Models of Binaural Interaction

Richard M. Stern
Constantine Trahiotis

I. INTRODUCTION: CROSS-CORRELATION MODELS OF BINAURAL PERCEPTION

The human binaural system has attracted the attention of auditory theorists since Lord Rayleigh formulated the duplex theory in 1907. The "modern era" of binaural modeling can be said to have begun in 1948 with Jeffress's prescient paper suggesting a neural coincidence mechanism to detect interaural time differences and, coincidentally, the original descriptions of the binaural masking level difference provided independently by Hirsch (1948) and Licklider (1948). The next 25 years witnessed an explosion in experimental studies in subjective lateralization, binaural detection, and interaural discrimination (as discussed in Chapter 9 of this volume, and in the complementary review by Hafter and Trahiotis, 1994). A number of significant efforts were made to describe these data in terms of quantitative models. For example, Sayers and Cherry (1957) formulated the first explicit cross-correlation model that was directly compared to experimental data. Webster (1951) and Jeffress, Blodgett, Sandel, and Wood (1956) called attention to the importance of stimulus variability through their seminal papers that highlighted the interaural time delay (ITD) produced by a vectorial combination of the target and masker components of stimuli used in binaural

Hearing
Copyright © 1995 by Academic Press, Inc. All rights of reproduction in any form reserved.

detection experiments. This "vector" model was later elaborated on by Hafter and others (e.g., Hafter, 1971; Yost, 1972). Other important models developed during that period include the equalization–cancellation model of Durlach (1963, 1972) and the model based on auditory-nerve activity of Colburn (1973, 1977), among several others. Two review chapters by Durlach and Colburn (1978) and Colburn and Durlach (1978), respectively, elegantly summarize most of the major experimental results and theoretical models up to 1972. Colburn (1995) has recently written a comprehensive review of binaural models that contains a thorough discussion of computational models of physiological processing.

Colburn and Durlach (1978) argued that all of the then-current binaural models could be thought of as a particular realization of the generic model of binaural interaction shown in Figure 1. This generic structure includes a series of peripheral processing steps consisting of bandpass filtering, rectification, stochastic neural representation of the signals, comparison of interaural timing information over a limited range of internal delays using a correlation or coincidence mechanism, consideration of interaural intensity differences of the outputs of the monaural processors, and a subsequent decision-making mechanism. Since 1978 there has been a general acceptance of this basic structure, and especially the cross-correlation mechanism used for the extraction of interaural timing information. The basic cross-correlation mechanism has been extended in a number of ways within more recent models. Other significant recent trends in binaural modeling include an increased reliance on computational (as opposed to analytical) approaches to predict the phenomena, as well as initial efforts to make use of head-related transfer functions to understand and reproduce out-of-head localization phenomena. In recent years our understanding of the binaural system has begun to be applied to the simulation of room acoustics for sound

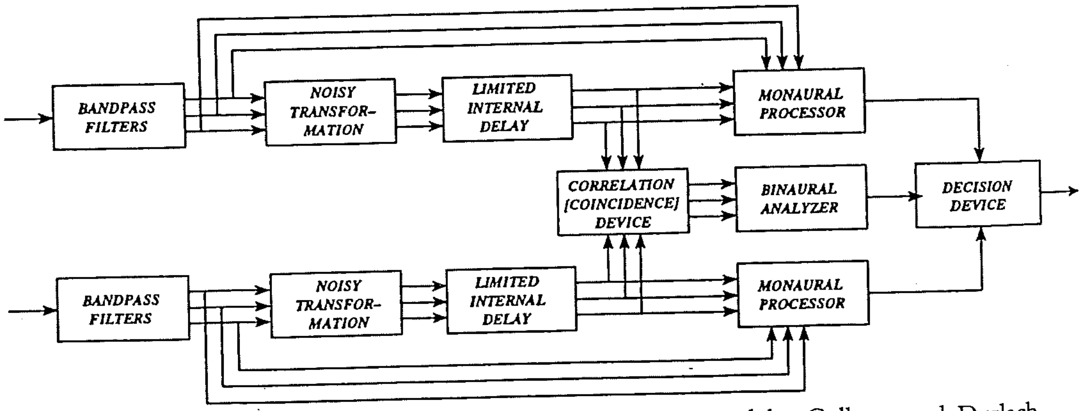


FIGURE 1 Generic model of binaural processing proposed by Colburn and Durlach (1978). The parallel sets of arrows indicate multiple parallel channels of information in the model.

presented through headphones (e.g., Bodden, 1993), to the reduction of error rates of speech recognition systems (e.g., DeSimio and Anderson, 1993; Sullivan and Stern, 1993), and in the simulation of out-of-head images for virtual environments and displays (e.g., Wenzel, 1992).

