ability of listeners to detect incoherence depends less on the individual noise-pairs and is better determined by the value of coherence itself. According to the hypothesis of this chapter, the second effect is the direct result of the first, and the main effect of a variation in bandwidth is to alter the distributions of interaural variances. Experiment 4 was designed to test this idea.

1.4.1 Method

To test the hypothesis, a subset of noise-pairs was assembled from a new collection of 1000 more pairs with a bandwidth of 14 Hz to make a "matched set" whose members were selected to best match the fluctuations in the noise-pairs from Experiment 2, which had a bandwidth of 108 Hz. Experiment 2 included 20 noise-pairs, ten for the phase set and ten for the level set, as determined by the five largest and five smallest

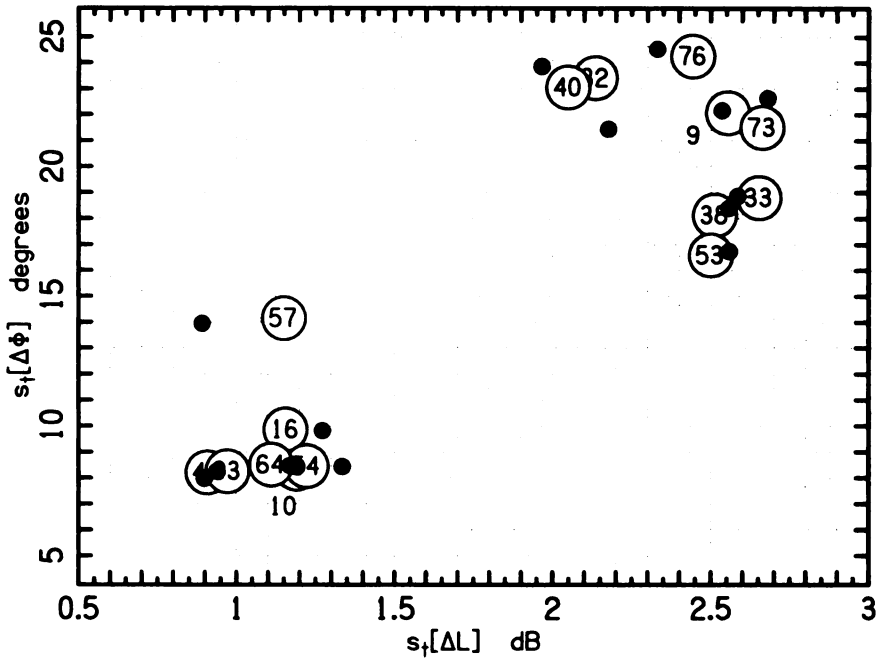

Figure 26: The matched set for Experiment 4. The 17 noise-pairs from phase set 1 and level set 1 in Experiment 2 were matched in fluctuations of IPD and fluctuations of ILD by the closest noise-pairs from Experiment 1, which are shown as black dots.

fluctuations. However, phase and level fluctuations tend to be correlated and five of the noise-pairs were common to the phase and level sets. Therefore, there were only 15 different noise-pairs in Experiment 2. For each of these 108-Hz bandwidth noise-pairs a 14-Hz bandwidth noise-pair was selected that best matched the fluctuations in phase and level. The selection is illustrated by the 15 open and filled circles in Figure 26. Phase and level sets using the matched noise-pairs formed the stimuli for Experiment 4, which was otherwise identical to the other experiments of this chapter.

1.4.2 Results

According to the working hypothesis, the detection scores for the matched 14-Hz sets from Experiment 4 ought to be identical to the detection scores in the 108-Hz sets from Experiment 2. The results of the comparison are shown in Figures 27–30 for the phase and level sets, and for the P_c and CAS values. The phase set data in

Figures 27 and 28 show that the values of P_c and CAS are comparable for the two bandwidths when the IPD fluctuations are matched for Listeners D and M. However, for Listeners P and W performance is better for the 14-Hz bandwidth than for the 108-Hz bandwidth.

The level set data in Figures 29 and 30 show comparable P_c and CAS values for the two bandwidths for all the listeners, though Listeners P and W still tend to show better performance at the smaller bandwidth - better on 15 of 20 possible comparisons for P_c and 14 of 20 possible comparisons for CAS. Therefore, the raw data, shown in Figures 27–30, offer modest support for the hypothesis that bandwidth should be unimportant if the sizes of fluctuations are matched.

The hypothesis can be further tested by examining the relative detectability of the noise-pairs with the largest interaural fluctuations vs. the noise-pairs with the smallest interaural fluctuations. These are, respectively, to the right and to the left of the vertical dashed line in Figures 27–30. A t-test of the hypothesis that P_c and CAS scores are higher for the five noise-pairs on the right led to the p-values in Table 11.

There, it can be seen that two of eight p-values are significant at the 0.05 level for the 14-Hz matched sets P_c values, and that two of eight p-values are significant at that level for the targeted 108-Hz sets. It can also be seen that three of eight p-values are significant at the 0.05 level for the 14-Hz matched sets CAS values, and that six of eight p-values are significant at that level for the targeted 108-Hz sets. By comparison, seven of eight p-values were significant at the 0.05 level for the 14-Hz bandwidth P_c data and all eight p-values were significant at the 0.02 level for the 14-Hz bandwidth CAS data in Experiment 1. Only one of the p-values were significant even at the 0.05 level for the wide bandwidth P_c data and none of the p-values were significant even at the 0.05 level for the wide bandwidth CAS data in Experiment 3. Thus, the matched phase and level sets appear to have approximately matched the

relative difference in performance between the noise-pairs in the Experiment 2 sets, consistent with the hypothesis that the size of interaural fluctuations determines the detection of incoherence.

However, there are differences between the results of Experiments 2 and 4. As noted above, there is a tendency for listeners P and W to score better on the 14-Hz bandwidth sets (Experiment 4). More impressive, a comparison of the inter-listener correlations in Tables 12 and 13 show that all twelve of the correlations are higher for the 14-Hz matched sets than for the correlations in Tables 7 and 9 for the 108-Hz bandwidth sets. Averaged over listener pairs and over phase and level sets, the inter-listener correlation was 0.601 for the 14-Hz bandwidth and only 0.171 for the 108-Hz bandwidth for the P_c results. Likewise, the averaged inter-listener correlation was 0.700 for the 14-Hz bandwidth and only 0.329 for the 108-Hz bandwidth for the CAS results.

Table 11: The p-values from a one-tailed t-test for matched phase and level data.

P_c	Phase	Level	CAS	Phase	Level
Listener	Set	Set	Listener	Set	Set
D	0.428	0.111	D	0.103	0.077
M	0.408	0.421	M	0.384	0.315
P	**0.025**	**0.050**	P	**0.006**	**0.038**
W	0.163	0.308	W	**0.042**	0.117

1.4.3 Discussion

Experiment 4 attempted to construct a set of noise-pairs with a 14-Hz bandwidth that would lead to the same patterns of detection performance that had been seen in Experiment 2, which used noise-pairs with a 108-Hz bandwidth. This was done by best matching the values of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ of the Experiment 2 noise-pairs

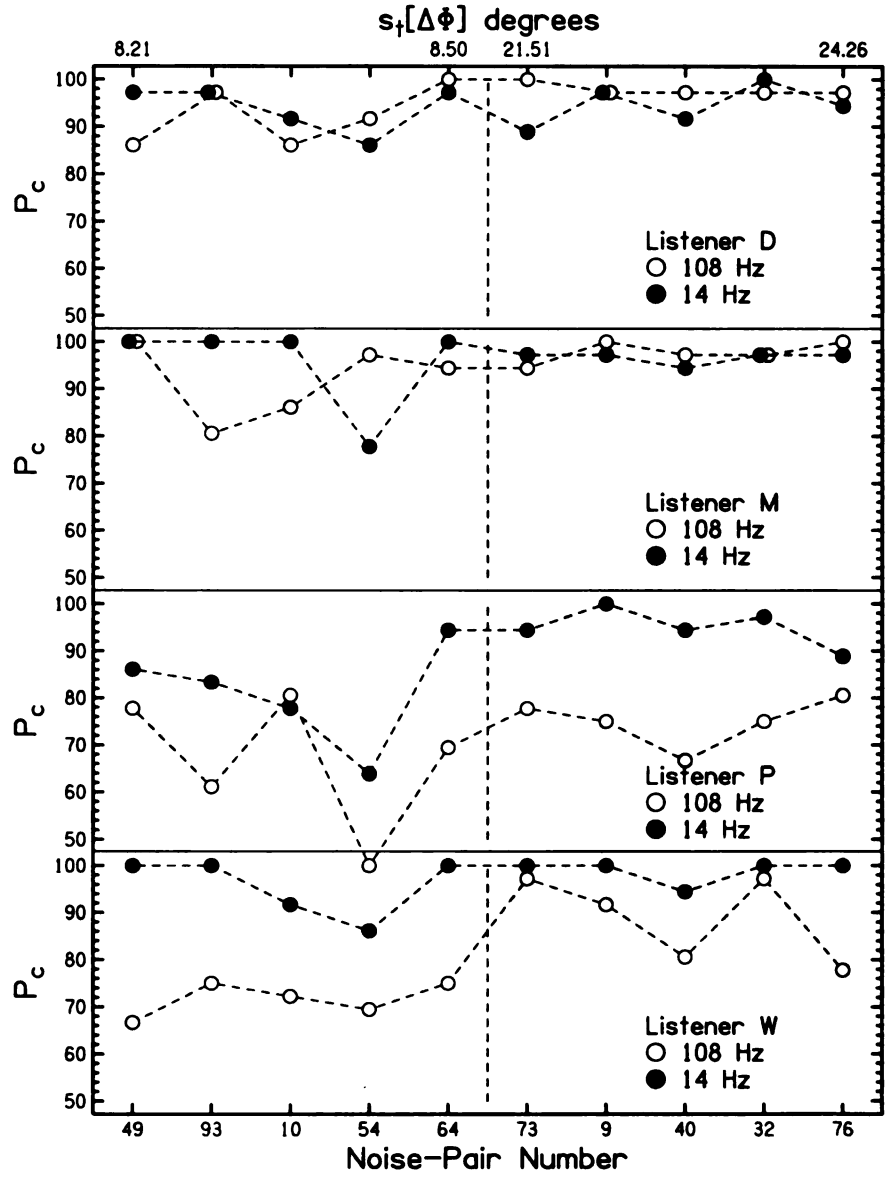

Figure 27: Matched phase set P_c data.

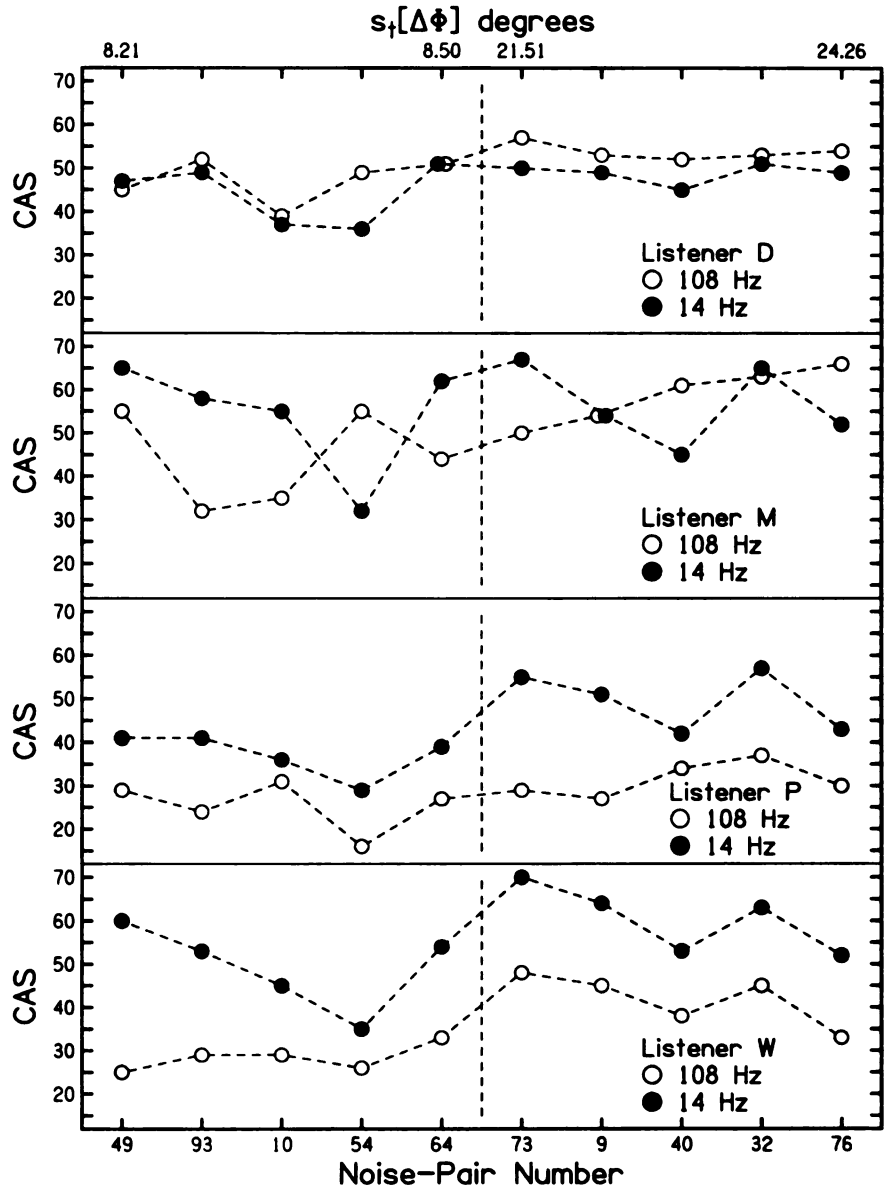

Figure 28: Matched phase set *CAS* data.

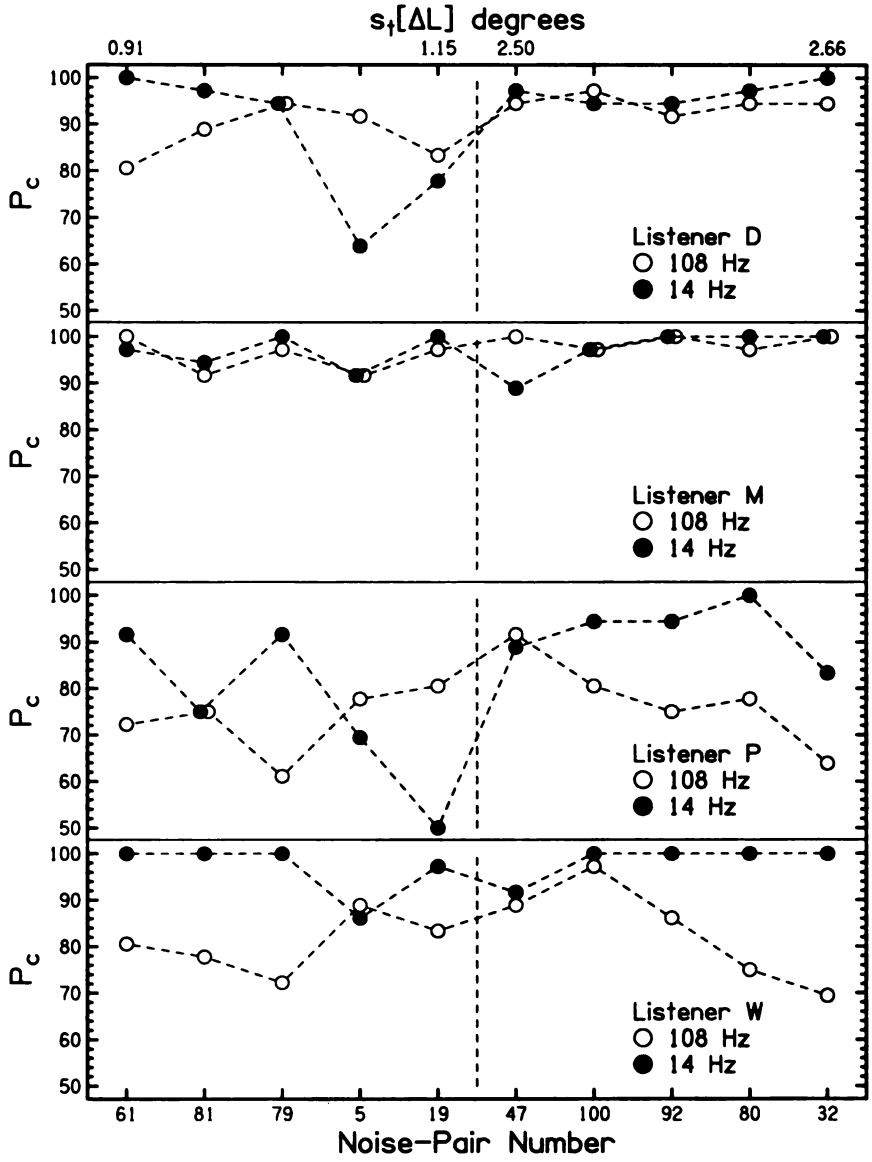

Figure 29: Matched level set P_c data.

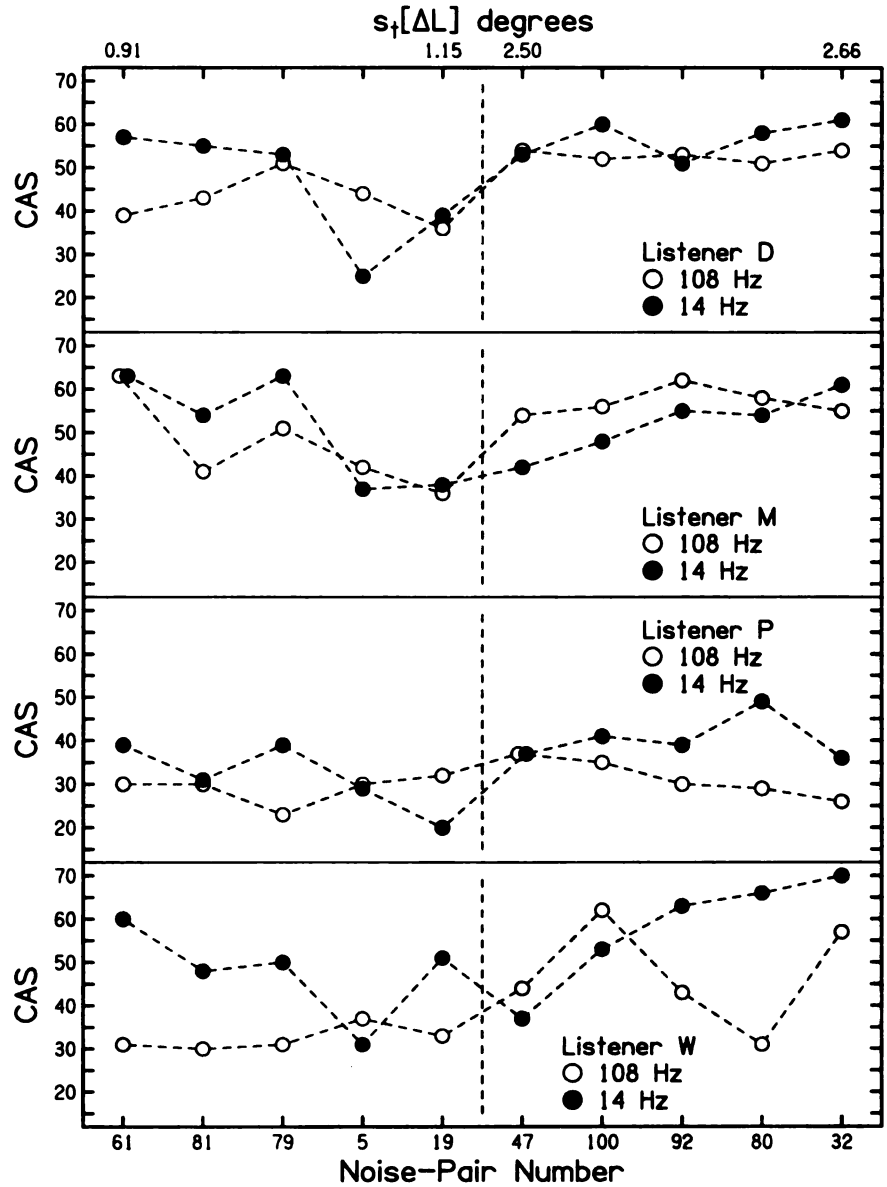

Figure 30: Matched level set *CAS* data.

Table 12: Inter-listener correlations for matched phase set.

P_c	D	M	P	W
D	1	0.681	0.599	0.750
M	–	1	0.644	0.783
P	–	–	1	0.804
W	–	–	–	1
Average	0.710			

CAS	D	M	P	W
D	1	0.708	0.737	0.808
M	–	1	0.661	0.799
P	–	–	1	0.912
W	–	–	–	1
Average	0.771			

with noise-pairs having a 14-Hz bandwidth. Experiment 4 was partially successful in that the number of significant p-values was comparable in Experiments 4 and 2, and quite different from Experiments 1 and 3. It is clear that the fluctuation matching procedure caused the 14-Hz bandwidth noise-pairs of Experiment 4 to behave more like the 108-Hz noise-pairs of Experiment 2 than like 14-Hz noise-pairs of Experiment 1.

Although the values of P_c and *CAS* for the 108-Hz bandwidth (Experiment 2) and the 14-Hz bandwidth (Experiment 4) were interleaved for two of the listeners, the values of P_c and *CAS* were higher for the 14-Hz bandwidth for the other two, and inter-listener correlations were higher overall for the 14-Hz bandwidth. Thus, there seems to be an advantage for the smaller bandwidth, whether the fluctuations are large or small. An advantage for smaller bandwidth was also found by Gabriel and Colburn (1981) where the jnd for detecting incoherence decreased with decreasing bandwidth. It is conjectured that the slowness of the fluctuations at the smaller bandwidth plays a major role in improving detection. Fluctuation speed was identified as an important factor in the MLD experiments by Zurek and Durlach (1987).

Table 13: Inter-listener correlations for matched level set.

P_c	D	M	P	W
D	1	0.256	0.681	0.752
M	–	1	0.115	0.758
P	–	–	1	0.387
W	–	–	–	1
Average	0.491			

CAS	D	M	P	W
D	1	0.714	0.659	0.679
M	–	1	0.561	0.691
P	–	–	1	0.464
W	–	–	–	1
Average	0.628			

1.5 EXPERIMENT 5: EFFECT OF MONAURAL CUES

During the course of the numerical study of interaural fluctuations it was decided to compare the interaural fluctuations in noise-pairs with the envelope fluctuations in the left-ear signal itself. (Left and right signals were so similar that it did not matter which was chosen.) *A priori* it seemed possible that noise-pairs with especially large (small) interaural fluctuations might often be derived from individual noise tokens with especially large (small) envelope fluctuations.

To make the comparison the envelope fluctuation was calculated,

$$s_t[E] = \sqrt{\frac{1}{M}\sum_{t=1}^{M}[E(t) - \overline{E}]^2} \tag{12}$$

where $E(t)$ is the envelope defined in Eq. 7, and $\overline{E}$ is the average over the time of the stimulus. Calculations with 1000 noise-pairs with 14-Hz bandwidth - stochastically identical to the pairs of Experiment 1 - showed a positive correlation between envelope fluctuations and interaural fluctuations. Specifically, the correlation between $s_t[E]$ and $s_t[\Delta L]$ was 0.48. The correlation between $s_t[E]$ and $s_t[\Delta\Phi]$ was 0.43. Since the level difference is calculated from the envelope, the first correlation seems intuitive. The latter correlation would be hard to understand were it not for the strong correlation between the standard deviations of interaural phase and level differences, as shown in Figs. 6, 15, and 21.

Next, the same calculations were made for the 20 noise-pairs actually used in Experiment 1. The interaural phase fluctuations and the interaural level fluctuations correlated with the monaural envelope fluctuations at levels of 0.59 and 0.65 respectively. Evidently, those waveforms with interaural fluctuations that are especially large or small particularly owe their binaural character to the envelope of the original generating noise token.

Given the positive correlation between interaural and monaural fluctuations, it

seemed possible that there might be information in the monaural signals that was used by listeners in performing the experiments of this chapter. Because of the evident importance of fluctuations in Experiment 1, attention centered on the stimuli used there, with a bandwidth of 14 Hz.

1.5.1 Diotic experiment method

Experiment 5 was identical to Experiment 1 except for the important difference that the left-ear signal of Experiment 1 was the signal for both ears in Experiment 5. In a second difference, the listeners in Experiment 5 had all completed Experiments 1 through 4 and therefore were highly experienced. The listeners were given three-interval sequences as before and were asked to apply the same strategy that they had used in the previous experiments. There was reason to believe that this approach might be successful because all the listeners volunteered that in the previous experiments they based their decisions on a sense of width, choosing the interval - either two or three - with the larger width. It seemed possible that the sense of roughness or other "action" associated with a diotic stimulus having large fluctuations could be associated with a sense of width. Consequently, it was expected that each trial of Experiment 5 would constitute a comparison between apparent "widths" for a particular identified noise and a different, randomly-chosen, noise from the set of ten. The stimulus sets were the phase set and level set from Experiment 1.

1.5.2 Results

Listeners made a negligible number of "confident" responses. Apparently the strong sensation of width elicited by some of the dichotic stimuli did not occur with any of the diotic noises. Therefore, the *CAS* had little value and results of Experiment 5 were plotted only in terms of the percentage of the trials on which a given noise was selected over other noises. The statistic will be called P_s, percent selected. It can be

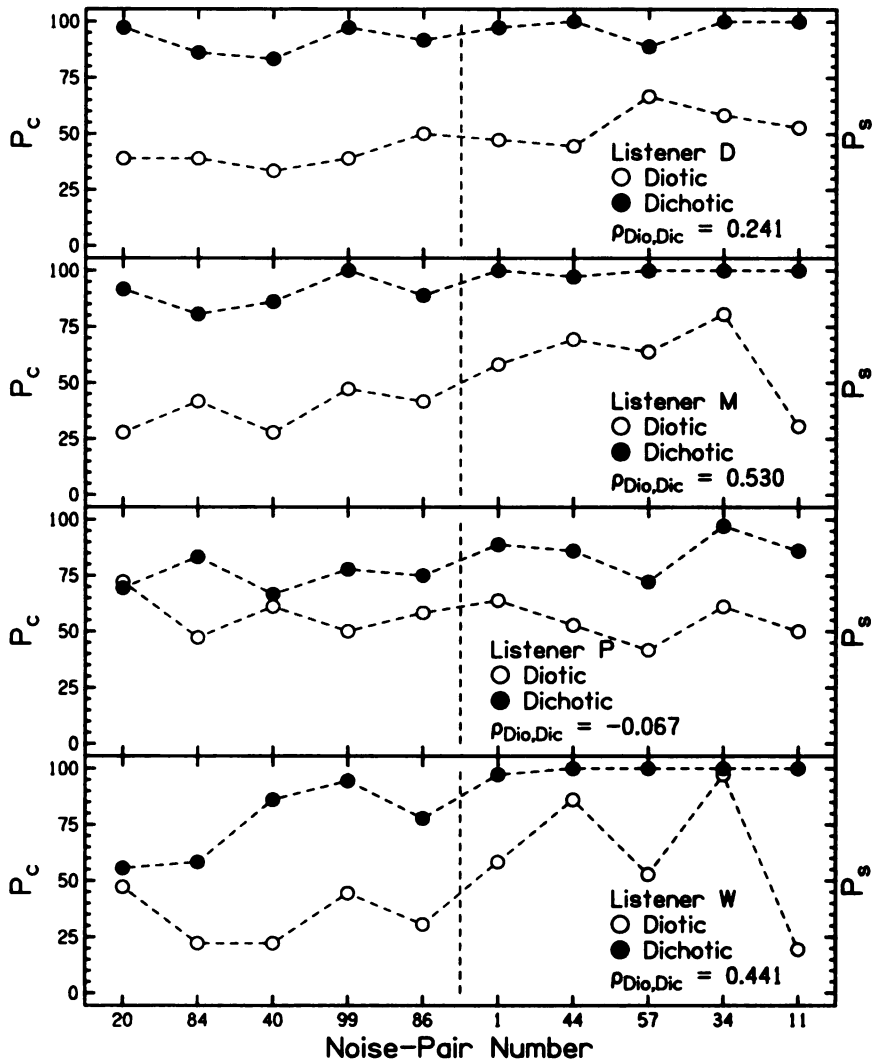

Figure 31: Comparison between P_c and P_s data for the phase set of Experiment 1 and the same noise-pairs with just the left channel presented diotically.

compared with P_c, the percent correct in the dichotic experiments.

1. Large fluctuation comparison

Particular interest centered on the five noise-pairs for which the interaural fluctuations were the greatest. By examining the scores from the diotic experiment it was expected to gain insight into the role that envelope fluctuations may have played in the dichotic experiment. The average values of P_s for those five noise-pairs for listeners D, M, P, and W were, respectively:

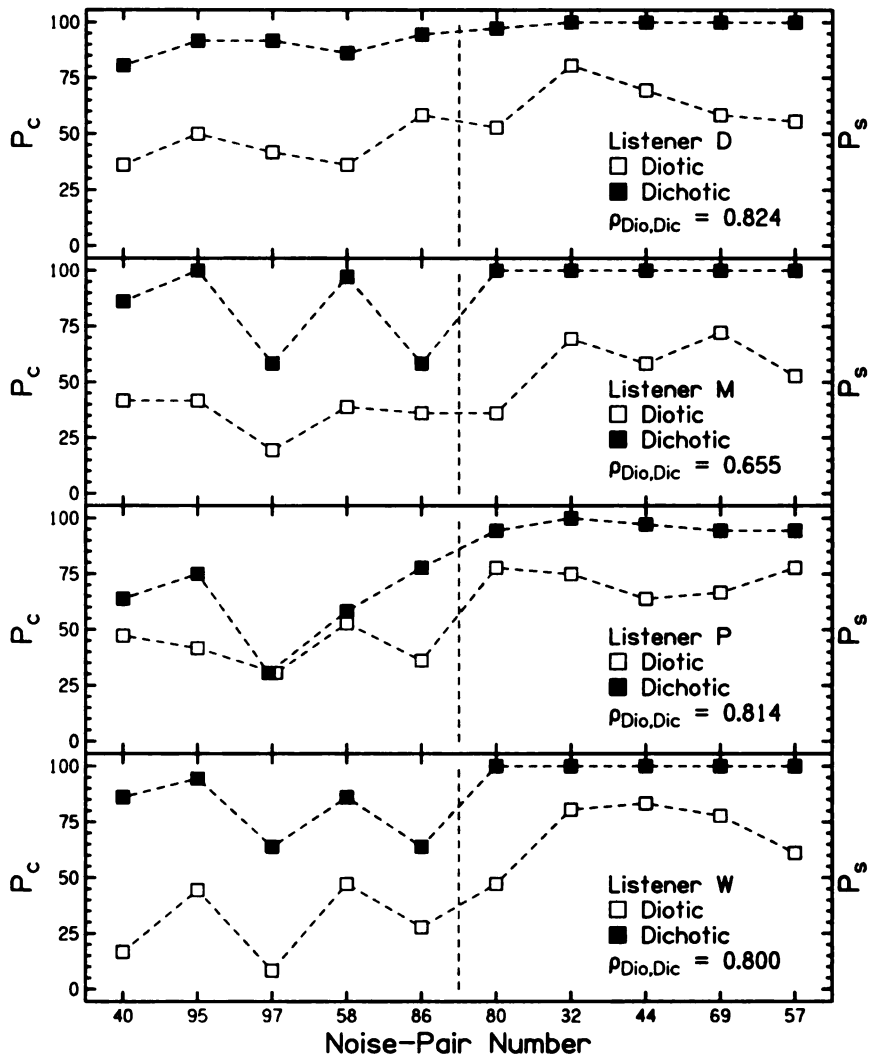

Figure 32: Comparison between P_c and P_s data for the level set of Experiment 1 and the same noise-pairs with just the left channel presented diotically.

For the phase set: 54, 61, 54, 63 %

For the level set: 63, 58, 72, 70 %.

Evidently, in the diotic experiment listeners chose the noises that had led to the largest interaural fluctuations clearly more than half the time. These numbers can be compared with the values of P_c in the dichotic experiment (Experiment 1) which averaged 88%.

2. Agreement between listeners

The agreement among the listeners was assessed by comparing values of P_s against noise serial number for listeners taken in pairs. Inter-listener correlations for D-M, D-P, D-W, M-P, M-W, and P-W were as follows:

For the phase set: 0.59, –0.38, 0.38, –0.22, 0.88, 0.17

For the level set: 0.70, 0.53, 0.76, 0.67, 0.89, 0.76.

The strongest correlation was between M and W. Listener P was responsible for the only negative correlations, both in the phase set. Correlations were clearly larger in the level set than in the phase set. The strong correlation indicates that listeners tended to agree about which fluctuations were salient.

3. Comparison with envelope fluctuation

A comparison between the listener selection of noises and fluctuation was assessed by comparing P_s with $s_t[E]$ as a function of the noise-pair serial number. Correlations for listeners D, M, P, and W were, respectively:

For the phase set: 0.38, 0.84, –0.03, 0.88

For the level set: 0.85, 0.66, 0.65, 0.70.

Again, P is responsible for the only negative correlation. The positive correlation indicates that the choices that listeners make can be predicted based on the physical envelope fluctuations, as measured by the standard deviation of the envelope over time, especially for the level set.

4. Comparison with Experiment 1

A comparison between the results of the corresponding diotic and dichotic experiments was made by comparing P_s on Experiment 5 with P_c on Experiment 1, both as functions of the noise-pair serial number. The data can be seen in Figures 31 and 32. Correlations for listeners D, M, P, and W were, respectively:

For the phase set: 0.24, 0.53, –0.07, 0.44

For the level set: 0.82, 0.66, 0.81, 0.80.

Again, correlations are larger for the level set.

1.5.3 Discussion

The correlations above are fairly impressive for being so large with possibly some exceptions for Listener P. These include the correlations between P_s in Experiment 5 and P_c in Experiment 1 as well as the correlations between P_s and monaural and dichotic fluctuations. There are several possible interpretations of these correlations.

Possibly the correlation between P_c and P_s scores only represents a chain of stimulus circumstances. For a narrow bandwidth like 14 Hz, every kind of stimulus fluctuation seems to correlate with every other kind. Interaural phase fluctuations correlate with interaural level fluctuations and both correlate with noise envelope fluctuations. In a dichotic experiment probing the detection of interaural incoherence listeners attend to the interaural fluctuations. In a diotic experiment probing an evaluation of roughness or other stimulus action listeners attend to the envelope fluctuations. The results of the two experiments, as functions of the stimulus serial number, are similar because the interaural and monaural fluctuations behave similarly with respect to serial number.

Alternatively, it is possible that the correlation between P_c and P_s scores arises because listeners in a dichotic experiment are misled by monaural envelope fluctuations that are particularly large or particularly small. Given the enormous difference in

the average P_c and P_s values for the five noise-pairs with large fluctuations, as noted in the results section above, it seems highly unlikely that monaural fluctuations *per se* contribute to listener judgements in the binaural experiment when the detection of interaural incoherence is easy. But when detection of interaural incoherence is difficult, or impossible, the cues from monaural envelope fluctuations (or the lack of them) may influence judgements and are probably responsible for P_c values in a dichotic experiment that are less than chance.

1.6 DISCUSSION AND CONCLUSION

Listeners are sensitive to small amounts of interaural coherence in noise. Given diotic noise as a comparison, listeners can detect a coherence change of 0.01, i.e. they are sensitive to the difference between 1.00 and 0.99 (Gabriel and Colburn, 1981). The goal of this chapter was to understand the origins of this remarkable sensitivity.

1.6.1 Detection of incoherence

Experiment 1 selected stimuli from an ensemble of 100 reproducible left-right noise-pairs, all of which had a bandwidth of 14 Hz and an interaural coherence of 0.9922. It was found that those pairs that had a large fluctuation in interaural phase difference (IPD) or large fluctuation in interaural level difference (ILD) were much more readily recognized as not perfectly coherent compared to pairs with small fluctuations. This result led to the conclusion that, for bandwidths as narrow as 14 Hz, the interaural coherence is not an adequate predictor of the ability to detect incoherence. Instead, the size of the interaural fluctuations matters.

Experiments 2 and 3 progressively increased the bandwidth and found that the ranges of fluctuations in IPD and ILD among different noise-pairs in an ensemble decreased with increasing bandwidth. (See Appendix B). This observation led to the expectation that the detectability of incoherence would exhibit less variation for

different noise-pairs with these wider bandwidths. Detection experiments similar to Experiment 1 showed an increasing uniformity in detectability for the incoherence in noises with increasing bandwidths, as expected.

It was conjectured that the only reason that detection performance for 14-Hz bandwidth was different from performance for 108-Hz bandwidth was that more extreme values of fluctuations (both very small and very large) were available in the ensemble with the narrow bandwidth. In Experiment 4 a comparison was made between performance on noise-pairs with 108-Hz bandwidth and performance on a matched set of noise-pairs with 14-Hz bandwidth. Noise-pairs in the matched set were selected to have approximately the same interaural fluctuations as the pairs of the 108-Hz set. The comparison showed that detectability differences among different noise-pairs in the matched set were reduced to about the same level as for the 108-Hz noises, consistent with the conjecture. The overall performance on the 14-Hz matched set was approximately equal on the 108-Hz set for two of the listeners; it was consistently higher for the other two listeners. These two results from the comparison suggest that differences in interaural fluctuations are responsible for differences in the detectability of incoherence for different noises, but that fluctuations of comparable size are more easily detected when the bandwidth is narrow. The most likely explanation for the advantage of narrow bands is that fluctuations are slower. The role of fluctuation speed will be addressed in the chapter on binaural modeling.

These conclusions differ from those of Breebaart and Kohlrausch (2001) who dismissed a specific role for IPD and ILD fluctuations in binaural detection because the distributions of those fluctuations, with or without a signal, failed to show a bandwidth dependence. By contrast their NρSπ detection experiments did show such a dependence. It is agreed that the $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ have mean values, averaged across waveforms, that are insensitive to bandwidth, as noted in Figs. 6, 15, and 21, and as shown in the widths of the distributions plotted by Breebaart and Kohlrausch. How-

ever, the variations in $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ among different waveforms depend strongly on bandwidth, and these variations are responsible for the large detectability differences that are reported.

Lastly, a check was done to ensure that the methods used in the first four experiments were valid. In Experiment 5, attempts were made to reproduce the results in the previous experiments with only diotic presentations. The results could not be fully reproduced, however, listeners showed positive correlations between monaural fluctuation and incoherence detection. This correlation might explain why some listeners performed below chance on certain noise-pairs.

1.6.2 Caveats - duration and bandwidth

All the stimuli used in this chapter had a duration of 500 ms. The stimuli were all constructed to have an interaural coherence of 0.9922 computed over this duration. Because different noise-pairs exhibited very different detectabilities, this chapter reached a main conclusion that the coherence measure is inadequate to predict detection for narrow bands. A problem that arises with this conclusion is that one expects that the temporal analysis windows that are appropriate for binaural perception are considerably shorter than 500 ms. Therefore, it is possible that, after all, the coherence is a perfectly adequate statistic to predict incoherence detection, but that the coherence must be calculated over the correct (shorter) time interval(s). This point of view would say that it was a mistake to compute a fixed coherence over an interval as long as 500 ms and expect it to be perceptually valid because the coherence varies from one momentary analysis window to the next.

There are two response to the above criticism. First, model calculations based on a running short-term cross-correlation have been performed (see Chapters 4 and 5) and it was found that this measure of coherence is inadequate to account for the detectability of incoherence in 500-ms noises with a 14-Hz bandwidth. Second, a

series of experiments with 14-Hz bandwidth and progressively shorter durations have been done (see Chapter 3). Coherence values calculated over those short durations were again required to be 0.9922. Stimuli were again selected based on large or small fluctuations over the duration, and listeners did discrimination experiments similar to Experiment 1. It was found that the results for durations of 100 ms and 50 ms were the same as the results for 500 ms in that noise-pairs with large fluctuations led to CAS values that were significantly greater compared to noise-pairs with the smallest fluctuations.

A second criticism pertains to the wideband noise-pairs investigated in Experiment 3. Specific noise-pairs were chosen with large or small fluctuations as computed over the entire bandwidth. It is possible, however, that listeners pay attention to some sub-band(s) as determined by auditory filters. If so, then the basis for choosing the pairs was faulty and it would not be surprising to find that there is no significant difference in CAS values for the different pairs. This criticism would weaken the case that is made here for the adequacy of the coherence measure as a predictor of incoherence detection in the wideband limit. This criticism could be removed if it could be shown that listeners did not make their decisions based on listening in sub-bands. One piece of evidence against sub-band listening is that the performance for the wide bands Experiment 3 is systematically worse than performance in the approximate critical bands of Experiment 2. If listeners were able to listen in sub-bands, one would expect comparable performance instead. A similar conclusion from an MLD experiment was reached by Evilsizer *et al.* (2002) who found that NoSπ detection was *not* determined by the masker within a critical band.

1.6.3 Binaural processing

The psychoacoustical literature often connects binaural capabilities, particularly binaural release from masking, with interaural cross-correlation (Wilbanks and Whit-

more, 1967; Domnitz and Colburn, 1976; Koehnke *et al.*, 1986). Wilbanks (1971) cited historical articles by Cherry, Licklider, and Jeffress "... supporting the notion that the binaural system is, logically speaking, a correlational detector."

The experiments of the present chapter agree with that conclusion, but only when the bandwidth is not small. A correlational model adequately reflects the fluctuations only in the wideband limit. Wideband noise signals tend to be ergodic wherein the statistical properties of an ensemble of noises become manifest in any given noise sample as that particular noise evolves in time. Then the fluctuations in any given sample of wideband noise are not much different from those in any other sample, and the size of the fluctuations appears to be a simple function of the coherence. The transition to that limit was seen in Experiments 1–3 where, with increasing bandwidth, the variance among different noises of IPD and ILD decreases and the ability of listeners to detect incoherence varies less among different noise samples. The wideband limit is approached but not reached for a bandwidth of 108 Hz, which is close to a one-third octave or critical bandwidth for 500 Hz. With a bandwidth that wide, a psychoacoustical experiment might easily miss the inhomogeneity of noises unless it were specifically designed to look for it.

For a narrow bandwidth, such as 14 Hz, each individual noise has so few spectral components that the interaural properties of an individual noise-pair can differ greatly from the ensemble mean properties. Improbable variations of this kind are responsible for the fact that a small amount of incoherence may be difficult to detect for one sample of noise but easy to detect for a another one, as reported in this chapter. It is likely that this effect was present in many historically important studies of binaural effects in narrow bands. In their coherence discrimination experiment Gabriel and Colburn (1981) found the just noticeable difference to decrease with decreasing bandwidth given a reference coherence of 1.0. The decrease may result from particularly favorable noise samples that occur for narrow bands.

There are parallels in MLD experiments in the detection of a low-frequency signal in the NoSπ condition. Van de Par and Kohlrausch (1999) found that as the bandwidth decreased, while remaining narrower than a critical band, signal thresholds remained approximately constant. The results of listeners D and M in Experiments 1 and 2 seem similar in that CAS values do not change much as the bandwidth changes from 14 to 108 Hz. By contrast Zurek and Durlach (1987) found that thresholds decrease as the bandwidth decreases, a result that seems consistent with listeners P and W in Experiments 1 and 2 presented here. Zurek and Durlach (1987) interpreted the decrease in threshold that they observe as the result of the slowness of fluctuations for narrow bands. The results of Experiment 4, using stimuli that matched the IPD and ILD fluctuations, but not the fluctuation rate, support this interpretation too.

Although there are parallels between interaural incoherence detection and MLD tasks, there are also important differences. An NoSπ stimulus with a sine signal and noise masker may include both static and dynamic interaural cues. An interaural incoherence detection experiment, as presented here, involves dynamic fluctuations only. Results can be different too. Using multiplied noise maskers and sine signals Breebaart *et al.* (1999) found that cross-correlation accounted for their signal detection data better than various measures of interaural fluctuations, a result that is entirely contradictory to the experience with incoherence detection. Further, the experimental data obtained by Breebaart *et al.* were best fitted by a model that used the energy in the binaural difference signal as a decision variable. For stimuli such as ours, constructed from orthogonalized waveforms with constant coherence, the energy in the difference signal is the same in all waveforms. Such a model would predict no difference in a listener's ability to detect the incoherence in any of the noise-pairs, contrary to experiments.

In an MLD experiment, the cues for detecting the signal are reported to be different depending on the bandwidth. For narrow bands an Sπ signal contributes a width

to the binaural image. For wide bands an $S\pi$ signal increases the tone-like strength (Evilsizer *et al.*, 2002). By contrast, in an incoherence detection experiment only the width cue occurs, whatever the bandwidth, though the rate of the fluctuations that establish the width does vary with bandwidth.

Experimental evidence in favor of this bandwidth effect on cues is that in an MLD experiment, performance for NoSo is correlated with performance for $NoS\pi$ for wide bands, but the correlation is significantly less for narrow bands (Evilsizer *et al.*, 2002.) [It should be noted that the bands called "narrow" by Evilsizer *et al.* are 100 Hz in width, equivalent to the "critical-band" noise-pairs.] See also Gilkey *et al.* (1985) and Isabelle and Colburn (1991) for similar evidence as described by Evilsizer *et al.* Presumably this difference in correlation arises because listeners are using the tonal cue for wide bands, which is similar for So and $S\pi$, but are using the width cue for narrow bands, which has a binaural contribution only for $S\pi$.

Similarly, Evilsizer *et al.* found that inter-listener correlation was strong for NoSo and also for wideband experiments for any combination of interaural noise and signal phases. Inter-listener correlations dropped and became negative for $NoS\pi$ when the band was narrow (100 Hz). This result is consistent with the idea that the tonal cue is similarly detected by different listeners, but that there are marked individual differences when the cue becomes an image width for narrow bands (Bernstein *et al.*, 1998). The situation for incoherence detection is just the opposite. Inter-listener correlations are strongest for narrow bands where the width cues vary greatly among different noise-pairs, and inter-listener correlations are weak for wide bands where the width cues are similar for different pairs. A corollary to the above argument is that Evilsizer *et al.* would have found better correlation among listeners for $S\pi$ in narrow bands had they used bands as narrow as ours.

In an MLD experiment with noise bands that are not narrow, a listener might listen in different sub-bands where the values of interaural coherence will be very

different. In an incoherence detection experiment the stimulus is simpler because the coherence is spectrally homogeneous and advantageous sub-band listening possibilities are limited to special cases for individual noise samples as deviations from the ensemble mean statistics. In my view, the incoherence detection experiment has the advantage that it extracts the essential binaural element from an MLD experiment and exposes it for observation.

Experiment 1 clearly showed the importance of the dynamic fluctuations in interaural parameters for narrow bands of noise, in contrast to the coherence measure of interaural differences. Experiment 1 gained its power from a selection of stimuli based on values of the standard deviation of IPD and ILD over time. These measures were plausible guesses about what is important to the binaural system, but there is no reason to expect them to be optimum measures of interaural fluctuations. In fact, it is clear that they are not optimum. Stimulus #57 in the 14-Hz phase set had one of the largest phase fluctuations and yet Figure 10 shows that all four listeners had a relatively difficult time with it. Chapter 4 will explore alternative models for the detection of incoherence in narrowband noises and attempts to determine the best characterization of fluctuations from a perceptual point of view.

2 FLUCTUATIONS WITH VARIED COHERENCE

In Chapter 1, it was discovered that the detection of incoherence in random noise-pairs is determined by the fluctuations of interaural phase and interaural level. It follows that the results of Chapter 1 may be reproduced by using noise-pairs with different values of coherence but similar fluctuations in IPD and ILD. This chapter will be similar to Chapter 1 with respect to the experiments performed (although additional experiments will be done). The value of this chapter is to validate the results of Chapter 1 with new listeners and new stimuli; it will come to few new conclusions. The figures and several tables have been omitted because of their similarity to Chapter 1 results, however the tables with p-values from t-tests remain.

2.1 EXPERIMENT 6: NARROW BANDWIDTH

Experiment 6 was identical to Experiment 1 except that the value of the coherence was not fixed at 0.9922. It will test to see if noise-pairs with comparable fluctuations yield similar data to that in the first chapter.

2.1.1 Method

As for Experiment 1, the center frequency was 500 Hz, with a bandwidth of 14 Hz, and the same raised-cosine shaping was applied to the amplitude spectrum. However, there were differences between experiments. The spectral spacing was 1 Hz instead of 2 Hz. The sample rate was 2000 sps and the buffer size was 2000 samples. Thus, the duration of the noise-pairs was initially 1 s, not 0.5 s as in Experiment 1. The same temporal shaping used in Chapter 1 was used on all the noise-pairs in Chapter 2. Noise-pairs were also generated with the same Tucker-Davis equipment. The signals were low-pass filtered at 700 Hz in this experiment (instead of 4000 Hz) to prevent

aliasing because the sample rate was lower.

The noise-pairs underwent the same orthogonalization and mixing as in Chapter 1. Initially, the value of coherence was 0.9922 for each of the 1-s noises that were generated for this experiment. However, the value of the coherence possibly changed because only the first 0.5 s of the stimuli were used. The spread of values of coherence was 0.969–0.998 as can be seen in Figure 33. The mean value of coherence was 0.991, slightly smaller the the initial 0.9922.

A phase set was selected as in Chapter 1, by choosing five noise-pairs with the largest phase fluctuations and five noise-pairs with the smallest phase fluctuations. Similarly, a level set was selected. The values of the fluctuations for individual noise-pairs can be seen in Figure 34. The means, standard deviations, and correlation of the phase and level fluctuations are comparable to those in the collection with fixed coherence seen in Figure 6. The means in this figure are slightly higher for the varied coherence noise-pairs compared to the fixed coherence noise-pairs. The reason may be the smaller frequency spacing. The fixed coherence noise-pairs (2-Hz spacing) have half the number of random variables compared to the varied coherence noise-pairs (1-Hz spacing) of this experiment. Therefore, for the 1-Hz spaced noise-pairs, there is greater variability in the interaural parameters and a slightly larger mean for the entire collection. The standard deviations and correlation between phase and level fluctuations remained the same.

2.1.2 Listeners

This experiment employed two female listeners, C and E, and three male listeners, M, W, and Z. Listeners C, E, and Z were between the ages of 20–35 and had normal hearing according to standard audiometric tests and histories. Listeners M and W also participated in experiments in Chapter 1.

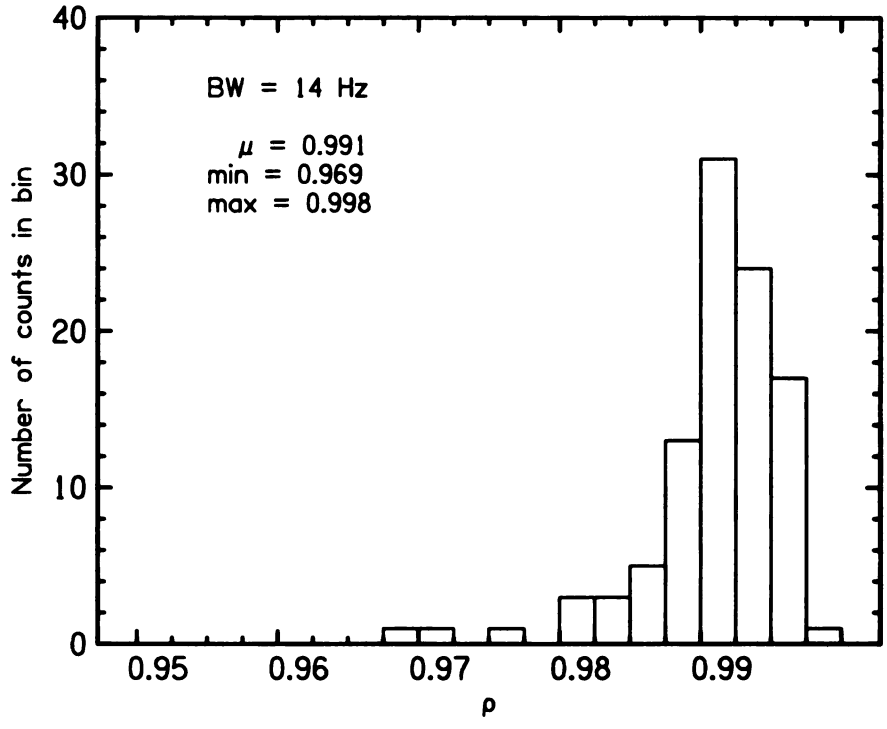

Figure 33: Range of values of coherence for the noise-pairs in Experiment 6.

2.1.3 Results

Two-sample t-tests were performed for the detection scores for the the five noise-pairs with the greatest fluctuations and the five noise-pairs with the smallest fluctuations. All ten differences were significant at the 0.05 level for the phase set and level set P_c data. All ten differences were significant at the 0.02 level for the *CAS* data. The individual p-values are shown in Table 14, comparable to the p-values in Table 2 in Experiment 1. The p-values for Experiment 1 were significant at the 0.05 level for the P_c data and at the 0.02 level for the *CAS* data.

The number of significant p-values can easily be compared for Experiment 1 and Experiment 6 in Table 15. The individual inter-listener correlations are not reported, but the average correlation is reported in Table 15. For the P_c values, the average inter-listener correlation for this experiment was larger than the inter-listener correlation for Experiment 1. The correlation was also larger for the phase set with *CAS*

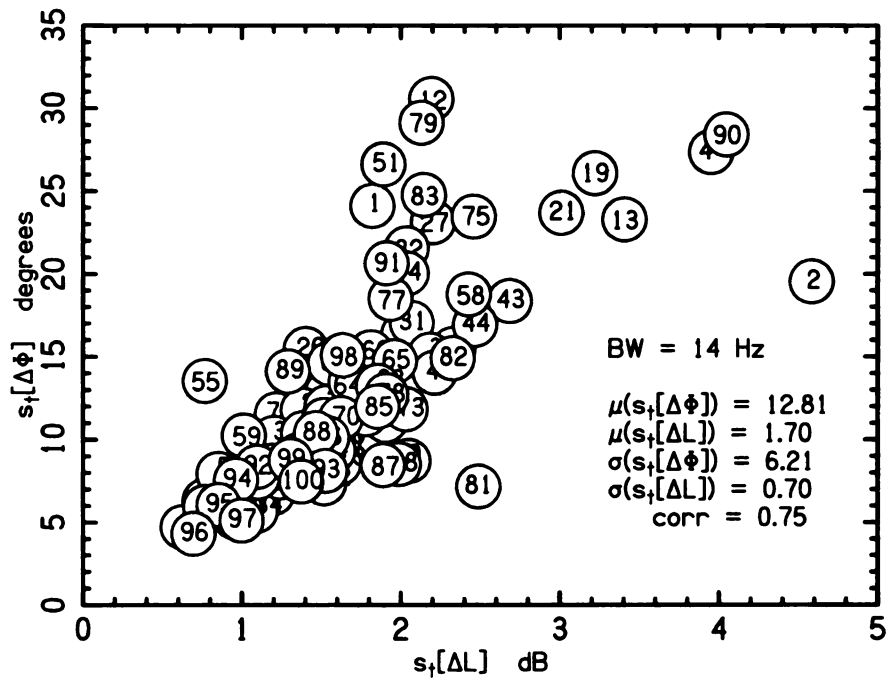

Figure 34: Fluctuations of IPD versus fluctuations of ILD for the first collection of 100 reproducible noise-pairs having a 14-Hz bandwidth, as used in Experiment 6. Each noise-pair is labeled by a serial number indicating only the order of creation. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. These values are similar to those in Figure 6 in Chapter 1, Experiment 1, but the means of the distributions are slightly larger in this collection.

values. It was smaller for the level with *CAS* values.

