Modeling the Dependence of Auditory Lateralization on Frequency and Bandwidth

Glenn D. Shear

Submitted to the Department of Electrical and Computer Engineering in Partial Fulfillment of the Requirements for the Degree of Master of Science at

Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

September 1987

Table of Contents

List of Figures

List of Tables

Abstract

Binaural interaction refers to the process by which the human auditory system infers information regarding differences in the signals at the two ears in order to perform a given perceptual task. Models that describe this process are studied in order to better understand the actual mechanisms involved. In this work we modify a particular model known as the position-variable model [Stern, R.M. and Colburn, H.S., J. Acoust. Soc. Am. **64**, 127-140 (1978)]. Major modifications to the model include (1) specification of a new frequency-dependent form of the function that describes the assumed distribution of interaural coincidence detectors as a function of their characteristic interaural delay, (2) replacement of the exponential rectifier of the original model with a less expansive, half-wave rectifier, and (3) exchange of the relative positions of the rectifier and lowpass filter used in the model's description of peripheral auditory processing. The modified model quantitatively describes the lateralization of tones as a function of interaural time delay (ITD) and frequency while remaining consistent with most previous predictions of the original position-variable model. The new model also qualitatively describes data for the lateralization of bandpass noise centered at 500 Hz as a function of bandwidth, and we have briefly considered how the model might be further modified to better describe these data. Two cues have proven useful in this regard. First, as the bandwidth of noise stimuli increases, the magnitude of the primary peak of the interaural correlation function increases relative to the magnitudes of secondary peaks. Second, as bandwidth increases, only those regions of the display near the true ITD remain consistent across characteristic frequency. The available psychoacoustical evidence is inadequate to allow us to determine which of these cues (or combination of cues) is actually exploited by the human auditory system.

Acknowledgements

A long, long time ago I wandered into Rich Stern's office but a wee lad of an engineer. Today I'm still a lad, but I'm a lad with a masters degree. I'd like to thank Rich for making it all possible and even quite enjoyable. I couldn't have picked a better person to be my advisor (even if I had had a choice).

I want to thank Laurel for feeding me ice cream in the bathroom, Matt for lending me his name (and his ear), Marko for teaching me how to make decisions, Jim for the therapeutic hack breaks, and the rest of the ECE community for not "talking technical" in front of me.

I am very grateful to my family for the long-distance moral support and for being so patient. By the way, I'll be home on Wednesday... No really, I will be.

Finally, special thanks to Grandma Bertha, Poppy Harry, Grandma Ruthy, and Poppy Sam for letting me grow up with them.

Chapter 1
Introduction

Psychoacoustics is the study of how the physical properties of an acoustical stimulus are related to our perception of that stimulus. It is helpful in such studies to distinguish between tasks that require the use of both ears, binaural hearing, and tasks that can be performed equally well with the use of only one ear, monaural hearing. A familiar example of binaural hearing is the localization of a sound source in space. Using only one ear, it is usually very difficult to perform such a task. However, if both ears are used, an externalized and fused "acoustic-image" is generally perceived in the vicinity of the source. An image is said to be *externalized* if it is perceived outside of the head, and it is said to be *fused* if it is spatially compact and unitary. The ability of the human auditory system to form such an estimate of the position of the sound source is based, in large part, on its ability to detect interaural time differences (ITDs) and interaural intensity differences (IIDs) present in the acoustical signals.

Many investigations have been conducted to examine the dependence of a number of spatial attributes of the perceived acoustic image on these two cues. In such studies, headphones are commonly employed in order to carefully and independently control the values of the various interaural disparities present in the binaural stimulus. Under these presentation conditions, the sound image is usually perceived within the subject's head (*i.e.*, it is not externalized), and the task of determining its position is referred to as *lateralization* rather than localization. Stimuli that are identical at both ears are said to be *diotic*, and the corresponding image is perceived in the center of the subject's head. Stimuli that are not diotic are referred to as *dichotic*, and the image is generally perceived closer to the ear receiving the signal that is more intense and/or leading in time. It is possible, however, to construct dichotic stimuli which produce a centered image by "trading" the effects of IID and ITD.

By proposing and testing hypothetical models of the processes involved in this and

other types of binaural interaction, we can begin to understand how information carried by the auditory nerves is processed and utilized by the more central mechanisms of the sensory pathway. This insight is important to the development of many devices that interface with the auditory sensory system, such as stereo systems and cochlear implants.

The purpose of this work is to modify a specific model of binaural interaction, the position-variable model (PV model) (Stern and Colburn, 1978), so as to better describe lateralization phenomena at a wide range of frequencies. While the current model is capable of describing a relatively large body of data at 500 Hz, it is not able to describe recently acquired data regarding the lateralization of pure tones as a function of frequency (Schiano, Trahiotis, and Bernstein, 1986). It is our intent to develop a modified model that can describe the data of Schiano, *et al.* while still remaining consistent with earlier, correct predictions of the position-variable model. In addition, we provide a cursory look at how the model could be further extended to describe the lateralization of bandpass noise, containing various interaural temporal disparities, as a function of bandwidth.

This report begins with a discussion in Chapter 2 of some relevant theoretical and experimental studies of binaural perception. In Chapter 3, we describe the modifications made to the position-variable model and the motivations for these changes. The procedure used to constrain the free parameters of the new model is presented in Chapter 4, and preliminary efforts to describe the bandpass noise data are discussed in Chapter 5. We conclude the report by considering the implications of the new predictions and by assessing the validity of the model based on its ability to describe the data as well as its physiological plausibility.

Chapter 2
Background

In this chapter, we discuss some selected topics in psychoacoustics which are directly related to the research described by this report. The first section is devoted to familiarizing the reader with a small body of psychoacoustic data which will be used in later chapters to constrain certain parameters of the new model. In the second section, we address a few general issues regarding models of binaural interaction and provide examples of these issues in the context of three crosscorrelation based models.

2.1. Some Relevant Psychoacoustical Data

Relations between the perceptual attributes of a sound generated from a binaural stimulus and those of a sound generated from the component monaural stimuli depend strongly on the particular attributes and stimuli considered. Many investigations have been conducted to determine the dependence of a number of these attributes (*e.g.*, pitch, binaural coherence, and subjective laterality) on a variety of stimulus parameters. It is well beyond the scope of this chapter to review all the data associated with these studies. Instead, we will focus on a small group of experiments whose results have proven problematical for existing models of binaural interaction. These studies were designed to reveal dependencies based on interaural timing information and certain spectral parameters of the stimuli. They do not examine effects of interaural intensity differences.

The investigations considered below can be divided into two categories: lateralization matching experiments, in which the subject is asked to "point" to the intracranial position of the sound image, and detection experiments, in which the subject is required to specify whether or not a target stimulus is present.

2.1.1. Lateralization Experiments

The general goal of the experiments described in this section is to quantitatively determine the lateral position of a sound image elicited by a controlled binaural stimulus.

The classes of stimuli considered here have been designed to test the dependence of subjective lateral position on interaural time and phase differences as well as certain spectral parameters of the stimulus. Specifically, the experiments make use of (1) low-frequency tones with various ITDs, and (2) bandpass noise centered at 500 Hz with a number of combinations of bandwidth, ITD, and interaural phase difference (IPD).

As was previously mentioned, the experiments we will review make use of a lateralization matching paradigm. The subject is presented a target stimulus and is instructed to adjust a graphical or acoustical pointer to match the lateral position of the target. This process constitutes one trial, and in general several trials are conducted for each target stimulus. Precautions are taken so that the subjects cannot effectively use results from previous trials to influence their judgement of the current target.[1] The lateralization matching task assumes that the images produced by the target and the acoustical pointer are fused and internalized. If the target does not produce a fused image, the task is more difficult and the results may vary greatly from trial to trial. Stimulus configurations for which such problems have been observed will be discussed later in this section.

2.1.1.1. Pure Tones

In this section, we discuss the dependence of subjective lateral position of tones on two simulus parameters, tone frequency and ITD. In addition, we present some simple arguments as to why the observed dependencies are intuitively appealing. It is helpful in trying to interpret the experimental results to first consider variations in only one parameter (while holding the other fixed) rather than considering the effects of simultaneous changes of ITD and frequency.

We will begin by describing the lateralization of a tone as a function of interaural time difference. The major source of evidence for this discussion is a study conducted by Sayers (1964), which made use of a visible scale set up in front of the subject to quantify

---

[1] See Appendix A for a more detailed description of one particular lateralization matching experiment.

judgements about the acoustic image position. Sayers found that each listener of seven exhibited generally similar behavior in the sense that although the results for individual subjects reveal different magnitudes of lateral displacement, suitable linear expansion of the judgement scale compensates for such differences in each case. Results for a typical subject are shown in Figure 2-1, and we consider the following to be the significant characteristics of the average-result curves depicted there:

- They are periodic (with respect to ITD) and have the same period T_0 as the stimulus tone.
- They are approximately linearly proportional to the value of the ITD τ_s when $|\tau_s| < T_0/4$.
- They achieve maximum displacement from center for values of $|\tau_s|$ near $T_0/3$.
- They indicate a centered image for ITDs of $\pm T_0/2$.

This behavior is consistent with the findings of Domnitz and Colburn (1977) at 500 Hz. One may note from Figure 2-1(a) that there are considerable differences between the individual judgement reports and the averaged reports for values of ITD near $\pm T_0/2$. This discrepency is related to observations that such stimuli do not produce well fused acoustic images and may even produce multiple images (one on each side of the head).

Consider now the dependence of the observed lateral position on tone frequency in the curves characterized above. Referring to Figure 2-1(b), we see that the lateralization curves for different frequency tones are essentially identical when plotted as a function of phase shift (where, for example, $\tau_s = T_0/4$ corresponds to a 90° phase shift). However, Sayers points out that this observation is only partially valid. It is true that the "shape" of these curves are similar, but it is not true that lateral position is a simple function of IPD. Specifically, listeners tend to automatically scale their judgements (for each block of trials) to fit within a certain preferred region of the visual scale. Sayers adjusted the experimental procedure to compensate for this tendency and thus was able to correctly compare relative image position across frequencies. He found that (1) the slope of perceived position computed with respect to time delay appears to be independent of frequency (up to 1200 Hz), (2) the magnitude of the maximum image displacement decreases with increases in signal frequency, and (3) at frequencies above about 1500 Hz, the extent of this maximum displacement diminishes even more rapidly with frequency.

These observations are consistent with the findings of Schiano, *et al.* (1986) who

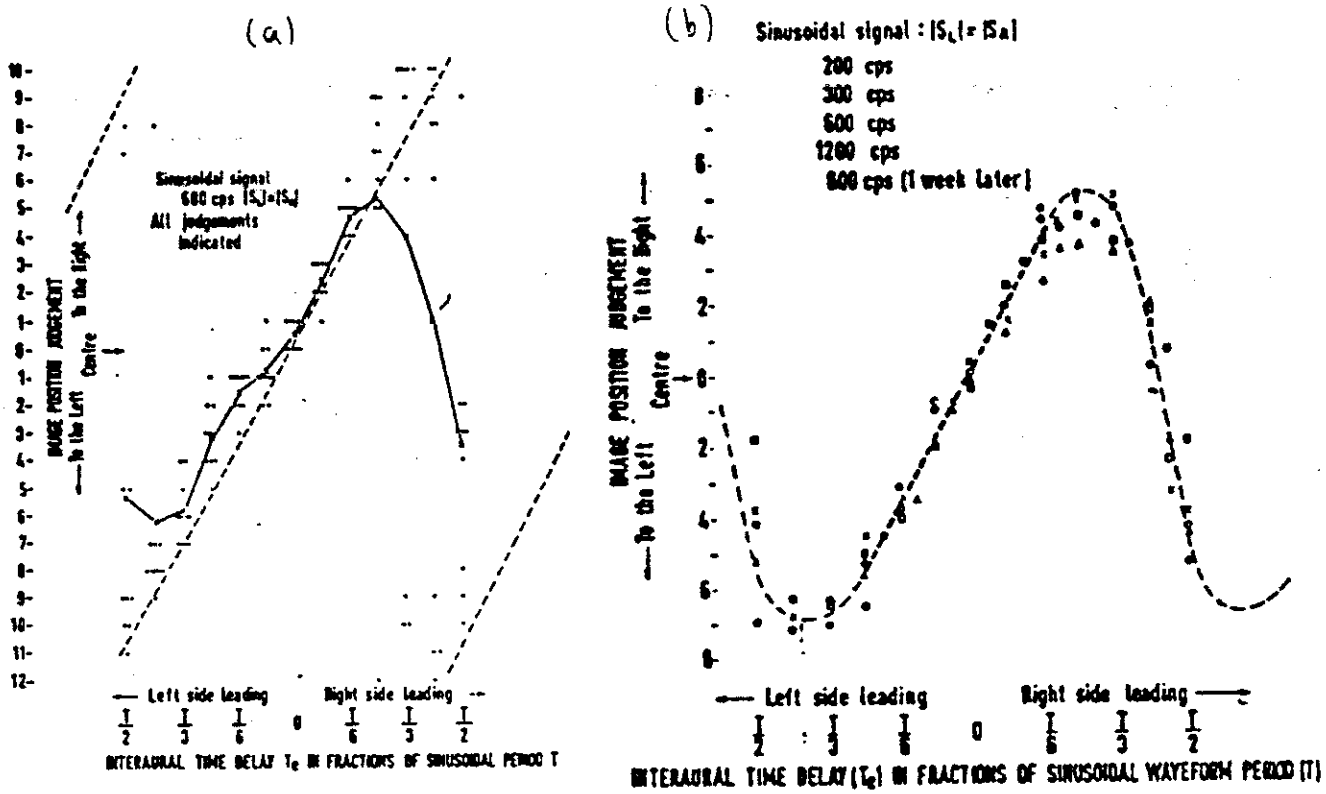

Figure 2-1: Lateralization data for tones as obtained by Sayers (1964). Relative lateral position determined with the aid of a visual scale is plotted as a function of ITD τ_a expressed in terms of the tone period T_0. Panel (a) shows individual responses and the average response curve for a typical subject at 600 Hz. Panel (b) shows average responses at the specified frequencies for a typical subject. [From Sayers, 1964.]

examined the lateralization of tones as a function of frequency for a few small ITDs. In these experiments, an acoustical pointer (bandpass noise with an adjustable IID) was employed, thereby eliminating the problem of arbitrary scale adjustments between blocks of trials for a single subject. Results from different listeners are, however, still subject to proportional scaling. The average responses of each of three subjects to tones with ITDs of $150\,\mu s$ are shown in Figure 2-2. (The different symbols in this figure represent results for different subjects.) The results obtained for an ITD of $100\,\mu s$ exhibit a similar dependence on tone frequency, and as we would expect they indicate an overall diminished lateral displacement with respect to the $150\,\mu s$ case. Complete experiments using longer time delays have not been conducted. However, on the basis of pilot studies by Schiano, *et al.* and the results of Sayers (1964) we expect that position vs. frequency

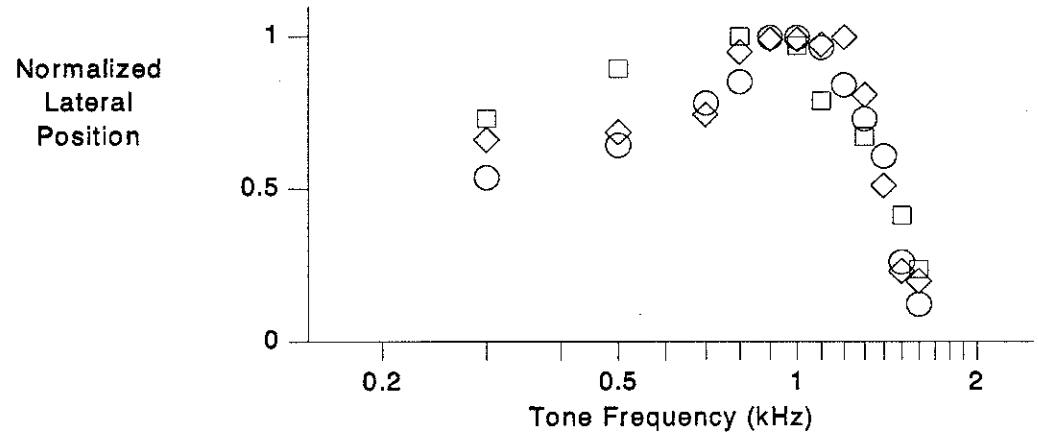

Figure 2-2: Lateralization data for tones as obtained by Schiano, *et al.* (1986). Relative lateral position of a tone with an ITD of 150 μs is plotted as a function of the tone frequency. Different symbols represent results for different subjects averaged over trials and scaled by the maximum deviation from center for each subject. [Normalized data from Schiano, *et al.*, 1986.]

curves for stimuli with larger ITDs (greater than $200\ \mu s$) will be similar to the ones shown in Figure 2-2, except that as signal frequency increases, relative lateral position will begin to decrease at some frequency less than 1200 Hz (approximately $\frac{1}{4\times\tau_s}$). In summary, the lateralization of tones with small interaural delays depends on the frequency of the tone and the ITD in the following three ways:

- Subjective lateral position is related to the value of the ITD through a proportionality factor which depends on the signal frequency.
- This proportionality factor is nearly constant for tone frequencies up to about 1200 Hz.
- The magnitude of this factor decreases quickly to near zero as the tone frequency is increased from 1200 Hz to 2000 Hz.

Expressed mathematically,

$$P(\tau_s, f_0) \approx C H_{data}(f_0) \times \tau_s, \quad \text{for } \tau_s < \frac{1}{4f_0} \tag{2.1}$$

where $P(\tau_s, f_0)$ represents subjective lateral position expressed as a function of the ITD τ_s and the tone frequency f_0, C is a proportionality constant, and $H_{data}(f)$ is a lowpass filter function with a corner frequency of about 1200 Hz. [We recognize that this expression

somewhat oversimplifies the frequency dependence exhibited by the data for tones less than 1200 Hz. However, Schiano, *et al.* (1986) have characterized these data in precisely the same manner and consider this to be a satisfactory description of the phenomena. Moreover, a simple closed-form expression such as the one in equation (2.1) facilitates the development of mathematical constraints for certain parameters of the model.]

We now present an intuitive argument as to why the lateralization phenomena described above should be correct, including an explanation of the "confusion" caused by antiphasic (or out-of-phase) tones. If we consider a naturally presented, low-frequency tone, it is reasonable to expect that a listener will localize the sound in the vicinity of the source and that the localization judgement will be based, in part, on the value of ITD detected in the binaural stimulus. Referring to the presentation configuration depicted in Figure 2-3, one can develop the following approximate relation between the ITD τ_s and the relative source angle θ (expressed in radians) by assuming a head radius of 8.75 cm and the speed of sound to be 34,000 cm/sec (Woodworth, 1938)

$$\tau_s \approx 257 \times (\theta + \sin\,\theta)\ \mu s \tag{2.2}$$

From this expression, we see that when the source is positioned directly in front of the subject, the value of the ITD is zero, and as the source is moved further to one side, the magnitude of the ITD increases, achieving a maximum magnitude of about $\pm 660\,\mu s$ when $\theta = \pm\pi/4$ radians. This limitation in the magnitude of naturally occuring ITDs is often referred to as the "headwidth constraint" since the maximum ITD is directly proportional to the radius of the listener's head.

Although the stimuli used by Sayers do not contain all the cues present in a naturally presented tone, we still expect the lateralization data to exhibit a dependence on τ_s similar to that exhibited by θ. As the data indicate, this is true to a certain extent. To understand why the image does not continue to move closer to the leading ear as the ITD is increased beyond $T_0/3$, we consider how the auditory system might estimate the value of τ_s from the acoustical signals. This is a common estimation problem in several areas of engineering, and crosscorrelation analysis has proven to be an effective solution.

In the present discussion we consider the interaural crosscorrelation function computed from the acoustical waveforms before *any* processing by the peripheral auditory system. Although we will primarily be concerned with crosscorrelation functions of the peripherally-transformed signals in following chapters, many of the concepts introduced below will be applicable to our later discussions.

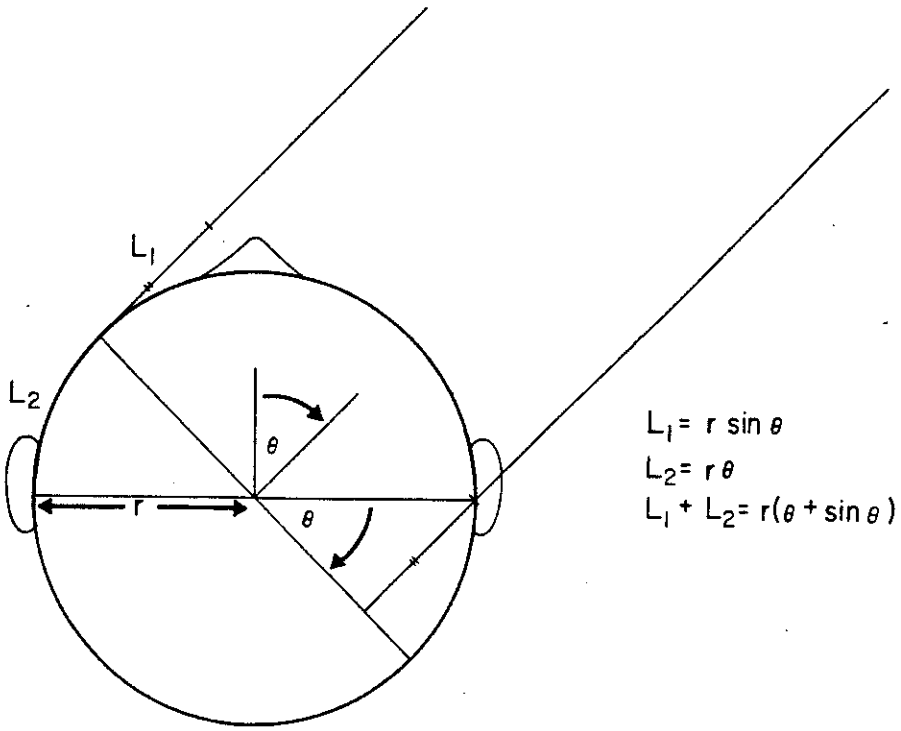

Figure 2-3: Path length difference ($L_1 + L_2$) between the two ears for a distant source at an azimuth of θ. [Adapted from Woodworth, 1938.]

The basic procedure of crosscorrelation analysis is to perform a time-averaged crosscorrelation of the two signals, and then to estimate the ITD from the resulting crosscorrelation function, recognizing that its primary (or largest) peak will usually occur when the function argument τ is equal to τ_s. A problem arises, however, when the input signals are continuous and periodic (*e.g.*, tones) because the resulting crosscorrelation function is also periodic. Thus, all peaks of this function are of equal magnitude, and it is impossible to determine which peak corresponds to the actual ITD. Based on the *a priori* knowledge that most interaural delays are in the range from $-660\,\mu s$ to $+660\,\mu s$, we argue that a good ITD-estimator would resolve this problem of ambiguity by computing a probabilistic estimate in which the peaks closest to the center (*i.e.*, closest to $\tau = 0$) are weighed more heavily than the other peaks. (The simplest of such schemes would be to merely pick the peak closest to the midline.) For any estimation procedure of this type, the predicted value of the ITD will be a periodic function of the true value of the ITD. It is for this reason that the lateralization of tones as a function of ITD is periodic with the same period as that of the tone.

The "arbitration rule" outlined above fails when the ITD corresponds to a $\pm 180^\circ$ phase difference. For such a stimulus it is impossible to determine which ear is receiving the signal that is leading in time. It is just as likely that the signal to the left ear is leading by half of a period as it is that the signal to the right ear is leading by the same amount. Thus, as the data indicate, lateralization judgements may be made to the left side, to the right side, or even centered. If we relate the degree of image fusion to how "certain" the auditory system is about its ITD estimate, we can explain why such antiphasic tones do not produce well fused binaural images.

Since the value of ITD associated with a given source location does not depend strongly on the frequency of the tone, we do not expect the auditory system's ITD estimate to depend on frequency, either. This is consistent with the experimental evidence at low frequencies. At higher frequencies (above about 1500 Hz), however, there is a considerable probability that a naturally occuring time delay will be estimated incorrectly, and thus a good localization mechanism should ignore timing cues at these frequencies and concentrate on other, more useful cues (such as interaural intensity difference). Electrophysiological evidence obtained from actual auditory-nerve fibers (in live cats) indicates, in fact, that neural units are incapable of transmitting timing information from high-frequency tones to the more central auditory pathways. In light of this evidence, it is understandable that the image produced by high-frequency tones is not displaced far from the midline for any value of ITD.

2.1.1.2. Bandpass Noise

The experiments considered here are intended to provide insight into how the auditory system might make use of temporal and spectral cues available in bandpass noise stimuli. Specifically, we examine the dependence of the subjective lateral position of bandpass noise centered at 500 Hz on bandwidth for four combinations of interaural time and phase differences, $(1500\,\mu s, 0^\circ)$, $(1000\,\mu s, 90^\circ)$, $(500\,\mu s, 180^\circ)$, and $(0\,\mu s, 270^\circ)$. These particular combinations have been chosen because they all produce interaural crosscorrelation maximae at the same values of τ for noise centered at 500 Hz. The significance of this restriction will be revealed in the discussion below.

The experimental evidence presented here comes from two pilot investigations. The first, by Bernstein and Trahiotis (unpublished)[2], was conducted using only the

[2]These data were obtained through personal correspondence with C. Trahiotis and L.R. Bernstein.

$(1500\,\mu s, 0^\circ)$ case. The second, by Shear and Stern (unpublished)[3], produced results for all four (ITD, IPD) pairs but involved only one subject. Both studies made use of an acoustical pointer, so results for each subject are consistently scaled across trials.

Consider now a binaural band of noise centered at 500 Hz for which the signal to the left ear is leading the signal to the right ear by $1500\,\mu s$. Averaged (and vertically scaled) judgements for one subject from each study are shown in Figure 2-4. For narrow bands of noise, the image is perceived on the left side of the head, but as the bandwidth increases, the image moves to the right side. Bandwidths of 100 to 200 Hz produce nearly centered images.

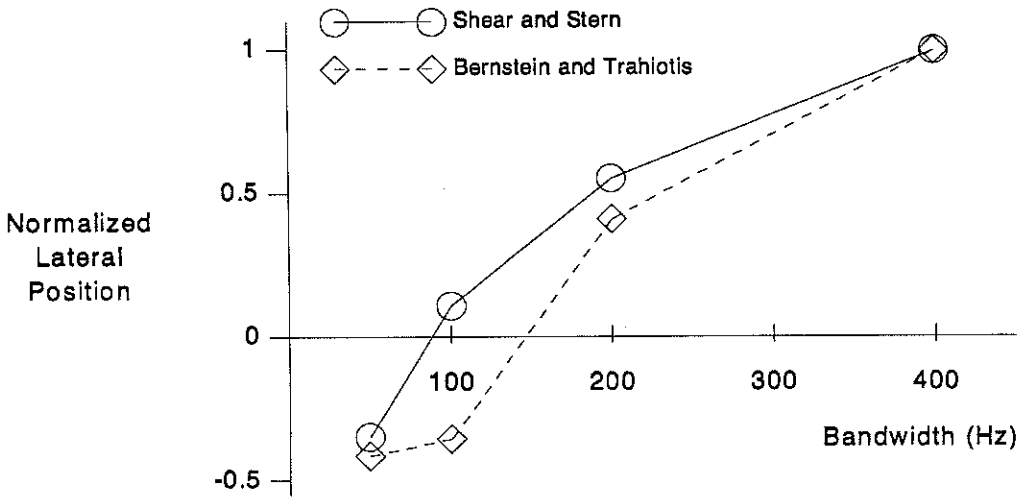

Figure 2-4: Average relative lateral position judgements for bandpass noise centered at 500 Hz with an ITD of $1500\ \mu s$ for two subjects plotted as a function of bandwidth. [Normalized, unpublished data of Shear and Stern and Bernstein and Trahiotis.]

These observations can be interpreted in the context of our previous, heuristic, discussion of interaural-delay estimation.

The crosscorrelation function for bandpass noise can be characterized in the following manner,

---

[3] See Appendix A.

$$R(\tau) = \sigma^2 \, \gamma(\tau - \tau_s) \cos \left[2\pi f_0 (\tau - \tau_s) - \phi_s\right] \tag{2.3}$$

where τ_s and ϕ_s are the values of ITD and IPD present in the binaural stimulus. The function $\gamma(\tau)$ is a gradually decaying "envelope" which achieves a maximum value of 1 at τ = 0 and which is symmetric about this point. The shape of this envelope depends on the bandwidth of the input noise. For very narrowband signals, $\gamma(\tau)$ is nearly constant and thus a narrowband stimulus centered at 500 Hz with an ITD of 1500 μs should be lateralized in much the same way as a similarly delayed 500 Hz tone. Since the value of the interaural delay is greater than half the period of the tone, our hypothetical ITD estimator will predict, incorrectly, that the value of the ITD is -500 μs, and the image will be lateralized towards the ear receiving the signal that is actually lagging in time. As the bandwidth of the stimulus increases, the envelope of the crosscorrelation function sharpens, attenuating all but the primary peak of $R(\tau)$, and thus the correct value of the ITD becomes more evident. In this way, broader-band stimuli provide more information which is useful in estimating ITDs.

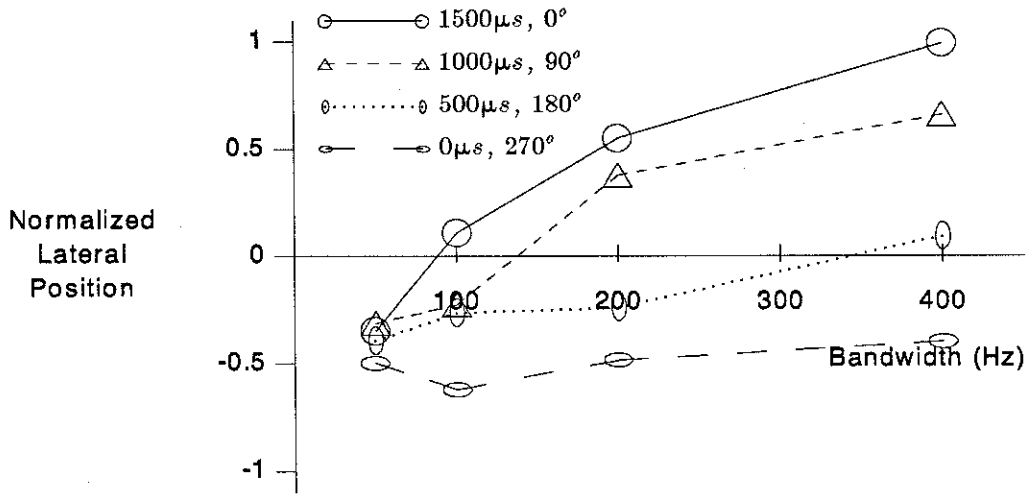

Figure 2-5: Average relative lateral position of bandpass noise centered at 500 Hz with several (ITD, IPD) combinations plotted as a function of bandwidth. [Normalized, unpublished data of Shear and Stern.]

The effect of phase differences on the lateralization judgements is illustrated in Figure 2-5. The observed decrease in dependence of position on bandwidth as larger phase shifts are introduced is intuitively appealing since IPDs are not generally present in

naturally occuring broadband sounds. Once again, our simple "model" is capable of qualitatively describing these phenomena. As equation (2.3) indicates, phase differences in the signals cause a shift of the fine structure of the crosscorrelation function, but not of its envelope. In light of this fact and considering the (0 μs, 270°) case, we see that the envelope of the crosscorrelation function will be centered at $\tau = 0$ causing the peak at -500 μs to be slightly larger than the one at 1500 μs and thus the image will be perceived on the lagging side of the head, almost independent of bandwidth. The other (ITD, IPD) combinations also yield peaks at -500 μs and 1500 μs, but the relative strength of these peaks depends on the position of the envelope, and consequently on the value of the ITD. Therefore, as $\tau_s \rightarrow 1500\mu s$, we expect the data to exhibit a stronger dependence on bandwidth.

2.1.2. Detection of Tones in Noise

In Chapter 1, we pointed out the benefits of having two ears when trying to localize the position of a sound source. Another commonly performed task that is also greatly aided by binaural hearing, is that of selectively listening to a particular sound that is presented simultaneously with other distracting sounds or noises. A familiar example of this selectivity is conversing with another person at a crowded party. In such an environment, each ear receives the summation of several signals from different sources. If, God forbid, only one ear is available, it is difficult to separate the acoustical signal into its components. However, when two ears are used, the sound of interest can be better detected or attended to by taking advantage of changes in the patterns of the interaural crosscorrelation function that occur when the sound is presented.

A great deal of effort has been devoted to examining how performance in this sort of detection task depends on various stimulus parameters. The enthusiasm is due, in large part, to the obvious applications of this information. Once again, we consider only a small subset of the available data, and our choice is motivated by our desire to have the new model remain consistent with the binaural phenomena correctly described by the original model. More specifically, the detection data considered here were used by Colburn (1977a) to constrain a particular function in the model, $p(\tau)$, which we attempt to re-specify in the current work.

The stimuli used in these studies are composed of a binaural tone added to a broadband noise masker, each with either a 0° or 180° IPD. Unlike the lateralization experiments of

the previous section, the task involved here does not require the listener to describe a complex perceptual attribute to the investigator. Instead, the subject need only indicate whether or not she or he can detect the presence of the tone in the masker. We define the detection threshold to be the intensity of the tone at which the listener responds correctly 75 per cent of the time. The notation $N_{\phi_n} S_{\phi_s}$ refers to a stimulus configuration for which ϕ_n and ϕ_s are the interaural phase differences present in the noise and tone, respectively. (At times, we will use this notation to refer to the actual detection threshold for the indicated stimulus configuration.)

Positive values of the ratio of $N_\pi S_0$ detection thresholds to $N_0 S_\pi$ detection thresholds (expressed in dB) indicate that it is more difficult to detect diotic tones in out-of-phase maskers than it is to detect out-of-phase tones in diotic maskers. (For convenience we will use the notation $\frac{N_\pi S_0}{N_0 S_\pi}$ to refer to this ratio.) As the results of a number of studies indicate (cf. Durlach and Colburn, 1978), this is the case for low-frequency tones. However, at higher frequencies performance is nearly the same for both stimulus conditions (see Figure 2-6).

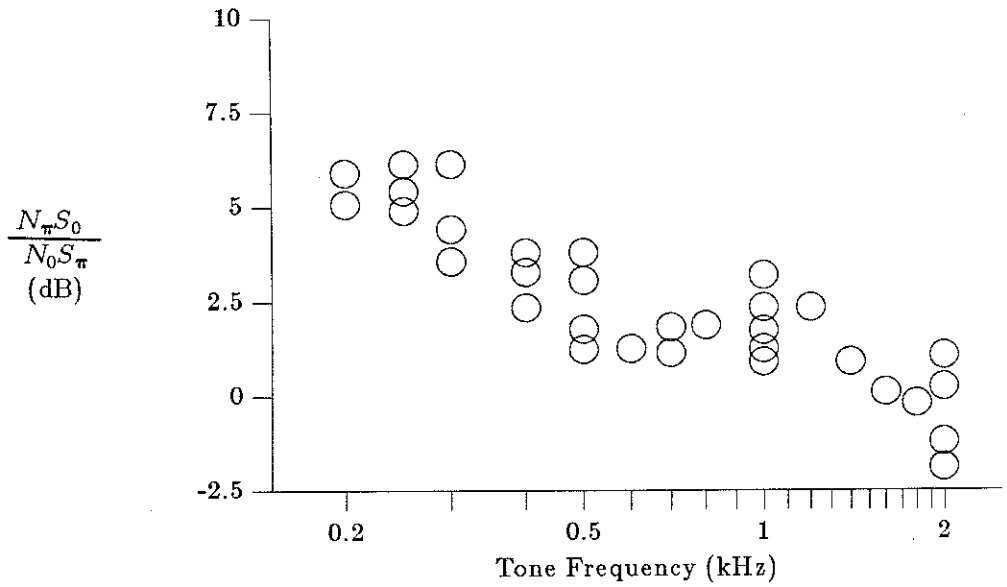

Figure 2-6: Relative detection thresholds $\frac{N_\pi S_0}{N_0 S_\pi}$ (dB) as a function of tone frequency. [These data were compiled by Colburn (1977a) from several studies.]

Reasons for this observed behavior are subtle, and a more thorough analysis is required to discuss the major issues involved in the perception of these stimuli. We will consider this phenomenon in detail in Section 4.1.2.

2.2. Models of Binaural Interaction

The purpose of this section is to provide the reader with an understanding of the similarities and differences among selected contemporary models of binaural interaction. In particular, we wish to make it clear that the implications of results obtained using the position-variable model should be applicable to other theories of binaural interaction.

As we have already seen, interaural crosscorrelation can be a powerful analytical tool in trying to understand why certain binaural phenomena occur. The usefulness of this technique has long been recognized, and it is a key component of many contemporary models of binaural interaction. The three models described below have been selected in order to help demonstrate the evolution of correlation-based theories over recent years. We will be using a slightly modified version of the position-variable model, (described in Section 2.2.2) in the following chapters.