The goal of this chapter is to provide the reader with an intuitive understanding of how cross-correlation-based binaural models work and an appreciation of their capabilities and limitations in describing a variety of binaural phenomena. This chapter is more tutorial but less comprehensive than the one recently written by Colburn (1994), and our expectation is that both types of discussion of this work will be useful, especially to new researches in the field. We review the initial formulations of the cross-correlation model in Section II and describe how this structure has been recently modified in Section III. In Section IV we describe how the general cross-correlation models have been applied to psychophysical data.

II. STRUCTURE OF BINAURAL CROSS-CORRELATION-BASED MODELS

A. The Original Forms of the Cross-Correlation Model

Modern binaural models are all based on Jeffress's (1948) original conception of a neural "place" mechanism that would enable the extraction of interaural timing information. Jeffress suggested that external interaural delays could be inferred by central units that record coincidences of neural impulses from pairs of more peripheral nerve fibers. Each central unit was presumed to compare information from the two ears after a series of internal time delays, as shown in the block diagram of Figure 2. The delay mechanism is commonly conceptualized in the form of a ladder-type delay line as in Figure 2, but such a structure is not the only possible realization. A key parameter in the analysis of the outputs of such a mechanism is the interaural difference of the total delay incurred by the two monaural signals arriving at a given coincidence detector. This variable will be referred to as the *net internal delay* τ for that particular unit. The short-term average of the set of coincidence outputs plotted as a function of their internal delay τ is an approximation to the short-term cross-correlation function of the neural signals arriving at the coincidence detectors. Licklider (1959) proposed that such a mechanism could also be used to achieve an autocorrelation of neural signals for use in models of pitch perception.

A different, analytical, model based on cross-correlation was developed by Sayers and Cherry (1957) to describe their early measurements of fusion and laterality. We briefly review the major features of this model because it was the first quantitative application of cross-correlation, and because it contains many of the elements of modern models of binaural perception. In

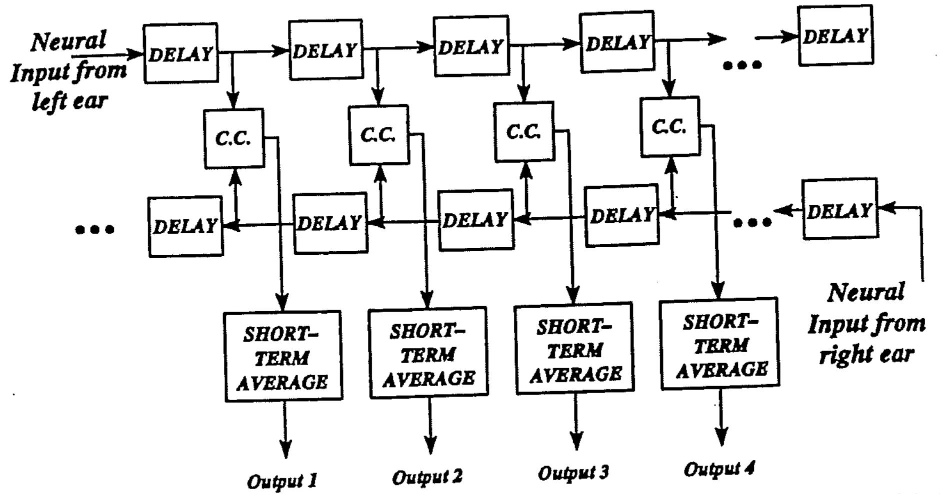


FIGURE 2 Schematic representation of the Jeffress place mechanism. The blocks labeled *C.C.* record coincidences of neural activity from the two ears (after the delays are incurred).