Table 14: The p-values from a one-tailed t-test for 14-Hz bandwidth data. The test compared the five noise-pairs with the largest fluctuations to the five noise-pairs with the smallest fluctuations. The p-values indicating at least a 95% level of confidence are in bold.

P_c	Phase	Level	*CAS*	Phase	Level
Listener	Set	Set	Listener	Set	Set
C	**0.023**	**0.002**	C	**0.003**	**<0.001**
E	**0.009**	**0.014**	E	**0.001**	**<0.001**
M	**<0.001**	**0.001**	M	**<0.001**	**0.002**
W	**0.036**	**0.008**	W	**0.002**	**<0.001**
Z	**<0.001**	**0.016**	Z	**<0.001**	**<0.001**

Table 15: Comparison of p-values and inter-listener correlations for Experiment 1 and Experiment 6, the 14-Hz bandwidth experiments.

P_c	Experiment 1 Fixed ρ	Experiment 6 Varied ρ
# of sig. p-values	7/8 (87.5%)	10/10 (100%)
Ave. phase set corr.	0.564	0.877
Ave. level set corr.	0.641	0.752
CAS	Experiment 1 Fixed ρ	Experiment 6 Varied ρ
# of sig. p-values	8/8 (100%)	10/10 (100%)
Ave. phase set corr.	0.857	0.939
Ave. level set corr.	0.917	0.906

2.1.4 Discussion

This experiment shows that similar fluctuations of phase and level yield similar detection data, whether the value of the coherence is fixed or allowed to vary. In comparing Experiment 6 to Experiment 1, both the P_c and *CAS* values show the same number of significant differences and the same level of significance for the phase and level sets.

However, the inter-listener correlations are smaller for the fixed coherence stimuli in Experiment 1 in three of four comparisons. The reason for this may be Listener P from Experiment 1, who had some of the smallest inter-listener correlations, brought down the average. Listener P was not in Experiment 6.

2.2 EXPERIMENT 7A: CRITICAL BANDWIDTH

Experiment 7A was similar to Experiment 2 except that the noise-pairs had varied coherence, the bandwidth was 136 Hz (instead of 108 Hz), and two collections of 100 noise-pairs were used.

2.2.1 Method

As in Experiment 6, the center frequency was 500 Hz and the spectral spacing was 1 Hz. Again, only the first 0.5 s of the 1 s noise-pair was used. Thus the value of the coherence could vary. The range of values of coherence can be seen in Figure 35. The mean correlation was 0.992, which was the same value as the originally generated noise-pairs. The range of values of coherence was 0.989–0.995 - much smaller than that in Experiment 6.

A raised-cosine shaping was applied to the amplitude spectrum, except that the edge width was multiplied by a factor of 10. Therefore, components between 450 and 550 Hz had unity amplitude; components in the ranges 400–450 and 550–600 Hz were shaped.

Two collections of 100 noise-pairs were used in this experiment. The serial numbers were 1–100 for the first collection and were 101–200 for the second collection. The interaural parameters $\Delta\Phi$ and ΔL can be seen as a function of time in Figure 36. A phase set and level set was chosen from each collection. Therefore, there were twice as many sets used in this experiment when compared to all the previous experiments.

Figure 37 shows $s_t[\Delta\Phi]$ versus $s_t[\Delta L]$ for the first collection of 136-Hz bandwidth

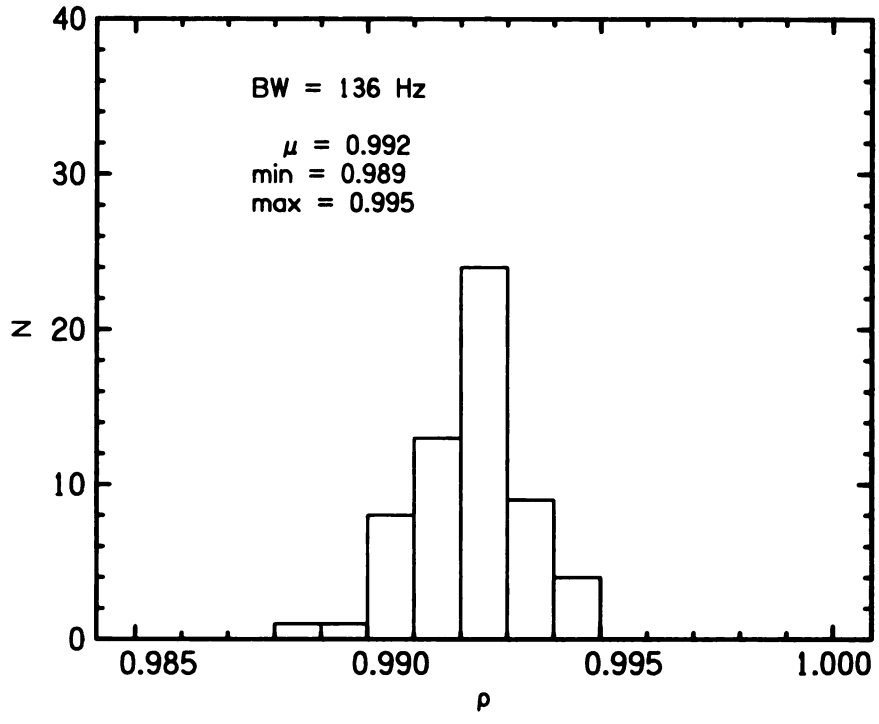

Figure 35: Range of values of coherence for the noise-pairs in Experiment 7A, 7B, and 7C.

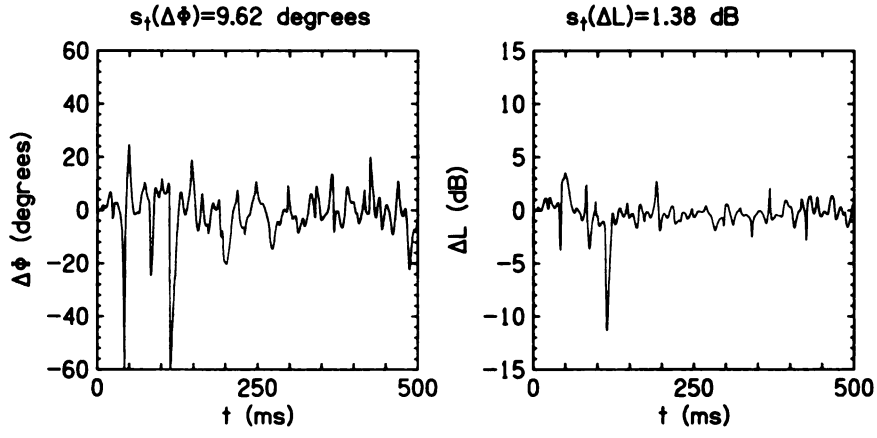

Figure 36: The interaural parameters of a typical 136-Hz bandwidth noise-pair. The fluctuations are comparable to those in the 108-Hz noise-pairs as seen in Figure 14.

noise-pairs. The black dots represent the noise-pairs with 14-Hz bandwidth from Experiment 6 for comparison. The mean of the distribution of fluctuations remains essentially the same as for the 14-Hz bandwidth noise-pairs. This is different from Chapter 1 where the means of the Experiment 1 noise-pair fluctuations were slightly smaller than the means of the collection in Experiment 2.

The standard deviations were lower for Experiment 7A compared to Experiment 2. The standard deviation of the phase fluctuations was 2.6 degrees for Experiment 7A and 3.7 degrees for Experiment 2. The standard deviation of the level fluctuations was 0.23 dB for Experiment 7A and 0.40 dB for Experiment 2. This can be explained by the bandwidth being about 25% larger for the noise-pairs in Experiment 7A.

In comparing Experiment 7A to Experiment 6, the standard deviation over the ensemble of the phase fluctuations decreased from 6.2 degrees to 2.6 degrees when the bandwidth increased from 14 Hz to 136 Hz. The standard deviation of the level fluctuations decreased from 0.70 dB to 0.23 dB. A decrease in standard deviation was also seen between Experiments 1 and 2.

The correlation between $s_t[\Delta\Phi]$ and $s_t[\Delta L]$, evaluated over the ensemble of 100 waveforms, decreased somewhat, from 0.75 to 0.67. The correlation in Chapter 1 increased across similar size bandwidths, but this was a statistical anomaly (see Appendix 2).

Expanded views of the two collections can be seen in Figure 38 and Figure 39. The noise-pairs on the two figures show comparable distributions of phase and level fluctuations, as one would expect.

2.2.2 Listeners

The five listeners from Experiment 6 plus two other male listeners participated in this experiment. Listeners DY and T were between the ages of 20–30 and had normal hearing according to standard audiometric tests and histories. All seven listeners

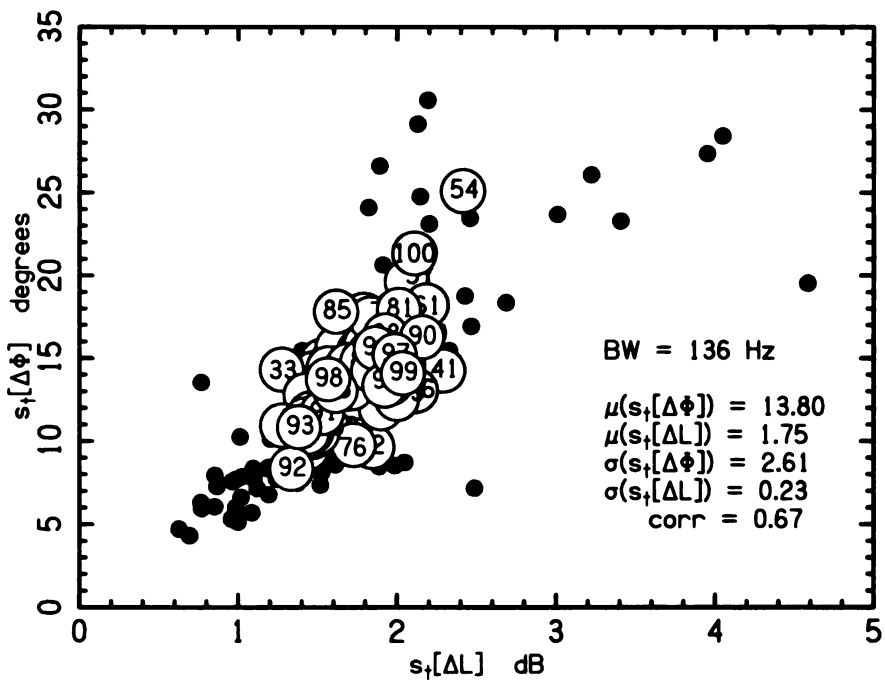

Figure 37: Fluctuations of IPD versus fluctuations of ILD for the first collection of 100 reproducible noise-pairs having a 136-Hz bandwidth, as used in Experiments 7A. Each noise-pair is labeled by a serial number indicating only the order of creation. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. The black dots represent the 14-Hz bandwidth noise-pairs from the collection in Experiment 6.

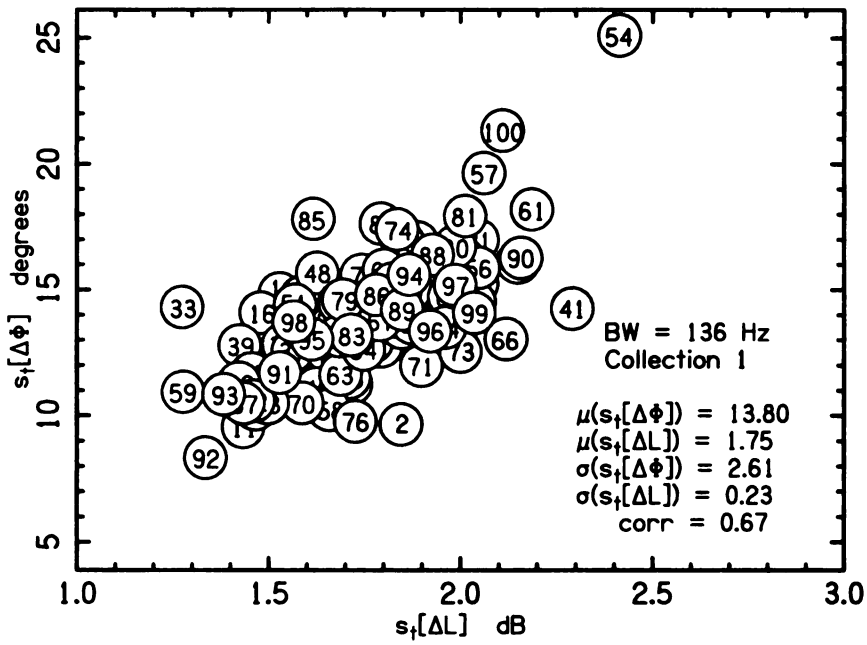

Figure 38: Expanded view of the first collection of 100 noise-pairs.

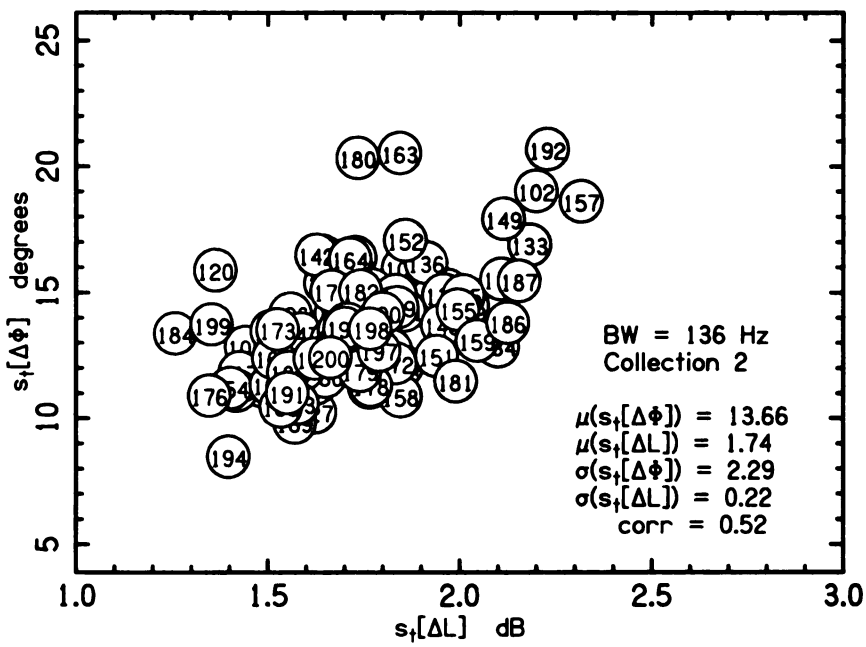

Figure 39: Expanded view of the second collection of 100 noise-pairs.

participated in detecting incoherence in the phase set and level set for Collection 1. Listeners DY, E, M, T, and W participated in detecting incoherence in the phase set and level set for Collection 2.

2.2.3 Results

Table 16 shows that only two of the twelve t-tests (seven listeners in set 1 and five listeners in set 2) from the phase sets P_c values led to differences that were significant at the 0.05 level between the results for the five largest fluctuations and the results for the five smallest fluctuations. This increased to three of twelve significant p-values when using the *CAS* values. The first phase set had all of the significant p-values even though the listeners with significant p-values in set 1 were also participants for set 2.

Table 16: The p-values from a one-tailed t-test for 136-Hz bandwidth data.

P_c	Phase	Phase	Level	Level
Listener	Set 1	Set 2	Set 1	Set 2
C	0.611	–	**0.044**	–
DY	**0.036**	0.265	**0.019**	**0.012**
E	0.165	0.557	**0.001**	**0.006**
M	0.094	0.403	**0.040**	0.280
T	0.110	0.545	0.236	0.250
W	**0.005**	0.228	**0.022**	**0.022**
Z	0.250	–	**0.019**	–

CAS	Phase	Phase	Level	Level
Listener	Set 1	Set 2	Set 1	Set 2
C	0.237	–	**0.039**	–
DY	**0.031**	0.215	**0.008**	**0.008**
E	0.658	0.275	**0.002**	0.106
M	0.057	0.432	0.070	0.077
T	0.057	0.516	0.240	0.158
W	**0.003**	0.057	**0.039**	**0.013**
Z	**0.019**	–	**0.018**	–

Nine of the twelve t-tests from the level sets for the P_c values led to differences that were significant at the 0.05 level. This dropped to seven of twelve significant p-values when using the *CAS* values. Both level sets showed significant p-values.

Table 17: Comparison of Experiment 2 and Experiment 7a p-values and inter-listener correlations.

P_c	Experiment 2 Fixed ρ	Experiment 7A Varied ρ
# of sig. p-values	2/8 (25%)	11/24 (45.8%)
Ave. phase set corr.	0.225	0.395
Ave. level set corr.	0.117	0.528
CAS	Experiment 2 Fixed ρ	Experiment 7A Varied ρ
# of sig. p-values	6/8 (75%)	10/24 (41.7%)
Ave. phase set corr.	0.379	0.571
Ave. level set corr.	0.278	0.643

Table 17 shows the comparison between Experiments 2 and 7A. The percentage of significant p-values is larger (45.8% compared to 25%) for the P_c data and smaller (41.7% compared to 75%) for the *CAS* data when comparing the varied coherence sets to the fixed coherence sets. However, all of these percentages are smaller than the 100% of significant p-values seen in Experiments 1 and 6. In general, the inter-listener correlations were larger for the varied coherence noise-pairs.

2.2.4 Discussion

This experiment tried to reproduce the results from Experiment 2 with slightly larger bandwidth noise-pairs (136 Hz compared to 108 Hz). The percentage of significant p-values was less than the 100% of Experiments 1 and 6, which used a 14-Hz bandwidth. The level of significance was also at the 0.05 level, as opposed near to the 0.02 level in Experiments 1 and 6 for the *CAS* scores. However, there is a difference when comparing Experiments 2 and 7A. The inter-listener correlations were much smaller for Experiment 2 than for Experiment 7A.

As for certain noise-pairs from Chapter 1, incoherence detection for some noise-pairs was clearly out of the expected order in this experiment. As the most striking example, all five listeners had difficulty detecting incoherence in noise-pair #192 even though both the phase set and level set from Collection 2 include this noise-pair as one of the largest fluctuations of IPD and ILD. This noise-pair is clearly out of order, and a more sophisticated means of ordering needs to be employed to account for listeners' performance on this noise-pair.

It seems possible that slightly altering the selection criteria for noise-pairs may yield an ordering that coincides better with incoherence detection. Thus, other possible stimulus sets for these two collections of noise-pairs were explored in Experiments 7B and 7C.

2.3 EXPERIMENT 7B: CRITICAL BANDWIDTH – TEMPORAL-AVERAGING

As a departure from trying to reproduce results from Chapter 1 with noise-pairs with a varied coherence, Experiment 7B attempted to reorder the noise-pairs by including temporal averaging to the interaural difference. This, in turn, has the possibility of having different noise-pairs chosen to make a different phase and level set.

2.3.1 Method

It is not apparent that the instantaneous fluctuations in IPD and ILD should be the best measure of incoherence detection performance. An alternative approach is to apply a temporal averaging window to the interaural differences so that large instantaneous peaks in IPD and ILD are smoothed. The filter applied is of the form

$$\hat{e}[f(t)] = \frac{\int_0^{T_D} f(t-t')e^{-t'/\tau}dt'}{\int_0^{T_D} e^{-t''/\tau}dt''} \quad (T_D < t), \tag{13}$$

where $f(t)$ is either the IPD or ILD. The time constant, τ, was set at 5 ms and the averaging window was terminated after the weight of the exponential function dropped to 0.1, which determined the upper limit of the integration T_D. Veimeister (1979) used a similar filter to describe monaural temporal modulation transfer functions and found that $\tau = 2.5$ ms was the optimal time constant. This experiment used $\tau = 5$ ms to exaggerate the effect of exponential time averaging. Also, longer integration times have been reported for binaural tasks when compared to monaural tasks. The effect, called "binaural sluggishness", reports integration times on the order of tens to hundreds of milliseconds (Grantham and Wightman, 1979; Grantham, 1982; Kollmeier and Gilkey, 1990; Boehnke *et al.*, 2002).

Figure 40 shows how the exponential averaging smooths brief instances of large IPDs and ILDs. In general, the values of s_t decrease when compared to the values in Experiment 7A. The same noise-pairs from the collections in Experiment 7A were used. However, they were reordered with respect to the new values of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ with the exponential averaging applied. The values of P_c and CAS were measured for two new phase sets and two new level sets. These sets were chosen in the same way as the previous experiments and the same procedure was employed. The same listeners from Experiment 7A participated in this experiment with the exception of Listener C and Z. Therefore, there were five listeners.

2.3.2 Results

Table 18 shows that there were three significant p-values of the ten possible for the two phase sets with P_c values. This increased to five of ten for the CAS values. The level set had only one significant p-value for P_c. This increased to three for CAS.

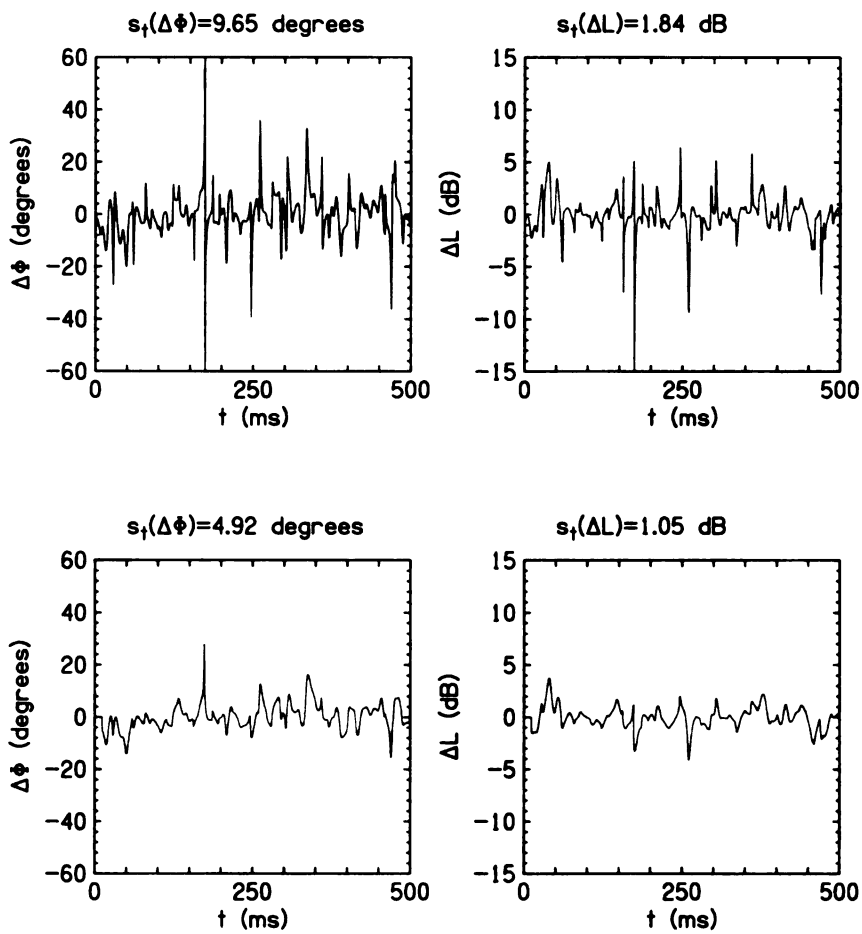

Figure 40: Interaural parameters with the exponential time averaging in Eq. 13. The weighting constant is $\tau = 5$ ms. The IPD and ILD are "smoother" as functions of time and the fluctuations are smaller.

Table 18: The p-values from a one-tailed t-test for 136-Hz bandwidth data with exponential averaging.

P_c	Phase	Phase	Level	Level
Listener	Set 1	Set 2	Set 1	Set 2
D	**0.014**	0.195	**0.010**	0.131
E	**0.019**	0.129	0.581	0.467
M	0.078	0.178	0.898	0.472
T	0.069	0.278	0.069	0.525
W	**<0.001**	0.076	0.171	0.070

CAS	Phase	Phase	Level	Level
Listener	Set 1	Set 2	Set 1	Set 2
D	**0.001**	0.159	**0.002**	0.100
E	**0.038**	0.129	0.220	0.197
M	0.060	0.142	0.350	0.143
T	**0.049**	0.256	**0.039**	0.378
W	**0.002**	**0.045**	0.307	**0.041**

2.3.3 Discussion

The results from this experiment were remarkably similar to those in Experiment 7A. Overall, only 40% of the p-values were significant for the *CAS* values. This can be compared to the 42% for the sets chosen in Experiment 7A. Since the results of these two experiments are so similar, it appears that sets of ten reproducible noise-pairs are insensitive to a change in ordering like the one made for this experiment. It was hypothesized that a more drastic change in ordering will be necessary for this bandwidth.

2.4 EXPERIMENT 7C: CRITICAL BANDWIDTH – TI TRADING

2.4.1 Method

In another attempt to reorder the noise-pairs in a sequence that better describes a listener's ability to distinguish incoherent from coherent noises, Experiment 7C used temporal averaging and time-intensity trading to form an auditory image position variable.

Hafter (1971) suggested the use of time-intensity trading to describe MLD data. Such a model suggests that a single image position is used for incoherence detection instead of two separate cues from $\Delta\Phi$ and ΔL. The form of time-intensity trading used in this experiment is

$$z(t) = \Delta\Phi(t) + c \cdot \Delta L(t) \tag{14}$$

where c is the time-intensity trading ratio. Yost and Hafter (1987) calculated $c =$ 10 degrees/dB. This experiment used a value of $c = 6$ degrees/dB because of the intuition of the experimenter at the time. Also, exponential temporal averaging was included in this experiment. The integration time constant was again $\tau = 5$ ms. The lateral position variable, $z(t)$, can be seen as a function of time in Figure 41.

The noise-pairs were selected from the same two collections in Experiments 7A and 7B. Two "time-intensity trading" sets were chosen for this experiment. The listeners were the same as those in Experiment 7B.

2.4.2 Results

The time-intensity trading set from Collection 1 led to one significant p-value for both the P_c values and CAS values, seen in Table 19. This is in contrast to the time-intensity trading set from Collection 2 that led to three significant p-values for

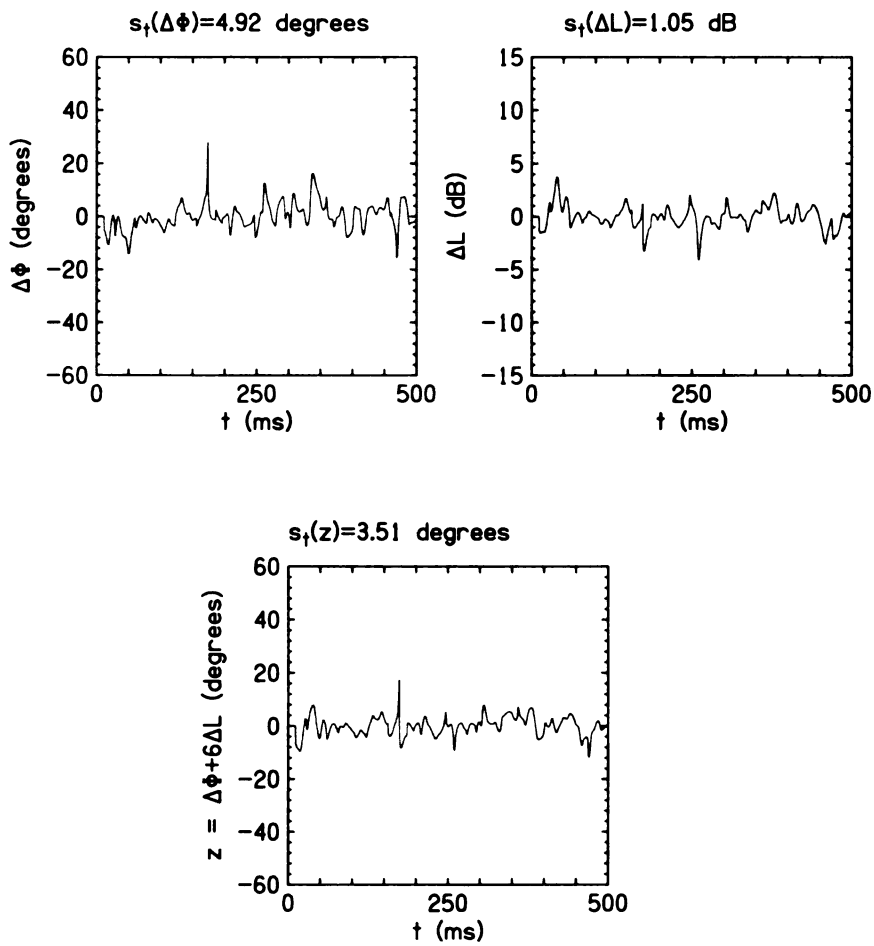

Figure 41: Interaural parameters combined with a 6 degree/dB time-intensity trading ratio to compute a time-dependent position of the auditory image. Exponential averaging is again included with a time constant of $\tau = 5$ ms.

P_c and four for *CAS* out of five possible tests.

Table 19: The p-values from a one-tailed t-test for 136-Hz bandwidth time-intensity trading data.

P_c Listener	Trading Set 1	Trading Set 2
D	**0.022**	**0.020**
E	0.440	0.075
M	0.357	**0.001**
T	0.553	0.097
W	0.276	**<0.001**

CAS Listener	Trading Set 1	Trading Set 2
D	**0.015**	**0.009**
E	0.415	**0.012**
M	0.160	**0.012**
T	0.422	0.069
W	0.316	**<0.001**

2.4.3 Discussion

Experiment 7C shows that the set chosen from Collection 2 had many significant p-values. However, the set chosen from Collection 1 did not. Considering both sets were selected in the same way, it appears that it was just coincidence that one set had many significant p-values.

When comparing the number significant p-values between Experiments 7A, 7B, and 7C, it appears that incoherence detection is more sophisticated that the most basic intuitions of the experimenter. However, there were always significant p-values in the sets chosen for each experiment. Therefore, it appears that fluctuations of interaural parameters (used separately or used when combined into a single auditory image) mediate incoherence detection at this bandwidth. Experiment 18 of Chapter 5 will discuss modeling the data from Experiments 7A, 7B, and 7C.

2.5 EXPERIMENT 8: WIDE BANDWIDTH

2.5.1 Method

Experiment 8 used a wideband, similar to Experiment 3 except that the bandwidth was narrowed, 636 Hz, when compared to the 2394-Hz bandwidth used in Chapter 1. The center frequency was 500 Hz and the coherence was 0.9922. The value of coherence had some variation because the noise-pairs were truncated from 1 s to 0.5 s. However, the coherence did not appreciably vary because of the large number of random variables in the wideband case. The range of coherence for the truncated noise-pairs was 0.9921–0.9923. Spectral components between 200 and 800 Hz had unity amplitude and components in the ranges 150–200 and 800–850 Hz were shaped by the raised-cosine window. (Neither exponential averaging nor time-intensity trading was used to order these stimuli.)

Figure 42 shows the distribution of interaural phase and level fluctuations for one-hundred 636-Hz bandwidth noises. The corresponding values for Experiments 6 and 7 (collection 1 only) are shown by the black dots. The means of the distribution remained about the same as in Experiments 6 and 7 in that the ensemble average phase fluctuation was about 13 degrees and the level fluctuation was about 1.7 dB. The standard deviations of the phase and level fluctuations decreased dramatically when the bandwidth was increased to 636 Hz, as shown by comparisons of the σ values in Figures Figure 42 and 21, respectively 1.34 degrees and 0.15 dB. The correlation between level and phase fluctuations decreased to 0.48.

2.5.2 Results

With five listeners and two sets there were ten possible significance tests, and Table 20 shows that three of them led to a significant difference at the 0.05 level for P_c; five for *CAS*.

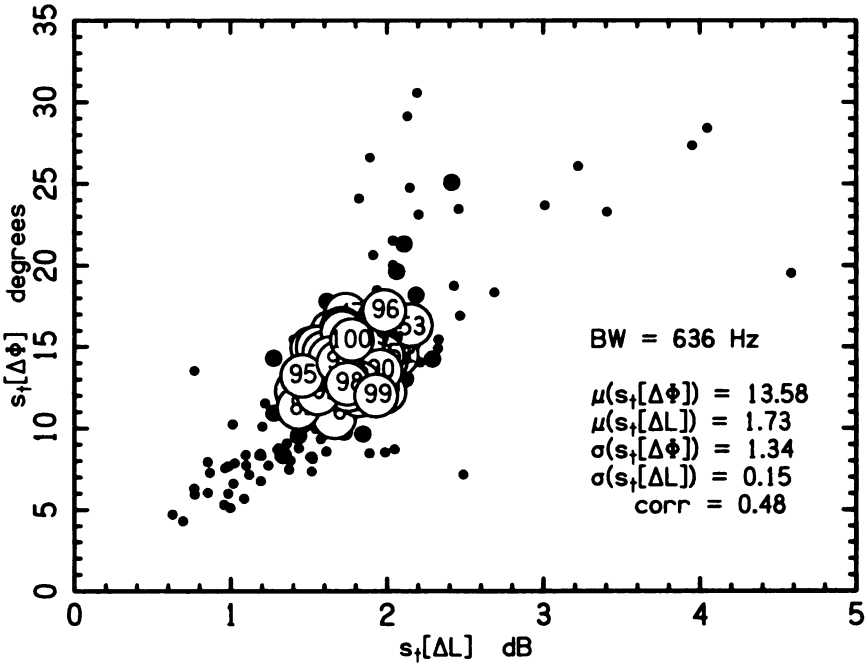

Figure 42: Fluctuations of IPD versus fluctuations of ILD for the first collection of 100 reproducible noise-pairs having a 636-Hz bandwidth, as used in Experiments 3. Each noise-pair is labeled by a serial number indicating only the order of creation. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. The small black dots represent the 14-Hz bandwidth noise-pairs from the collection in Experiment 6. The large black dots represent the 136-Hz bandwidth noise-pairs from Collection 1 in Experiment 7A.

Table 20: The p-values for t-tests on the 636-Hz bandwidth data.

P_c	Phase	Level	CAS	Phase	Level
Listener	Set	Set	Listener	Set	Set
C	0.860	**0.048**	C	0.860	**0.048**
E	**0.023**	0.231	E	**0.018**	0.201
M	**0.047**	0.053	M	**<0.001**	**0.045**
W	0.109	0.952	W	0.131	0.991
Z	0.440	0.107	Z	**0.044**	0.093

Table 21: Comparison of p-values and inter-listener correlations for Experiment 3 (fixed value of coherence with a 2394-Hz bandwidth) and Experiment 8 (varied values of coherence with a 636-Hz bandwidth), the wideband experiments.

P_c	Experiment 3 Fixed ρ	Experiment 8 Varied ρ
# of sig. p-values	1/8 (12.5%)	3/10 (30%)
Ave. phase set corr.	0.208	0.241
Ave. level set corr.	–0.106	0.143
CAS	Experiment 3 Fixed ρ	Experiment 8 Varied ρ
# of sig. p-values	0/8 (0%)	5/10 (50%)
Ave. phase set corr.	0.252	0.348
Ave. level set corr.	–0.105	0.026

Table 21 shows the comparison of the wideband experiments, Experiments 3 and 8. The varied coherence noise-pairs with a smaller bandwidth showed 30% of the p-values to be significant for the P_c data and 50% for the CAS data. Like Experiments 6 and 7A, average listener correlations are higher for the Experiment 8 noise-pairs when compared to the Experiment 3 noise-pairs.

2.5.3 Discussion

In this experiment, it was interesting that the phase set had three significant p-values for the CAS data and the level set had two significant p-values whereas Experiment

3 has no significant p-values for the *CAS* data. The reason for this difference is that, unlike Experiment 3, the bandwidth of Experiment 8 has not yet reached the wideband limit, even though it is much larger than a critical bandwidth centered at 500 Hz.

2.6 EXPERIMENT 9: THE ROLE OF BANDWIDTH

Experiment 4 started with two observations: 1) the ranges of fluctuations in IPD and ILD become narrower and 2) the ability of listeners to detect incoherence depends less on the individual noise-pairs and is better determined by the value of coherence itself. It was hypothesized that the second effect is the direct result of the first, and the main effect of a variation in bandwidth is to alter the distributions of interaural variances. This was shown in Experiment 4 to be the case. Experiment 9 attempts to reproduce the results of Experiment 4.

2.6.1 Method

As for Experiment 4 in Chapter 1, a subset of noise-pairs from the collection with a bandwidth of 14 Hz was assembled to make a "matched set" whose members were selected to best match the noise-pairs from Collection 1 of Experiment 7A, which had a bandwidth of 136 Hz. The first collection of Experiment 7A included 20 noise-pairs, ten for the phase set and ten for the level set, as determined by the five largest and five smallest fluctuations. However, phase and level fluctuations tend to be correlated and three of the noise-pairs were common to the phase and level sets. Therefore, there were only 17 different noise-pairs in the first collection of Experiment 2A. For each of these 136-Hz bandwidth noise-pairs, a 14-Hz bandwidth noise-pair was selected that best matched the fluctuations in phase and level. The selection is illustrated by the 17 open and filled circles in Figure 43. Unlike Experiment 4, new noise-pairs were not generated to best match the values of the fluctuations in phase and level. Instead,

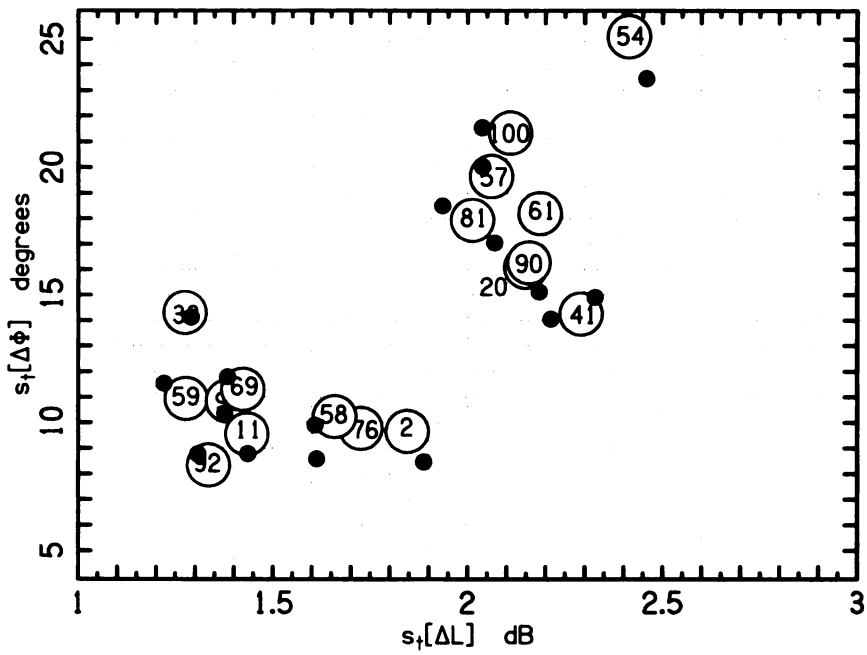

Figure 43: The matched set for Experiment 9. The 17 noise-pairs from phase set 1 and level set 1 in Experiment 7A were matched in fluctuations of IPD and ILD by the closest noise-pairs from Experiment 6, which are shown as black dots.

noise-pairs from the 14-Hz bandwidth collection from Experiment 6 were used to match the 136-Hz bandwidth noise-pairs. Listeners E, M, W, and Z participated.

2.6.2 Results

Using the working hypothesis of Experiment 4, the variation in detectability within the matched 14-Hz sets from Experiment 9 ought to be identical to the variation in detectability within the 136-Hz set from Experiment 7A. The p-values appear in Table 22. There, it can be seen that three of eight (37.5%) p-values are significant at the 0.05 level for the 14-Hz matched set for the P_c data. This can be compared to the 8 of 14 (57.1%) p-values that are significant for the P_c data for the phase set and level set from Collection 1 of Experiment 7A. The number of significant p-values increases to six of eight (75%) for the matched set *CAS* data. This can be compared to the 8 of 14 (57.1%) p-values that are significant for the *CAS* data for the phase set and

level set of Experiment 7A as seen in Table 16. Thus, the matched phase and level sets appear to have approximately matched the relative difference in performance between the noise-pairs in the Experiment 7A sets when compared to Experiment 6. Experiment 4 also approximately matched the difference in detection performance between the narrow and critical bandwidth.

Figures 44 and 46 show the P_c data for the four listeners in Experiment 9. Listeners M and Z have data that correlate amazingly well across bandwidths. However, they are near the ceiling for all the points, and differences between different bandwidth data would be difficult to see. Listeners E and W show that detecting incoherence is much easier for the 14-Hz noise-pairs than for the 136-Hz noise-pairs. This was also seen in Experiment 4 of Chapter 1, where Listeners P and W performed consistently better for the narrowband noise-pairs.

The slopes of the *CAS* values in Figures 45 and 47 are about the same for the two bandwidths. However, listeners' *CAS* were consistently higher for the 14-Hz matched sets, with only eight exceptions from the 60 possible comparisons.

Table 22: The p-values for the matched sets.

P_c	Phase	Level	*CAS*	Phase	Level
Listener	Set	Set	Listener	Set	Set
E	**0.027**	0.148	E	**0.011**	0.235
M	0.059	0.296	M	**0.010**	0.116
W	**0.019**	**0.024**	W	**0.017**	**0.009**
Z	0.077	0.064	Z	**<0.001**	**0.003**

Table 23 shows the comparison of the matched set experiments, Experiments 4 and 9. The percentage of significant p-values is too large for the P_c data (25% compared to 37.5%) and for the *CAS* data (37.5% compared to 75%) when comparing the fixed coherence sets to the varied coherence sets. However, all of these percentages are

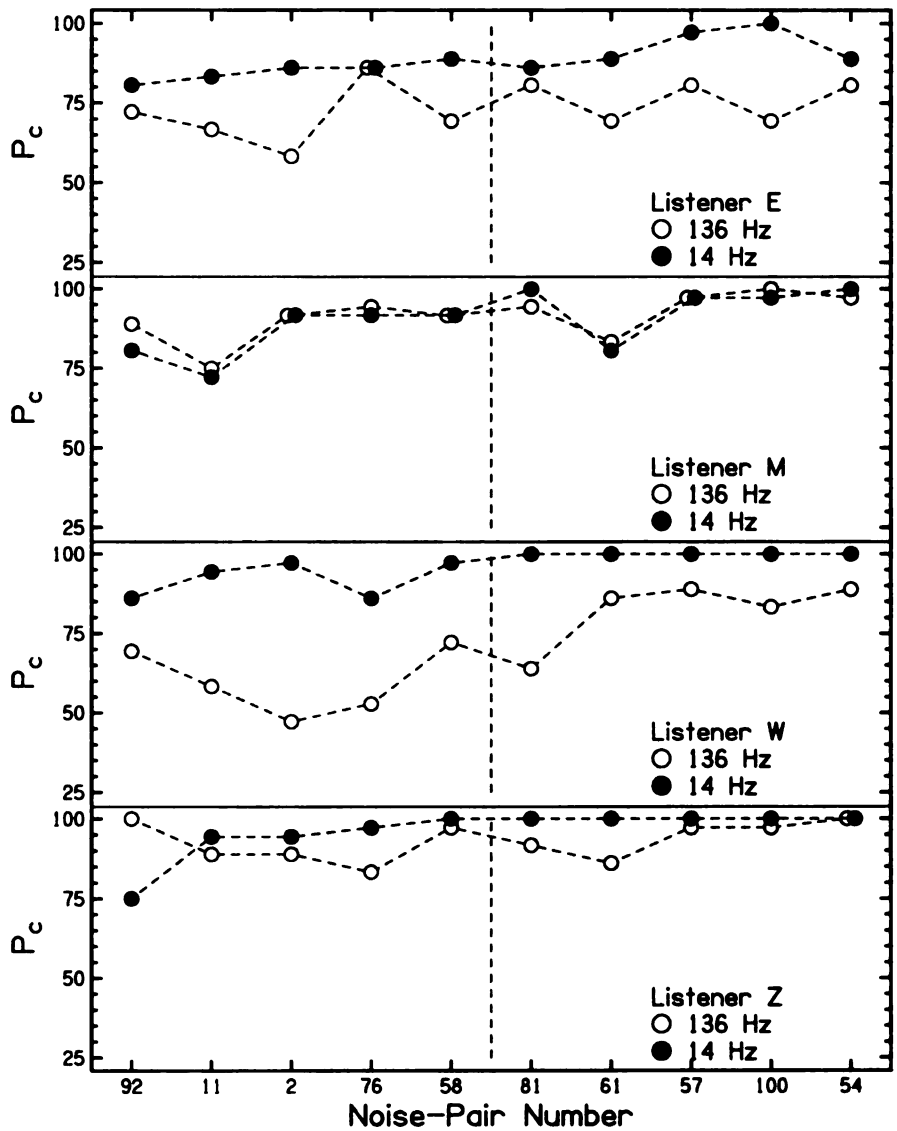

Figure 44: Matched phase set P_c data. The numbers on the horizontal axis correspond to noise-pairs chosen in Experiment 7A.

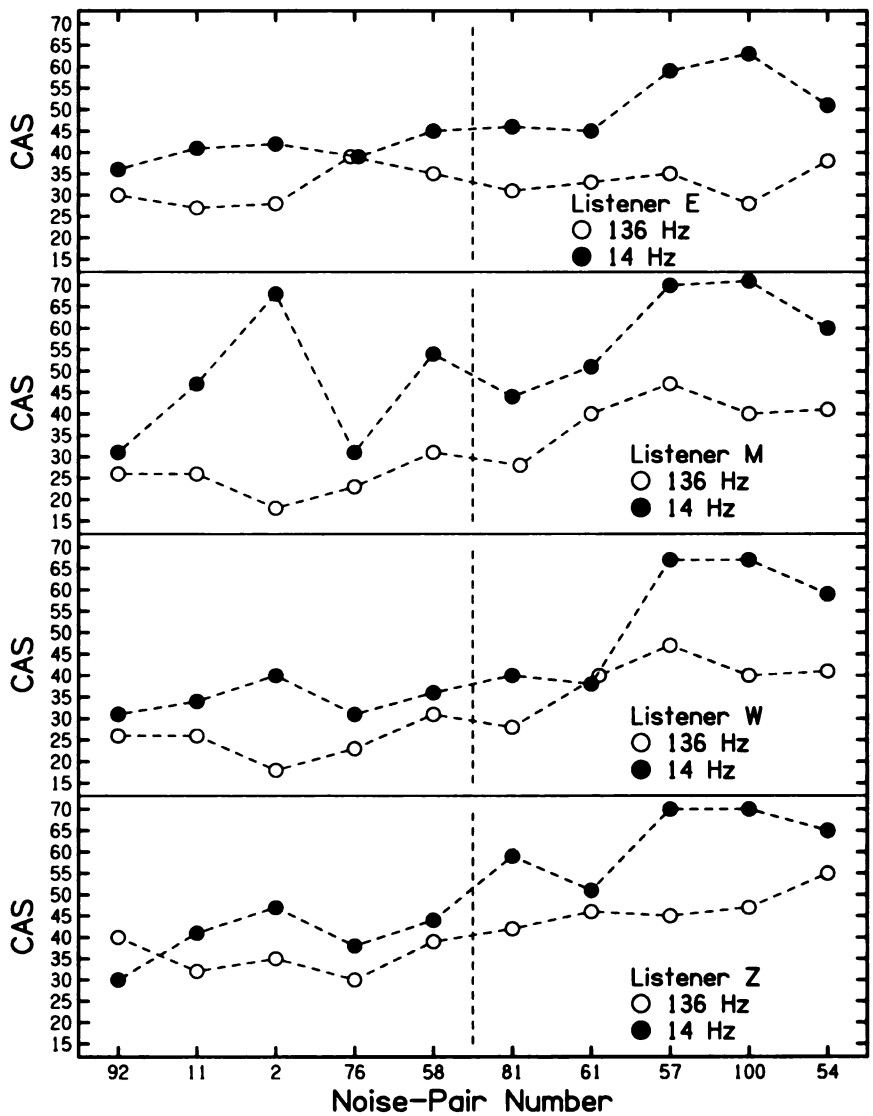

Figure 45: Matched phase set *CAS* data. The numbers on the horizontal axis correspond to noise-pairs chosen in Experiment 7A.

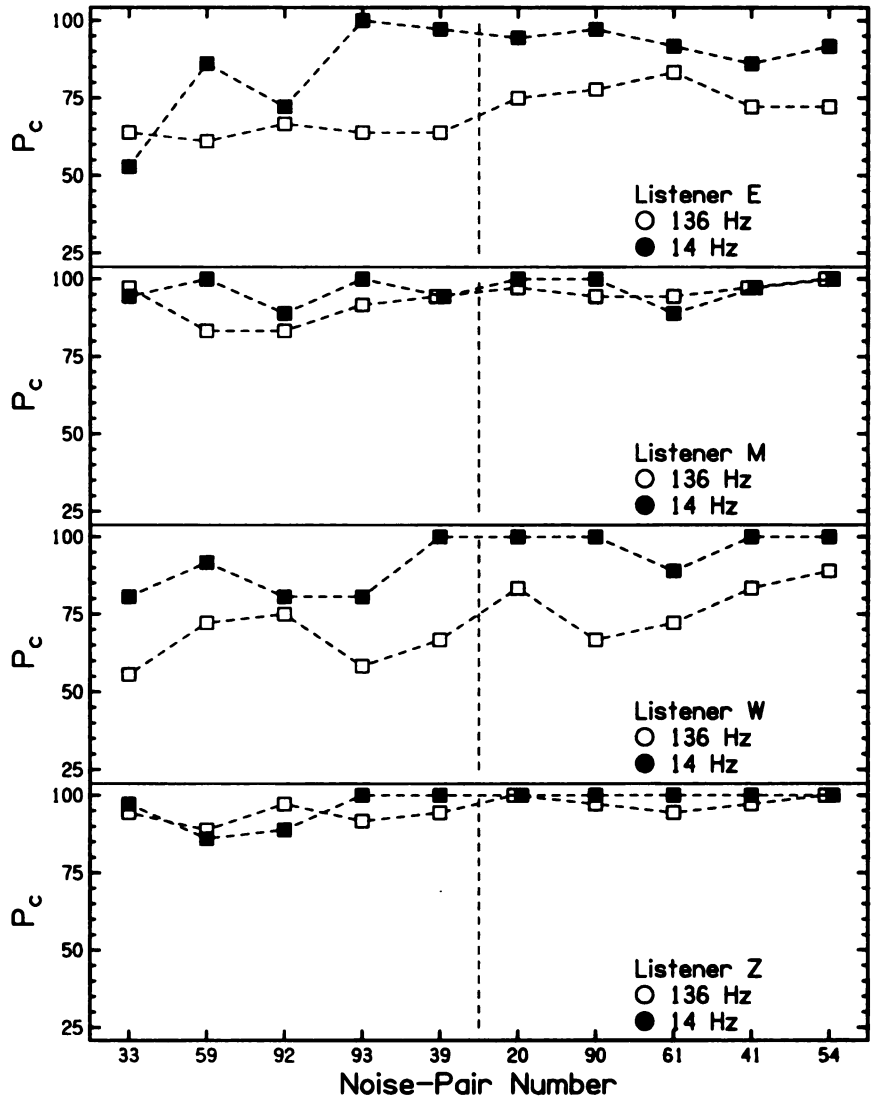

Figure 46: Matched level set P_c data. The numbers on the horizontal axis correspond to noise-pairs chosen in Experiment 7A.

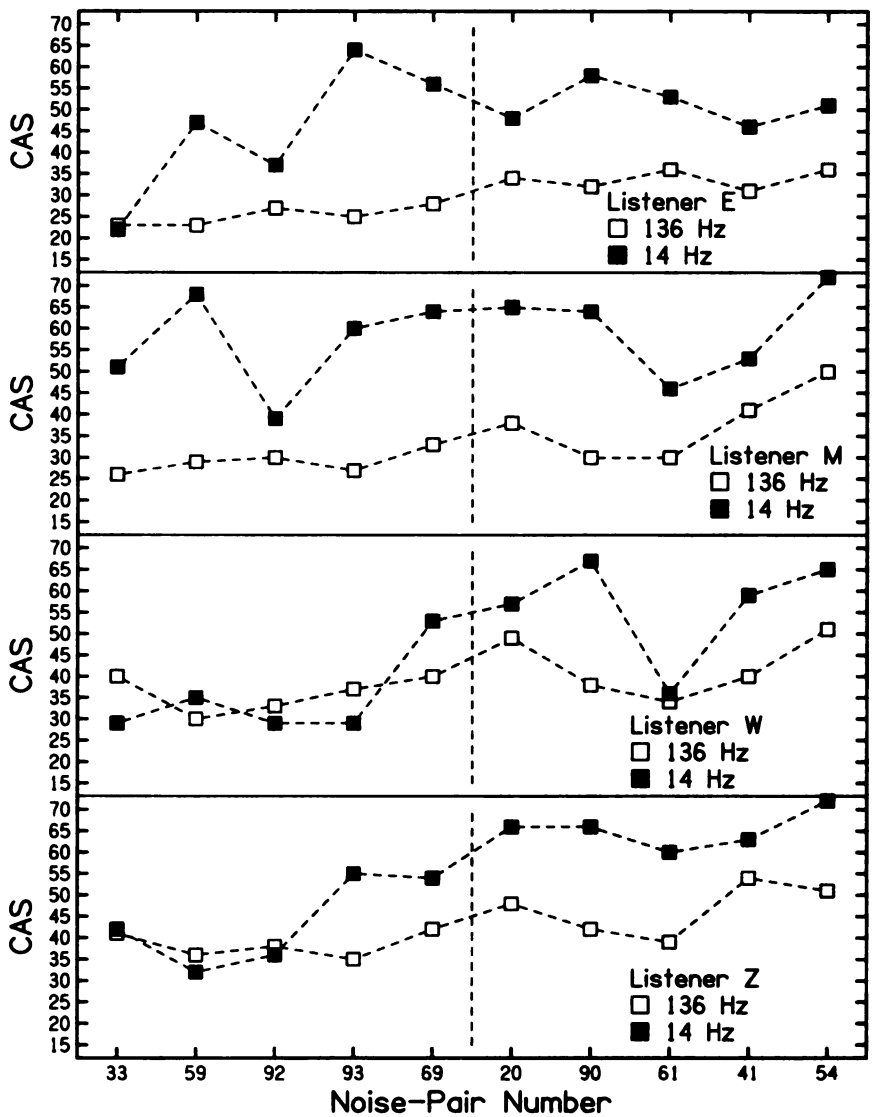

Figure 47: Matched level set *CAS* data. The numbers on the horizontal axis correspond to noise-pairs chosen in Experiment 7A.

Table 23: Comparison of p-values and inter-listener correlations for Experiment 4 (fixed) and Experiment 9 (varied), the matched set experiments.

P_c	Experiment 4 Fixed ρ	Experiment 9 Varied ρ
# of sig. p-values	2/8 (25%)	3/8 (37.5%)
Ave. phase set corr.	0.710	0.573
Ave. level set corr.	0.491	0.402
CAS	Experiment 4 Fixed ρ	Experiment 9 Varied ρ
# of sig. p-values	3/8 (37.5%)	6/8 (75%)
Ave. phase set corr.	0.771	0.912
Ave. level set corr.	0.628	0.579

smaller than the 100% seen in Experiments 1 and 6, the experiments that also used 14-Hz bandwidth noise-pairs. Like previous experiments, average listener correlations are smaller for three of the four comparisons for the varied coherence noise-pairs when compared to the fixed coherence noise-pairs.