2.2.1. The running crosscorrelation model

The modelling work of Sayers and Cherry (1957) represents one of the first serious efforts to produce a quantitative, crosscorrelation-based description of lateralization and binaural fusion phenomena. As was pointed out in their original paper, interaural correlation seems to be a promising mechanism to explain these phenomena for two reasons. First, it provides a measure of the similarity between the signals at the two ears, and such a measure is apparently important in the binaural fusion process since an image is generally more fused the more alike the binaural inputs are. Second, interaural crosscorrelation allows for easy estimation of the ITD present in a binaural stimulus, and this cue, as we have already seen, is relevant to the formation of lateral position judgements.

A pictorial representation of the model is shown in Figure 2-7. The signal at each ear is first "rectified" by adding a dc term proportional to its intensity level, and then it is compared with the signal from the opposite ear by means of a running correlation. The output of the correlator is weighted by two factors. The first is a double-sided exponential

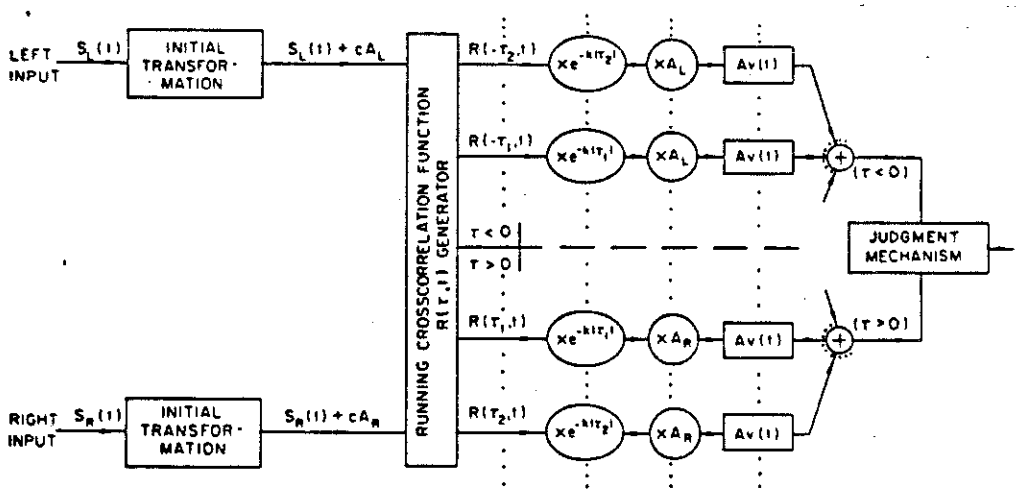

Figure 2-7: Schematic representation of the running cross-correlation model of Sayers and Cherry (1957). [From Colburn and Durlach, 1978.]

function of the correlation argument τ that "windows" the correlation function in order to focus the attention of the "binaural analyzer" on those values of τ that are smaller in magnitude. The second provides the means by which interaural intensity differences are incorporated in the image formation process. The "left" half of the crosscorrelation plane, which is associated with images on the left side of the head, is multiplied by the averaged level of the signal to the left ear. Similarly, the "right" half-plane is multiplied by the averaged level of the signal to the right ear. The resulting running crosscorrelation function is averaged over time, with the most recent values of the function being weighed most heavily. The areas underneath this final function in each half-plane are computed, and they are denoted by I_L and I_R.

Sayers and Cherry used the magnitude of the judgement criterion given by

$$J = \frac{I_L - I_R}{I_L + I_R} \tag{2.4}$$

as a measure of the probability that a listener will perceive the image of a binaural sound on the left side of the head, for positive J, and on the right side of the head, for negative J.

By this means, the model is able to successfully describe the (experimentally determined) probability that a listener will correctly lateralize a given sound for a number of binaural stimulus configurations including simple speech waveforms (Sayers and Cherry, 1957).

Several slightly modified versions of this model have been used over the years to describe a number of psychoacoustic phenomena (as reviewed briefly by Colburn and Durlach, 1978). In one version, the integration with respect to τ and the judgement mechanism of equation (2.4) were replaced with a mechanism that relates subjective lateral position to the centroid (along the τ-axis) of the time-averaged correlation function. Sayers (1964) argues that predictions of this modified model should parallel his findings for the lateralization of tones.

2.2.2. The position-variable model

The position-variable model described by Stern and Colburn (1978) was one of the first models to quantitatively describe the subjective phenomenon of lateralization in terms of a detailed description of the response of the auditory periphery to binaural stimuli. The models of the auditory-nerve response and the binaural timing display (described below) are based on the work of Colburn (1969, 1973, and 1977), and the auditory-nerve model is an extension of the work of Siebert (1968 and 1970).

A block diagram of the original position-variable model is shown in Figure 2-8. The peripheral processors are intended to model the transducing effects of the inner and middle ears. Each processor takes as its input the acoustical signal at one ear and produces a probabilistic description of the temporal firing patterns on all the fibers of the corresponding auditory-nerve. A key property of these processors is that they function, in part, like a bank of bandpass filters whose center frequencies span the range of frequencies useful in human hearing. This has the effect of spectrally decomposing the input signal to each ear. The resulting representation of neural activity is then examined by higher centers, indicated by the monaural processors and the timing displayer of Figure 2-8.

The specific properties of the monaural processors are not addressed by Stern and Colburn (1978), nor are they important to the current investigation (see Siebert, 1968 and 1970 for a detailed discussion). We are concerned only with the fact that the intensity function generator is able to estimate the value of the IID from the outputs of these

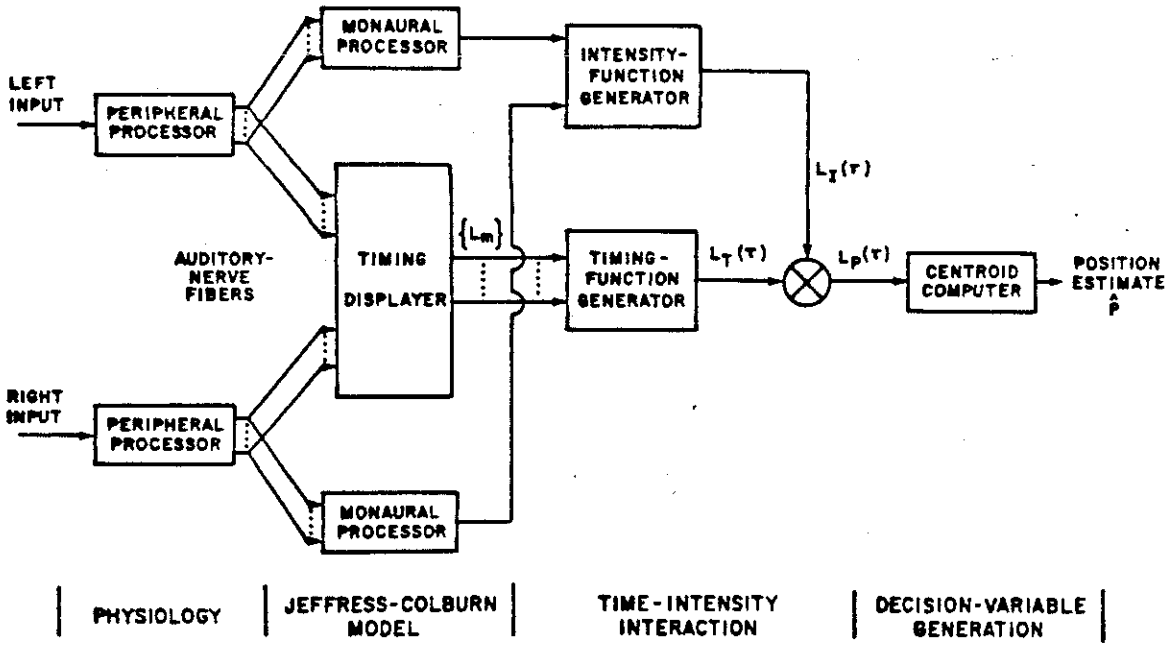

Figure 2-8: Schematic representation of the position-variable model.[Taken from Stern and Colburn, 1978.]

processors. This estimated IID is then used to determine the position along the τ-axis of the intensity function $L_I(\tau)$, which is assumed to be a deterministic Gaussian pulse (as in Figure 2-9(d,f)). This function provides the means by which timing and intensity information can be compared.

The timing displayer of Figure 2-8 processes the information carried by the two auditory nerves and produces a two-dimensional display which is related to the interaural crosscorrelation of the spectrally decomposed auditory signals. Jeffress (1948) has proposed a neural "coincidence counting" network which effectively accomplishes the crosscorrelation computation. The output of the timing displayer is processed by the timing-function generator in order to compute the function $L_T(\tau)$. An important step in the formation of this function is the weighting of the interaural correlation function $L_m(\tau)$ by the density function $p(\tau)$, which represents the assumed distribution of interaural delays associated with the binaural coincidence-counting units (see Figure 2-9(a,b,c)). The density function $p(\tau)$ is discussed extensively in Chapter 4, and in its original form is a double-sided exponential with a constant plateau for values of τ between ± 150 μs.

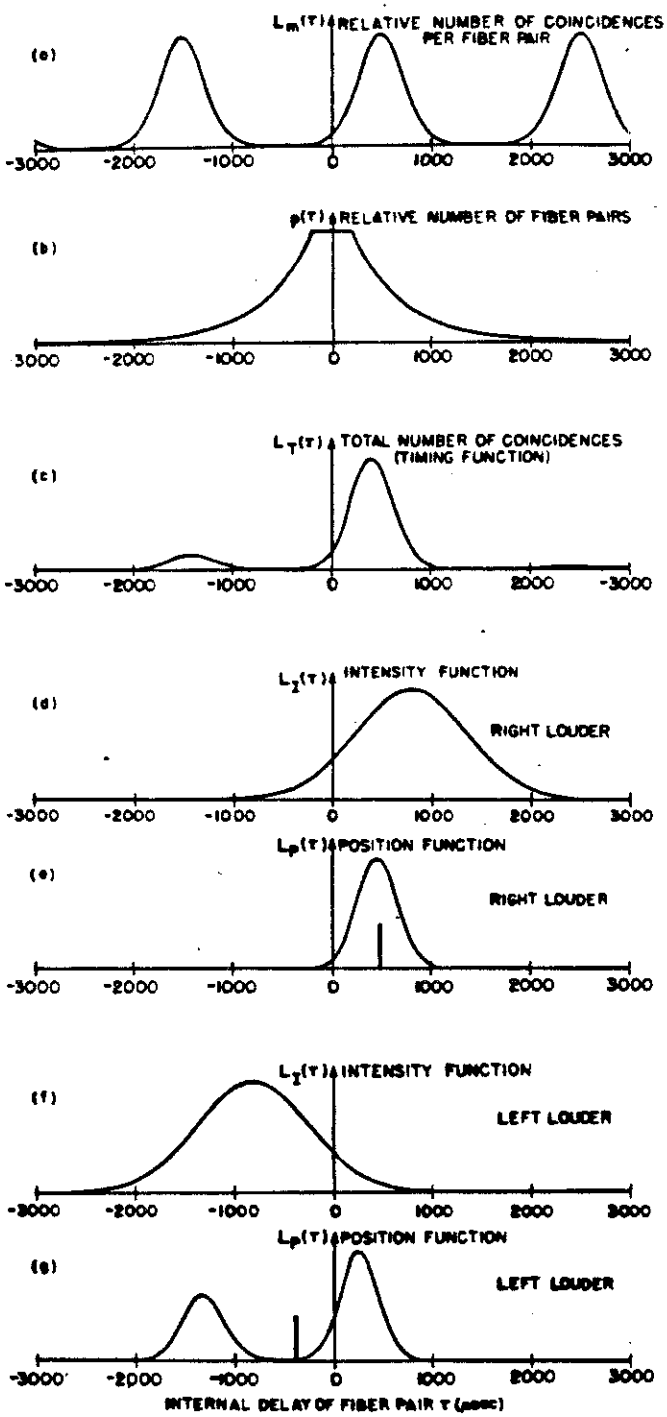

Figure 2-9: Generation of the position variable $\hat{P}$ from the timing and intensity functions for a 500-Hz tone with a 500 μs ITD and both (d,e) a reinforcing IID and (f,g) a cancelling IID. [From Stern and Colburn, 1978.]

Finally, the timing and intensity functions are multiplicatively combined and the centroid of the resulting function is computed with respect to τ (as in Figure 2-9(e,g)). The value of the centroid is assigned to the variable $\hat{P}$, and it is postulated to be monotonically related to the subjective lateral position of the binaural image.

The position-variable model has been used to describe nearly all available lateralization data for 500-Hz tones. In addition, it has been successful in describing binaural detection and discrimination phenomena for stimulus conditions in which lateral position cues are available.

2.2.3. The Lindemann model

Lindemann (1986) describes an "extended" running crosscorrelation model based on earlier work of Blauert and Cobben (1978). The peripheral transformations used in this model are very similar to those used in the postion-variable model. However, unlike the position-variable model, predictions for Lindemann's model are obtained from computer simulations of the model's operations on actual stimulus waveforms. After peripheral processing, the signals from each ear are used as inputs to multiply-tapped delay lines as shown in Figure 2-10(a). The outputs from the left and right delay lines are used to compute a running crosscorrelation function. The magnitudes of the interaural delays are limited by

$$|\tau| \leq \begin{cases} 1/(2\,fc), & \text{for } fc \leq 800 \text{ Hz} \\ 0.625 \text{ ms}, & \text{for } fc > 800 \text{ Hz} \end{cases} \tag{2.5}$$

where fc is the center frequency of the bandpass filter associated with a given delay line. This restriction imposes a rectangular window centered about τ equals zero on the crosscorrelation function. The effect of this window is similar to that of the distribution $p(\tau)$ in the position-variable model; it focuses the attention of the central processor, which examines the resulting crosscorrelation function, on those values of interaural delay near zero.

Lindemann has proposed a scheme of "contralateral inhibition" to describe both stationary effects of interaural intensity differences and dynamic effects of subsequent inputs to each channel of the model. This inhibition network is attractive because it provides a physiologically plausible realization of the "black-box" intensity function of the position-variable model. The ability of this network to describe dynamic phenomena, such as "the law of the first wavefront", is also quite impressive.

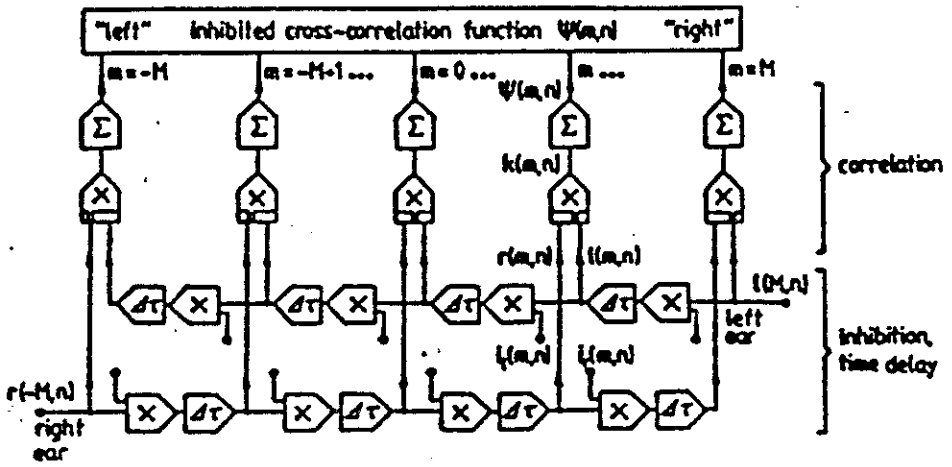

Figure 2-10: Block diagram of the crosscorrelation meachanism of Lindemann's model. [From Lindemann, 1986.]

For lateralization tasks, Lindemann uses the centroid of the running-crosscorrelation function as a measure of subjective lateral position. This model has been used to succesfully describe the lateralization of tones as function of ITD and IID. It has also been used to describe certain dynamic lateralization phenomena.

2.2.4. Similarities Among Models

The three models of binaural interaction described above are very similar in the operations they perform on a binaural stimulus and in the way lateral position judgements are formed from the interaural crosscorrelation display. The two most recent models include banks of peripheral bandpass filters and rectifiers before the crosscorrelation mechanism. Although the model described by Sayers and Cherry (1957) does not include bandpass filters, later versions of this model did adopt such structures (cf. Colburn and Durlach, 1978). In each model, the output of the interaural crosscorrelator is windowed by

some symmetric function of τ. Furthermore, each model uses the centroid, along the τ-axis, of the windowed correlation function as a measure of subjective lateral position. The particular mechanism by which interaural intensity differences are incorporated into the crosscorrelation display is different for each model, but their effect on the position estimate is similar for all three models.

2.2.5. Criteria for Evaluating Models

Over the years, a few basic criteria for the qualitative evaluation of models have evolved. While these criteria are almost self-evident, we feel it is worthwhile to mention them at this time. First, and perhaps most obvious, is the ability of a model to describe a wide body of psychoacoustic data (using as few free parameters as possible). In general, the more data a model is capable of describing, the better it is. One must be wary, however, of models which include a sufficiently large number of free parameters that they effectively perform a transformation of coordinates (from physical-parameter space to perceptual-response space). Second, models which contain physiologically identifiable components and intermediate signals are preferred to those which employ so called "black-box" elements. Since little physiological evidence is avalaible to characterize the most central mechanisms (*e.g.*, the binaural analyzer and the decision device), all current models include some "black-box" elements. Third, models which involve relatively simple signal processing techniques are usually considered more laudable than those which incorporate contrived or exotic schemes. This is due, in part, to the fact that simpler models generally provide added insight into the relationships between the physical parameters of the stimulus and the perceptual attributes which it elicits. The inclination to develop simpler models is not always warranted, however. For example, the relatively complex, yet accurate, description of peripheral transduction used by Colburn (1973) allowed him to draw several important conclusions regarding the relative significance of peripheral and central mechanisms in limiting performance in interaural time and intensity discrimination experiments.

2.2.6. Summary and Conclusions

In light of the measures of merit described above, we see that the position-variable model is generally considered to be attractive. The PV model includes a fairly accurate model of the auditory periphery, and the assumptions regarding the structure of the binaural timing displayer are physiologically plausible. In fact, units that behave similarly to

the hypothesized coincidence counters have been observed in the physiology (cf. Erulkar, 1972). Furthermore, the formation of the position variable can be accomplished by very simple operations. Most importantly, the model is able to describe nearly all available lateralization data at 500 Hz, including those involving interactions of ITD and IID.

As we have seen, there are many similarities between the position-variable model and other crosscorrelation-based models. One similarity that is particularly noteworthy is that all three models considered have used a centroid-based position estimate to describe subjective lateral position phenomena. Furthermore, each model includes some "window" function, similar in effect to $p(\tau)$, that focuses the attention of the centroid computer on values of interaural delay near zero. In the current work, we address the question of whether or not a slightly modified version of the PV model is capable of describing the lateralization data for tones as discussed in Section 2.1.1.1. Thus, the findings of our study should impact not only on future applications of the PV model, but also on the work conducted using a number of other crosscorrelation-based models of subjective lateral position.

Chapter 3
The Modified Position-Variable Model

In this chapter, we describe the assumptions and bases for the model of binaural interaction used in our research. A block diagram of the model is shown in Figure 3-1 and can be divided into three major components: (1) the peripheral mechanisms which describe the transduction of acoustical waveforms to activity on the auditory-nerve fibers, (2) the more central processes which compare the activity of the fibers from the two ears and produce a display of interaural crosscorrelation, and (3) the decision mechanism which uses the information encoded in the crosscorrelation display (and the two monaural variables) to yield a response for the given task. Each of the first three sections of the chapter is devoted to describing one of these model components, highlighting the modifications of the original PV model that were employed in the present calculations. The last section summarizes the differences between the new and original models.

3.1. Auditory-Nerve Model

In this section, we describe the peripheral processors of the model. We begin in the paragraph below with a few introductory comments and then proceed in the first subsection by listing the assumptions which define these processors. The second subsection discusses the functional implications of these assumptions, and the final subsection describes the differences between the original and new versions of the peripheral processors.

The model of peripheral transduction characterizes the neural activity of each fiber of the two auditory nerves in response to a given stimulus. It takes as its input the acoustical waveforms at each ear and produces as its ouput a probabilistic description of the temporal firing patterns on each auditory-nerve fiber. The characterization defined below

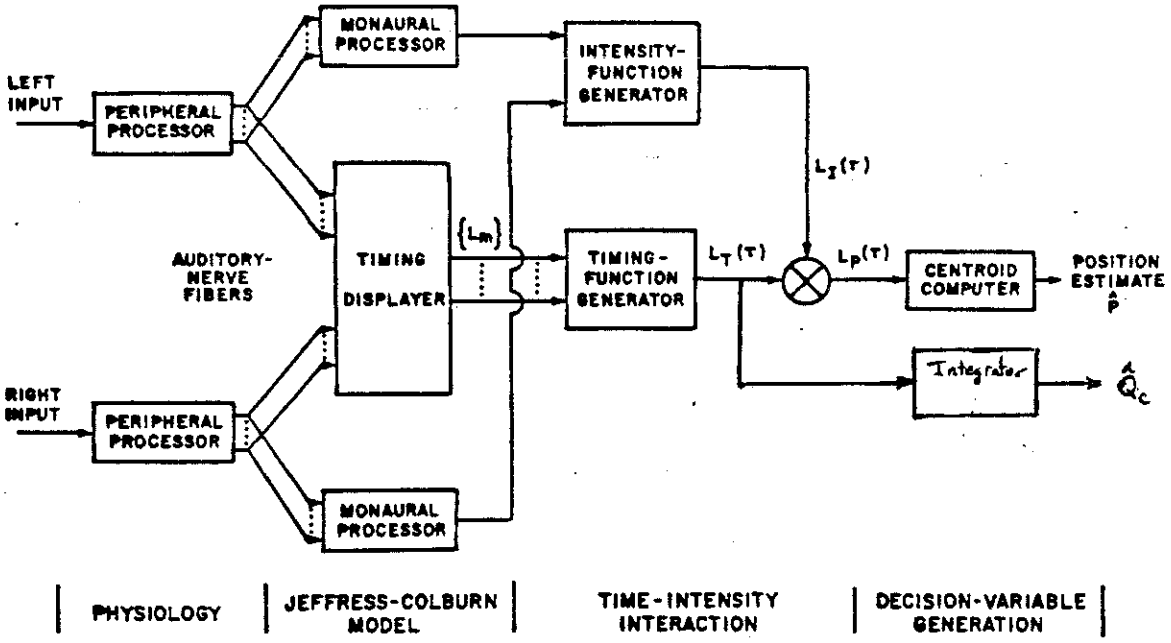

Figure 3-1: Block diagram of the modified position-variable model used for the current work.

is a slightly modified version of the one specified by Colburn (1973, 1977a,b), and it is based on the electrophysiological data obtained from cat by Kiang and his associates (Kiang *et al.*, 1965; Kiang, 1968) and Johnson (1974). Since such data are not available for humans we assume that the cat-based model applies. While it is generally accepted that this is a reasonable approach to the problem, it does ignore certain known differences between the human and cat auditory systems and at times it is instructive to consider qualitative adjustments to the model based on such differences.

3.1.1. Defining Assumptions

Below we list the set of assumptions which define the auditory-nerve model. Assumptions **AN1-AN4** and **AN6** are taken directly from Colburn (1973), and assumption **AN5** reflects the use of a discontinuous automatic gain control element as discussed by Colburn (1977a). The last assumption, **AN7**, includes **all** modifications introduced for this research.

(AN1) Firing patterns of individual auditory-nerve fibers are sample functions from stochastic point processes.

(AN2) The processes at the two ears are statistically identical and conditionally statistically independent given the stimulus waveforms.

(AN3) The processes corresponding to different individual fibers are conditionally statistically independent when the stimulus waveforms are given.

(AN4) The processes describing firing patterns from individual fibers are nonhomogeneous Poisson processes (defined in Parzen, 1962).

(AN5) Each auditory nerve is composed of 3×10^4 fibers, each characterized by a pair of numbers, a *characteristic frequency* fc_m and a *sensitivity constant* K_m. Characteristic frequencies are spaced uniformly on a logarithmic scale between 20 Hz and 20 kHz and sensitivity constants of fibers with characteristic frequency fc_m are spaced uniformly on a logarithmic scale over a range of 40 dB so that the curve described by the minimum values of K_m as a function of fc_m has the same shape as the threshold-of-hearing curve for tones (Kiang *et al.*, 1965, p. 89).[4]

(AN6) With each characteristic frequency fc_m, there is associated a filter with impulse response $h_m(t)$. We specify $h_m(t)$ through $H_m(f)$, the magnitude of the corresponding frequency response, and $\theta_m(f)$, the minimum-phase characteristic consistent with $H_m(f)$. $H_m(f)$ is given by

$$H_m(f) = \begin{cases} (f/fc_m)^{\alpha(fc_m)}, & \text{for } 0 \le f \le fc_m \\ (fc_m/f)^{2\alpha(fc_m)}, & \text{for } f > fc_m \end{cases} \tag{3.1}$$

where $\alpha(fc)$ is specified by the equation

$$\alpha(fc) = \begin{cases} 4, & \text{for } 0 \le fc \le 800 \text{ Hz} \\ 4(fc/800), & \text{for } fc > 800 \text{ Hz} \end{cases} \tag{3.2}$$

(AN7) Intensity functions $r_m(t)$ for an arbitrary stimulus $x(t)$ are specified by the relations

---

[4] Model predictions were obtained using the piece-wise approximation to this curve given by

$$\text{THRESHOLD}(f) \begin{cases} = 4.5 + 44.846 \log(500/f), & \text{for } f < 500 \text{ Hz} \\ = 4.5 - 14.9847 \log(f/500), & \text{for } 500 \le f < 1000 \text{ Hz} \\ = 0, & \text{for } 1000 \le f < 2500 \text{ Hz} \\ = 28.7044 \log(f/2500), & \text{for } f \ge 2500 \text{ Hz} \end{cases}$$

where f is in Hz and THRESHOLD (f) is in dB SPL.

$$r_m(t) = \begin{cases} a_m \mathcal{R}_\nu[h_m(t) * x(t)] * g(t), & \text{for } \mathbf{MS}[h_m(t) * x(t)] > K_m \\ 50, & \text{otherwise} \end{cases} \tag{3.3}$$

where **MS** is a short-time mean-square operator, a_m is chosen so that the time-averaged rate $\overline{r_m}$ is 200 per second, $g(t)$ is the impulse response of a low-pass filter which has a frequency response with magnitude

$$\begin{aligned} G(f) &= 1, && \text{for } 0 \leq f \leq 1200 \text{ Hz} \\ &= \frac{1 - f/5600}{1 - 1.2/5.6}, && \text{for } 1200 < f \leq 5600 \text{ Hz} \\ &= 0, && \text{for } f > 5600 \text{Hz}, \end{aligned} \tag{3.4}$$

and $\mathcal{R}_\nu[z]$ is a ν^{th}-law half-wave rectifier defined by

$$\mathcal{R}_\nu[z] = \begin{cases} z^\nu, & \text{for } z \geq 0 \\ 0, & \text{for } z < 0 \end{cases} \tag{3.5}$$

The first assumption is a restatement of the widely accepted all-or-none description of neural firing. The stochastic nature of the process is in agreement with the observation that repeated presentations of a deterministic stimulus do not produce repetitions of the detailed firing pattern for a given fiber. The variation in the response is a direct result of "noise", presumably, in the transduction mechanism, and it is assumed that this is the only source of internal noise for our model. Although there are no directly applicable physiological data to support (or disprove) assumption **AN2**, it is intuitively appealing and can be inferred from the physiological evidence which supports assumption **AN3**. The remaining assumptions are included to describe specific physiological phenomena, and the reader is referred to Colburn (1973) for a more complete discussion. The differences between the original version of assumption **AN7** and the current one will be discussed in Section 3.1.3.

Although the model's description of neural firing statistics is specified in terms of an arbitrary stimulus, the model is best suited to describing responses to low- to moderate-intensity tones, noise, and combinations of the two. While the model does not describe certain phenomena associated with high-intensity tones or tone pairs (Johnson, 1974), these classes of stimuli are not important to the psychoacoustic studies considered in this research. A more serious inaccuracy of the model is its failure to describe the refractory period between neural firings; this problem stems from the Poisson process assumption,

AN4. While altering this assumption to include refractoriness would provide a better description of auditory-nerve data, it would also greatly complicate the statistical analysis (which takes advantage of certatin properties of Poisson point processes). Moreover, Colburn (1969, 1973) argues that the lack of refractoriness in the model does not strongly influence most predictions involving tones. Noise stimuli, on the other hand, have proven to be a problem for the original model in this respect, however the new model provides an alternate solution that allows us to keep the Poisson process assumption without sacrificing the accuracy of the model predicitions. (See the discussion on page 34.)

3.1.2. Functional Interpretation

We now consider the functional effects of the assumptions defined above. It is helpful to conceptualize each peripheral processor as consisting of a bank of similarly specified units operating on the input waveform in parallel, each unit producing its own output. As assumption **AN5** indicates, one major difference between each unit is its frequency selectivity. Thus, the peripheral processor operates in a manner similar to a bank of bandpass filters (with different center frequencies) in order to spectrally decompose the acoustical signal. The additional processing performed along each channel is nearly identical for all channels and is described below.

We now consider the processes described by **AN6** and **AN7** which relate the stimulus to the instantaneous firing rate of a particular fiber, as illustrated in Figure 3-2. The first component of this model is the bandpass filter $H_m(f)$ which reflects the frequency selectivity of the basilar membrane and has a maximum value at the characteristic frequency fc_m of the fiber. The amount of stimulus power which passes through this filter determines whether or not the fiber is "stimulated". The automatic gain control (AGC) element measures this power and adjusts its own output so that fibers which are *stimulated*, or *active*, exhibit a time-averaged rate of firing of 200/s while fibers which are *unstimulated*, or spontaneously firing, exhibit a constant firing rate of 50/s, independently of the stimulus waveform. The next functional block in Figure 3-2 is the rectifier which is necessary to insure a nonnegative intensity function for the Poisson process generator. Finally, the lowpass filter describes the observed loss of synchrony between the detailed timing structure of the stimulus and the firing patterns of auditory-nerve fibers for high frequency stimuli. Thus, for low frequency stimuli, the instantaneous rate of firing is phase-locked to the stimulus waveform, but for high frequency stimuli the fiber fires at a

rate which is independent of the detailed time structure of the stimulus (though it may depend on the low-frequency envelope of the stimulus if one is present).

3.1.3. Changes from the Original Model

In this section, we describe the differences between the original version of the auditory-nerve model and the version defined above. We also discuss the motivations for these changes and the procedure used to specify the new lowpass filter function $G(f)$.

3.1.3.1. The rectifier

As Figure 3-3 shows, the new model makes use of a ν^{th}-law half-wave device in place of an exponential rectifier. This was motivated by the comments of Colburn (1969, 1977b) regarding the inaccurate statistical representation of the original model in response to noise stimuli. The original model predicts a much greater short-term firing rate for noise stimuli than is observed in the physiology. This inaccurate description of the nerve data is caused by two properties of the original model. First, it does not account for the observed refractory period between neural firings; and second, its description of the instantaneous rate of firing is dominated by very large, very improbable values of the input noise process. While one remedy to this problem would be to incorporate refractoriness into the the model, this greatly complicates the statistical analysis. Another solution, is to specify an intensity function which does not weigh large input values as heavily as the original model does. This can be accomplished by replacing the exponential rectifier with another rectifier that exhibits less expansive transfer characteristics, especially for large input values. The use of the half-wave ν^{th}-law rectifier (with small ν) provides such a solution without sacrificing much analytical tractability of the model. While we recognize that the new model still does not describe the refractory nature of neural firing, it no longer exhibits the statistical inaccuracies discussed by Colburn.

Although the arguments reviewed above suggest that the exponential rectifier should be abandoned, we present psychoacoustical evidence in Chapter 5 that supports the use of just such a rectifier. For this reason, we presently do not ignore the possibility that an exponential rectifier should be incorporated in the model.

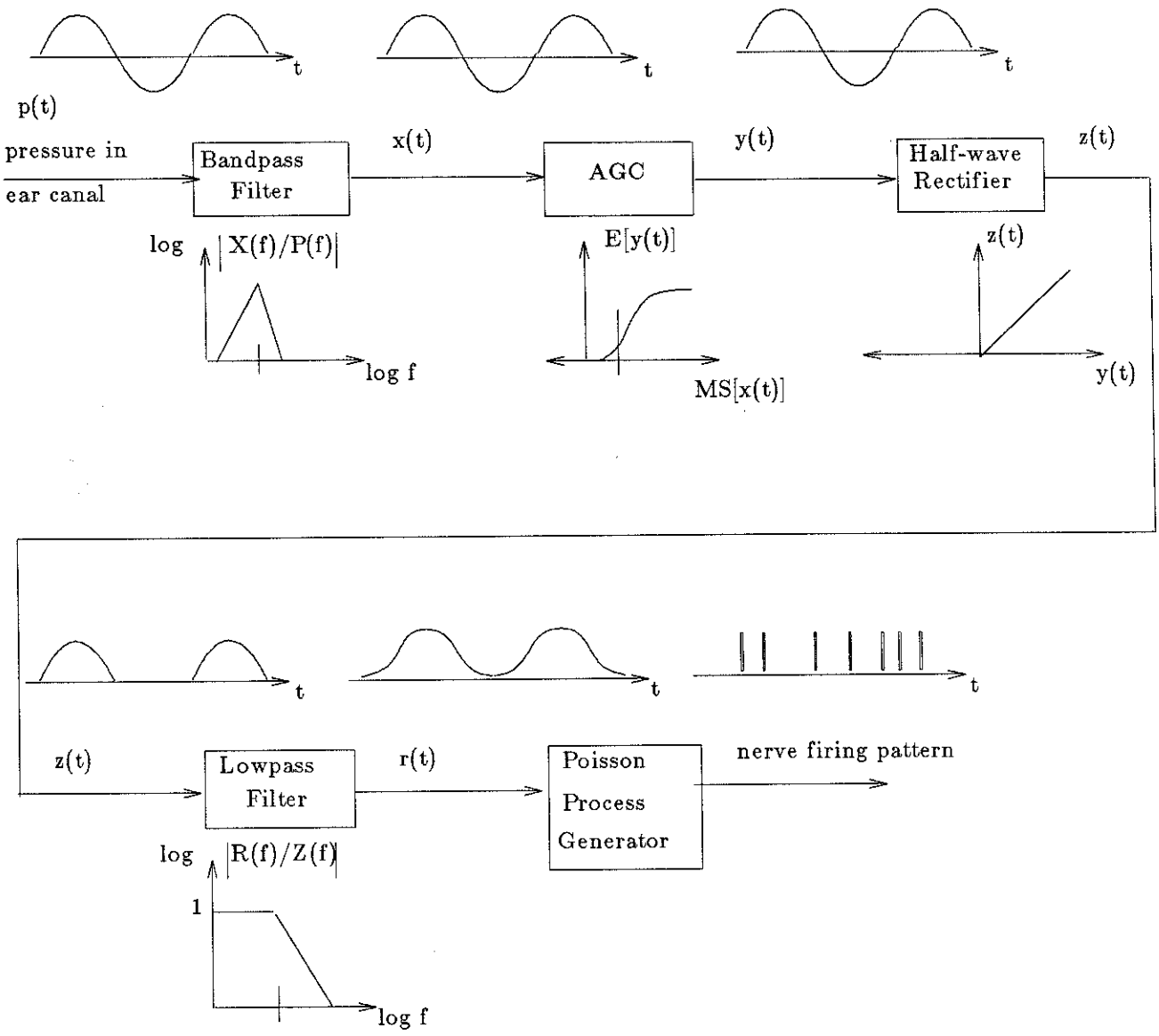

Figure 3-2: Functional representation of the new auditory-nerve model used for this work.