Sayers and Cherry's formulation the *short-term running cross-correlation function R* (τ, t) is formed from the signals to the two ears, according to the equation

$$R(\tau,t) = \int_{-\infty}^{t} x_L(\alpha)x_R(\alpha - \tau)w(t - \alpha)p(\tau)\, d\alpha$$

where $x_L(t)$ and $x_R(t)$ are the signals to the left and right ears. The function $w(t)$ represents the temporal weighting of the short-term correlation operation, and typically took on an exponential form in most of Sayers and Cherry's calculations. The function $p(\tau)$ was typically of the form $e^{-k|\tau|}$ and served to emphasize the contributions of internal delays of small magnitude. We later refer to this type of emphasis as *centrality*. As is the case with all cross-correlation-based models, an additional mechanism is needed to account for the *effects of interaural intensity difference* (IID). Sayers and Cherry added a constant proportional to the intensity of the left-ear signal to values of $R(\tau,t)$ for which τ was less than 0 and a (generally different) constant proportional to the intensity of the right-ear signal to values of $R(\tau,t)$ for which τ was greater than 0. A *judgment mechanism* then extracted subjective lateral position using the statistic

$$\hat{P} = \frac{I_L - I_R}{I_L + I_R}$$

where I_L and I_R are the integrals of the intensity-weighted short-term cross-correlation function over negative and positive values of τ, respectively.

It is important to note that the Sayers and Cherry model is based on processing the cross-correlation of the *original* signals to the two ears, rather than the cross-correlation of the neural representation of these signals after filtering by the auditory periphery. Figure 3 shows examples of such cross-correlation function, for a 500-Hz tone and for an ideal bandpass noise with center frequency 500 Hz and bandwidth 200 Hz. Each signal contains an ITD of 500 μs (0.5 ms). Both functions exhibit positive peaks spaced at the reciprocal of the center frequency, and the amplitudes of the peaks for the bandpass noise decrease at a rate that is proportional to the signal bandwidth. In each case the maximum value of the cross-correlation function occurs at the internal delay equal to the ITD of the stimulus, but for the sine wave, this maximum value also occurs at several other locations along the τ axis.

B. Colburn's Auditory-Nerve-Based Model

An influential quantification of the Jeffress hypothesis was formulated by Colburn (1973, 1977) who compared the information that could be extrac-

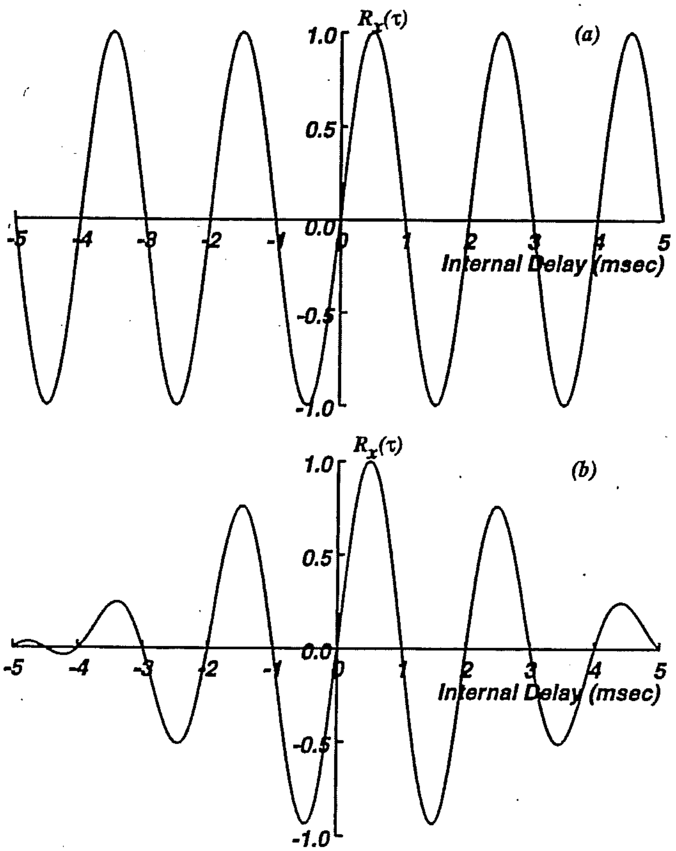


FIGURE 3 Examples of cross-correlation functions for (a) a 500-Hz pure tone and (b) bandpass noise with a center frequency of 500 Hz and bandwidth of 200 Hz. In each case the signals are presented with an ITD of 0.5 ms.