2.6.3 Discussion

Experiment 9 succeeded in approximately matching the number of significant p-values and level of significance for the critical bandwidth noise-pairs in Experiment 7A by using noise-pairs from Experiment 6. This result also occurred in Experiment 4, where the number of significant p-values and level of significance for the critical bandwidth noise-pairs were matched by using narrow bandwidth noise-pairs. In Experiment 9, the inter-listener correlation was higher for the 14-Hz matched phase set compared to the matched level set and the 136-Hz phase and level sets, which was also seen in Experiment 4. As postulated in the discussion of Experiment 4, the higher inter-listener correlations may be due to the slowness of the fluctuations with this small bandwidth.

One difference between this experiment and Experiment 4 is that performance was

consistently higher on the 14-Hz bandwidth noise-pairs compared to the 136-Hz noise-pairs for some of the P_c data and almost all of the *CAS* data. Experiment 4 found two listeners that performed better for *CAS* and two listeners that had comparable performance for *CAS*.

2.7 EXPERIMENT 10: EFFECT OF MONAURAL CUES

To test the results of Experiment 5, another numerical study of interaural fluctuations was done to compare the interaural fluctuations in noise-pairs with the envelope fluctuations in the left-ear signal itself.

To make the comparison, the envelope fluctuation was calculated per Eq. 12. As stated in Experiment 4, calculations with 1000 noise-pairs with 14-Hz bandwidth - stochastically identical to pairs of Experiment 1 - showed a positive correlation. The correlation between $s_t[E]$ and $s_t[\Delta L]$ was 0.48. The correlation between $s_t[E]$ and $s_t[\Delta\Phi]$ was 0.43.

The same calculations were made for the 20 noise-pairs actually used in Experiment 6. The interaural phase fluctuations and the interaural level fluctuations correlated with the monaural envelope fluctuations at levels of 0.67 and 0.72 respectively. These correlations are higher than those reported in Experiment 5. In the fixed-coherence experiment, the interaural phase fluctuations and the interaural level fluctuations correlated with the monaural envelope fluctuations at levels of 0.59 and 0.65 respectively.

2.7.1 Diotic experiment method

Experiment 10 was identical to Experiment 6 (the 14-Hz noise-pairs with varied coherence) except for the important difference that the left-ear signal of Experiment 1 was the signal for both ears in Experiment 10. Experiment 10 was identical to Experiment 5, except that different noises were used. The stimulus sets were the phase

set and level set from Experiment 6. Listeners E, M, and W participated.

2.7.2 Results

Like Experiment 5, listeners made a negligible number of "confident" responses. Again, the *CAS* had little value and results of Experiment 10 were plotted only in terms of the percentage of the trials on which a given noise was selected over other noises, called P_s. It will again be compared with P_c, the percent correct in the dichotic experiments.

1. Large Fluctuation Comparison

Particular interest centered on the five noise-pairs for which the interaural (and monaural) fluctuations were the greatest. The average values of P_s for listeners E, M, and W were, respectively:

For the phase set: 45, 54, 57 %

For the level set: 48, 62, 58 %.

The highest value of P_s ever seen (among 30) was 78%. These numbers can be compared with the values of P_c in the dichotic experiment which averaged 97%. The lowest value of P_c ever seen (among 30) was 78%.

Experiment 5 also showed small values of P_s compared to P_c. Listeners M and W had mostly higher values of P_s in Experiment 5. For the phase set, Listener M had $P_s = 61\%$ and Listener W had $P_s = 63\%$. For the level set, Listener M had $P_s = 58\%$ and Listener W had $P_s = 70\%$. The reason for the higher values of P_s is probably due to the higher correlation of monaural fluctuations to interaural fluctuations for Experiment 10 when compared to Experiment 5.

2. Agreement between listeners

The agreement among the listeners was assessed by comparing values of P_s against noise serial number for listeners taken in pairs. Inter-listener correlations for E-M, M-W, and E-W were as follows:

For the phase set: –0.07, 0.76, –0.23

For the level set: –0.33, 0.85, –0.35.

Evidently there was strong correlation between M and W but anticorrelation with listener E. Experiment 5 also showed both positive and negative inter-listener correlations. The two common listeners, M and W, showed a strong correlation in Experiment 5, 0.88 for the phase set and 0.89 for the level set.

3. Comparison with envelope fluctuation

A comparison between the listener selection of noises and the fluctuations was assessed by comparing P_s with $s_t[E]$ as a function of the noise-pair serial number. Correlations for listeners E, M, and W were, respectively:

For the phase set: –0.33, 0.67, 0.72

For the level set: 0.22, 0.81, 0.46

Like Experiment 5, Listeners M and W had high correlations between P_s and $s_t[E]$. For the phase set, Listener M had a correlation of 0.84 and Listener W had a correlation of 0.88. For the level set, Listener M had a correlation of 0.66 and Listener W had a correlation of 0.70.

4. Comparison with Experiment 6

A comparison between the results of the corresponding diotic and dichotic experiments was made by comparing P_s on Experiment 10 with P_c on Experiment 6, both as functions of the noise-pair serial number. Correlations of P_s and P_c for listeners E, M, and W were, respectively:

For the phase set: –0.62, 0.58, 0.67

For the level set: –0.38, 0.85, 0.45.

Like Experiment 5, Listeners M and W had high correlations between P_s and P_c. For the phase set, Listener M had a correlation of 0.53 and Listener W had a correlation of 0.44. For the level set, Listener M had a correlation of 0.66 and Listener W had a correlation of 0.80.

2.7.3 Discussion

The above correlations for Listeners M and W are fairly impressive. These include the correlations between P_s in Experiment 10 and P_c in Experiment 6 as well as the correlations between P_s and monaural and dichotic fluctuations. Listeners M and W are also impressively consistent between Experiments 5 and 10. There appears to be no contradiction between this experiment and the one that used fixed coherence noise-pairs.

2.8 EXPERIMENT 11: CONTEXTUAL EFFECTS

During the course of the investigations of incoherence detection with reproducible noises, the question of contextual effects arose. Would it be the case that a particular noise-pair would yield a different level of detection if it was presented against different noise-pairs than those chosen for the phase and level sets? It was seen in Experiments 7A, 7B, and 7C that certain noise-pairs kept their individual detection tendencies. For example, it was always difficult for listeners to detect incoherence in noise-pair #192 independent of which noise-pairs were presented coherently against it.

2.8.1 Method

The procedure was the same as previous experiments, except that the noise-pairs were grouped in near-random contexts (not sequential), not phase and level sets. This was done by listening to all 100 noises in the collection from Experiment 6 (the 14-Hz noise-pairs with varied coherence). Therefore, ten sets of ten noise-pairs were heard in this experiment. The seventeen noise-pairs fell into seven different sets out of the possible ten sets.

The P_c and CAS values of this experiment were compared to the matched phase and level sets of Experiment 9. The matched set noise-pairs of Experiment 9 were used because the difference in the values of the fluctuations were not as large as

the noise-pairs in Experiment 6. A large difference between the largest and smallest fluctuations noise-pairs might be less sensitive to contextual effects. Listeners E, M, and W participated.

2.8.2 Results and Discussion

Figures 48–51 show the results for the contextual effects experiment. Correlations between the matched sets and the random sets are reported on the figures as ρ_{MR}. It can be seen that these correlations are high for the CAS data ($\geq$ 0.71 for the phase set, $\geq$ 0.64 for the level set). They are also high for Listeners E and W for the P_c data ($\geq$ 0.78 for the phase set, $\geq$ 0.64 for the level set). Listener M was near the ceiling for the data, which probably explains why his correlations were smaller than the other listeners (0.28 for the phase set, –0.04 for the level set). Considering the results of Experiment 5, it seems highly likely that the small changes in detectability for noise-pairs in different contexts is due entirely to monaural cues. In conclusion, considering the high values of ρ_{MR}, it seems that contextual effects play a only small role when listening to reproducible noise-pairs.

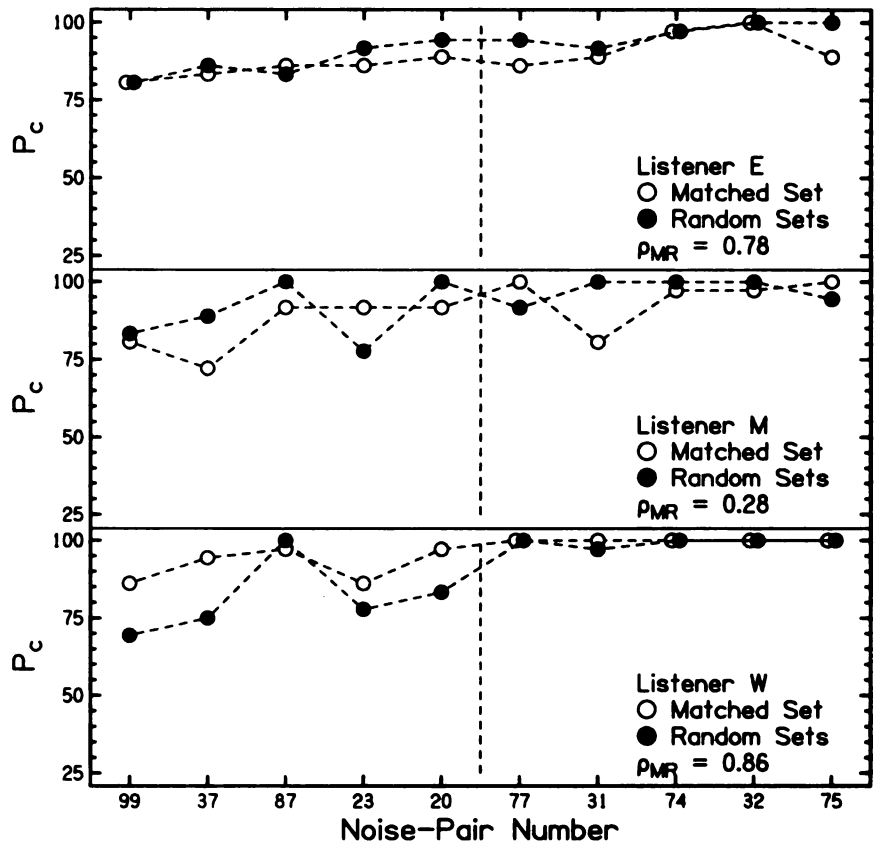

Figure 48: The P_c scores for the matched phase set selected from Experiment 9 with 14-Hz noise pairs are shown by open circles. They are plotted together with the data from this experiment, the random sets, which presented the same sounds in a different context, shown by the closed circles.

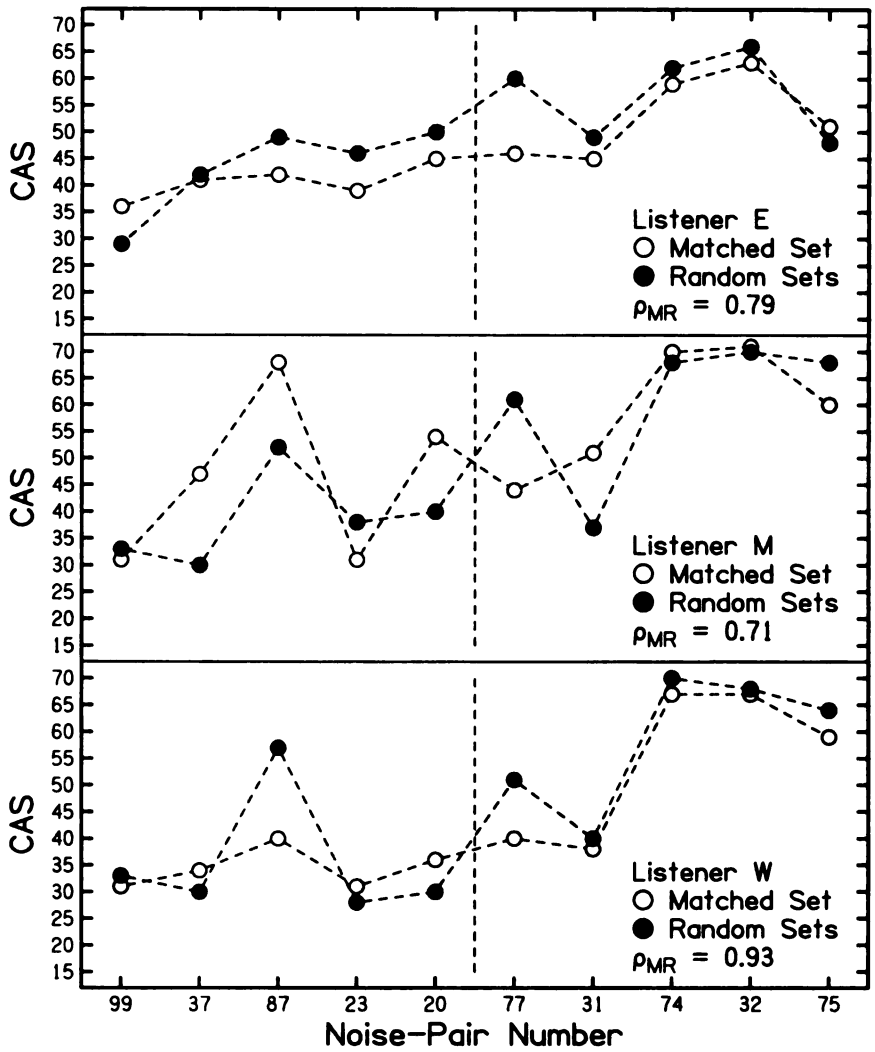

Figure 49: The confidence adjusted scores (CAS) for the matched phase set selected from Experiment 9 with 14-Hz noise pairs are shown by open circles. They are plotted together with the data from this experiment, the random sets, which presented the same sounds in a different context, shown by the closed circles.

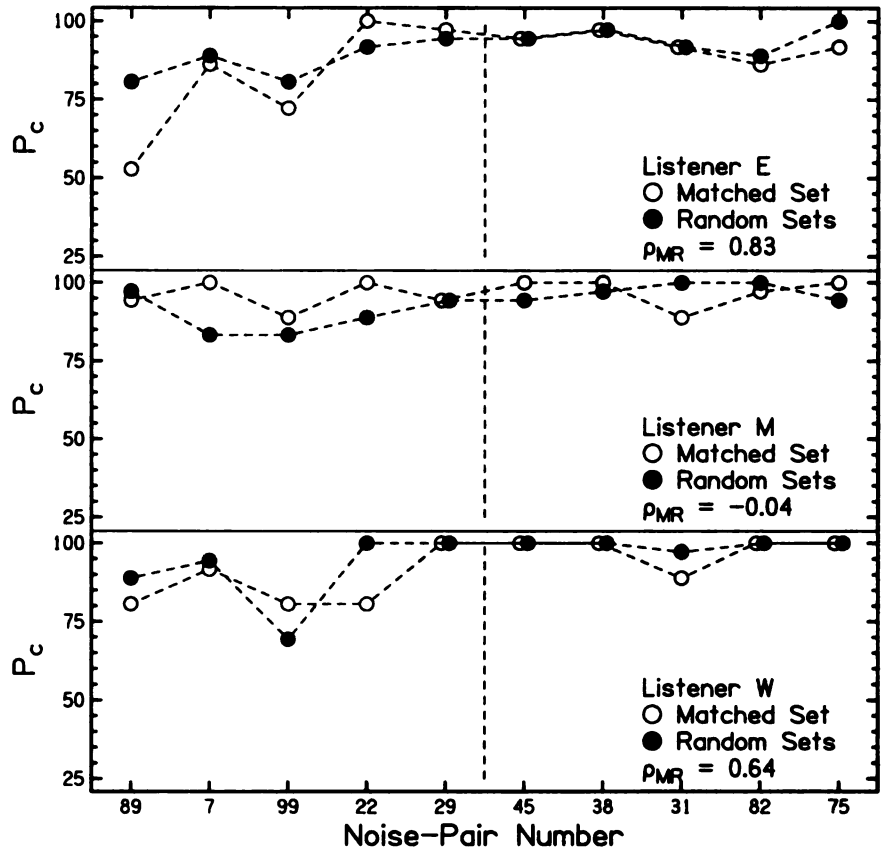

Figure 50: The P_c scores for the matched level set selected from Experiment 9 with 14-Hz noise pairs are shown by open circles. They are plotted together with the data from this experiment, the random sets, which presented the same sounds in a different context, shown by the closed circles.

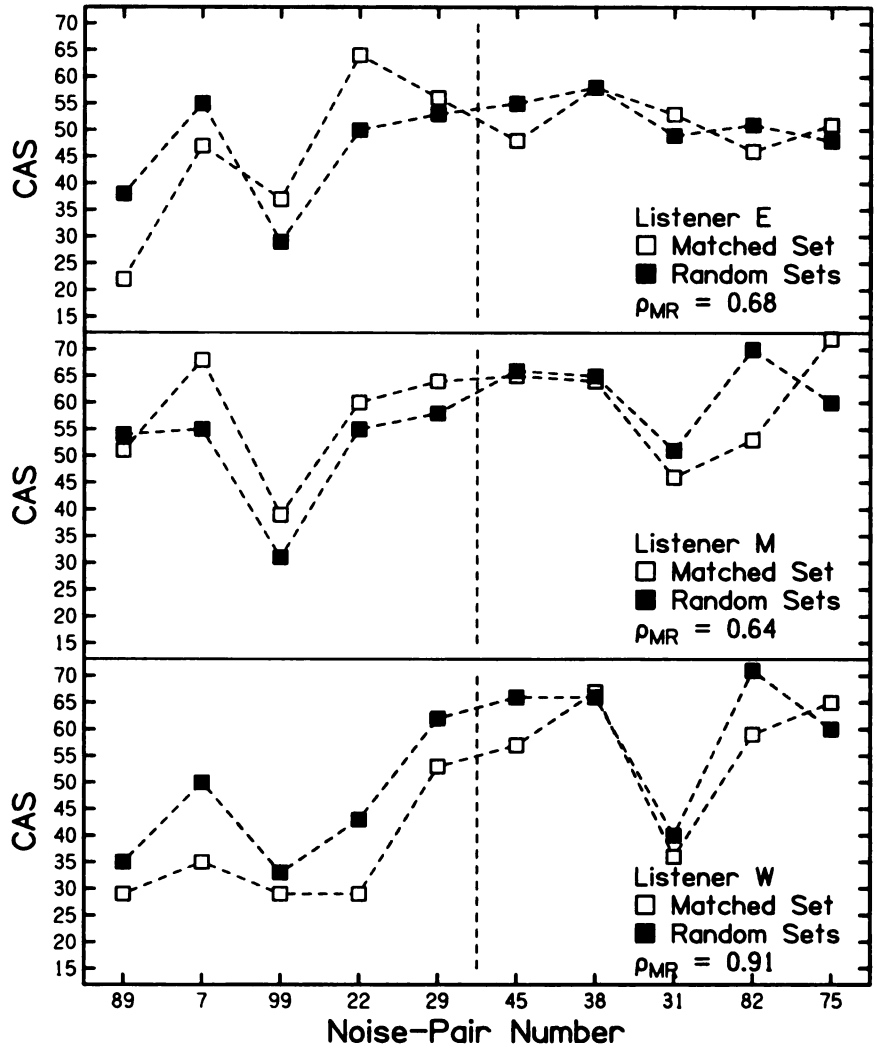

Figure 51: The confidence adjusted scores (CAS) for the matched level set selected from Experiment 9 with 14-Hz noise pairs are shown by open circles. They are plotted together with the data from this experiment, the random sets, which presented the same sounds in a different context, shown by the closed circles.

2.9 CONCLUSION

Experiments in this chapter attempted to reproduce the results of Chapter 1 with noise-pairs that had a coherence that was not necessarily fixed at 0.9922. The results of Experiments 1, 2, 4, and 5 were reproduced by Experiments 6, 7A, 9, and 10. Experiment 8 failed to reproduce the results of Experiment 3 because Experiment 8 had significant p-values and Experiment 3 did not. The reason for this failure is probably that the bandwidth of 636 Hz used in Experiment 8 was not large enough to reach the wideband limit. The wideband limit was reached in Experiment 3 using a bandwidth of 2394 Hz.

Some additional experiments were performed, like Experiments 7B, that used a temporal averaging filter, and 7C, that used a time-intensity trading variable to reorder the noise-pairs for different noise-pairs in the phase and level sets. It was found that these reorderings did not appreciably change the percentage or level of p-values. Therefore is is unclear how the fluctuations truly mediate incoherence detection. Chapters 4 and 5 will address this uncertainty in a more systematic way.

Experiment 11 studied the contextual effects on incoherence detection of reproducible noise-pairs. It was found that under two different contexts of listening to the same reproducible noise-pairs, scores correlated well both in P_c (when ceiling effects were not playing a role) and CAS. The reason that perfect correlation was not achieved was probably due to the monaural cues (see Chapters 5 and 10) that occurred during the experiment because the coherent noise-pairs were presented randomly against an incoherent noise-pair.

3 THE EFFECTS OF DURATION ON INCOHERENCE DETECTION

Reproducible noise-pairs in Chapters 1 and 2 were 500 ms in duration. The purpose of this chapter is explore the effects of duration on incoherence detection. Specifically, it will try to determine whether the value of the coherence over small durations, like 100 ms, mediates incoherence detection. Therefore, the results of this chapter will have implications for the validity of results from Chapters 1 and 2, which found that fluctuations in interaural parameters mediated incoherence detection, not the value of the coherence function.

3.1 EXPERIMENT 12: NARROW BANDWIDTH

3.1.1 Stimuli

Three collections of 100 noise-pairs with 14-Hz bandwidth were created for Experiment 12. The three collections were different from each other because they had different durations - 25, 50, and 100 ms.

As for the noise-pairs from Chapter 1, each noise-pair was constructed from equal-amplitude random-phase components that spanned a frequency range of 490–510 Hz with a frequency spacing of 2 Hz. Components between 495–505 Hz had equal amplitudes of unity. Frequencies below 495 Hz and above 505 Hz were attenuated with a raised-cosine window. Consequently there were nine components in the stimuli that had non-zero amplitudes. The 3-dB bandwidth was 14 Hz. An orthogonalization procedure guaranteed that the interaural coherence of each noise-pair was precisely 0.9922.

Noise stimuli originally had a duration of 500 ms like those in Chapter 1. The noise-pairs in this chapter were different from those in Chapter 1 because these noise-pairs had a raised-cosine window applied with a 10-ms rise-fall time and a total

non-zero duration of 25, 50, or 100 ms. After the temporal shaping, the value of the coherence was checked. Noise-pairs were only accepted if the value of the coherence was 0.992 ± 0.001. For the 100 accepted 25-ms noise-pairs 1852 noise-pairs were rejected, 1575 were rejected for 50-ms noise-pairs, and 12112 were rejected for 100-ms noise-pairs. This requirement on the value of coherence is more lenient than used in the Chapter 1 (± 0.0001 was used before). However, from the result of the experiments in Chapter 2, it was found that a small change in coherence should not cause a change in the detection data. This is because the noise-pairs in Chapter 2 with varied values of coherence reproduced the results of Chapter 1 with a fixed value of coherence.

The goal of the stimulus generation technique was to pack a number of components into a narrow band to create a noise with complicated fluctuations, and yet it is evident that there is a formal inconsistency in the technique as described. A stimulus 100 ms long is correctly synthesized with components that are separated by 10 Hz, and a stimulus that is 25 ms long implies components that are separated by 40 Hz, assuming periodic waveforms and an 8 kHz sample rate. Neither is consistent with nine components separated by 2 Hz, for an overall bandwidth of 14 Hz. The result of the inconsistency is that noise stimuli had larger bandwidths than 14 Hz. Actual bands computed using the synthesis technique, including the temporal envelope, are shown in Figure 52 for the phase sets and Figure 53 for the level sets.

Averaging the bar lengths in Figures 52 and 53 show that, for a 100-ms noise, 90 percent of the energy is contained in a band 24 Hz (±11) wide and 99 percent is contained in a band 76 Hz wide (±22). For a 50-ms noise, the 90 percent and 99 percent bandwidths are 39 (±8) Hz and 106 (±16) Hz. For a 25-ms noise, the corresponding bandwidths are 74 (±4) Hz and 126 (±16) Hz.

Figures 52 and 53 show that the synthesis technique enables the complexity of multiple components while maintaining a reasonably narrow bandwidth. For example,

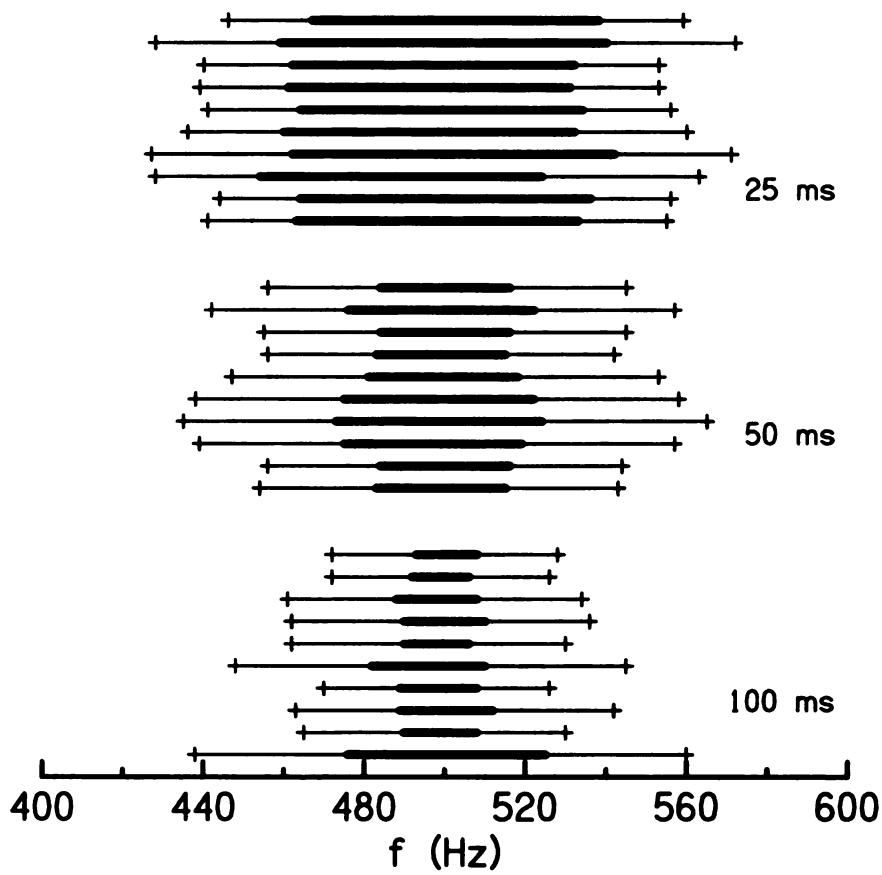

Figure 52: For each of three stimulus durations, the waveforms in the phase sets were computed using the synthesis technique of the stimuli section, Section 3.1.1. The thin horizontal lines show bands that contain 99 percent of the noise energy because the + signs at the ends of each line are drawn at the 0.5 and 99.5 percent points for cumulative energy with increasing frequency. The heavy horizontal lines show the bands that contain 90 percent of the energy because the lines end at the 5% and 95% points for cumulative energy.

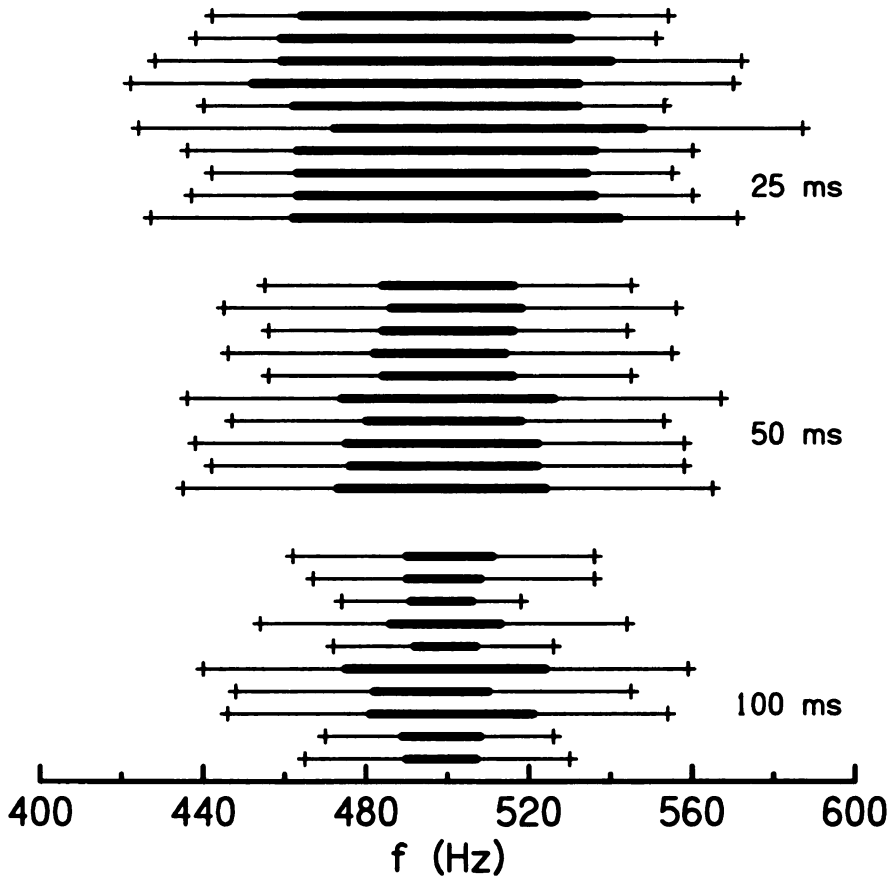

Figure 53: For each of three stimulus durations, the waveforms in the level sets were computed using the synthesis technique of the stimuli section, Section 3.1.1. The thin horizontal lines show bands that contain 99 percent of the noise energy because the + signs at the ends of each line are drawn at the 0.5 and 99.5 percent points for cumulative energy with increasing frequency. The heavy horizontal lines show the bands that contain 90 percent of the energy because the lines end at the 5% and 95% points for cumulative energy.

99 percent of the energy is found in a band that is 126 Hz wide for a duration of 25 ms. By contrast, a formally correct synthesis would require a bandwidth of 320 Hz, more than twice as wide.

Although the inconsistency in the technique results in a bandwidth that is larger than the nominal bandwidth, the inconsistency does not affect the precise interaural coherence because it does not affect the orthogonality. The inconsistency has a potential affect on the fluctuation statistics for interaural level and phase because it changes the computation of the Hilbert transform of the signal.

The real noise-pairs, as presented to the listeners, are computed by adding cosines. The Hilbert transforms of the left-right noise-pairs are computed by adding the corresponding sines. The noise-pairs and their Hilbert transforms are the real and imaginary parts of the analytic signals. When the time waveforms have a duration that is consistent with the frequency spacing in the spectrum the real and imaginary parts have similar properties. The overlap integral between left channel and initial right channel is the same for real and imaginary parts, and so the orthogonality computation involves the same weighting. Also, when the orthogonalized right-channel waveform is rescaled to have the same power as the left-channel waveform, the real and imaginary parts have the same scale factor because the total energy is the same in both parts. These two common features are lost when the synthesis is inconsistent. To ameliorate this problem and to make the imaginary part of the waveform most closely resemble the Hilbert transform of the real part, the overlap integral and power scaling factors for the imaginary part were taken to be those for the real part. The resulting inconsistencies in the imaginary part were computed. The computation shows that when the durations are 100 or 50 ms, the power in the imaginary parts may differ by a few percent from the power in the real signal for both left and right channels. Also, the value of coherence as computed for the imaginary part may differ by a few parts in 10,000 from the desired value of 0.9922. When the duration is

25 ms, the coherence in the imaginary part may differ by a few parts in 1000 from 0.9922. These inconsistencies are not expected to have an effect on the computation of fluctuations because the corrected scaling of the imaginary parts means that the imaginary parts are correct Hilbert transforms of the real parts.

Table 24 gives the mean and standard deviation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ for the noise pairs used in the experiments. However, when the durations are short, different 100-noise ensembles may exhibit quite different value for the mean and standard deviation. Consequently the values given in Appendix 3 were computed over 5000 noise-pairs.

Table 24: Values of the mean and standard deviation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ for noise-pairs with a nominal bandwidth of 14 Hz and three nominal durations: 25 ms, 50 ms, and 100 ms. Correlation between the standard deviations is also given. Each value is based on 100 noise-pairs.

BW (Hz)	Duration (ms)	$\mu(s_t[\Delta\Phi])$ (degrees)	$\mu(s_t[\Delta L])$ (dB)	$\sigma(s_t[\Delta\Phi])$ (degrees)	$\sigma(s_t[\Delta L])$ (dB)	corr
14	25	3.62	0.41	3.56	0.36	0.53
14	50	5.85	0.73	6.83	0.65	0.74
14	100	7.80	1.06	6.83	0.68	0.75

As for Chapter 1, noise-pairs were computed by a Tucker-Davis AP2 array processor (System II) and converted to analog form by 16-bit DACs (DD1). The buffer size was 4000 samples per channel and the sample rate was 8 ksps. The noise was lowpass filtered with a corner frequency of 4000 Hz and a –115 dB/octave rolloff. The noises were presented at 70 dB $\pm$ 3 dB with levels determined by programmable attenuators (PA4) operating in parallel on the two channels. The level was randomly chosen in 1-dB increments, for each of the three intervals within a trial to discourage the listener from trying to use level cues to perform the task.

3.1.2 Procedure

The noise-pairs for the three collections were made into phase sets and a level sets as was done in Chapters 1 and 2. This was done by choosing the five noise-pairs with the largest fluctuations and five noise-pairs with the smallest fluctuations, for both phase and level. Six runs were devoted to listening to a group of ten reproducible noise-pairs. Listeners completed each group before moving on to the next group. The structure of runs, trials within a run, and the data collection procedure was the same as that in Chapters 1 and 2. Experiments in this chapter employed three male listeners from Chapter 1 - D, M, and W.

3.1.3 Results

Figures 54–57 show the P_c and CAS values for the phase and level sets for the 25, 50, and 100-ms noise-pairs. The vertical dashed line shows the division between the five smallest fluctuation noise-pairs and the five largest fluctuation noise-pairs, as in Chapters 1 and 2. The horizontal dashed line shows the value corresponding to guessing.

In Figures 54–57, the 50-ms and 100-ms conditions show higher values of P_c and CAS for the five noise-pairs to the right of the horizontal dashed line compared to the five noise-pairs to the left. These noise-pairs with the largest fluctuations have values of P_c and CAS that are near the ceiling ($P_c = 100\%$ or $CAS = 72$). This is different from the noise-pairs with a 25-ms duration, which had few noise-pairs with P_c values greater than 75% and no noise-pairs with CAS values greater than 36.

Figures 54–57 are like Figures 9–12 in Chapter 1. However, there are some differences, most easily seen in the P_c data. The average value of P_c over phase and level sets over all three listeners was: 61% for 25 ms, 65% for 50 ms, and 68% for 100 ms. These values of P_c are notably smaller than the average $P_c = 92\%$ for the 500-ms noise-pairs from Chapter 1. Figures 54 and 56 show several 50-ms and 100-ms

noise-pairs below guessing for the five noise-pairs left of the horizontal dashed line. There are 30 small-fluctuation noise-pair data points (3 listeners $\times$ 2 sets $\times$ 5 small-fluctuation noise-pairs). For the 500-ms noise-pairs in Experiment 1, one had $P_c <$ 50%. For the 100-ms noise-pairs, 23 had $P_c < 50\%$. For the 50-ms noise-pairs, 23 had $P_c < 50\%$. For the 25-ms noise-pairs, 11 had $P_c < 50\%$.

As in Chapters 1 and 2, t-tests were performed to test the hypothesis that detection of incoherence more readily occurs in the five noise-pairs with the largest fluctuations in phase or level when compared to the five noise-pairs with the smallest fluctuations in phase or level. The p-values from these t-tests are in Table 25. For the 100-ms sets, six of six p-values were significant at the 0.02 level for P_c and at the 0.02 level for *CAS*. The 100-ms *CAS* data shows the same number of significant p-values and level of significance when compared to the 500-ms *CAS* data. However, the 100-ms P_c data shows a larger number of significant p-values when compared to the 500-ms P_c data. For the 50-ms sets, six of six p-values were significant at the 0.02 level for P_c and at the 0.02 level for *CAS*, much like the 100-ms sets. For the 25-ms sets, three of six p-values were significant at the 0.05 level for P_c and at the 0.05 level for *CAS*. Clearly, the number of significant p-values and level of significance for the 25-ms sets did not match the number of significant p-values and level of significance for 50, 100, and 500-ms sets.

All three listeners reported that they changed the cue used to detect incoherence for the 25-ms noise-pairs. They began to use a lateralization cue instead of a width cue. In addition, two listeners reported that two auditory images were sometimes perceived. Listener D reported that he sometimes chose the incoherent stimulus by identifying the noise-pair with two different spatial positions and/or two different temporal onsets. Listener M reported the movement or spatial position of two different images, but not two different temporal onsets. Listener W did not report two perceived images in the task.

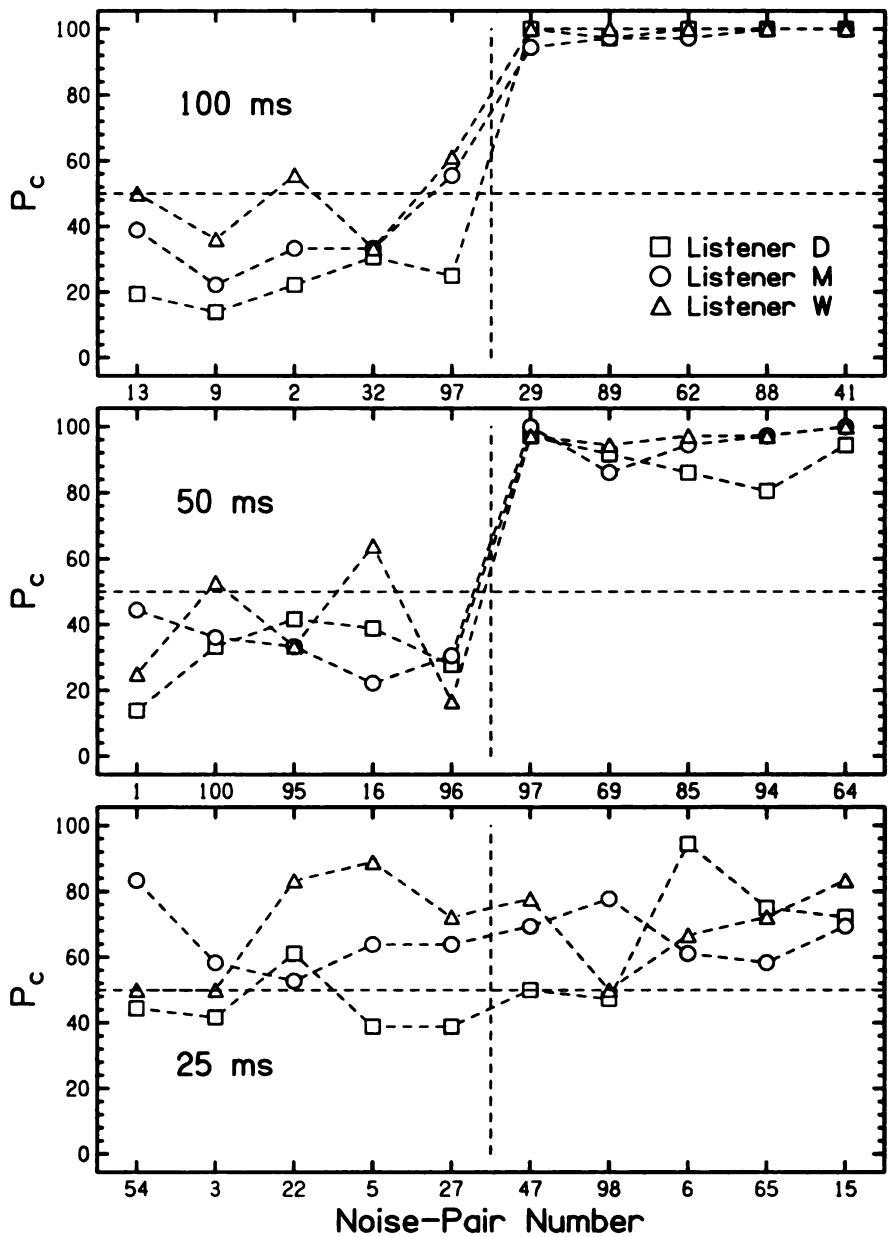

Figure 54: The percent correct (P_c) for three listeners for the *phase set* from the 25-ms, 50-ms, and 100-ms collections of Experiment 12. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta\Phi]$ in the collection of 100 noise-pairs. The noise-pairs are rank ordered by increasing $s_t[\Delta\Phi]$ along the horizontal axis. The vertical dashed line represents 90 unused reproducible noise-pairs. The horizontal dashed line represents the level of guessing. Several of the 50 and 100-ms noise-pairs to the left of the dashed line are below the level of guessing.

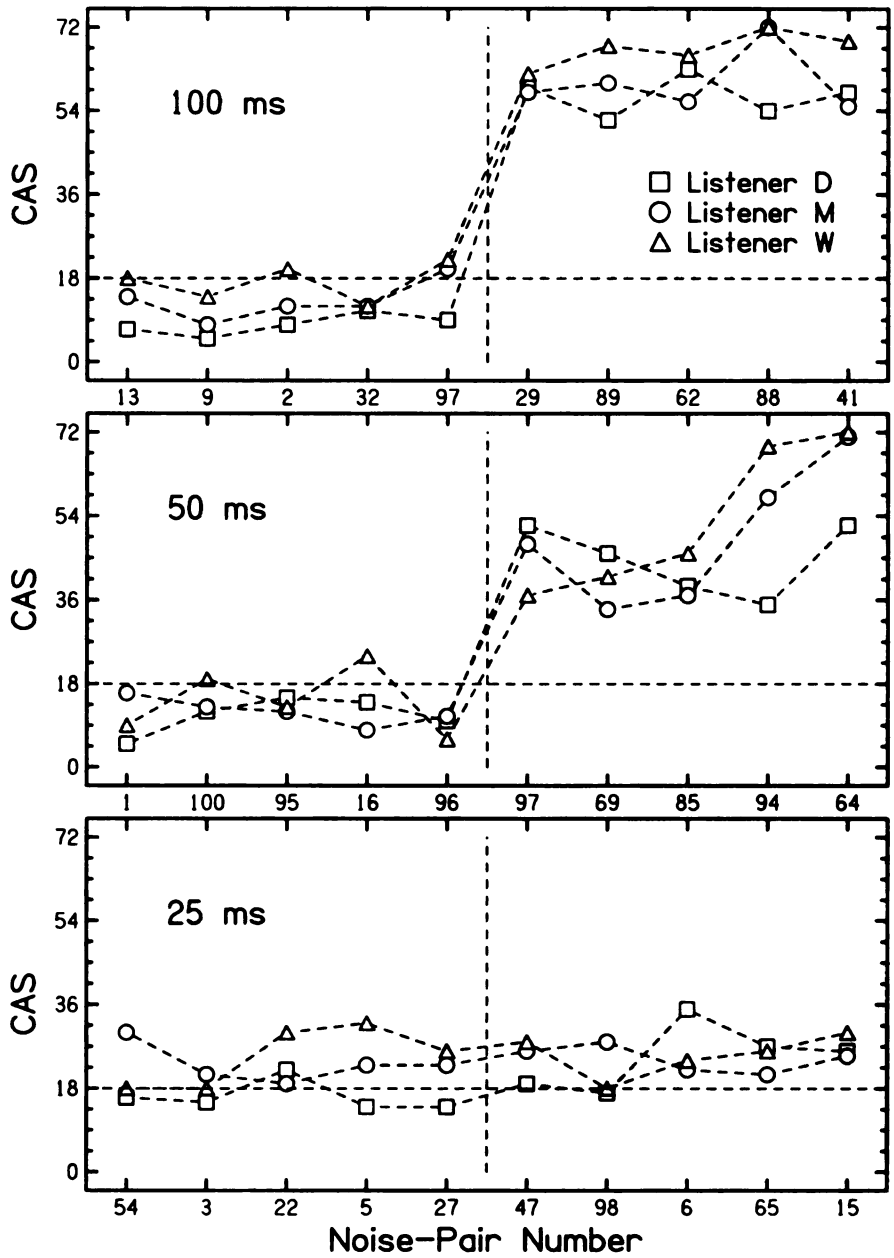

Figure 55: The *CAS* for three listeners for the *phase set* from the 25-ms, 50-ms, and 100-ms collections of Experiment 12. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta\Phi]$ in the collection of 100 noise-pairs. The noise-pairs are rank ordered by increasing $s_t[\Delta\Phi]$ along the horizontal axis. The vertical dashed line represents 90 unused reproducible noise-pairs. The horizontal dashed line represents the level of guessing. Several of the 50 and 100-ms noise-pairs to the left of the dashed line are below the level of guessing.

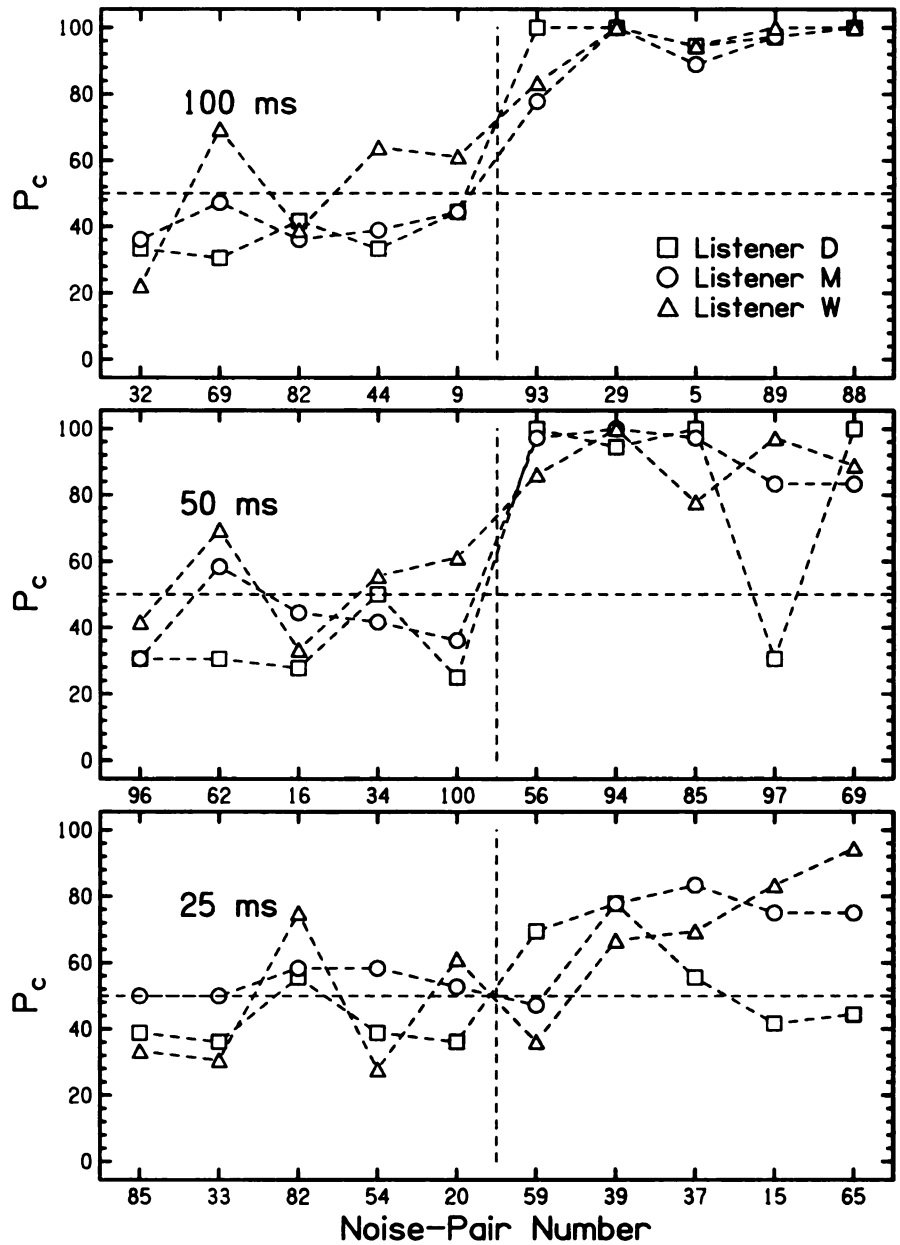

Figure 56: The percent correct (P_c) for three listeners for the *level set* from the 25-ms, 50-ms, and 100-ms collections of Experiment 12. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta L]$ in the collection of 100 noise-pairs. The noise-pairs are rank ordered by increasing $s_t[\Delta L]$ along the horizontal axis. The vertical dashed line represents 90 unused reproducible noise-pairs. The horizontal dashed line represents the level of guessing. Several of the 50 and 100-ms noise-pairs to the left of the dashed line are below the level of guessing.

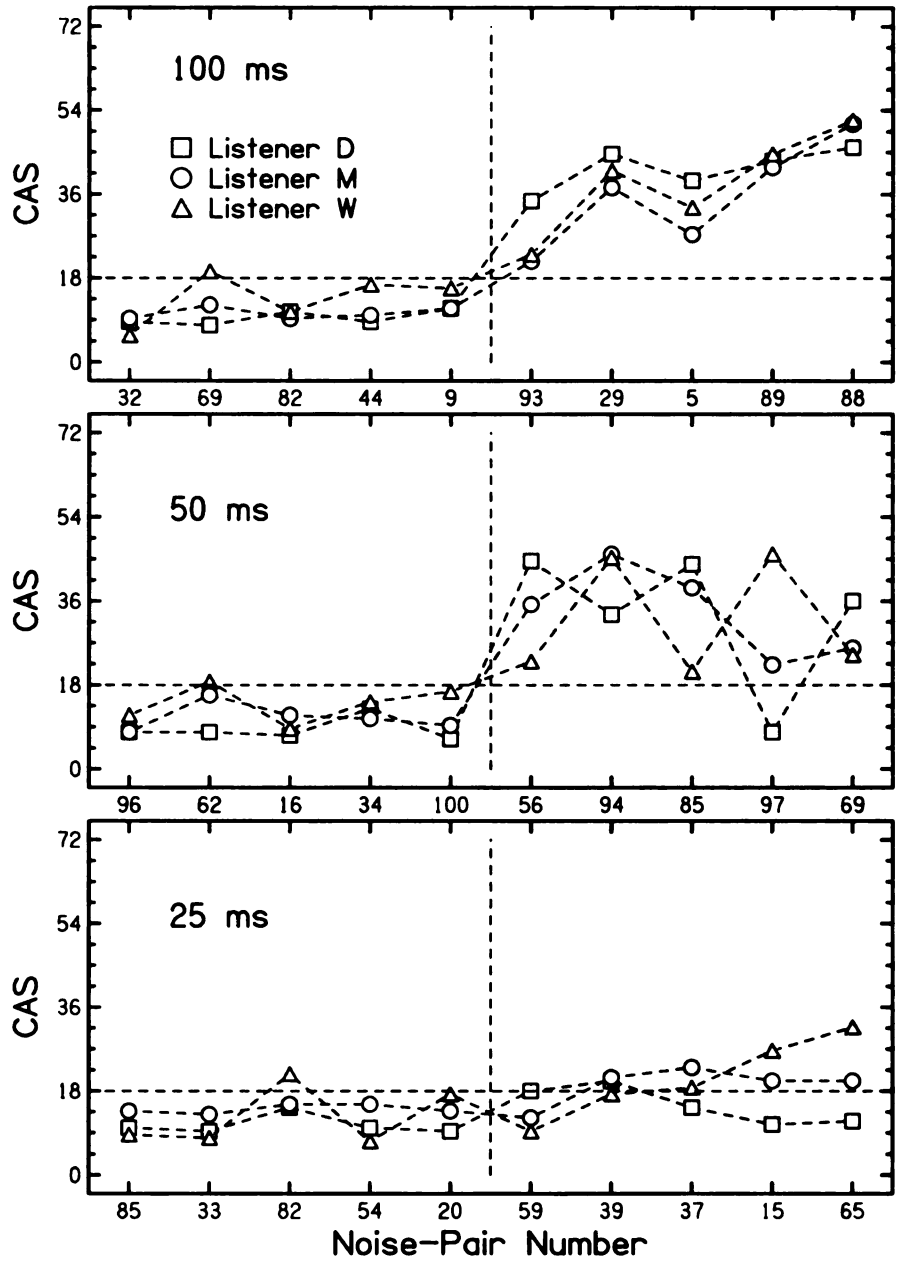

Figure 57: The *CAS* for three listeners for the *level set* from the 25-ms, 50-ms, and 100-ms collections of Experiment 12. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta L]$ in the collection of 100 noise-pairs. The noise-pairs are rank ordered by increasing $s_t[\Delta L]$ along the horizontal axis. The vertical dashed line represents 90 unused reproducible noise-pairs. The horizontal dashed line represents the level of guessing. Several of the 50 and 100-ms noise-pairs to the left of the dashed line are below the level of guessing.

Table 25: The p-values for the phase and level sets with a bandwidth of 14 Hz and four durations: 25 ms, 50 ms, 100 ms, and 500 ms. The 500 ms p-values are taken from Chapter 1.

P_c	25 ms		50 ms		100 ms		500 ms	
Listener	Phase	Level	Phase	Level	Phase	Level	Phase	Level
D	0.029	0.039	<0.001	0.008	<0.001	<0.001	0.095	0.006
M	0.334	0.027	<0.001	<0.001	<0.001	<0.001	0.021	0.047
W	0.457	0.056	0.001	<0.001	<0.001	0.003	0.015	0.014

CAS	25 ms		50 ms		100 ms		500 ms	
Listener	Phase	Level	Phase	Level	Phase	Level	Phase	Level
D	0.027	0.039	<0.001	0.010	<0.001	<0.001	<0.001	<0.001
M	0.306	0.023	0.002	0.002	<0.001	0.004	0.002	0.002
W	0.457	0.065	0.002	0.015	<0.001	0.002	0.001	<0.001

3.1.4 Discussion

The detection data of Experiment 12 for the 50 and 100-ms noise-pairs were much like those of Experiment 1 in Chapter 1 for the 500-ms noise-pairs. There was a difference in the detection scores for the smallest fluctuation noise-pairs, which were often below chance detection. The reason for this dramatic drop in P_c for the short durations is possibly the monaural envelope fluctuations as were studied in Experiments 5 and 10. For this experiment, the noise-pairs with the largest fluctuations in phase and level consistently had larger envelope fluctuations than the noise-pairs with the smallest fluctuations in phase and level. It seems that listeners are more apt to confuse envelope fluctuations with interaural fluctuations for the durations of 50 to 100 ms than for 500 ms. In the limit that noise-pairs with large interaural fluctuations always have large monaural fluctuations and noise-pairs with large monaural fluctuations are always chosen over noise-pairs with small interaural fluctuations in an incoherence detection task, values of P_c should be approximately 25% for the small interaural fluctuation noise-pairs. This is because small interaural fluctuation noise-

pairs will be presented against large monaural fluctuation noise-pairs about half of the time. This, in turn, would reduce the P_c values of these noise-pairs from guessing level, which is 50%, to half of guessing level, which is 25%. As can be seen in Figures 54 and 56, many small interaural fluctuation noise-pairs with a duration of 50 or 100 ms have scores of P_c near 25%. Noise-pairs that were 25 ms in duration did lead to as many values of P_c and CAS below chance level. These stimuli are mostly likely too short to elicit perceptible fluctuations, monaural or binaural.