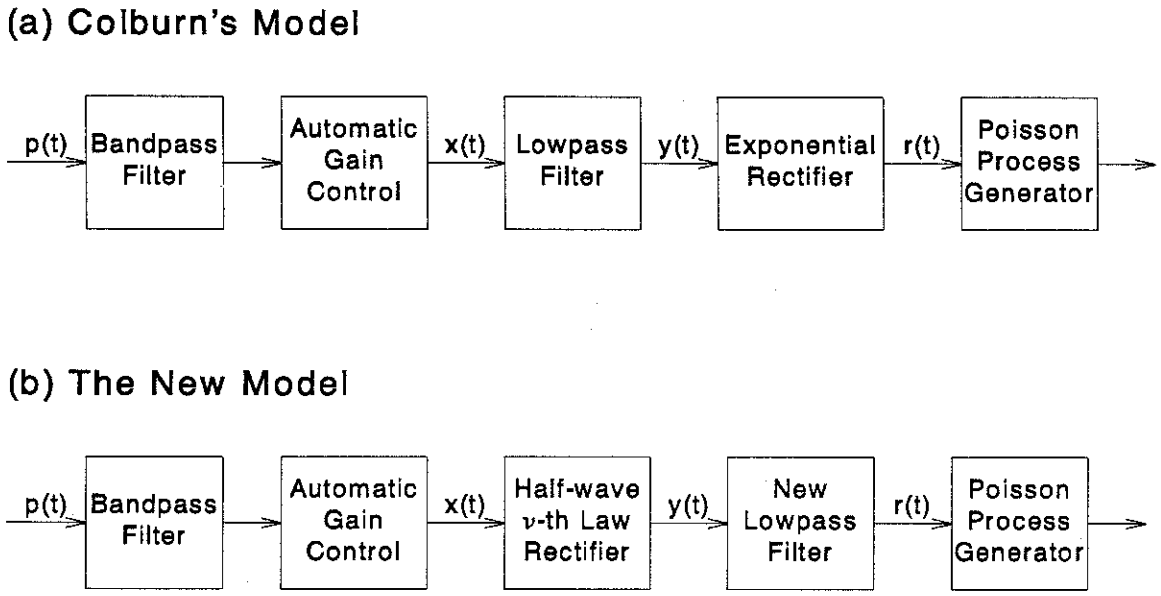

Figure 3-3: Block diagrams of (a) the original auditory-nerve model as specified by Colburn (1973) and (b) the new version of the auditory-nerve model used for this work.

3.1.3.2. The lowpass filter

Referring to Figure 3-3, note that the relative positions of the lowpass filter and the rectifier have been changed in the new version of the model. The placement of the lowpass element after the rectifier was necessary to describe the saturation effects of the automatic gain control (AGC) element without greatly complicating the analysis of the model given stochastic stimuli. This configuration of rectifier followed by lowpass filter is used in several other models of auditory physiology and perception (*e.g.*, (Duifhuis, 1973; Blauert and Cobben, 1978; Lindemann, 1986) and provides a good description of the auditory-nerve data. In addition, the new configuration is preferred over the old because it has envelope detecting properties, which are expected to help explain the lateralization of certain complex, high-frequency stimuli (Henning, 1974; Nuetzel and Hafter, 1976; McFadden and Pasanen, 1976; Henning, 1980; Nuetzel and Hafter, 1981). Personal correspondence with Colburn has revealed that his original decision to position the lowpass filter before the rectifier was somewhat arbitrary.

A commonly used characterization of auditory-nerve firing patterns is the post-stimulus-time (PST) histogram which provides a measure of the instantenous rate of firing for a given neural unit. The expected values of the model's intensity functions and the expected values of the PST histograms should be proportional, and experimentally determined PST responses to tones were originally used to specify the rate functions of the auditory-nerve model (Colburn, 1973). One important characteristic of these histograms is their loss of synchronization to the input stimulus as the tone frequency is increased. In the model, this loss of synchrony is characterized by the lowpass filter, and our expression for $G(f)$ in Equation (3.4) was chosen to describe the physiological data obtained by Johnson (1974) (which were unavailable for use by Colburn when the original model was specified). Johnson defines the *synchronization index* S_f to be the magnitude of the first (or fundamental) coefficient of the Fourier series for a given PST histogram normalized by the average firing rate for the same histogram. The data reported by Johnson describe the dependence of this index on the frequency of the tone, and they are shown in Figure 3-4. The trend of the data exhibited in this figure can be approximated by the following expression,

$$S_f(f_0) \approx 0.85 \times G(f_0)$$

where f_0 is the frequency of the tone and $G(f)$ is given by Equation (3.4). This index can be computed in a similar manner for an intensity function $r_m(t)$, and the shape of $G(f)$ was chosen so that the predicted values of $S_f(f_0)$ are consistent with the data collected by Johnson (1974).

As the above expression indicates, the frequency response of the lowpass filter of the new auditory-nerve model is directly proportional to the frequency dependence of the synchronization index. This is because the lowpass element is assumed to be located after the rectifying nonlinearity in the input transformation. It is worth noting that this choice of $G(f)$ can be used to accurately characterize the loss of synchrony at frequencies above 1200 Hz for any auditory-nerve model in which the lowpass filter follows the rectifier (assuming that no other elements alter the spectral content of the predicted intensity function). However, the ability of the model to describe the synchronization index data at low frequencies is related to the type of rectifier used. For the ν^{th}-law half-wave device defined by Equation (3.5) the values $\nu = 1, 2, 3$ yield predictions for the low-frequency synchronization indices of 0.79, 0.85, and 0.88, respectively. Each of these choices for ν is considered to be acceptable since the actual

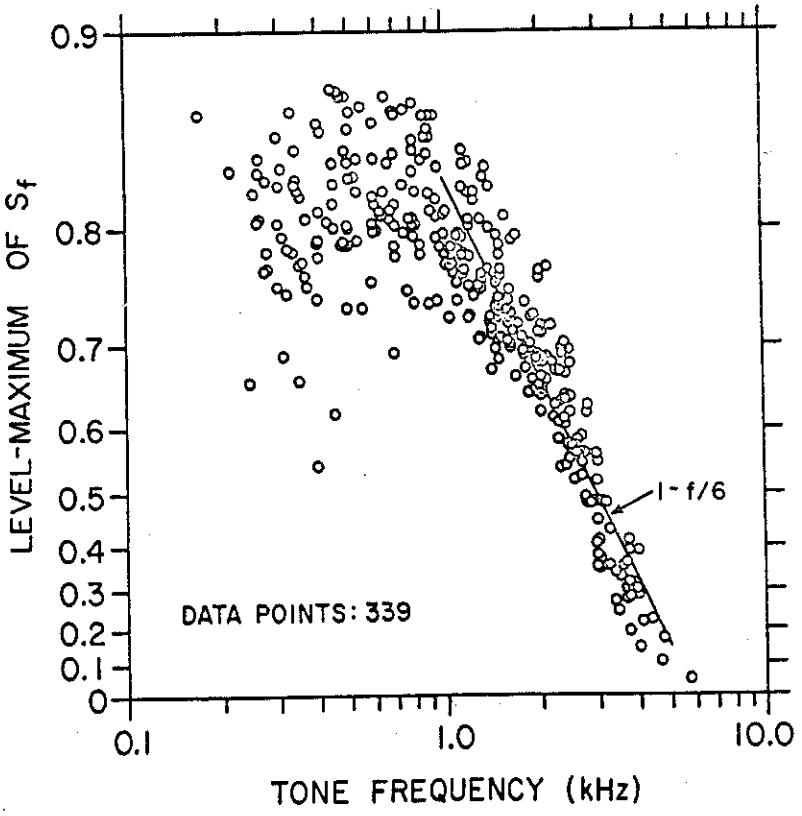

Figure 3-4: Synchronization indices estimated from the responses of 233 neural units to stimulation by a single tone are shown as a function of tone frequency. [From Johnson, 1974.]

data, shown in Figure 3-4, exhibit a fair amount of variability over the range from 0.78 to 0.88. Although the synchronization indices predicted by slightly higher-order nonlinearities (*e.g.*, ν equal to 4 or 5) do not lie strictly within this range, these rectifiers still provide an adequate description of the data.

3.2. Binaural Displayer

In this section, we describe the "binaural timing displayer" mechanism depicted in Figure 3-5. We begin with a brief introduction and then proceed in the first subsection by listing the assumptions which define the timing displayer. The second subsection discusses the functional implications of these assumptions, and the third subsection describes the differences between the original and new versions. The final subsection is devoted to providing a statistical analysis of the ouputs of the timing displayer.

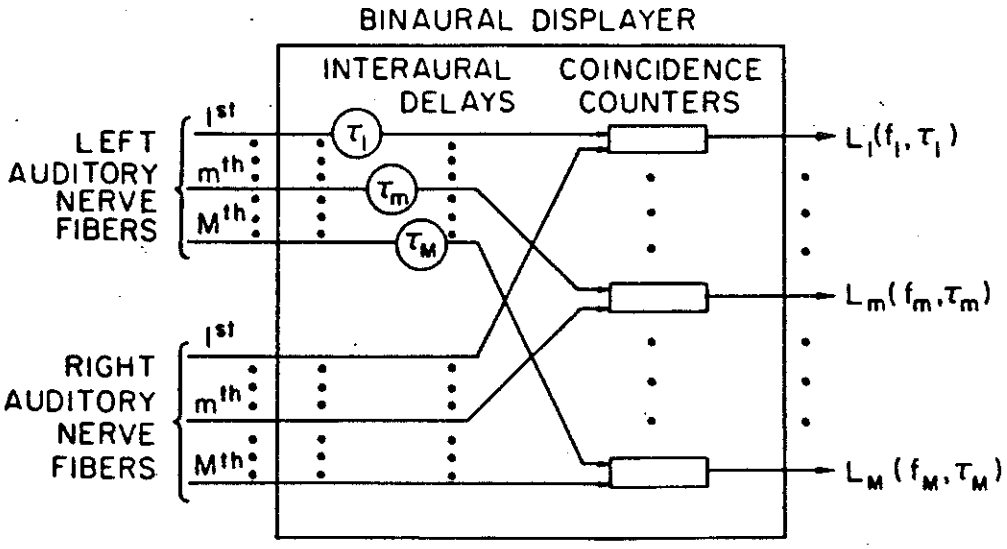

Figure 3-5: Diagram of the binaural timing displayer. [From Colburn, 1977a.]

The binaural timing displayer describes one possible mechanism by which the auditory system may process the firing patterns of the auditory-nerve fibers in order to extract information regarding various interaural stimulus parameters. The displayer is assumed to be located beyond the auditory nerve in the sensory pathways, and it is also assumed to act in a manner independent of stimulus type and psychophysical task. Thus, this element is considered to be a "hard-wired" physiological structure, and it is assumed that only the interpretation of its output can be altered by higher-level, cognitive processes. While the motivation for the formulation described below comes from the psychoacoustical modelling work of Colburn (1973, 1977a), there is some physiological evidence available which supports this formulation (as reviewed by Erulkar, 1972) , and it is generally considered to be physiologically plausible. It is worth noting that since the bases for the binaural displayer are psychophysical, and not physiological, the various parameters which help define this component of the model may be adjusted to better describe psychoacoustic data. This displayer mechanism has proven useful in describing a wide variety of binaural phenomena (*e.g.*, (Colburn, 1973; Colburn, 1977a; Stern and Colburn, 1978; Bachorski, 1983; Stern and Colburn, 1985).

3.2.1. Defining Assumptions

The assumptions that define the current implementation of the binaural timing displayer have been adapted from Colburn (1977a), and they are listed below. A more intuitive description of these assumptions is provided in Section 3.2.2.

(BD1) The binaural displayer of Figure 3-5 is specified by the equation

$$L_m = \sum_{i=1}^{N_{Lm}} \sum_{j=1}^{N_{Rm}} f(t_i^{Lm} - t_j^{Rm} - \tau_m) \text{ for } m = 1 \text{ to } M \tag{3.6}$$

where

N_{Lm} is the number of firings of the m^{th} fiber of the left auditory nerve during the stimulus presentation,

t_i^{Lm} is the time (in seconds) of the i^{th} firing on the m^{th} fiber of the left auditory nerve during the stimulus presentation,

N_{Rm} and t_j^{Rm} are defined similarly,

$f(z)$ is a rectangular-shaped function centered about 0 with $f(0) = 1$ and an area of 10^{-5} (seconds),

τ_m is a fixed, internal interaural delay (expressed in seconds) associated with the m^{th} fiber pair,

and the m^{th} fibers from the left and right auditory nerves have the same characteristic frequency fc_m.

(BD2) The values of τ_m are distributed over all the fiber pairs independently of all parameters other than characteristic frequency, and their distribution is specified by the conditional density function $p(\tau \mid fc)$, which is an even, pulse-shaped[5] function of τ for a given value of fc.

(BD3) The appropriate expression for the variance of the displayer output L_m is assumed to be

$$Var[L_m] = E\{Var[L_m \mid x(t)]\}, \tag{3.7}$$

where $Var[L_m \mid x(t)]$ denotes the conditional variance of L_m given the stimulus waveforms $x_R(t)$ and $x_L(t)$. (The remaining expectation is computed over the sample functions of $x(t)$.)

---

[5] Specifically, $p(\tau \mid fc)$ is a nonincreasing function of $|\tau|$ and approaches zero as $|\tau| \rightarrow \infty$ for $20 \leq fc \leq 20000$ Hz.

The assumption in **BD1** that restricts comparisons to coincidence counts between fibers of the same characteristic frequency was motivated by the theoretical work of Colburn (1973) in which he considered performance in interaural time discrimination experiments. The further assumptions (in **BD1**) regarding the size and shape of the coincidence window $f(z)$ have been made to simplify the analysis of the displayer outputs (in Section 3.2.4).

The portions of assumption **BD2** regarding the shape of the distribution of internal delay parameters are consistent with the lateral symmetry of the human body and the fact that most naturally occuring ITDs are less than 660 μs (as was noted on page 12).

Assumption **BD3** is included mainly to facilitate variance computations. The statistically correct expression for the variance of L_m is given by

$$Var\{L_m\} = V_{AN} + V_{STIM},$$

where $V_{AN} = E\{Var[L_m \mid x(t)]\}$ is the variance due to the auditory-nerve process and $V_{STIM} = Var\{E[L_m \mid x(t)]\}$ is the variance due to the stimulus. Colburn (1969) argues that for an accurate characterization of auditory-nerve activity, V_{STIM} is negligible compared to V_{AN} and thus the above assumption is justified. However, psychoacoustic experiments were later conducted to test this hypothesis, and they revealed that the variance due the stimulus does play some role in determining overall performance in binaural detection experiments (Siegel and Colburn, 1983). It is not clear from their study how V_{STIM} affects relative detection thresholds (such as those considered here), and it is conceivable that this variance causes approximately the same degree of decreased performance in both $N_\pi S_0$ and $N_0 S_\pi$ experiments. In any event, it is very difficult to correctly compute the variance of L_m and for the present time we have resigned ourselves to using Equation (3.7). Consequently, predictions that do involve computing the variance of L_m must be regarded as approximations.

As a final comment we note that the realization of the binaural displayer considered above is technically only applicable to stimuli which are statistically stationary (ignoring brief transients at the beginning and end of their presentation) since it assumes an ideal integrator in the coincidence counting mechanism rather than the commonly assumed "leaky" integrator (Sayers and Cherry, 1957; Blauert and Cobben, 1978; Bachorski, 1983; Lindemann, 1986). This issue is not important for the current study since the lateralization and detection experiments considered in this paper do employ stationary stimluli.

3.2.2. Functional Interpretation

The inputs to the displayer shown in Figure 3-5 are pulse trains representing firing patterns of auditory-nerve fibers. Each fiber's pulse train is assumed to be delayed by a fixed amount before being compared with the pulse train of another fiber (from the opposite ear). A coincidence is registered each time pulses from the two delayed trains occur "almost simultaneously" (as in Eq. (3.6)), and a count L_m of the number of coincidences for the m^{th} fiber pair is maintained. The interaural delay introduced between the two fibers being compared is denoted by τ_m, and it is assumed that comparisons are only made between fibers with the same characteristic frequencies. Thus, the output of the binaural displayer can be interpreted as a two-dimensional matrix $L_m(\tau_m, fc_m)$ with the first coordinate representing the internal interaural delay and the second representing the characteristic frequency of the input fiber pair. It is instructive to consider the output of the displayer in response to a tonal stimulus with frequency f_0 and ITD τ_s. Fiber pairs with characteristic frequencies nearer to the tone frequency are stimulated to higher firing rates, and thus the outputs of the displayer will reflect a larger number of coincidences for these values of fc_m. Similarly, values of τ_m near τ_s (and $\tau_s \pm \frac{k}{f_0}$) will produce larger displayer outputs than other values of τ_m provided the firing rates of the fibers are phase locked to the detailed time structure of the stimulus. Thus, the displayer output has properties similar to (and can be interpreted as) a display of interaural crosscorrelation of the peripherally transformed stimuli. For fiber pairs in which one or both fibers are not synchronized to the stimulus waveform, the output of the display will be a constant that is independent of internal interaural delay. If both fibers are active (or stimulated), we refer to the fiber pair as *doubly active.*

3.2.3. Changes from the Original Model

The only modifications introduced in the binaural displayer for this research are related to the distribution of the internal delays. The distribution specified by Colburn (1977a) was chosen to describe data for the detection of tones in noise. His assumption that the density function is independent of the characteristic frequency of the fiber pair was made because there were insufficient psychophysical data to meaningfully constrain a frequency-dependent form of the function. It appears that recently acquired data on the lateralization of tones as a function of frequency (Schiano, Trahiotis, and Bernstein, 1986) may lend further insight into the shape of the distribution and its dependence on fc.

The specification of the distribution constitutes a major part of our research, and we will consider this issue in detail in Chapter 4. Presently, we do not assume any particular form of $p(\tau \mid fc)$, and we merely state that the assumed function should yield predictions for lateralization and detection experiments that are consistent with the two bodies of data reviewed in Sections 2.1.1.1 and 2.1.2.

3.2.4. Statistical Analysis

In order to determine the conditional variance and conditional mean of L_m given the stimulus waveforms, we begin by noting that $L_m(t)$, expressed here as a function of time since the stimulus onset, exhibits the following three characteristics:

1. It takes on integer values at discrete times since the coincidence window is given by

$$f(z) = \begin{cases} 1, & \text{for } |z| \leq T_w/2 \\ 0, & \text{otherwise} \end{cases}$$

   where T_w is the area of the window (*i.e.*, 10 μs).

2. It has independent increments since both $N_{Lm}(t)$ and $N_{Rm}(t)$ have independent increments.

3. The probability of having more than one coincidence within any time period of length T_w is negligible for the classes of stimuli considered, as ensured by our narrow choice for $f(z)$.

These are sufficient conditions to qualify $L_m(t)$ as a nonhomogeneous Poisson counting process, and we can estimate its intensity function as follows,

$$r(t) = \lim_{h \to 0} \frac{1 - \text{Pr}\,\{L_m(t+h) - L_m(t) = 0\}}{h} \approx T_w\, r_{Lm}(t - \tau_m)\, r_{Rm}(t) \tag{3.8}$$

and thus,

$$\text{E}\,[\,L_m(t) \mid x(t)] = \text{Var}\,\{\,L_m(t) \mid x(t)\,\} \approx T_w \int_0^t r_{Lm}(z - \tau_m)\, r_{Rm}(z)\, dz$$

If we assume the stimulus to be of duration T_S and take the final expectation over the binaural inputs $x(t)$, we obtain

$$\begin{aligned} \text{E}\,[\,L_m\,] &= \text{E}\,[\,\text{Var}\,\{\,L_m \mid x(t)\,\}] \approx T_w \int_0^{T_S} \text{E}\,[\,r_{Lm}(z - \tau_m)\, r_{Rm}(z)\,]\, dz \\ &= T_w\, T_S\, R_{RLm}(\tau_m) \end{aligned} \tag{3.9}$$

where $R_{RLm}(\tau_m)$ is the time-averaged (or ensemble-averaged) interaural crosscorrelation function of the peripherally transformed deterministic (or stochastic) stimulus. The expectations in the above expressions are taken with respect to the neural point processes and the binaural input $x(t)$. (For the present discussion, we assume that the characterizing parameters fc_m, τ_m, etc. have fixed, known values.)

[Note that the above expressions for the mean and variance of L_m should actually be considered accurate only to within some arbitrary constant. This is due to the dependence of these expressions on the specific assumptions regarding the shape of the coincidence window $f(z)$ and the properties of the "leaky-integrator" or coincidence counter. While this degree of inaccuracy does not affect most calculations of interest, it does prevent us from generating meaningful predictions for absolute detection thresholds. This issue will be discussed in further detail in Section 3.3.3.2.]

If either fiber from the m^{th} pair is inactive (*i.e.*, spontaneously firing), then the interaural correlation function of Equation (3.9) is simply equal to the product of the two mean firing rates. That is to say, $R_{RLm}(\tau_m) = (50)(50)$ when both fibers are inactive and $R_{RLm}(\tau_m) = (200)(50)$ when one of the fibers is active and the other is not.

If the fiber pair is doubly active, then the form of $R_{RLm}(\tau_m)$ will depend on the type of stimulus employed, as well as on the characteristic frequency fc_m and interaural delay τ_m. In Appendix B, we derive expressions for this correlation function when the binaural input is a pure tone, a Gaussian noise, and an additive combination of the two. One important result for the case of a pure binaural tone is that the interaural correlation function, for doubly active fiber pairs, is (1) independent of characteristic frequency and (2) periodic with respect to the delay argument τ_m (having period equal to that of the tone). These properties will be exploited in the next section in order to derive a frequency-domain expression for predictions of subjective lateral position.

3.3. Decision-Making Device

In this section, we describe the decision-making device that is assumed to be used for the two psychoacoustical tasks of lateralization and detection. We begin in the paragraph below with a brief introduction and then proceed in the first subsection by listing the assumptions that define the decision device. The second subsection discusses the functional implications of these assumptions, and the final subsection is devoted to

providing a statistical analysis of the decision variables and a means of computing a performance index for the detection of tones in noise.

As is indicated by Figure 3-1, the decision mechanism makes use of the information provided by the binaural displayer and two monaural processors in order to produce a response for a given perceptual task. In the current work, we are not immediately interested in the effects of interaural intensity differences, and we simply ignore the possible role that these cues may play in binaural interaction. There is no psychoacoustical justification for this assumption, and it is proposed only to facilitate the specification of $p(\tau|fc)$. While we believe that the time-intensity interactions in future versions of the postition-variable model will be similar in form to those in the original model, the exact assumptions regarding these interactions will undoubtedly be related to our choice of $p(\tau|fc)$. We also believe that the findings of this study will be directly applicable to future investigations that focus on specifying the mechanisms which process interaural intensity differences.

3.3.1. Defining Assumptions

Our assumptions regarding the decision making mechanisms involved in lateralization and detection experiments are specified below. These assumptions are based on the work of Stern and Colburn (Stern, 1976; Colburn, 1977b; Stern and Colburn, 1978).

(DM1) Two internal variables $\hat{P}$ and $\hat{Q}$ are combined in an optimal fashion to arrive at a final decision for a given task. The variable $\hat{P}$ is postulated to be monotonically related to subjective lateral position, and it is the fundamental variable used in making lateral position judgements.

(DM2) The internal decision variable $\hat{P}$ is specified by

$$\hat{P} = \frac{\sum_{m \in \mathbf{Z}_{task}} \tau_m L_m}{\sum_{m \in \mathbf{Z}_{task}} L_m} \tag{3.10}$$

where $\mathbf{Z}_{task}$ describes the set of displayer outputs (specified below) which are considered by the decision mechanism. The decision variable $\hat{Q}$ may be specified in one of the two following ways

$$\hat{Q}_o = \sum_{m \in \mathbf{Z}_{task}} c_m L_m \tag{3.11}$$

or

$$\hat{Q}_c = C \sum_{m \in \mathbf{Z}_{task}} L_m \tag{3.12}$$

In the first case, each coefficient c_m is chosen to yield optimal performance in the detection task, and its value may depend on the characteristics of the m^{th} fiber pair and the stimulus. The latter specification, $\hat{Q}_c$, describes a simpler decision statistic that uses a constant coefficient C.[6]

(DM3) The set of coincidence counts deemed significant in performing a given task is restricted by the characteristic frequency of the fiber pair in the following manner

$$\mathbf{Z}_{task} = \{ m \mid m = 1 \text{ to } M \text{ and } fc_m \in \mathbf{R}_{task} \},$$

where M is the number of fibers in each auditory nerve (3×10^4) and $\mathbf{R}_{task}$ defines a range of characteristic frequencies which is chosen optimally for each task.

The formulation given above may be used to predict behavior in a number of psychoacoustic tasks. In this study we focus on two tasks, lateralization and detection. When considering a lateralization experiment, the concept of "optimal performance" is somewhat ambiguous. Clearly there is no single, correct answer to the question "Where do you hear the sound?", and thus, it is impossible to define the optimal means of responding. In order to resolve this problem, we assume (in **DM1**) that the variable $\hat{P}$ is fundamental to the formation of position estimates (and that $\hat{Q}$ is not important in this process). The relative significance of the two decision variables in a detection experiment depends on the type of stimuli involved. If the perceived position of the target stimulus is different from that of the masker, $\hat{P}$ may be useful. However, for the stimuli considered in this investigation, such lateral position cues are not present, and the listener must concentrate on small changes in the "quality" of the binaural image. The variable $\hat{Q}$ is presumably a good indication of these changes. Throughout this report, we use the decision variable $\hat{P}$ exclusively for lateralization predictions, and we use the decision variable $\hat{Q}$ exclusively for detection predictions.

---

[6] The subscript "o" in $\hat{Q}_o$ refers to optimal coefficients, and the subscript "c" in $\hat{Q}_c$ refers to constant coefficients.

Before proceeding further, it is helpful to give our two decision variables names. For obvious reasons, $\hat{P}$ has been dubbed the *position variable*. Because $\hat{Q}$ provides a measure of the degree of correlation between the firing patterns on the two auditory nerves, we refer to $\hat{Q}$ as the *correlation variable*.

In **DM1**, we assume that $\hat{P}$ is monotonically related to subjective lateral position, but we do not specify what the exact relationship is. Based on the definition of $\hat{P}$ and our understanding of how ITDs generally influence position judgements, it is reasonable to assume that this monotonic relationship exhibits odd symmetry, with positive values corresponding to one side of the head, negative values corresponding to the other side, and zero indicating a centered image. In other words, any given value of $\hat{P}$ will correspond to a certain displacement of the perceived image towards one ear while $-\hat{P}$ will correspond to a displacement of the same magnitude towards the opposite ear. Although we believe that the function that relates the postion variable to lateral position most likely saturates for some value of $\hat{P}$ (since a person's head is finite in extent), we have not attempted to constrain this function, and for simplicity, we assume $\hat{P}$ to be directly proportional to subjective lateral position. For the majority of the stimuli considered in this report such a description is adequate since listeners do not report extreme displacements of the binaural image.

In **DM2**, we described two possible ways of computing the correlation variable. This provision was included to allow a comparison of the two decision statistics $\hat{Q}_o$ and $\hat{Q}_c$. The former was used by Colburn (1969, 1977a) in his evaluation of the model for detection and discrimination experiments. Stern (1976) argued that the modified statistic $\hat{Q}_c$ would yield nearly optimal performance for binaural detection experiments. This argument relied in part on the extremely expansive nature of the original exponential rectifier. Since we have replaced this rectifier with a lower-order nonlinearity in the new model, it is possible that Stern's argument is no longer completely valid. In the current work we compute predictions for the ratio of detection thresholds $\frac{N_\pi S_0}{N_0 S_\pi}$ using both decision variables and we evaluate their relative performance. Since the constant-coefficient assumption ($\hat{Q}_c$) provides a much simpler decision scheme than an optimal processor, it is intuitively more appealing (but only if it is capable of describing the data).

In considering **DM3**, we must specify the range of characteristic frequencies $\mathbf{R}_{lat}$ considered by the decision making mechanism (where $\mathbf{R}_{lat}$ equals $\mathbf{R}_{task}$ for lateralization

experiments). Stern and Colburn (1978) assumed that this range was chosen to optimize performance in interaural time discrimination experiments. They might have instead chosen the range which produced the best performance for an IID discrimination task. Another reasonable assumption might be to choose the range of characteristic frequencies which yields the largest shift in $\hat{P}$ for a given ITD (or IID). A number of other plausible restrictions can be imagined, but it would not prove very instructive to include a complete list at this time. Instead, we defer this discussion until Chapter 4 where we will be able to make more insightful comments regarding the effects of $\mathbf{R}_{lat}$ on lateralization predictions (see Section 4.3.1).

For detection experiments the concept of optimal performance is well defined, and the presumed correct choice for $\mathbf{R}_{det}$ can be determined empirically (where $\mathbf{R}_{det}$ refers to $\mathbf{R}_{task}$ for detection experiments). In general, this optimal range of characteristic frequencies will depend on the choice of $\hat{Q}_c$ vs. $\hat{Q}_o$ as well as on the frequency of the tone and on a number of parameters describing the masker.

3.3.2. Functional Interpretation

In this section, we attempt to provide some insight into the means by which the two decision variables are generated from the outputs of the binaural displayer. We begin with the computation of the position variable $\hat{P}$.

If we re-write the definition of the position variable (from Equation (3.10)) in the form of

$$\hat{P} = \sum_{m \in \mathbf{Z}_{lat}} \mathsf{P}_m \tau_m \tag{3.13}$$

where

$$\mathsf{P}_m = \frac{L_m}{\sum_{m \in \mathbf{Z}_{lat}} L_m},$$

it is apparent that $\hat{P}$ is simply a weighted sum of the internal delays. Since the outputs of the displayer are non-negative, the weighting factors P_m take on values in the range 0 to 1 and their sum is obviously 1. This suggests that the set of weighting factors can be interpreted as a discrete probability distribution, and the position variable is the expected value of this distribution. In order to interpret the significance of these "probabilities", it is

helpful to recall our earlier observation that the outputs of the binaural displayer exhibit the largest values when τ_m is close to the ITD (or a multiple period of the ITD) present in the stimulus (see page 42). With this in mind, we further note that each weighting factor P_m is proportional to the value of the display output L_m, and thus, it seems reasonable to think of these weighting factors as a measure of the probability that τ_m is "equal" to the interaural time delay in the stimulus.

The distribution of characteristic frequencies and internal delays can be considered to be a mechanism that adjusts our probability measure P_m to consider the information at certain characteristic frequencies and delay values to be more significant or reliable. The basis for this adjustment is presumably some *a priori* knowledge regarding the spectral and temporal properties of commonly encountered stimuli. If we are aware of the fact that most naturally occuring stimuli contain ITDs which are fairly small, we can include this information in our estimated probability function by having more fiber pairs with fairly small values of τ_m than there are fiber pairs with larger values.[7]

In light of the above observations, the position variable can be interpreted as a probabilistic estimate of the ITD present in the binaural stimulus; an estimate which is based on information extracted from the current input waveforms as well as preconceived notions regarding commonly encountered stimuli. This way of conceptualizing $\hat{P}$ does not contradict or preclude our previous centroid-based interpretation; it merely presents a slightly different perspective on the same set of operations.

While the correlation variable $\hat{Q}$ is not as easily related to a familiar perceptual attribute such as lateral position, it can be thought of as a crude measure of how alike the two input waveforms are. In general, this similarity measure will depend on both the target-to-masker ratio of the stimuli and the interaural disparities present in the tone and the noise. Since the outputs of the displayer are related to interaural correlation, they provide a description of how similar the left and right inputs are on a point by point basis. As the intensity of the tone is increased, the values of these outputs will vary. The amount by which a given L_m is affected by a given change in the signal level depends on the values of τ_m, fc_m, and interaural differences in the stimulus. Some displayer outputs will exhibit a

[7] The effect of the distribution of τ_m on $\hat{P}$ can best be demonstrated by considering the extreme case in which all fiber pairs have the same τ_m. Referring to Equation (3.13), we see that under these circumstances $\hat{P}$ will equal this value of τ_m regardless of the output of the binaural displayer.

strong dependence on the signal level and others will be relatively insensitive to this parameter. Overall performance in the detection task will depend on how well $\hat{Q}$ reflects these changes caused by the presence of the tone in the masker. While it is obvious that an optimal choice of the coefficients in Equation (3.11) will predict detection thresholds that are less than or equal to those predicted using a constant value for these coefficients (as in Equation (3.12)), it is not obvious as to how the predicted ratio of thresholds $(\frac{N_\pi S_0}{N_0 S_\pi})$ will depend on our choice for these coefficients. We must rely on actual calculations to expose this dependency.

3.3.3. Statistical Analysis

In this section, we derive a stimulus-independent expression for the expected value of the position variable $\hat{P}$ and we describe a performance index based on the use of the correlation variable $\hat{Q}$ in a typical detection experiment.

3.3.3.1. The Position Variable

In order to compute the expected value of $\hat{P}$, we make the following approximation,

$$\hat{P} \approx \frac{\sum_{m \in \mathbf{Z}_{lat}} \tau_m L_m}{\mathsf{E}[\sum_{m \in \mathbf{Z}_{lat}} L_m]} \tag{3.14}$$

which is justified by the observation that the standard deviation of the term in the denominator is much smaller than its mean for the stimuli considered. Thus, the mean of the position variable is given by

$$E[\hat{P}] = \frac{\mathsf{E}[\sum_{m \in \mathbf{Z}_{lat}} \tau_m L_m]}{\mathsf{E}[\sum_{m \in \mathbf{Z}_{lat}} L_m]} = \frac{\sum_{m \in \mathbf{Z}_{lat}} \tau_m \, \mathsf{E}[L_m]}{\sum_{m \in \mathbf{Z}_{lat}} \mathsf{E}[L_m]} \tag{3.15}$$

In the above expressions, it is assumed that we are using some particular realization of the binaural displayer with fixed, known values of the characterizing parameters fc_m, τ_m, and the sensitivity constants for the left and right fibers, K_{Lm} and K_{Rm}. In actuality we do not know their values, and it is necessary to compute the expectation over these parameters.

This is justified by the fact that the set of characterizing parameters for a given realization of the timing displayer is necessarily representative of the assumed distribution of these parameters since the assumed distribution was inferred from physiological and psychoacoustical data. Thus, the mean of the position estimate is given by

$$E[\hat{P}] = \frac{\int_{\mathbf{R}_{lat}} p(fc) \int_{-\infty}^{\infty} \tau L(\tau, fc)\, p(\tau \,|fc)\, d\tau\, dfc}{\int_{\mathbf{R}_{lat}} p(fc) \int_{-\infty}^{\infty} L(\tau, fc)\, p(\tau \,|fc)\, d\tau\, dfc}\,, \tag{3.16}$$

with

$$L(\tau, fc) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathsf{E}\left[L_m \,|\, \tau_m = \tau, fc_m = fc, K_{Lm} = K_L, K_{Rm} = K_R \right] p(K_L, K_R \,|fc)\, dK_L\, dK_R\,. \tag{3.17}$$

The remaining expectation is computed with respect to the stimulus and the auditory-nerve model, and it is computed as in Equation (3.9). The above expression can be simplified by noting that the sensitivity constants only determine whether or not a fiber is active. There are three distinct cases to be considered: both fibers active, only one active, and neither active. Thus, $L(\tau, fc)$ can be expressed as,

$$L(\tau, fc) = \eta_2(fc)\, \mathcal{L}_2(\tau, fc) + \eta_1(fc)\, \mathcal{L}_1 + \eta_0(fc)\, \mathcal{L}_0 \tag{3.18}$$

where

$$\mathcal{L}_i(\tau, fc) \triangleq E[\, L_m | \tau_m = \tau, fc_m = fc, \text{ "}i \text{ fibers are active"}\,] \quad \text{for } i = 0, 1, 2 \tag{3.19}$$

and $\eta_i(fc)$ is defined to be the fraction of fiber pairs with characteristic frequency fc that have i active fibers ($i = 0, 1, 2$).

If either fiber in a pair is firing spontaneously, the two intensity functions are statistically independent and the interaural crosscorrelation $R_{RLm}(\tau_m)$ is simply the product of the two mean firing rates. Thus, $\mathcal{L}_1 = T_w T_S(50)(200)$ and $\mathcal{L}_0 = T_w T_S(50)(50)$, independently of the stimulus properties (and the characteristics of the fiber pair). On the other hand, the function $\mathcal{L}_2(\tau, fc)$ does depend on the type of input involved.

As was already noted in Section 3.2.4, the interaural correlation function for tonal inputs is periodic and independent of characteristic frequency; hence, $\mathcal{L}_2(\tau, fc)$ can be written as $\mathcal{L}_2(\tau)$ for tonal stimuli. In Appendix C, we exploit these properties of the interaural correlation function to derive a frequency-domain expression for $E[\hat{P}]$ when tonal stimuli are used.

3.3.3.2. A Performance Index for Detection Experiments

In Section 2.1.2, we reviewed the findings of several studies examining relative detection thresholds. Each of these studies employed a different experimental setup and even a different definition of threshold performance. Empirically, it appears that such differences among studies are not important since only relative comparisons of thresholds are made. As long as a consistent experimental procedure is maintained throughout each investigation, the findings should be directly comparable.

In order to calculate predicted performance for detection experiments, we assume that a symmetric, two-interval, two-alternative-forced-choice (2I-2AFC) paradigm is used. The stimulus is composed of a low intensity tone added to a broadband Gaussian noise. During one of the two presentation intervals, the tone level is zero (stimulus *a*), while during the other interval it is set to A (stimulus *b*). The order of presentation of these two stimuli is randomly selected with equal probability that (*a*, *b*) or (*b*, *a*) will occur. After each presentation of the stimulus pair, listeners must indicate which interval contained the tone (stimulus *b*).