ted from the responses of populations of auditory-nerve fibers to performance in binaural detection and interaural discrimination experiments. Colburn's model consisted of two parts: a model of auditory-nerve activity and a central processor that analyzes and displays comparisons of firing times from ear to ear.

1. The Model of Auditory-Nerve Activity

The model of auditory-nerve activity used in the original Colburn model was adapted from an earlier formulation by Siebert (1970). It consists of a bandpass filter, a lowpass filter, and an exponential rectifier, followed by a mechanism that generates firing times of a nonhomogeneous Poisson process at a rate that is proportional to the output of the rectifier. Stern and Shear (1995) later modified this model by changing the shape of the nonlinear rectifier and interchanging the order of the rectifier and the low-pass filter. Functional models of similar form have been used in the work of several other researchers including Duifhuis (1973), Blauert and Cobben (1978), and Lindemann (1986a).

Colburn used the nonhomogeneous Poisson process to characterize the response of auditory-nerve fibers to sound because it is the simplest stochastic process that can realistically be applied to model the neural firing times. Each fiber is characterized by a *rate function,* $r(t)$, that describes the instantaneous rate of firing assumed to be produced by that fiber. This time-varying function depends on both the characteristic frequency (CF) of the nerve fiber and the spectral–temporal characteristics of the stimulus. Using an explicit analytical model like the Poisson process, one can calculate means and variances of the predicted outputs of the coincidence counters. These statistics can then be used to predict discrimination and detection thresholds either by application of the Cramer–Rao bound (cf. Van Trees, 1968), or by direct prediction of performance obtained by assuming that the decision variable is normally distributed. The general success of this approach notwithstanding, constructing and evaluating quantitative neurally based models is inevitably a compromise between analytical tractability and faithfulness to the known physiological results. For example, it is well known that the peripheral auditory system is both time varying (due to the refractory nature of the auditory response) and nonlinear. The Poisson-process model ignores the refractoriness in the response and analytical predictions can be developed easily only for a limited set of stimuli with quasi-static interaural differences (including pure tones, tones in noise, and bandpass noise). Furthermore, predictions are easily developed only for a limited set of assumed peripheral nonlinearities (including exponential and half-wave power-law rectifiers).

Because of the difficulty in developing analytical predictions for many interesting stimuli, including signals with significant transients and stimuli that give rise to the precedence effect, more computationally oriented (and more physiologically accurate) models of the peripheral auditory response to sound are now becoming increasingly popular (e.g., Carney, 1993; Payton, 1988; Meddis, Hewitt, & Shackleton, 1990). For example, the Meddis et al. model of auditory nerve activity has been incorporated into the binaural processing model of Shackleton, Meddis, & Hewitt (1992).

2. The Model of Central Processing

Colburn first considered for the central processing model a binaural analyzer that used general comparisons of timing information from the responses of the fibers from the two ears, but he found that more information was available to such a model than appeared to be used by humans. He found that predictions of a model that compares interaural timing information emanating only from fibers with the same CF, and with only a single internal interaural delay, were consistent with performance observed in several interaural time and intensity discrimination tasks (Colburn, 1973). The more restricted model described the processing that would be provided by the ensemble of coincidence-counting units originally proposed by Jeffress. Colburn subsequently compared the predictions of the restricted model with available data on binaural detection thresholds and found that the model could predict virtually all of those data as well (Colburn, 1977).