All of the noise-pairs in this experiment (25, 50, and 100 ms) had the value of the coherence fixed at 0.9922 calculated over the duration of the stimulus, as was done for the 500-ms noise-pairs in Chapters 1 and 2. The p-values in Table 25 for the 50 and 100-ms conditions are much like those in Tables 2 and 14 for the 500-ms condition in the number of significant differences and levels of significance. Fewer significant p-values were seen for the 25-ms condition. Since the detection scores showed significant differences between the smallest fluctuation noise-pairs and the largest fluctuation noise-pairs, the coherence function is not used in incoherence detection, even for short durations. Hence, the results of Chapters 1 and 2 are not misleading in that fluctuations of interaural parameters are used for incoherence detection for durations as short as 50 ms.

All three listeners reported that they began to use a lateralization cue because there seemed to be no width cue for most of the 25-ms noise-pairs. Interaural fluctuations of a 14-Hz bandwidth noise are expected to vary over $1/14 = 71$ ms. Therefore, it is expected that the IPD and ILD do not vary much over 25 ms. The interaural parameters of three noise-pairs used in the 25-ms sets can be seen in Figure 58. This figure clearly illustrates that the IPD and ILD do not vary much over a duration of 25 ms for these noise-pairs. Hence, it seems appropriate for listeners to change their incoherence detection strategy to use a lateralization cue. It would then be useful to use the mean instead of the standard deviation as a statistic for incoherence detection

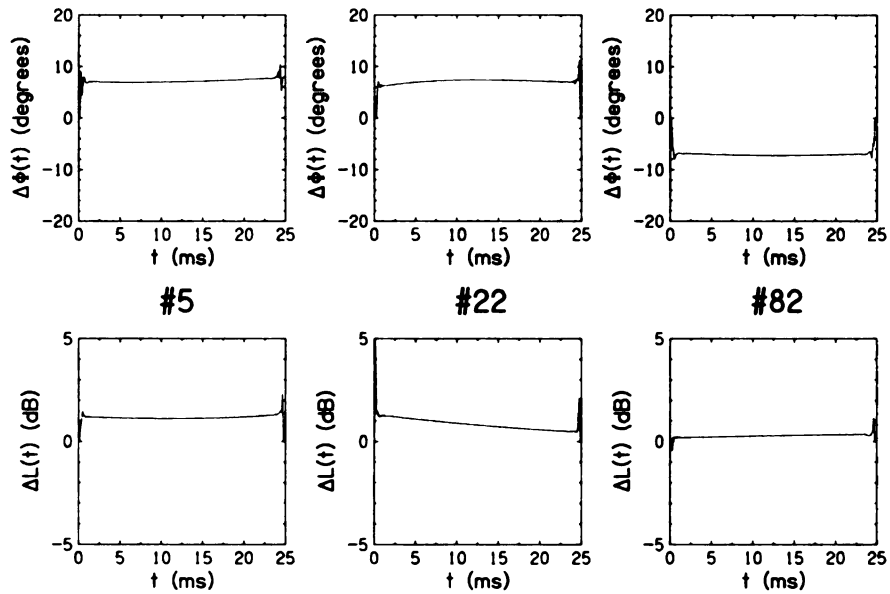

Figure 58: The interaural phase difference and interaural level difference for three noise-pairs used in the 25-ms sets in Experiment 12. For these noise-pairs, the interaural parameters do not vary much over the duration of the stimulus.

in the 25-ms noise-pairs.

The change in detection strategy from a width cue to a lateralization cue could describe the P_c and CAS values and results of the t-tests. Noise-pairs were sorted by the s_t statistic, a statistic associated with a width cue. When listeners used a lateralization cue on sets that were sorted by a statistic associated with a width cue, p-values did not show as many significant differences between the noise-pairs.

Two listeners reported that two auditory images were sometimes perceived. It may be noteworthy to note that Listener D had four significant p-values, Listener M had two significant p-values, and Listener W had zero significant p-values for the 25-ms stimuli. Therefore, listeners that perceived two images had more significant differences between the largest and smallest fluctuation noise-pairs.

This phenomenon of multiple auditory images has been seen before in Hafter and Jeffress (1968), Young and Levine (1977), and Ruotolo *et al.* (1979). These experiments were time-intensity trading experiments that used short-duration tones.

However, there are issues with the interpretation of the dual-image experiments which are addressed in Trahoitis and Kappauf (1978).

From the results of this experiment, incoherence detection experiments with a duration of 25 ms may relate to these time-intensity trading experiments. Time-intensity trading experiments use static ITDs and ILDs. Some of the 25-ms noise-pairs in this experiment have quasi-static $\Delta\Phi(t)$ and $\Delta L(t)$, as calculated from the analytic signal (see Figure 58 for examples). In both types of experiments, listeners use a lateralization cue. In both types of experiments, two images are perceived only in the case of short durations. Therefore it seems plausible that the movement of two images is used during incoherence detection and gives the sensation of width. However, since no formal experiment has been done to test the validity of the perceived dual images, this will remain a conjecture for the time being.

3.1.5 Conclusion

This chapter found that the value of the coherence over short durations was not used to detect incoherence in noise-pairs. Listeners still used the fluctuations in phase and level to detect incoherence. For the durations of 50 ms and 100 ms, the detection data and levels of significance were comparable to those found in Chapters 1 and 2. For the duration of 25 ms, the detection data and levels of significance were different from those found in Chapters 1 and 2. The reason for this discrepancy was that listeners used a lateralization cue to detect incoherence, but the noise-pairs were sorted by the s_t function, which is associated with a width cue. Listeners needed to use a lateralization cue because the time scale of interaural fluctuations for a 14-Hz bandwidth was much slower than the stimulus duration. Lastly, two listeners reported two perceived images for the shortest stimulus duration.

4 BINAURAL MODELING: FIXED COHERENCE

Chapters 1–3 studied the ability of listeners to detect interaural incoherence. The experiments showed that for the narrowest bandwidth, listeners found it much easier to detect incoherence when the fluctuations in interaural phase or level were large. As the bandwidth increased, the relative advantage of large fluctuations diminished so that the coherence itself became a better predictor of detectability.

The stimulus sets for the experiments of the previous chapters were constructed by selecting noises based on fluctuations in interaural phase or level. The hypothesis in constructing the phase set, for example, was that noise-pairs with the largest phase fluctuations would be more evidently incoherent than noise-pairs with the smallest phase fluctuations. A similar hypothesis motivated the level set. The clear difference perceived by the listeners, especially seen in the 14-Hz bandwidth data, indicated that these hypothetical criteria have some measure of perceptual validity. However, it is possible, even likely, that some other measure of stimulus fluctuation would correlate better with the human perception of incoherence. As an example, noise-pair #57 had one of the largest phase fluctuations and yet Figure 10 of Chapter 1 showed that all four listeners in the experiment found the incoherence relatively hard to detect. It is obvious that something was missing from the prediction of detectability based on stimulus parameters.

Although the experiments in previous chapters selected stimuli based on IPD fluctuations (phase sets) and ILD fluctuations (level sets), it is not clear that IPD and ILD fluctuations should be considered as comparably important. The experiments showed that the phase sets and level sets led to comparable data, but phase and level fluctuations are so strongly correlated within an ensemble of noise-pairs that comparable data do not clearly demonstrate comparable importance. For instance, it is possible that listeners only responded to interaural phase fluctuations. Constructing a set based on level fluctuations then led to significant distinctions only because the

level fluctuations are so strongly correlated with phase fluctuations.

The purpose of this chapter is to address the uncertainties left from Chapters 1–3 by testing a variety of different binaural detection models against incoherence detection data. The binaural detection models were derived from models previously used to study the masking-level difference (MLD). This was a sensible approach because the MLD is closely related to incoherence detection, which was argued in the Introduction.

Domnitz and Colburn (1976) summarized the two major types of binaural models historically presented to explain the MLD phenomenon. The first type uses interaural parameter differences, IPD and ILD. For example, the vector model (Jeffress *et al.*, 1956) predicts the largest release from masking for a signal phase difference (NoSπ) of 180°. Another example is the lateralization model (Hafter, 1971), in which a signal is detected by a shifted lateral image that is formed by combining IPD and ILD. The second type of model includes energy and cross-correlation models. When adding an out-of-phase tone to noise, the interaural correlation of the entire stimulus is reduced. Models such as the equalization-cancellation (EC) model by Durlach (1963, 1966, 1972) and the correlation model of Osman (1971, 1973) fall into this category.

There is still debate as to which type of model best describes binaural detection. Gilkey, Robinson, and Hanna (1985) found that wideband reproducible noise masking data were incompatible with several interaural parameter models. On the other hand, Colburn, Isabelle, and Tollin (1997) showed that the EC model was incompatible with reproducible noise data from Isabelle and Colburn (1991).

From the results of Chapters 1–3, it is evident that incoherence detection cannot be described by a correlation model because all of the reproducible noises had the same value of coherence and yet detection varied enormously among the different noises. Therefore, the present chapter will focus on interaural parameter models.

4.1 EXPERIMENT 13: NARROW BANDWIDTH

The purpose of Experiment 13 was to obtain incoherence detection data from a large set of narrowband noises that were randomly generated and unselected so as to be a fair representation of all noises with the given bandwidth. The detection data were collected in order to test the models presented in this chapter.

4.1.1 Stimuli

The collection of 100 noise-pairs with 14-Hz bandwidth from Experiment 1 in Chapter 1 was used for Experiment 13. In the present experiment all 100 noise-pairs were used to avoid any bias. As in Chapter 1, attention focused on the interaural phase difference $\Delta\Phi$ and the interaural level difference ΔL. Fluctuations in these interaural differences were initially defined in terms of their standard deviations over time, indicated by function s_t. The fluctuations $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ are shown in Figure 6.

4.1.2 Procedure

The 100 noise-pairs were listened to in sets of ten as ordered by serial number (order of creation). Thus, the first set had noise-pairs 1–10, the second set had noise-pairs 11–20, and so on. Six runs were devoted to listening to a set of ten reproducible-noise pairs. Listeners completed each set before moving on to the next set. The structure of runs, trials within a run, and the data collection procedure was the same as that in Chapters 1–3. Experiments in this chapter employed three male listeners from Chapter 1 - D, M and W.

4.1.3 Results

The results from listening to all 100 noise-pairs can be seen in Figure 59. The open symbols show the CAS while the solid symbols show the number of correct responses, essentially equivalent to the P_c. These values are plotted as a function of $s_t[\Delta\Phi]$ only

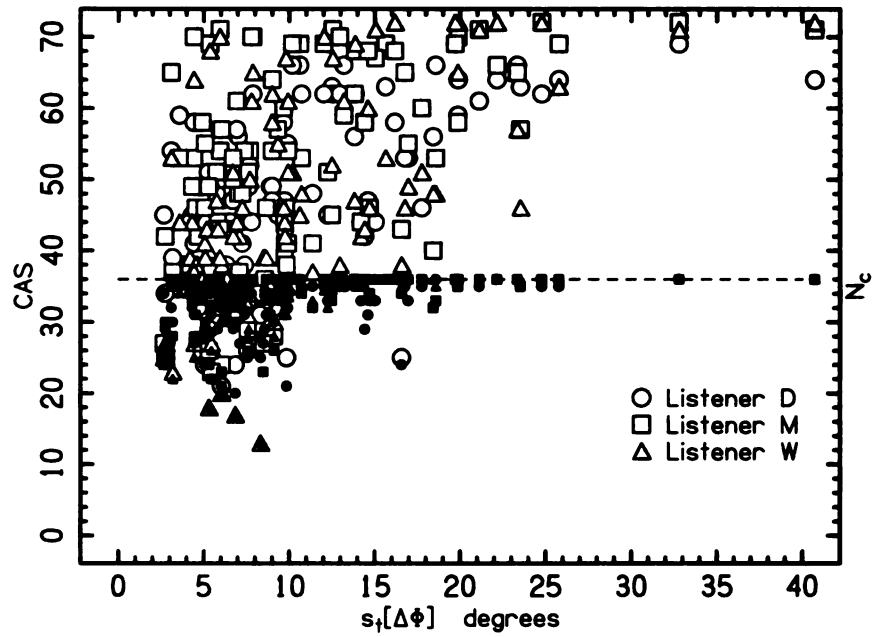

Figure 59: All the detection data for the 100 noise-pairs from Experiment 13 for three listeners, D, M, and W, are plotted twice, once as the number correct - on a scale from 0 to 36, and once as *CAS* - on a scale from 0 to 72. The data are plotted as a function of the standard deviation of the interaural phase in an attempt to give some order to the plot.

to give some order to the plot, not because $s_t[\Delta\Phi]$ is thought to be the best model for detection. This figure illustrates the advantage of using the *CAS* instead of P_c because the number of correct responses reaches a ceiling, especially for listener M. The *CAS* increases the dynamic range of the experiment.

Agreement between the listeners for individual noise-pairs is difficult to see in Figure 59, but agreement is actually good. The inter-listener correlation was 0.73 for D and M, 0.71 for D and W, and 0.80 for M and W. These inter-listener correlations are somewhat smaller than those reported for the ten noise-pairs in Chapter 1 - approximately 0.9 on average. The reason for the difference is probably that members of the entire collection of 100 noise-pairs are less distinctive than are the five largest and five smallest fluctuation noise-pairs used in Chapter 1.

4.2 MODELS FOR INCOHERENCE DETECTION

In order to discover the stimulus features that best predict human perception of incoherence, models of perception were constructed using transformed interaural parameters, and the models were tested against the large set of perceptual data from Experiment 13. Because the clearest distinction between noise-pairs occurs for the smallest bandwidths, it was natural for attention to focus on the 14-Hz bandwidth. Therefore the 14-Hz data will be treated first. The remainder of this section describes the models.

4.2.1 Model preprocessing assumptions

Several assumptions, common to all the models, were made to reflect auditory preprocessing of the complex incoherent stimuli. Two free parameters are introduced in this section as well as a scale of lateralization for the $\Delta\Phi(t)$ and $\Delta L(t)$.

1. Compression

A small static interaural difference displaces the lateral position of the auditory image from a centered position. A greater interaural difference leads to a greater displacement. But increasing interaural differences produce diminishing returns because the lateral position is a compressive function of interaural differences. A perceptual model for fluctuations can easily adopt this effect from static experiments. The compression functions used in the present analysis were derived from Yost's 1981 experiments. They are of the form

$$\Psi^{\circ}_{\Delta\Phi}(t) = 10 \cdot \mathrm{sgn}[\Delta\Phi(t)] \cdot (1 - e^{-|\Delta\Phi(t)|/40}) \tag{15}$$

and

$$\Psi^{\circ}_{\Delta L}(t) = 10 \cdot \mathrm{sgn}[\Delta L(t)] \cdot (1 - e^{-|\Delta L(t)|/8}) \tag{16}$$

where $\Psi^{\circ}_{\Delta\Phi}(t)$ and $\Psi^{\circ}_{\Delta L}(t)$ are on a scale of lateral position that ranges from −10 to

10. In Eq. 15, the weighting constant of the exponential is 40 degrees. In Eq. 16, the weighting constant of the exponential is 8 dB. These functions correspond to the lateral position of a sine-tone at a frequency of 500 Hz, the center frequency of the noise-pairs. A further benefit of the compressive lateral position transformation is that IPD and ILD are put on the same scale so that they can be easily combined in mathematical models.

2. Temporal averaging

The fluctuation measures used to construct stimulus sets for Chapters 1–3 were based on instantaneous values of interaural differences as they appeared with the 8000 Hz sample rate. But it is not evident that, for example, a large interaural difference with a duration of only 0.125 ms would receive much respect from the binaural system. Therefore, the present models include a parametric temporal averaging operation, following other models, e.g. Viemeister (1979), in using an exponential averaging window to represent a temporal modulation transfer function of the form

$$\hat{e}[\Psi^\circ(t)] = \frac{\int_0^{T_D} \Psi^\circ(t-t')e^{-t'/\tau}dt'}{\int_0^{T_D} e^{-t''/\tau}dt''} \quad (T_D < t), \tag{17}$$

where Ψ° represents a transformed interaural difference, the IPD from Eq. 15 or the ILD from Eq. 16. The time constant τ was a free parameter and the averaging window was terminated after the weight of the exponential function dropped to 0.1, which determined the upper limit of the integration T_D.

3. Critical envelope value weighting

Maxima can occur in the IPD, $\Delta\Phi(t)$, if the envelope in one ear is very small relative to the other. These large IPDs can occur especially if one of the channels (left or right) has an envelope that is near zero. But if the envelope is near zero, the listener may not be able to detect this fluctuation in $\Delta\Phi(t)$. Therefore, it would be wrong for a model to give much weight to this phase fluctuation. Following a suggestion by

Colburn (2004), it was sought to reduce the problem by discounting phase fluctuations that coincided with very small envelope values by employing a weighting function,

$$w_g(t) = \begin{cases} 1 & \text{if } E_L(t) \text{ and } E_R(t) \geq g\ E_{\text{RMS}} \\ 0 & \text{if } E_L(t) \text{ or } E_R(t) < g\ E_{\text{RMS}} \end{cases} \tag{18}$$

where E_L and E_R are Hilbert envelopes for left and right channels, and E_{RMS} actually indicates a comparison with corresponding left or right channels RMS values. Parameter g is the critical envelope fraction - a free parameter. If the envelope in either channel is less than g times the RMS envelope, then the weight is set to zero. Otherwise, the weight is set to one.

After all the modeling assumptions, the *transformed* IPD and ILD are described by the notation

$$\Psi_{\Delta\Phi}(t) = \hat{e}[\Psi^{\circ}_{\Delta\Phi}(t)] \cdot w_g(t) \tag{19}$$

and

$$\Psi_{\Delta L}(t) = \hat{e}[\Psi^{\circ}_{\Delta L}(t)]. \tag{20}$$

Because the allowed values of the preprocessing parameters τ (exponential averaging) and g (envelope weighting) include the entire physical range, the transformed interaural differences admit the possibility of no transformation. The exception is in the compression, which was always applied to models 1–7 below.

4.2.2 Models for binaural combination

Ten different binaural combination models with adjustable parameters were studied. Each model produced a decision statistic intended to predict the detectability of incoherence. The models and their parameters were then independent variables in regressions comparing predictions with listener detection performance.

The models tested three different hypotheses concerning binaural combination: (1)

the independent interaural difference or independent centers model, (2) the lateral position or lateralization model, and (3) the short-term cross-correlation model. In models of the independent difference type, averaged fluctuations in IPD and in ILD are combined with a relative weighting parameter a. In models of the lateral position type, an image location is calculated based on IPD and ILD values that are combined with a time/intensity trading parameter b. The decision statistic is then based on fluctuations in that location. In the short-term cross-correlation models, only the IPD is used, so no weighting or trading parameter is necessary.

Model 1: Sum of interaural differences

A simple model of the independent-interaural-difference type hypothesizes that incoherence is detected on the basis of a linear combination of the standard deviations computed over the duration of the noise in IPD and ILD. The standard deviations of a transformed interaural difference is

$$s_t[\Psi] = \sqrt{\frac{1}{T - T_D} \int_{T_D}^{T} [\Psi(t) - \overline{\Psi}]^2 dt} \tag{21}$$

where $\Psi(t)$ is either the transformed IPD or ILD. Therefore, the sum of transformed standard deviation of IPD and ILD is

$$d_1 = a \cdot s_t[\Psi_{\Delta\Phi}(t)] + (1 - a) \cdot s_t[\Psi_{\Delta L}(t)]. \tag{22}$$

This model has three free parameters: a, τ, and g. The nontransformed fluctuations in IPD and ILD were, in fact, the basis for choosing stimuli in Chapters 1–3. There, it was found that larger values of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ correlated with a greater detectability of incoherence in noise-pairs for a given value of coherence.

Model 2: Sum of mean square variations

As a close relative to the decision statistic d_1, an independent differences model could use the square of the fluctuation. The sum of mean square variations model

was introduced by Isabelle and Colburn (1987) in connection with a masking level difference experiment with reproducible stimuli,

$$d_2 = a^2 \cdot s_t^2[\Psi_{\Delta\Phi}(t)] + (1-a)^2 \cdot s_t^2[\Psi_{\Delta L}(t)]. \tag{23}$$

This model has the same free parameters as model 1. This model "...intended to capture the subjective increase in image width caused by the addition of a target tone to the narrowband masker..." (Isabelle and Colburn, 2004).

Model 3: Sum of integrations

An alternative decision statistic is based on an integration of the absolute value of the fluctuations over the duration of the stimulus. In this model the contributions of the IPD and ILD are computed separately. There is no known precedent for such an independent absolute value model in the literature.

$$d_3 = a \cdot \frac{1}{T - T_D} \int_{T_D}^{T} \left|\Psi_{\Delta\Phi}(t)\right| dt + (1-a) \cdot \frac{1}{T - T_D} \int_{T_D}^{T} \left|\Psi_{\Delta L}(t)\right| dt. \tag{24}$$

This model has the same free parameters as model 1.

Model 4: Sum of threshold deviations

A fourth kind of decision statistic measures the fraction of the time that interaural differences are far from zero (the center position). This thresholded statistic is defined as

$$d_4 = a \cdot \frac{1}{T - T_D} \int_{T_D}^{T} W[h, \Psi_{\Delta\Phi}(t)] dt + (1-a) \cdot \frac{1}{T - T_D} \int_{T_D}^{T} W[h, \Psi_{\Delta L}(t)] dt. \tag{25}$$

where

$$W[h, \Psi(t)] = \begin{cases} 1 & \text{if } \left|\Psi(t)\right| \geq h \\ 0 & \text{if } \left|\Psi(t)\right| < h. \end{cases} \tag{26}$$

Besides having the same free parameters as other models, model 4 has an extra free

parameter (four free parameters total), h, to set the level of threshold. Since both transformed interaural parameters are on the same scale of lateral position, it was assumed that the threshold was the same for both interaural differences.

Webster (1951) proposed a similar model that used only deviations of IPD to determine the influence of interaural phase on masking thresholds. This model predicts that large deviations in either IPD or ILD are the basis for incoherence detection. Model 4 reduces to Webster's model for $a = 1$.

Model 5: Standard deviation of the lateral position

Model 5 comes from a model by Hafter (1971) that asserts detection of a signal might occur from a shift in the lateral position of an image formed by combining IPD and ILD with a time-intensity trading ratio. Model 5 is the first model of three in this chapter based on a time-varying lateral position, and it hypothesizes that the standard deviation of fluctuations in the lateral position best describes incoherence detection. In a lateral position model, a fluctuation in phase can cancel a fluctuation in level. Cancellation is not possible in an independent center model. The lateral position itself can be defined as,

$$\Psi_z(t) = b \cdot \Psi_{\Delta\Phi}(t) + (1 - b) \cdot \Psi_{\Delta L}(t), \tag{27}$$

where b is a dimensionless time-intensity trading parameter for transformed interaural differences. Then the measure of the fluctuation of the lateral position becomes,

$$d_5 = s_t[\Psi_z(t)] = s_t[b \cdot \Psi_{\Delta\Phi}(t) + (1 - b) \cdot \Psi_{\Delta L}(t)], \tag{28}$$

where there are three free parameters: b, τ, and g.

Model 6: Integration of the lateral position

The particular model that Hafter proposed in 1971 was actually a model based on the absolute value of lateral-position incorporating time-intensity trading. Converted

to use transformed variables, the model gives

$$d_6 = \frac{1}{T - T_D} \int_{T_D}^{T} \left| \Psi_z(t) \right| dt = \frac{1}{T - T_D} \int_{T_D}^{T} \left| b \cdot \Psi_{\Delta\Phi}(t) + (1 - b) \cdot \Psi_{\Delta L}(t) \right| dt. \quad (29)$$

Here, the instantaneous lateral position corresponds to the fluctuation because it is assumed that the undisplaced position corresponds to $z = 0$. This model has the same free parameters as model 5.

Model 7: Threshold deviation of the lateral position

Deviations that exceed a threshold value constitute events, and the durations of these events are summed in a decision statistic given by

$$d_7 = \frac{1}{T - T_D} \int_{T_D}^{T} W[h, \Psi_z(t)] dt. \quad (30)$$

As for model 4, W has the value 1 if Ψ_z is greater than h and is zero otherwise. Besides having the same free parameters as other lateral position models, model 7 has an extra free parameter (four free parameters total), h, to set the level of threshold.

Model 8: Standard deviation of the short-term cross-correlation function

In connection with the MLD, Osman (1971) proposed a model based on the interaural cross-correlation computed over the entire observation interval. An alternative model computes the cross-correlation as a function of running time t. It can be shown that for slowly varying $E_L(t)$, $E_R(t)$, $\Phi_L(t)$, and $\Phi_R(t)$ the running cross-correlation reduces to the cosine of the IPD (Isabelle and Colburn, 2004). The cross-correlation function, written as a function of time is

$$\gamma_{\mathrm{st}}(t) = \frac{\int_{t-\Delta t}^{t} x_L(t') x_R(t') dt'}{\sqrt{\int_{t-\Delta t}^{t} x_L^2(t_1) dt_1 \int_{t-\Delta t}^{t} x_R^2(t_2) dt_2}}, \quad (31)$$

where $x_L(t')$ is the left channel waveform and $x_R(t')$ is the right channel waveform. This cross-correlation function is evaluated at zero lag because an incoherence de-

tection experiment includes no offset ITD. The integration window Δt is brief. For instance, Isabelle and Colburn (2004) take it to be the inverse of the center frequency of the noise band.

This is the same function as Eq. 1 in the Introduction except that it is assumed that the binaural system evaluates the function on a smaller time scale than the entire duration of the stimulus; therefore it does not vary the value of the lag to maximize the value of the cross-correlation function. Rewritten in terms of the Hilbert envelope and phase, the running cross-correlation is

$$\gamma_{\mathrm{st}}(t) = \frac{\int_{t-\Delta t}^{t} E_L(t')E_R(t')\cos[\omega t' + \Phi_L(t')]\cos[\omega t' + \Phi_R(t')]dt'}{\sqrt{\int_{t-\Delta t}^{t} |E_L(t_1)|^2 \cos^2[\omega t_1 + \Phi_L(t_1)]dt_1 \int_{t-\Delta t}^{t} |E_R(t_2)|^2 \cos^2[\omega t_2 + \Phi_L(t_2)]dt_2}}. \tag{32}$$

Isabelle and Colburn (2004) showed that $\gamma_{\mathrm{st}}(t)$ is approximately given by the cosine of the instantaneous interaural phase difference when the bandwidth is small, as for Experiment 13. For the bandwidth of 14 Hz, the Hilbert envelope and phase should vary on the time-scale of $1/14 = 74$ ms. For the center frequency of $f_c = 500$ Hz, this time-scale is much slower than the period, $\Delta t = 2$ ms. It then can be assumed that E_L, E_R, Φ_L, and Φ_R are approximately constant over the integration intervals in the above equation. Over one period of the stimulus the denominator reduces to $E_L E_R$, which then cancels the envelope factors in the numerator. Therefore, the short-term cross-correlation function can be approximated as

$$\gamma_{\mathrm{st}}(t) \approx \cos \Delta\Phi(t). \tag{33}$$

The deviation from the diotic value is $1 - \gamma(t)$ and the transformed deviation is

$$\Psi_{\mathrm{CC}}(t) = \hat{e}\{1 - \cos[\Delta\Phi(t)]\} \cdot w_g(t). \tag{34}$$

The value of $\Psi_{\mathrm{CC}}(t)$ ranges from 0 to 2. The transformed deviation includes temporal

averaging which potentially reduces the effectiveness of brief lateral excursions, and it incorporates envelope weighting wherein a deviation from perfect correlation is not noticed if an envelope becomes too small.

The variability of the transformed deviation then forms a decision statistic

$$d_8 = s_t[\Psi_{CC}(t)] = \left[\frac{1}{T - T_D}\int_{T_D}^{T} \Psi_{CC}(t)^2 dt\right]^{1/2}. \tag{35}$$

Like the other models, the short-term cross-correlation incorporates exponential temporal windowing, and critical envelope weighting. Unlike the other models, it does not include compression so that the interaural phase remains in units of radians. Because d_8 does not include any form of ILD, it does not include IPD-ILD weighting and has only two free parameters, τ and g.

Another feature of model 8 is that the meaning of "s_t" was slightly modified. Whereas Eq. 21 defines s_t as a deviation from the mean value, the decision statistic d_8 was computed as a deviation of Ψ_{CC} from zero to better represent the idea of deviation from a diotic noise.

Model 9: Integration of the short-term cross-correlation function

Just as models 3 and 6 integrated the absolute value of model percepts, model 9 integrates the absolute deviation.

$$d_9 = \frac{1}{T - T_D}\int_{T_D}^{T} \Psi_{CC}(t) dt. \tag{36}$$

By definition, $\Psi_{CC} \geq 0$. Model 9 has the same free parameters as model 8, τ and g. Also, like Model 8, compression was not included in the transformed variable so that $\Delta\Phi(t)$ is in radians.

Model 10: Threshold of the short-term cross-correlation function

Just as models 4 and 7 were based on thresholded values of model percepts, model

10 integrates a thresholded short-term cross-correlation.

$$d_{10} = \frac{1}{T - T_D} \int_{T_D}^{T} W[h, \Psi_{CC}(t)]dt. \tag{37}$$

Model 10 has three parameters, τ, g, and h, where h sets the *magnitude* of the threshold deviation. The magnitude of the deviation can be as large as 2, the difference between cos(0) and cos(π). Like the other short-term cross-correlation models, the compression was omitted.

4.2.3 Models vs. Experiment 13

The ten models presented above were tested against the data from Experiment 13. A linear regression of the form $y = mx + b$ was used to evaluate the effectiveness of a model to describe incoherence detection. The y-variable was the *CAS* for the individual listeners or for an average over listeners. The x-variable was d_n from one of the ten models. The linear correlation coefficient, r, was used to compare the results of the regressions. The maximum r, r_{max}, was found by independently varying all the free parameters over a reasonable space. For example, model 1 has three free parameters a, τ, and g. The range of a was 0 to 1 with a 0.01 increment; the range of τ was 0 to 6 ms with a 0.5 ms increment (tests with larger values of τ will be described later); the range of g was 0 to 1 with an increment of 0.01. Therefore, for model 1, 480,000 linear regressions were performed (100 $\times$ 12 $\times$ 100 $\times$ 4 listeners). For the threshold models 4 and 7, the range of h was 0 to 10 with a 0.25 increment. (Recall that the compressed IPD and ILD are on a scale of -10 to 10.) For threshold model 10, the range of h was 0 to 2 with a 0.01 increment. A power law regression equation was also used to fit the data, but it did not improve the correlation between the experimental data and the models.

1. Comparison of model types

The results of the regressions are shown in Figures 60 and 61 for P_c and CAS for the best combination (largest r) of all the parameters for each model. The figure presents results for individual listeners and for the average listener. Therefore, the $r_{\max}$ of the averaged data is not the average $r_{\max}$ of the listeners. It is interesting that the most successful models agree better with the average listener than they do with any single listener. A similar result was found by Isabelle and Colburn (1991) in modeling MLDs for reproducible noise. Given that the models are simple signal processing algorithms whereas human listeners have complicated individual tendencies, that is the sort of result that one would expect from a model that correctly represents the general population. The modeling results from the CAS data is consistently better than the P_c data.

Figures 60 and 61 show that the threshold processing type (models 4, 7, and 10) yield no distinction between the independent centers, lateral position, and short-term cross-correlation (STCC) models. However, there were distinctions for the other processing types. The models of the independent center type (models 1–3) were more successful than the lateral position type (models 5 and 6). The independent center and lateral position models described the data better than the short-term cross-correlation models (models 8 and 9). Since the STCC depends entirely on $\Delta\Phi(t)$, this may be evidence for an important contribution of $\Delta L(t)$ to incoherence detection.

Model 1 had the largest $r_{\max}$ for all three listeners and for the averaged data. For the averaged data, $r_{\max} = 0.69$ for the P_c data and $r_{\max} = 0.87$ for the CAS data. The values of $r_{\max}$ for each type of model (independent centers, lateral position, and STCC) for the average listener can be more easily compared on Figures 62 and 63. These figures omit model 2 because model 2 has no analog among the lateral position and STCC models. Also, the performance of model 2 is very similar to that of model 1, except that the $r_{\max}$ is always slightly smaller for model 2.

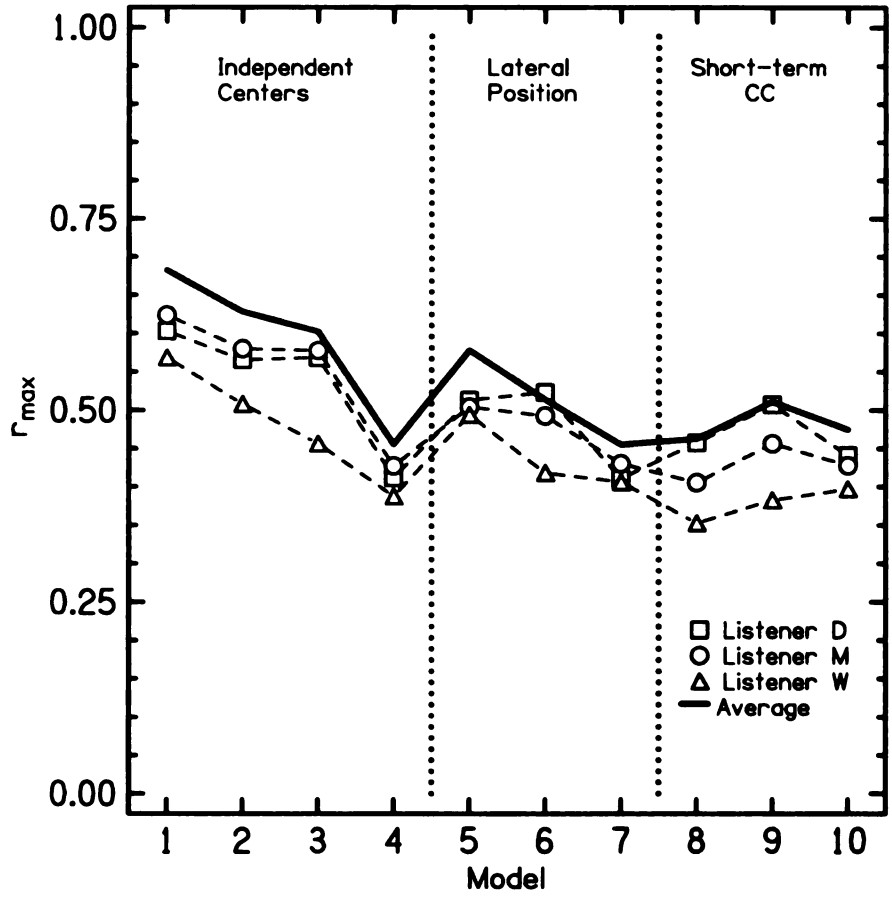

Figure 60: The comparison of P_c scores for the 14-Hz noise-pairs of Experiment 13 with ten models. The value of $r_{\max}$ shows the correlation between the experimental *CAS* scores for the 100 noise-pairs and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. It is not the average of $r_{\max}$ averaged over the listeners. Model 1 performs best with $r_{\max} = 0.68$. Models 2–3 perform almost as well. Models 5–6 do not correlate as well with the data as models 1–3. Models 8–9 do not correlate as well with the data as models 5–7. There is little difference between models 4, 7, and 10.

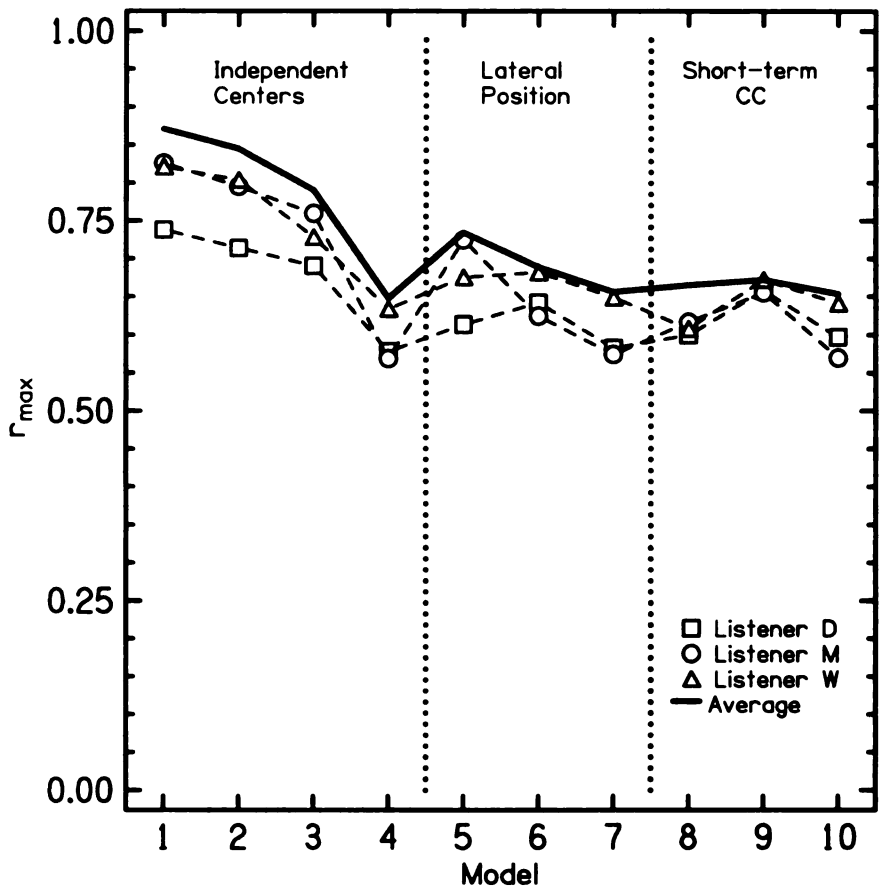

Figure 61: The comparison of *CAS* scores for the 14-Hz noise-pairs of Experiment 13 with ten models. The value of r_{max} shows the correlation between the experimental *CAS* scores for the 100 noise-pairs and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. It is not the average of r_{max} averaged over the listeners. Model 1 performs best with $r_{max} = 0.87$. Models 2–3 perform almost as well. Models 5–6 do not correlate as well with the data as models 1–3. Models 8–9 do not correlate as well with the data as models 5–7. There is little difference between models 4, 7, and 10.

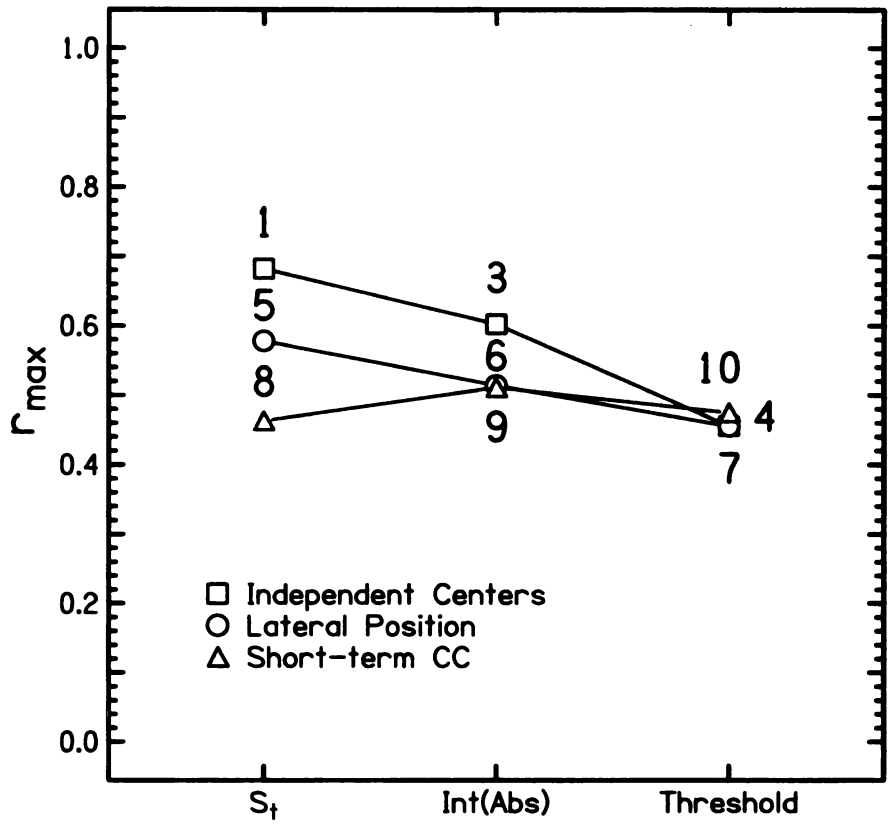

Figure 62: The comparison of each mathematical model (s_t ...) for each type of binaural processing (independent centers ...) in the modeling of the 14-Hz bandwidth data for the average listener for the P_c data. Model 2 is omitted from this plot.

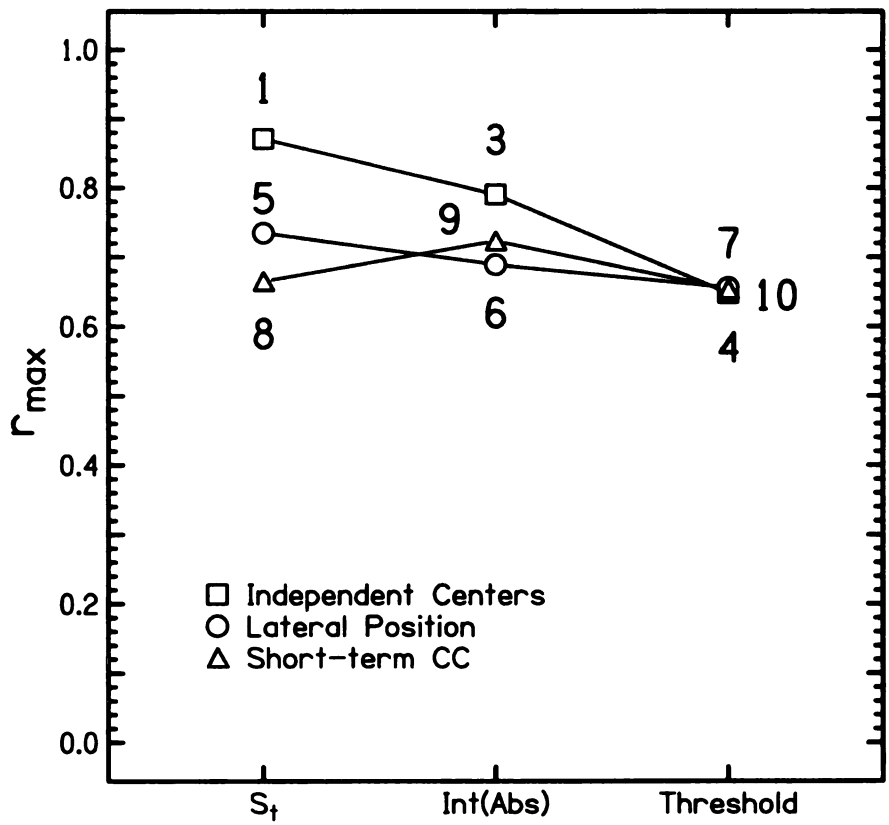

Figure 63: The comparison of each mathematical model (s_t ...) for each type of binaural processing (independent centers ...) in the modeling of the 14-Hz bandwidth data for the average listener for the *CAS* data. Model 2 is omitted from this plot.

Tables 26 and 27 show the values of the free parameters for the maximized r for the 14-Hz bandwidth modeling. A value of Tables 26 and 27 is that they show that the parameters are similar for different listeners over the different types of models (independent center, lateral position, and STCC). Consequently, the fits to the average listener shown in Figures 60–63 are meaningful. Tables 26 and 27 also show that fitting parameters that optimize r are similar across types of models, to the extent that the models permit them to be compared.

Table 26: Values of free parameters that optimize r in modeling the P_c detection results of Experiment 13 with 100 noise-pairs with a bandwidth of 14 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_1	D	1.5	0.49	0.19	-
	M	2.5	0.54	0.33	-
	W	0.5	0.47	0.32	-
	Ave	0.0	0.49	0.27	-
d_2	D	1.5	0.50	0.21	-
	M	2.0	0.54	0.33	-
	W	0.0	0.50	0.27	-
	Ave	0.0	0.51	0.27	-
d_3	D	2.0	0.54	0.15	-
	M	2.0	0.53	0.16	-
	W	0.0	0.38	0.15	-
	Ave	1.5	0.48	0.15	-
d_4	D	0.5	0.98	0.00	4.00
	M	0.5	1.00	0.00	3.75
	W	0.5	0.96	0.00	5.50
	Ave	0.5	0.97	0.00	3.75
d_5	D	0.5	0.00	0.00	-
	M	0.0	0.00	0.00	-
	W	0.0	0.00	0.00	-
	Ave	0.0	0.00	0.00	-
d_6	D	0.5	1.00	0.04	-
	M	2.5	1.00	0.04	-
	W	3.0	0.29	1.00	-
	Ave	1.0	1.00	0.04	-
d_7	D	0.5	1.00	0.00	4.00
	M	0.0	0.89	0.00	3.00
	W	0.0	0.94	0.00	2.00
	Ave	0.5	0.94	0.00	3.50
d_8	D	9.0	-	0.39	-
	M	8.5	-	0.39	-
	W	9.0	-	0.39	-
	Ave	9.0	-	0.39	-
d_9	D	9.5	-	0.36	-
	M	9.0	-	0.36	-
	W	11.5	-	0.36	-
	Ave	10.0	-	0.36	-
d_{10}	D	6.5	-	0.00	0.05
	M	0.0	-	0.00	0.05
	W	10.0	-	0.00	0.04
	Ave	9.0	-	0.00	0.05

Table 27: Values of free parameters that optimize r in modeling the *CAS* detection results of Experiment 13 with 100 noise-pairs with a bandwidth of 14 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_1	D	1.5	0.53	0.17	-
	M	2.5	0.45	0.21	-
	W	0.5	0.53	0.23	-
	Ave	2.0	0.50	0.19	-
d_2	D	1.5	0.51	0.17	-
	M	2.0	0.48	0.23	-
	W	0.5	0.52	0.23	-
	Ave	1.5	0.50	0.23	-
d_3	D	0.5	0.62	0.04	-
	M	2.0	0.44	0.15	-
	W	1.5	0.51	0.12	-
	Ave	1.5	0.50	0.12	-
d_4	D	0.5	1.00	0.00	4.00
	M	0.5	0.99	0.00	3.75
	W	0.5	0.94	0.00	4.75
	Ave	0.5	0.99	0.00	4.00
d_5	D	0.5	0.00	0.00	-
	M	2.0	0.00	0.00	-
	W	2.5	0.00	0.00	-
	Ave	1.5	0.00	0.00	-
d_6	D	0.5	0.97	0.04	-
	M	1.0	0.11	0.15	-
	W	0.0	0.42	0.15	-
	Ave	0.5	0.91	0.04	-
d_7	D	0.5	0.81	0.00	3.25
	M	0.0	0.84	0.00	3.50
	W	0.0	0.77	0.00	3.25
	Ave	0.5	0.80	0.00	3.25
d_8	D	9.5	-	0.39	-
	M	9.0	-	0.38	-
	W	9.5	-	0.36	-
	Ave	9.5	-	0.38	-
d_9	D	10.0	-	0.36	-
	M	11.5	-	0.36	-
	W	7.5	-	0.29	-
	Ave	10.5	-	0.35	-
d_{10}	D	6.5	-	0.00	0.05
	M	0.0	-	0.00	0.06
	W	4.0	-	0.00	0.08
	Ave	3.0	-	0.00	0.06

2. Optimized parameters for model 1

The most successful model was model 1, and Figure 64 shows how the free parameters change against each other to maximize r for the average listener in that model for the *CAS* data. Results of other conditions (P_c vs. *CAS*, individual listeners vs. average listener) produced very similar plots to the extent that it is necessary to show only the most general condition. Figure 64 was generated by varying two parameters and keeping the other constant. Plots of a vs. τ assume that $g = 0.19$, plots of g vs. τ assume that $a = 0.50$, and plots of a vs. g assume that $\tau = 2$.

The results of Figure 64 can be summarized as follows:

(a) Integration time:

The integration time constant τ was found to be unimportant. The greatest r occurs for $\tau = 2$ ms, however changing τ leads to little difference in r. The characteristic stimulus fluctuations, expected to be of order 1/14 seconds, are slow compared to the time constants tested in detail, 0–6 ms.

In addition to the detailed test for integration times less than 6 ms, longer integration times were specifically tested. Longer times were potentially indicated because the oscillating coherence experiments of Grantham and Wightman (1978) led to binaural time constants as long as 64 ms (−3 dB response at 2.5 Hz), an effect commonly called "binaural sluggishness." MLD experiments using a masker with temporally varying coherence by Grantham and Wightman (1979) led to time constants that were even longer. Therefore, spot checks were performed, trying to fit *CAS* data with longer values of τ. For increasing values of $\tau = 0$, 25, 50, 75, 100, 125, and 150 ms, the value of r decreased monotonically. The value of $r_{\max}$ dropped from $r_{\max} \approx 0.8$ for $\tau = 0$ ms, to $r_{\max} \approx 0.6$ for $\tau = 75$ ms, to $r_{max} \approx 0.3$ for $\tau = 150$ ms. It was concluded that there is no useful role for binaural sluggishness in model 1 - the best model.

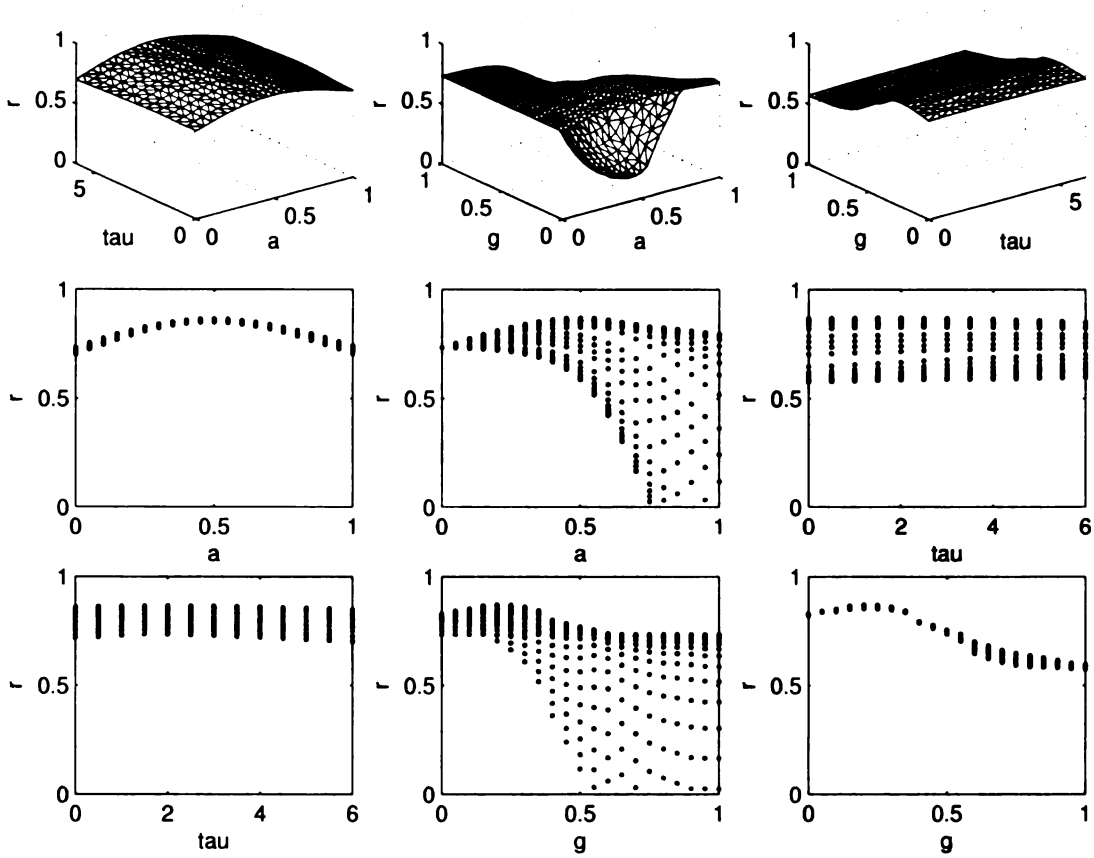

Figure 64: The free parameter surfaces for fitting model 1 to the average listener data from Experiment 13 - one-hundred noise-pairs with 14-Hz bandwidth. In the upper-left panel, r is plotted against a and τ for $g = 0.19$. In the upper-middle panel, r is plotted against g and τ for $a = 0.5$. In the upper-right panel, r is plotted against a and g for $\tau = 2$ ms. The two panels below an upper panel flatten one of the free parameter dimensions. The variations of free parameters create smooth surfaces. At this bandwidth, the dependence upon τ is negligible.

(b) Critical envelope weighting:

According to the regression analysis, the best value of g for model 1 is about 0.19, though the r_{max} is insensitive to g in this vicinity ($0.1 < g < 0.3$). This result indicates that there is a modest benefit on the average of ignoring phase differences that coincide with a particularly small envelope in one or both of the ears.

A greater benefit from envelope weighting is seen when one tries to predict detection for individual waveforms. The weighting omits large phase fluctuations that occur during the onset and offset of the stimulus, where the temporal shaping is applied and the envelope is small. One would expect that even large phase fluctuations during these times would often be missed by the listener because they occur at the very beginning or end of the stimulus.

The envelope weighting applied in Eq. 18 is a simple on/off type. Other envelope weighting functions were tried - linear envelope weighting of the IPD and squared envelope weighting of the IPD - but including these functions did not describe the data as well as including the on/off type envelope weighting.

(c) Relative IPD-ILD contributions:

The best value of a for model 1 for the average data is 0.50, which means that transformed IPD and ILD values contribute equally to the sensation of incoherence. Because the transformed values were scaled by Eq. 15 and Eq. 16, this near equality is interpreted to mean that the scaling correctly represents the relative perceptual importance of IPD and ILD. Because the scaling was derived from Yost's steady-state sine experiment results, it was concluded that it is valid to extend the steady-state data to the case of fluctuating interaural differences.

3. Optimized parameters for other models

(a) Lateral position models:

Tables 26 and 27 show that the first lateral position model (5) favors the transformed ILD over the transformed IPD in fitting the average listener data ($b = 0$). The other

lateral position models (6 and 7) mostly favor the transformed IPD ($b \approx 1$). This is in contrast to the best three independent center models (1–3), that weigh IPD and ILD as equally important. However, a lateral position model that uses only IPD ($b = 1$) or only ILD ($b = 0$) is equivalent to an independent centers model that uses only IPD ($a = 1$) or only ILD ($a = 0$). For example, an independent centers model that incorporates IPD fluctuations separately must lead to an r value that is at least as large as the r for the lateral position model. Therefore, as long as a and b are near zero or unity, model 1 will perform as least as well as model 5 in Figures 62 and 63.

(b) Short-term cross-correlation:

The STCC models (models 8–10) correlated least well with the data, possibly because only the IPD is used in these decision statistics. Integration time constants were noticeably larger for the STCC models ($\tau \approx 10$ ms) when compared to the other models ($\tau \approx 1$ ms). Model 10 produced an interesting result in that the threshold magnitude is approximately 0.06, which means that listeners are detecting instantaneous decorrelations from unity at coherence values of 0.94. This is larger than the jnds found by Gabriel and Colburn (1981), however, they studied the jnd for the entire duration of the stimulus and not decorrelations over a short time interval. Therefore, this seems like a reasonable result. Even though the STCC models correlate with the data least well, it appears that the approximation that yields Eq. 33, is not a bad approximation for this bandwidth.