The detection threshold is defined to be the value of A for which the index Q_d has unit value (when there is no bias). The quantity Q_d is defined by

$$Q_d = \frac{(E[\hat{Q} \mid a] - E[\hat{Q} \mid b])^2}{Var\{\hat{Q}\}}, \tag{3.20}$$

where $E[\hat{Q} \mid a]$ and $E[\hat{Q} \mid b]$ represent the expected values of $\hat{Q}$ conditioned on alternatives *a* and *b*, respectively. [Colburn (1977a) assumes that the variance of $\hat{Q}$ given stimulus *a* is approximately equal to the variance of $\hat{Q}$ given stimulus *b* for the types of stimuli considered. While this is not always the case in the current study, we maintain this assumption for the present time.] If we assume that the underlying conditional distributions are Gaussian, a unit value of Q_d is indicative of aprroximately 75% correct performance. Such an assumption is reasonable if one considers that the decision statistic $\hat{Q}$ (by either definition) consists of the sum of many independent random variables.

An important property of the performance index described above is that the value of Q_d for an optimally weighted sum of uncorrelated decision variables is simply the sum of the Q_d's computed for each of the individual decision variables. This implies that for the optimal correlation variable $\hat{Q}_o$, performance can be measured by,

$$Q_d = \sum_{m \in \mathbf{Z}_{det}} \frac{(E[L_m|a] - E[L_m|b])^2}{Var\{L_m\}} = \sum_{m \in \mathbf{Z}_{det}} \frac{(E[L_m|a] - E[L_m|b])^2}{E\{L_m\}} \tag{3.21}$$

The latter part of this equation takes advantage of assumption **BD3** and the fact the mean and variance of a Poisson-distributed random variable are equal.

It can easily be shown that by defining the coefficients c_m to be

$$c_m = \frac{E[L_m|a] - E[L_m|b]}{Var\{L_m\}} \tag{3.22}$$

Equation (3.20) reduces to Equation (3.21) for $\hat{Q} = \hat{Q}_o$. This is in agreement with our intuition that the most significant outputs of the binaural displayer for detection are those which are most strongly affected by the addition of the target to the masker.

Considering now the constant-coefficient statistic $\hat{Q}_c$ and again making use of assumption **BD3**, we see that the performance index Q_d is simply given by,

$$Q_d = \frac{\left(\sum_{m \in \mathbf{Z}_{det}} E[L_m|a] - \sum_{m \in \mathbf{Z}_{det}} E[L_m|b]\right)^2}{\sum_{m \in \mathbf{Z}_{det}} E[L_m]} . \tag{3.23}$$

Thus far, we have assumed that the variance of the decision variable $\hat{Q}$ is the same for both stimulus configurations, *a* and *b*. However, predictions indicate that the variance terms in the denominators of Equations (3.21) and (3.23) do in fact depend on which stimulus alternative is presented. In order to account for this dependence when computing predictions we chose to geometrically average the values of the two conditional variances given stimulus *a* or *b*. Another solution would be to use an arithmetic average of the conditional variances. Comparisons of predictions, for the relative detection thresholds considered, obtained using both averaging techniques reveal differences of only a fraction of a dB, which is much less than the variability exhibited by the data. Thus, it appears that relative threshold predictions are somewhat insensitive to minor changes in the specification of Q_d.

We now compute the expectation of L_m with respect to the characterizing parameters of the model. In addition, we introduce the geometrically averaged conditional variance terms discussed above. Thus, the performance index based on $\hat{Q}_c$ becomes

$$Q_d = \frac{\left[\int_{\mathbf{R}_{det}} \int_{-\infty}^{\infty} L(\tau, fc \mid a)\, p(\tau, fc)\, d\tau\, dfc \; - \int_{\mathbf{R}_{det}} \int_{-\infty}^{\infty} L(\tau, fc \mid b)\, p(\tau, fc)\, d\tau\, dfc \right]^2}{\sqrt{\int_{\mathbf{R}_{det}} \int_{-\infty}^{\infty} L(\tau, fc \mid a)\, p(\tau, fc)\, d\tau\, dfc} \; \sqrt{\int_{\mathbf{R}_{det}} \int_{-\infty}^{\infty} L(\tau, fc \mid b)\, p(\tau, fc)\, d\tau\, dfc}} \tag{3.24}$$

where $L(\tau, fc \mid a)$ and $L(\tau, fc \mid b)$ are the conditional means of L_m, as defined in Equation (3.17), but further conditioned on the presentation of stimulus *a* or *b*.

Before performing a similar operation to obtain Q_d for the optimally weighted correlation variable, we note that for the broadband noise maskers considered, all fiber pairs of interest should be doubly active. This implies that $L(\tau, fc)$ is equal to $L_2(\tau, fc)$, and we can therefore write

$$Q_d = \int_{\mathbf{R}_{det}} \int_{-\infty}^{\infty} \left[\frac{\left(L_2(\tau, fc \mid a) - L_2(\tau, fc \mid b) \right)^2}{\sqrt{L_2(\tau, fc \mid a)\, L_2(\tau, fc \mid b)}} \right] p(\tau, fc)\, d\tau\, dfc \tag{3.25}$$

where $L_2(\tau, fc \mid a)$ is the mean of L_m given (1) that the m^{th} fiber pair is doubly active, (2) that stimulus *a* is presented, (3) that $\tau_m = \tau$, and (4) that $fc_m = fc$. $L_2(\tau, fc \mid b)$ is similarly defined.[8]

In Section 3.2.4 we pointed out that it is not realistic to expect predictions for the mean and variance of L_m to be entirely accurate; they will (at least) differ from the actual statistics by a constant factor, which is related to our assumptions about the coincidence window $f(z)$ and the time constant of the coincidence counter. With this in mind, we further note that both Equations (3.24) and (3.25) are sensitive to multiplicative factors in the expressions for $L(\tau, fc)$ and $L_2(\tau, fc)$. Specifically, the value of the erroneous factor will multiply the performance index Q_d in both of these expressions. This suggests that it is difficult to obtain absolute detection threshold predictions since the value of this factor is unknown. In order to generate relative threshold predictions, we use one absolute threshold data point to determine the value of the constant factor, and we are then able to generate threshold predictions for other stimulus configurations, measured relative to the first data point.

---

[8] When generating predictions using Equation (3.25), it is possible for the denominator to be zero for certain values of τ and fc. For these particular values, an arithmetic average (rather than a geometric average) was used. Based on our previous observation regarding the insensitivity of relative threshold predictions to the type of averaging used, we don't feel that this implementational consideration significantly affects predictions.

3.4. Summary

In this chapter we have described the detailed assumptions that define the model of binaural interaction used in the current study. This model is based very closely on the position-variable model described by Stern and Colburn (1978), with only a few modifications. First, the description of the peripheral transduction process has been altered to allow for a more accurate description of the neural response to noise waveforms. This was accomplished by making use of a less expansive half-wave rectifier in place of the original exponential rectifier. Second, the envelope detecting capabilities of the the auditory-nerve model have been improved by relocating the lowpass filter element at a point beyond the rectifier in the input transformation. This should result in a more accurate characterization by the model of lateralization phenomena for complex, high-frequency stimuli that contain temporal cues in their low frequency envelopes. Third, the assumed distribution of internal, interaural delay parameters over the binaural coincidence-counting units has been left unspecified in order that it may be constrained by certain lateralization and detection data in the next chapter.

Chapter 4

Predictions for the Lateralization and Detection of Tones

In this chapter, we compare predictions of the new model to experimental data describing the lateralization and detection of tones as discussed in Sections 2.1.1.1 and 2.1.2. Our two major goals in this effort are: (1) to gain insight into the dependence of the predictions on the various free parameters of the model (especially ν, $p(\tau|fc)$, and $\mathbf{R}_{lat}$); and (2) to specify at least one set of these free parameters that provides good agreement between predictions and data in order to prove that a model of this type (which uses centroid analysis) is capable of describing the lateralization and detection phenomena considered.

We begin this section by proposing a few forms for $p(\tau|fc)$ and examining how certain properties of the proposed functions affect the ability of the model to describe the two bodies of psychoacoustic data mentioned above. In the second and third sections we discuss the selection of ν and $\mathbf{R}_{lat}$, respectively. In the fourth section we present a brief discussion of predictions for detection thresholds of N_0S_π stimuli as a function of target frequency. We conclude in the fifth section by summarizing our observations.

4.1. Specifying the Distribution of Internal Delays

We now address the problem of specifying a distribution of internal delays which will allow the modified model to describe both the lateralization of tones as a function of frequency and the difference between $N_\pi S_0$ and N_0S_π detection thresholds as a function of target frequency. The motivation for this portion of our work and the approach used to achieve the goals cited above are outlined in the first subsection. The second subsection provides a brief discussion of how predictions for the detection experiments are generally affected by the shape of the internal-delay distribution. In the third subsection we derive a

set of mathematical constraints which must be satisfied by $p(\tau \mid fc)$ in order to describe the lateralization data of Schiano, *et al.* (1986). We then propose a few parametrically defined density functions and examine how their parameters must be chosen in order to comply with these constraints. In the fourth subsection we present predictions which were obtained using the new functions and we evaluate the model's performance in describing the data.

While the discussions presented in this section are organized in a manner similar to the actual research, there are two major exceptions worth noting. First, we do not discuss the entire set of density functions $p(\tau \mid fc)$ that were studied; we describe only those which have proven most successful and those failures which tend to contradict certain intuitively appealing ideas. Second, in order to concentrate on the effects caused by changes in $p(\tau \mid fc)$, we assume for this discussion that the other free parameters of the model are fixed. More specifically, the predictions presented in this section have been generated using a rectifier power ν equal to three and by considering only those fiber pairs with characteristic frequencies equal to the tone frequency (*i.e.*, $\mathbf{R}_{lat} = \{fc \mid fc = f_0\}$). During our original consideration of the proposed density functions, predictions were generated for several choices of ν and $\mathbf{R}_{lat}$. The values mentioned above have been selected because they yield (somewhat) typical predictions for both experiments considered. (A more complete analysis of the dependence of predictions on the rectifier power and the range of characteristic frequencies is given in Sections 4.2 and 4.3.)

4.1.1. Motivation and Strategy

The work described in this section is motivated by the inability of the original position-variable model to describe the lateralization of tones with small ITDs as a function of frequency (Schiano, Trahiotis, and Bernstein, 1986). Referring to Figure 4-1, we see that the original model predicts perceived position to be strongly dependent on the frequency of the tone. However, the findings of Schiano, *et al.* indicate that the image position remains relatively constant for tones with frequencies up to about 1200 Hz, and only at higher frequencies does the image shift towards the center of the head. Throughout this chapter we consider the data for low and high frequencies as separate cases since we feel that different mechanisms are responsible for the observed behavior over these two frequency ranges.

The inability of subjects to lateralize high frequency tones purely on the basis of

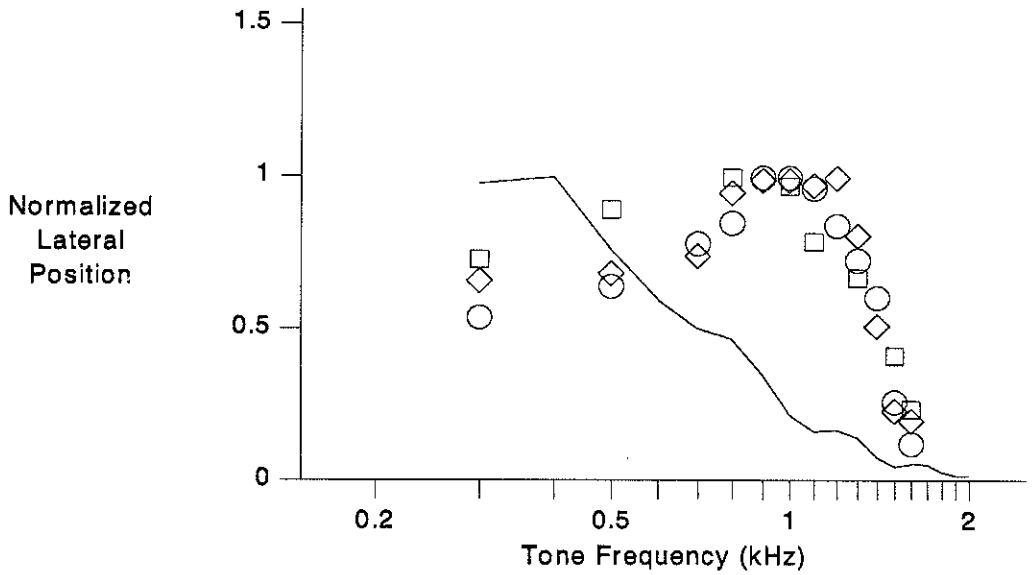

Figure 4-1: Predictions of the original position-variable model for the lateralization of tones with ITDs of 150 μs as a function of frequency. [Normalized data points have been adapted from Schiano, *et al.*, 1986.]

interaural time delay is believed to be a result of the loss of synchrony between neural firing patterns and the detailed timing structure of the stimulus, making it impossible to estimate the value of the ITD present in the binaural stimulus. This observation is (qualitatively) consistent with predictions of the auditory-nerve model, and we do not expect the data for higher frequencies to be problematical.

In order to understand the bases within the original position-variable model which prevent it from correctly describing the data at low frequencies, we consider a binaural tone of frequency f_0 (and no IID) for which the signal in the left ear is leading the signal in the right by a small amount τ_s. As is illustrated in Figure 4-2(a), the interaural crosscorrelation function $L_2(\tau)$ is periodic with maxima occuring at $\tau = \tau_s + \frac{k}{f_0}$, where k is any integer. The effect of $p(\tau|fc)$ in the position calculation is to greatly attenuate all but the most central of these maxima (see panel (b)). For ITDs less than half the tone period, the primary peak of $L_T(\tau)$ occurs when $\tau = \tau_s$ and the secondary peak occurs when $\tau = \tau_s - \frac{1}{f_0}$. Referring to panel (c), we see that as the stimulus frequency increases, the secondary peak of $L_2(\tau)$ moves closer to the midline while the primary peak remains at $\tau = \tau_s$. Thus, the relative significance of the secondary peak in the centroid computation increases and the position estimate moves towards the center.

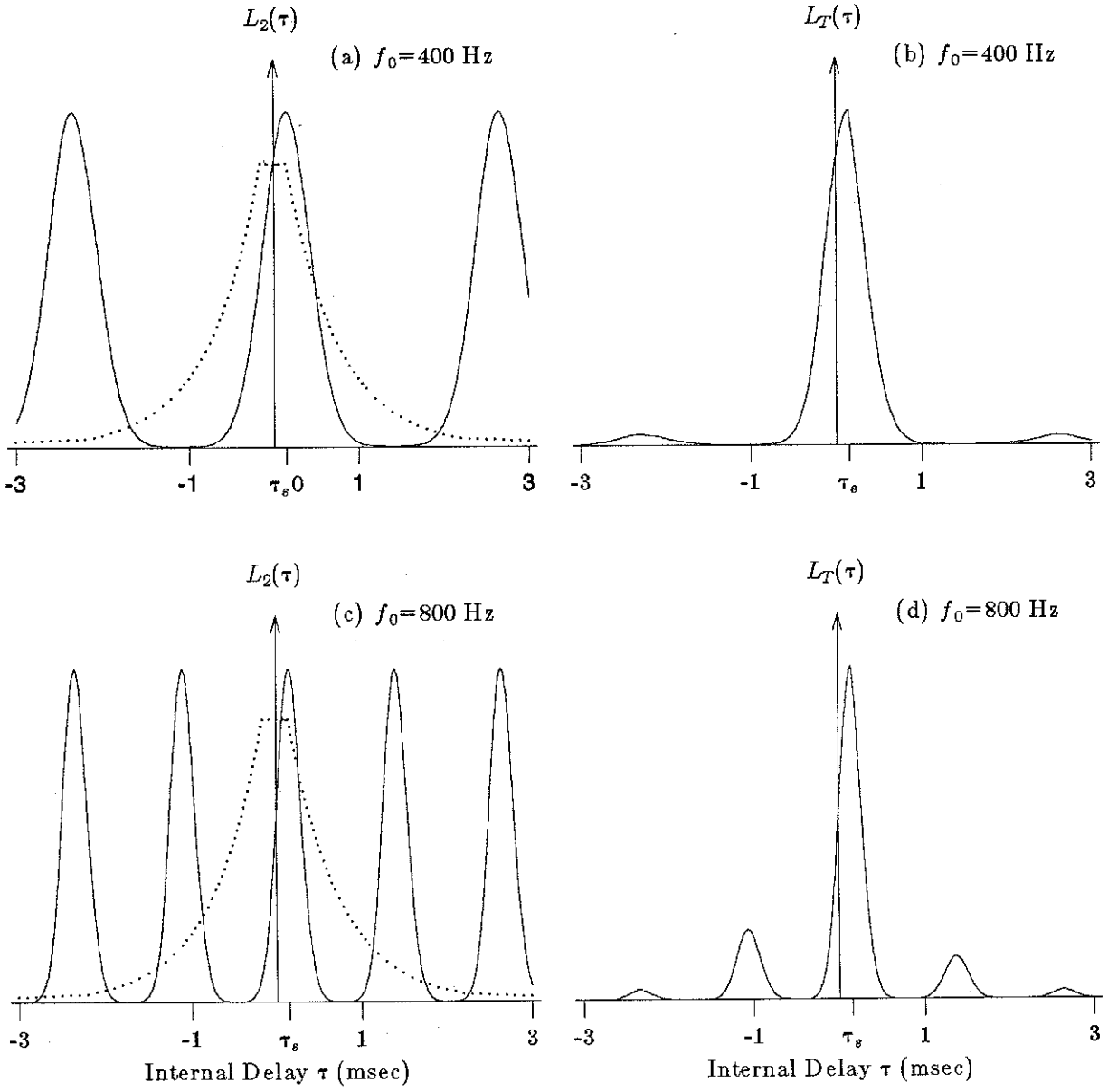

Figure 4-2: The functions $L_2(\tau)$ and $L_T(\tau)$ of the original position-variable model are plotted as a function of the internal delay parameter τ for (a,b) 400-Hz and (c,d) 800-Hz tones with ITDs of 150 μs (marked by τ_s). The dotted line in panels (a) and (c) describes the relative distribution of internal delays, as specified by Colburn (1977a).

In order to maintain $E[\hat{P}]$ constant for low frequencies, it appears that $p(\tau|fc)$ must become narrower as the secondary peak approaches the midline or, in other words, as f_0 (and presumably fc of interest) increases. This narrowing of $p(\tau|fc)$ at higher

characteristic frequencies corresponds to an increase in the relative number of fiber pairs with small internal delays at these frequencies. Such an increase is intuitively correct for the lateralization of tones since little information is gained from internal delays of magnitude greater than half the period of the tone.

The above observations imply that although the original position-variable model does not correctly describe these particular lateralization phenomena, it may be possible to correct this inconsistency between model predictions and data by respecifying the assumed distribution of internal delays. Since the original distribution was chosen by Colburn (1969, 1977a) to explain the relative dependence of certain masking level differences on tone frequency, we feel it is important that the new distribution also be able to describe this dependence.

In our efforts to attain this goal, we first made certain simplifying assumptions and approximations which allowed us to analytically determine the constraints imposed on $p(\tau \mid fc)$ by the lateralization data. Several forms of the density function which satisfied these constraints were proposed, and we verified their ability to predict the lateralization data. These functions were also used to generate predictions for the detection experiments. We then adjusted the remaining free parameters and relaxed the lateralization constraints in an attempt to obtain a satisfactory description of both bodies of data. Bear in mind that only those apsects of our work that ultimately proved most effective in realizing this goal and those aspects that yielded counterintuitive results are described in the following sections.

4.1.2. Detection Considerations

The purpose of this discussion is to provide the reader with a basic understanding of how the shape of $p(\tau \mid fc)$ affects predictions for the difference between $N_\pi S_0$ and $N_0 S_\pi$ detection thresholds. While we do not derive a strict set of restrictions which must be imposed on the form of $p(\tau \mid fc)$ in order to describe these detection data, we do develop a sense of how a given density function should be altered in order to more accurately predict these phenomena.

We begin by considering the task of detecting the presence of a 250-Hz, antiphasic tone in a diotic noise masker (*i.e.*, an $N_0 S_\pi$ configuration). The solid line of Figure 4-3(a) describes the interaural crosscorrelation function $L_2(\tau, fc)$ when fc=250 Hz and when

there is no target signal present.[9] As the target-to-masker ratio is increased, this function is altered, as indicated by the dashed line of the same figure. The major differences between these two curves appear to occur at the peaks (although some changes can also be noted in the "valleys"), and we expect that the decision statistic $\hat{Q}$ will reflect these differences between the signal-present and signal-absent cases. If we now consider panel (b), which describes the same functions for the $N_\pi S_0$ configuration, we see that similar changes occur when the target is present. However, the locations of the peaks (and valleys) have been shifted by half the period of the tone. Since the distribution $p(\tau|fc)$ indicates that there are more coincidence units with τ_m near zero, those peaks which lie closer to the midline can be considered more reliable than those which occur at larger values of τ. Thus, we expect better performance under those stimulus conditions for which the peaks of $L_2(\tau)$ coincide with the dominant regions of $p(\tau|fc)$. For the stimuli considered here, we predict that the out-of-phase target in the diotic masker is more easily detected than the diotic target in the out-of-phase masker.

As the tone frequency is increased (see Figure 4-3(c,d)), the peaks for the $N_\pi S_0$ case move closer to the midline, and thus the difference in the detection thresholds should decrease. The exact rate at which this difference in thresholds decreases with respect to target frequency is dependent on the shape of $p(\tau|fc)$. If the distribution of internal delays is broad with respect to the period of the tone ($\sim\frac{1}{fc}$), then there will be only a small difference between the thresholds for the two cases. Similarly, if the density function indicates that the majority of the fiber pairs have associated delays with magnitudes less than half of the period of the tone, then the $N_0 S_\pi$ threshold should be significantly less than the $N_\pi S_0$ threshold.

At very high frequencies, the loss of synchrony on the auditory nerve fibers causes the interaural correlation functions to be constants, independent of τ. For these stimuli, detection is dominated by monaural mechanisms, and interaural time disparities do not significantly affect threshold levels (cf. Durlach and Colburn, 1978). Thus, the data at frequencies above about 2000 Hz are not particularly useful in constraining $p(\tau|fc)$.

---

[9] The curves in this figure are presented only for positive values of τ since these functions are symmetric about $\tau=0$.

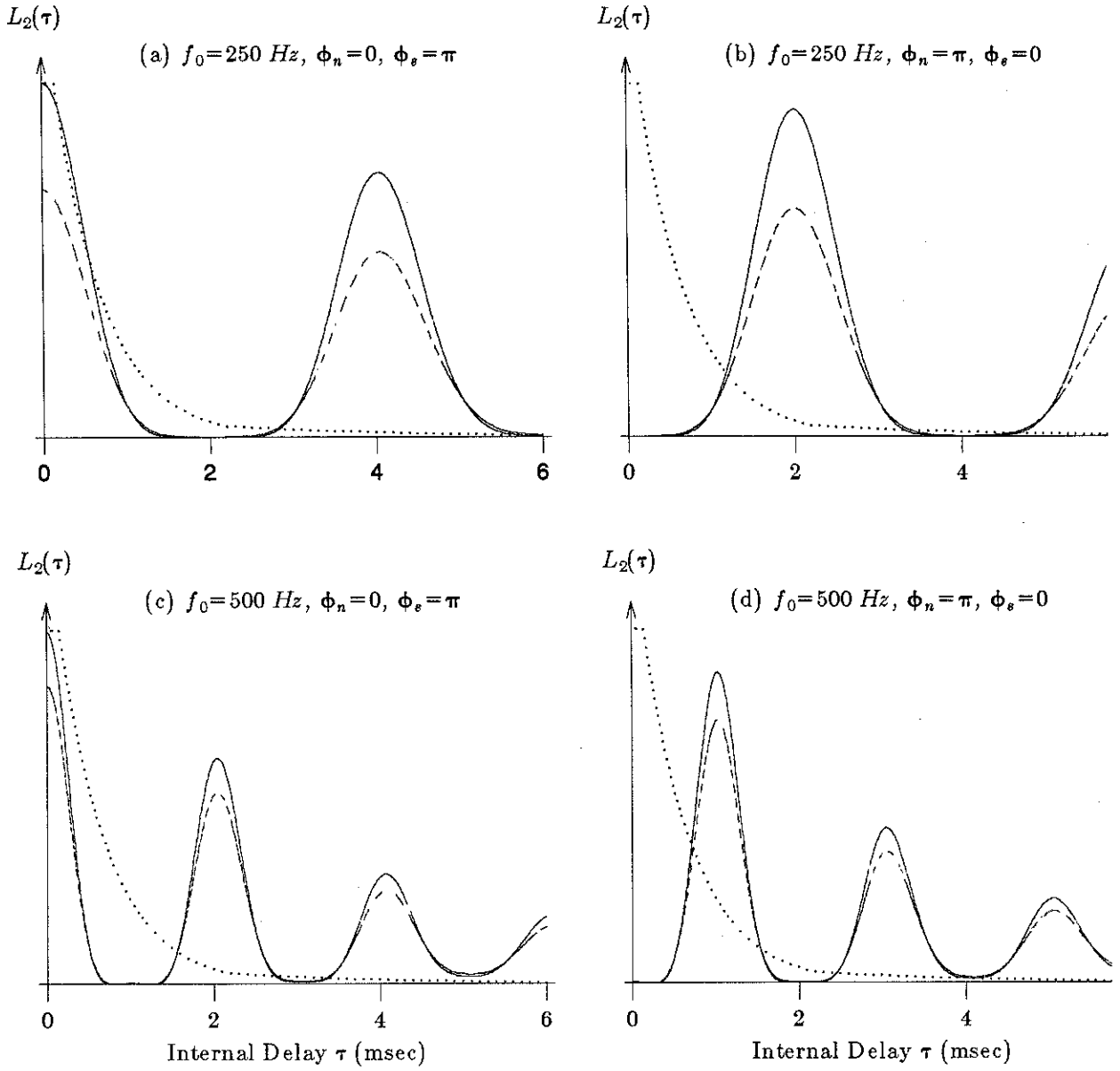

Figure 4-3: The statistic $L_2(\tau, fc)$ is plotted as a function of τ when $fc = f_0$ for 250-Hz and 500-Hz tones presented in a broadband noise. The solid line represents the case when the tone is absent, and the dashed line indicates the case when the tone is presented near threshold. The dotted curve describes the shape of the original $p(\tau)$ function specified by Colburn (1977).

4.1.3. Lateralization Constraints

In Section 4.1.1, we provided a heuristic interpretation of the dependence of lateralization predictions on the shape of $p(\tau|fc)$. We now attempt a more thorough investigation of this issue.

We begin this discussion by restating the trends and characteristics of the laterality matching data that we intend for the new model to successfully describe. The experiments conducted by Schiano, *et al.* (1986) made use of fairly long presentations of a binaural stimulus similar to the one given by

$$x_L(t) = A\cos\ 2\pi f_0 t$$

$$x_R(t) = A\cos\ 2\pi f_0 (t-\tau_s)$$

where A is chosen to correspond to a level of 70 dB SPL (for each channel). As was pointed out in Section 2.1.1.1, the position judgements obtained in these experiments can be summarized by

$$E[\hat{P}] \approx C\,H_{data}(f_0) \times \tau_s, \quad \text{for } \tau_s \leq \frac{1}{4\,f_0} \tag{4.1}$$

where C is a proportionality constant and $H_{data}(f)$ is a lowpass filter function with a corner frequency of about 1200 Hz (see Figure 2-2).

It is not obvious from Equation (3.16) what forms of $p(\tau|fc)$, if any, will allow the model to predict this behavior. Thus, before proceeding further, we must first derive a simple expression for $E[\hat{P}]$ which resembles Equation (4.1) in form. By noting that for tonal stimuli the interaural crosscorrelation function $\mathcal{L}_2(\tau, fc)$ is periodic in τ and independent of fc and by exploiting certain properties of the Fourier transform, we demonstrate in Appendix C that the ratio in Equation (3.16) can be equivalently expressed as

$$E[\hat{P}] = \frac{\dfrac{-200^2}{\pi}\displaystyle\sum_{n=1}^{\infty} G^2(nf_0)\,S_n^2 \sin(2\pi\, nf_0\tau_s)\int_{\mathbf{R}_{lat}} \eta_2(fc)\,\dot{\phi}_\tau(nf_0|fc)\,p(fc)\,dfc}{D_0 + 2\,(200)^2\displaystyle\sum_{n=1}^{\infty} G^2(nf_0)\,S_n^2 \cos(2\pi\, nf_0\tau_s)\int_{\mathbf{R}_{lat}} \eta_2(fc)\,\phi_\tau(nf_0|fc)\,p(fc)\,dfc} \tag{4.2}$$

where

$$D_0 \triangleq \int_{\mathbf{R}_{lat}} [\,\eta_2(fc)\,200^2 + \eta_1(fc)\,(200)\,(50) + \eta_0(fc)\,50^2\,]\,p(fc)\,dfc\ , \tag{4.3}$$

$\phi_\tau(f\,|fc)$ is the Fourier transform (or characteristic function) of $p(\tau\,|fc)$, $\dot{\phi}_\tau(f\,|fc)$ is the derivative of $\phi_\tau(f\,|fc)$ taken with respect to f, and S_n is the magnitude of the n^{th} Fourier series coefficient for the output of the rectifier normalized by the average firing rate (200 per second).

While this expression in its current form provides a more computationally efficient means of calculating the predicted position, it does not shed much additional light on the problem at hand. However, by further assuming that (1) $\eta_2(fc)$ is equal to one for all fc in $\mathbf{R}_{lat}$, (2) the contribution from all higher harmonics of the fundamental frequency of $L_2(\tau)$ is negligible, and (3) the remaining sine and cosine terms may be replaced by the appropriate first order approximations (which are valid for $\tau_s \leq \frac{1}{5\,f_0}$), we obtain the following relation

$$E[\hat{P}] \approx \left[\frac{-2\,S_1^2\,G^2(f_0)\,f_0 \int_{\mathbf{R}_{lat}} \dot{\phi}_\tau(f_0|fc)\,p(fc)\,dfc}{\int_{\mathbf{R}_{lat}} p(fc)\,dfc\ +\ 2S_1^2\,G^2(f_0) \int_{\mathbf{R}_{lat}} \phi_\tau(f_0|fc)\,p(fc)\,dfc}\right] \times\ \tau_s,$$

$$\text{for}\ \ \tau_s \leq \frac{1}{5f_0} \tag{4.4}$$

which is in a form similar to that of Equation (4.1).

These observations imply that the term in brackets above must be directly proportional to $H_{data}(f_0)$ in order to describe the experimental results. In other words, the mathematical constraint imposed on the model's parameters by the lateralization data (under the simplifying assumptions stated above) is given by

$$H_{data}(f_0) = C\ \times \left[\frac{-2\,S_1^2\,G^2(f_0)\,f_0 \int_{\mathbf{R}_{lat}} \dot{\phi}_\tau(f_0|fc)\,p(fc)\,dfc}{\int_{\mathbf{R}_{lat}} p(fc)\,dfc\ +\ 2S_1^2\,G^2(f_0) \int_{\mathbf{R}_{lat}} \phi_\tau(f_0|fc)\,p(fc)\,dfc}\right] \tag{4.5}$$

where C is some constant. For tone frequencies less than 1200 Hz, $H_{data}(f_0)$ is considered to be constant and it does not appear that for an arbitrary range of characteristic frequencies $\mathbf{R}_{lat}$ this relationship can be satisfied exactly. Consequently, we now consider two approximations that allow us to further simplify the constraint.

In the first case we assume that the central processor only uses those fiber pairs with characteristic frequencies (nearly) equal to f_0 when computing the position estimate $\hat{P}$. While such a choice for $\mathbf{R}_{lat}$ seems somewhat unlikely, we have already noted that there is no definitive range of characteristic frequencies for lateralization and it is not inconceivable that this assumption is valid. For this choice of $\mathbf{R}_{lat}$, Equation (4.5) can be simplified to

$$H_{data}(f_0) = C \times \left[\frac{-2\, S_1^2\, G^2(f_0)\, f_0\, \dot{\phi}_\tau(f_0|f_0)}{1 + 2S_1^2\, G^2(f_0)\, \phi_\tau(f_0|f_0)} \right] \tag{4.6}$$

where C is some constant. It can be demonstrated that any conditional distribution that can be expressed in the form

$$p(\tau\,|\,fc) = [\,C\,fc\,]\; p_\theta(C\,\tau\,fc)\;, \quad \text{for } 300 \leq fc \leq 1200 \text{ Hz}, \tag{4.7}$$

will satisfy Equation (4.6). In the above expression, C is some positive constant and $p_\theta(\theta)$ is a valid density function that fulfills the requirements established by assumption **BD2** (on page 40). We refer to this type of conditional density function as a *phase-based* distribution since it is simply a function of the product $\tau\, fc$, or phase. This result is in agreement with our earlier, intuitive argument that $p(\tau\,|\,fc)$ should become "narrower" as fc increases.

For the second solution we make no assumptions regarding the range of characteristic frequencies considered. In order for the right hand side of Equation (4.5) to be independent of $\mathbf{R}_{lat}$, it appears that the assumed distribution of delays should not depend heavily (or at all) on characteristic frequency. Noting that the numerator of the term in brackets is more strongly dependent on f_0 than the denominator, we arrive at the following approximate solution to the original constraint.

$$\dot{\phi}_\tau(f\,|\,fc) \approx \frac{-C}{f}, \quad \text{for } 300 \leq f \leq 1200 \text{ Hz} \tag{4.8}$$

where C is an arbitrary, positive constant. This implies that the characteristic function $\phi_\tau(f\,|\,fc)$ should be linearly related to the logarithm of f between 300 and 1200 Hz. For ease of reference, functions that satisfy this constraint will be called *log based* distributions.

Consider now the model's response, as described by Equation (4.4), to input tones of frequency greater than 1200 Hz. If we assume that $p(\tau\,|\,fc)$ satisfies either of the two constraints just derived, even for frequencies above 1200 Hz, we see that $E[\hat{P}]$ is

dominated by the lowpass filter $G(f_0)$ since the magnitude of a characteristic function is necessarily less than or equal to one (cf. Lukacs, 1970) and since the normalized Fourier coefficient S_1^2 is also less than one. Because this behavior describes the basic trend exhibited by the lateralization data we presently do not attempt to further constrain $p(\tau|fc)$ for characteristic frequencies greater than 1200 Hz. A more detailed discussion of this issue is provided later in the chapter.

4.1.4. Predictions Obtained Using the Proposed Density Functions

Several density functions which satisfy the two constraints described above were proposed and used to generate predictions for the two experiments of interest. We now describe predictions for a few of the functions considered that are helpful in demonstrating the general trends and issues observed while studying the more complete set of density functions.

The predictions in this section were generated under the assumptions that $\nu = 3$ and that only those characteristic frequencies (nearly) equal to the tone frequency are significant in the formation of the decision statistics $\hat{P}$ and $\hat{Q}$. (The effects of these two assumptions are discussed later in the chapter.)

4.1.4.1. Phase Based Distributions

In order to demonstrate the ability (or inability) of phase-based distributions to describe both the lateralization and detection data of interest, we consider the frequency-dependent distribution $p_G(\tau|fc)$

$$p_G(\tau|fc) = \frac{1}{\sqrt{2\pi}\,\sigma(fc)} \exp\left[\frac{-\tau^2}{2\,\sigma^2(fc)}\right] \tag{4.9}$$

with

$$\sigma(fc) \triangleq \begin{cases} C_\sigma / fc^\gamma, & \text{for } fc < 1200 \text{ Hz} \\ C_\sigma / 1200^\gamma, & \text{for } fc \geq 1200 \text{ Hz} \end{cases} . \tag{4.10}$$

When γ is equal to one, the above distribution satisfies the "phase" constraint of Equation (4.7). We have introduced this parameter in order to examine how the model's predictions are affected by the rate at which $p(\tau|fc)$ "narrows" as fc increases. There are two important issues regarding the frequency-dependent standard deviation $\sigma(fc)$ that should be discussed before proceeding.

First, Equation (4.7) suggests that the value of the constant C_σ does not affect relative lateral position predictions. However, our actual predictions do not entirely agree with this assertion. The source of this discrepancy lies in our earlier approximation regarding the contribution from higher order harmonics in Equation (4.2). In general, the characteristic function $\phi_\tau(f|fc)$ resembles a lowpass filter function, and the magnitude of its derivative has a maximum at the primary inflection point of this function (see Figure 4-4). It is possible to specify a sufficiently narrow or sharp density function so that the maximum of $|\phi'_\tau(f|fc)|$ occurs at some higher harmonic of f_0 (or fc). Under these circumstances, the predicted position estimate is dominated by the terms involving this higher harmonic rather than those involving the fundamental frequency, as was previously assumed. This results in poor correspondence between predicted and experimentally determined curves of position vs. ITD for a tone of given frequency. Thus, we should add the further stipulation to Equation (4.10) that the constant be sufficiently large so as to ensure that the inflection point of $\phi_G(f|fc)$ lies at a frequency well below $2fc$. For values of the constant within this range, relative lateralization predictions are in fact fairly insensitive to the exact choice of this parameter, and it is free to be constrained by the detection data.