The response of each coincidence counter of the Colburn model is characterized by two parameters: the internal delay and CF. Colburn (1977) assumed that a coincidence is achieved only when the firings from the two input fibers are nearly simultaneous. This enabled him to assume that the output of the coincidence counter is also a Poisson process, with mean

$$E[L(\tau,f)] = T_W \int_0^T r_L(\mathrm{t}) r_R(t - \tau)\, \mathrm{d}t$$

where $L(\tau,f)$ represents the number of coincidences recorded by a unit with internal delay and CF equal to τ and f, respectively. The functions $r_L(t)$ and $r_R(t)$ are the Poisson rate functions of the two input fibers, T_W represents the duration of the coincidence window, and T represents the duration of the stimulus. For a given CF, the expected number of coincidences plotted as a function of the internal delay parameter (τ) describes the *cross-correlation of the neural representation* of the binaural signal, as determined by the rate functions of the fibers from the two ears at that CF. The function $E[L(\tau,f)]$ can also be thought of as a special case of the corresponding running cross-correlation function of the Poisson rate functions $r_L(t)$ and $r_R(t)$,

$$E[L(\tau,f)] = E[L(\tau,t,f)] = \int_{-\infty}^{t} r_L(\alpha) r_R(\alpha - \tau) w_c(t - \alpha) p(\tau)\, d\alpha$$

The function $w_c(t)$ in this equation is a temporal weighting function and emphasizes the most recent values of the cross-correlation of $r_L(t)$ and $r_R(t)$, just as the function $w(t)$ used by Sayers and Cherry provided temporal weighting for the cross-correlation of the original stimulus. For most of the research of Colburn, Stern, and colleagues, $w_c(t)$ is assumed to be a constant for $0 \leq t \leq T$ and 0 otherwise. Stern and Bachorski (1983) have developed some predictions for the statistics of the coincidence-counter outputs using an exponentially shaped $w_c(t)$, similar to the exponential function $w(t)$ proposed by Sayers and Cherry (1957).

Colburn and Durlach (1978) have noted that Colburn's auditory-nerve-based model can also be regarded as a generalization of the equalization–cancellation (EC) model of Durlach (1963). The EC model has been most successful in predicting the results of binaural detection experiments. Predictions are obtained by applying a combination of ITD and IID that produce the best "equalization" of the masker components of the stimuli presented to each of the two ears and allow "cancellation" of the resulting signals by subtraction of one from the other. Compensation for ITD has a greater impact on predictions than compensation for IID in the equalization stage of the EC model. The internal interaural delays of the fiber pairs of the Jeffress–Colburn model perform the same function as the ITD-equalizing operation of the EC model. As a result, many detection-threshold predictions for the two models are similar in form. Minor differences in predictions occur because the EC model assumes that only a single best delay is available in the equalization operation, whereas the Jeffress–Colburn structure implies that many delays are simultaneously available for processing the signals.

C. Physiological Support for Cross-Correlation in the Binaural System

A number of physiological studies have described cells that are likely to be relevant to binaural processing, as summarized by the discussion in Chapter 3, along with the review chapters of Kuwada and Yin (1987) and Colburn (1995). For example, cells that appear to record IIDs have been reported in the superior olivary complex, inferior colliculus, and other sites (e.g., Boudreau & Tsuchitani, 1968; Goldberg & Brown, 1969). Of particular interest to the developers of models based on the Jeffress–Colburn coincidence mechanism are cells first reported by Rose, Geisler, and Hind (1966) in the inferior colliculus that appear to be maximally sensitive to signals presented with a specific interaural delay, independent of frequency. This delay is referred to as a *characteristic delay*. Cells with a similar response have been reported by others in the medial superior olive (e.g., Goldberg & Brown, 1969; Crow, Rupert, & Moushegian, 1978; Yin & Chan, 1990), and the dorsal nucleus of the lateral lemniscus (e.g., Brugge, Anderson, & Aitkin, 1970). A series of measurements has been performed that characterizes the distribution of ITD-sensitive cells in the inferior colliculus (Yin & Kuwada, 1984; Kuwada, Stanford, & Batra, 1987), and the medial geniculate body (Stanford, Kuwada, & Batra, 1992). Although most cells exhibit characteristic delays that fall within the maximum delay possible for a point source in a free field for a particular animal, a substantial number of ITD-sensitive cells have characteristic delays that fall outside the "physically plausible" range.