(c) The advantage of preprocessing:

The results of modeling the *CAS* data with the preprocessing removed (no compression, no temporal averaging, and no critical envelope weighting) yielded $r_{\max} = 0.69$ for the independent centers model for the averaged listener data to be compared with 0.87 with preprocessing included, i.e. model 1. For the P_c data, $r_{\max} = 0.50$ without preprocessing compared to $r_{\max} = 0.68$ with preprocessing. Thus, these preprocessing assumptions prove beneficial in the modeling attempts. Without preprocessing r was

maximized by weighting ILD and IPD fluctuations by the ratio of 0.14 dB/degree for the *CAS* data (0.06 dB/degree for P_c). Yost and Hafter (1987) reported that the trading of intensity and phase should be 0.10 dB/degree for interaural phases less than 90° and should be 0.08 or 0.10 dB/degree for interaural phases greater than 90°. The ratio for the *CAS* data is higher than that of Yost and Hafter, although it does not seem to be excessively high. The difference may be because this experiment uses noise-pairs with dynamic fluctuations, whereas Yost and Hafter analyzed static interaural differences.

4.3 EXPERIMENT 14: CRITICAL BANDWIDTH

After testing the ten models against the 14-Hz bandwidth noise-pairs and coming to some preliminary conclusions, model performance was tested for the wider bandwidth of 108-Hz.

4.3.1 Method

The standard deviations of the 100 noise-pairs used in Experiment 14 can be seen in Figure 15 along with the means, standard deviations, and correlation of interaural parameters. It was the same collection with 108-Hz bandwidth from which particular noise-pairs were selected in Experiment 2 from Chapter 1. The same three listeners participated and the same procedure was used.

4.3.2 Results

The ten models tested with Experiment 13, were also tested with Experiment 14. The results of Experiment 14 are shown in Figure 65, entirely parallel to Figure 59 for Experiment 13. For Experiment 14, the *CAS* does not come close to saturating, but the percentage of correct responses does show a ceiling effect. Unlike Experiment 13, where the ceiling was reached for particular noise-pairs for all the listeners, the

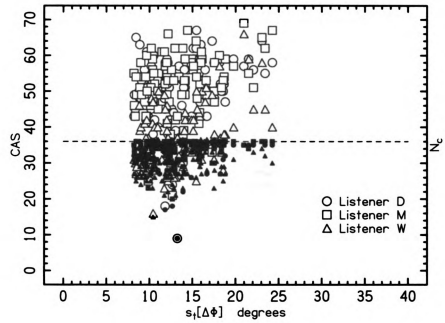

Figure 65: All the detection data for the 100 noise-pairs from Experiment 14 for three listeners, D, M, and W, are plotted twice, once as the number correct - on a scale from 0 to 36, and once as *CAS* - on a scale from 0 to 72. The data are plotted as a function of the standard deviation of the interaural phase for consistency with Figure 59.

ceiling in Experiment 14 is a factor for listeners D and M, but not necessarily W.

1. Comparison of model types

The results of the regression analysis for Experiment 14 are shown in Figures 66 and 67, parallel to Figures 60 and 61 for Experiment 13. The values of $r_{\max}$ are smaller than in Figure 61 because there is less variation in detectability for a band as wide as 108 Hz compared to a band with a 14-Hz width. Model 4, the independent center threshold-type model, gave the largest $r_{\max}$ for all three listeners and for the averaged data. For the averaged data, $r_{\max} = 0.55$.

Tables 28 and 29 show the values of the free parameters that maximized r for the 108-Hz bandwidth modeling. A value of Tables 28 and 29 is that they show that the parameters for the models at the 108-Hz bandwidth are mostly similar for different listeners. Consequently, the fits to the average listener shown in Figure 8

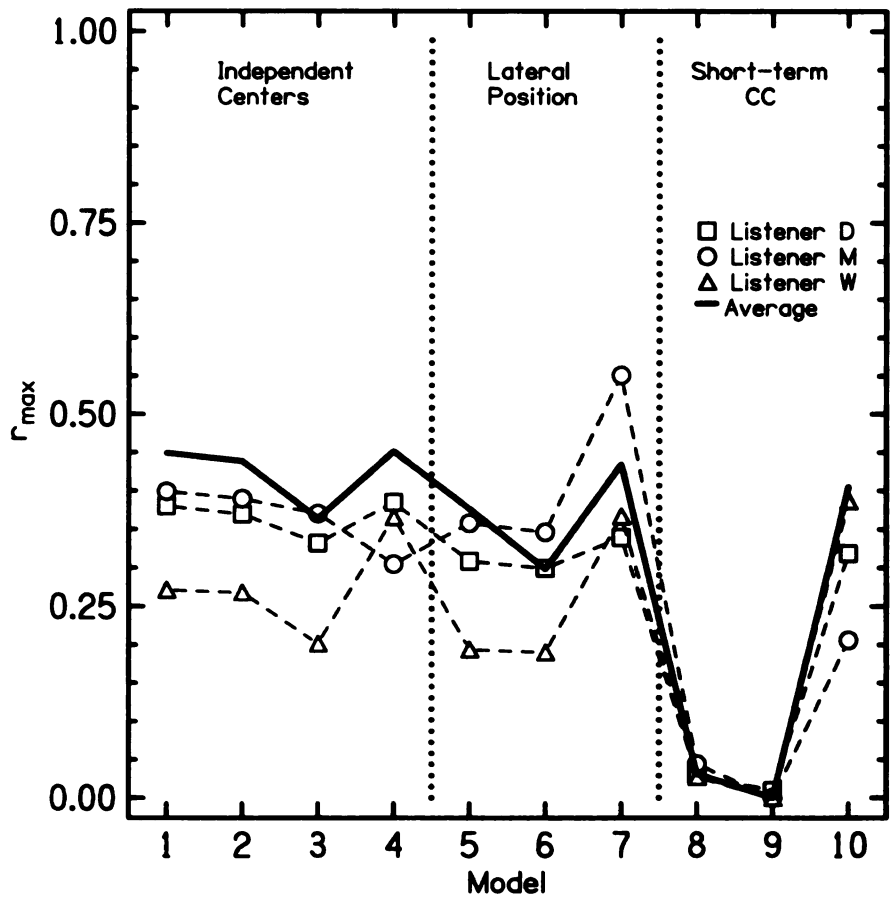

Figure 66: The comparison of P_c scores for the 108-Hz noise-pairs of Experiment 14 with ten models. The value of r_{max} shows the correlation between the experimental *CAS* scores for the 100 noise-pairs and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. The values of r_{max} are not as high as for the 14-Hz data shown in Figure 60, except that model 10 improves.

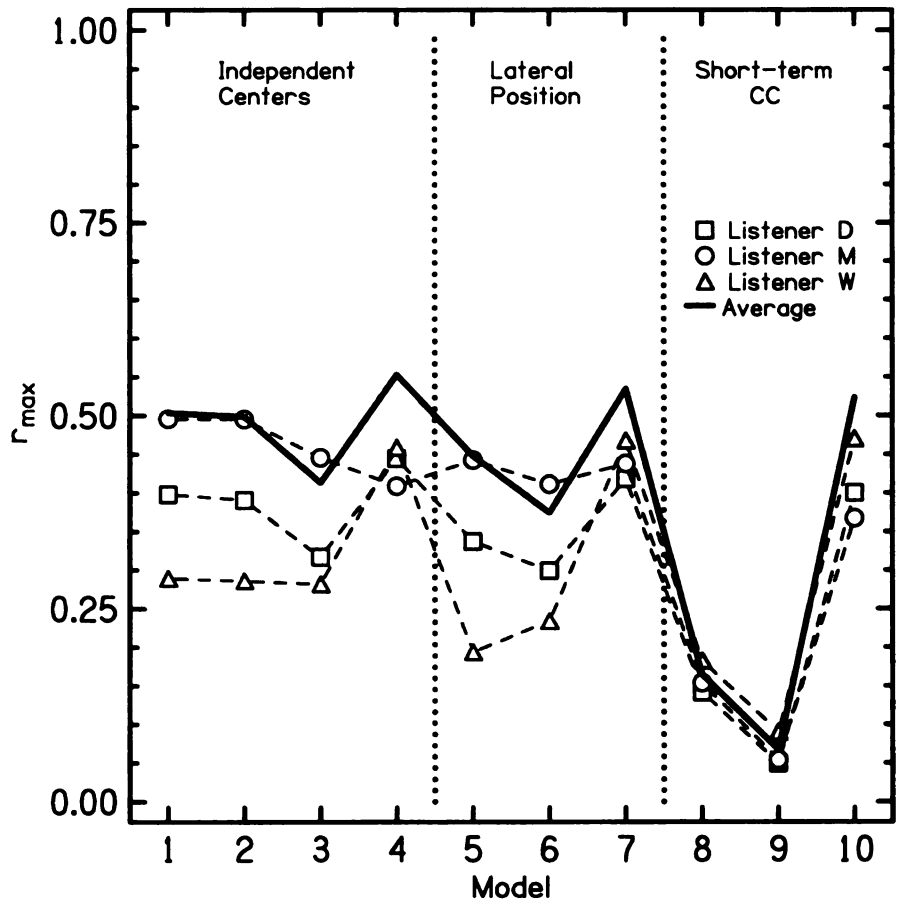

Figure 67: The comparison of *CAS* scores for the 108-Hz noise-pairs of Experiment 14 with ten models. The value of r_{max} shows the correlation between the experimental *CAS* scores for the 100 noise-pairs and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. The values of r_{max} are not as high as for the 14-Hz data shown in Figure 63, except that model 10 improves.

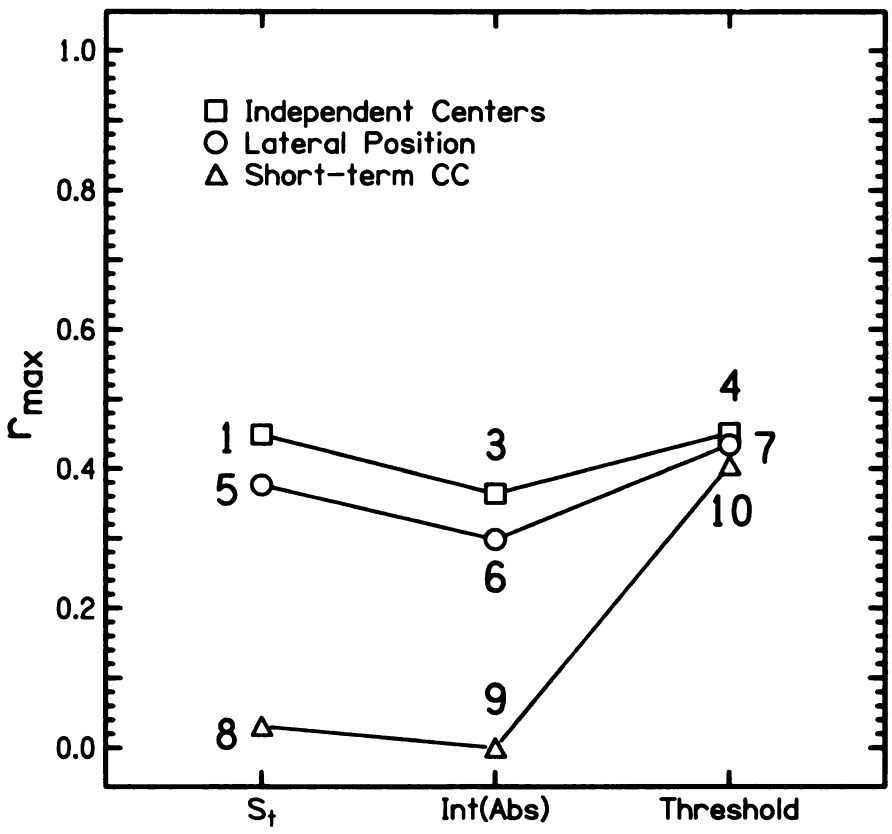

Figure 68: The comparison of each mathematical model for each type of binaural processing in the modeling of the 108-Hz bandwidth P_c data for the average listener. Model 2 is omitted from this plot.

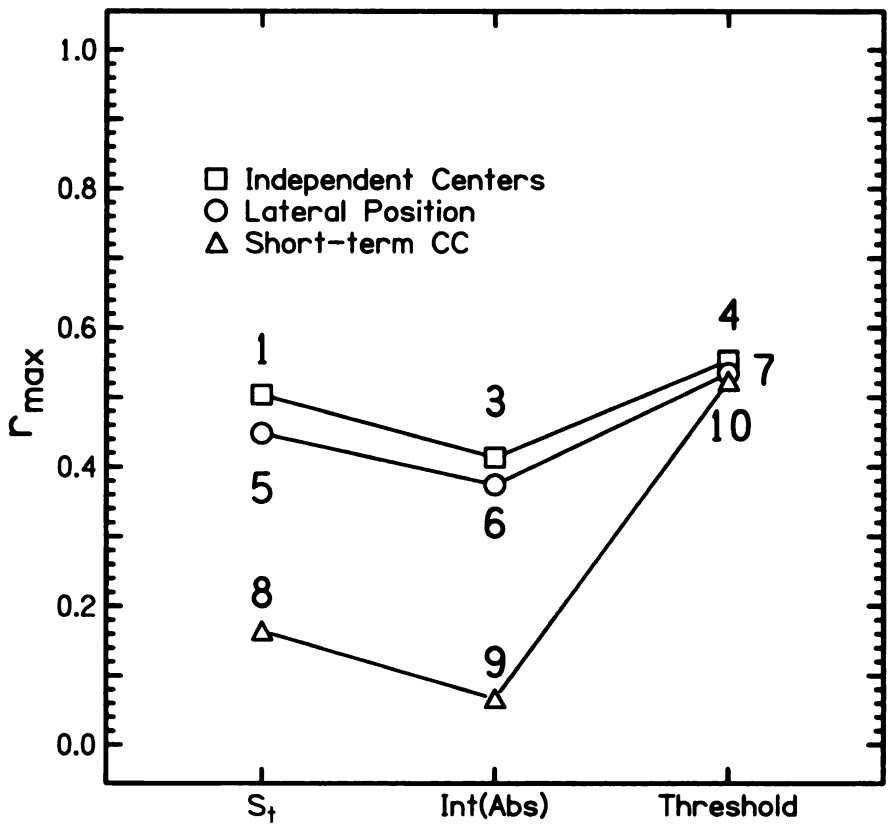

Figure 69: The comparison of each mathematical model for each type of binaural processing in the modeling of the 108-Hz bandwidth *CAS* data for the average listener. Model 2 is omitted from this plot.

are meaningful, although less convincing than the 14-Hz bandwidth fits.

What is interesting about the results shown in Figures 66 and 67 is that the independent centers models still outperform the lateral position models, though only marginally, and both types outperform the STCC models. As shown in Figures 68 and 69, all the threshold processing type models (4, 7, and 10) outperform the s_t processing type models, even model 1 - the most successful model in Experiment 13. The reason for this may be that models 4 and 7 have an extra free parameter. However, model 10 has the same number of free parameters as model 1. This could be evidence that a threshold statistic is used for detecting incoherence in larger bandwidth stimuli. Again, the models account better for the average listener than for any individual listener, with a few exceptions.

Table 28: Values of free parameters that optimize r in modeling the P_c detection results of Experiment 14 with 100 noise-pairs with a bandwidth of 108 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_1	D	2.0	0.77	0.03	-
	M	0.5	0.45	0.12	-
	W	0.5	0.60	0.70	-
	Ave	1.0	0.62	0.04	-
d_2	D	2.0	0.62	0.04	-
	M	0.5	0.46	0.12	-
	W	0.5	0.59	0.70	-
	Ave	1.0	0.54	0.04	-
d_3	D	1.0	0.53	0.06	-
	M	1.0	0.33	0.01	-
	W	0.5	0.55	0.70	-
	Ave	1.5	0.49	0.03	-
d_4	D	0.5	0.84	0.00	4.75
	M	0.5	0.90	0.00	6.25
	W	0.5	0.99	0.00	5.50
	Ave	0.5	0.91	0.00	5.50
d_5	D	0.5	0.00	0.00	-
	M	0.5	0.00	0.00	-
	W	2.5	0.00	0.00	-
	Ave	0.5	0.00	0.31	-
d_6	D	2.0	1.00	0.03	-
	M	0.5	0.27	0.07	-
	W	6.0	0.28	0.82	-
	Ave	1.5	1.00	0.03	-
d_7	D	1.0	0.91	0.00	3.75
	M	1.0	0.59	0.00	3.75
	W	0.5	0.98	0.00	5.25
	Ave	1.0	0.59	0.00	3.75
d_8	D	0.0	-	0.24	-
	M	0.0	-	0.24	-
	W	0.0	-	0.19	-
	Ave	0.0	-	0.24	-
d_9	D	0.0	-	0.58	-
	M	0.0	-	0.00	-
	W	0.0	-	0.00	-
	Ave	0.0	-	0.00	-
d_{10}	D	0.0	-	0.19	0.13
	M	4.0	-	0.20	0.02
	W	0.0	-	0.20	0.15
	Ave	0.0	-	0.19	0.15

Table 29: Values of free parameters that optimize r in modeling the *CAS* detection results of Experiment 14 with 100 noise-pairs with a bandwidth of 108 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_1	D	2.5	0.78	0.02	-
	M	0.5	0.52	0.02	-
	W	0.5	0.68	0.03	-
	Ave	0.5	0.51	0.03	-
d_2	D	2.5	0.63	0.02	-
	M	0.5	0.49	0.02	-
	W	0.5	0.56	0.03	-
	Ave	0.5	0.48	0.03	-
d_3	D	0.0	0.39	0.70	-
	M	0.5	0.29	0.02	-
	W	0.5	0.51	0.83	-
	Ave	0.5	0.32	0.03	-
d_4	D	0.5	0.91	0.00	5.00
	M	0.5	0.96	0.00	6.25
	W	0.5	0.97	0.00	5.50
	Ave	0.5	0.95	0.00	5.75
d_5	D	0.5	0.00	0.00	-
	M	0.0	0.00	0.00	-
	W	0.5	0.00	0.37	-
	Ave	0.5	0.00	0.00	-
d_6	D	2.5	1.00	0.02	-
	M	0.5	0.01	0.13	-
	W	0.0	0.60	0.83	-
	Ave	0.5	0.00	0.00	-
d_7	D	1.5	0.93	0.00	3.50
	M	2.0	0.51	0.00	3.50
	W	0.0	0.81	0.00	5.50
	Ave	0.5	0.99	0.00	6.00
d_8	D	0.0	-	0.24	-
	M	0.0	-	0.20	-
	W	0.0	-	0.25	-
	Ave	0.0	-	0.24	-
d_9	D	0.0	-	0.58	-
	M	0.0	-	0.19	-
	W	0.0	-	0.21	-
	Ave	0.0	-	0.20	-
d_{10}	D	0.0	-	0.20	0.13
	M	6.0	-	0.20	0.15
	W	0.0	-	0.11	0.19
	Ave	0.0	-	0.09	0.20

2. Optimized parameters for models 1 and 4

(a) Model 1:

Since model 1 was the best model for the 14-Hz data (Experiment 13) and model 1 still described the 108-Hz data (Experiment 14) well, model 1 will be analyzed in detail together with model 4, the most successful model at 108 Hz. For model 1, for the *CAS* data of Experiment 14, the best value of g is 0.03, which is smaller than the value of 0.19 found in Experiment 13. The best value of a is 0.51, which means that the transformed IPD and the transformed ILD still contribute equally, in good agreement with the value of $a = 0.50$ for the narrower bandwidth.

The comparison between free parameter values for model 1 can also be seen in Figure 70. Figure 70, which is parallel to Figure 64, shows how the free parameters change against each other when the goal is to maximize r for the average listener in the 108-Hz bandwidth experiment using model 1. Again, when a parameter does not explicitly appear in a plot, the best value of the parameter was used. The results of Figure 70 can be summarized as follows: Unlike Experiment 13, the fluctuations in Experiment 14 are fast enough to allow the experiment to determine an optimum integration time (recall that model 1 was insensitive to changes in τ for the 14-Hz bandwidth, as seen in Figure 64). The optimum integration time constant τ turns out to be 0.5 ms. This value is smaller than the value of 2.5 ms found by Viemeister (1979) in a monaural modulation transfer function measurement. It does not show any evidence of binaural sluggishness (Grantham and Wightman, 1978). This result indicates that the reason Experiment 13 found no value of τ was that the relevant temporal integration is much faster than the slow fluctuations present in the narrow-bandwidth noise-pairs.

Figure 70 also shows that the free parameter g may not be important for the larger bandwidth data. The figure shows that $r_{\max}$ occurs for $g = 0.03$. With a few exceptions, Table 29 shows that the values of g that yielded the best fit were smaller

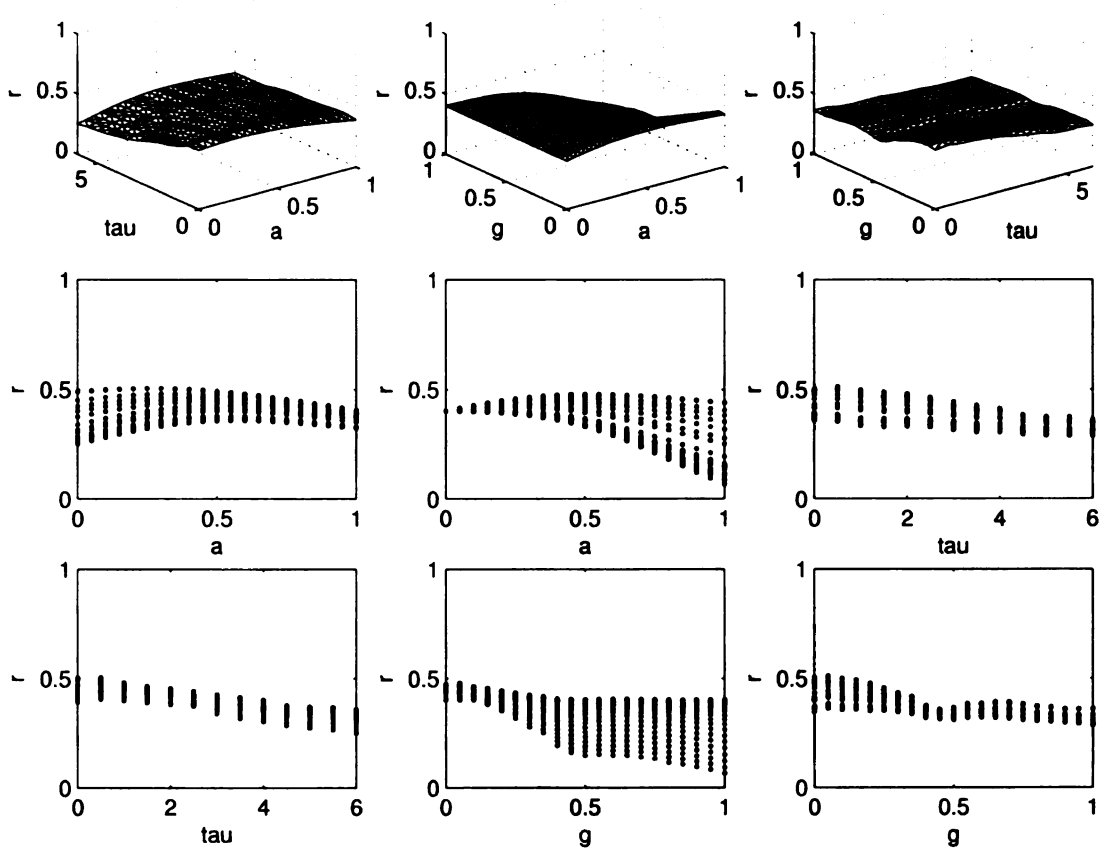

Figure 70: The free parameter surfaces for fitting model 1 to the average listener data from Experiment 14 with 108-Hz bandwidth. The plots are in the same form as Figure 64. Fixed parameters, not appearing along the axes, were given the optimum values for the average listener for model 1 in Table 29. At this bandwidth, a dependence upon τ becomes noticeable.

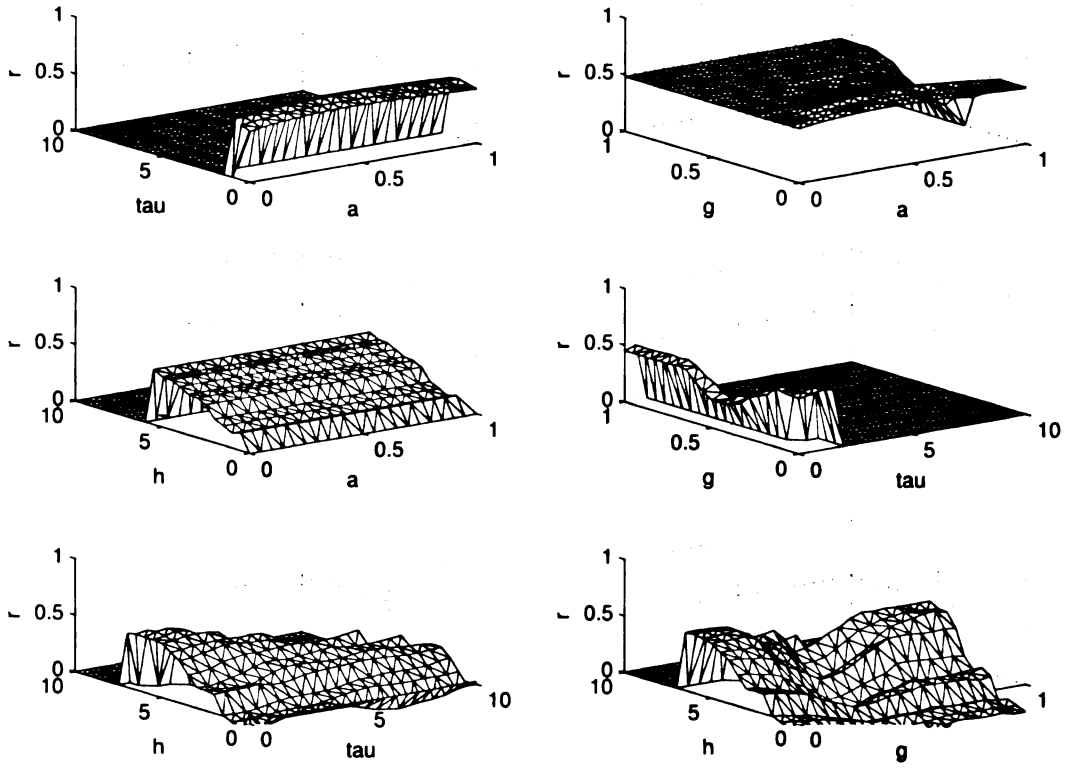

Figure 71: The free parameter surfaces for fitting model 4 to the average listener data from Experiment 14 with 108-Hz bandwidth. The free parameters are the trading parameter a, the integration time constant τ, the envelope weighting factor g, and the threshold level h. Fixed parameters, not appearing along the axes, were given the optimum values for the average listener for model 4 in Table 29. Results for model 4 show a strong interaction between h and g, a strong interaction between h and τ, and insensitivity to a.

than those in Experiment 13.

(b) Model 4:

Figure 71 shows how the four free parameters change for different values in model 4, the best model from Experiment 14. Each panel shows how two free parameters change while two others are kept constant. Like Figures 64 and 70, the constant parameters match the free parameters that yield $r_{\max}$. For model 4, when kept constant, $a = 0.95$, $\tau = 0.5$ ms, $g = 0$, and $h = 5.75$.

The upper-left panel shows a vs. τ. It is possible to see that this model is

insensitive to changes in a. Also, the value of τ can only be small (1.5 ms or less). A model that is insensitive to changes in a means that the binaural system cannot profit from using both IPD and ILD fluctuations. The reason for this may be because of the fluctuation rate of the interaural parameters.

In the upper-right panel, a vs. g, the insensitivity to a is seen, except for large values of g. Also, values of g near 0 yield the largest values of r. The middle-row-left panel shows that the optimum a is insensitive to different values of h, but a near 1 yields the best $r_{\mathbf{max}}$. Also, it is possible to see the sharp drop off of r for values of h greater than 6. The middle-row-right panel shows, that the best value of τ is 0.5 milliseconds and only values of g that are small can describe the detection data.

The lower-left panel shows that for no temporal averaging, model 4 can describe the data for large values of h up to 7.5. However, as the noise-pairs become more smoothed for larger values of τ (and therefore have less peaks above a high threshold), the model can only describe the data for values of h up to 3. This interaction between τ and h can be seen in all the panels of Figure 71 any time the value of r drops rapidly to zero.

Finally, the lower-right panel shows that large values of h can be used to describe the data when g is near zero. The largest values of r are seen in this vicinity. However, there is also a peak in r for large values of g and values of h around 4. Therefore, the interaction between g and h is strong and this effect can be seen in the upper four panels of Figure 71 when one of these free parameters is fixed.

3. Optimized parameters for other models

(a) Lateral position models:

Once again, Tables 26 and 27 show that models 5–7 at 108 Hz have values of b that are near 0 or 1 for most of the listeners. Therefore, these models best fit the data when they use only IPD or ILD information. However, a lateral position model that makes no use of one of the interaural differences is indistinguishable from an independent

centers of binaural processing model. In fact, lateral position models with $b = 1$ are identical to independent center models with $a = 1$. Therefore, in the case of the 108-Hz bandwidth data, it seems that independent centers may again be the better type of model, even if there is little distinction between models 1–7 by the values of $r_{\max}$.

(b) Short-term cross-correlation:

The STCC models follow the same shape as the other models in Figures 68 and 69, but they describe the data the much less successfully (except for model 10). Even though the best model (4) showed an insensitivity to the weighting of IPD and ILD, the STCC models do not seem able to use just the IPD fluctuations to describe the data as well as the other models with both IPD and ILD. For model 10, the magnitudes of the best thresholds were larger for the 108-Hz bandwidth when compared to the 14-Hz bandwidth. Also, model 10 is the only threshold model to have a non-zero value for g.

(c) The advantage of preprocessing:

As for Experiment 13, the preprocessing assumptions were removed to gauge their effect. For model 1 and average listener *CAS* data, $r_{\max} = 0.47$ without preprocessing can be compared with $r_{\max} = 0.50$ with preprocessing included. For model 1 and average listener P_c data, $r_{\max} = 0.37$ without preprocessing can be compared with $r_{\max} = 0.45$ with preprocessing included. As for Experiment 13, the preprocessing assumptions of compression, temporal averaging, and critical envelope weighting improved the agreement between model and data. However, it was a marginal gain when compared to the gain of using preprocessing at the narrow bandwidth. The best fit of *CAS* data without preprocessing was obtained by weighting ILD and IPD fluctuations in the ratio of 0.08 dB/degree, which impressively matches the ratio suggested in the review by Yost and Hafter (1987), namely 0.08–0.10 dB/degree. The P_c data led to a trading ratio of 0.02 dB/degree.

4.4 MULTIPLE PARAMETER MODEL

Model 1 appears to be one of the best models for both bandwidths, therefore attempts were made to improve upon this model. Since fluctuation rate appears to be an important factor (see Chapters 1–3), the fluctuation of the time derivative of the transformed interaural difference functions were added to the sum of interaural differences model. The fluctuation of the time derivative of the transformed interaural difference is defined as

$$s_t[\Psi'] = \sqrt{\frac{1}{T - T_D}\int_{T_D}^{T}\left[\frac{d}{dt}\Psi(t)\right]^2 dt} \tag{38}$$

where Ψ is either the transformed interaural phase or level. Also, since envelope fluctuations were found to be possibly influencing some listeners decisions (see Experiments 5 and 10), $s_t[E]$ and $s_t[E']$ were also included in the model, where

$$s_t[E'] = \sqrt{\frac{1}{T}\int_{0}^{T}\left[\frac{d}{dt}E(t)\right]^2 dt}. \tag{39}$$

In the end, six stimulus variables were considered to characterize the detection of incoherence in a noise-pair: $s_t[\Psi_{\Delta\Phi}]$, $s_t[\Psi'_{\Delta\Phi}]$, $s_t[\Psi_{\Delta L}]$, $s_t[\Psi'_{\Delta L}]$, $s_t[E]$, and $s_t[E']$. Thus, model 11 was of the form:

$$d_{11} = c_1 \cdot s_t[\Psi_{\Delta\Phi}] + c_2 \cdot s'_t[\Psi_{\Delta\Phi}] + c_3 \cdot s_t[\Psi_{\Delta L}] + c_4 \cdot s'_t[\Psi_{\Delta L}] + c_5 \cdot s_t[E] + c_6 \cdot s'_t[E]. \tag{40}$$

A multiple regression for model 11 was performed on the *CAS* data from Experiment 13. For this regression, $r = 0.88$. This is a small gain compared to model 1 which had $r = 0.87$ for just two variables, $s_t[\Psi_{\Delta\Phi}]$ and $s_t[\Psi_{\Delta L}]$.

Separate linear regressions were done for each of the six variables in model 11 for the *CAS* data for the 100 noise-pairs with a 14-Hz bandwidth. The values of r were: 0.76, 0.41, 0.73, 0.11, 0.66, and 0.33 for $s_t[\Psi_{\Delta\Phi}]$, $s_t[\Psi'_{\Delta\Phi}]$, $s_t[\Psi_{\Delta L}]$, $s_t[\Psi'_{\Delta L}]$, $s_t[E]$, and

$s_t[E']$ respectively. The derivative of the fluctuation functions do not describe the detection data as well as the fluctuation functions. Therefore, it appears that the derivative of the fluctuation functions can be omitted to simplify the model.

A multiple regression with $s_t[\Psi_{\Delta\Phi}]$, $s_t[\Psi_{\Delta L}]$, and $s_t[E]$ (model 11 with $c_2 = c_4 = c_6 = 0$) was performed on the *CAS* data from Experiment 13. No improvement was made from the previous value of $r = 0.87$ that was found in model 1. Therefore, it seems that model 1, which used just $s_t[\Psi_{\Delta\Phi}]$ and $s_t[\Psi_{\Delta L}]$, describes incoherence detection adequately when compared to the more complicated model 11.

4.5 DISCUSSION AND CONCLUSION

The goal of the experiments reported in this chapter was to determine the binaural characteristics that enable a listener to detect small amounts of incoherence in bands of noise.

4.5.1 Summary

Experiments 13 and 14 used 100 reproducible noise-pairs, with bandwidths of 14 and 108 Hz respectively, to test binaural models of incoherence detection. Each model was rated by varying its parameters (two, three, or four parameters) to find the best agreement with detection data. The data from Experiment 13 were the most informative. They showed that the independent center models outperformed the lateral position models, which outperformed the short-term cross-correlation models. The results of Experiment 14 were less informative because the models had smaller differences between the values of $r_{\max}$.

In the end, the best of the best-fitting models for 14 Hz was of the following form: *The binaural system detects incoherence on the basis of fluctuations in independently processed IPD and ILD channels, as though IPD and ILD were encoded at different centers without regard for the relative timing of their fluctuations. Nevertheless, there*

is a residual IPD-ILD interaction in that IPD fluctuations have no effect if the envelope in the left or right channel becomes smaller than about 20 percent of the RMS envelope value. The IPD and ILD fluctuations are compressed, and the compression established in steady-state experiments on lateral position is adequate to describe the compression of fluctuations. The processing centers register compressed fluctuations in IPD and ILD as measured by standard deviations over time, as integrated through an exponential window with a time constant of about 2 ms. The registered fluctuations are added on a lateral position scale at a more central site to form a decision statistic used to detect incoherence. For narrow bands near 500 Hz, IPD and ILD fluctuations are added with approximately equal weight. For wider bandwidth the relative weighting is unclear. As the bandwidth grows, different noises have increasingly similar fluctuations, and the detection of incoherence is better determined by the value of coherence itself.

The above bold statement about the best binaural model is based on a literal interpretation of the correlation coefficient values and must be taken with a grain of salt. It is doubtful that it is really possible to say that combining independent IPD and ILD fluctuations (model 1) is appreciably more successful than combining independent IPD and ILD mean square fluctuations (model 2).

A model that uses IPD and ILD independently might be supported by previous experiments. The dynamic fluctuations of the incoherent stimuli in this dissertation have interaural differences that change over short durations. The tracking of multiple images over short durations has been seen before in Hafter and Jeffress (1968) and Ruotolo *et al.* (1979), although never tied to incoherence detection.

The most interesting comparison is between models 1–4, which consider fluctuations in IPD and ILD independently, and models 5–7, which consider a fluctuation of the position of the lateral image. Figure 63 for 14-Hz bandwidth shows that the independent center models outperform the lateral position models for all three listeners

and for the average listener. Thus, a simple sign test favors models of the independent processing type. Table 29 shows that several values of b for the lateral position models were near a value of 0 or 1, so that only the IPD or only the ILD contributes to detection. That result means that in the optimizing process the lateral position models become unstable and become equivalent to independent binaural centers models. Nevertheless, the difference between the best independent centers model and the best lateral position model is not great, and the data do not permit a firm conclusion. The informal opinions of listeners in the experiments were that fluctuations in the perceived laterality of the noise image were the basis for incoherence detection. Whatever conclusion one reaches about models should not be in obvious disagreement with that opinion. However, listeners may not readily distinguish momentary image shifts, caused by ILD and IPD fluctuations in the same direction, from image broadening, caused by fluctuations in opposite directions.

4.5.2 Binaural processing

Like the experiments of Chapters 1–3, the experiments presented here conclude that coherence is an inadequate measure when the bandwidth is narrow. Coherence may be adequate in the wideband limit. Three results from Chapter 1 and the present chapter indicate features of the wideband limit. With increasing bandwidth: (1) The variance among different noises of the fluctuations of IPD and the fluctuations of ILD decreases. (2) The ability of listeners to detect incoherence varies less among different noise samples. (3) Different models of incoherence detection make predictions that are increasing similar, consistent with the prediction of Domnitz and Colburn (1976).

An important difference between the experiments with 14-Hz bandwidth and experiments with 108-Hz bandwidth is the speed of the fluctuations. On the basis of the experiments and modeling, there is agreement with Zurek and Durlach (1987) about the advantage of slow fluctuations, but it is not agreed that binaural slug-

gishness plays a role. The best fitting model leads to a binaural time constant of 0.5 ms for the 108-Hz bandwidth, commensurate with modulation transfer functions seen in such monaural tasks as the detection of amplitude modulation of broadband noise (Viemeister, 1979). The best fitting model found an insensitivity to τ in the region of 0.5 ms for the 14-Hz bandwidth. This time constant is not inconsistent with the matched-noises experiment in Chapter 1 wherein the slow fluctuations at 14-Hz bandwidth proved advantageous compared to the fluctuations at 108 Hz.

By contrast, binaural sluggishness is associated with time constants of tens, or even hundreds, of milliseconds. As suggested in the last paragraph of Hall *et al.* (1998), binaural sluggishness seems to arise in situations where both the masker and the signal plus masker contain dynamical interaural cues. If the masker is interaurally stable the binaural system can take advantage of events in brief epochs. The detection of a small amount of incoherence as a contrast to a diotic noise, as in the experiments, is well modeled as a stable masker (No) and a noise-like signal with a different phase relationship. A rapid response for such a task is consistent with other experiments cited by Hall *et al.*

The best integration time of 0.5 ms can be compared with the integration time of 300 ms found to be best in the loudness meter model of localization as calculated by Hartmann and Constan (2002). Thus, it seems that the binaural auditory system is capable of employing either short or long integration times depending on which better suits the task. When the task is to lateralize an image based on a binaural cue (loudness meter) the integration time is long. When the task is to detect rapid fluctuations in binaural cues, as in the work, the time is short. A similar point of view was taken with respect to monaural listening by Eddins and Green (1995) wherein integration times of several hundreds of milliseconds are possible for detecting the presence of a signal but times as short as several milliseconds are possible for detecting rapid signal variations.

5 BINAURAL MODELING: VARIED COHERENCE

This chapter is entirely parallel to Chapter 4 because it performs the same binaural modeling of fluctuations for large sets of incoherence detection data. However, it uses the noise-pairs from Chapter 2, the noise-pairs with varied values of coherence. It is predicted that the results of this chapter will be the same as Chapter 4, just as Chapters 1 and 2 had the same results.

5.1 EXPERIMENT 15: NARROW BANDWIDTH

5.1.1 Stimuli

The collection of 100 dual-channel noises with 14-Hz bandwidth from Experiment 6 in Chapter 2 was used in this experiment. The values of coherence ranged from 0.969 to 0.998, which can be seen in Figure 33. In the present experiment all 100 noise-pairs were used to avoid any bias. The fluctuations $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ are shown in Figure 34.

5.1.2 Procedure

The 100 noise-pairs were randomly divided into ten groups of ten. It was shown in Experiment 11 of Chapter 2 that context changes listeners' detection of incoherence only slightly for reproducible noises. This experiment employed three listeners from Chapter 2 - two males, M and W, and one female, E.

5.1.3 Results

The results from listening to all 100 noise-pairs can be seen in Figure 72, which is entirely parallel to Figure 59. The open symbols show the CAS while the solid symbols show the number of correct responses, essentially equivalent to the P_c. This

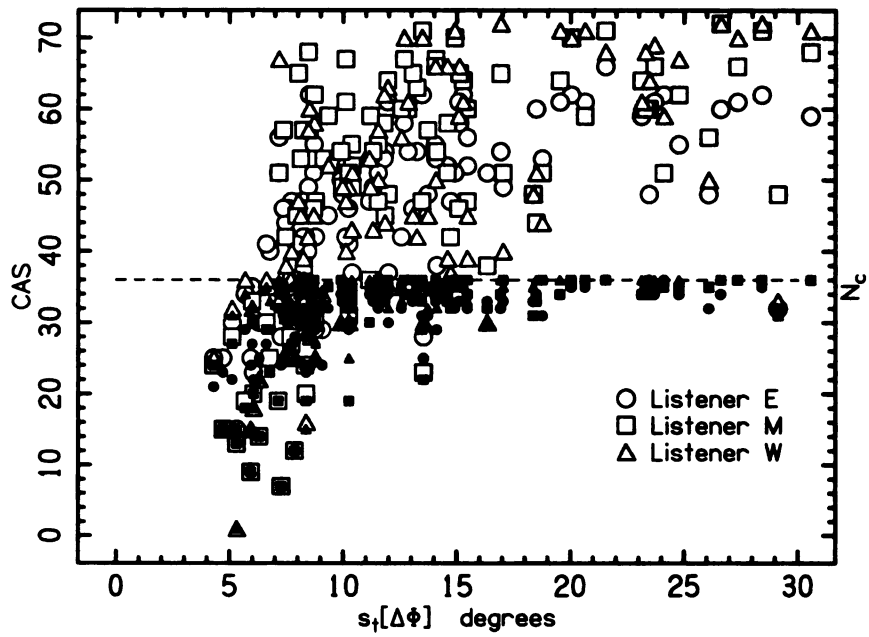

Figure 72: All the detection data for the 100 noise-pairs from Experiment 15 for three listeners, E, M, and W, are plotted twice, once as the number correct - on a scale from 0 to 36, and once as *CAS* - on a scale from 0 to 72. The data are plotted as a function of the standard deviation of the interaural phase in an attempt to give some order to the plot. Note that the scale of this plot is different from that in 59.

figure illustrates the advantage of using the *CAS* instead of P_c because the number of correct responses reaches a ceiling, especially for Listener M.

Agreement between the listeners for individual noise-pairs is difficult to see in Figure 72, but agreement is actually good. The inter-listener correlation for *CAS* for E and M was 0.76, for E and W was 0.83, and for M and W was 0.79. These values of the correlation are approximately the same as those reported in Experiment 13 (0.71, 0.73, and 0.80). These inter-listener correlations are somewhat smaller than those reported for the ten noise-pairs in Experiment 6 of Chapter 2 - approximately 0.9.

5.2 MODELS FOR INCOHERENCE DETECTION

The preprocessing functions and models used in this chapter are exactly the same as those used in Chapter 4.

5.2.1 Models vs. Experiment 15

The ten models presented above were tested against the data from Experiment 15. A linear regression of the form $y = mx + b$ was used to evaluate the effectiveness of a model to describe incoherence detection. The y-variable was the P_c or CAS values for the individual listeners or for an average over listeners. The x-variable was d_n from one of the ten models. The linear correlation coefficient, r, was used to compare the results of the regressions. The maximum r, $r_{\max}$, was found by independently varying all the free parameters over a reasonable space.

5.2.2 Comparison of model types

The results of the regressions are shown for the P_c and CAS data in Figures 73 and 74 for the very best combination (largest r) of all the free parameter values for each model. The r-values from the CAS data are consistently larger than the r-values from the P_c data. Like the experiments with fixed coherence noise-pairs, the most successful models agree better with the average listener than they do with any single listener for the CAS data, where ceiling effects were minimized. There are a few exceptions to this. For example, Figure 73 shows that the models for Listener E sometimes performs better than the models for the average listener, but this occurs only for the P_c data.

Figures 73 and 74 shows that models of the independent interaural difference type (models 1–4) were more successful than lateral position models (models 5–7). In fact, the worst independent centers model (model 3) described the CAS data better than the best lateral position model (model 5) for the average listener. This is different

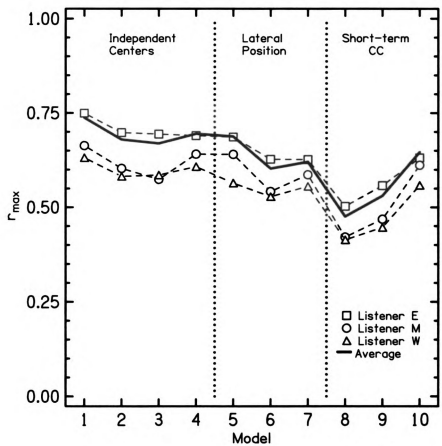

Figure 73: The comparison of P_c scores for the 14-Hz noise-pairs of Experiment 15 with ten models. The value of r_{max} shows the correlation between the experimental P_c scores for the 100 noise-pairs and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. It is not the average of r_{max} averaged over the listeners. Model 1 performs best with $r_{max} = 0.75$. Models 2–4 perform almost as well. Models 5–7 do not correlate as well with the data as models 1–4. Models 8–10 do not correlate as well with the data as models 5–7.

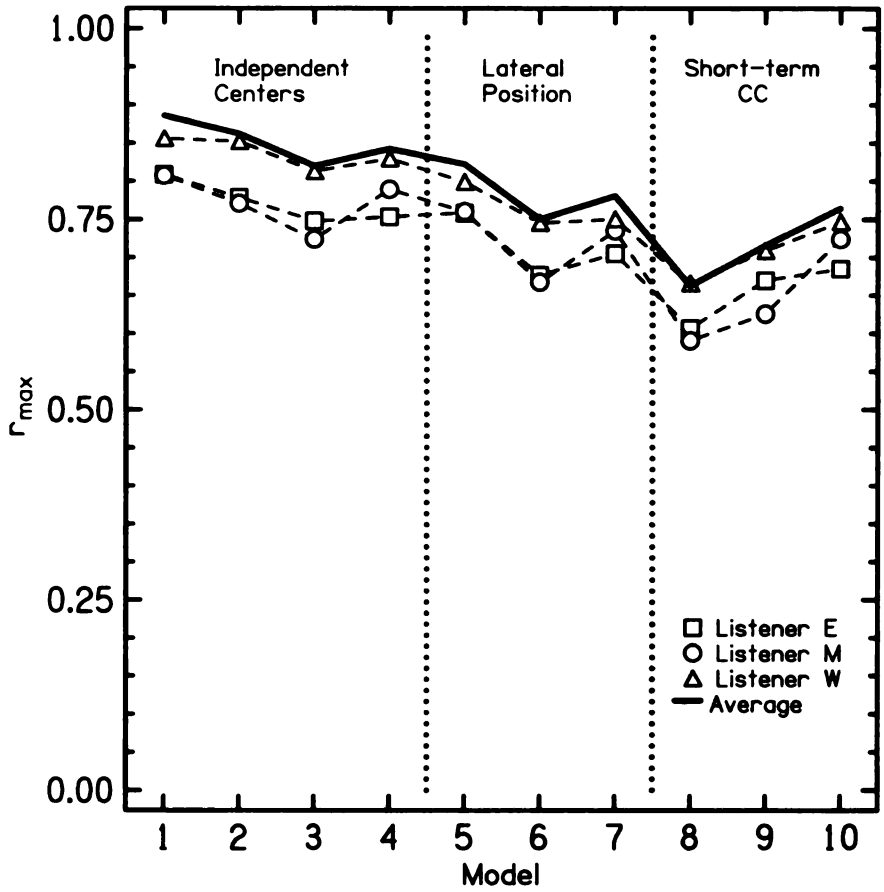

Figure 74: The comparison of *CAS* scores for the 14-Hz noise-pairs of Experiment 15 with ten models. The value of r_{max} shows the correlation between the experimental *CAS* scores for the 100 noise-pairs and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. It is not the average of r_{max} averaged over the listeners. Model 1 performs best with $r_{max} = 0.89$. Models 2–4 perform almost as well. Models 5–7 do not correlate as well with the data as models 1–4. Models 8–10 do not correlate as well with the data as models 5–7.

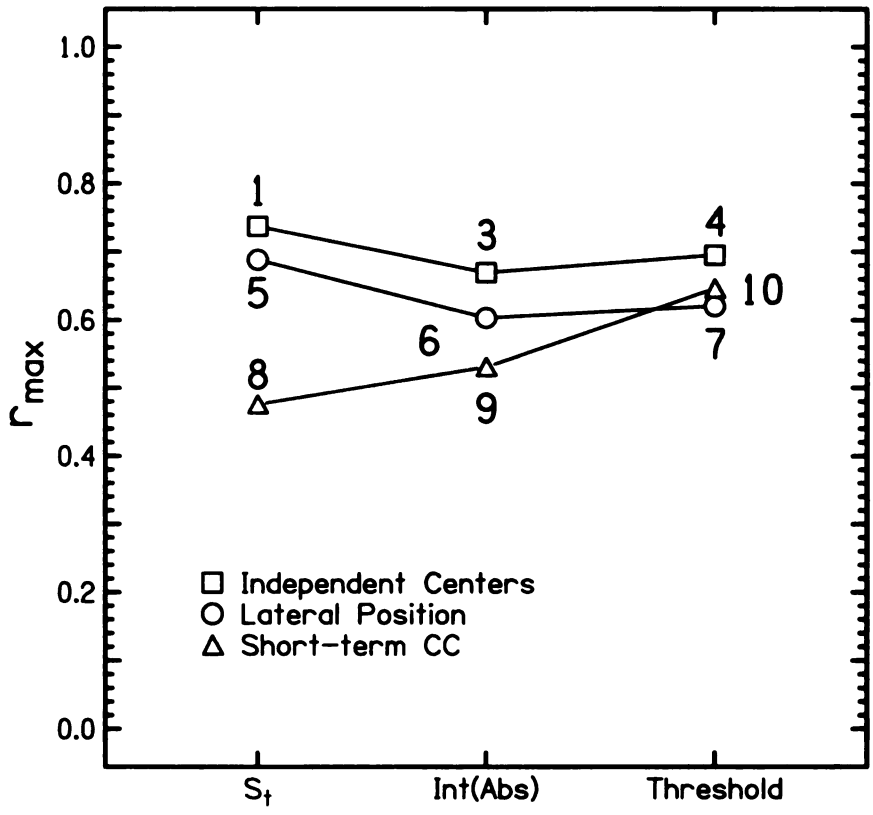

Figure 75: The comparison of each mathematical model (s_t ...) for each type of binaural processing (independent centers ...) in the modeling of the 14-Hz bandwidth P_c data. Model 2 is omitted from this plot.

than the result of the previous chapter that found the independent center threshold type model (model 4) had a value of r_{max} that was smaller than the r_{max} of lateral position models 5 and 6. In the present experiment, all seven of the independent center and lateral position models described the data better than the short-term cross-correlation (STCC) models. The correlation between the model results of the *CAS* data of Experiments 13 and 15 was 0.65. The correlation would have been larger had there been fewer changes in the order of the threshold models.

Model 1 had the largest r_{max} for all three listeners and for the averaged data. For the averaged P_c data, $r_{max} = 0.75$. For the averaged *CAS* data, $r_{max} = 0.89$. This can be compared to 0.68 and 0.87 for the P_c and *CAS* data of Chapter 4. These values show remarkably good agreement. The values of r_{max} for each type of model

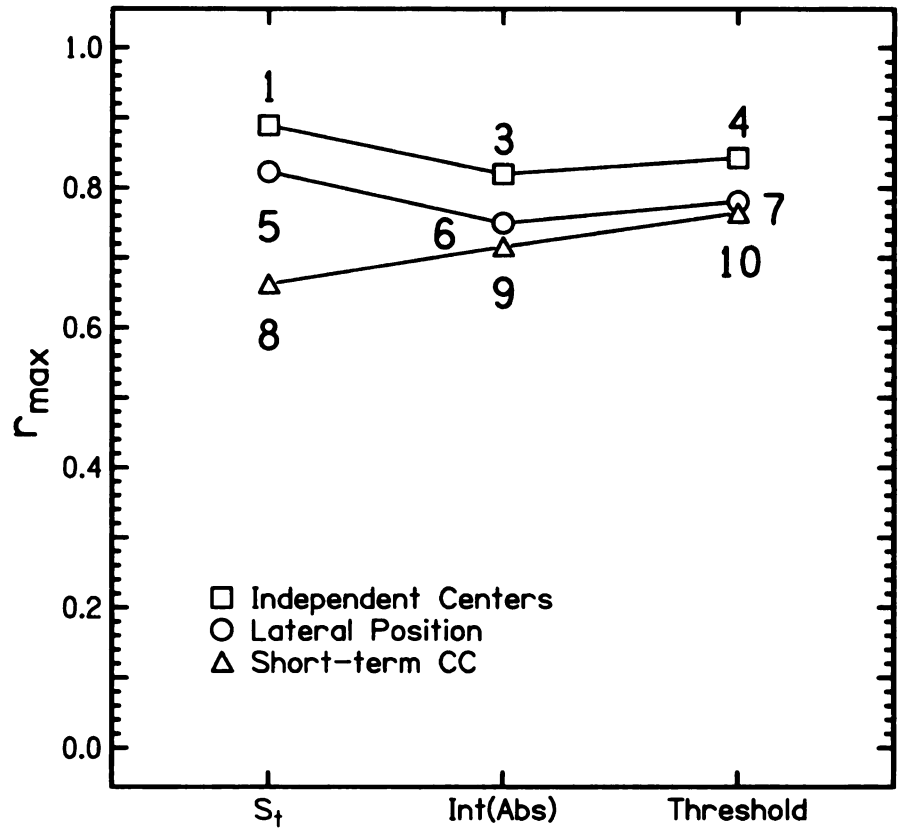

Figure 76: The comparison of each mathematical model (s_t ...) for each type of binaural processing (independent centers ...) in the modeling of the 14-Hz bandwidth *CAS* data. Model 2 is omitted from this plot.

(independent centers, lateral position, and STCC) can be more easily compared on Figures 75 and 76. The ordering of the models for the P_c data compared to the *CAS* data is consistent except that model 10 performs relatively better for the P_c data.

Tables 30 and 31 show the values of the free parameters for the maximized r for the 14-Hz bandwidth modeling. A value of Tables 30 and 31 is that they show that the parameters are similar for different listeners over the different types of models (independent center, lateral position, and STCC). Consequently, the fits to the average listener shown in Figures 73–76 are meaningful. Tables 30 and 31 also show that fitting parameters that optimize r are similar across models, to the extent that the models permit them to be compared. Tables 30 and 31 also show that fitting parameters that optimize r are similar across P_c and *CAS* data. The best fitting parameters of Experiment 15 (Tables 30 and 31) are comparable to those found in Experiment 13 (Tables 26 and 27).