Second, we have defined $\sigma(fc)$ in such a way that it is independent of characteristic frequency above 1200 Hz. Although such a change in the dependence of the distribution on fc may initially seem somewhat contrived, it is in fact justifiable. First, we note that this dependence was initially introduced to explain the lateralization of tones with ITDs at low frequencies. Since the psychoacoustical evidence indicates that high-frequency tones cannot be lateralized on the basis of interaural time delay alone, it is not reasonable to assume that this dependence must be maintained at higher frequencies. Second, it is generally accepted that listeners are capable of making use of ITDs in the envelopes of complex, high-frequency stimuli (*e.g*, (Henning, 1974; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1976; Henning, 1980; Nuetzel and Hafter, 1981). If $p(\tau|fc)$ is assumed to be a function of phase even at high frequencies, the model will predict only very small displacements, if any, of the sound image elicited by such stimuli. Although we have not actually generated predictions for lateralization experiments of this type, we feel that the definition of $\sigma(fc)$ in Equation (4.10) should provide a better description of these data. Third, predictions for the lateralization experiments considered in this chapter provide a better description of the sharp shift in lateral position observed at frequencies above 1200 Hz if we assume that $\sigma(fc)$ is not a function of phase at these frequencies. While a more

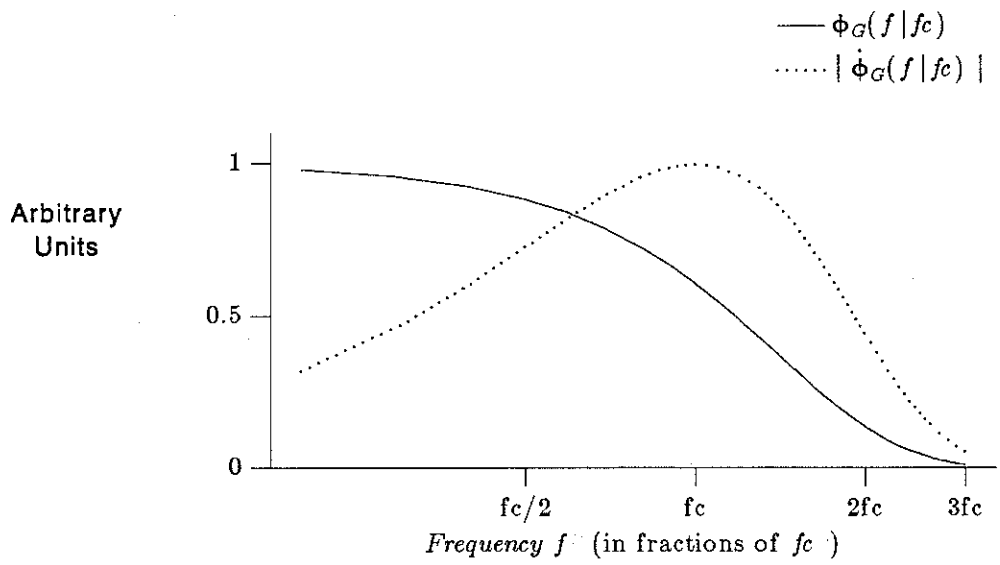

Figure 4-4: Plots of the characteristic function $\phi_G(f\,|fc)$ (solid line) and the vertically scaled magnitude of its derivative $|\dot{\phi}_G(f\,|fc)|$ (dotted line) as functions of frequency when $C_\sigma = 0.159$ and $\gamma = 1$.

gradual decrease in the dependence of σ on fc at high frequencies might be more plausible, the above characterization is satisfactory for the purposes of this discussion.

Figure 4-5 shows predictions for the two experiments obtained using $p_G(\tau\,|fc)$ with three different pairs of parameter values. When γ is set to 1.2, we obtain a good description of the lateralization data in panel (a). As the value of γ is decreased to 0.9, the ability of the model to describe the data at frequencies below 1200 Hz diminishes. The theoretical curve corresponding to γ equal to one, or a phase-based distribution, indicates that satisfying the phase constraint (of Equation (4.7)) does result in nearly constant lateral position predictions below 1200 Hz. (Unfortunately, it also points out that this characterization of the data is not very accurate.)

Figure 4-5(b) shows predictions for the detection experiment obtained using the same realizations of $p_G(\tau\,|fc)$ and using the optimal correlation variable $\hat{Q}_o$. (Predictions obtained using the constant-coefficient variable $\hat{Q}_c$, under the current assumptions, are almost identical to those in the figure, and therefore we do not include them.) Considering the curve generated using γ equal to one, we note that the predicted difference in detection

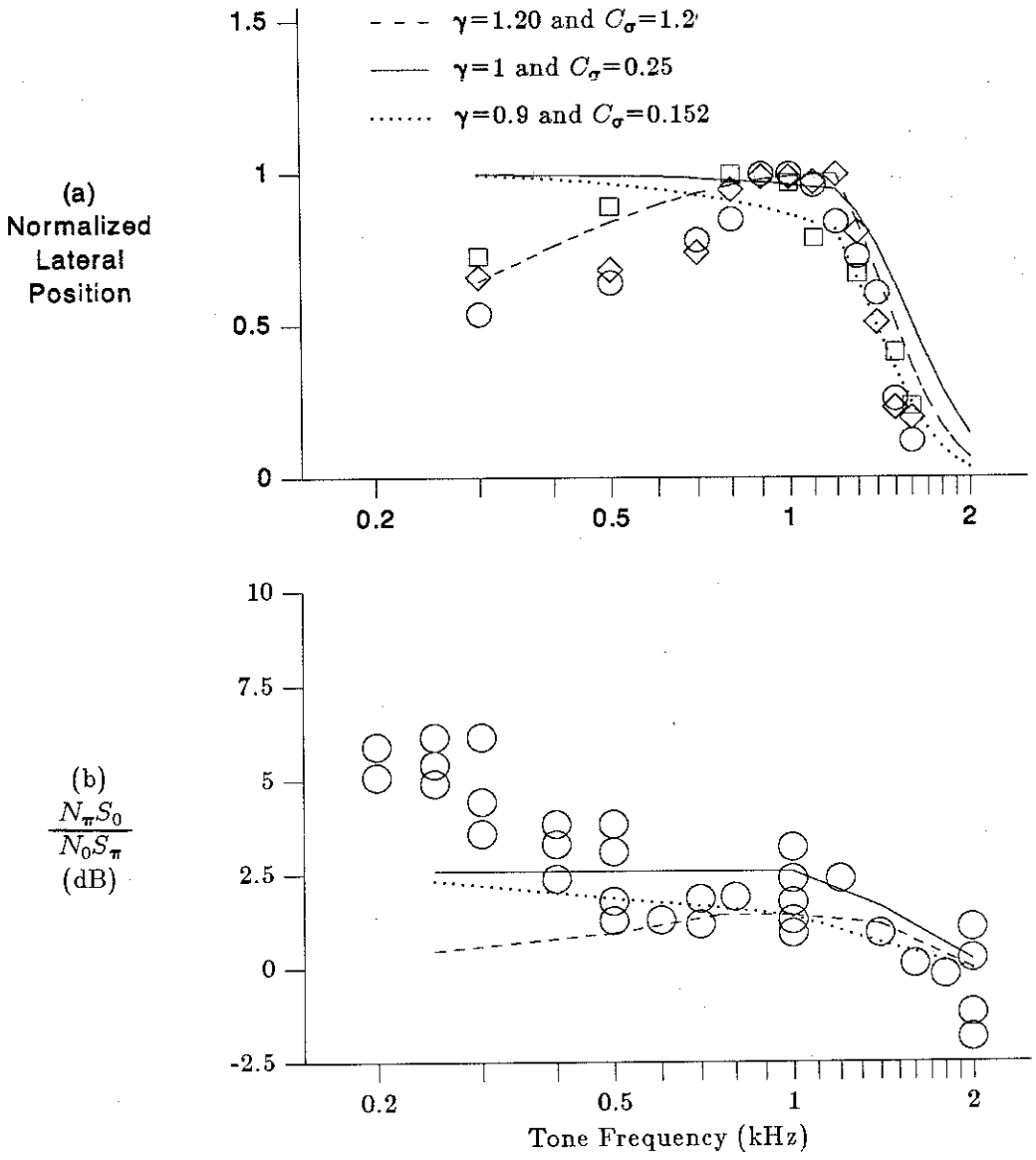

Figure 4-5: Predictions obtained using $p_G(\tau|fc)$ for (a) the lateralization of tones with an ITD of $150\mu s$ as a function of frequency and (b) the ratio of $N_\pi S_0$ thresholds to $N_0 S_\pi$ thresholds as a function of target frequency. The above predictions were generated under the assumptions that $\nu = 3$, $fc = f_0$, and $\hat{Q} = \hat{Q}_o$. [The lateral position data have been reproduced from Figure 2-2, and the relative threshold data have been reproduced from Figure 2-6.]

thresholds is essentially constant for frequencies less than 1200 Hz.[10] These results are understandable in the context of our discussion in Section 4.1.2 since both $p(\tau\,|\,fc{=}f_0)$ and $\mathcal{L}_2(\tau, f_0)$ are functions of phase. As the parameter γ is decreased below one, the qualitative trend exhibited by the detection predictions at low frequencies begins to more closely resemble that of the data.

In general, we found that values of γ less than one are required to provide a good description of the detection data while values greater than one are necessary to provide a good description of the lateralization phenomena. Thus, we feel that it is not possible to specify a single density function of the form given by Equation (4.9) that will describe the available psychophysical evidence for the two experiments considered. Similar observations were made for other forms of $p(\tau\,|\,fc)$, for example a double-sided exponential with a "plateau", resembling the original $p(\tau)$ specified by Colburn (1977).

4.1.4.2. Log-Based Distributions

We now address the problem of finding a characteristic function that is consistent with our previous assumptions regarding the shape of $p(\tau\,|\,fc)$ (as noted in **BD2** on page 40) and that also exhibits the properties described by Equation (4.8) over the range of frequencies of interest. Several such characteristic functions exist. The function $\phi_L(f)$ given below was selected because it has a corresponding density function $p_L(\tau)$ that is analytic in τ.

$$\phi_L(f) = \frac{1}{2\,\ln(f_h/f_l)}\,\ln\left[\frac{f^2+f_h^2}{f^2+f_l^2}\right] \tag{4.11}$$

$$\updownarrow$$

$$p_L(\tau) = \frac{1}{2\,\ln(f_h/f_l)}\,\frac{e^{-2\pi f_l|\tau|} - e^{-2\pi f_h|\tau|}}{|\tau|} \tag{4.12}$$

where f_l and f_h are constants which can be selected in order to best describe the lateralization and detection data. Due to the dependence of $\dot{\phi}_L(f)$ on these constants and our desire to satisfy Equation (4.8), f_l should generally be less than 200 Hz and f_h should generally be greater than 1200 Hz.

---

[10] The curves of Figure 4-5(b) actually indicate a "corner" at 1000 Hz, but this is due to the fact that we have only predicted values for tones of 250, 500, 700, 1000, 1400, and 2000 Hz and then linearly interpolated between them. Undoubtedly, the actual corner is at 1200 Hz where $G(f)$ begins to decrease.

Predictions for the lateralization and detection experiments considered are shown in Figure 4-6. The solid lines in this figure were generated using $p_L(\tau)$ when $f_l = 50$ Hz and $f_h = 4000$ Hz. These values were selected to jointly minimize the discrepancies between predictions and data for both sets of data. Considering first the lateralization predictions in panel (a), we see that, at low frequencies, there is fairly good correspondence between the model's predictions and the experimental results. At higher frequencies, the theoretical curve does not exhibit as sharp a decrease in image displacement as do the data.

For the same distribution of internal delays, the solid lines of Figure 4-6(b) show the predicted ratio of detection thresholds $(\frac{N_\pi S_0}{N_0 S_\pi})$ as a function of target frequency. These curves indicate threshold differences which are larger than those exhibited by the data at almost all frequencies shown. In order to more accurately describe these data, $p_L(\tau)$ should indicate fewer fiber pairs with associated delays near zero. Towards this end, we propose "clipping" this function so that it is constant for delays of magnitude less than $200\mu s$. The modified density function $p_{L'}(\tau)$ is given by

$$p_{L'}(\tau) = \begin{cases} C_{L'}\, p_L(\tau_L), & \text{for } |\tau| \leq \tau_L \\ C_{L'}\, p_L(\tau), & \text{otherwise} \end{cases} \tag{4.13}$$

where τ_L equals $200\mu s$ and $C_{L'}$ is chosen so that $p_{L'}(\tau)$ has unit area. (Note that when τ_L is zero, this function reduces to $p_L(\tau)$.)

As the dashed lines in Figure 4-6(b) indicate, the "clipped" density function does indeed provide a better description of the detection data. Unfortunately, the lateralization predictions shown in panel (a) are not as accurate as those for the unmodified $p_L(\tau)$ (*i.e.*, τ_L equals zero). Predicted position begins to shift towards the center at only 500 Hz while actual judgements remain relatively constant out to 1200 Hz.

When examining predictions for various configurations of $p_{L'}(\tau)$ we found that, in general, the detection predictions were most sensitive to the parameter τ_L. Moderate changes in the values of f_l and f_h did not seem to drastically affect the ability of the model to describe these data. On the other hand, predictions for the lateralization experiment are closely related to the values of all three parameters. Furthermore, typical values of τ_L required to describe the detection data result in a poor description of the lateralization data.

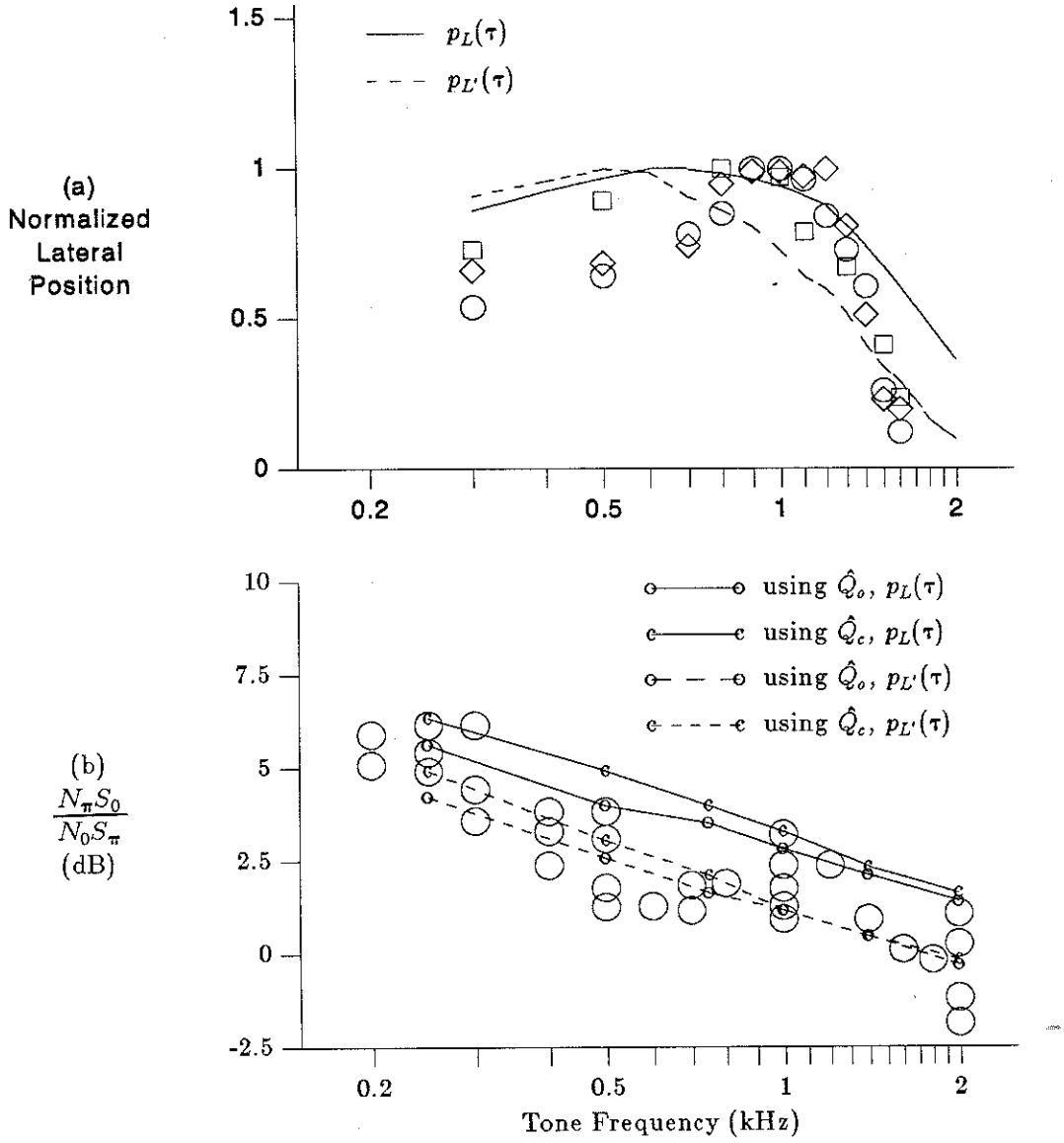

Figure 4-6: Predictions obtained using $p_L(\tau)$ and $p_{L'}(\tau)$ for (a) the lateralization of tones with an ITD of $150\mu s$ as a function of frequency and (b) the ratio of $N_\pi S_0$ thresholds to $N_0 S_\pi$ thresholds as a function of target frequency. The above predictions were generated under the assumptions that $\nu = 3$ and $fc = f_0$. [The lateral position data were reproduced from Figure 2-2, and the relative threshold data were reproduced from Figure 2-6.]

These observations suggest that by allowing the parameters f_l and/or f_h to depend on characteristic frequency, one might be able to jointly describe both sets of data. The modified density function $p_{LF}(\tau\,|\,fc)$ shown below has proven effective in this regard.

$$p_{LF}(\tau\,|\,fc) = \begin{cases} C_{LF}(fc)\,, & \text{for } |\tau| \leq 200\mu s \\ C_{LF}(fc)\ (e^{-2\pi f_l(fc)|\tau|} - e^{-2\pi f_h|\tau|})\ /\ |\tau|, & \text{otherwise} \end{cases} \tag{4.14}$$

where

$$f_l(fc) = \begin{cases} 0.1\ fc^{1.1}\,, & \text{for } fc \leq 1200 \text{ Hz} \\ 0.1\ (1200)^{1.1}\,, & \text{for } fc > 1200 \text{ Hz} \end{cases}\,, \tag{4.15}$$

$f_h = 3000$ Hz, and $C_{LF}(fc)$ is chosen so that $p_{LF}(\tau\,|\,fc)$ is a valid density function.

Although $f_l(fc)$ is nearly proportional to fc (for frequencies less than 1200 Hz), this expression does not describe a phase-based distribution since both τ_L and f_h are independent of characteristic frequency. This point is illustrated in Figure 4-7, which compares $p_{LF}(\tau\,|\,fc)$, evaluated at fc equal to 500 and 1000 Hz, with the original distribution proposed by Colburn (1977). As characteristic frequency increases, $p_{LF}(\tau\,|\,fc)$ indicates an increase in the relative number of fiber pairs with associated delays less than 200 μs (in magnitude) as compared to the number of pairs with delays larger than 200 μs. However, the distribution of delays between -200 μs and +200 μs remains constant, independent of characteristic frequency.

Predictions obtained using $p_{LF}(\tau\,|\,fc)$ for both experiments of interest are shown in Figure 4-8. The top panel indicates that the introducion of the frequency-dependent parameter $f_l(fc)$ has enabled us to accurately describe the lateralization data while maintaining τ_L equal to 200 μs. Consequently, the detection predictions of panel (b) are also in agreement with the experimental evidence.

Although the lateralization predictions (in panel (a)) do not indicate as strong a shift in image position with increasing frequency as do the data (above 1200 Hz), it is possible that a second mechanism other than the distribution of internal delays is responsible for this descrepancy between predictions and data. This issue is discussed in further detail in the following section.

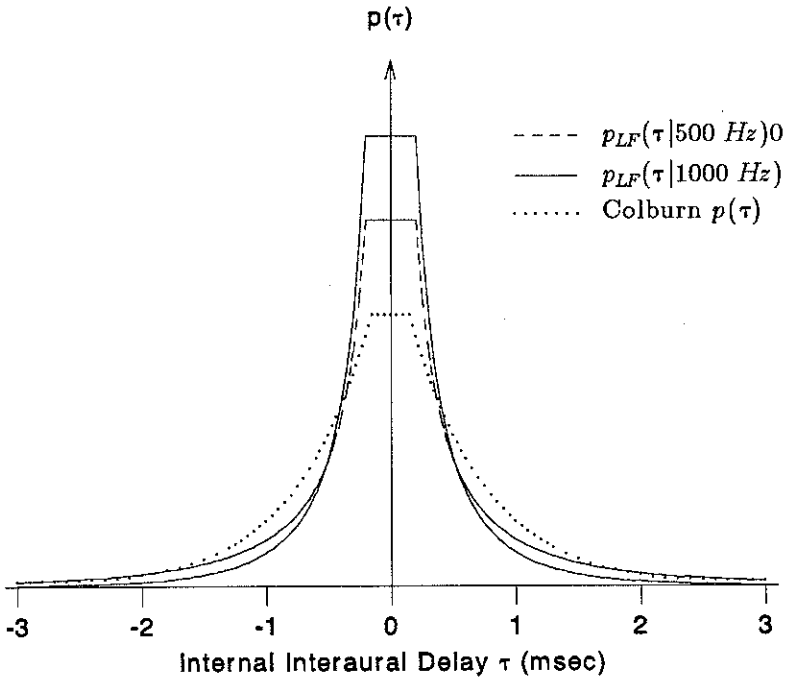

Figure 4-7: Comparison of the shapes of $p_{LF}(\tau|fc)$ and the original density function specified by Colburn (1977).

4.1.5. Summary and Conclusions

Two solutions to the lateralization constraints derived in Section 4.1.3 were considered. We found that phase-based distributions, in general, were incapable of jointly describing both sets of experimental evidence. Although satisfactory descriptions of the lateralization data could be obtained, predictions for $N_\pi S_0$ vs. $N_0 S_\pi$ detection thresholds did not exhibit as strong a dependence on target frequency as can be observed in the data. Attempts to improve predictions for these detection data degraded the ability of the model to describe the lateralization phenomena.

On the other hand, we were able to specify a log-based function $p_{LF}(\tau|fc)$ that allows the model to describe both the lateralization and detection phenomena. This function is similar in form to the original function proposed by Colburn (1977) except that (1) the tails of $p_{LF}(\tau|fc)$ decay more rapidly with respect to τ than the tails of the original function and (2) the rate at which these tails decay is dependent on characteristic frequency.

For both types of density functions considered, lateralization predictions for tones

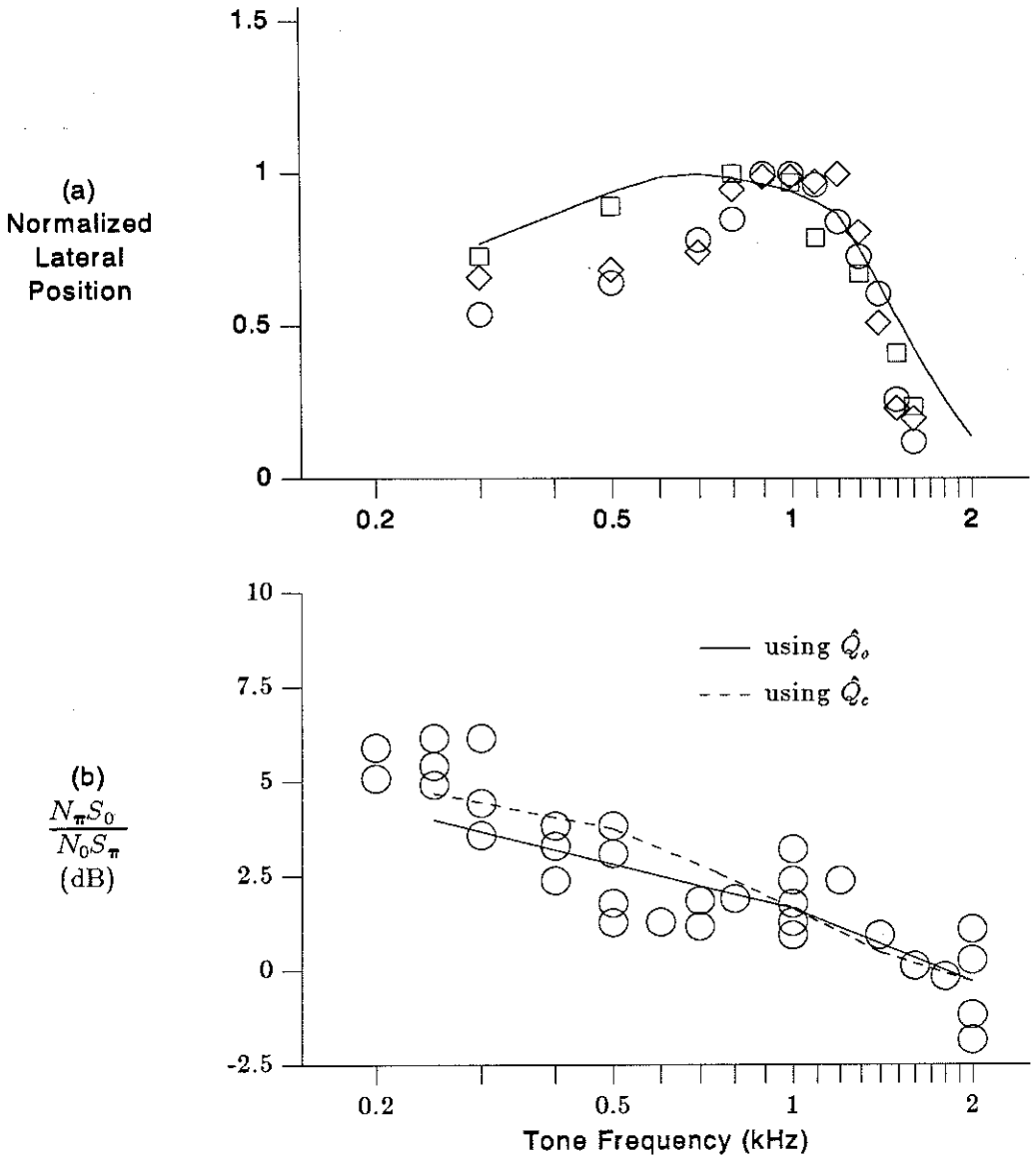

Figure 4-8: Predictions obtained using $p_{LF}(\tau|fc)$ for (a) the lateralization of tones with an ITD of 150 μs as a function of frequency and (b) the ratio of $N_\pi S_0$ thresholds to $N_0 S_\pi$ thresholds as a function of target frequency. The above predictions were generated under the assumptions that $\nu = 3$ and $fc = f_0$. [The lateral position data have been reproduced from Figure 2-2, and the relative threshold data have been reproduced from Figure 2-6.]

above 1200 Hz generally indicate a larger displacement of the binaural image from center than do the data. While it is possible to develop (fairly elaborate) forms of $p(\tau \mid fc)$ that can more accurately describe the sharp shift in image position above 1300 Hz, we feel that a more plausible cause for this discrepancy between predictions and data is the lowpass filter function of the auditory-nerve model, $G(f)$. As noted earlier, lateralization predictions for tones at high frequencies are strongly dependent on the shape of this function, which was specified to describe the loss of synchrony exhibited by nerve fibers in cats. Since the human auditory system is sensitive to a smaller range of frequencies than the auditory system of cats, it is reasonable to assume that the shape of $G(f)$ in humans and/or its stopband might also be more compressed with respect to frequency. Such a lowpass filter function (with a sharper stop band) would better describe the lateralization data at frequencies above 1300 Hz. Modifying $G(f)$ in this way should also allow the model to more accurately describe the observed inability of listeners to detect differences in interaural time delays at tone frequencies above 1300 Hz (Klumpp and Eady, 1956). While we believe that these observations warrant a further consideration of how the lowpass filter should be specified, we do not attempt such an investigation in this work. It is sufficient to note that the current model, at the very least, predicts the general trends of the data at high frequencies.

While Lindemann (1986) has made use of a phase based distribution, he has not published predictions for the lateralization or detection data considered here. In light of our results, we believe that Lindemann's model is not able to completely describe the lateralization data or detection data, and we suggest that the inclusion of this type of density function in his model (or any centroid-based model of binaural interaction) be re-considered.

4.2. Dependence of Predictions on the Rectifier Power

In Chapter 3, we demonstrated that the ν^{th}-law half-wave rectifier of the auditory-nerve model should be of order 1, 2, or 3 to best describe the maximum synchronization index for low frequency tones (as measured by Johnson, 1974). We now examine how predictions of the model for the lateralization and detection of tones are affected by choices of ν in this range. While the examples provided below are only for the distributions $p_{LF}(\tau \mid fc)$ and $p_{L'}(\tau)$, the trends discussed appear to be relatively insensitive to the exact shape of the density function.

(The predictions presented in this section were computed using only characteristic frequencies approximately equal to the tone frequency.)

4.2.1. Subjective Lateral Position

Figure 4-9 shows predictions for the lateralization experiment obtained using first, second, and third order half-wave rectifiers. The bottom panel was generated using $p_{LF}(\tau \mid fc)$ and the top panel was generated using $p_{L'}(\tau)$ (with τ_L = 200 μs). In both graphs, the model configuration employing a first-order rectifier best describes the data. As the rectifier becomes more expansive, the significance of harmonics greater than the fundamental frequency of $L_2(\tau)$ increases, and the predictions for tones of frequency above about 700 Hz begin to shift towards zero (since 150 μs is approximately one fourth the period of a 1400 Hz tone).

For the values of ν considered here, the extent to which the model's predictions degrade is negligible. However, for higher-order rectifiers (*e.g.*, ν greater than 5), the observed discrepancies could become significant. Thus, these data for the lateralization of tones do not further restrict the selection of the rectifier power ν; values between 1 and 3 all provide an accurate description of the data. (This does not imply that all lateralization predictions are unaffected by the order of the rectifying nonlinearity. As we will see in the next chapter, this parameter can significantly alter lateralization predictions for bandpass noise.)

4.2.2. Relative Detection Thresholds

Unlike the lateralization predictions for tones, the theoretical difference between $N_\pi S_0$ and $N_0 S_\pi$ detection thresholds is sensitive to the value of the rectifier power ν. However, the ways in which these predictions are affected by changes in ν depends on which decision statistic is used, $\hat{Q}_o$ or $\hat{Q}_c$.

Before discussing the actual threshold predictions, it is instructive to first examine the dependence of the timing displayer output on ν. For the stimulus conditions of interest, increasing the order of the nonlinearity has the following effects on the interaural crosscorrelation function $L_2(\tau, f_0)$. First, the peaks of this function increase in magnitude and become narrower with respect to τ. At the same time, the differences between the signal-present and signal-absent cases are magnified at the peaks. Consequently,

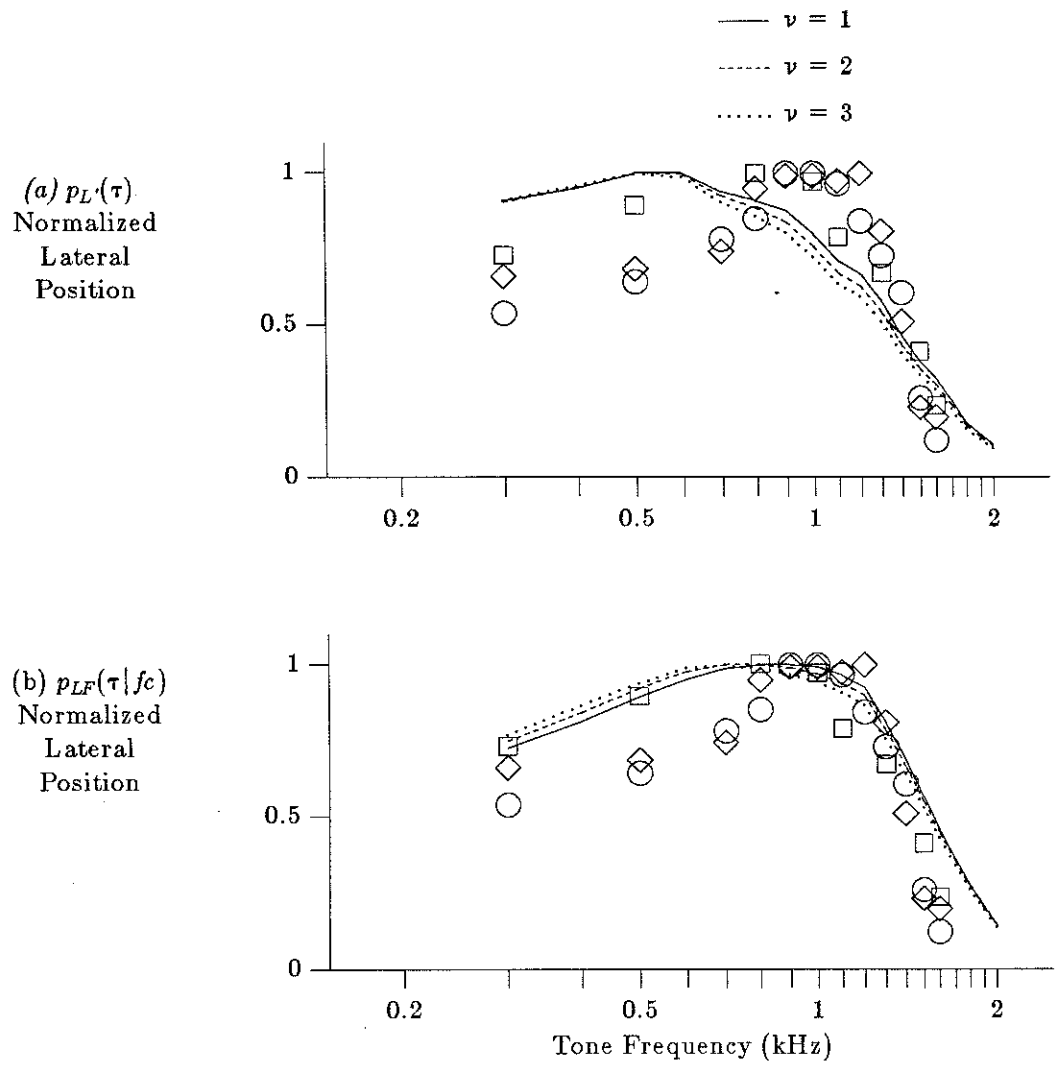

Figure 4-9: Predictions obtained using various rectifier configurations for the lateralization of tones with ITDs of 150 μs as a function of frequency. [The data points were reproduced from Figure 2-2.]

outputs of the binaural displayer with associated internal delays near $\frac{\phi_n \pm 2k\pi}{2\pi f_0}$ (where k is an integer) become the best indicators for detecting the presence of the tone.

For the optimal detection variable $\hat{Q}_o$, Figure 4-10(a) indicates that an increase in the rectifier power results in an increase in the predicted difference between the thresholds for the two interaural phase configurations (especially at low frequencies). This is caused

by the fact that for small values of ν, the differences which occur in the "valleys" of the interaural crosscorrelation functions are almost as signficant as the differences which occur at the peaks. Subsequently, the effect of $p(\tau \mid fc)$ in computing the decision statistic is not strong. However, since larger values of ν cause the peaks to dominate the detection process, the distribution of internal delays does become significant and predicted N_0S_π performance improves relative to $N_\pi S_0$ performance.

Figure 4-10(b) illustrates that the opposite dependence on ν is observed if we assume that the simpler, constant-coefficient variable $\hat{Q}_c$ is used. The difference between thresholds for the two stimulus configurations is predicted to decrease for larger rectifier powers. It is not obvious as to why $\hat{Q}_c$ should behave oppositely to $\hat{Q}_o$ in this respect, and presently we are unable to provide an answer to this question.

For most of the density functions examined, setting ν equal to three yielded similar predictions using either definition of the correlation variable. Since we are not aware of any evidence that suggests which definition of $\hat{Q}$ should be adopted, it is convenient to designate one set of model parameters (ν and $p(\tau \mid fc)$) that yields similar predictions for both $\hat{Q}_o$ and $\hat{Q}_c$.

4.2.3. Conclusions

The observations described above do not provide a strong indication of the value of ν that is most likely to be "correct". It is generally accepted that the peripheral transformation is somewhat expansive and therefore we basically reject the use of a half-wave linear rectifier (cf. Kiang *et al.*, 1965; Kiang, 1968; Johnson, 1974). For the reasons stated in the previous paragraph, we prefer using a third-order device, but this is mainly an issue of convenience. In the next chapter we present psychoacoustical evidence, based on data for the lateralization of bandpass noise, that suggests that a more expansive rectifier should possibly be adopted.