The anatomical origin of the characteristic delays has been the source of some speculation. Many physiologists believe that the delays are of neural origin, caused either by slowed conduction velocity or synaptic delays (e.g., Smith, Joris, & Yin, 1993; Carr & Konishi, 1990; Young & Rubel, 1983). Schroeder (1977) proposed an alternative hypothesis, suggesting that the characteristic delays could also be obtained if higher processing centers compare timing information derived from auditory-nerve fibers with different CFs. This hypothesis is also a part of the model proposed more recently by Shamma, Shen, and Gopalaswamy (1989), which they call the *stereausis model*. Shamma's model for central processing is very similar to the general structure proposed by Jeffress and quantified by Colburn, and it has been implemented as an integrated circuit by Lazzaro, Mead, and colleagues (Lazzaro, 1991; Mead, Arreguit, & Lazzaro, 1991). The stereausis model is not currently specified in sufficient detail to enable its predictions to be compared critically either to predictions of other models or to the corresponding experimental data. In general, the predictions of binaural models are unaffected by whether the internal delays are assumed to be caused by neural or mechanical phenomena.

D. Temporal Integration of the Coincidence Display

Although the binaural system can be shown to resolve static ITDs as small as tens of microseconds (e.g., Klumpp & Eady, 1956; Zwislocki & Feldman, 1956), experiments measuring responses to *time-varying* ITDs (e.g., Licklider, Webster, & Hedlun, 1950; Grantham & Wightman, 1978) indicate a much more "sluggish" response, with limits on the order of tens of milliseconds, as discussed in Chapter 9 of this volume. To understand these apparently diverging sets of results, one must recall that results in static ITD-discrimination experiments reflect changes in place of activity of the coincidence-counting units along the internal-delay axis. Hence, thresholds would reflect the density of fiber pairs with respect to internal delay at each

CF. Resolution of time-varying interaural differences, on the other hand, reflects temporal integration or the averaging of instantaneous responses over *running time* (as opposed to interaural time delay).

To understand binaural "sluggishness," it is helpful to think of the temporal averaging of the matrix of coincidence-counting units as the output of a linear filter that has an impulse response equal to the temporal weighting function $w(t)$ and an input equal to the instantaneous interaural cross-correlation of the neural response to the signals, $r_L(t)r_R(t - \tau)$. Figure 4 demonstrates how $E[L(t,\tau,f)]$, the expected value of the instantaneous number of coincidences, varies as a simultaneous function of internal delay (τ) and running time (t) with and without temporal integration. The stimulus in each case is a 500-Hz tone, and responses are depicted for fibers with that frequency as the CF. In the upper panel the instantaneous value of $r_L(t)r_R(t - \tau)$ is plotted without averaging over running time. Note that peaks appearing in this panel are limited to particular intervals of the running time, t (as well as occurring at particular values of τ). The lower panel of Figure 4 shows the same function, but *after* integration by convolution