Table 30: Values of free parameters that optimize r in modeling the P_c detection results of Experiment 15 with 100 noise-pairs with a bandwidth of 14 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_1	E	0.0	0.42	0.13	-
	M	0.0	0.32	0.12	-
	W	1.0	0.42	0.27	-
	Ave	0.0	0.38	0.12	-
d_2	E	0.0	0.46	0.13	-
	M	0.0	0.43	0.12	-
	W	0.5	0.48	0.27	-
	Ave	0.5	0.46	0.24	-
d_3	E	0.5	0.51	0.00	-
	M	0.5	0.45	0.00	-
	W	0.5	0.51	0.00	-
	Ave	0.5	0.49	0.00	-
d_4	E	2.5	0.45	0.00	3.00
	M	1.5	0.47	0.00	3.00
	W	0.5	0.47	0.00	1.75
	Ave	1.5	0.48	0.00	3.00
d_5	E	1.0	0.27	0.38	-
	M	0.5	0.24	0.42	-
	W	1.0	0.22	0.36	-
	Ave	0.5	0.24	0.41	-
d_6	E	1.5	1.00	0.00	-
	M	0.5	0.30	0.41	-
	W	1.5	1.00	0.00	-
	Ave	1.0	0.28	0.34	-
d_7	E	3.5	0.51	0.30	1.75
	M	2.5	0.80	0.00	2.75
	W	3.0	0.95	0.00	2.25
	Ave	2.5	0.76	0.00	2.50
d_8	E	1.0	-	0.27	-
	M	3.0	-	0.27	-
	W	0.0	-	0.36	-
	Ave	0.0	-	0.34	-
d_9	E	6.0	-	0.25	-
	M	5.0	-	0.25	-
	W	4.5	-	0.27	-
	Ave	5.0	-	0.25	-
d_{10}	E	5.0	-	0.00	0.02
	M	5.0	-	0.00	0.03
	W	4.0	-	0.00	0.02
	Ave	5.0	-	0.00	0.03

Table 31: Values of free parameters that optimize r in modeling the *CAS* detection results of Experiment 15 with 100 noise-pairs with a bandwidth of 14 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_1	E	0.0	0.43	0.12	-
	M	0.0	0.41	0.15	-
	W	1.0	0.45	0.21	-
	Ave	0.5	0.43	0.17	-
d_2	E	0.5	0.46	0.12	-
	M	0.0	0.47	0.18	-
	W	1.0	0.49	0.22	-
	Ave	0.5	0.47	0.18	-
d_3	E	0.0	0.50	0.00	-
	M	0.5	0.54	0.00	-
	W	0.5	0.54	0.00	-
	Ave	0.5	0.53	0.00	-
d_4	E	2.5	0.47	0.00	3.00
	M	0.0	0.46	0.08	3.50
	W	2.5	0.54	0.26	4.25
	Ave	0.5	0.47	0.11	3.75
d_5	E	1.0	0.35	0.30	-
	M	1.0	0.40	0.18	-
	W	3.0	0.76	0.01	-
	Ave	2.5	0.74	0.01	-
d_6	E	1.0	0.32	0.30	-
	M	2.5	1.00	0.00	-
	W	1.5	1.00	0.00	-
	Ave	2.0	1.00	0.00	-
d_7	E	4.0	0.70	0.02	2.50
	M	2.5	0.72	0.00	2.50
	W	1.5	0.84	0.00	3.25
	Ave	2.5	0.71	0.00	2.50
d_8	E	10.5	-	0.32	-
	M	0.5	-	0.25	-
	W	0.5	-	0.25	-
	Ave	0.5	-	0.25	-
d_9	E	6.5	-	0.25	-
	M	4.0	-	0.23	-
	W	5.0	-	0.25	-
	Ave	4.5	-	0.24	-
d_{10}	E	4.5	-	0.01	0.06
	M	5.0	-	0.00	0.03
	W	0.0	-	0.00	0.07
	Ave	2.5	-	0.00	0.06

5.2.3 Optimized parameters for model 1

The most successful model was model 1, and Figure 77 shows how the free parameters change against each other to maximize r for the average listener *CAS* data in that model. Figure 77 was generated by varying two parameters and keeping the other constant. Plots of a vs. τ assume that $g = 0.17$, plots of g vs. τ assume that $a = 0.43$, and plots of a vs. g assume that $\tau = 0.5$. Results of other conditions (P_c vs. *CAS*, individual listeners vs. average listener) produced very similar plots to the extent that it is necessary to show only the most general condition.

The results of Figure 77 can be summarized as follows:

(a) Integration time:

As for Experiment 13, the integration time constant τ was found to be unimportant. The characteristic stimulus fluctuations, expected to be of order 1/14 seconds, are slow compared to the time constants tested in detail, 0–6 ms.

In addition to the detailed test for integration times less than 6 ms, longer integration times were specifically tested. Therefore spot checks were performed trying to fit *CAS* data with longer values of τ. For increasing values of τ = 0, 25, 50, 75, 100, 125, and 150 ms, the value of r decreased monotonically. The value of $r_{\max}$ dropped from $r_{\max} \approx 0.8$ for $\tau = 0$ ms, to $r_{\max} \approx 0.6$ for $\tau = 25$ ms, to $r_{\max} \approx 0.4$ for τ = 150 ms. As in Experiment 13, it is concluded that there is no useful role for binaural sluggishness in model 1 - the best model.

(b) Critical envelope weighting:

According to the regression analysis, the best value of g for model 1 is about 0.17, though the $r_{\max}$ is insensitive to g for g in this vicinity. Applying critical envelope weighting actually solved the problem of noise-pair #79 from Experiment 6 in Chapter 2 because the major IPD feature in this noise-pair is a large peak during the last 30 ms of the signal when the envelope is small.

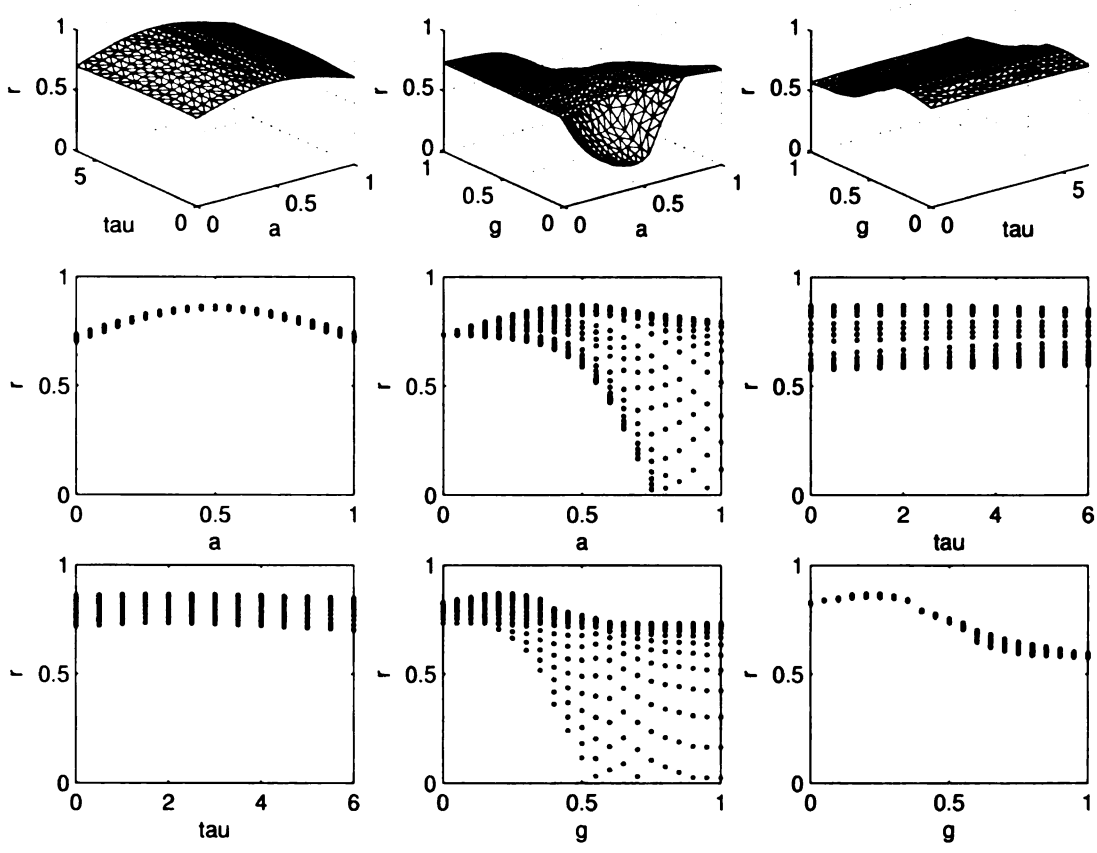

Figure 77: The free parameter surfaces for fitting model 1 to the average listener *CAS* data from Experiment 15 - one-hundred noise-pairs with 14-Hz bandwidth. In the upper-left panel, r is plotted against a and τ for $g = 0.17$. In the upper-middle panel, r is plotted against g and τ for $a = 0.43$. In the upper-right panel, r is plotted against a and g for $\tau = 0.5$. The two panels below an upper panel flatten one of the free parameter dimensions. The variations of free parameters create smooth surfaces. At this bandwidth, the dependence upon τ is negligible.

(c) Relative IPD-ILD contributions:

The best value of a for model 1 is 0.43, approximately 0.5, which means that transformed IPD and ILD values contribute about equally to the sensation of incoherence. This can be compared to $a = 0.50$ for the fixed coherence noise-pairs in Experiment 13.

5.2.4 Optimized parameters for other models

(a) Lateral position models:

Tables 30 and 31 show that the lateral position models (5–7) favor the transformed IPD over the transformed ILD in fitting the average listener data. This is in contrast to the four independent center models (1–4), that weigh IPD and ILD as about equally important. In the extreme case of model 6, only the IPD appears ($b = 1$) for two of the listeners and for the averaged CAS data. However, an lateral position model that uses only IPD ($b = 1$) is equivalent to an independent centers model that uses only IPD ($a = 1$). Therefore, an independent centers model that incorporates ILD fluctuations separately must lead to an r value that is at least as large as the r for the lateral position model. Hence, model 3 must perform as least as well as model 6 in Figures 60 and 61. This result was also seen in Chapter 4.

(b) Short-term cross-correlation:

The STCC models (models 8–10) correlated least well with the data, possibly because only IPD is used in these decision statistics. Model 10 produced an interesting result in that the threshold magnitude is approximately 0.02 for P_c and 0.05 for CAS, which means that listeners are detecting instantaneous decorrelations from unity at coherence values of 0.95–0.98. The magnitude of the decorrelations are slightly smaller than those found in Chapter 4.

(c) The advantage of preprocessing:

The *CAS* results of modeling the data with the preprocessing removed (no compression, no temporal averaging, and no critical envelope weighting) yielded $r_{max} = 0.72$ for the independent centers model for the averaged listener data to be compared with 0.89 with preprocessing included, i.e. model 1. Thus, these preprocessing assumptions prove beneficial in the modeling attempts. Without preprocessing r was maximized by weighting ITD and IPD fluctuations by the ratio of 0.08 dB/degree. Yost and Hafter (1987) reported that the trading of intensity and phase should be 0.10 dB/degree for interaural phases less than 90° and should be 0.08 or 0.10 dB/degree for interaural phases greater than 90°. The agreement is impressive.

5.2.5 Summary

Experiment 15 modeled detection data from noise-pairs with different values of coherence for three listeners, one of whom (Listener E) did not participate in Experiment 13. The results of the modeling of the P_c and *CAS* data in Experiment 15 were essentially the same as Experiment 13. Model 1 was the best model with $r_{max} = 0.89$ (0.87 for Experiment 13) for the average listener *CAS* data. The values of the free parameters for the *CAS* data were $\tau = 0.5$ ms (2 ms for Experiment 13), $a = 0.43$ (0.50 for Experiment 13), and $g = 0.17$ (0.19 for Experiment 13). The results of models between Experiments 13 and 15 correlated at a value of 0.65. Other notable results were also consistent between Experiments 13 and 15, such as the independent center models outperforming the lateral position models outperforming the STCC models, the insensitivity to changes in τ, and the advantage of preprocessing.

5.3 EXPERIMENT 16: CRITICAL BANDWIDTH

A great deal of data had been collected for noise bandwidths approximately equal to a critical bandwidth. During an extensive course of incoherence detection experiments,

listeners heard many different sets of noises with a bandwidth of 136 Hz. From the two collections used in Experiments 7A, 7B, and 7C, data were collected for 60 different noise-pairs.

5.3.1 Method

The 60 noise-pairs used in Experiment 16 can be seen in Figure 78 along with the means, standard deviations, and correlation of interaural parameters. The noise-pairs of Experiment 16 could be considered to be a random set of noises with a bandwidth of 136 Hz similar to the narrow band set of Experiment 15. Such a description would not be entirely fair because each of the 60 noise-pairs in Experiment 16 were selected according to some criterion or another (see Experiments 7A–7C), depending on the intuitions of the experimenter at the time. However, for a bandwidth as large as 136 Hz, the process of selection does not lead to large differences among waveforms. The means and IPD-ILD correlations of the 60 noise-pair set (Figure 78) are similar to those for the unselected set of 100 noise-pairs shown in Figure 15 of Chapter 2. The standard deviations of the 60 noise-pairs are only 40 percent greater. Experiment 16 was the methodologically the same as Experiment 15. The listeners DY, E, M, T, and W participated.

Another difference between Experiments 15 and 16 arises from using the data from Experiments 7A, 7B, and 7C from Chapter 2. Some noise-pairs were presented in several different sets. Noise-pairs from Experiments 13–15 were guaranteed to be in only one set. However, for this experiment, some noise-pairs were in as many as four different sets. For these noise-pairs, the average was taken for the P_c or CAS value.

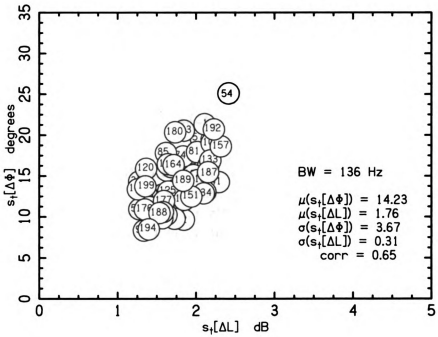

Figure 78: Fluctuations of IPD versus fluctuations of ILD for the 60 reproducible noise-pairs having a 136-Hz bandwidth, as used in Experiment 16. Each noise-pair is labelled by the serial number assigned in Experiments 7A, 7B, and 7C. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. The means of the interaural fluctuations are the similar to those shown in Figure 15; the standard deviations are slightly larger.

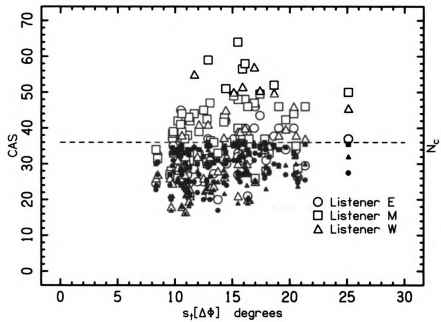

Figure 79: All the detection data for the 60 noise-pairs from Experiment 16 for three listeners, E, M, and W, are plotted twice, once as the number correct - on a scale from 0 to 36, and once as *CAS* - on a scale from 0 to 72. The data are plotted as a function of the standard deviation of the interaural phase for consistency with Figure 72. Listeners DY and T were omitted from the plot for the sake of clarity. Note that the scale of this plot is different from that in Figure 65.

5.3.2 Results

The results of Experiment 16 are shown in Figure 79, entirely parallel to Figure 72 for Experiment 15. For Experiment 16, the *CAS* does not come close to saturating, but the percentage of correct responses does show a ceiling effect. Unlike Experiment 15, where the ceiling was reached for particular noise-pairs, the ceiling in Experiment 16 is a factor for Listener M, but not necessarily Listeners E and W.

5.3.3 Comparison of model types

The results of the regression analysis for Experiment 16 are shown in Figures 80 and 81, parallel to Figures 73 and 74 for Experiment 15. The values of $r_{\mathbf{max}}$ are smaller than in Figures 73 and 74 because there is less variation in detectability for a band

as wide as 136 Hz compared to a band with a 14-Hz width. One can conjecture that if the 60 noise-pairs of this experiment had not been selected then the value of r_{max} might have been even smaller. Smaller values of r were seen for the unselected 100 noise-pairs in Experiment 14 compared to Experiment 16. Model 4 gave the largest r_{max} for all three listeners and for the averaged data. For the averaged *CAS* data, $r_{max} = 0.77$. Unlike Experiments 13 and 15, but like Experiment 14, the P_c and *CAS* modeling results yield approximately the same value of r_{max}. From this it is concluded that the *CAS* does not offer appreciably more information when the majority of listeners are not near the ceiling of P_c.

Tables 32, 34, 36 show the values of the free parameters that maximized r for the 136-Hz bandwidth modeling of the P_c data. Tables 33, 35, 37 show the values of the free parameters for the *CAS* data. A value of these tables is that they show that the parameters for the models at the 136-Hz bandwidth are similar for different listeners. Consequently, the fits to the average listener shown in Figures 80–83 are meaningful.

What is interesting about the results shown in Figures 80 and 81 is that the independent center models still outperform the lateral position models, though only marginally, and both types outperform the STCC models. However, the differences between models, best seen in Figures 82 and 83, are often small, and they do not permit clear statements about the ordering. Several threshold models, models 4 and 7, outperform the s_t models, models 1 and 5 - the most successful models in Experiment 15. The reason for this may be that models 4 and 7 have an extra free parameter. Again, the models often account better for the average listener than for any individual listener, especially for the *CAS* data. The correlation between different models' r_{max} for the *CAS* data in Experiments 14 and 16 was 0.95.

In comparing Figures 82 and 83, there is a slight, but small, increase in r_{max}. This difference between modeling the P_c and *CAS* data was much larger for the smaller bandwidth data of Experiments 13 and 15, but comparable for Experiment 14.

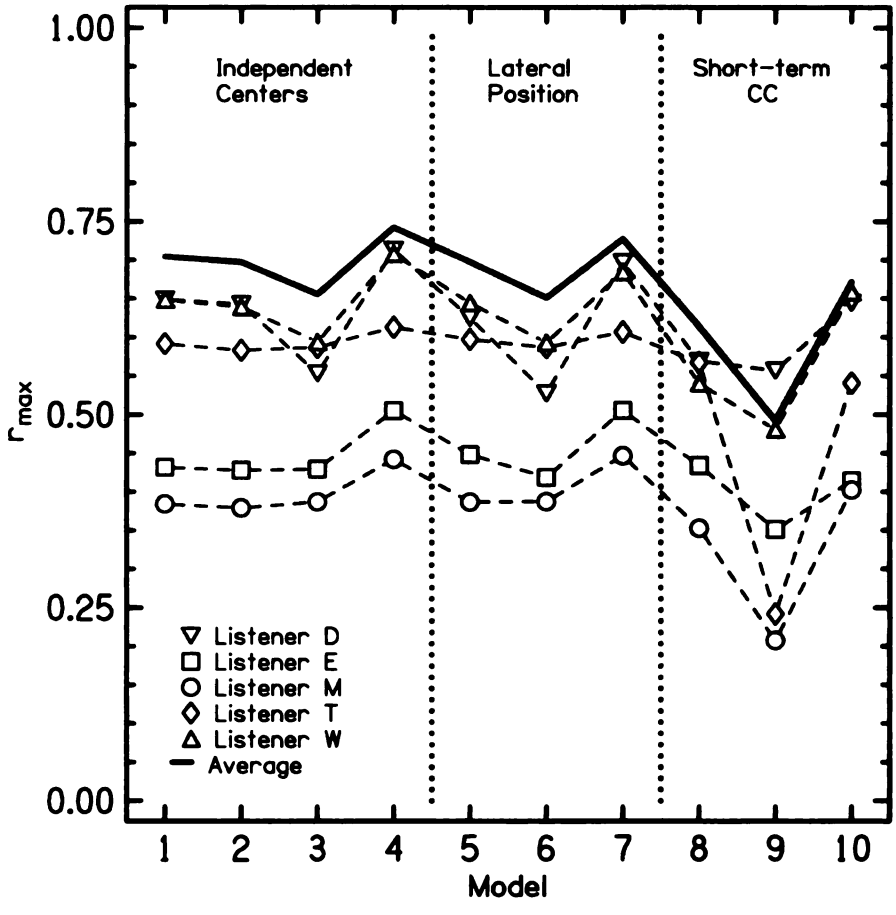

Figure 80: The comparison of P_c scores for the 136-Hz noise-pairs of Experiment 16 with ten models. The value of r_{max} shows the correlation between the experimental P_c scores for the 60 noise-pairs and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. The values of r_{max} are not as high as for the 14-Hz data shown in Figure 60, except that Models 8 and 9 improve. The difference between most models appears to be small.

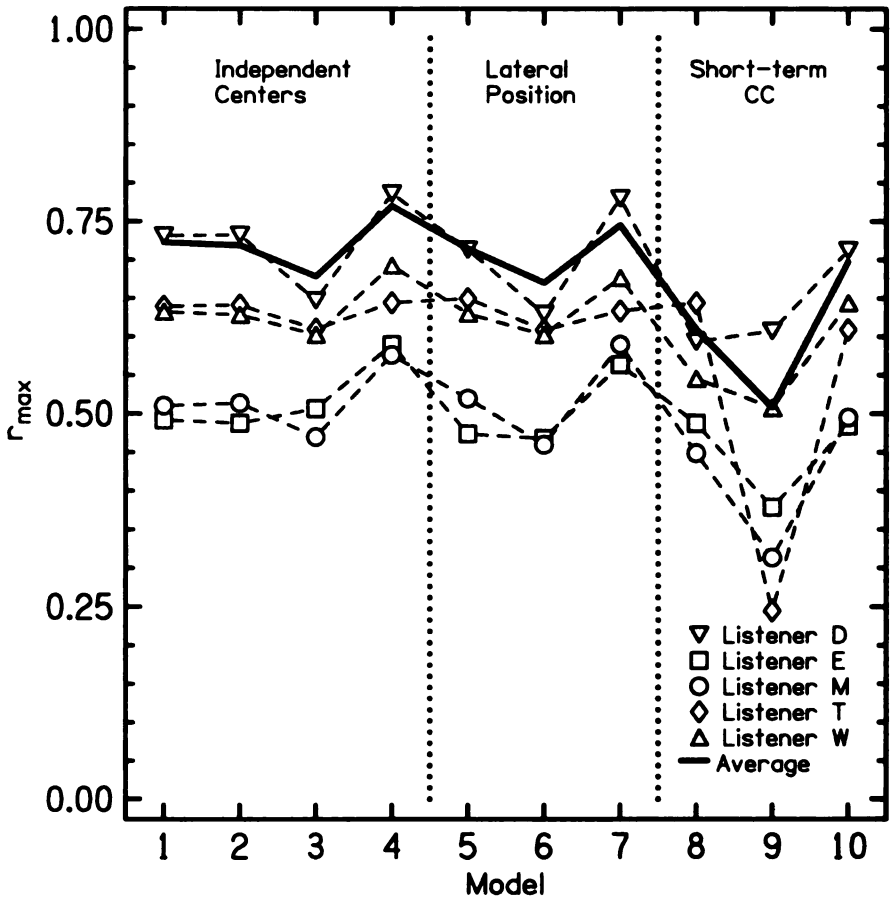

Figure 81: The comparison of *CAS* scores for the 136-Hz noise-pairs of Experiment 16 with ten models. The value of r_{max} shows the correlation between the experimental *CAS* scores for the 60 noise-pairs and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. The values of r_{max} are not as high as for the 14-Hz data shown in Figure 74, except that Models 8 and 9 improve. The difference between most models appears to be small.

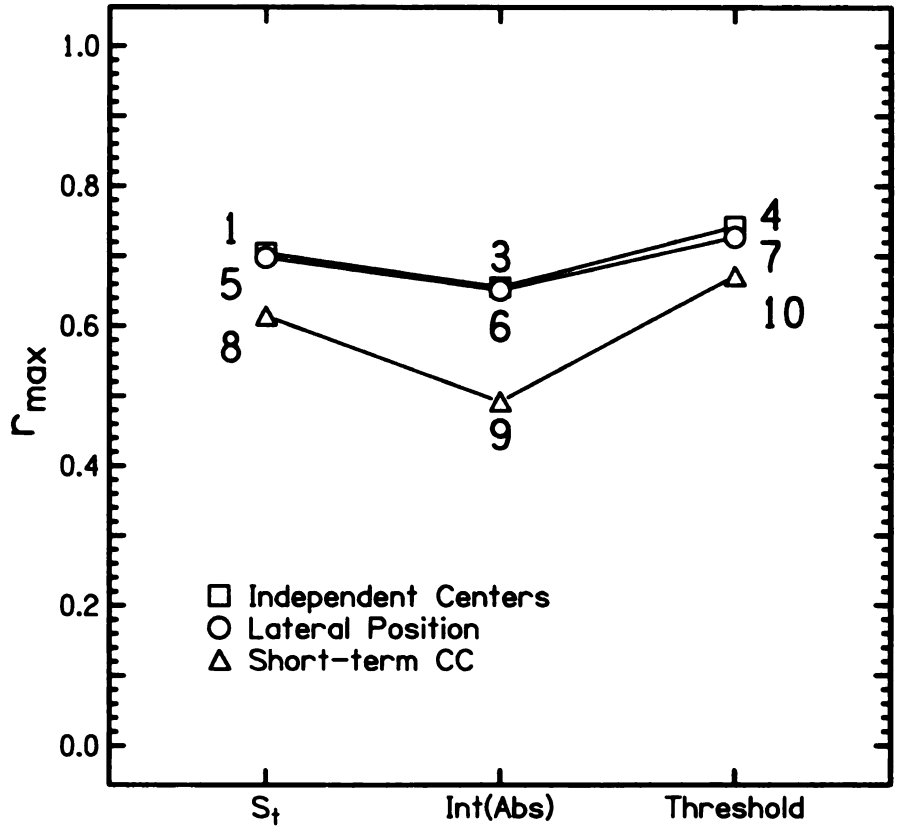

Figure 82: The comparison of each mathematical model (s_t ...) for each type of binaural processing (independent centers ...) in the modeling of the 136-Hz bandwidth P_c data. Model 2 is omitted from this plot.

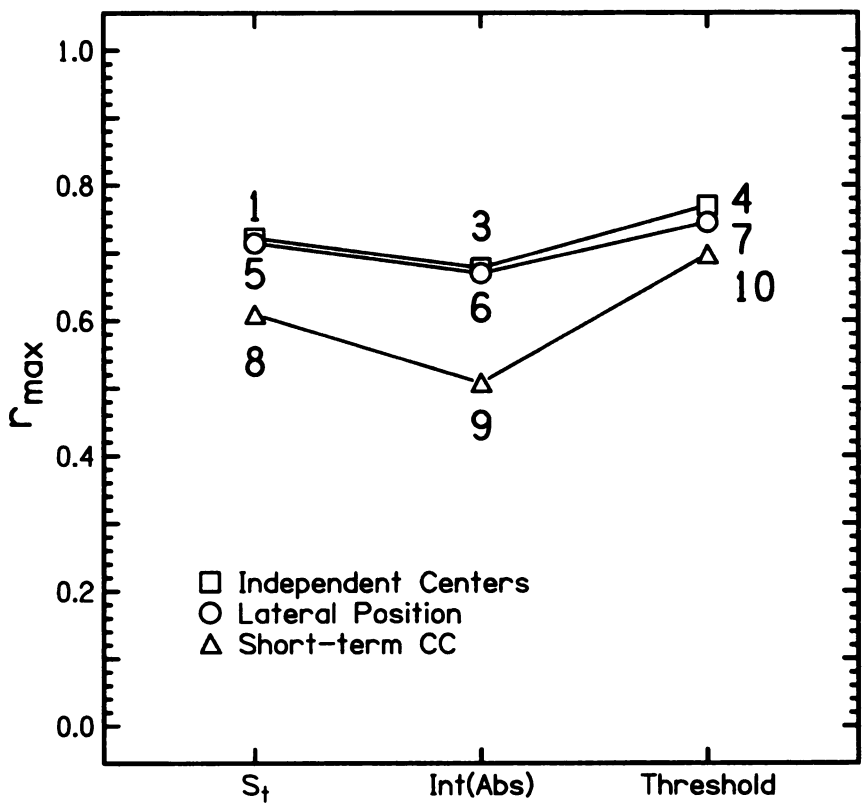

Figure 83: The comparison of each mathematical model (s_t ...) for each type of binaural processing (independent centers ...) in the modeling of the 136-Hz bandwidth *CAS* data. Model 2 is omitted from this plot.

Table 32: Values of free parameters that optimize r in modeling the P_c detection results of Experiment 16 with 60 noise-pairs with a bandwidth of 136 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_1	DY	1.0	0.61	0.06	-
	E	0.0	0.40	0.13	-
	M	1.0	1.00	0.22	-
	T	1.0	1.00	0.36	-
	W	0.5	0.83	0.13	-
	Ave	1.0	0.82	0.14	-
d_2	DY	1.0	0.51	0.06	-
	E	0.0	0.43	0.13	-
	M	1.0	1.00	0.22	-
	T	1.0	1.00	0.36	-
	W	1.0	0.62	0.14	-
	Ave	1.0	0.65	0.14	-
d_3	DY	1.0	0.56	0.02	-
	E	4.5	0.66	0.02	-
	M	2.0	1.00	0.20	-
	T	1.5	1.00	0.21	-
	W	1.0	0.91	0.06	-
	Ave	1.5	0.84	0.06	-
d_4	DY	1.5	0.65	0.04	3.75
	E	3.5	0.75	0.07	2.25
	M	0.0	0.53	0.18	8.75
	T	1.0	1.00	0.33	3.00
	W	1.0	0.65	0.28	3.25
	Ave	1.0	0.72	0.26	3.50

5.3.4 Optimized parameters for models 1 and 4

Since model 1 was the best model for the 14 Hz data (Experiment 15) and model 1 still described the 136 Hz data (Experiment 16) well, model 1 will be analyzed in detail together with model 4, the most successful model at 136 Hz. Again only the averaged CAS will be analyzed. For model 1, the best value of g is 0.13, close to the value of 0.17 found in Experiment 15. The best value of a is about 0.75, which means that the transformed IPD contributes somewhat more than the transformed ILD. This result can be compared with the value of $a = 0.43$ for the narrower bandwidth.

Table 33: Values of free parameters that optimize r in modeling the *CAS* detection results of Experiment 16 with 60 noise-pairs with a bandwidth of 136 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_1	DY	1.0	0.65	0.06	-
	E	1.0	0.66	0.13	-
	M	2.0	0.67	0.30	-
	T	0.5	0.98	0.36	-
	W	1.0	0.87	0.09	-
	Ave	1.0	0.74	0.17	-
d_2	DY	1.0	0.54	0.06	-
	E	1.0	0.56	0.13	-
	M	3.0	0.56	0.74	-
	T	0.5	0.86	0.36	-
	W	1.0	0.69	0.09	-
	Ave	1.0	0.61	0.14	-
d_3	DY	1.5	0.68	0.02	-
	E	6.0	0.56	0.06	-
	M	2.0	0.68	0.22	-
	T	1.0	0.89	0.30	-
	W	1.0	1.00	0.05	-
	Ave	1.5	0.79	0.09	-
d_4	DY	0.5	0.42	0.17	5.25
	E	6.0	0.64	0.13	2.00
	M	3.5	0.32	0.09	2.75
	T	1.0	1.00	0.31	3.00
	W	3.0	0.80	0.06	2.75
	Ave	3.0	0.61	0.08	2.75

This is unlike the result from the fixed coherence noise-pairs, which was also found to be near $a = 0.5$ for the 108-Hz bandwidth.

For model 4, the values of a and g are similar to those for model 1. However, there is a difference in τ as model 4 has longer integration times, 3–6 ms, compared 1–2 ms for model 1.

The comparison between free parameter values for models 1 and 4 can also be seen in Figures 84 and 85 respectively. Figure 84 shows how the free parameters change against each other when the goal is to maximize r for the average listener in the 136-

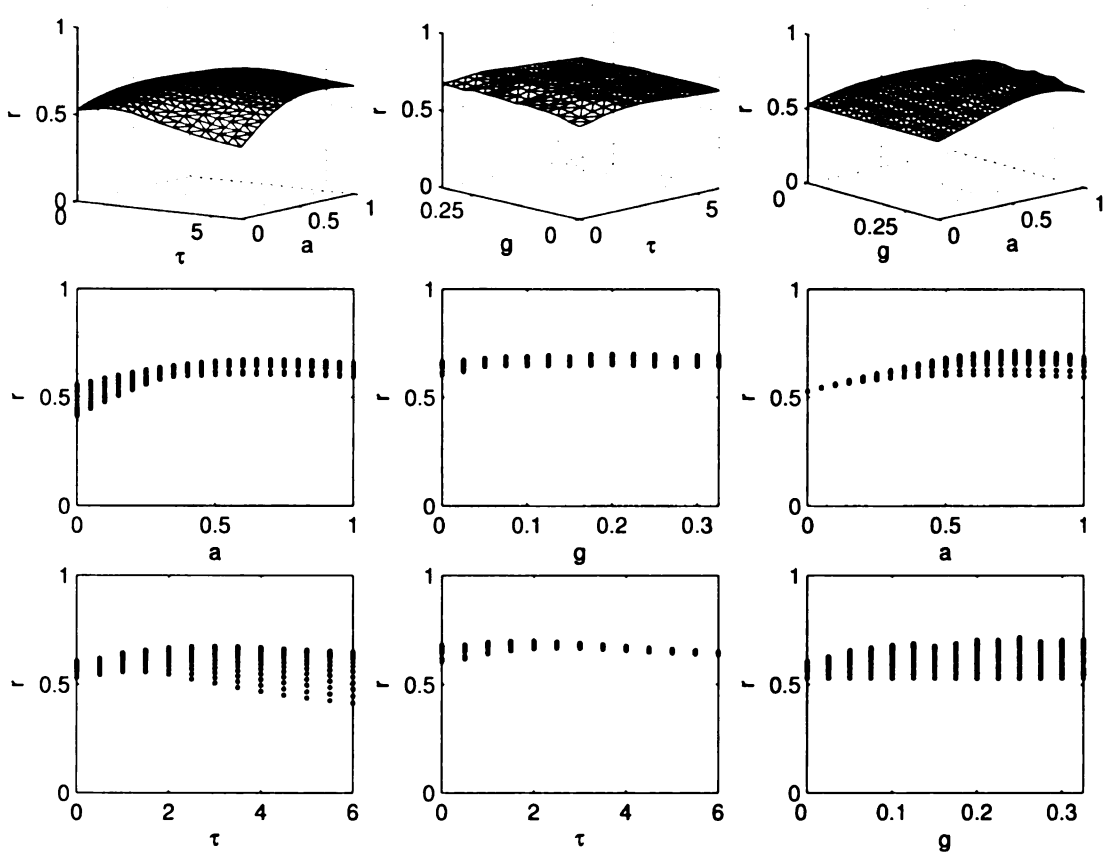

Figure 84: The free parameter surfaces for fitting model 1 to the average listener *CAS* data from Experiment 16 - 60 noise-pairs with 136-Hz bandwidth. The plots are in the same form as Figure 77. Fixed parameters, not appearing along the axes, were given the optimum values for the average listener for model 1 in Table 33. At this bandwidth, a dependence upon τ becomes noticeable.

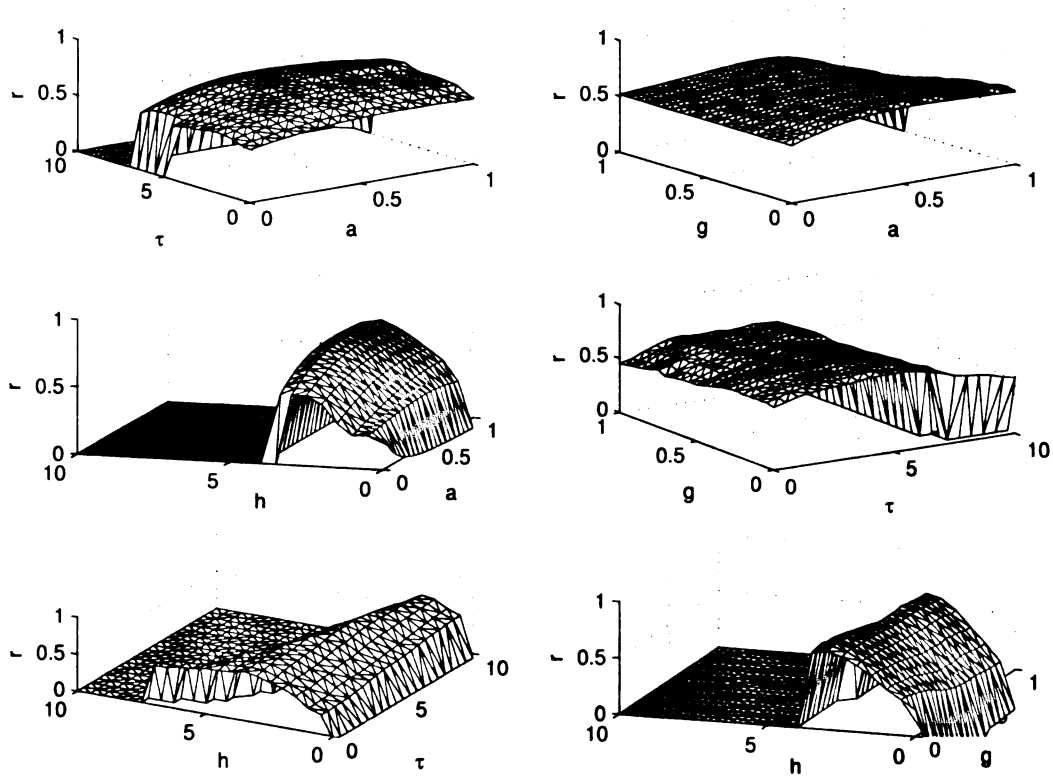

Figure 85: The free parameter surfaces for fitting model 4 to the average listener *CAS* data from Experiment 16 - 60 noise-pairs with 136-Hz bandwidth. Fixed parameters, not appearing along the axes, were given the optimum values for the average listener for model 4 in Table 33. Results for model 4 are not inconsistent with model 1.

Table 34: Values of free parameters that optimize r in modeling the P_c detection results of Experiment 16 with 60 noise-pairs with a bandwidth of 136 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_5	DY	1.0	1.00	0.06	-
	E	0.0	0.43	0.36	-
	M	1.0	0.90	0.22	-
	T	1.0	0.88	0.36	-
	W	0.5	0.93	0.13	-
	Ave	1.0	1.00	0.14	-
d_6	DY	1.0	1.00	0.02	-
	E	1.5	1.00	0.05	-
	M	2.0	0.96	0.20	-
	T	1.5	1.00	0.21	-
	W	1.0	1.00	0.06	-
	Ave	1.5	1.00	0.06	-
d_7	DY	1.5	1.00	0.04	3.75
	E	1.5	0.97	0.12	3.25
	M	0.5	0.56	0.55	3.00
	T	2.0	0.88	0.30	2.75
	W	2.0	0.97	0.09	3.00
	Ave	1.5	0.94	0.10	3.25

Hz bandwidth experiment using model 1. This figure is entirely parallel to Figure 77. Again, when a parameter does not explicitly appear in a plot, the best value of the parameter was used. The results of Figure 84 can be summarized as follows: Unlike Experiments 13 and 15, but like Experiment 14, the fluctuations in Experiment 16 are fast enough to allow the experiment to determine an optimum integration time. The optimum integration time constant τ turns out to be 1 or 2 ms. This value agrees with the value found by Viemeister (1979) in a monaural modulation transfer function measurement. This result indicates that the reason Experiment 15 found no value of τ was that the relevant temporal integration is much faster than the slow fluctuations present in the narrow-bandwidth noise-pairs.

Figure 85 shows how the four free parameters change for different values in model **4**, the best model from Experiment 16. Each panel shows how two free parameters

Table 35: Values of free parameters that optimize r in modeling the *CAS* detection results of Experiment 16 with 60 noise-pairs with a bandwidth of 136 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_5	DY	1.0	0.94	0.02	-
	E	1.0	1.00	0.13	-
	M	0.5	0.65	0.17	-
	T	0.5	0.84	0.36	-
	W	1.0	1.00	0.09	-
	Ave	0.5	0.76	0.17	-
d_6	DY	1.5	1.00	0.02	-
	E	1.0	1.00	0.05	-
	M	2.0	0.82	0.22	-
	T	1.0	1.00	0.21	-
	W	1.0	1.00	0.05	-
	Ave	1.5	1.00	0.09	-
d_7	DY	1.5	0.73	0.02	3.25
	E	0.5	0.24	0.93	1.75
	M	0.5	0.55	0.55	3.00
	T	0.0	0.71	0.66	2.25
	W	2.0	0.97	0.08	3.00
	Ave	2.0	0.97	0.08	3.25

change while two others are kept constant. Like Figures 77 and 84, the constant parameters match the free parameters that yield $r_{\max}$. For model 4, when kept constant, $a = 0.60$, $\tau = 3$ ms, $g = 0.08$, and $h = 2.75$. The upper-left panel shows a vs. τ. It is possible to see $r_{\max}$ occur for values of a between 0.5 and 1.0 and τ about 3 ms. In the upper-right panel, a vs. g, $r_{\max}$ occurs for a greater than 0.5 and values of g tending towards 0. The middle-row-left panel shows that the optimum a is relatively insensitive to different values of h, but a greater than 0.5 yields the best $r_{\max}$. Also, it is possible to see the sharp drop off of r for values of h greater than 3. The middle-row-right panel shows, although difficult to see, that the best value of τ is 3 ms. The lower-left panel shows that for no temporal averaging, model 4 can describe the data for large values of h up to 7.5. However, as the noise-pair becomes more smoothed for larger values of τ (and therefore the interaural differences have

Table 36: Values of free parameters that optimize r in modeling the P_c detection results of Experiment 16 with 60 noise-pairs with a bandwidth of 136 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_8	DY	0.0	-	0.14	-
	E	0.0	-	0.13	-
	M	1.0	-	0.44	-
	T	0.0	-	0.46	-
	W	0.0	-	0.35	-
	Ave	0.0	-	0.14	-
d_9	DY	15.0	-	0.00	-
	E	14.5	-	0.00	-
	M	2.0	-	0.00	-
	T	2.5	-	0.00	-
	W	12.0	-	0.00	-
	Ave	8.0	-	0.00	-
d_{10}	DY	0.0	-	0.04	0.12
	E	0.0	-	0.06	0.03
	M	0.0	-	0.17	0.04
	T	0.0	-	0.37	0.03
	W	0.0	-	0.26	0.05
	Ave	0.0	-	0.13	0.04

fewer instances of time above a high threshold), the model can only describe the data for values of h up to 2.5. This interaction between τ and h can be seen in all the panels of Figure 85 any time the value of r drops rapidly to zero. Finally, the lower-right panel shows that the optimum value of g is relatively constant for a given h. The differences between Figures 71 and 85 are due to the different fixed parameters. For example, the value of h is smaller in this experiment and model 4 is sensitive to changes in h.

5.3.5 Optimized parameters for other models

(a) Lateral position models:

As for Experiments 13–15, models 5–7 for the 136-Hz bandwidth have values of b that are 1 or near 1 for some listeners. Therefore, these models best fit the data when

Table 37: Values of free parameters that optimize r in modeling the *CAS* detection results of Experiment 16 with 60 noise-pairs with a bandwidth of 136 Hz.

Model	Listener	τ (ms)	a, b	g (% of RMS)	h (LP)
d_8	DY	0.0	-	0.26	-
	E	0.0	-	0.13	-
	M	0.0	-	0.30	-
	T	0.0	-	0.46	-
	W	0.0	-	0.50	-
	Ave	0.0	-	0.44	-
d_9	DY	15.5	-	0.00	-
	E	10.5	-	0.00	-
	M	15.0	-	0.00	-
	T	2.5	-	0.00	-
	W	13.5	-	0.00	-
	Ave	15.0	-	0.00	-
d_{10}	DY	0.0	-	0.04	0.06
	E	0.0	-	0.06	0.03
	M	0.0	-	0.27	0.14
	T	0.0	-	0.32	0.05
	W	0.0	-	0.26	0.06
	Ave	0.0	-	0.26	0.07

they use only IPD information and no ILD information. However, a lateral position model that makes no use of one of the interaural differences is indistinguishable from an independent centers of binaural processing model. In fact, lateral position models with $b = 1$ are identical to independent center models with $a = 1$. Therefore, in the case of the 136-Hz bandwidth data, it seems that independent centers may again be the better type of model, even if there is little distinction between models 1–7 by the values of $r_{\max}$.

(b) Short-term cross-correlation:

The STCC models follow the same shape as the other models in Figure 81, but they describe the data the much less successfully. Even though the independent center models favor IPD fluctuations at the 136-Hz bandwidth, the STCC models do not seem able to use just the IPD fluctuations to describe the data as well as the other

models with both IPD and ILD. For model 10, the magnitudes of the best thresholds were comparable for the 14-Hz and 136-Hz bandwidths.

(c) The advantage of preprocessing:

As for Experiment 15, the preprocessing assumptions were removed to gauge their effect. For model 1 and average listener *CAS* data, $r_{max} = 0.56$ without preprocessing can be compared with $r_{max} = 0.73$ with preprocessing included. As for Experiment 15, the preprocessing assumptions of compression, temporal averaging, and critical envelope weighting improved the agreement between model and data. The improvement of the modeling without preprocessing compared to with preprocessing is much larger for this experiment when compared to Experiment 16. The best fit without preprocessing was obtained by weighting ILD and IPD fluctuations in the ratio of 0.05 dB/degree which is slightly lower than suggested in the review by Yost and Hafter (1987).

5.3.6 Summary

Experiment 14 from Chapter 4 and Experiment 16 from this chapter had many of the same major results. For example, model 4 was the best model type, followed by model 1 and the ordering of the models was also the same. The results of the models correlated at a value of 0.95.

There were four notable differences in the experimental method when comparing Experiments 14 and 16. First, only 60 selected noise-pairs were used, not 100 unselected noise-pairs. Second, some noise-pairs were presented in multiple sets in this experiment. Third, there were different listeners and a different number of listeners between experiments. Fourth, the bandwidth was 136 Hz instead of 108 Hz. These four differences may explain some of the differences in the results for Experiments 14 and 16. First, the values of r_{max} were about 0.2 higher for Experiment 16 when compared to Experiment 14. Model 1 had $r_{max} = 0.67$ (0.50 for Experiment 14) and

model 4 had $r_{max} = 0.74$ (0.55 for Experiment 14). Model 1 maximized r for $\tau = 1$ ms (0.5 for Experiment 14), $a = 0.74$ (0.51 for Experiment 14), and $g = 0.13$ (0.04 for Experiment 14). Model 4 maximized r for $\tau = 3$ ms (0.5 for Experiment 14), $a = 0.60$ (0.95 for Experiment 14), $g = 0.08$ (0.00 for Experiment 14), and $h = 2.75$ (5.75 for Experiment 14). Lastly, the advantage of using preprocessing yielded a gain of 0.17 in r (0.03 for Experiment 14).

5.4 CONCLUSION

The results of this chapter were very similar to those in Chapter 4 results, as expected. Modeling the noise-pairs with varied coherence makes little difference to modeling the noise-pairs with fixed coherence as long as the distributions of the fluctuations are the same. The value of this chapter is that it shows the results of Chapter 4 are reproducible with different listeners.

6 PROBING MODELS WITH SELECTED NOISE-PAIRS

The modeling experiments of Chapters 4 and 5 were motivated by the observation that some noise-pairs in Chapters 1 and 2 were clearly out of the order expected from choosing noise-pairs with the largest and smallest fluctuations in phase and level. In this chapter, the best model from Chapters 4 and 5 (model 1) will be used to select noise-pairs for certain properties. For example, noise-pairs with only phase fluctuations or only level fluctuations will be chosen. Also, stimuli will be chosen to test the independent centers model against the lateral position model.

6.1 EXPERIMENT 17: MINIMUM FLUCTUATION SETS

The modeling of the incoherence detection data finds that the processing of IPD is independent from the processing of ILD. However, it was seen in Figures 6 and 34 that fluctuations in phase and level were highly correlated. This experiment will use sets with minimum fluctuations in phase or level to see if detection of fluctuations in IPD is truly independent from ILD.

6.1.1 Stimuli

Noise-pairs were generated like those in Chapter 2. The center frequency was still 500 Hz. The coherence started at 0.9922, but varied over the range of 0.969–0.998 because the signal was truncated from 1 s to 0.5 s. The same temporal shaping with 30 ms rise-fall time was used. The noise-pairs in this experiment had a 14-Hz (narrow) bandwidth. No other bandwidths were used in this chapter because the modeling results described the incoherence detection data best for the narrow bandwidth.

A collection of 1000 noise-pairs was generated and $s_t[\Psi_{\Delta\Phi}]$ and $s_t[\Psi_{\Delta L}]$ for the noise-pairs were plotted in the upper panel of Figure 86. The difference between this

plot of phase fluctuations vs. level fluctuations and the ones shown in the previous chapters is that preprocessing was included to make this plot. The preprocessing values were chosen from the best model in Chapters 4 and 5, which was model 1, the sum of interaural differences model. The preprocessing included compression and envelope weighting, which were shown to be useful for describing the data. The value of the envelope weighting was $g = 0.15$, near the optimum values of 0.17 and 0.19. Temporal averaging was omitted because of the insensitivity to changes in τ at this bandwidth. The value of the trading parameter, a, was set at 0.5 in agreement with modeling results. The use of $a = 0.5$ does not apply to Figure 86, but does apply to latter figures in this chapter.

In Figure 86, it can be seen that the minimum values of $s_t[\Psi_{\Delta\Phi}]$ and $s_t[\Psi_{\Delta L}]$ are approximately 1.0 for this bandwidth and level of mixing. It was encouraging that the means and standard deviations of the fluctuations in phase and level were approximately equal after the compression was added. As conjectured in Chapter 4, this may be evidence that the saturating exponentials (derived from Yost 1981) used to perform the compression were well chosen. The mean and standard deviation of phase fluctuations were smaller than the mean and standard deviation of the level fluctuations. This is probably due to the inclusion of the envelope weighting, which ignored phase fluctuations at any instant when the envelope fell below 15% RMS-normalized signal value.

After finding the minimum value of the fluctuations for phase and level for the collection of 1000 noise-pairs, two more collections of 1000 noise-pairs were generated; one called the *minimum phase fluctuation collection* and the other called the *minimum level fluctuation collection.* These collections can be seen in the lower panel of Figure 86. For the minimum phase fluctuation collection, it was required that $s_t[\Psi_{\Delta\Phi}] = 1 \pm 0.1$. Likewise, for the minimum level fluctuation collection, it was required that $s_t[\Psi_{\Delta L}] = 1 \pm 0.1$. Since the phase fluctuations are so small in the min-

imum phase fluctuation collection, incoherence detection in these stimuli is expected to rely heavily, if not completely, on the level fluctuations. Similarly, incoherence detection in the stimuli in the minimum level fluctuation collection is expected to rely on phase fluctuations.

Next, a *minimum phase fluctuation set* was picked from the minimum phase fluctuation collection. Twenty noise-pairs were chosen for this set, and they spanned the range of $s_t[\Psi_{\Delta L}]$ in this collection. Noise-pairs in the minimum phase fluctuation set were rank ordered (1–20) with respect to decreasing level fluctuations. The fluctuation values of these noise-pairs can be seen in Table 38. Likewise, a *minimum level fluctuation set* of twenty noise-pairs was picked from the minimum level fluctuation collection. Noise-pairs in the minimum level fluctuation set were rank ordered (1–20) with respect to decreasing phase fluctuations. The fluctuation values of these noise-pairs can be seen in Table 39.

6.1.2 Procedure

The procedure used in this experiment was essentially the same as the one that was used in all the previous experiments. However, there is one notable difference; there were 20 noise-pairs to listen to in each set. Therefore, two subsets of ten noise-pairs were made. The first subset contained the odd-rank-numbered noise-pairs from the set. Therefore, the subset spanned the entire range of fluctuation values (for this bandwidth and level of mixing) because the noise-pairs were rank ordered with respect to decreasing fluctuations. The second subset had the even-rank-numbered noise-pairs. Thus, the second subset had a range of fluctuations comparable to the first subset.

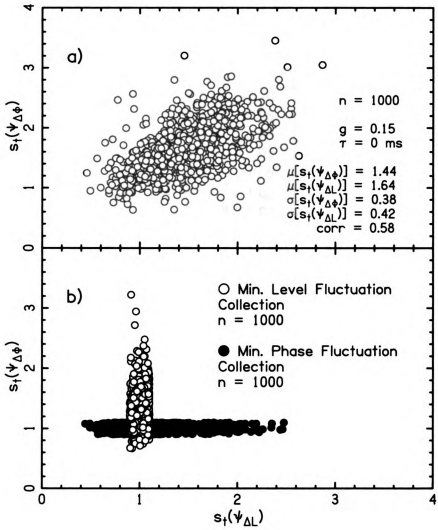

Figure 86: a) The phase-level plot for the first collection of 1000 noise-pairs in Experiment 17, the minimum flucutation experiment. Note that the axes are now in units of the psychological variables $s_t[\Psi_{\Delta\Phi}]$ and $s_t[\Psi_{\Delta L}]$, not the physical variables $s_t[\Delta\Phi]$ and $s_t[\Delta L]$. From this collection, it was determined that the minimum phase and level fluctuations for this bandwidth (14 Hz) and level of mixing (α=0.125) were approximately $s_t[\Psi_{\Delta\Phi}] = 1$ and $s_t[\Psi_{\Delta L}] = 1$. b) The plot for the minimum phase fluctuation collection and minimum level fluctuation collection. Each slice has 1000 noise-pairs. From these collections, the minimum phase fluctuation and minimum level fluctuation sets were chosen. Each set had 20 noise-pairs that spanned the values in fluctuations.

Table 38: The fluctuations of phase and level for the minimum phase fluctuation set. Fluctuations are on the scale of lateral position units (see Chapter 4). Fluctuations in $s_t[\Psi_{\Delta\Phi}]$ were required to be 1 ± 0.1. The noise-pairs were rank ordered with respect to decreasing fluctuations in level.

Serial Number	Rank Number	$s_t[\Psi_{\Delta\Phi}]$	$s_t[\Psi_{\Delta L}]$
19	1	1.10	2.48
895	2	0.99	2.26
576	3	0.98	2.08
981	4	1.02	2.05
459	5	1.06	1.83
489	6	1.02	1.81
321	7	1.03	1.68
885	8	1.09	1.63
125	9	1.06	1.52
475	10	0.97	1.48
17	11	1.08	1.36
211	12	1.08	1.28
558	13	1.04	1.20
819	14	0.94	1.08
814	15	1.02	1.04
583	16	1.09	0.88
276	17	0.93	0.88
152	18	1.05	0.72
435	19	0.96	0.68
744	20	1.09	0.38

6.1.3 Listeners

Listeners E and M from the previous chapters participated in this experiment. A new listener, Listener DA, also participated in this experiment. Listener DA was 22 years old and had normal hearing.