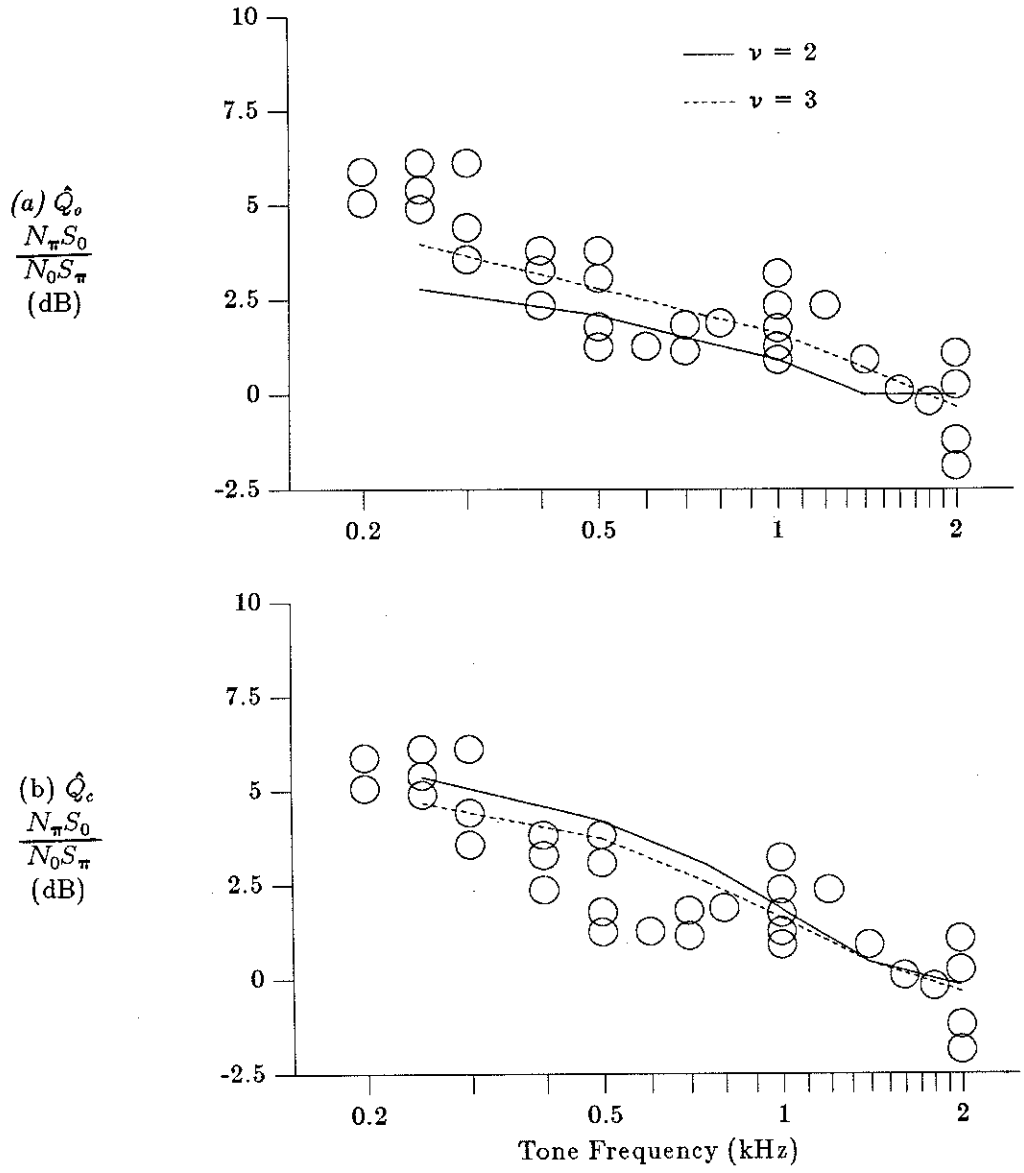

Figure 4-10: Predictions for the ratio of $N_\pi S_0$ detection thresholds to $N_0 S_\pi$ detection thresholds as a function of frequency obtained using $p_{LF}(\tau \mid fc)$, (a) $\hat{Q}_o$, and (b) $\hat{Q}_c$.

4.3. Dependence of Predictions on the Range of Characteristic Frequencies

Thus far, all predictions presented have been generated under the assumption that only fiber pairs with characteristic frequencies equal to f_0 are significant in the formation of the various decision statistics. We now examine the validity of this assumption for the two tasks considered and describe how other choices for $\mathbf{R}_{task}$ affect predictions.

4.3.1. Subjective Lateral Position

There are two factors that determine the extent to which the specification of $\mathbf{R}_{lat}$ affects relative lateralization predictions. The first is the fraction of fiber pairs considered by the central processor that are doubly active. As this fraction decreases, the predicted position of the sound image shifts towards the center of the head. This result is reasonable since spontaneously firing fibers do not contribute any information regarding the ITD present in a binaural stimulus. The second factor is the dependence of the distribution of internal delays on characteristic frequency. The ways in which these two factors influence relative lateralization predictions for tones depends on how $\mathbf{R}_{lat}$ is specified.

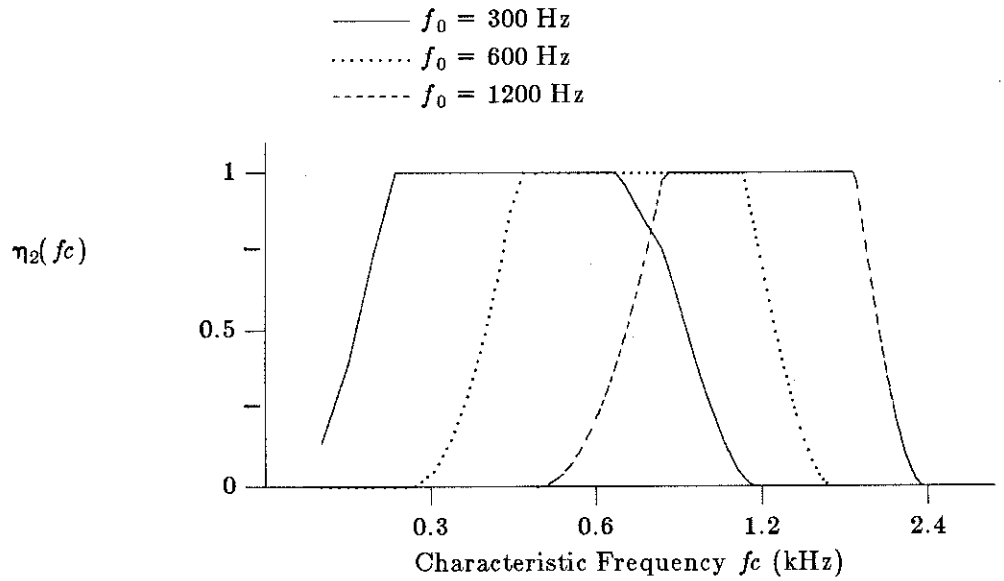

Figure 4-11: $\eta_2(fc)$ is plotted as a function of fc for tones of various frequencies presented at 70 dB SPL.

In order to examine the relative contribution from each factor described above, we consider the following three ways of defining the range of characteristic frequencies to be considered.

1. Only characteristic frequencies approximately equal to the tone frequency f_0 are considered.

2. Only those characteristic frequencies for which $\eta_2(fc)$ is equal to one or is at its maximum are considered.

3. All characteristic frequencies within a fixed, broad range are considered. For now we arbitrarily designate this range to be 190 Hz to 2500 Hz.

For the signal presentation level (70 dB SPL) considered throughout this chapter, our assumption that $fc \approx f_0$ further implies that all fiber pairs considered are doubly active. The second case also insures that this will be true, but, as Figure 4-11 indicates, the characteristic frequencies of fibers considered are no longer necessarily equal to f_0. Thus, this case can be used to isolate the effects of $p(\tau \mid fc)$ on relative position predictions. Predictions generated using the final specification given above should include effects from both factors being studied when the distribution of internal delays is assumed to be frequency dependent, and they should include effects from the fraction of doubly-active fiber pairs alone if the distribution is independent of characteristic frequency.

We now consider the effects of the fraction of doubly-active fiber pairs alone by using the frequency independent distribution $p_{L'}(\tau)$. Figure 4-12(a) shows predictions for the lateralization experiment obtained using the three definitions of $\mathbf{R}_{lat}$ described above. As we would expect, there is no difference between the predictions for the first two specifications of $\mathbf{R}_{lat}$ (since $\eta_2(fc)$ is equal to one in both cases). For the fixed range of characteristic frequencies, there is only a small change in the shape of the theoretical curve as compared to the other two cases. This indicates that the fraction of doubly-active fiber pairs with characteristic frequencies between 190 Hz and 2500 Hz does not vary significantly with tone frequency.

The effects of $\mathbf{R}_{lat}$ on predictions obtained using a phase based distribution function are demonstrated in Figure 4-12(b). These predictions were generated using $p_G(\tau \mid fc)$ with $\gamma = 1.2$ and $C_\sigma = 1.2$. It is apparent from the difference between the curves for the $fc = f_0$ case and the $\eta_2(fc) = 1$ case that predictions are significantly affected by broadening the range of characteristic frequencies considered. This is understandable since a key assumption in the derivation of the phase constraint is that only characteristic frequencies

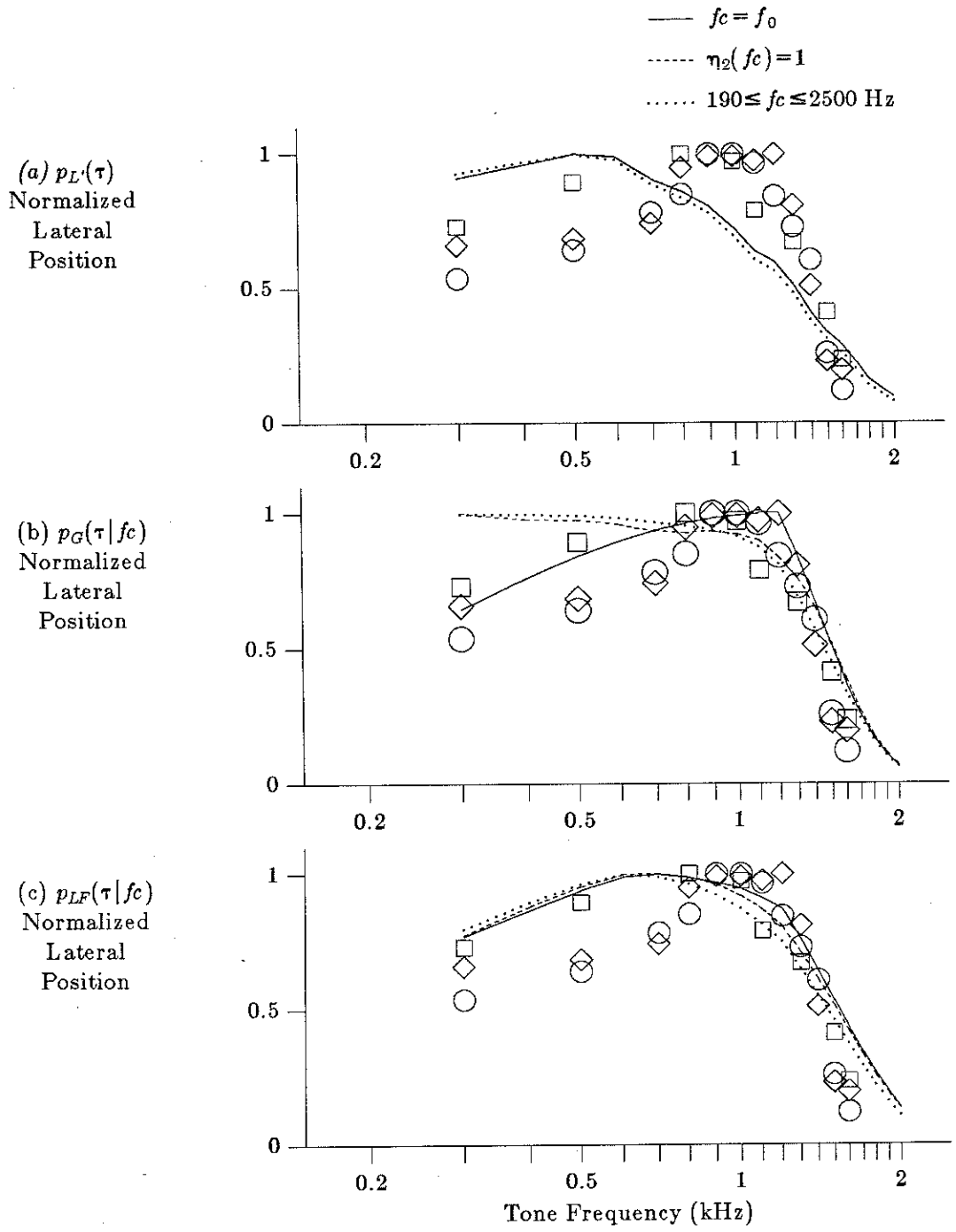

Figure 4-12: Predictions for the lateralization of tones with an ITD of 150 μs obtained using various specifications of $\mathbf{R}_{lat}$ and (a) $p_{L'}(\tau)$, (b) $p_G(\tau|fc)$ (with $\gamma = C_\sigma = 1.2$), and (c) $p_{LF}(\tau|fc)$.

nearly equal to f_0 are considered. The magnitude and direction of the change in predictions caused by broadening $\mathbf{R}_{lat}$ is not necessarily the same for all phase based distributions; these properties depend on the particular shape of $p(\tau \mid fc)$. Comparing now the predictions for the fixed range of characteristic frequencies and the $\eta_2(fc) = 1$ case, we note (once again) that the effects due to variations in the fraction of doubly-active fiber pairs are very small.

Finally, Figure 4-12(c) shows lateralization predictions obtained using $p_{LF}(\tau \mid fc)$. These results indicate that the frequency dependence exhibited by this function does cause some decrease in the ability of the model to accurately describe the data as $\mathbf{R}_{lat}$ is broadened, but the degree of this decreased performance is not very significant.

The findings presented above indicate that for distributions that exhibit only a moderate (or no) dependence on characteristic frequency, the lateralization predictions are fairly insensitive to the specification of $\mathbf{R}_{lat}$.[11] However, model predictions obtained using distributions that are strongly dependent on characteristic frequency (*e.g.*, phase based distributions) can be significantly affected by the range of characteristic frequencies considered.

4.3.2. Relative Detection Thresholds

Unlike $\mathbf{R}_{lat}$, the range of characteristic frequencies considered by the central processor for detection is well specified by the defining assumptions of the model. Assumption **DM3** (on page 46) states that this range should be chosen to provide optimal performance for the given task. If the optimal correlation variable $\hat{Q}_o$ is used, this range includes all outputs of the binaural displayer since the coefficients $\{c_m\}$ allow the central processor to simply ignore those units which provide no useful information. However, we have found that only those fiber pairs with characteristic frequencies within ± 75 Hz of f_0 contribute significantly to improving performance. This result is understandable in light of the curves shown in Figure 4-13, which indicate that the target-to-masker ratio decreases sharply for fiber pairs with characteristic frequencies not equal to f_0. Similar observations have

---

[11] This statement strictly only applies to the definitions of $\mathbf{R}_{lat}$ considered here. However, it can be generalized to other such definitions provided that they treat tones of various frequencies somewhat "equally". Obviously, defining $\mathbf{R}_{lat}$ to be $\{fc \mid fc \approx 1200 \text{ Hz}\}$ will result in predictions that exhibit the greatest lateral image displacement for tones of frequency near 1200 Hz, (almost) independently of the form assumed for $p(\tau \mid fc)$.

been made regarding the optimal range for $\hat{Q}_c$, with one exception; performance actually begins to degrade if $\mathbf{R}_{det}$ is made too broad. This is due to the fact that all outputs of the binaural displayer are weighed equally in the formation of $\hat{Q}_c$ and that only those units with characteristic frequencies near f_0 are actually useful.

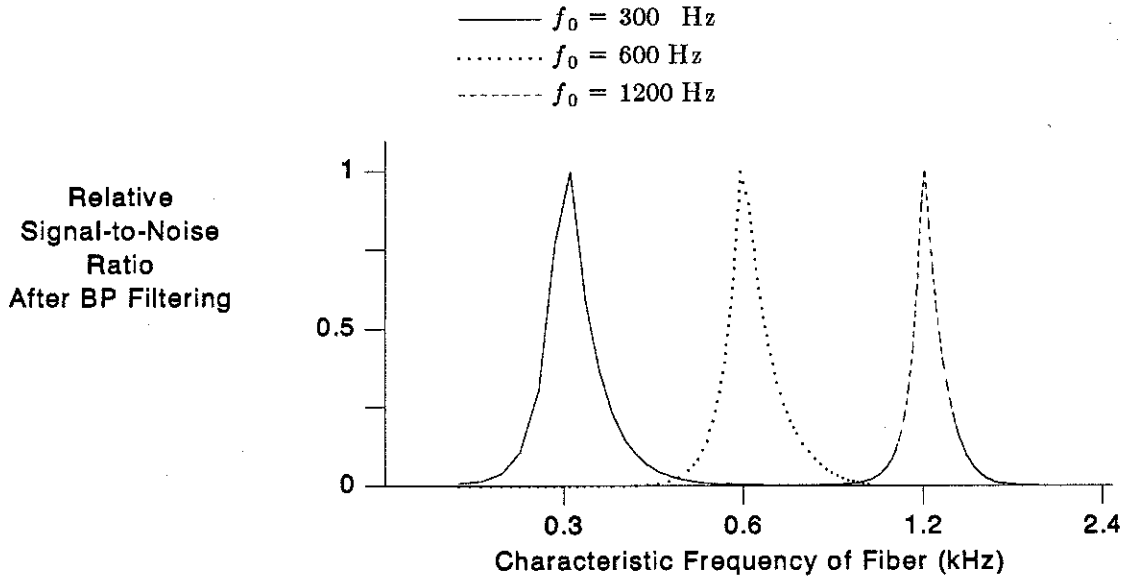

Figure 4-13: Relative target-to-masker ratios computed after peripheral bandpass filtering are shown as a function of the characteristic frequency of the peripheral filter for 300-Hz, 600-Hz, and 1200-Hz target tones.

For both $\hat{Q}_o$ and $\hat{Q}_c$, we have found that predictions for the differrence between $N_\pi S_0$ and $N_0 S_\pi$ detection thresholds obtained using only characteristic frequencies nearly equal to f_0 differ by only a fraction of one dB from predictions obtained using the optimal range of frequencies $\mathbf{R}_{det}$. This discrepancy is much less than the measured standard deviation of the data ($\sim$ 2 dB) and therefore we consider this approximation acceptable and have adopted it to expedite relative detection threshold computations.

4.4. Predictions for Binaural Detection Thresholds as a Function of Frequency

Up to this point, we have only considered how target frequency affects differences between $N_\pi S_0$ and $N_0 S_\pi$ detections thresholds, and we have ignored how this stimulus paramater affects the overall detectability of a tone. This is due in part to the fact that our predictions for $N_0 S_\pi$ thresholds as a function of frequency have proven to be almost entirely independent of the specification of ν, $p(\tau|fc)$, or the selection of $\hat{Q}$. These predictions are depicted by the solid line of Figure 4-14, and they indicate that the model fails to correctly describe the data at frequencies above about 500 Hz. As we have already noted, this failure of the model is apparently not related to any of the model components which we have studied, and we presently do not attempt a thorough investigation of the problem. Instead, we merely discuss a few preliminary observations which should prove useful in the future evolution of the model.

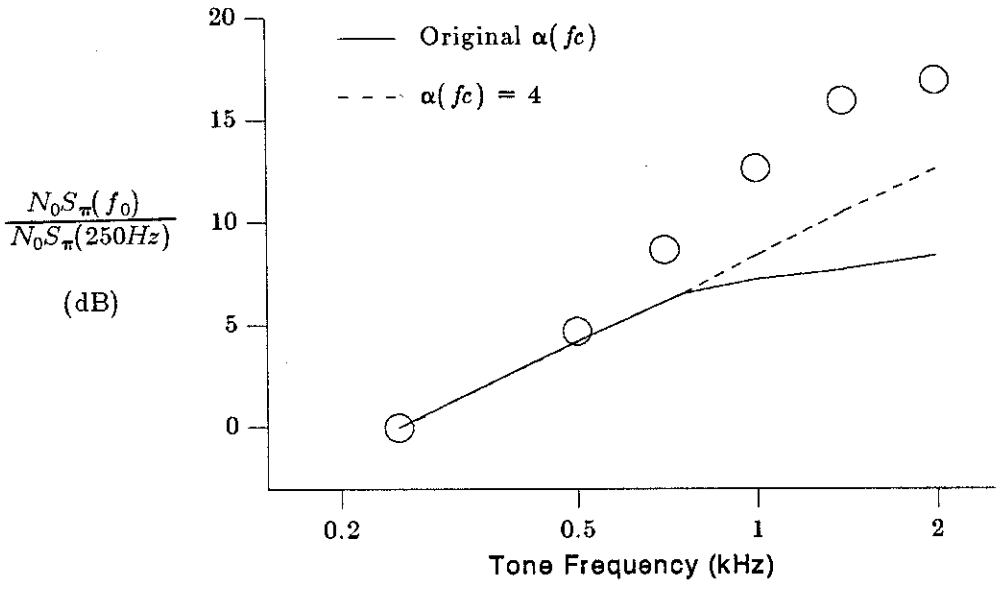

Figure 4-14: $N_0 S_\pi$ detection thresholds are plotted as a function of frequency relative to the $N_0 S_\pi$ threshold for a 250-Hz target. Predictions are shown for two bandpass filter configurations. The data have been taken from Hirsch and Burgeat, 1958.

A possible source of the discrepancy between data and predictions is the

characterization of the tuning curves of auditory-nerve fibers (given by Equation (3.1)). Consider for the moment a single fiber with characteristic frequency equal to the tone frequency. If we increase the bandwidth of the tuning filter associated with this unit, the contribution of neural activity from the noise masker increases with respect to the contribution from the target tone. Consequently, the level of tone required to achieve threshold performance should also increase. In order to test this hypothesis, we have generated predictions under the assumption that $\alpha(fc)$ of Equation (3.1) is equal to 4 at all frequencies. This modification causes the bandpass filters with characteristic frequencies above 800 Hz to broader than the original filters. Predictions obtained using the modified filters are plotted as the dashed line in Figure 4-14, and they do indicate an increase in threhold level for frequencies above 800 Hz. This result is consistent with our comments regarding the filter bandwidth.

Although the new predictions do more closely follow the trend of the data, we do not suggest that this particular modification to the model be adopted. Instead, we feel that it would be better to completely respecify $H_m(f)$ based on more recent physiological data (*e.g.*, Kiang and Moxon, 1974) which indicate that the low frequency tails of the tuning curves are not necessarily as steep as those indicated by $H_m(f)$. Broader filters of this type would presumably allow the model to describe the data of Figure 4-14 more accurately. (As we will see in Chapter 5, this modification would also slightly improve performance in describing the lateralization of bandpass noise.)

4.5. Summary

In this chapter we have considered the ability of the modified position-variable model to describe two types of psychoacoustical phenomena involving tones: the lateralization of tones as a function of frequency, and the difference between $N_\pi S_0$and $N_0 S_\pi$ detection thresholds as a function of frequency. The dependence of the model's predictions for these experiments on the distribution of internal delays, the order of the rectifier, and the range of characteristic frequencies $\mathbf{R}_{task}$ was discussed.

A set of mathematical constraints which must be satisfied by a centroid-based model in order to describe the lateralization of tones as a function of frequency was derived. Phase-based density functions satisfy this constraint if $\mathbf{R}_{lat}$ is narrow about the tone frequency. Functions of this type are capable of providing a good description of the

lateralization data, but they yield predictions for the relative detection threshold experiment that are not in agreement with the data. A more restricted class of density functions, referred to as log-based distributions, satisfies the lateralization constraint independently of the specification of $\mathbf{R}_{lat}$. At least one function of this type, $p_{LF}(\tau|fc)$, allows the model to describe both the lateralization and detection data considered.

We pointed out in Chapter 3 that in order to accurately characterize data describing the instantaneous firing rate of auditory-nerve fibers stimulated by pure tones, the order of the model's rectifier ν should be in the range 1 to 3. The results presented in this chapter indicate that the lateralization and detection data of interest do not further constrain the value of ν within this range. We have decided to use a value of three for convenience, since predictions for the detection data are similar using both correlation variables, $\hat{Q}_c$ and $\hat{Q}_o$.

If the specified distribution of internal delays is not strongly dependent on characteristic frequency, then the predictions of the model for the lateralization of tones are fairly insensitive to the range of characteristic frequencies $\mathbf{R}_{lat}$ considered by the central processor. In general, we consider the restriction that only doubly-active fiber pairs be used by the decision mechansism to be a reasonable assumption since little or no timing information can be extracted from other regions of the correlation display.

Predictions for the relative detection threshold data considered are also somewhat insensitive to the range of characteristic frequencies considered. In particular, predictions obtained considering only fibers for which $fc \approx f_0$ are almost identical to predictions obtained using the optimal range of characteristic frequencies.

Finally, we note two issues that should be addressed by future studies. First, lateralization predictions suggest that the lowpass filter function $G(f)$ should possibly exhibit a sharper cut-off above 1200 Hz. Psychoacoustic data describing the inability of listeners to discriminate between high-frequency tones on the basis of ITD (*e.g*, Klumpp and Eady, 1956) should be helpful in considering this issue. Second, the theoretical predictions for N_0S_π thresholds as a function of frequency appear to be sensitive to the specification of the peripheral bandpass filters of the auditory-nerve model. The tendency of the model to underestimate the values of thresholds for targets above about 500 Hz suggests that these filters should possibly be "broader" for characteristic frequencies above 500 Hz.

Chapter 5
Lateralization of Bandpass Noise

The original position-variable model succesfully describes nearly all available lateralization data for 500-Hz tones. In the previous chapter, we concerned ourselves with how this model might be modified to describe phenomena for tonal stimuli at other frequencies as well. As we have seen, tonal stimuli produce a display of interaural crosscorrelation which is independent of characteristic frequency (over the significant region of the display), and the major issue in describing the related phenomena is specifying how this display should be "windowed" with respect to interaural time in order to obtain consistent ITD estimates for different stimulus frequencies. Thus, one can view the work in Chapter 4 as extending the current theory so that the independent processing of information at different frequencies is consistent with (certain) observed phenomena.

In this chapter, we present a cursory look at the problem of how information from different frequency bands might be processed simultaneously in order to describe observed behavior. Towards this end, we consider the lateralization of bandpass noise, which produces an interaural correlation display that is (usually) dependent on characteristic frequency. It is not the purpose of this investigation to rigorously develop a set of modifications to the model that enable it to quantitatively describe the phenomena considered. (This would require both more experimental evidence and time than are currently available.) Instead, we merely attempt to gain insight into the types of modifications that might be helpful in describing the lateralization of broadband stimuli.

In the next section, we describe the motivations for this study. In addition, we consider the interaural correlation display produced by the stimuli of interest in order to identify those attributes of the display that are important to the formation of "correct" lateral position judgements. We then examine how the position-variable model can be modified to exploit these attributes.

5.1. Motivation

Recently, Bernstein and Trahiotis noted that the subjective lateral position of an interaurally delayed band of Gaussian noise may depend strongly on the bandwidth of the noise. For example, we saw in Section 2.1.1.2 that a band of noise centered at 500 Hz with an ITD of 1.5 ms is perceived on one side of the head for narrow bandwidths, but moves to the other side as the bandwidth is increased to only 400 Hz (see Figure 5-1 below).

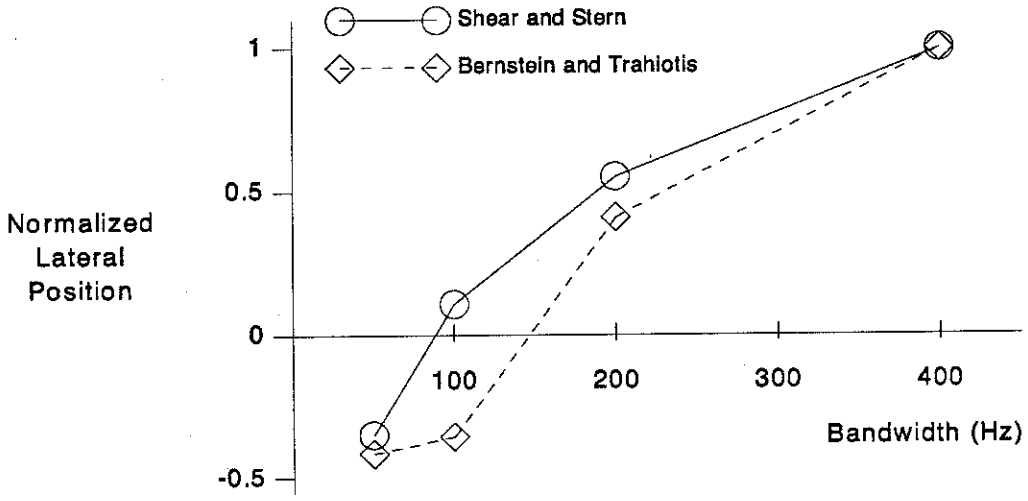

Figure 5-1: Relative lateral position of bandpass noise centered at 500 Hz with an ITD of 1.5 ms is plotted as a function of bandwidth. [Unpublished, normalized data of Bernstein and Trahiotis and Shear and Stern.]

In retrospect, it is easy to understand why this phenomenon is observed. For very narrow bandwidths the noise source closely resembles a pure tone of 500 Hz. Consider a 500-Hz tone presented with the signal to the right ear time delayed relative to the signal to the left ear by 1.5 ms. This stimulus is equivalent to one in which the signal to the right ear is leading the signal to the left ear by 0.5 ms (since the period of a 500-Hz tone is 2 ms). Consequently, the image of the binaural sound is perceived toward the right side of the head. As bandwidth is increased, however, the signal becomes less and less tonelike. In the limiting case, one is presented with broadband noise with the signal to the left ear leading the signal to the right ear by 1.5 ms, and this stimulus would certainly be

lateralized toward the left (or "leading") side of the head. This raises the question of how the auditory system is able to determine that the signal to the left ear is leading the signal to the right ear by 1.5 ms when the signal is sufficiently broad in bandwidth, even though it is "fooled" when the bandwidth is smaller.

One possible answer to this question is easily understood in terms of the current model of binaural interaction. Consider Figure 5-2 which depicts the mean of the internal correlation display $L(\tau, fc)$ elicited by the noise stimulus described above when the bandwidth is 50 Hz. As was previously noted, it is difficult to determine which peak of the correlation function is associated with the actual value of time delay present in the binaural stimulus. When the surface of Figure 5-2 is weighted by $p(\tau|fc)$ and its centroid along the τ-axis is computed, the most central peak (*i.e.*, the one at -0.5 ms) dominates the calculation. In this case, this peak does not represent the true ITD, and the image is "incorrectly" perceived on the right side of the head.

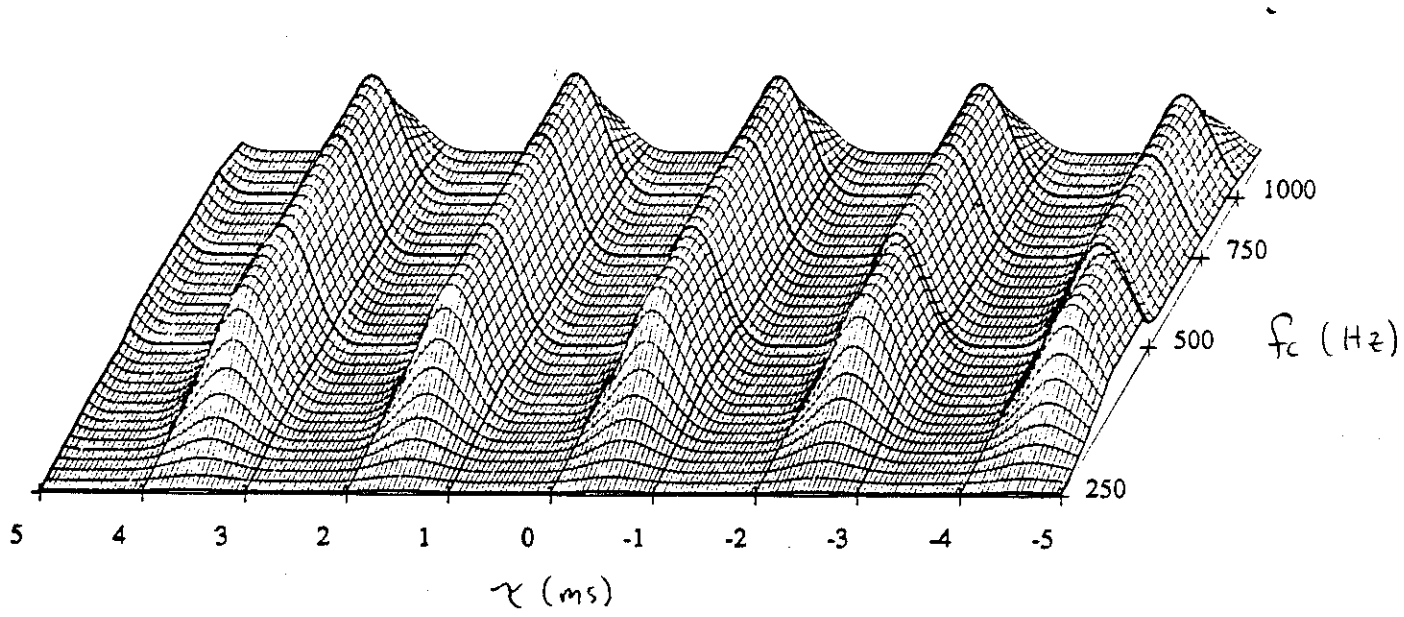

Figure 5-2: Plot of the internal correlation function $L(\tau, fc)$ elicited by a 50 Hz band of noise centered at 500 Hz with an ITD of 1.5 ms.

Figure 5-3 shows $L(\tau, fc)$ when the bandwidth of the input noise is 400 Hz. The true value of the ITD is more readily identifiable in this case for two reasons. First, the peaks of

$L(\tau, fc)$ that occur at τ = 1.5 ms (the true ITD) are larger than the local maxima that occur at other values of τ. (This is most evident in Figure 5-3 if one considers the contour of the crosscorrelation surface at high characteristic frequencies.) Second, the correlation function $L(\tau, fc)$ is independent of characteristic frequency for values of τ near the actual ITD. In other words, $L(\tau, fc)$ is "consistent" across frequency for interaural delays near the actual ITD.

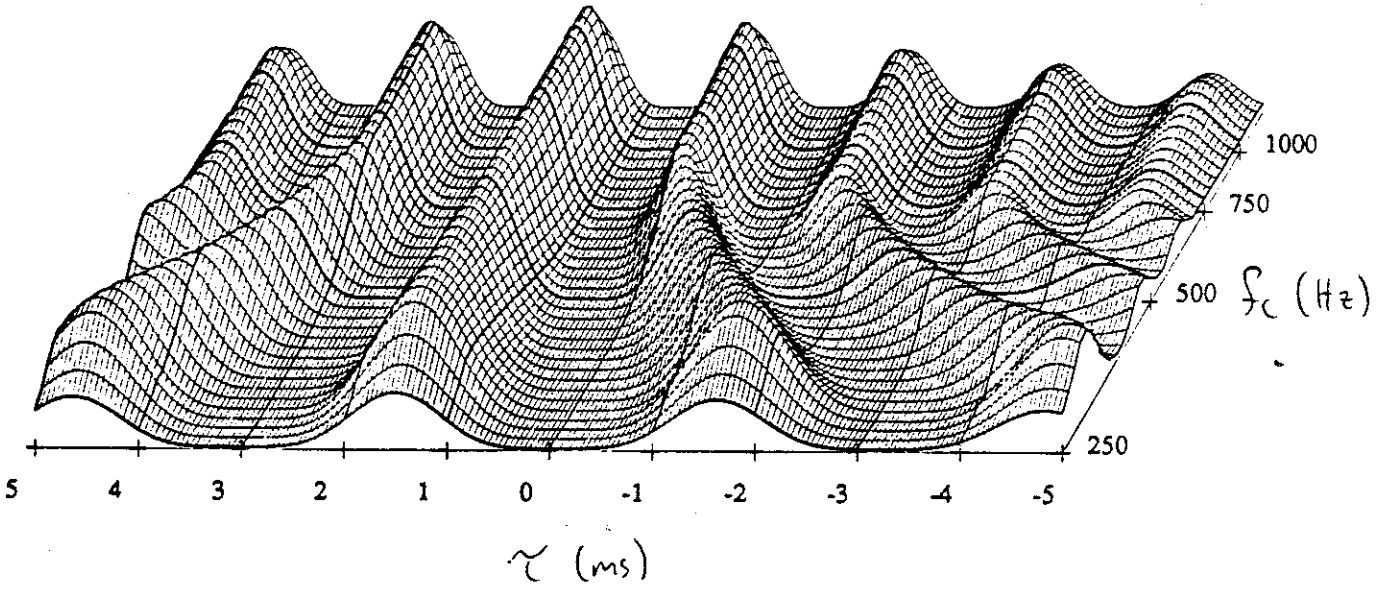

Figure 5-3: Plot of the internal correlation function $L(\tau, fc)$ elicited by a 400 Hz band of noise centered at 500 Hz with an ITD of 1.5 ms.