6.1.4 Results

1. Detection Data

Figures 87–89 show the incoherence detection scores for the three listeners in this experiment. The horizontal axis is the sum of interaural fluctuations with equal

Table 39: The fluctuations of phase and level for the minimum level fluctuation set. Fluctuations are on the scale of lateral position units (see Chapter 4). Fluctuations in $s_t[\Psi_{\Delta L}]$ were required to be 1 ± 0.1. The noise-pairs were rank ordered with respect to decreasing fluctuations in phase.

Serial Number	Rank Number	$s_t[\Psi_{\Delta\Phi}]$	$s_t[\Psi_{\Delta L}]$
499	1	3.22	0.91
614	2	2.94	0.96
9	3	2.74	1.06
25	4	2.72	0.95
655	5	2.49	1.04
501	6	2.37	1.04
173	7	2.32	0.97
623	8	2.18	0.91
70	9	2.10	0.98
886	10	1.90	1.06
30	11	1.90	0.96
19	12	1.70	1.10
516	13	1.64	1.01
738	14	1.50	1.04
755	15	1.38	1.02
958	16	1.30	1.04
7	17	1.12	1.05
637	18	1.10	1.06
140	19	0.90	0.98
186	20	0.86	0.95

weight ($a = 0.5$). Because the IPD and ILD are now compressed, it is possible to evaluate the fluctuations of both phase and level with a single variable. The vertical axis shows the *CAS*. Since the use of the *CAS* is well-justified from the results of previous chapters and Appendix 1, the P_c results will be omitted from this chapter.

The data for Listener DA are in Figure 87. It can be seen that Listener DA had few noise-pairs near the ceiling and only one noise-pair below chance. Detection data for the minimum phase fluctuation set show higher scores than the minimum level fluctuation set when comparing noise-pairs with equal fluctuations as measured by $s_t[\Psi_{\Delta\Phi}]$ and $s_t[\Psi_{\Delta L}]$. There is a trend that noise-pairs with more fluctuations have

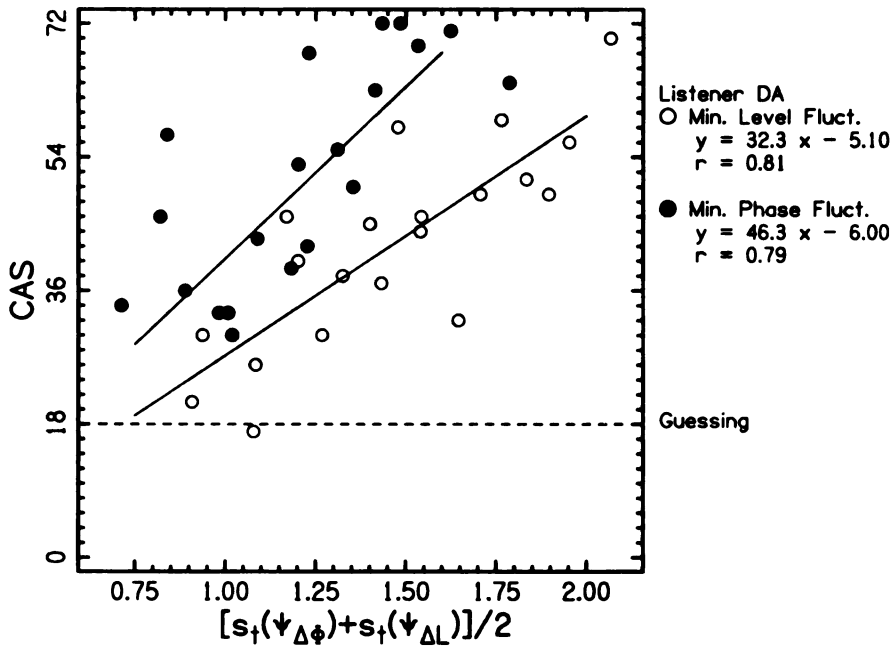

Figure 87: Percent *CAS* score vs. the combined fluctuations of phase and level for Listener DA. The open circles show the minimum level fluctuation set. The closed circles show the minimum phase fluctuation set. The equation of the line of best fit and the linear correlation coefficient are reported.

larger detection scores and that this relationship is fairly linear.

The data for Listener E are in Figure 88. Listener E has no noise-pairs near the *CAS* ceiling and none less than chance. Like Listener DA, Listener E shows higher scores for the minimum phase fluctuation set.

The data for Listener M are in Figure 89. Compared to the other two listeners, Listener M had the most noise-pairs at the ceiling (seven with $CAS = 72$) and the most noise-pairs below chance (four with $CAS < 18$). Unlike Listeners DA and E, Listener M had *CAS* scores that were comparable for the minimum phase fluctuation and minimum level fluctuation sets.

2. Regression Comparison

A linear regression was done on both sets of data for each listener. The equation of best fit and linear correlation coefficient are reported to the right of Figures 87–89.

For Listener DA, the slope for the minimum level fluctuation set was smaller than

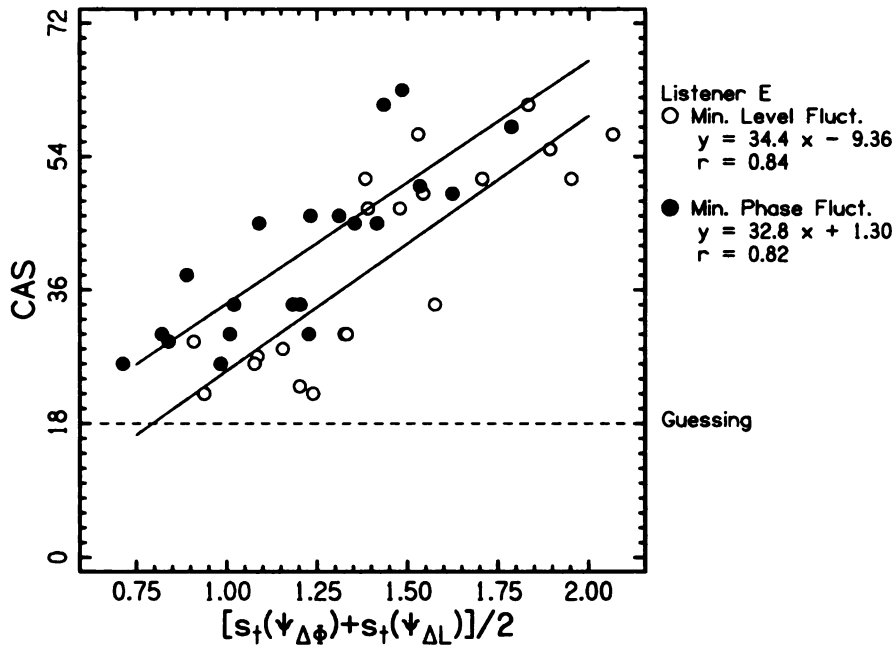

Figure 88: Percent *CAS* score vs. the combined fluctuations of phase and level for Listener E. The open circles show the minimum level fluctuation set. The closed circles show the minimum phase fluctuation set. The equation of the line of best fit and the linear correlation coefficient are reported.

the minimum phase fluctuation set. The linear correlation coefficient of the regression was $r = 0.81$ for the minimum level fluctuation set and $r = 0.79$ for the minimum phase fluctuation set.

For Listener E, the linear regressions of the data had nearly equal slopes for the best fitting lines. Compared to Listener DA, the slope of the line for the minimum level fluctuation set was about equal (slope ≈ 33). However, Listener E had a smaller slope for the minimum phase fluctuation set compared to Listener DA. The lines describe the data points about equally well for each set. The linear correlation coefficient was $r = 0.84$ for the minimum level fluctuation set and $r = 0.82$ for the minimum phase fluctuation set. These coefficients can be compared to the coefficient obtained from the modeling that was done in Chapter 5. For model 1, Listener E had $r = 0.81$ for the 100 noise-pairs with varied coherence.

For Listener M, the slopes of the regression lines were much steeper when com-

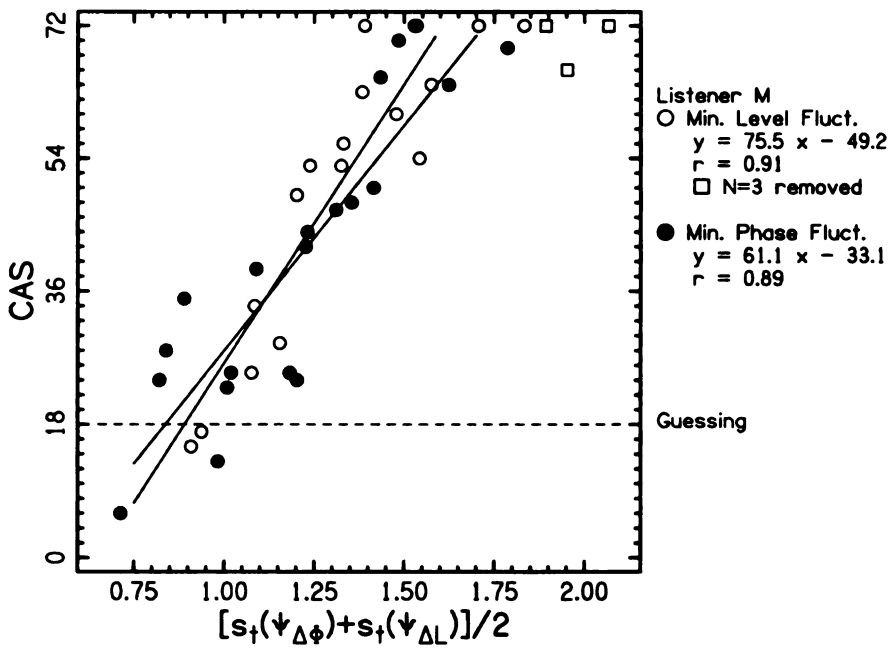

Figure 89: Percent *CAS* score vs. the combined fluctuations of phase and level for Listener M. The open circles show the minimum level fluctuation set. The closed circles show the minimum phase fluctuation set. The equation of the line of best fit and the linear correlation coefficient are reported. Three points were removed from the regression in the minimum phase fluctuation set. These three points are shown with open boxes. The reason that these points were removed was because Listener M had clearly reached the ceiling of *CAS*, unlike the other listeners.

pared to the other listeners; therefore Listener M is more sensitive to incoherence than Listeners DA and E. (Note that three noise-pairs were removed for the linear regression performed on the minimum level fluctuation set because Listener M clearly reached the ceiling for these noise-pairs). The lines also describe the data points about equally well for each set. The linear correlation coefficient was $r = 0.91$ for the minimum level fluctuation set and $r = 0.89$ for the minimum phase fluctuation set. These coefficients can be compared to the coefficient from the modeling that was done in Chapter 5. For model 1, Listener M had $r = 0.81$ for the 100 noise-pairs with varied coherence. Therefore, it the selected noise-pairs of this experiment yield higher correlation coefficients.

Table 40: Values of the linear correlation coefficient, r, for the minimum phase and level fluctuation sets for the *CAS* data. Linear regressions were done on the noise-pairs with preprocessing included and with no preprocessing included. The values of r are larger for four of the six regressions that used preprocessing.

Listener	Min. phase Preprocessing	Min. level Preprocessing	Min. phase No Preprocessing	Min. level No Preprocessing
DA	0.79	0.81	0.71	0.88
E	0.82	0.84	0.69	0.84
M	0.89	0.91	0.85	0.79

Linear regressions were performed by using no preprocessing (no compression or envelope weighting) in the fluctuation calculation and presented in Table 40. Four of the six possible regression comparisons were higher when preprocessing was included. One set (Listener E, Minimum Level) had the same value of r whether preprocessing was included or not. One set (Listener D, Minimum Phase) had a smaller value of r when preprocessing was included.

6.1.5 Discussion

As can be seen from the figures, listeners can use only phase fluctuations or level fluctuations to detect incoherence, as predicted by the models in Chapters 4 and 5. This is because listeners, for the minimum level fluctuation set, had increasing detection scores for increasing fluctuations in phase when the level fluctuations were kept very small ($s_t[\Psi_{\Delta L}] \approx 1$). Likewise, listeners, for the minimum phase fluctuation set, had increasing detection scores for increasing fluctuations in level when the phase fluctuations were kept very small ($s_t[\Psi_{\Delta\Phi}] \approx 1$).

Two of the three listeners were more sensitive to fluctuations in ILD than in IPD when they have the same units of lateral position. This is because the detection scores were higher for the minimum phase fluctuation set, the set that held the fluctuations in phase at a small, constant value but varied the fluctuations in level. This result for dynamic interaural differences agrees with the work by McFadden *et al.* (1971) which showed that some subjects were more sensitive to static ILDs than to static IPDs (and vice versa).

Chapters 4 and 5 were motivated by the result from Chapters 1 and 2 that some noise-pairs were clearly out of order with respect to detectability scores. Noise-pairs in those previous experiments were picked by the two physical variables, $s_t[\Delta\Phi]$ and $s_t[\Delta L]$. By transforming the fluctuation variables with psychologically relevant transformations, the results of Chapters 4 and 5 can be directly applied to pick noise-pairs with fluctuations that correlate better with detectability scores, which can be seen in Table 40.

Lastly, the correlation coefficients were higher in this experiment than in Experiment 15 in Chapter 5, the modeling of the 100 unselected noise-pairs with varied coherence and 14-Hz bandwidth. This might seem like an unexpected result since this experiment used only 20 noise-pairs and Experiment 15 used 100 noise-pairs. The reason for this may be that the noise-pairs in this experiment were selected for

particular properties. This is not the only occurrence of higher correlation coefficients for selected noise-pairs; the values of r were higher for the modeling in Experiment 16 that had 60 selected noise-pairs compared to the modeling in Experiment 14 that had 100 unselected noise-pairs for essentially the bandwidth and distribution of fluctuations in phase and level.

6.2 EXPERIMENT 18: ADDITION AND CANCELLATION SETS

Experiments in Chapters 4 and 5 found that an independent centers model performs better than a lateral position model to describe incoherence detection. This experiment is an additional test of the independent centers versus lateral position model.

6.2.1 Stimuli Selection

Noise-pairs were generated like the previous experiment. The preprocessing functions of compression and envelope weighting were included in the sorting and selecting of noise-pairs. For this collection, it was required that $s_t[\Psi_{\Delta\Phi}] = 1 \pm 0.1$ and $s_t[\Psi_{\Delta L}] = 1 \pm 0.1$. The total number of noise-pairs generated for this collection was 10,000. From the collection, the noise-pairs were sorted with respect to the lateral position variable, $s_t[\Psi_z] = s_t[\Psi_{\Delta\Phi}/2 + \Psi_{\Delta L}/2]$.

Twenty noise-pairs with the largest $s_t[\Psi_z]$ were chosen from the collection of 10,000 as the *addition set.* It was expected that the addition set would have noise-pairs with large IPDs and ILDs in the same lateral direction that coincided temporally. Next, twenty noise-pairs with the smallest $s_t[\Psi_z]$ were chosen from the collection of 10,000 as the *cancellation set.* It was expected that the cancellation set would mostly have noise-pairs with IPDs and ILDs in opposite lateral directions that coincide temporally. An example of a noise-pair from the addition set and a noise-pair from the cancellation set can be seen in Figure 90.

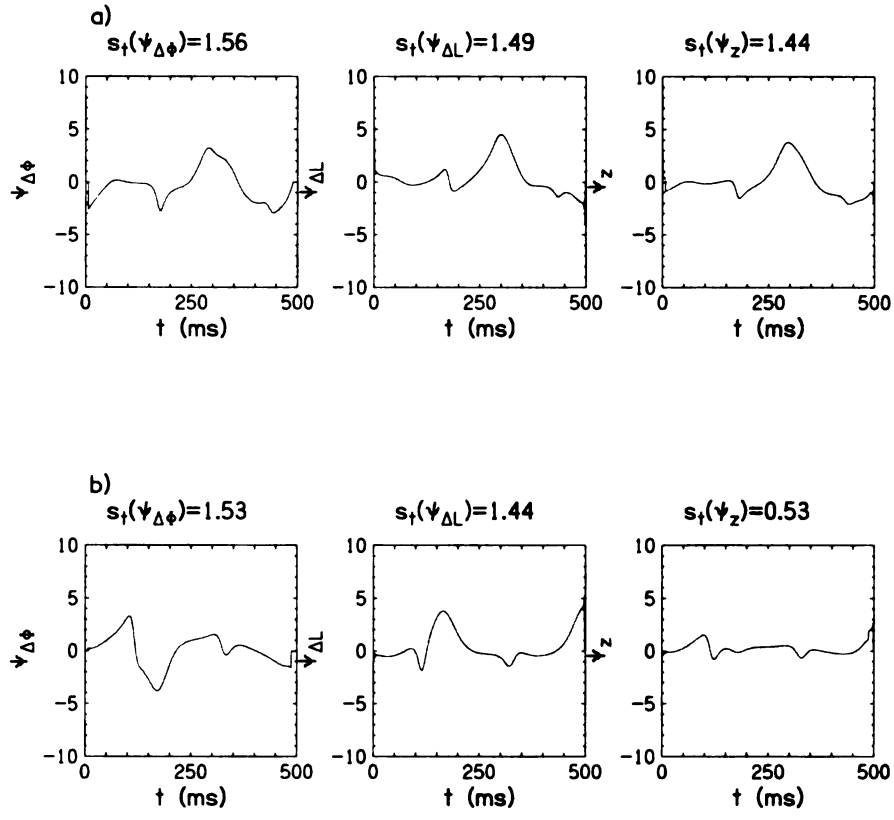

Figure 90: a) The fluctuations in phase, level, and lateral position for a sample noise-pair from the addition set. This noise-pair has a large interaural phase and level at approximately 300 ms. By the lateral position model, these interaural differences add to give a large fluctuation of the lateral position. b) The fluctuations in phase, level, and lateral position for a sample noise-pair from the cancellation set. This noise-pair has a large interaural phase and level at approximately 150 ms, but in opposite directions. By the lateral position model, these interaural differences cancel each other to give a small fluctuation of the lateral position.

Two more collections of 10,000 noise-pairs were similarly generated. One collection had $s_t[\Psi_{\Delta\Phi}] = 1.25 \pm 0.1$ and $s_t[\Psi_{\Delta L}] = 1.25 \pm 0.1$, and the other collection had $s_t[\Psi_{\Delta\Phi}] = 1.5 \pm 0.1$ and $s_t[\Psi_{\Delta L}] = 1.5 \pm 0.1$. An addition set and a cancellation set were similarly picked from each of these two collections. These three values of $(s_t[\Psi_{\Delta\Phi}] + s_t[\Psi_{\Delta L}])/2 = 1$, 1.25, 1.5 were chosen because they avoided the ceiling for Listener M, which was seen to be reached for $(s_t[\Psi_{\Delta\Phi}] + s_t[\Psi_{\Delta L}])/2 \approx 1.75$ in Figure 89.

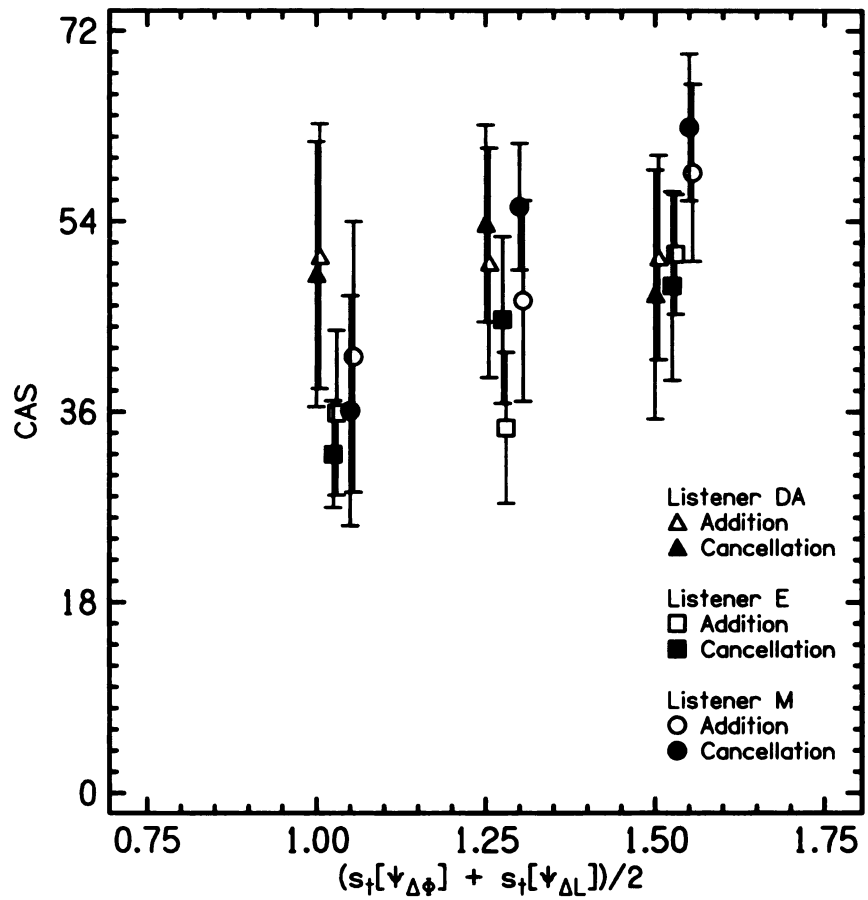

Figure 91: The addition sets (open symbols) and cancellation sets (closed symbols) for Experiment 18. Error bars are two standard deviations in overall length. All three listeners show no significant difference between addition or cancellation sets, supporting the hypothesis that the lateral position model is not used by listeners to detect incoherence.

6.2.2 Results

Figure 5 shows the means and standard deviations of the *CAS* for the twenty noise-pairs in the addition set (open symbols) and cancellation set (closed symbols). Symbols are jogged for each listener around the values of $(s_t[\Psi_{\Delta\Phi}] + s_t[\Psi_{\Delta L}])/2 = 1$, 1.25, and 1.5. All three listeners showed little increase in the mean for each set with respect to the standard deviations of the points. The error bars are large for each data point in this experiment.

There was no significant difference between noise-pairs in the addition set and

cancellation set for three listeners as measured by a two-sample t-test. All the t-tests between the scores for the addition and cancellation sets showed no significant difference at the 0.05 level. In fact, for the sets with fluctuations of about 1.25, it was seen that the cancellation sets performed consistently better than the addition sets.

6.2.3 Discussion and Conclusion

In this experiment, noise-pairs with nearly the same values of $s_t[\Psi_{\Delta\Phi}]$ and $s_t[\Psi_{\Delta L}]$ were generated. They were then sorted with respect to $s_t[\Psi_z]$ to test if there was any difference between noise-pairs that had interaural differences that added or cancelled in the lateral position model. No difference was seen, supporting the hypothesis that the independent centers model is used for incoherence detection.

It might be argued that this experiment is fundamentally flawed because the preprocessing used (model 1) should favor the independent centers model. However, this is not entirely true. All the modeling done in Chapter 4 used the same compression functions for models 1–7. As for temporal averaging, it was omitted from this experiment because it did little to change the modeling results for the 14-Hz bandwidth. Lastly, the envelope weighting has different values of g for these two types of models. The envelope weighting was found to be unnecessary for the best lateral position model in Chapter 4 ($g \approx 0$). However, if a value of $g = 0.15$ is used in the lateral position model, the description of the detection data does not dramatically change ($r = 0.76$ for $g = 0.15$, $r = 0.81$ for $g = 0$). Therefore it appears that preprocessing only slightly favors an independent centers model over a lateral position model.

Experiments in Chapters 4 and 5 showed values of the trading parameter b near 0 or 1 for the lateral position models. This means that listeners use only IPD fluctuations or only ILD fluctuations for incoherence detection. It was argued that this result supports an independent centers model of fluctuation processing and incoherence detection. But, the current experiment gave the lateral position model a second

chance. No part of the preprocessing changed the sign of the IPD or ILD. Therefore, noise-pairs that add $s_t[\Psi_{\Delta\Phi}]$ and $s_t[\Psi_{\Delta L}]$ in the same lateral direction will, in general, always add. Likewise, noise-pairs with cancelling interaural differences will always cancel somewhat. Thus, the addition sets must have noise-pairs with larger values of $s_t[\Psi_z]$ than noise-pairs in the cancellation sets. Hence, it seems that any value of b would provide an adequate test of the lateral position model. Choosing $b = 0.5$ possibly gave the lateral position model the best chance to succeed in this experiment.

Because of the way the stimuli were picked, an additional conclusion can be made from this experiment. Since all of the noise-pairs had nearly the same values of phase and level fluctuations, the standard deviations of the data points are predicted to be small for this experiment. But, in fact, the standard deviations of the data points are quite large. Although some of the size of the error bars can be contributed to listener uncertainty, this may be evidence that the simple signal processing algorithm used in Chapters 4 and 5 still needs to be refined to describe incoherence detection. Or it is possible that a more sophisticated model is necessary; such as one that models the neural firing patterns of the brainstem nuclei.

7 CONCLUSIONS

The work described in this dissertation began with the suspicion that the detection of interaural incoherence is not entirely described by the coherence function. The coherence function is a time-averaged function. However, the perception of incoherence over headphones is dynamic and varies as a function of time.

In order to resolve this discrepancy, experiments in Chapters 1 and 2 were performed. Reproducible noise-pairs were generated with a fixed coherence of 0.9922. It was found that the standard deviation of the variation in the interaural parameters over time varies greatly between individual noise-pairs for narrow bandwidths. However, the standard deviation of the interaural parameters over time does not vary much between individual noise-pairs for wide bandwidths. Sets were constructed by selecting stimuli with the largest and smallest fluctuations in interaural parameters. It was found that listeners detect incoherence by using fluctuations in interaural phase and level. Since the fluctuations are similar for only wide band cases, the use of the coherence function to predict detection seems only appropriate in the case of wide bandwidths.

Experiments with shorter duration stimuli were used in Chapter 3 to test the possibility that a coherence function shorter than the duration of the 500-ms stimuli is used to detect incoherence. It was found that listeners could detect interaural fluctuations for durations as small as 50 ms for a 14-Hz bandwidth. Similar to the results of Chapter 1, the detection scores showed significant differences between the largest fluctuation noise-pairs and the smallest fluctuation noise-pairs, despite the noise-pairs having a fixed value of coherence. Therefore, listeners do not use a short-term coherence function. For the sets with 25-ms noise-pairs, listeners began using a lateralization cue of one or two auditory images.

Chapters 4 and 5 developed a black-box model of incoherence detection by using models originally presented to describe MLD detection data. It was found that the

best model added the standard deviation over time of the IPD and the standard deviation over time of the ILD. These functions of phase and level were added with equal weight after the interaural parameters were compressed to a lateral position scale. Envelope weighting and temporal averaging were also included in the model. This model may support the perception of two lateral images that listeners experienced for the 25-ms noise-pairs in Chapter 3.

Chapter 6 attempted to use the modeling results to confirm two previous results from Chapters 4 and 5: 1) listeners can use just phase fluctuations or just level fluctuations to detect incoherence and 2) listeners track independent IPD and ILD as a function of time and not a single auditory image made of the combination of the two interaural parameters. These results were confirmed.

From these experiments, it seems that incoherence is detected by using the interaural phase difference and interaural level difference as calculated from the analytic signal. The use of these lateralized interaural parameters from the center position in the head produces the sense of width often associated with incoherence detection. Before this dissertation, van de Par and Kohlrausch (1999) found that the coherence function wholly describes incoherence detection. The reason for the discrepancy between experiments is due to the different interpretations of the bandwidth dependence of interaural parameters. Van de Par and Kohlrausch did not find a bandwidth dependence, whereas there was one found in this dissertation. With respect to other experiments in the literature, the reason that the coherence function has been used to describe incoherence detection is possibly due to the stimuli that have been used in previous experiments. Namely, critical bandwidth noises of about 100 Hz (Koehnke *et al.*, 1986; Culling et al., 2001) and wide bandwidth noises (Pollack and Trittipoe, 1959a,b; Boehnke *et al.*, 2002) have been used for experiments where individual features in the interaural parameters are often too fast to recognize.

The implications of the results in this dissertation are important to MLD experi-

ments because it is thought that incoherence detection is used to detect out-of-phase tones in noise (Wilbanks and Whitmore, 1967; Koehnke *et al.*, 1986; Durlach *et al.*, 1986; Bernstein and Trahiotis, 1992; van de Par and Kohlrausch, 1995). Two prevalent models to describe MLD data are the EC model and the correlation model. However, from the results of these incoherence experiments, both of these models should only be applied to stimuli that have wide bandwidths and similar fluctuations. This is because the EC model and the correlation model are energy-based models that have no mechanism to analyze interaural parameters over short duration windows.

Despite all of the progress made in this dissertation to understand incoherence detection, there is more work to be done. For example, it was found that there is room for improvement on the modeling of the detection data, especially in the case of modeling the critical bandwidth data (about 100 Hz wide for a center frequency of 500 Hz). In addition, there were confusions between monaural envelope fluctuations and interaural fluctuations in Experiments 5 and 10. It would be interesting to explore how these confusions relate to false alarm rates in single-interval MLD experiments with reproducible stimuli. In addition, some experiments should be done to explore the effects of changing the center frequency of the noises. For increasing center frequency, it is expected that listeners will not be able to use interaural phase differences. This is because the auditory system cannot hear phase differences for frequencies above about 2 kHz. This type of experiment might give insight into high-frequency MLD experiment results (e.g. Bernstein and Trahiotis, 1992). In addition, some experiments should be done to explore the effects of changing the value of coherence (level of mixing) of the noises. Fluctuations should increase dramatically as the value of the coherence drops to 0. Do the results in this dissertation apply to lower values of coherence? Lastly, a more sophisticated model than the ones presented in Chapters 4 and 5 can be applied to the detection data. There has been plenty of

work done to model firing patterns from the MSO and LSO, the brainstem nuclei that process interaural parameters (Colburn *et al.*, 1990; Han and Colburn, 1993; Brughera *et al.*, 1995; Dasika *et al.*, 2005; Zhou and Colburn, 2005). However, in all of these models, the outputs of the MSO and LSO are never addressed together and never applied to incoherence detection data. Therefore, the data and results in this dissertation might help in modeling brainstem nuclei firing patterns.

APPENDIX A: CONFIDENCE ADJUSTED SCORES

A.1 HISTORY

The history of incorporating rating methods in binary decision tasks, such as yes-no (YN) and two-alternative forced-choice (2AFC) tasks, has been studied briefly and some time ago. The reason to attempt using rating methods in psychophysical tasks is best expressed by Pollack and Decker (1958): "At first examination, the binary-decision procedure appears inefficient. Often listeners report that they can make finer discriminations than the simple yes-no judgment of message acceptance or message rejection." Another reason to use rating methods is that the experimenters might need to extend the usable range of data to avoid ceiling and floor effects. This would be especially important for experiments with frozen stimuli.

There are at least two distinctly different applications of the theory of signal detection to rating methods. The first is using a "Type 2" psychophysical task in which the listener rates the correctness of a previous response. The other is using a rating of confidence in the listener's ability to perform the given task. Although the two rating methods sound similar (and are often confused for one another), they are distinctly different in the signal detection analysis.

A.1.1 Type 2 Tasks

An example of a Type 1 psychophysical task is the detection of a signal in noise by either a YN or 2AFC task. The difference between a Type 1 and Type 2 task, by definition, is that a Type 2 task rates the correctness of the previous Type 1 task. This rating must happen after the Type 1 decision; the two tasks cannot happen simultaneously. Type 2 psychophysical tasks were studied for a short time during the experiments by Egan and Clark (1956), Egan *et al.* (1956), and Pollack and Decker

(1958). These were source-receiver experiments in which the task was to rate the correctness of the receiver's response in identifying a word presented over a noisy channel. It was quickly realized that this task was fundamentally different from the normal detection of a signal in noise.

Although Type 2 tasks have been used in memory studies, calibration studies, and perception without awareness studies, a recent article by Glavin *et al.* (2003) has shown Type 2 tasks to be potentially very difficult to use in a controlled way. The major problem occurs when the probability density function (pdf) of the signal is transformed for the Type 2 task. There is a chance, in fact it is shown that it is often preferable, that the pdf of the Type 1 task is transformed to perform the Type 2 task. Because the transformation of the pdf might be unpredictable, it would seem wise to avoid using a Type 2 psychophysical task unless necessary. This problem was only realized in hindsight by Pollack and addressed in a letter (Pollack, 1959); this is probably the reason why the use of Type 2 tasks has been limited for the past 50 years.

The question is whether the methods used in this dissertation qualify as a Type 2 task. Glavin *et al.* (2003) says, "Another possible confusion could arise if an observer were asked to give a binary choice between a signal and noise and to follow this with a rating of confidence that the trial was the signal trial. The fact that the second decision is a rating and follows a binary Type 1 decision does not make it a Type 2 decision. If the second decision is a rating of confidence in the signal event, rather than in the correctness of the first decision, then it is a Type 1 rating, no matter when it occurs." It is apparent that using a rating system and using the confidence adjusted scores (CAS) in this study is not akin to performing a Type 2 task.

A.1.2 Rating Methods

Confidence adjusted scores are clearly not equivalent to Type 2 tasks. Therefore, the use of theory of signal detection must be applied to the rating methods and *CAS*. It is possible to find other derivations in the literature on confidence ratings and multi-point decision scales. Egan *et al.* (1959) showed that in a multi-point (6 points), ordinal decision scale in a YN task was consistent with the theory of signal detection. This was followed up by Schulmann and Mitchell (1965) who showed that a multi-point rating task could be also be used for a 2AFC task.

Since the task used in these experiments were 2AFC with a confidence rating, it is consistent with the theory of signal detection. However, in the analysis of the data for these experiments, the point system used (1 for correct, 2 for correct and sure) is not necessarily supported by any previous study that is known. Nevertheless, because of the nature of the experiment, namely the use of frozen-noise pairs, it may be possible to evaluate the acceptability and effectiveness of using the adopted point system.

A.2 CAS VS. Pc SCORES

As was stated in the data collection section of Chapter 1, a four-button box was used in data collection so that the listeners could respond to the correct interval with a confidence estimate. The buttons from left to right were 2!, 2, 3, and 3!, representing confident second interval, second interval, third interval, and confident third interval respectively. Listeners were instructed to use a confident response only if there was no uncertainty as to which interval was incoherent. The *CAS* is defined as the number of times the listener responded correctly plus the number of times that the listener was confident about the correct response. Since an individual noise-pair was heard 36 times, it was possible for a listener to get a score of 72 if the listener was able to respond correctly and confidently for all 36 presentations.

In order to test the validity of the *CAS*, it may be instructive to first observe when

listeners use confident responses. The scores used were from the entire collection experiments of Chapters 4 and 5, not the phase and level sets of Chapters 1 and 2. Figure 92 shows values of CAS vs. P_c for the 100 individual noise-pairs from 14-Hz bandwidth collection of Chapter 1. Likewise, Figure 93 shows the 100 noise-pairs for 14-Hz collection of Chapter 2, Figure 94 shows the 100 noise-pairs for 108-Hz collection of Chapter 1, and Figure 95 shows 60 noise-pairs for 136-Hz collections of Chapter 2.

Figure 92 shows that Listeners M had 39 noise-pairs with scores at the the ceiling ($P_c = 100\%$). Listener W had 46 noise-pairs with $P_c = 100\%$. Listener D also reached the ceiling for 21 noise-pairs. Listeners D and M tend to start responding confidently at $P_c = 75\%$. Listener W uses confident responses more sparingly, using them when $P_c > 80\%$. Similar results are seen in Figure 93.

For the 108-Hz bandwidth, Listeners D and M show scores near the ceiling in Figure 94. Listener W does not have many noise-pairs scores near the ceiling, however, he is still using the confident responses for the highest values of P_c, and thus increasing the dynamic range of the experiment.

For the 136-Hz bandwidth, Figure 95 shows that the CAS vs. P_c values of DY, E, T, and W are like those of W in Figure 94. The only two listeners that were common between the 108-Hz collection and the 136-Hz collection were Listeners M and W. Listeners M and W show that the CAS vs. P_c results are alike between the two bandwidths in number of noise-pairs near $P_c = 100\%$ and when confident responses were used. This result is expected because the difference between these two bandwidths is small; they are both approximately a critical bandwidth for a center frequency of 500 Hz.

A comparison across bandwidth shows that the 14-Hz bandwidth task has many more data points near the ceiling of $P_c = 100\%$ than for the 108-Hz or 136-Hz bandwidth task. This can be seen when comparing values of P_c in Figures 92 and 93 with

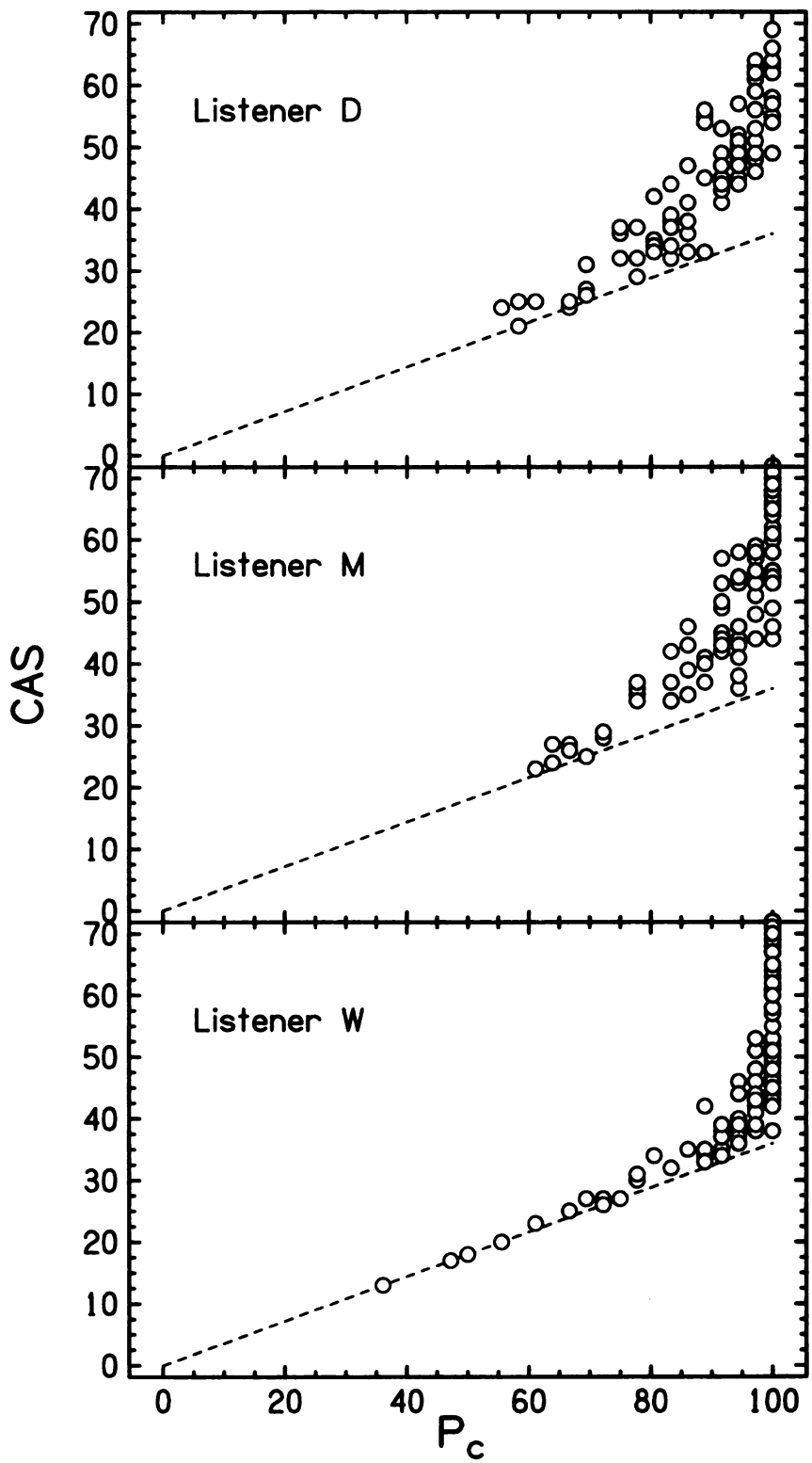

Figure 92: The values of *CAS* vs. P_c for the 100 noise-pairs of the 14-Hz bandwidth in Experiment 1 of Chapter 1.

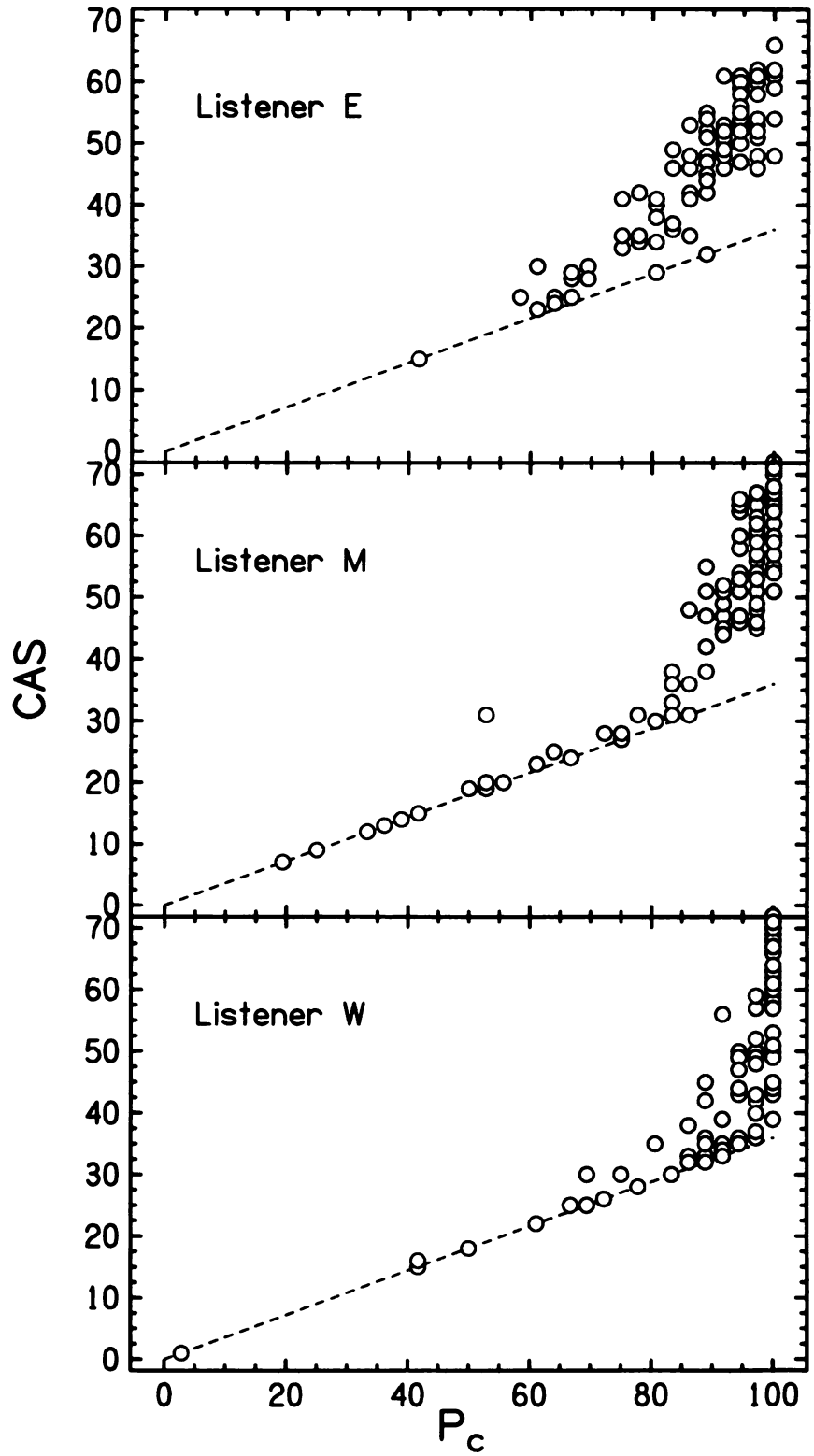

Figure 93: The values of CAS vs. P_c for the 100 noise-pairs of the 14-Hz bandwidth in Experiment 6 of Chapter 2.

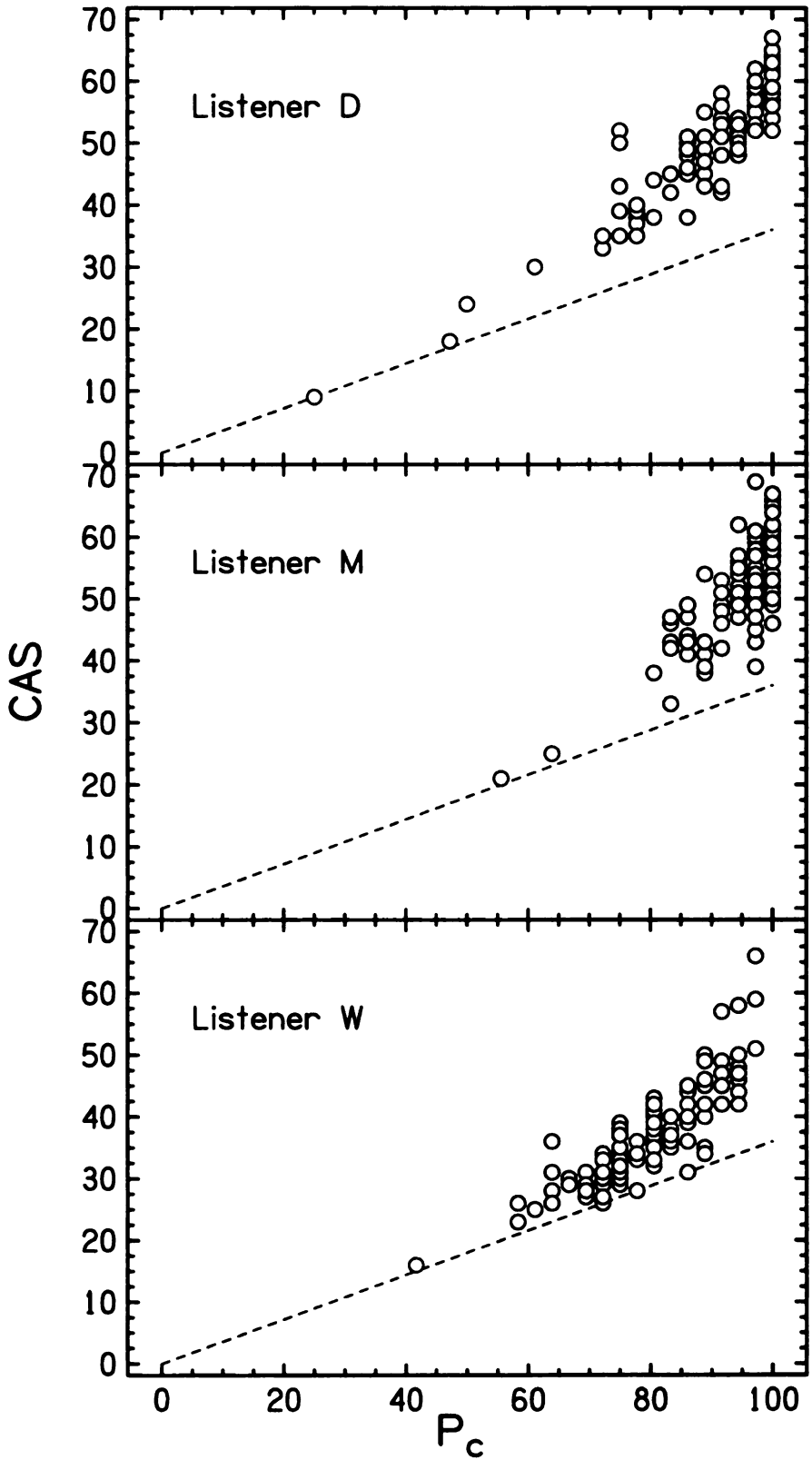

Figure 94: The values of *CAS* vs. P_c for the 100 noise-pairs of the 108-Hz bandwidth in Experiment 2 of Chapter 1.

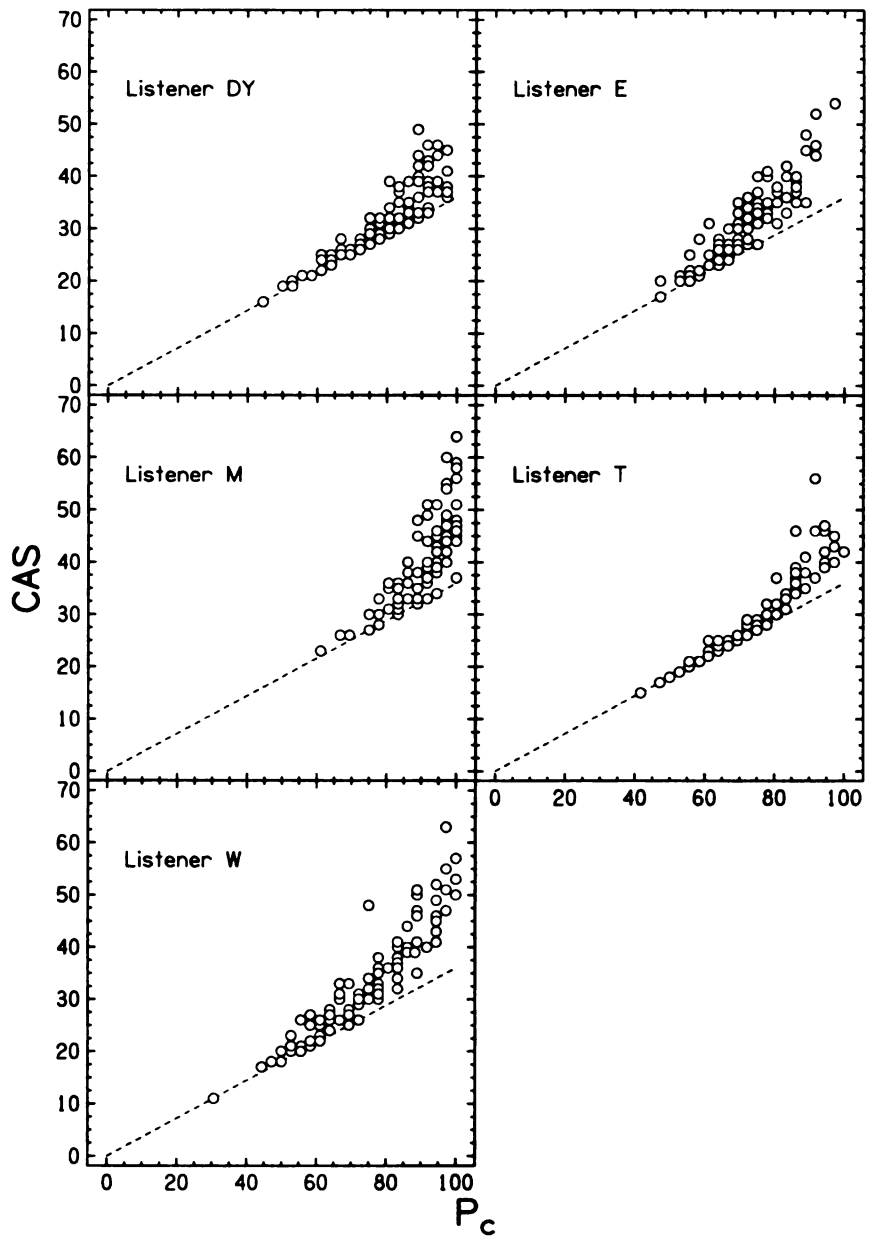

Figure 95: The values of *CAS* vs. P_c for the 60 noise-pairs of the 136-Hz bandwidth Experiment 7 in Chapter 2.

Figures 94 and 95 for common listeners. Therefore, it is easier to detect the incoherence in the 14-Hz bandwidth noise-pairs compared to the wider bandwidth noise-pairs for the same value of coherence. One of the reasons the CAS was introduced was to remove the effects of the ceiling of P_c, especially for the 14-Hz bandwidth where incoherence detection is easiest.

A.3 WEIGHTING

The weight given to the number of confident responses (N_{conf}) compared to number of correct responses will now be varied to study the dependency of CAS results to different weights. The rest of the figures in this appendix show p-values, inter-listener correlations, and model correlations plotted against the weight of the confident responses in the equation

$$CAS = N_{\text{corr}} + \text{weight} \times N_{\text{conf}} \tag{41}$$

where N_{corr} is the number of correct responses and the weight is on a pseudo-log scale. Note that N_{corr} is equivalent to P_c. The weights used were 0 (corresponding to P_c results), 0.25, 0.5, 1 (the weight used throughout the dissertation for the CAS), 2, 4, and ∞.

A.3.1 P-values

Since the same noise-pairs were presented to listeners, it is expected that the p-values should be smallest if all the listeners adopt the same strategy for incoherence detection. The p-values of a one-tailed t-test were found for P_c and CAS values in phase and level sets in Chapters 1 and 2.

The results of varying the weight for the phase and level set t-tests are seen in Figures 96–99. Figure 96 shows the p-values for the phase and level sets for the four

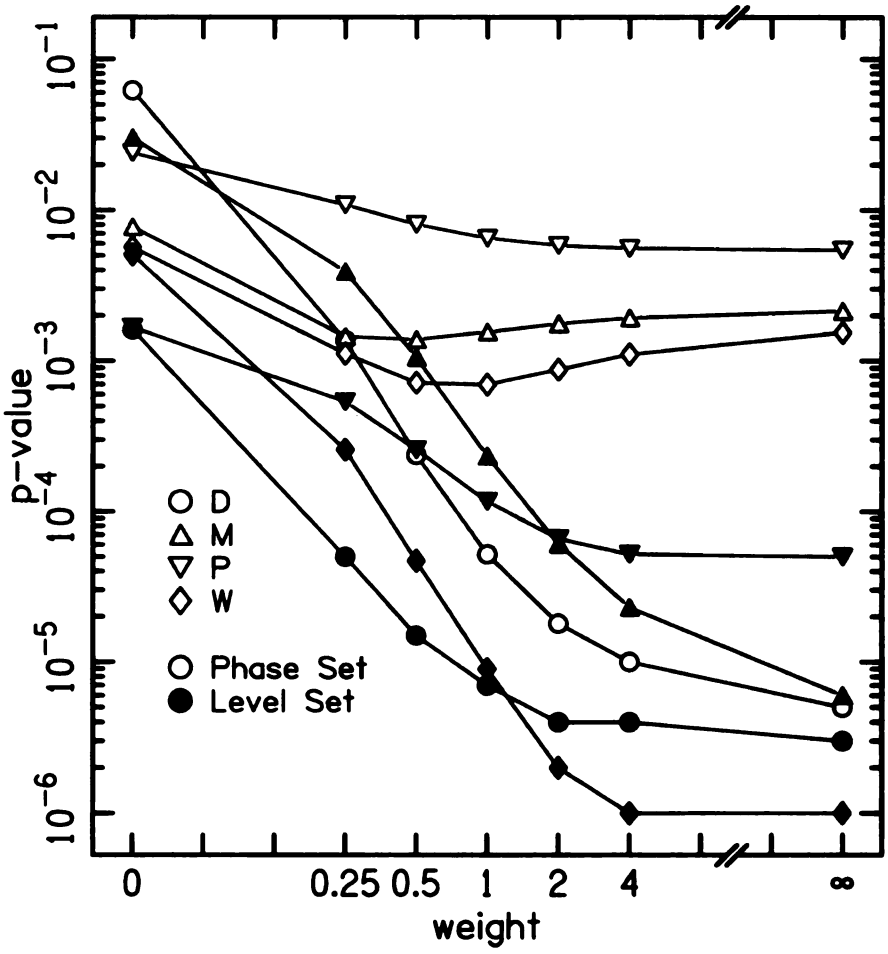

Figure 96: The p-values as a function of weight for an one-tailed t-test. The p-values are from the phase and level sets of the 14-Hz bandwidth noise-pairs from Experiment 1 in Chapter 1.

listeners of Experiment 1, the 14-Hz bandwidth. The weighting curves are relatively flat for four of the curves in the vicinity of a weight equal to 1; the other four curves continue to decrease near 1. In Figure 97, four curves are flat near 1, four curves decrease, and two curves increase. Figure 98 shows all the curves relatively flat and near-minimum around a weight of 1. Figure 99 shows mostly flat curves except one that continues to decrease (Listener D phase set) and one curve that has a sharp minimum at 1 (Listener M level set).

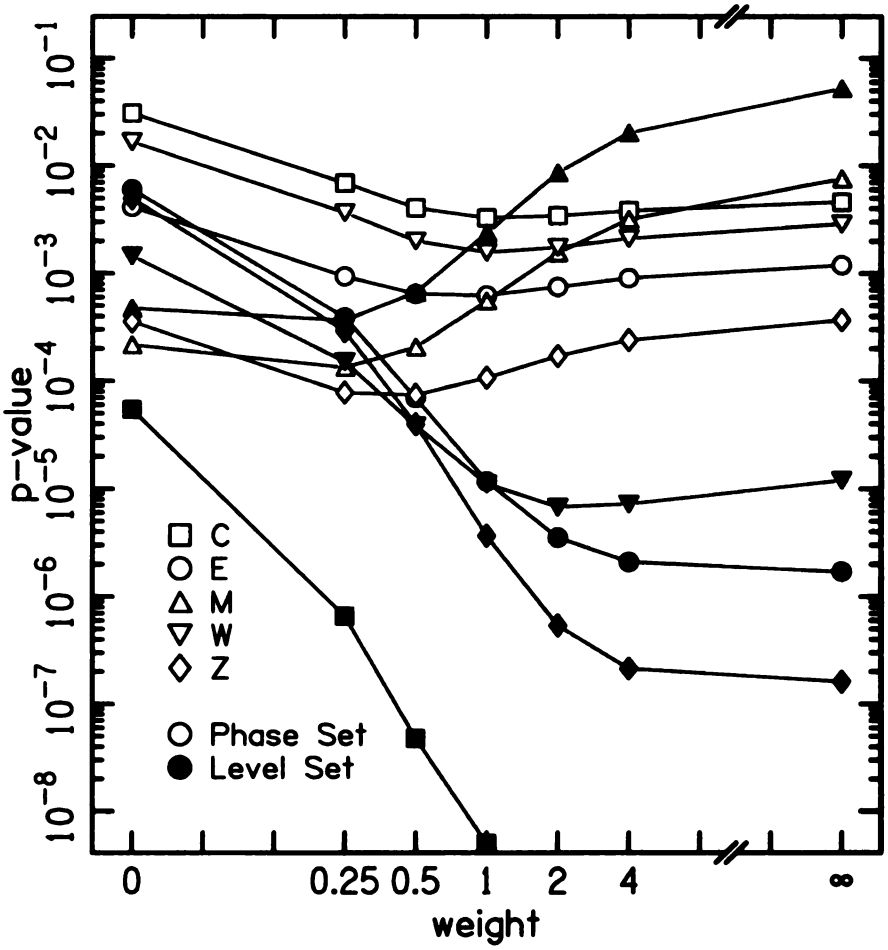

Figure 97: The p-values as a function of weight for an one-tailed t-test. The p-values are from the phase and level sets of the 14-Hz bandwidth noise-pairs from Experiment 6 in Chapter 2.

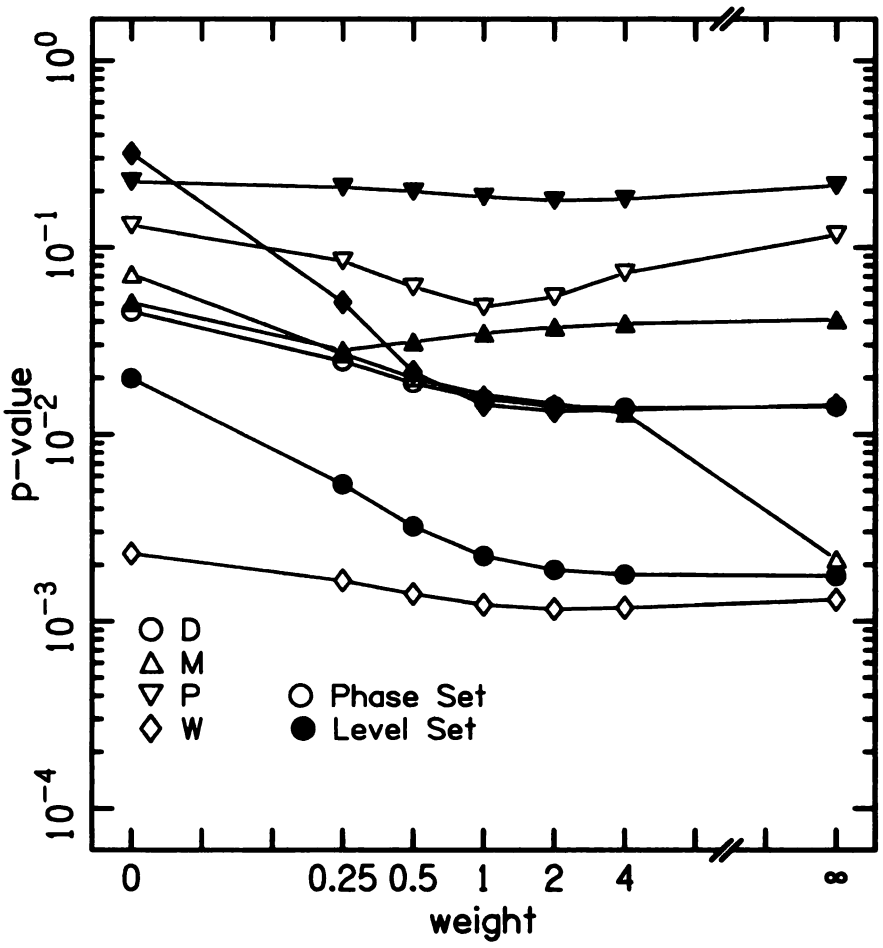

Figure 98: The p-values as a function of weight for an one-tailed t-test. The p-values are from the phase and level sets of the 108-Hz bandwidth noise-pairs from Experiment 2 in Chapter 1.

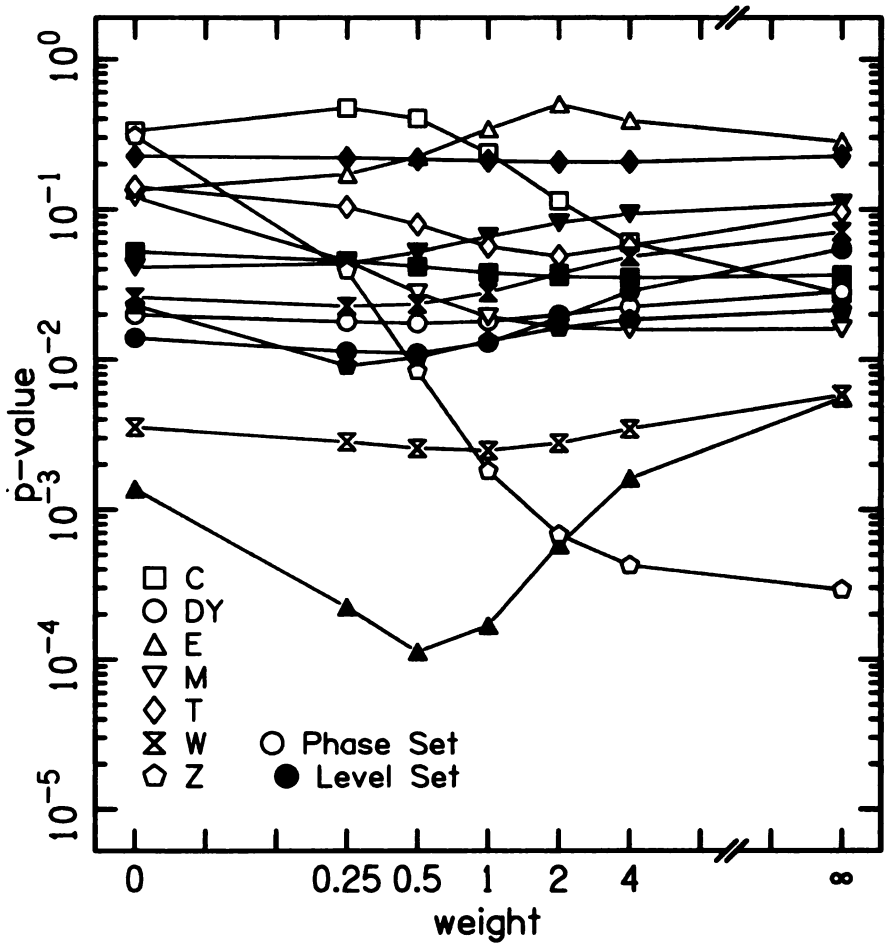

Figure 99: The p-values as a function of weight for an one-tailed t-test. The p-values are from the phase and level sets of the 136-Hz bandwidth noise-pairs from Experiment 7 in Chapter 2.

A.3.2 Inter-listener Correlation

Next, average inter-listener correlation (over all the listeners in the experiment) was plotted as a function of weight.

1. Phase and Level Sets

Figures 100 and 101 shows average inter-listener correlation for the two 14-Hz experiments. Figure 100 shows both the phase and level sets monotonically increase with increasing weight. However, the maximum is almost reached in both cases for a weight equal to 1. This means that the use of N_{corr} is not very consistent over observers relative to N_{conf}. This is unlike Figure 101 which has a maximum at 0.5 for both the phase and level sets average inter-listener correlation, although a weight of 1 is almost as good. Note that the correlation is relatively flat for values near 1. Figure 102 shows a maximum at 2 for the phase set and a general increase for increasing weight for the level set. Figure 103 shows maxima at 1 for both the phase and level sets.

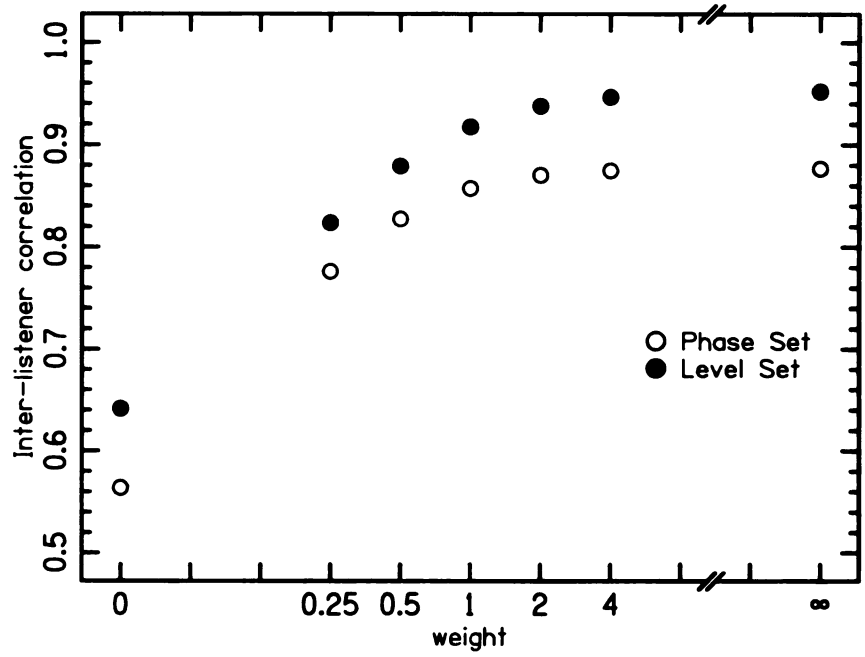

Figure 100: The average inter-listener correlation for the phase set and level set in Experiment 1.

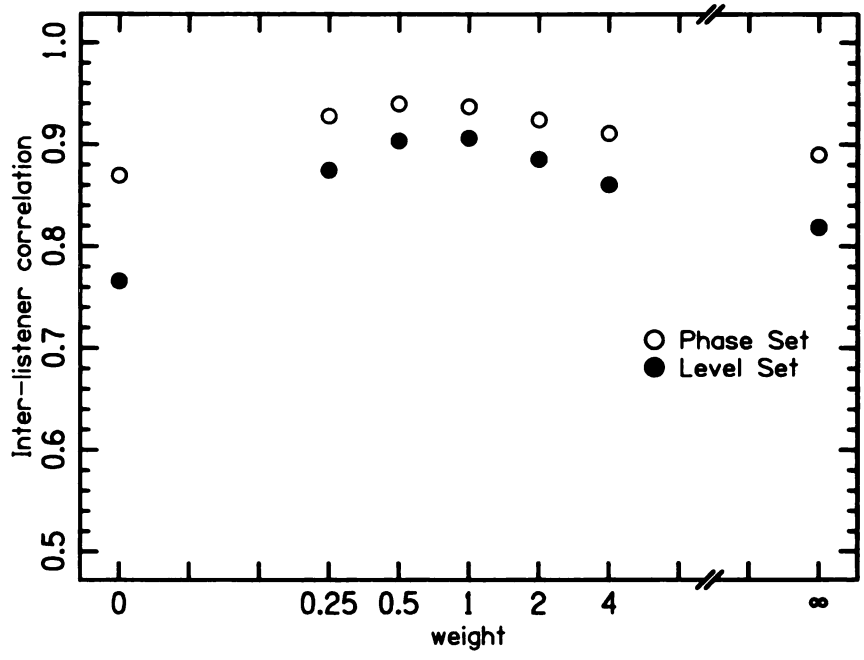

Figure 101: The average inter-listener correlation for the phase set and level set in Experiment 6.

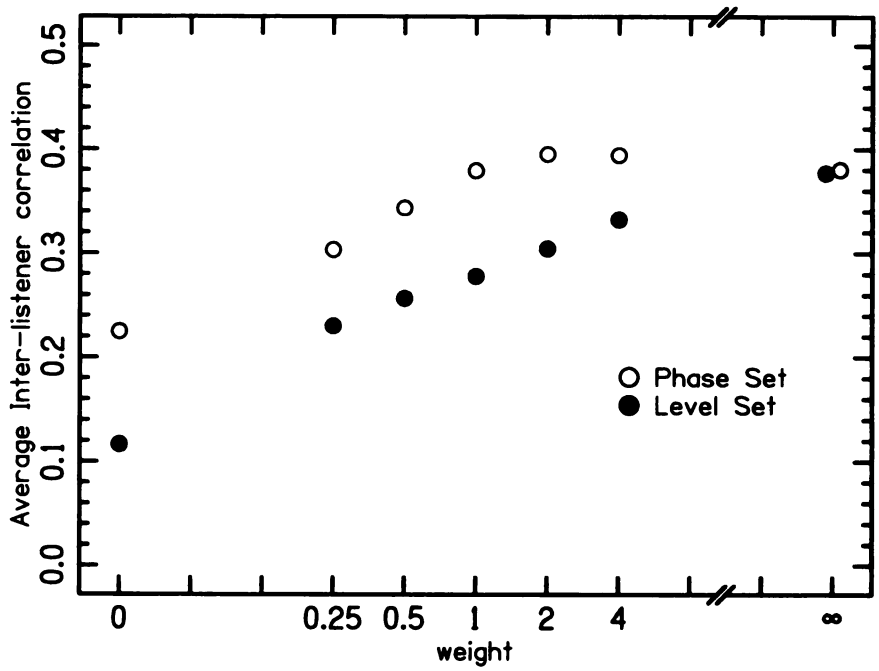

Figure 102: The average inter-listener correlation for the phase set and level set in Experiment 2. Note that the vertical scale is the same size as Figure 100, just translated downward.

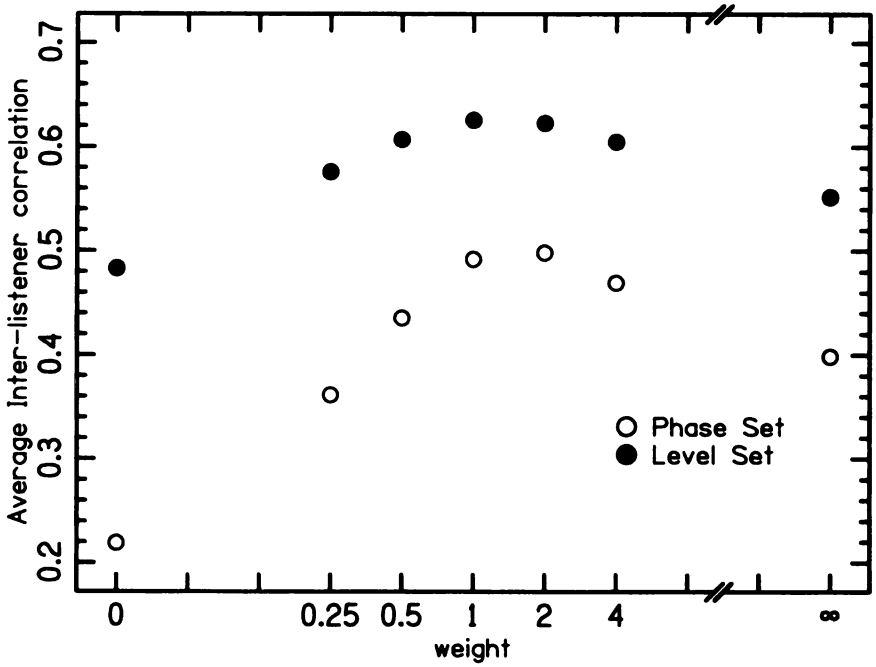

Figure 103: The average inter-listener correlation for the phase set and level set in Experiment 7. Note that the vertical scale is the same size as Figure 100, just translated downward.

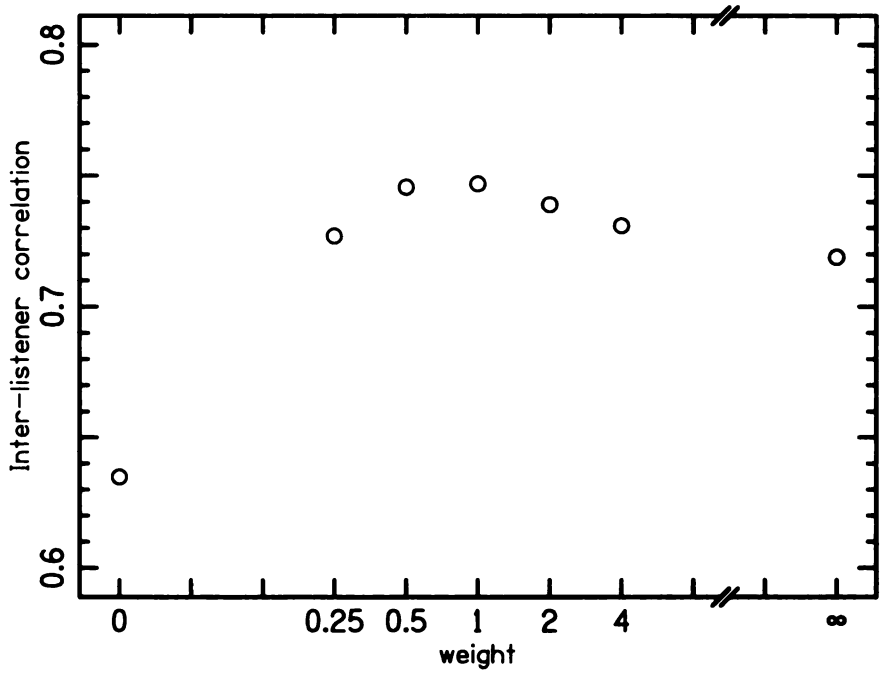

Figure 104: The weight of the confident responses is varied to show the effect on the average inter-listener correlation for the 14-Hz bandwidth data in Experiment 1. The function shows a maximum near a value of 1, with relative insensitivity in this region. The minima of the function are at the extreme weights of 0 (using only P_c) and ∞ (using only confident).

2. Entire Collection

Now, the average inter-listener correlation was calculated as a function of weight for the entire collection of 100 noise-pairs that were used in Experiments 1 (14 Hz), 2 (108 Hz), and 6 (14 Hz). The 136-Hz collections in Experiment 7 were also used, but only 60 of 200 noise-pairs were included.

Figures 104–107 show the results. All four figures show maximum inter-listener correlation for weights of 0.5 and 1. For the cases where the maximum is 0.5, the average inter-listener correlation at 1 is near the maximum value.

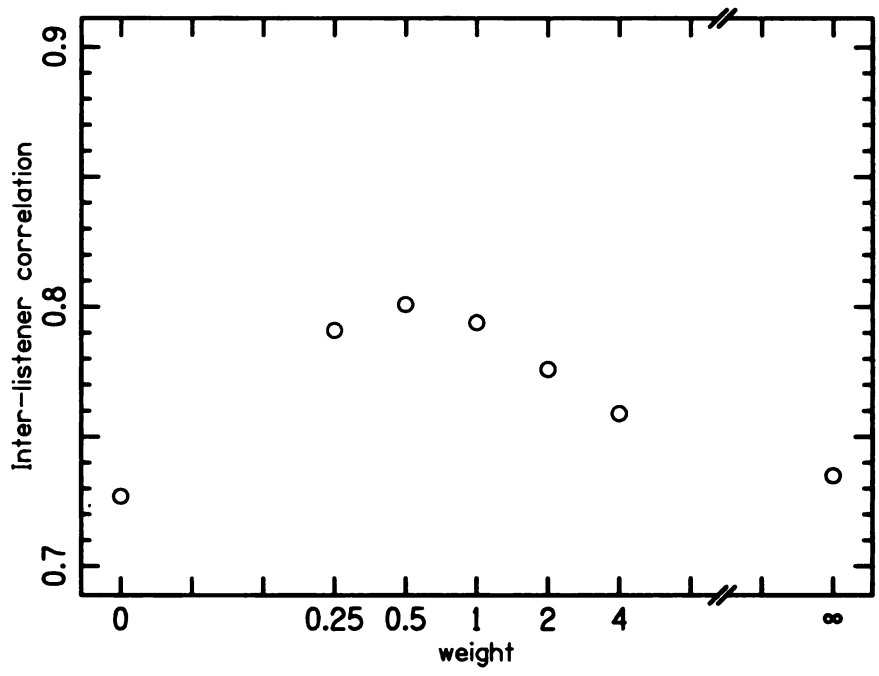

Figure 105: The weight of the confident responses is varied to show the effect on the average inter-listener correlation for the 14-Hz bandwidth data in Experiment 6. The function shows a maximum near a value of 1, with relative insensitivity in this region. The minima of the function are at the extreme weights of 0 (using only P_c) and ∞ (using only confident response). Note that the vertical scale of this figure is the same size, but translated upward when compared to Figure 104.

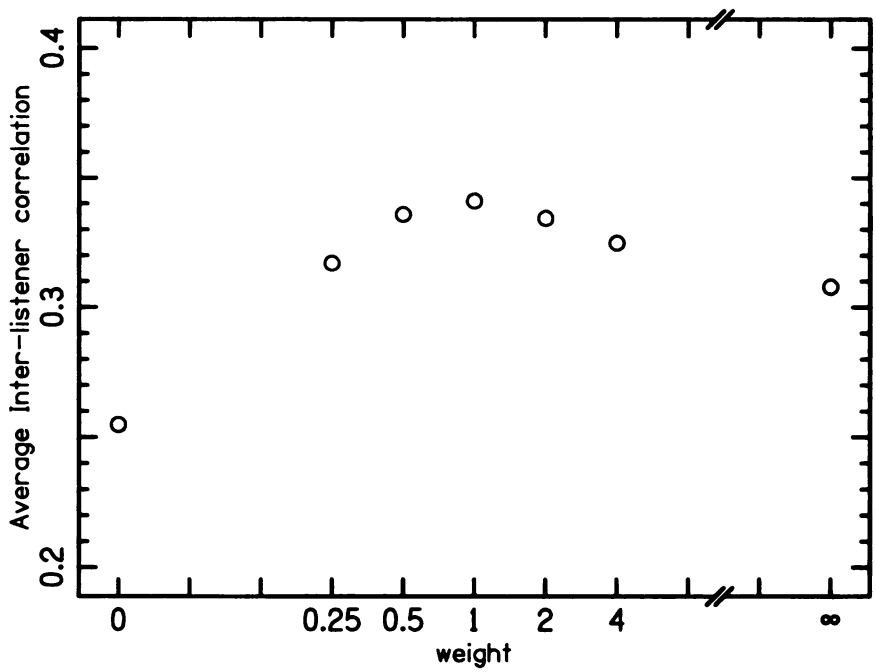

Figure 106: The weight of the confident responses is varied to show the effect on the average inter-listener correlation for the 108-Hz bandwidth data in Experiment 2. The minima of the function are at the extreme weights of 0 (using only P_c) and ∞ (using only confident). Note that the vertical scale of this figure is the same size, but translated downward when compared to Figure 104.

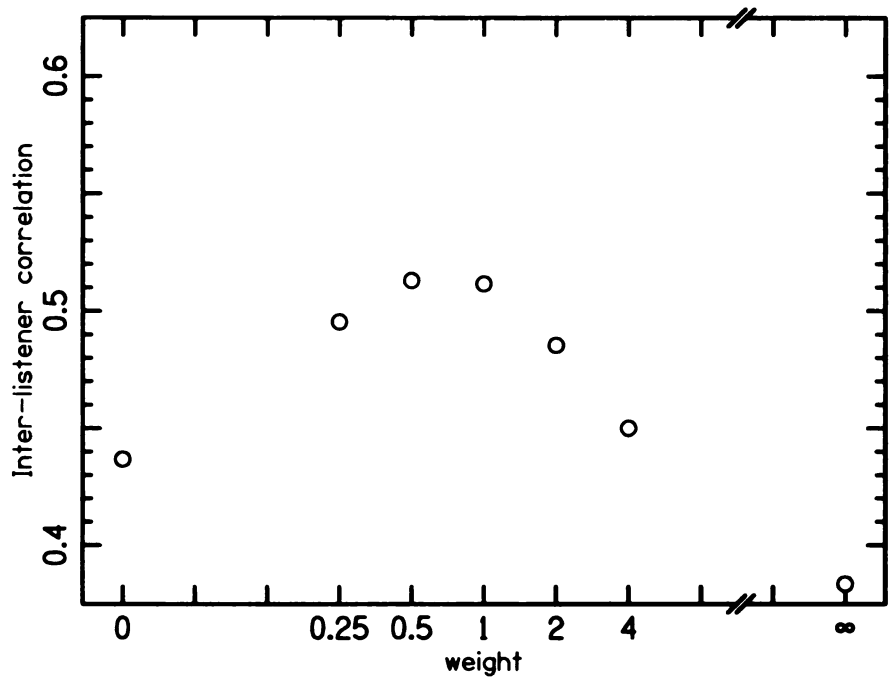

Figure 107: The weight of the confident responses is varied to show the effect on the average inter-listener correlation for the 136-Hz bandwidth data in Experiment 7. The function shows a maximum near a value of 1, with relative insensitivity in this region. The minima of the function are at the extreme weights of 0 (using only P_c) and ∞ (using only confident). Note that the vertical scale of this figure is the same size, but translated downward when compared to Figure 104.

A.3.3 Modeling Results

The last test to evaluate the validity of the *CAS* was performed by remodeling of the detection data from Chapters 4 and 5. The difference between the previous modeling of the *CAS* data and this analysis is the weight of N_{conf} was varied. The weight of the N_{conf} was varied for model 1, the best model for the 14-Hz data and nearly the best model for the 108-Hz and 136-Hz data. Except for varying the weight of the confident response, the modeling was performed exactly the same as in Chapter 4, including all the preprocessing and free parameters.

For the 14-Hz data with fixed coherence, Figure 108 shows r_{max} for the individual listeners and for the average of the listener data. It shows that confident responses are very important to include when modeling the data. The weighting curves increase for increasing weight and are near the maximum for a weight of 1 for all the listeners and averaged data.

For the 14-Hz data with varied coherence, the results are seen in Figure 109. For Listener M and the averaged data, the value of r_{max} increases for weights up to 2 and remains relatively constant for weights of 4 and infinity. Listeners E and W show a maximum at weights of 0.5 and 2 respectively, although, r_{max} decreases very gradually for higher weights. The importance of using confident responses at this bandwidth is not surprising as much of the data was near the ceiling for P_c, as shown in Figure 92.

For the 108-Hz data, the results can be seen in Figure 110. Listener M and the averaged data show a maximum at a weight of 1. Listener D shows a maximum at 0.5. Listener W shows the maximum r_{max} at a weight of ∞ that is not much larger than the r_{max} at a weight of 1. All the curves are relatively flat in the vicinity of a weight equal to 1.

For the 136-Hz data, the results can be seen in Figure 111. Three listeners, D, E, and T show the peak of r_{max} at a weight near 1. Listener M shows a monotonic

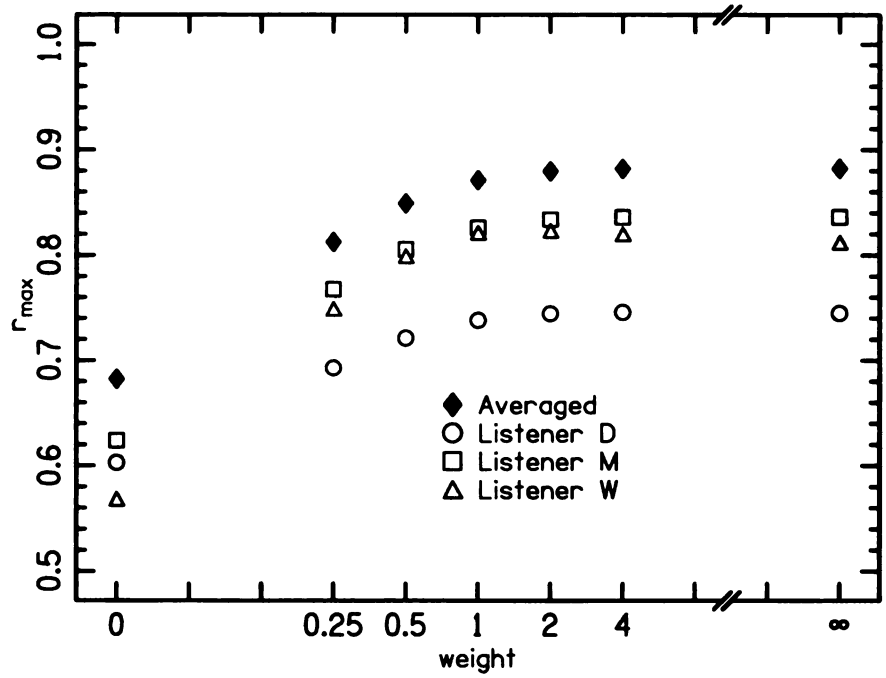

Figure 108: The weight of the confident responses is varied to show the effect on the modeling predictions for model 1 for the 14-Hz bandwidth data with fixed coherence.

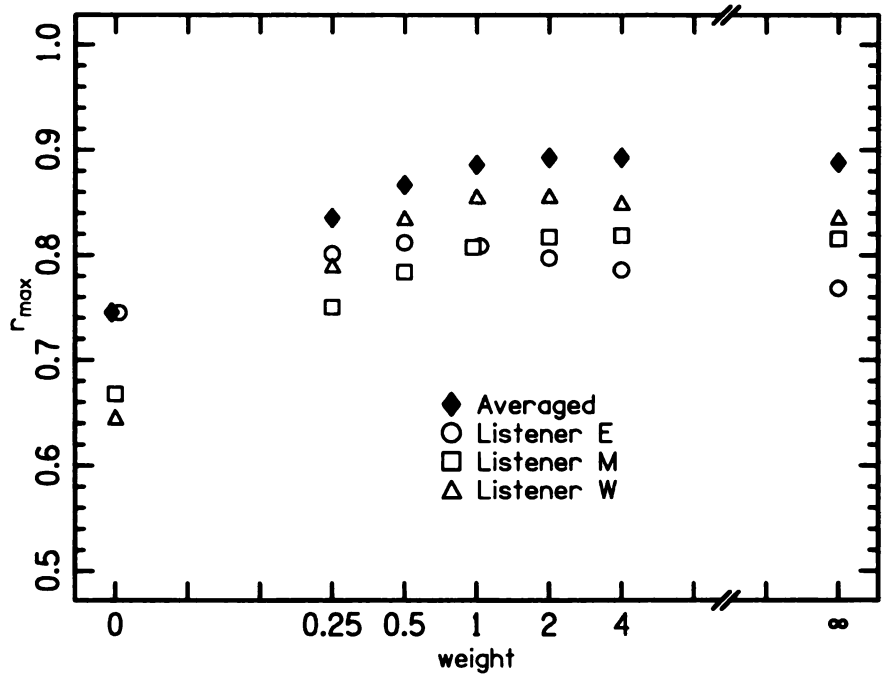

Figure 109: The weight of the confident responses is varied to show the effect on the modeling predictions for model 1 for the 14-Hz bandwidth data with varied coherence.

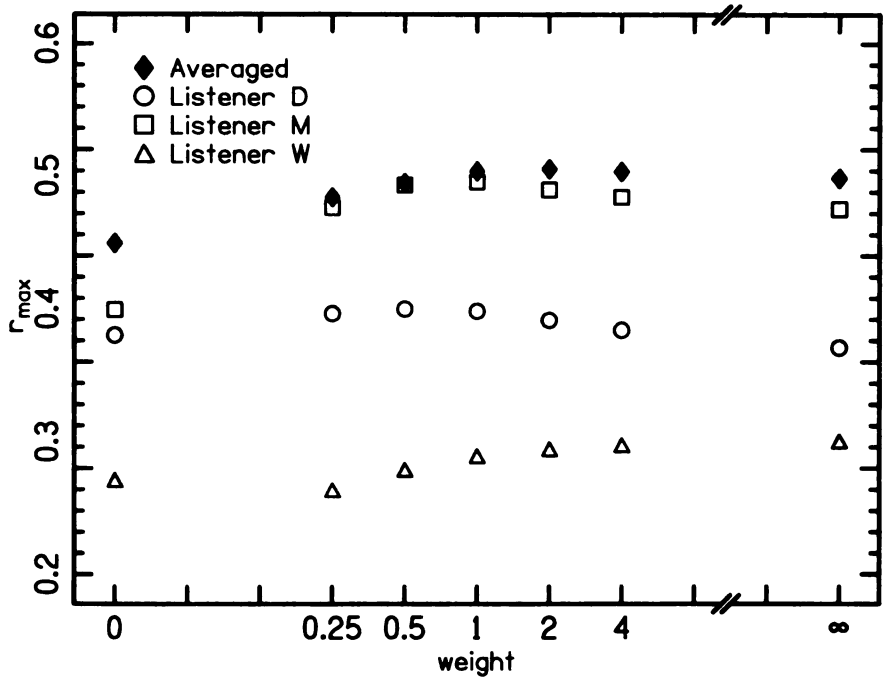

Figure 110: The weight of the confident responses is varied to show the effect on the modeling predictions for model 1 for the 108-Hz bandwidth data with fixed coherence.

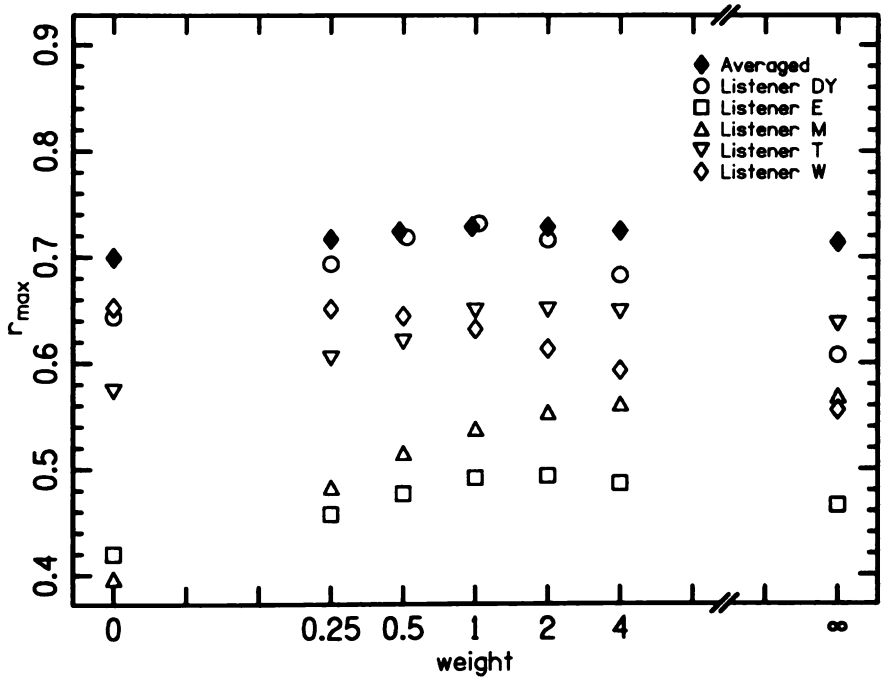

Figure 111: The weight of the confident responses is varied to show the effect on the modeling predictions for model 1 for the 136-Hz bandwidth data with varied coherence.

increase for increasing weight. This may be expected for Listener M because he was often at the ceiling for P_c is this experiment as can be seen in Figure 95. Listener W was the opposite of M, showing a monotonic decrease for increasing weight. This shows that the decision strategy of Listener W was probably non-optimal or that he was inconsistent when using the confident responses. Although it is not reported anywhere in this dissertation, Listener W had the most incorrect-confident responses of all the listeners and had the largest number of runs omitted due to these mistakes over all the cases. Nevertheless, even with Listener W's decrease in $r_{\max}$ for increased weight of N_{conf}, the average 136-Hz data shows that the general population has a maximum in the $r_{\max}$ function at a weight of 1.

A.4 DISCUSSION AND CONCLUSION

The results of this analysis are encouraging for several reasons. Of particular merit are the weighting curves that are mostly concave downward for all of the inter-listener correlations and modeling correlations. The weighting curves often showed maxima near a weight of 1, which was used as the weight for the *CAS* in this dissertation. If the maxima for the weighting curve was not at 1, it was often the case that a weight of 1 was near the maximum (or minimum, in the case of the t-tests) value and the curve had a near-zero slope in this vicinity, indicating an insensitivity to the exact weight chosen for the confident responses.

In conclusion, the validity of using confidence adjusted scores was tested in this appendix. Given that the use of the *CAS* is not a Type 2 psychophysical task (even though this task may be mistaken for one), but rather a multi-point decision scale that is supported by the theory of signal detection, from a historical standpoint, it appears using the *CAS* is a legitimate psychophysical method. Experiments varying the weight of the confident responses show that a weight of 1 is a near optimum choice for the weight for this task. Therefore, the psychophysical methods used in

this dissertation seem valid and justified.

APPENDIX B: FLUCTUATIONS AS A FUNCTION OF BANDWIDTH

The mean values and standard deviations of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ are given in Figures 6, 15, and 21 for a collection of 100 noise-pairs and bandwidths of 14 Hz, 108 Hz, and 2394 Hz respectively. It can be seen that as bandwidth increases, the standard deviation decreases. The purpose of this appendix is to provide more precise values of the means and standard deviations and to fit a functional form to the variation of the standard deviation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ as a function of bandwidth.

For each value of bandwidth, 5000 noise-pairs were generated with a fixed value of coherence 0.9922 as described in Experiments 1–3. These noise-pairs included the same spectral envelope and temporal windowing as in the experiments. The results are given in Table 41.

BW (Hz)	$\mu(s_t[\Delta\Phi])$ (degrees)	$\mu(s_t[\Delta L])$ (dB)	$\sigma(s_t[\Delta\Phi])$ (degrees)	$\sigma(s_t[\Delta L])$ (dB)	corr
14	10.78	1.45	6.95	0.77	0.80
108	13.19	1.71	3.40	0.35	0.73
2394	13.56	1.75	1.04	0.12	0.41

Table 41: Values of the mean and standard deviation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ for noise-pairs with three bandwidths: 14 Hz, 108 Hz, and 2394 Hz. Correlation between the standard deviations is also given. Each value is based on 5000 noise-pairs.

A power regression was used to find the a power law describing the variation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ as a function of bandwidth. The power regression was of the form $y = mx^p$ where the bandwidth is the x-variable and the standard deviation of $s_t[\Delta\Phi]$ or $s_t[\Delta L]$ is the y-variable. The line of best fit was

$$y = 18.6x^{-0.37} \tag{42}$$

for $s_t[\Delta\Phi]$ and

$$y = 1.93x^{-0.36} \tag{43}$$

for $s_t[\Delta L]$. Both fits described over 99% of the variance of the points. Therefore, the dependence of the fluctuations in IPD and ILD over time vary as approximately the inverse cubed-root of the bandwidth.

APPENDIX C: FLUCTUATIONS AS A FUNCTION OF DURATION

The purpose of this appendix is to provide more precise values of the means and standard deviations of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$. Values of the mean and standard deviation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ for noise-pairs with three bandwidths: 14 Hz, 108 Hz, and 2394 Hz and four durations: 25 ms, 50 ms, 100 ms, and 500 ms. Correlation between the standard deviations is also given. Each value is based on 5000 noise-pairs. Differences between values from Appendix B and Appendix C are due to the different temporal shaping used for the collections. Appendix B used a 30 ms rise/fall time and Appendix C used a 10 ms rise/fall time.

BW (Hz)	Duration (ms)	$\mu(s_t[\Delta\Phi])$ (degrees)	$\mu(s_t[\Delta L])$ (dB)	$\sigma(s_t[\Delta\Phi])$ (degrees)	$\sigma(s_t[\Delta L])$ (dB)	corr
14	25	4.01	0.46	4.24	0.46	0.75
14	50	5.94	0.77	5.91	0.72	0.80
14	100	8.57	1.15	6.92	0.79	0.78
14	500	12.14	1.60	5.48	0.57	0.73
108	25	10.35	1.38	6.88	0.76	0.79
108	50	11.66	1.54	5.78	0.62	0.73
108	100	12.51	1.64	4.70	0.50	0.72
108	500	13.35	1.73	2.40	0.25	0.69
2394	25	13.30	1.70	3.27	0.34	0.41
2394	50	13.42	1.72	2.30	0.25	0.41
2394	100	13.50	1.74	1.66	0.18	0.41
2394	500	13.58	1.75	0.72	0.08	0.38

BIBLIOGRAPHY

Barron, M. (**1983**). "Objective measures of spatial impression in concert halls," *Proc. 6th International Congress on Acoustics, Paris*, Vol. **7**, 105-108.

Barron, M. (**2004**). "The current status of spatial impression in concert halls," Proc. 18th International Congress on Acoustics, Kyoto, Th2.B1.1, **IV**, 2449-2452.

Barron, M. and Marshall, A.H. (**1981**). "Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure," J. Sound and Vibration, **77**, 211-232.

Beranek, L.L. (**2004**). *Concert Halls and Opera Houses - Music, Acoustics, and Architecture*, Springer, New York.

Bernstein, L.R. and Trahiotis, C. (**1992**). "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies," J. Acoust. Soc. Am. **91**, 306-316.

Bernstein, L.R., Trahiotis, C., and Hyde, E.L. (**1998**). "Inter-individual differences in binaural detection of low-frequency tonal signals masked by narrowband or broadband noise," J. Acoust. Soc. Am. **103**, 2069-2078.

Blauert, J. and Lindemann, W. (**1986**). "Auditory spaciousness: Some further psychoacoustic analyses," J. Acoust. Soc. Am. **80**, 533-542.

Boehnke, S.E., Hall, S.E., and Marquardt, T. (**2002**). "Detection of static and dynamic changes in interaural correlation," J. Acoust. Soc. Am. **2002**, 1617-1626.

Bradley, J.S. and Soulodre, G.A. (**1995**). "Objective measures of listener envelopment," J. Acoust. Soc. Am. **98**, 2590-2597.

Brand, A., Behred, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature, **417**, 543-547.

Breebart, J., van de Par, S., and Kohlrausch A. (**1999**). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," J. Acoust. Soc. Am. **106**, 979-992.

Breebaart, J. and Kohlrausch A. (**2001**). "The influence of interaural stimulus uncertainty on binaural signal detection," J. Acoust. Soc. Am. **109**, 331-345.

Bronkhorst, A.W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508-1516.

Brughera, A.R., Stutman, E.R., Carney, L.H., and Colburn, H.S. (**1996**). "A model with excitation and inhibition for cells in the medial superior olive," Auditory Neuroscience, **2**, 219-233.

Carhart, R., Tillman, T.W., and Greetis, K.R. (**1969a**). "Release from multiple maskers: Effects of interaural time disparities," J. Acoust. Soc. Am. **45**, 411-418.

Carhart, R., Tillman, T.W., and Greetis, K.R. (**1969b**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694-703.

Colburn, H.S., Han, Y., and Culotta, C.P. (**1990**). "Coincidence model of MSO responses," Hear. Res. **49**, 335-346.

Colburn, H.S., Isabelle, S.K., and Tollin, D.J. (**1997**). "Modelling binaural detection performance for individual masker waveforms." In R. H. Gilkey & T. Anderson (Eds.), *Binaural and spatial hearing.* L.A. Erlbaum, Englewood Cliffs, N.J.

Culling, J.F., Colburn, H.S., and Spurchise, M. (**2001**). "Interaural correlation sensitivity," J. Acoust. Soc. Am. **110**, 1020-1028.

Dasika, V.K., White, J.A., Carney, L.H., and Colburn, H.S. (**2005**). "A network model of avian brainstem: Effects of superior olive nucleus inhibition," (abst) ARO Midwinter Meeting.

Domnitz, R.H. and Colburn, H.S. (**1976**). "Analysis of binaural detection models for dependence on interaural target parameters," J. Acoust. Soc. Am. **59**, 598-601.

Durlach, N.I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206-1218.

Durlach, N.I. (**1966**). "On the application of the EC model to interaural jnds," J. Acoust. Soc. Am. **40**, 1392-1397.

Durlach , N.I. (**1972**). "Binaural Signal Detection: Equalization and Cancellation Theory," in *Foundations of Modern Auditory Theory*, edited by J.V. Tobias (Academic, New York).

Durlach, N.I., Colburn, H.S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination: II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548-1556.

Eddins, D.A. and Green, D.M (**1995**). "Temporal integration and temporal resolution," *Handbook of Perception and Cognition - Hearing* second ed. ed. B.C.J. Moore, Academic Press, pp 207-242.

Egan, J. P., and Clarke, F. R. (**1956**). "Source and receiver behavior in the use of a criterion," J. Acoust. Soc. Am. **28**, 1267-1269.

Egan, J. P., Clarke, F. R., and Carterette, E. C. (**1956**). "On the transmission and confirmation of messages in noise," J. Acoust. Soc. Am. **28**, 536-550.

Egan, J.P., Schulman, A.I., and Greenberg, G.Z. (**1959**). "Operating characteristics determined by binary decisions and by ratings," J. Acoust. Soc. Am. **31**, 768-773.

Evilsizer, M.E., Gilkey, R.H., Mason, C.R., Colburn, H.S., and Carney, L.H. (**2002**). "Binaural detection with narrowband and wideband reproducible noise maskers: I. Results for human," J. Acoust. Soc. Am. **111**, 336-345.

Gabriel, K.J. and Colburn, H.S. (**1981**). "Interaural correlation discrimination: I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394-1401.

Gilkey, R.H., Robinson, D.E., and Hanna, T.E. (**1985**). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Am. Soc. **78**, 1207-1219.

Gilkey, R.H. and Robinson, D.E. (**1986**). "Models of auditory masking: A molecular psychophysical approach," J. Acoust. Am. Soc. **79**, 1499-1510.

Galvin, S. J., Podd, J. V., Drga, V., and Whitmore, J. (**2003**). "Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions," Psychonomic Bulletin & Review, **10**, 843-876.

Goldberg, J.M. and Brown, P.B. (**1968**). "Functional organization of the dog superior olivary complex: an anatomical and electrophysiological study," J. Neurophysiol. **31**, 639-656.

Goldberg, J.M. and Brown, P.B. (**1969**). "Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization," J. Neurophysiol. **32**, 613-636.

Grantham, D.W. (**1982**). "Detectability of time-varying interaural correlation in narrow-band noise stimuli," J. Acoust. Soc. Am. **72**, 1178-1184.

Grantham, D.W. and Wightman, F.L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511-523.

Grantham, D.W. and Wightman, F.L. (**1979**). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. **65**, 1509-1517.

Hafter, E.R. (**1971**). "Quantitative evaluation of a lateralization model of masking-level differences," J. Acoust. Soc. Am. **50**, 1116-1122.

Hafter, E.R. and Jeffress, L.A. (**1968**) "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563-569.

Hall, J.W., Grose, J.H., and Hartmann, W.M. (**1998**). "The masking level difference in low-noise noise," J. Acoust. Soc. Am. **103**, 2573-2577.

Han, Y. and Colburn, H.S. (**1993**). "Point-neuron model for binaural interaction in the MSO," Hear. Res. **68**, 115-130.

Hartmann, W.M. "How we localize sound," Phys. Today, November 1999.

Hartmann, W.M. and Constan, Z.A. (**2002**) "Interaural level differences and the level meter model," J. Acoust. Soc. Am. **112**, 1037-1045.

Isabelle, S.K. and Colburn, H.S. (**1987**). "Effects of target phase in narrowband frozen noise detection data," J. Acoust. Soc. Am. **82 (S1)**, S109.

Isabelle, S.K. and Colburn, H.S. (**1991**). "Detection of tones in reproducible narrowband noise," J. Acoust. Soc. Am. **89**, 352-359.

Isabelle, S.K. and Colburn, H.S. (**2004**). "Binaural detection of tones masked by reproducible noise: Experiment and models," *personal communication.*

Jain, M., Gallagher, D.T., Koehnke, J., and Colburn, H.S. (**1991**). "Fringed correlation discrimination and binaural detection," J. Acoust. Soc. Am. **90**, 1918-1926.

Jeffress, L.A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35-49.

Jeffress, J.A., Blodgett, H.C., Sandel, T.T., and Wood, C.L. (**1956**). "Masking of tonal signals," J. Acoust. Soc. Am. **28**, 416-426.

Koehnke, J., Colburn, H.S., and Durlach, N.I. (**1986**). "Performance in several binaural-interaction experiments," J. Acoust. Soc. Am. **79**, 1558-1562.

Kollmeier, B. and Gilkey, R.H. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709-1719.

Knudsen, E.I. and Konishi, M. (**1979**). "Mechanisms of sound localization in the barn owl," J. Neurophysiol. **41**, 687-695.

Levitt, H., and Rabiner, L.R. (**1967a**). "Binaural release from masking for speech and gain in intelligibility," J. Acoust. Soc. Am. **42**, 601-608.

Levitt, H., and Rabiner, L.R. (**1967b**). "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am. **42**, 820-829.

Licklider, J.C.R. (**1948**). "The influence of interaural phase relations upon the masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150-159.

McFadden, D., Jeffress, L.A., and Ermey, H.L. (**1971**). "Differences of Interaural Phase and Level in Detection and Lateralization: 250 Hz," J. Acoust. Soc. Am. **50**, 1484-1493.

Møller, A.R. (**1965**). "An experimental study of the acoustic impedance of the middle ear and its transmission properties," Acta Otolar. **60**, 129-149.

Nix, J. and Hohmann, V. (**2001**). "Statistics of interaural parameters in real sound fields and its application to sound source localization," JASA preprint.

Osman, E. (**1971**). "A correlation model of binaural masking level differences," J. Acoust. Am. Soc. **50**, 1494-1511.

Osman, E. (**1973**). "Correlation model of binaural detection: interaural amplitude ratio and phase variation for signal," J. Acoust. Am. Soc. **54**, 386-389.

Pollack, I. (**1959**). "On indices of signal and response discriminability," J. Acoust. Soc. Am. **31**, 1031.

Pollack, I. and Decker, L.R. (**1958**). "Confidence ratings, message reception, and the receiver operator characteristic," J. Acoust. Soc. Am. **30**, 286-292.

Pollack, I. and Trittipoe, W.J. (**1959a**). "Binaural listening and interaural noise cross correlation," J. Acoust. Soc. Am. **31**, 1250-1252.

Pollack, I. and Trittipoe, W.J. (**1959b**). "Internal noise correlation: examination of variables," J. Acoust. Soc. Am. **31**, 1616-1618.

Ruotolo, B.R., Stern, R.M., and Colburn, H.S. (**1979**). "Discrimination of symmetric time-intensity traded binaural stimuli," J. Acoust. Soc. Am. **66**, 1733-1737.

Schulman, A.I. and Mitchell, R.R. (**1965**). "Operating characteristics from yes-no and forced-choice procedures," J. Acoust. Soc. Am. **40**, 473-477.

Smith, P.H., Joris, P.X., and Yin, T.C.T. (**1993**). "Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: Evidence for delay lines to the medial superior olive," J. Comp. Neurology. **331**, 245-260.

Trahiotis, C. and Kappauf, W.E. (**1978**). "Regression interpretation of differences in time-intensity trading ratios obtained in studies of laterality using the method of adjustment," J. Acoust. Soc. Am. **64**, 1041-1047.

Tsuchitani, C. (**1977**). "Functional organization of lateral cell groups of cat superior olivary complex," J. Neurophysiol. **40**, 296-318.

Tsuchitani, C. and Boudreau, J.C. (**1966**). "Single unit analysis of cat superior olive S-segment with tonal stimuli," J. Neurophysiol. **29**, 296-318.

van de Par, S. and Kohlrausch, A. (**1995**). "Analytic expressions for the envelope correlation of certain narrow-band stimuli," J. Acoust. Soc. Am. **98**, 3157-3169.

van de Par, S. and Kohlrausch, A. (**1999**). "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," J. Acoust. Soc. Am. **106**, 1940-1947.

Viemeister, N.F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364-1380.

Webster, F.A. (**1951**). "The influence of interaural phase on masked thresholds. I. The role of interaural time-deviation," J. Acoust. Soc. Am. **23**, 452-462.

Wightman, F.L. (**1971**). "Detection of binaural tones as a function of masker bandwidth." J. Acoust. Soc. Am. **50**, 623-636.

Wilbanks, W.A. (**1971**). "Detection of a narrow-band noise as a function of the interaural correlation of both signal and masker," J. Acoust. Soc. Am. **49**, 1814-1817.

Wilbanks, W.A. and Whitmore, J.K. (**1967**). "Detection of monaural signals as a function of interaural noise correlation and signal frequency," J. Acoust. Soc. Am. **43**, 785-797.

Yin, T.C.T., and Chan, J.C.K. (**1990**) "Interaural time sensitivity in medial superior olive of cat," J. Neurophysiol. **64**, 465-488.

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

Young, L.L. and Levine, J. (**1977**). "Time-intensity trades revisited," J. Acoust. Soc. Am. **61**, 607-609.

Yost, W.A. and Hafter, E.R. (**1987**). Lateralization: In: *Directional Hearing.* Yost, W.A. and Gourevitch, G. (eds.). New York: Springer-Verlag. 49-84.

Zhou, Y. and Colburn, H.S. (**2005**). "A model study of the effect of the afterhyperpolarization on the sound-level sensitivity of the LSO chopper units," (abst) ARO Midwinter Meeting.

Zurek, P.M. (**1991**). "Probability distributions of interaural phase and level differences in binaural detection stimuli," J. Acoust. Soc. Am. **90**, 1927-1932.

Zurek, P.M. and Durlach, N.I. (**1987**). "Masker-bandwidth dependence in homophasic and antiphasic tone detection," J. Acoust. Soc. Am. **81**, 459-464.