Based on the above observations, we can now generalize the changes which occur in $L(\tau, fc)$ as the bandwidth of the input noise is increased. For narrowband stimuli, all peaks of $L(\tau, fc)$ have nearly the same magnitude and $L(\tau, fc)$ is independent of characteristic frequency. As bandwidth increases, the secondary peaks of $L(\tau, fc)$ decrease in magnitude with respect to the "true" peak. In addition, $L(\tau, fc)$ becomes dependent on characteristic frequency, except for values of internal delay at (or near) the actual ITD. Figures 5-4 and 5-5 help to illustrate these points.

The current definition of the position variable $\hat{P}$ (given by Equation (3.10)) does exploit the changes in relative peak height that occur with increases in bandwidth, and lateral position predictions for the stimuli of Figure 5-1 do indeed shift towards the left side of the head as bandwidth increases. However, as Figure 5-6 indicates, the extent of the predicted shift in image position is generally much less than the shift exhibited by the data. One reason for this discrepancy may be the distribution of internal delays. If the distribution is assumed to be broad relative to 1.5 ms, then the model does predict a large shift in position as bandwidth increases. But if a narrow distribution is assumed, the shift is fairly small and may not cross the midline. This observation is demonstrated by the differences between the two sets of predictions in Figure 5-6.[12] The solid line was generated using $p_{LF}(\tau \,|\, fc)$. This function is, on the average, "narrower" than $p_{L'}(\tau)$, which was used to generate the dashed line.

While it is important to recognize that the model's predictions do depend on the shape of $p(\tau \,|\, fc)$, it is also worthwhile to consider how the model might be modified to take better advantage of the cues available in broadband stimuli.

In the next two sections we examine several modifications to the model which help to improve the accuracy of its ITD estimate using these cues. Two types of modifications are considered: those to the model of peripheral transduction and those to our assumptions regarding the formation of the position variable.

Predictions presented in these sections have been generated using the density function $p_{LF}(\tau \,|\, fc)$ and a rectifier power ν equal to 3, unless otherwise indicated. Calculations were carried out over only those characteristic frequencies for which all fiber pairs are doubly active.

---

[12] These predictions were generated using a rectifier power $\nu = 3$, and the calculations were computed over only those characterisitic frequencies for which all fiber pairs are doubly-active (*i.e.* $\eta_2(fc) = 1$).

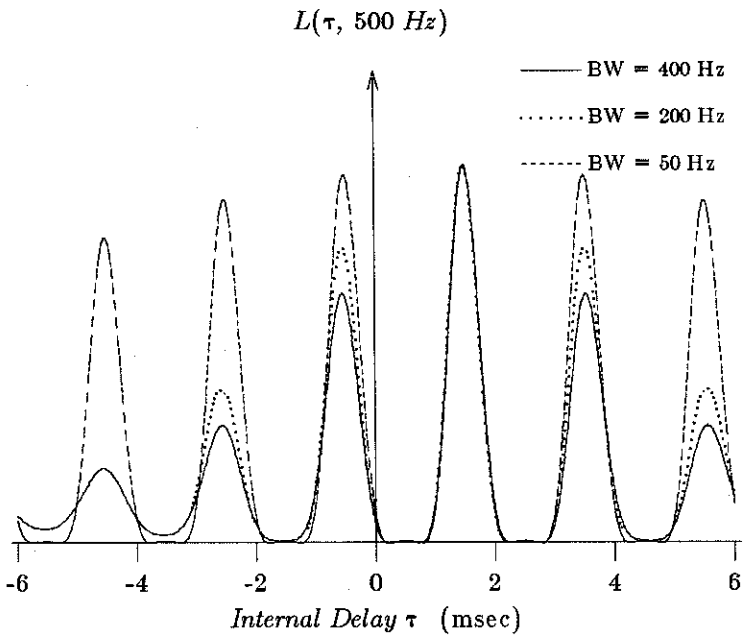

Figure 5-4: Plot of the internal correlation function $L(\tau, fc)$ at a characteristic frequency of 500 Hz. The stimulus used is bandpass noise centered at 500 Hz with an ITD of 1.5 ms and various bandwidths.

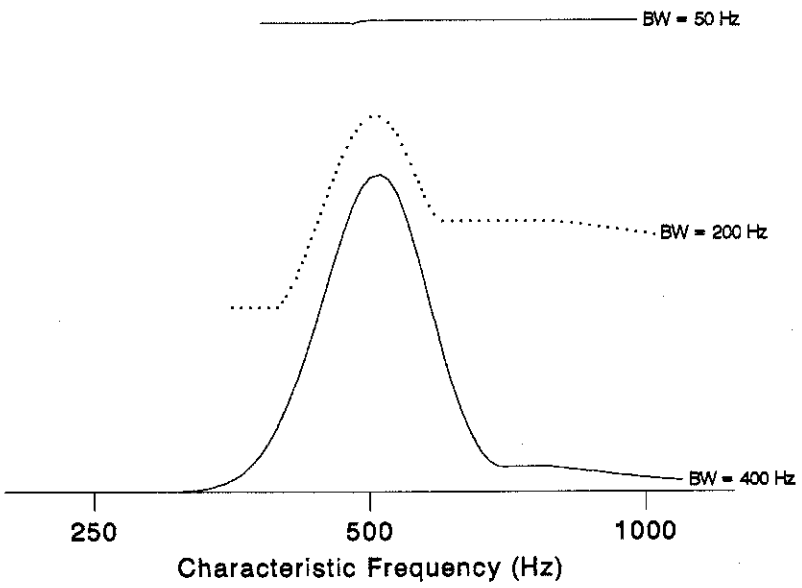

Figure 5-5: Plot of the internal correlation function $L(\tau, fc)$ when τ = -0.5 ms. The stimulus used is bandpass noise centered at 500 Hz with an ITD of 1.5 ms and various bandwidths.

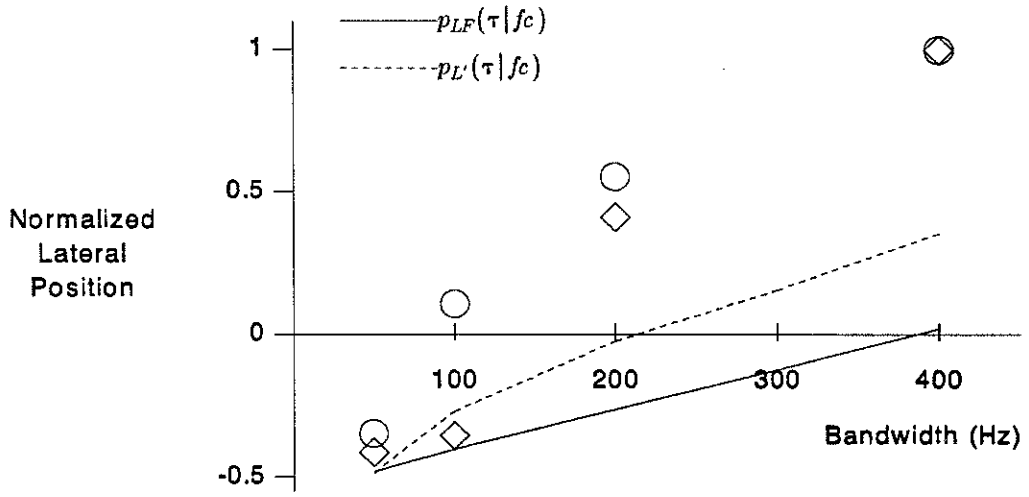

Figure 5-6: Predictions of the new PV model for the data of Figure 5-1 obtained using $p_{LF}(\tau|fc)$ (solid line) and $p_{L'}(\tau)$ (dashed line) with ν= 3. (Circles represent unpublished data of Shear and Stern, and diamonds represent unpublished data of Bernstein and Trahiotis.)

5.2. Modifications to the Auditory-Nerve Model

The purpose of this discussion is to examine how the auditory-nerve model might be modified in order to amplify the attributes of the interaural correlation display that are associated with values of τ near the ITD of the stimulus. This approach is attractive because it has the potential of improving the ability of the model to describe the lateralization of bandpass noise while still using the simple, centroid-based position variable. One is restricted, however, in the extent to which the auditory-nerve model can be modified since these changes will also affect earlier predictions for both physiological and psychoacoustical phenomena.

Since the auditory-nerve model processes input waveforms in parallel with respect to characteristic frequency, it is not possible to explicitly make use of the attribute of consistency across characteristic frequency. It is possible, however, to increase the amplitude of the true peak relative to the other peaks at each characteristic frequency. In Section 2.1.1.2 we saw that the crosscorrelation function for bandpass noise can be expressed as the product of an "envelope" times a sinusoid. As the bandwidth of the

noise increases, the envelope becomes sharper about the true ITD, and the other peaks of the overall correlation function decrease in magnitude with respect to this primary peak. A similar analysis can be applied to the peripherally transformed waveforms, and the final result is that in order to accentuate the primary peak of the correlation function, we wish to increase the bandwidth of the intensity function for each fiber.

5.2.1. Bandwidth of the Peripheral Filters

The most obvious mechanism for increasing the bandwidth of an intensity function $r_m(t)$ is to broaden the spectral response of the associated bandpass filter $H_m(f)$. Figure 5-7 shows predictions for the data of interest when the constant factor 4 in Equation (3.2) is replaced by 0, 2, and 4. As the value of this factor decreases, the effective bandwidth of the peripheral filters increases, with $\alpha(fc)$ equal to 0 corresponding to all-pass filters. As the results indicate, the predictions of the model are sensitive to this parameter and the predictions generated using broader peripheral filters do more closely resemble the data. However, such drastic changes in the shape of the peripheral filters conflict with all physiological evidence. While minor modifications to the filters might be acceptable from a physiological standpoint, they would not significantly affect the lateralization predictions for bandpass noise. Thus, this does not appear to be a viable solution to the problem.

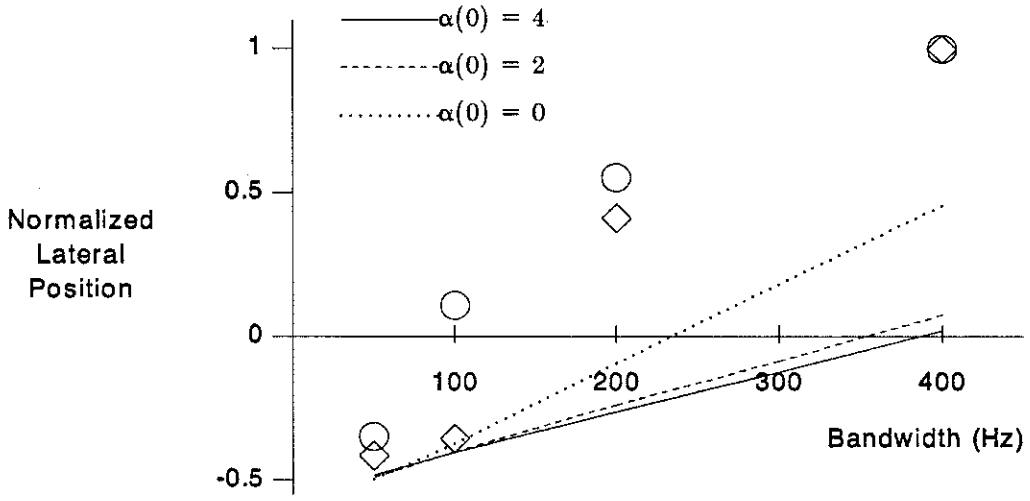

Figure 5-7: Predictions of the PV model for the data of Figure 5-1 generated using various peripheral bandpass filter configurations (see text).

5.2.2. Expansiveness of the Rectifier

Another means of increasing the bandwidth of the intensity process is to increase the order of the nonlinear rectifier. The autocorrelation function of the output of such a device can be expressed as a power series of the input autocorrelation function (see Section B.1). The n-th term of such a series corresponds to an n-fold convolution (in the frequency domain) of the input spectral density function. Thus, as the device is made more expansive, the coefficients of the higher-order terms in the power series become larger and the spectral density of the output process becomes broader.

Expansive nonlinearities also tend to sharpen the peaks and broaden the valleys of the interaural correlation function. If we consider $L(\tau, fc)$ for a given value of τ equal to τ_i (as in Figure 5-5) the effect of increasing the expansive nature of the rectifier will be to reduce the area under $L(\tau_i, fc)$ with respect to the area under $L(\tau_n, fc)$, where τ_n is the value of ITD present in the stimulus. Thus, when $\hat{P}$ is generated the relative weighting of the true ITD, eqaul to τ_n, compared to an arbitrary delay τ_i is increased. This is especially true if the means of generating the position variable explicitly takes into account the degree of consistency across frequency exhibited by $L(\tau, fc)$ for a given value of τ.

Figure 5-8 shows predictions for the data of Figure 5-1 using a rectifier power ν equal to 4 and 5 and using an exponential rectifier. The results indicate a distinct improvement in the ability of the model to describe the data as the expansivity of the rectifier is increased.

In Section 4.2.1, we discussed the dependence of lateralization predictions for tones on the properties of the peripheral rectifier. We believe that for values of ν much greater than 3 the discrepancies between data and predictions become significant. For this reason, such high order rectifiers are not acceptable, even though they may be helpful in describing the bandpass noise data. The exponential rectifier, on the other hand, provides a good fit to both sets of data (see Figures 5-8 and 5-9). This may be somewhat surprising upon initial consideration since the exponential rectifier is generally considered to be more expansive than ν^{th}-law devices. However, the derivative of e^z is less than the derivative of z^ν for small values of z. For this reason, the exponential rectifier behaves in a manner similar to that of only a third-order half-wave rectifier in response to tonal stimuli and predictions of the model are in agreement with the tone data. However, for noise inputs the operating point in the nonlinear transformation is higher and large values of the noise process are amplified significantly more by the exponential rectifier than they are by

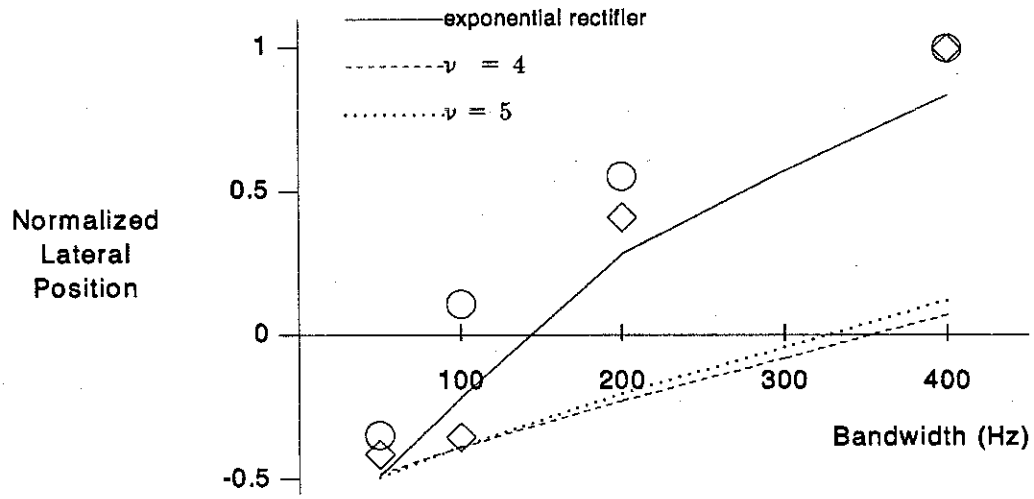

Figure 5-8: Predictions for the data of Figure 5-1 obtained using $p_{LF}(\tau|fc)$ and a fourth- and fifth-order half-wave rectifier and an exponential rectifier.

even high order (*e.g.*, ν = 10) ν^{th}-law rectifiers. This results in broader-band intensity functions for the exponential rectifier compared to those for the ν^{th}-law rectifier when the input is noise but similar intensity functions for tones.

In Chapter 3, we presented arguments that the exponential rectifier did not provide an accurate description of auditory-nerve data in response to noise inputs. These arguments motivated the replacement of the exponential rectifier by a less expansive device in the current version of the model. While the arguments of Chapter 3 are generally convincing, they are not conclusive. For this reason, the implications of the results presented here are sufficient to suggest that perhaps the use of an exponential rectifier should be reconsidered. Such an investigation is beyond the scope of this project and is left as a suggestion for future research.

The observations presented above indicate that the specification of the rectifier in the peripheral transduction mechanism of the model can strongly influence the extent to which bandwidth affects the predicted lateral position of bandpass noise. The use of a very expansive rectifier generally provides a good description of bandpass noise data. However, if the device also causes extreme sharpening of the intensity functions for tones, predictions for the lateralization of tones as a function of frequency tend to differ from the data.

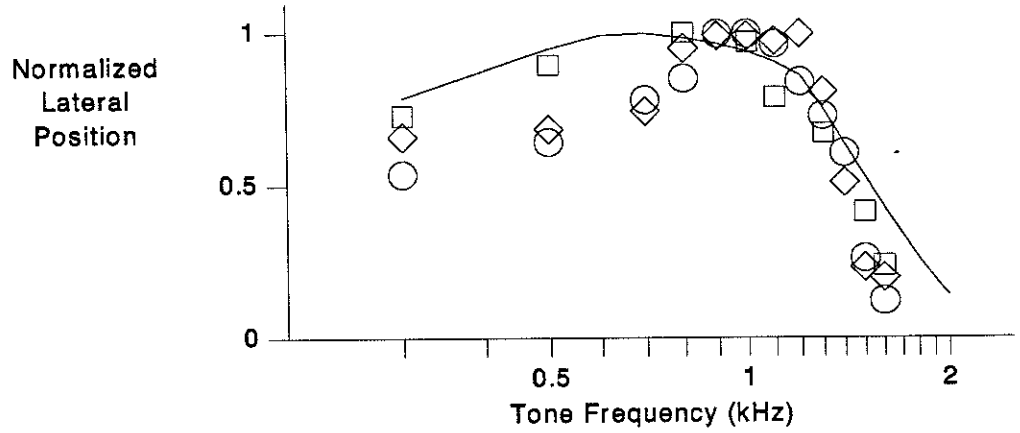

Figure 5-9: Predictions of the modified position-variable model for the lateralization of tones with an ITD of 0.15 ms obtained using an exponential rectifier.

Once again, it is important to point out that the need for a very expansive rectifier is strongly dependent on the assumed shape of $p(\tau|fc)$. For example, if we set ν equal to 5 and use the density function $p_L(\tau)$, the model provides a fair description of the data of Figure 5-1. However, for the same value of ν and the density function $p_{LF}(\tau|fc)$, predictions of the model do not even cross to the "correct" side of the head until the input bandwidth is increased to almost 400 Hz.

5.3. Modifications to the Position Variable

In this section we consider how the position variable $\hat{P}$ might be redefined in order to better exploit the attributes of the interaural correlation display that indicate the true value of the ITD present in a stimulus. Unlike the modifications considered in the previous section, this approach allows us to take advantage of both the cues of consistency across frequency and absolute maxima. In addition, the modifications introduced here are, strictly speaking, not subject to physiological constraints, although hypotheses that are physiologically plausible are generally considered to be more attractive.

The work presented here represents only a preliminary investigation of several "position variables" and should not be considered a thorough treatment of all the issues involved.

In order to simplify the analysis and quickly gain insight into the effects of several types of modifications to how information is interpreted from the internal representation of interaural correlation, we assume that position estimates are obtained from direct operations on the mean of this display, $L(\tau, fc)$. In general this approach does not provide a statistically correct method of computing the mean of the position variable for atleast two reasons. First, many of the modifications considered involve passing $L(\tau, fc)$ through some nonlinear transformation, and it is well known that, for a given random variable z and nonlinear transformation $W(z)$, $E[W(z)]$ is generally not equal to $W(E[z])$. Second, the techniques used to average over the distribution of internal delays and characteristic frequencies are incorrect because, for certain modifications, the distribution functions are also passed through nonlinear functions. For the sake of expedience and insight we ignore these inaccuracies. In many of the cases considered, there is little difference between the statistically correct expression and the one used here. In those cases where there may be significant discrepancies, we attempt to qualitatively predict how the correct calculations may differ from the ones presented below. Future efforts should attempt a more rigorous treatment of the statistical analysis.

5.3.1. Attenuating "False" Peaks

While many different processing schemes can be imagined to attenuate the secondary peaks of $L(\tau, fc)$ with respect to the primary or true peak, nearly all such schemes employ an expansive, nonlinear transformation to increase the effective bandwidth of the interaural correlation function. In general, the processing schemes only differ in the specification of the nonlinear transformation and the stage of the computation at which it is applied.

Consider the modified position variable $\hat{P}_1$ given by,

$$\hat{P}_1 = \frac{\sum_{m \in \mathbf{Z}_{lat}} \tau_m W(L_m)}{\sum_{m \in \mathbf{Z}_{lat}} W(L_m)} \tag{5.1}$$

where $W(z)$ is some expansive, nonlinear transformation (*e.g.*, $W(z) = z^\rho$, $\rho > 1$). The above definition is very similar in effect to increasing the order of the peripheral rectifier, and the observations made in Section 5.2.2 are applicable here. For this reason, we do not

discuss predictions of this model in detail and merely point out a few important observations.

First, one should be careful in specifying the transformation $W(z)$ so that it does not significantly affect predictions for tones. This is more difficult to accomplish for the above formulation than when the rectifier itself is modified since the transformation occurs after the peripheral lowpass filter. This means that high-order harmonics introduced by $W(z)$ into the correlation function for tones are not attenuated by the lowpass filter. These terms can significantly detract from the ability of the model to describe the data for tones.

The mean of the position variable specified by Equation (5.1) depends on the shape of $W(z)$. If we assume that only doubly-active fiber pairs are considered and that $W(z)$ is expansive, the approximation

$$E[\hat{P}_1] \approx \frac{\int_{\mathbf{R}_{lat}} p(fc) \int_{-\infty}^{\infty} \tau W[L(\tau, fc)]\, p(\tau \,|\, fc)\, d\tau\; dfc}{\int_{\mathbf{R}_{lat}} p(fc) \int_{-\infty}^{\infty} W[L(\tau, fc)]\, p(\tau \,|\, fc)\, d\tau\; dfc} \tag{5.2}$$

will generally yield a result that is further from the true ITD than the statistically correct mean value of $\hat{P}_1$ (for noise stimuli with pure ITDs). Thus, the incorrect expression used here underestimates the ability of the model to identify the true ITD. This implies that a less expansive transformation is actually required to describe the bandpass noise data than is indicated by predictions obtained using the above approximation.

The "additional" processing included in the definiton of $\hat{P}_1$ occurs at the output of each concidence counter, or alternately it could be considered an actual property of these units. This type of modification is physiologically quite plausible.

Another similar modification would be a scheme in which the nonlinear transformation is applied after integrating across frequency. In other words we are proposing a second modified position variable, cleverly named $\hat{P}_2$, whose mean can be approximated by

$$E[\hat{P}_2] \approx \frac{\int_{-\infty}^{\infty} \tau\, p(\tau)\, W\!\left[\int_{\mathbf{R}_{lat}} L(\tau, fc)\, p(fc)\, dfc\right] d\tau}{\int_{-\infty}^{\infty} p(\tau)\, W\!\left[\int_{\mathbf{R}_{lat}} L(\tau, fc)\, p(fc)\, dfc\right] d\tau} \tag{5.3}$$

Comparing the above expression with Equation (5.2), it is apparent that a less expansive nonlinearity should be required to describe the bandpass noise data using Equation (5.3). This observation is supported by the predictions shown in Figure 5-10(a), which were generated with $W(z) = z^3$ and a rectifier power ν equal to 2. It is interesting to note that for a simple power-law transformation of this type, predictions for tones generated using the two schemes are nearly identical. Thus, it appears that one is able to achieve a better description of the noise data using Equation (5.3) than using Equation (5.2) for a given realization of $W(z)$. At the same time, the (moderate) degradation in the fit of predictions to data describing the lateralization of pure tones as a function of frequency is the same using both equations (see Figure 5-10(b)).

5.3.2. Detecting Consistency Across Frequency

All the processing techniques considered up to this point take advantage of the fact that the interaural correlation function $L(\tau, fc)$ is largest for values of τ near the true ITD. However, none of these techniques make use of the fact that $L(\tau, fc)$ is also constant over a range of frequencies when τ equals the ITD. Thus, there is a large amount of information contained in the interaural display which is ignored by the previously considered processing schemes.

One can imagine a number of ways to measure "consistency" across frequency and a number of ways of incorporating such a measure into the specification of a position variable. In this section we present results for two examples of (ad hoc) processing schemes in an attempt to demonstrate the usefulness and reliability of the "consistency" cue in describing the lateral position data for noise (and tones).

5.3.2.1. Gradient-weighted position variable

The first technique makes use of the partial derivative of $L(\tau, fc)$ taken with respect to the log of characteristic frequency. Consider the mean of a *gradient-weighted* position variable given by

$$E[\hat{P}_3] = \frac{\int_{-\infty}^{\infty} \int_{\mathbf{R}_{lat}} \tau\, L(\tau, fc)\, W_g(\tau, fc)\, p(\tau, fc)\, d fc\; d\tau}{\int_{-\infty}^{\infty} \int_{fc \in \mathbf{R}_{lat}} L(\tau, fc)\, W_g(\tau, fc)\, p(\tau, fc)\, d fc\; d\tau} \tag{5.4}$$

where

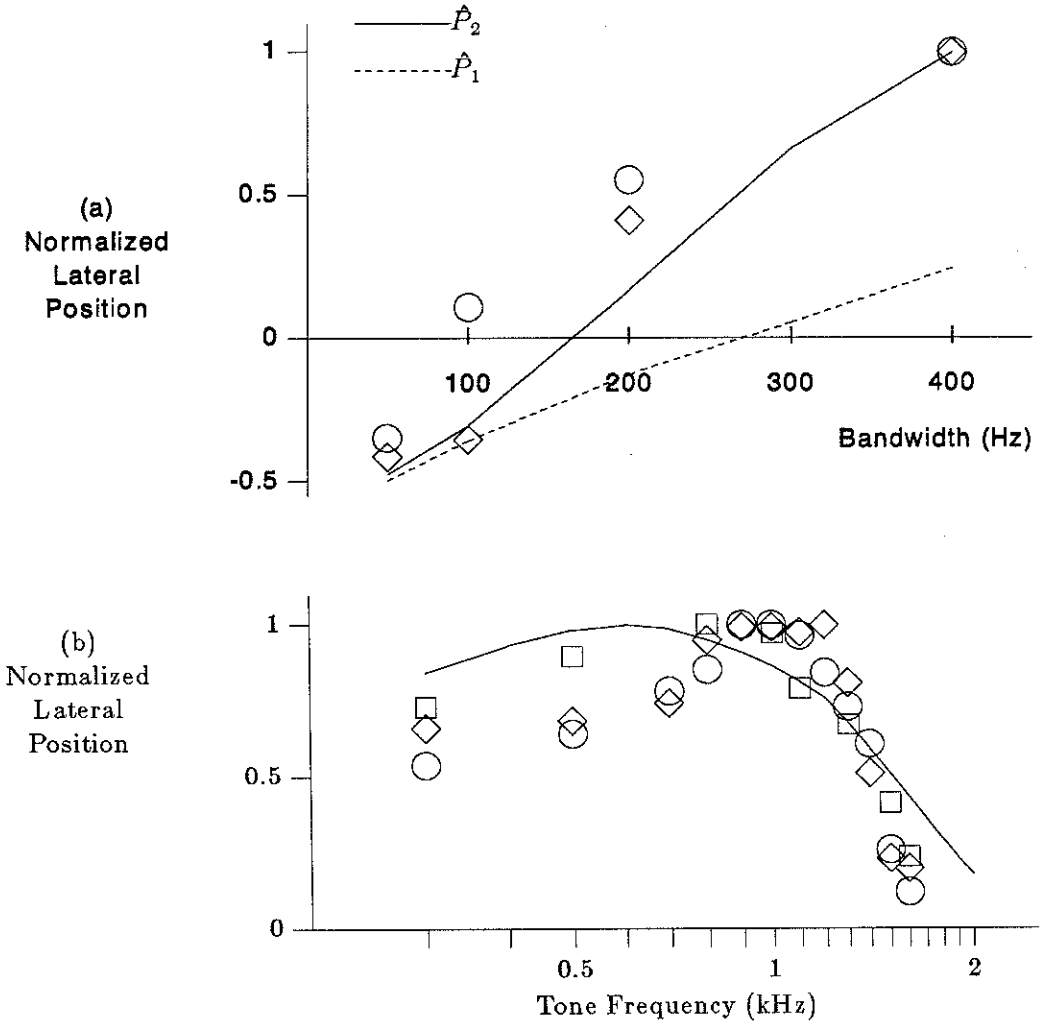

Figure 5-10: Predictions for the lateralization of (a) bandpass noise centered at 500 Hz with an ITD of 1.5 ms and (b) tones with ITDs of 0.15 ms obtained using $p_{LF}(\tau|fc)$ and ν = 2. The position estimates were were generated using the nonlinear processing schemes for $\hat{P}_1$ and $\hat{P}_2$ with $W(z) = z^3$.

$$W_g(\tau, fc) \triangleq \frac{L(\tau, fc)}{[L(\tau, fc) / C] + |\partial L(\tau, fc) / \partial \log\ fc|} \tag{5.5}$$

and C is some positive constant. For values of τ near the true ITD, the magnitude of the partial derivative in the above expression should be very nearly zero, independent of characteristic frequency and the bandwidth of the input noise. Consequently, the weighting function $W_g(\tau, fc)$ reduces to approximately C at the primary peak. For interaural delays not equal to the ITD, the magnitude of the partial derivative will depend on both the characteristic frequency and the bandwidth. Referring to Figure 5-5, we see that for a typical value of τ not equal to the ITD, the average magnitude of the partial derivative will increase as the bandwidth of the noise increases. Thus, as bandwidth increases, the average value of $W_g(\tau, fc)$ for a given τ will decrease, except when τ is equal to the true ITD.

If the input waveform is a tone, the interaural correlation function $L(\tau, fc)$ is independent of characteristic frequency and thus the magnitude of the partial derivative in the expression above is zero (for all values of τ). Consequently, the weighting function reduces to a simple constant and the mean of this position variable is identical to that of the original position variable (see assumption **DM2** in Chapter 3). This implies that the predictions for tones are not affected by the modifications to the model described above.

Figure 5-11 shows predictions for the lateralization of bandpass noise as a function of bandwidth generated using the expression in Equation (5.4). The parameter C was selected to maximize the shift in predicted position as bandwidth increases.[13] While the predictions in this figure do not provide an accurate quantitative description of the data, they are in better agreement with the data than the predictions generated using the original specification of $\hat{P}$ (see the solid line of Figure 5-6). This confirms that the cue of consistency across frequency is useful in describing the bandpass noise data. However, it appears that the weighting function $W_g(\tau, fc)$ does not exploit this cue to the extent necessary to describe the data. Many other measures of consistency which also make use of the partial derivative can be imagined, and it is quite probable that some of these will provide a better description of the data.

---

[13]Originally, we expected predictions to monotonically approach the true ITD as C was increased, but we found that this was not the case. We feel that this unexpected trend in the predictions is probably caused by computational artifacts rather than a fundamental deficiency in the definition of $W_g(\tau, fc)$. For example, the computer-generated samples of the function $L(\tau, fc)$ are not entirely independent of characteristic frequency when τ is equal to the ITD, as they should be. This results in a non-zero partial derivative term in Equation (5.5) at the true ITD. Furthermore, the truncated power-series approximation used to compute $L(\tau, fc)$ yielded some negative samples that were "clipped" to zero, causing certain regions of the display to be "falsely" consistent across frequency.

Finally, we wish to point out that the partial derivative operation used above could be physiologically plausible, making this a particularly promising approach which should be more closely examined in future studies.

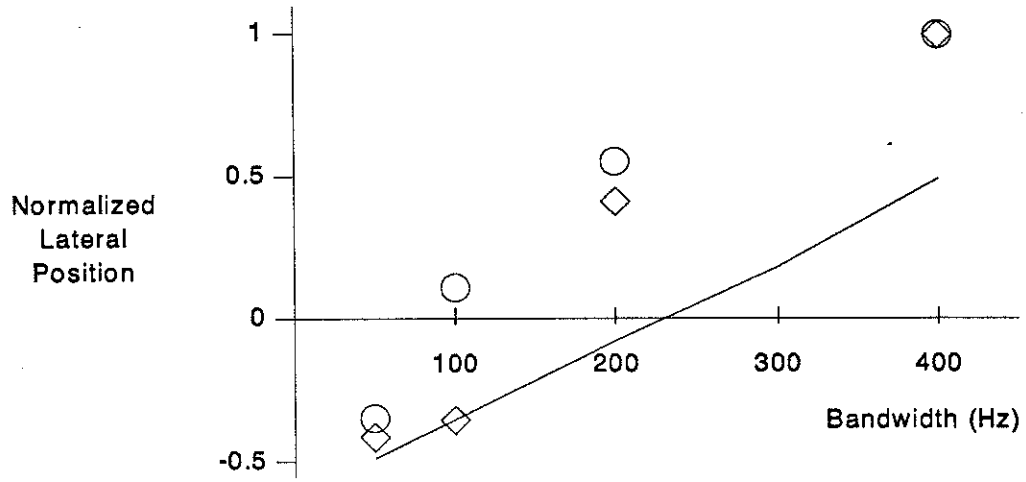

Figure 5-11: Predictions for the data of Figure 5-1 using the gradient-weighted position variable specified by Equation (5.4), $p_{LF}(\tau \,|\, fc)$, and ν equal to 3.

5.3.2.2. Variability-weighted position variable

Another measure of the consistency of $L(\tau, fc)$ across frequencies is related to the "variability" of $L(\tau, fc)$ for a given value of τ. We define the *variability* $V_T^2(\tau)$ as follows,

$$V_T^2(\tau) \triangleq \frac{\int_{\mathbf{R}_{lat}} [\, L(\tau, fc) - \mu_T(\tau) \,]^2 \, p(fc)\, dfc}{\int_{\mathbf{R}_{lat}} p(fc)\, dfc} \tag{5.6}$$

where

$$\mu_T(\tau) \triangleq \frac{\int_{\mathbf{R}_{lat}} L(\tau, fc)\, p(fc)\, dfc}{\int_{\mathbf{R}_{lat}} p(fc)\, dfc} \tag{5.7}$$

If τ is equal to the value of the ITD present in the binaural sound, $V_T^2(\tau)$ will be zero. For

nearly all other values of τ, the variability $V_T^2(\tau)$ will be non-zero and will increase as the bandwidth of the input increases.

We now incorporate this measure of consistency (or the lack of consistency) into a "variability-weighted" position variable whose mean is assumed to be given by

$$E[\hat{P}_4] = \frac{\int_{-\infty}^{\infty} \tau \left[\frac{L_T(\tau)}{C + V_T^2(\tau)} \right] d\tau}{\int_{-\infty}^{\infty} \left[\frac{L_T(\tau)}{C + V_T^2(\tau)} \right] d\tau} \tag{5.8}$$

where

$$L_T(\tau) = \int_{R_{lat}} L(\tau, fc)\, p(\tau, fc)\; dfc \tag{5.9}$$

and C is some positive constant. The term in brackets above provides a measure of the probability that τ is equal to the ITD present in the stimulus.[14] The function $L_T(\tau)$ adjusts this probability measure on the basis of the distribution of internal delays and the relative height of the peaks of $L(\tau, fc)$ (in a manner very similar to the unmodified definition of $\hat{P}$). The term involving $V_T^2(\tau)$ reflects the fact that values of τ for which the correlation function $L(\tau, fc)$ is not consistent across frequency are not likely to reflect the true ITD. For such values of τ, the variability term $V_T^2(\tau)$ will be large and τ will be weighed lightly in the position estimate. On the other hand, the true ITD will be weighed more heavily than other values of τ since $V_T^2(\tau)$ will be zero at the ITD.

The results shown in Figure 5-12 indicate that the above formulation is capable of describing the bandpass noise data of interest. Predictions for the lateralization of tones as a function of frequency are also in agreement with the data since Equation (5.8) describes the mean of the original position variable when $L(\tau, fc)$ is independent of characteristic frequency.

Thus, we see that the cue of consistency across frequency can be used in the formation of position estimates in a manner that is consistent with the observed dependence of the lateralization of bandpass noise, with pure ITDs, on the bandwidth of the input noise.

---

[14] This interpretation is similar to our discussion in Section 3.3.2.

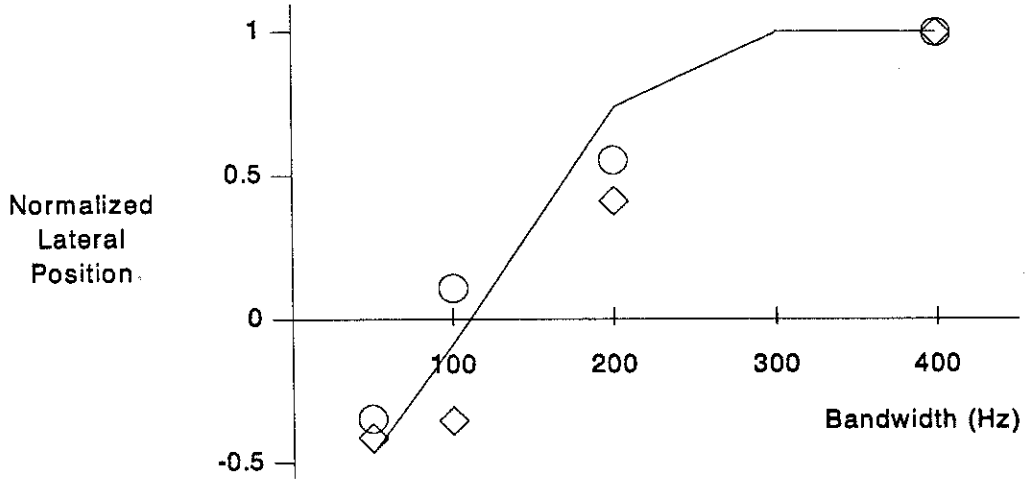

Figure 5-12: Predictions for the data of Figure 5-1 using the variability-weighted position variable specified by Equation (5.8), $p_{LF}(\tau \mid fc)$, and ν equal to 3.

5.4. Joint Dependence of Predicted Position on Bandwidth, ITD, and IPD

In Section 2.1.1.2 we discussed results of a (pilot) lateralization experiment similar to the one considered thus far, with the exception that the binaural noise stimulus was phase delayed as well as time delayed before presentation. Combinations of ITD and IPD were selected so that the interaural correlation functions computed for the various stimulus configurations are nearly identical for narrow bandwidths. The effect of the IPD is to reduce the strength of both the "consistency" and "maximum peak" cues that help to indicate which ear is actually receiving the signal that is leading in time.

Experimental results from this pilot study are shown in Figure 5-13 below. They indicate that as the magnitude of the phase delay increases, the extent to which position judgements depend on bandwidth decreases. This observation is intuitively appealing since most naturally occuring broadband sounds do not contain interaural phase delays and we do not expect the auditory system to be able to make efficient use of such cues.

Predictions of the modified position-variable model obtained using all configurations discussed thus far describe at least the qualitative trends of these data. Furthermore,

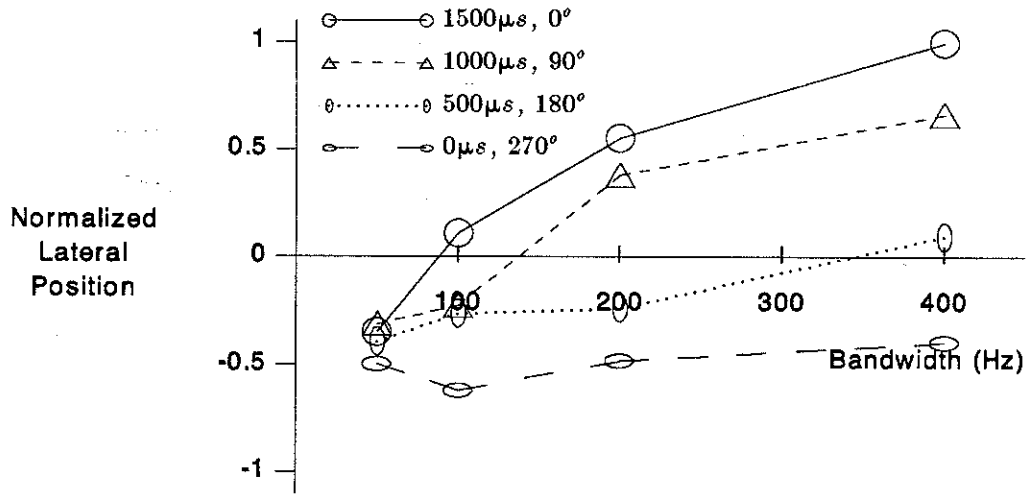

Figure 5-13: Perceived lateral position of bandpass noise centered at 500 Hz with a number of ITD, IPD combinations is plotted as a function of bandwidth. [Data from Shear and Stern (unpublished).]

those model configurations which quantitatively describe results for the (1500 μs, 0°) case also describe the results for the other combinations of ITD and IPD (that are noted in Figure 5-13). An example of typical model predictions for these data, obtained using the variability-weighted position variable, are shown in Figure 5-14 below.

To further our understanding of how position estimates are formed it would be desirable to empirically determine the relative roles that the cues of consistency across frequency and the height of the true peak play in this judgement process. Unfortunately, consideration of the model's response to the bandpass noise stimuli described above has not provided much insight into this question. It appears that those stimuli that elicit internal correlation displays that are consistent across frequency near the true ITD also yield the displays with the largest peaks near the true ITD. Thus far we have not developed stimulus configurations that can be used to reliably separate the effects of these cues.

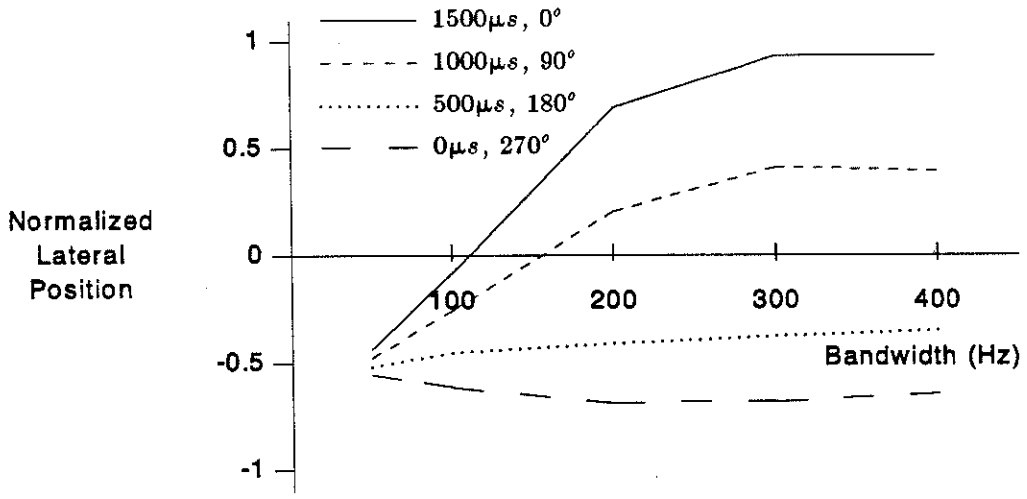

Figure 5-14: Predictions of the variability-weighted position-variable model for the lateralization of bandpass noise centered at 500 Hz with a number of ITD, IPD combinations are plotted as a function of bandwidth. (These predictions were obtained using $p_{LF}(\tau|fc)$ and ν = 3.)

5.5. Summary and Conclusions

The results presented in this chapter indicate that our current theories of peripheral transduction and binaural interaction do describe the basic trends exhibited by experimental evidence for the lateralization of bandpass noise centered at 500 Hz with various interaural time and phase differences as a function of bandwidth. We have identified two attributes of the interaural crosscorrelation display that can each provide a progressively better indication of the true ITD present in a binaural stimulus as the bandwidth of the stimulus increases. The first of these cues is the relative height of the peaks of the interaural crosscorrelation function, with the largest peak indicating the true ITD. The second cue is the degree of consistency across frequency exhibited by the correlation display for a given value of interaural delay, with the most consistent regions of the display located near the true value of the ITD.

Modifications to the model that better exploit each of these cues have been considered. Our results indicate that such changes do allow the model to more accurately quantitatively describe the lateralization phenomena of interest. However, we have been

unable to determine precisely which type of modification or which combination of modifications is most consistent with the actual mechanisms of the human auditory system. This is due in part to the fact that there is remarkably little experimental evidence describing the lateralization of bandpass noise available to constrain the models considered. Lateralization phenomena elicited by broadband sounds possess great potential for improving our understanding of binaural interaction, and we feel that such phenomena are highly worthy of consideration in future theoretical and experimental efforts.

Chapter 6
Summary

The original position-variable model described by Stern and Colburn (1978) describes nearly all available lateralization data for 500-Hz tones. This model can also be used to describe the difference between $N_\pi S_0$ and $N_0 S_\pi$ detection thresholds as a function of frequency. However, predictions of the original position-variable model fail to accurately describe the following three sets of phenomena:

- The response patterns of auditory-nerve fibers to noise.
- The lateralization of complex, high-frequency stimuli with ITDs in their low-frequency envelopes.
- The lateralization of pure tones as a function frequency.

In this report we describe a modified version of the position-variable model that more accurately describes the phenomena noted above. The major modifications to the model are:

- The use of a low-order ν^{th}-law half-wave rectifier in place of the original exponential rectifier.
- The placement of the lowpass filter element after the rectifier in the model of peripheral auditory transduction.
- The specification of a new form for the function $p(\tau|fc)$ that describes the assumed distribution of internal, interaural delays over the model's binaural coincidence units.

The first modification above is intended to provide a more accurate statistical characterization of neural firing patterns elicited by noise stimuli. While we have not generated predictions of the modified model for the lateralization of complex, high-frequency stimuli, we feel that the second modification noted above should greatly improve the ability of the model to describe these phenomena since such a configuration provides a means by which the envelopes of these stimuli can be extracted. The

respecification of the density function $p(\tau|fc)$ is necessary to allow the model to accurately describe the data for the lateralization of tones as a function of frequency.

The modified model, with the new density function $p_{LF}(\tau|fc)$ (given by Equation (4.14)), provides a good description of both the lateralization and detection data for tones considered here.

In our efforts to determine a new density function, we observed several dependencies exhibited by predictions of the model that are worth noting. First, phase-based distributions, or those distributions that are functions of the product $\tau\, fc$, generally provide a fair description of the lateralization phenomena but are not well suited for describing the detection data of interest. Second, lateralization predictions for tones are more sensitive to the exact shape of the distribution of delays than are predictions for the relative detection thresholds considered. Third, both the lateralization and detection predictions for high-frequency tones are dominated by the lowpass filter function $G(f)$, which characterizes the observed loss of synchrony between neural firing patterns and the detailed timing structure of such stimuli. We feel that minor discrepancies between the model's predictions and available lateralization data at these frequencies are probably related to an inaccuracy in our description of this function. Finally, predictions for N_0S_π detection thresholds, as a function of target frequency, are not strongly dependent on the specification of $p(\tau|fc)$ or the order of the rectifier. However, these predictions are sensitive to the shape of the peripheral bandpass filters. We feel that the inability of the model to quantitatively describe N_0S_π detection thresholds at frequencies above about 500 Hz is related to the current characterization of the shapes of these filters.

In this report we also present results from a preliminary investigation into the modified model's ability to describe certain lateralization phenomena for bandpass noise with various combinations of ITD and IPD as a function of bandwidth. We found that the new position-variable model describes the qualitative trends exhibited by the data, but quantitative descriptions are strongly dependent on a number of the model's parameters, with the two most significant parameters being the distribution of internal delays and the order of the peripheral rectifier. For the values of these parameters used to describe the phenomena for tones, predictions of the new model do not quantitatively describe the bandpass noise data considered.

We have noted that the interaural correlation dipslay is consistent across characteristic

frequency for values of τ near the ITD present in a noise stimulus. Moreover, the height of the peak in the display at this value of τ increases relative to the other peaks in the display as bandwidth increases. Modifications to the mechanism used to generate the position estimate were proposed in order to determine the usefulness of these two cues in "correctly" lateralizing broadband sounds. Our results indicate that both cues are helpful in this respect, but the available psychoacoustical evidence is inadequate to allow us to determine which cue or combination of cues is actually exploited by the human auditory system.

Although we have demonstrated the ability of the modified position-variable model to describe only a small set of binaural phenomena, we feel that this model is potentially a very powerful tool that can be used to improve our understanding of a much larger class of phenomena.

6.1. Suggestions for Future Work

In its present configuration the modified position-variable model can be easily applied to psychoacoustic evidence describing the lateralization of complex, high-frequency stimuli. A study of the model's predictions for such stimuli is likely to provide useful insight into many of the questions left unanswered by this report. Specifically, the data obtained by Henning (1974, 1980) should be useful in more closely constraining $p(\tau|fc)$ at characteristic frequencies above 1200 Hz. These data may also provide a means of determining the relative roles of "consistency" and "absolute maxima" cues in the process of forming lateral position judgements.

We are somewhat surprised by the lack of psychoacoustic data for the lateralization of bandpass and broadband noise. The findings of such experiments promise to be very useful in furthering our understanding of how timing information encoded in the internal correlation display might be processed by the higher centers of the human auditory system. We suggest that closely related theoretical and experimental research efforts focusing on the lateralization of noise stimuli with temporal disparities be conducted. The theoretical efforts should, in part, consider a more rigorous analysis of some of the "post-processing" schemes described in Section 5.3 of this report.

The modified PV model used in our work does not include a mechanism by which interaural intensity differences can be incorporated in the position estimate. While we

suspect that a mechanism similar to the one used in the original position-variable model (cf. Stern and Colburn, 1978) should be adequate, a complete study needs to be conducted. It is probable that such a study would require data for the lateralization of tones at various frequencies with combinations of ITD and IID. Such data for 500-Hz tones obtained by Domnitz and Colburn (1977) were used by Stern and Colburn (1978) to specify the intensity mechanisms of the original model. It might also be interesting to try to adapt the "lateral-inhibition" scheme of Lindemann (1986) to the modified PV model.

There are two issues regarding the underlying mechanisms of the model that should be addressed. First, an attempt should be made to determine the correct expression for the variance of the displayer outputs $\{L_m\}$ for the modified model configuration. Second, we have also noted that there is reason to re-examine the shapes of the peripheral lowpass and bandpass filters on the basis of atleast two sets of psychoacoustical results.

Appendix A
Lateralization of Bandpass Noise Experiment

A.1. Introduction

In this appendix we describe the methods and results of a pilot lateralization experiment. The stimuli considered were bands of Gaussian noise centered at 500 Hz with bandwidths of 50, 100, 200, and 400 Hz and (ITD, IPD) combinations of (1500 μs, 0°), (1000 μs, 90°), (500 μs, 180°), and (0 μs, 270°). These (ITD, IPD) pairs were chosen because they all produce interaural crosscorrelation functions $L(\tau, fc)$ that exhibit a maximum at τ = 1500 μs and fc = 500 Hz. This facilitates the observation of the roles of consistency and absolute maxima cues in the formation of lateral position judgements. As the magnitude of the IPD increases, the less helpful these two cues become.

The results of the study described below have already been presented in the body of this report (see Figure 5-13).

A.2. Procedure

The method of Schiano, *et al.* (1986) was replicated to the extent possible given the differences in laboratory facilities. The primary task of the experiment was the positioning of an acoustical pointer to match the intracranial position of a target stimulus. The pointer consisted of a 200-Hz band of noise centered at 500 Hz with a variable IID that was controlled by the subject. The position matching task was performed a minimum of three times for each set of stimulus parameters and the mean and standard deviation of the IID were recorded.[15]

---

[15]Sample standard deviation was computed using the number of data points minus one in the denominator of the expression.

The stimuli were presented in six 100-ms pulses separated by 50-ms intervals. The rise and fall times for each pulse were 10 ms. The first three presentations consisted of the target and the last three consisted of the pointer. Each pulse train was separated by an 800-ms pause during which time the subject had the option of not responding, responding left, responding right, or indicating a match. If the subject did not respond, the same set of target and pointer parameters was used on the next cycle. If the subject wished to move the pointer in either direction, the IID was incremented or decremented accordingly before the next presentation. The subject signalled a match to stop the process, and the final IID was recorded. The initial pointer position for each trial was selected at random so that previous trial results would not influence the subject.

The target stimuli described above were generated on a DEC PDP 11/23 computer using the expression given by

$$n_L(t) = \sum_{i=0}^{249} A_i \cos(2\pi f_i t + \theta_i)$$

$$n_R(t) = \sum_{i=0}^{249} A_i \cos[2\pi f_i (t - \tau_n) - \phi_n + \theta_i]$$

where the coefficients $\{A_i\}$ were drawn from a Rayleigh distribution, and the angles $\{\theta_i\}$ were drawn from a uniform distribution over the interval $[0, 2\pi)$. Each frequency f_i was selected from a uniform distribution over the range from $f_l + \Delta f(i-1/2)$ to $f_l + \Delta f(i+1/2)$, where f_l is the lowest frequency in the target band and Δf is equal to the bandwidth divided by 250.

The targets were presented at a level of 70 dB SPL and the pointer at a nominal (*i.e.*, with an IID of 0 dB) level of 73 dB SPL. The pointer was presented at a higher intensity to prevent loudness matching rather than position matching. A piecewise-linear approximation to empirical curves of constant loudness vs. IID (cf. Keen, 1972) was implemented to reduce the salience due of loudness cues.

A pointer bias factor was determined for each block of trials by presenting a diotic target. This bias IID was subtracted from each of the averaged data points in the block of trials and the adjusted IID was used to represent the perceived position of the target. Schiano, *et al.* (1986) found that accounting for this bias improved the reproducibility of

results from session to session. One possible cause of this variable bias is the positioning of the headphones. For this reason the headphones were not moved during a given block of trials.

A.3. Results

The mean and standard deviation of the pointer IID used to match each set of stimulus configurations are listed in Table A-1. While we consider data points with standard deviations up to only about 2.5 dB to be completely acceptable, we feel the averaged results given below still provide a fair indication of subjective lateral position for the stimuli considered. It is likely that with more training the variability of the data could be reduced significantly.

Average / Standard Deviation of Pointer IID (R/L dB)				
ITD, IPD	50 Hz	100 Hz	200 Hz	400 Hz
1500 μs, 0^o	3.16 / 2.75	-0.17 / 2.43	-5.80 / 1.82	-11.42 / 2.98
1000 μs, 90^o	5.66 / 1.72	3.42 / 2.05	-1.42 / 2.83	-4.75 / 4.72
500 μs, 180^o	5.50 / 4.05	4.32 / 3.26	6.42 / 4.66	0.50 / 2.21
0 μs, 270^o	5.00 / 2.69	5.56 / 2.89	7.00 / 2.43	5.83 / 1.59

Table A-1: Average and standard deviation of pointer IIDs used by subject GS to match perceived lateral position of bands of noise centered at 500 Hz with various combinations of bandwidth, ITD, and IPD.

Appendix B

Response of the Binaural Displayer to Tones, Noise, and Combinations of the Two

In this appendix we present expressions for the interaural correlation function $\mathcal{L}_2(\tau, fc)$ when the input stimulus is an additive combination of a tone and Gaussian noise.

B.1. Response to a Tone with Additive Noise

Consider the binaural stimulus given by

$$x_L(t) = n(t) + A \cos 2\pi f_0 t$$

$$x_R(t) = n_R(t) + A \cos 2\pi f_0 (t - \tau_s),$$

where the noise process $n(t)$ has the one-sided spectral density function $N(f)$. The noise component to the right ear $n_R(t)$ is obtained by time- and phase- delaying $n(t)$ by τ_n and ϕ_n.

When considering the response of the m^{th} fiber pair to this input, it is helpful to define the following quantities.

$$N_m(f) = H_m^2(f)\, N(f)$$

$$\sigma_{n\,m}^2 = \int_0^\infty H_m^2(f)\, N(f)\, df$$

$$R_{n\,m}(\tau) = \mathcal{F}^{-1}\{ N_m(f) \exp(-j2\pi f\tau_n - j\phi_n) \}$$

$$\sigma_{s\,m}^2 = \frac{H_m^2(f_0)\, A^2}{2}$$

$$\sigma_m^2 = \sigma_{s\,m}^2 + \sigma_{n\,m}^2$$

where $\mathcal{F}^{-1}\{\}$ represents the inverse Fourier transform operation (accounting for the fact that $N_m(f)$ is one sided).

If both fibers are active, then the interaural correlation $R_{RLm}(\tau_m)$ will depend on the values fc_m and τ_m, as well as on the properties of the binaural stimulus. In order to determine this value, we must compute the crosscorrelation of the outputs of the ν^{th}-law half-wave rectifiers from each ear. Using the results of Davenport and Root (1958, pp. 277-308) we obtain[16]

$$R_{RLm}(\tau_m) = \left[\sum_{i=0}^{\infty} \sum_{k=0}^{\infty} \frac{\epsilon_i h_{ik}^2}{k!} R_{n\,m}^k(\tau_m) \cos\,[2\pi i f_0(\tau_m - \tau_s) - \phi_s] \right] * g(\tau_m) * g(-\tau_m) \qquad (B.1)$$

where ϵ_i is the Neumann factor $\epsilon_0 = 1$, $\epsilon_i = 2$ (i = 1, 2, ...),

$$h_{ik} = \frac{200\,(\sigma_{s\,m}^2/\sigma_{n\,m}^2)^{i/2}\,\Gamma(1+\frac{\nu}{2})\,2^{k/2}}{i!\,\Gamma[1-(i+k-\nu)/2]\,\sigma_{n\,m}^k}\;{}_1F_1\left(\frac{i+k-\nu}{2}; i+1; -\frac{\sigma_{s\,m}^2}{\sigma_{n\,m}^2}\right), \qquad (B.2)$$

and ${}_1F_1(a; c; z)$ is the confluent hypergeometric function defined by the series

$${}_1F_1(a; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k\, z^k}{(c)_k\, k!} = 1 + \frac{a\,z}{c\,1!} + \frac{a(a+1)\,z^2}{c(c+1)\,2!} + \cdots \qquad (B.3)$$

The expression for the crosscorrelation given above reduces to the simple power series

$$R_{RLm}(\tau_m) = 200^2\,\{\,1 + 2\sum_{k=1}^{\infty}\left[\frac{2^k\,\Gamma^2(1+\frac{\nu}{2})}{k!\,\Gamma^2[1-(k-\nu)/2]}\right][R_{n\,m}(\tau_m)/\sigma_{n\,m}^2]^2\,\} * g(\tau_m) * g(\tau_m) \qquad (B.4)$$

when there is no tone present (*i.e.*, $\sigma_{s\,m}^2$ equals zero).

B.2. Response to a Pure Tone

Equation (B.2) does not converge when the stimulus is a pure tone (*i.e.*, $\sigma_{n\,m}^2$ equals zero). This stimulus configuration must be considered as a special case.

As was already mentioned, the mean of L_m is a constant when either of the input fibers to the coincidence counter is inactive. If we consider a single fiber pair and assume that the

---

[16]Davenport and Root (1958) derive an expression for the autocorrelation of the output of the rectifier. However, since the intensities of the component stimuli of interest are identical in each ear, these results can be applied directly to the crosscorrelation considered.

input tone is sufficiently intense to activate both fibers, it is apparent that the automatic gain control element causes the intensity function for each fiber to be independent of the characteristic frequency of the fiber. In addition, each intensity function is periodic with fundamental frequency f_0. These observations suggest the following means of expressing the crosscorrelation $R_{RLm}(\tau_m)$,

$$R_{RLm}(\tau_m) = (200)^2 \{ 1 + 2 \sum_{n=1}^{\infty} S_n^2 G^2(n f_0) \cos [2\pi n f_0 (\tau_m - \tau_s) - \phi_s] \} \qquad (B.5)$$

where S_n is the magnitude of the n^{th} coefficient of the Fourier series of the output of the rectifier normalized by the mean firing rate (200 per second). The coefficients $\{S_n\}$ depend on the order of the half-wave rectifier being used. Table B-1 gives the values of the first eight coefficients when ν equals 1, 2, and 3. The more expansive the rectifier is (*i.e.*, the larger ν is), the larger the coefficients of the second and third order terms are.

	S_n^2		
n	$\nu = 1$	$\nu = 2$	$\nu = 3$
1	0.61685028	0.72050619	0.78070113
2	0.11111111	0.25	0.36
3	0.0	0.02882025	0.08674457
4	0.00444444	0.0	0.00734694
5	0.0	0.00058817	0.0
6	0.00081633	0.0	0.00009070
7	0.0	0.00006535	0.0
8	0.00043403	0.0	0.00000675

Table B-1: Normalized coefficients of the Fourier series of the output of a ν^{th}-law half-wave rectifier when the input is a tone.

Appendix C

Position Calculations in the Frequency Domain

In this appendix we derive a general expression for the mean of the position variable $\hat{P}$ using simple properties of the Fourier transform. The resulting expression is most useful for the special case when the interaural correlation function $\mathcal{L}_2(\tau, fc)$ is periodic with respect to τ. We simplify the expression for tonal inputs.

We begin the derivation with Equation (3.16), which is replicated below.

$$E[\hat{P}] = \frac{\int_{\mathbf{R}_{lat}} p(fc) \int_{-\infty}^{\infty} \tau L(\tau, fc)\, p(\tau \,|\, fc)\, d\tau \; dfc}{\int_{\mathbf{R}_{lat}} p(fc) \int_{-\infty}^{\infty} L(\tau, fc)\, p(\tau \,|\, fc)\, d\tau \; dfc}, \tag{C.1}$$

where $L(\tau, fc)$ can be expressed as

$$L(\tau, fc) = \eta_2(fc)\, \mathcal{L}_2(\tau, fc) + \eta_1(fc)\, \mathcal{L}_1 + \eta_0(fc)\, \mathcal{L}_0 . \tag{C.2}$$

Using two well known properties of the Fourier transform,

$$\int_{-\infty}^{\infty} x(t)\, y(t)\, dt \quad = \quad \int_{-\infty}^{\infty} X(f)\, Y^{*}(f)\, df$$

$$t\, x(t) \quad \leftrightarrow \quad \frac{-1}{j2\pi} \frac{dX(f)}{df}$$

and taking advantage of the fact that $p(\tau \,|\, fc)$ is an even function of τ, we obtain

$$E[\hat{P}] =$$

$$\frac{\int_{\mathbf{R}_{lat}} p(fc) \int_{-\infty}^{\infty} (-j2\pi)^{-1} \eta_2(fc)\, S_{\mathcal{L}_2}(f, fc)\, \dot{\phi}_\tau(f\,|fc)\; df\; dfc}{\int_{\mathbf{R}_{lat}} p(fc)\, [\, \eta_0(fc)\mathcal{L}_0 + \eta_1(fc)\, \mathcal{L}_1 + \int_{-\infty}^{\infty} \eta_2(fc)\, S_{\mathcal{L}_2}(f, fc)\, \phi_\tau(f\,|fc)\, df\,]\, dfc} \qquad (C.3)$$

where $\phi_\tau(f\,|fc)$ is the Fourier transform of $p(\tau\,|fc)$, $\dot{\phi}_\tau(f\,|fc)$ is the derivative of $\phi_\tau(f\,|fc)$ taken with respect to f, and $S_{\mathcal{L}_2}(f, fc)$ is the Fourier transform of $\mathcal{L}_2(\tau, fc)$.

If the input is a binaural tone, $\mathcal{L}_2(\tau, fc)$ will be periodic and can be expressed as

$$\mathcal{L}_2(\tau, fc) = T_w T_S 200^2\, [\, 1 + 2 \sum_{n=1}^{\infty} S_n^2\, G^2(nf_0) \cos\; 2\pi n\, f_0\, (\tau - \tau_s)\,] \qquad (C.4)$$

where f_0 is the frequency of the input tone (in Hertz), τ_s is the amount by which the right channel is delayed relative to the left channel, and S_n is the magnitude of the n^{th} Fourier series coefficient for the ouput of the rectifier normalized by the average rate of firing (200 per second).

Taking the Fourier transform of Equation (C.4) to obtain $S_{\mathcal{L}_2}(f, fc)$, substituting the result into Equation (C.3), and simplifying, we obtain

$$E[\hat{P}] =$$

$$= \frac{\frac{-200^2}{\pi} \sum_{n=1}^{\infty} G^2(nf_0)\, S_n^2 \sin(2\pi\, nf_0\tau_s) \int_{\mathbf{R}_{lat}} \eta_2(fc)\, \dot{\phi}_\tau(nf_0|fc)\, p(fc)\, dfc}{D_0 + 2\,(200)^2 \sum_{n=1}^{\infty} G^2(nf_0)\, S_n^2 \cos(2\pi\, nf_0\tau_s) \int_{\mathbf{R}_{lat}} \eta_2(fc)\, \phi_\tau(nf_0|fc)\, p(fc)\, dfc} \qquad (C.5)$$

with

$$D_0 \triangleq \int_{\mathbf{R}_{lat}} [\, \eta_2(fc)\, 200^2 + \eta_1(fc)\, (200)\, (50) + \eta_0(fc)\, 50^2\,]\, p(fc)\, dfc\ . \qquad (C.6)$$

In general, this expression provides a more efficient and accurate means of computing $E[\hat{P}]$ than the time-domain expression given by Equation (C.1).

Appendix D

Implementation Notes on the Generation of Predicitons

In this appendix we describe the details of the computer algorithms used to generate the predictions presented in this report.

All integrals were computed as discrete Riemann sums with a time step of 0.025 ms and a log-frequency step of 0.01 $\log_{10}$(Hz). The limits on the time summations were between $\pm$8 ms for detection experiments and $\pm$12.75 ms for lateralization experiments.

For detection experiments, threshold was declared when the square root of Q_d was between 0.975 and 1.025.

In order to determine samples of $R_{n\,m}(\tau)$ (as defined in Appendix B), a 4096-point inverse discrete Fourier transform was used. Similarly, lowpass filtering was accomplished by means of a 4096-point discrete Fourier transform.

The infinite summations of Equation (B.1) were truncated to include only terms 0 through 14. For pure tones, the Fourier series of Equation (B.5) was truncated to include only terms 0 through 8. (Values of theses coefficients for ν equals 1, 2, and 3 are given in Table B-1.) Terms of the confluent hypergeometric function given by Equation (B.3) were computed and summed until the absolute value of a given term divided by the summation at that point was less than 10^{-5}.

References

Bachorski, S.J. (1983). Perception of Binaural Stimuli which have Time-Varying Interaural Parameters. Master's thesis, Elec. and Comp. Eng. Dept., CMU, December, 1983.

Blauert, J., and Cobben, W. (1978). Some Consideration of Binaural Crosscorrelation Analysis. *Acoustica*, 1978, *39*, 96-103.

Colburn, H.S. (1969). *Some Physiological Limitations on Binaural Performance.* PhD thesis, Elec. Eng. Dept., MIT, September, 1969.

Colburn, H.S. (1973). Theory of Binaural Interaction Based on Auditory-Nerve Data. I. General Strategy and Preliminary Results on Interaural Discrimination. *J. Acoust. Soc. Am.*, 1973, *54*(6), 1458-1470.

Colburn, H.S. (1977a). Theory of Binaural Interaction Based on Auditory-Nerve Data. II. Detection of Tones in Noise. *J. Acoust. Soc. Am.*, 1977, *61*(2), 525-533.

Colburn, H.S. (1977b). Theory of Binaural Interaction Based on Auditory-Nerve Data. II. Detection of Tones in Noise. Supplementary Material. *AIP Document No. PAPS JASMA-61-525-98.*, 1977, *61*, 1-81.

Colburn, H.S. and Durlach, N.I. (1978). *Handbook of Perception.* Volume IV: Models of Binaural Interaction. In Carterette, E.C. and Friedman, M.P. (Ed.), *Hearing,* New York, NY: Academic Press, 1978.

Davenport, W.B. , Jr. and Root, W.L. (1958). *An Introduction to the Theory of Random Signals and Noise.* New York:McGraw-Hill Book Co., 1958.

Domnitz, R.H. and Colburn, H.S. (1977). Lateral Position and Interaural Discrimination. *J. Acoust. Soc. Am.*, 1977, *61*, 1586-1598.

Duifhuis, H. (1973). Consequences of peripheral frequency selectivity for nonsimultaneous masking. *J. Acous. Soc. Am.*, 1973, *54*(6), 1471-1488.

Durlach, N.I. and Colburn, H.S. (1978). *Handbook of Perception.* Volume IV: Binaural Phenomena. In Carterette, E.C. and Friedman, M.P. (Ed.), *Hearing,* New York, NY: Academic Press, 1978.

Erulkar, S.D. (1972). Comparative aspects of spatial localization of sound. *Physiological Review*, 1972, *52*, 237-360.

Green, D.M. (1976). *An Introduction to Hearing.* Hillsdale, New Jersey:Lawrence Erlbaum Associates, Inc., 1976.

Henning, G.B. (1974). Detectability of Interaural Delay in High-frequency Complex Waveforms. *J. Acoust. Soc. Am.*, 1974, *55*, 84-90.

Henning, G.B. (1980). Some Observations on the Lateralization of Complex Waveforms. *J. Acoust. Soc. Am.*, 1980, *68*, 446-453.

Hirsh, I.J. and Burgeat, M. (1958). Binaural Effects in Remote Masking. *J. Acoust. Soc. Am.*, 1958, *30*, 827-832.

Jeffress, L.A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psych.*, 1948, *41*, 35-39.

Johnson, D.H. (1974). *The Response of Single Auditory-Nerve Fibers in the Cat to Single Tones: Synchrony and Average Discharge Rate.* PhD thesis, Elec. Eng. Dept., MIT, September, 1974.

Keen, K. (1972). Preservation of Constant Loudness with Interaural Amplitude Asymmetry. *J. Acoust. Soc. Am.*, 1972, *52*(4), 1193-1196.

Kiang, N.Y.S. (1968). A Survey of Recent Developments in the Study of Auditory Physiology. *Ann. Otol. Rhinol. Laryngol.*, 1968, *77*, 656-675.

Kiang, N.Y.S., Watanabe, T., Thomas, E.C., and Clark, L.F. (1965). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve.* Cambridge, Mass.:MIT Press, 1965. Research Monograph 35.

Kiang, N.Y.S. and Moxon, E.C. (1974). Tails of tuning curves of auditory-nerve fibers. *J. Acoust. Soc. Am.*, 1974, *55*(3), 620-630.

Klumpp, R.G. and Eady, H.R. (1956). Some Measurements of Interaural Time Difference Thresholds. *J. Acoust. Soc. Am.*, 1956, *28*(5), 859-860.

Lindemann, W. (1986). Extension of a binaural cross-correlation model by contralateral inhibition.I. Simulation of lateralization for stationary signals. *J. Acoust. Soc. Am.*, December 1986, *80*(6), 1608-1622.

Lukacs, E. (1970). *Characteristic Functions.* London:Griffin, 1970.

McFadden, D. and Pasanen, E.G. (1976). Lateralization at High-frequencies Based on Interaural Time Differences. *J. Acoust. Soc. Am.*, 1976, *59*, 634-639.

Nuetzel, J.M. and Hafter, E.R. (1976). Lateralization of Complex Waveforms: Effects of Fine Structure, Amplitude, and Duration. *J. Acoust. Soc. Am.*, 1976, *60*, 1339-1346.

Nuetzel, J.M. and Hafter, E.R. (1981). Discrimination of Interaural Delays in Complex Waveforms: Spectral Effects. *J. Acoust. Soc. Am.*, 1981, *69*, 1112-1118.

Parzen, E. (1962). *Stochastic Processes.* San Francisco:Holden-Day, 1962.

Sayers, B.McA. (1964). Acoustic-Image Lateralization Judgements with Binaural Tones. *J. Acoust. Soc. Am.*, 1964, *36*, 923-926.

Sayers, B.McA. and Cherry, E.C. (1957). Mechanism of Binaural Fusion in the Hearing of Speech. *J. Acoust. Soc. Am.*, 1957, *29*, 973-987.

Schiano, J.L., Trahiotis, C., and Bernstein, L.R. (1986). Lateralization of low-frequency tones and narrow bands of noise. *J. Acoust. Soc. Am.*, May 1986, *79*(5), 1563-1570.

Siebert, W.M. (1968). Stimulus transformations in the peripheral auditory system. In P. Kolers and M. Eden (Ed.), *Recognizing Patterns,* Cambridge, MA: MIT Press, 1968.

Siebert, W.M. (1970). Frequency Discrimination in the Auditory System: Place or Periodictiy Mechanisms. *Proc. IEEE*, 1970, *58*, 723-730.

Siegel, R.A. and Colburn, H.S. (1983). Internal and external noise in binaural detection. *Hearing Research*, 1983, *11*, 117-123.

Stern, R.M., Jr. (1976). *Lateralization, Discrimination, and Detection of Binaural Pure Tones.* PhD thesis, Elec. Eng. Dept., MIT, December, 1976.

Stern, R.M., Jr., and Colburn, H.S. (1978). Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position. *J. Acoust. Soc. Am.*, July 1978, *64*(1), 127-140.

Stern, R.M. and Colburn, H.S. (1985). Lateral-position-based models of interaural discrimination. *J. Acoust. Soc. Am.*, 1985, *77*, 753-755.

van Trees, H.L. (1968). *Detection, Estimation, and Modulation Theory. Part I.* New York:John Wiley and Sons, Inc., 1968.

Wakeford, O.S. and Robinson, D.E. (1974). Lateralization of tonal stimuli by the cat. *J. Acoust. Soc. Am.*, 1974, *55*(3), 649-652.

Woodworth, R.S. (1938). *Experimental Psychology.* New York:Holt, 1938.