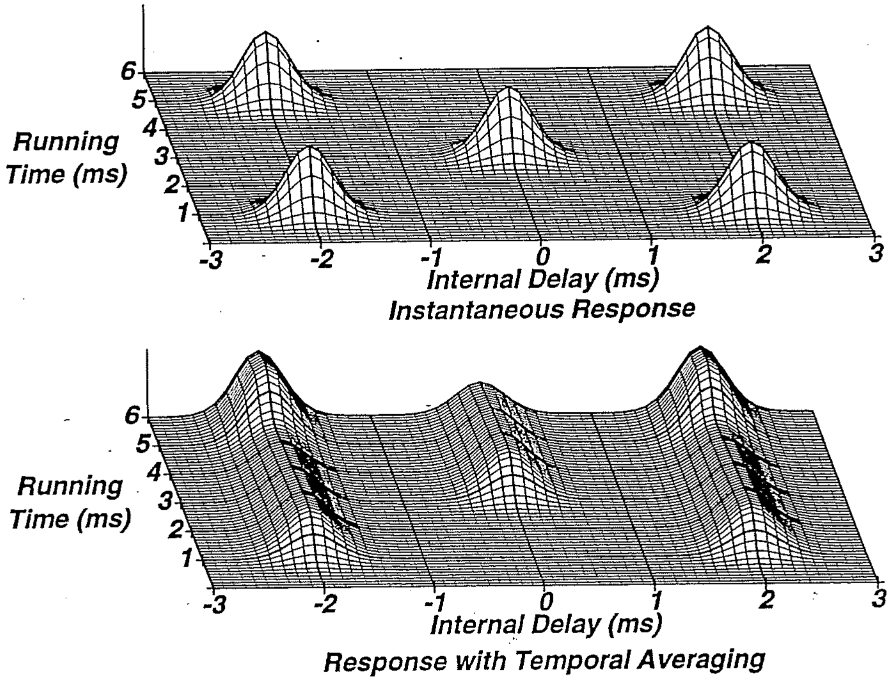


FIGURE 4 The expected value of the instantaneous number of coincidences, $E[L(t, \tau, f)]$, as a simultaneous function of running time t and internal delay τ to a 500-Hz tone with zero ITD. The response is shown using no temporal integration (upper panel) and using temporal integration by an exponentially shaped temporal weighting function with an effective cutoff frequency of 5 Hz in the frequency domain.

with an exponential time window of the form $w(t) = e^{-10\pi t}$ for positive values of t. This is the impulse response of a single-pole low-pass filter with cutoff frequency 5 Hz, which is typical of the types of integrating filters considered by Grantham and Wightman (1978) and others for accounting for sensitivity to temporal modulation of ITDs. Note that temporal integration causes the isolated peaks in the instantaneous cross-correlation shown in the upper panel to be transformed to smoother ridges that are parallel to the t-axis. We believe that this enables the binaural system to provide a stable spatial representation of the acoustic world.

III. EXTENSIONS OF THE CROSS-CORRELATION APPROACH

Since the time of Colburn's original formulation, several research groups have extended the structure and application of the frequency-dependent cross-correlation analyzer in a number of different ways. We summarize a number of these extensions to the Jeffress–Colburn model in this section.

A. Extensions by Stern, Colburn, and Trahiotis

The goals of the work of Stern and his colleagues (Stern & Colburn, 1978, 1985; Stern & Bachorski, 1983; Stern, Shear, & Zeppenfeld, 1988a; Stern & Trahiotis, 1992; Stern & Shear, 1995) have been to determine ways in which the subjective lateral position of binaural stimuli can be related to the activity of the coincidence-counting units and in turn to examine the extent to which objective interaural discrimination and binaural detection results can be related to changes in predicted lateral position. Colburn's coincidence-counting mechanism was extended by adding explicit assumptions concerning time–intensity interaction as well as a mechanism for extracting subjective lateral position from the modified display (Stern & Colburn, 1978). This extension of the Jeffress–Colburn model is referred to as the *position-variable model* by Stern and his colleagues. To predict quantitatively various data concerning lateral position, Stern and colleagues (Stern et al., 1988a; Stern & Shear, 1995) also slightly modified the characterization of auditory-nerve activity and the description of the function that specifies the distribution of internal delays. In addition they proposed a second coincidence-based mechanism that emphasizes the impact of ITDs that are consistent over a range of frequencies (Stern and Trahiotis, 1992).

The function specifying the distribution of internal delays plays an important (but frequently unrecognized) role in developing predictions of subjective lateral position. Colburn (1969, 1977) originally assumed that the density function for internal delays, called $p(\tau)$, was independent of frequency, and he fitted the shape of $p(\tau)$ to predict the relative masking level differences for stimuli in the N_0S_π *vs.* $N_\pi S_0$ conditions, as defined in Chapter 9 of this volume. More recently, Stern and Shear (1995) made this function weakly dependent on frequency and changed its shape slightly to describe the observed dependence of the lateralization of tonal stimuli with a fixed ITD on stimulus frequency (Schiano, Trahiotis, & Bernstein, 1986).

The effect of the frequency-dependent density function for internal delay, $p(\tau,f)$, on the representation of a 500-Hz pure tone with an ITD of +0.5 ms is demonstrated in Figure 5. This figure shows the average total number of coincidences recorded by the coincidence-counting units as a joint function of internal delay (along the horizontal axis) and CF (along the oblique axis). Note that even though the stimulus is tonal, the spread of excitation result-

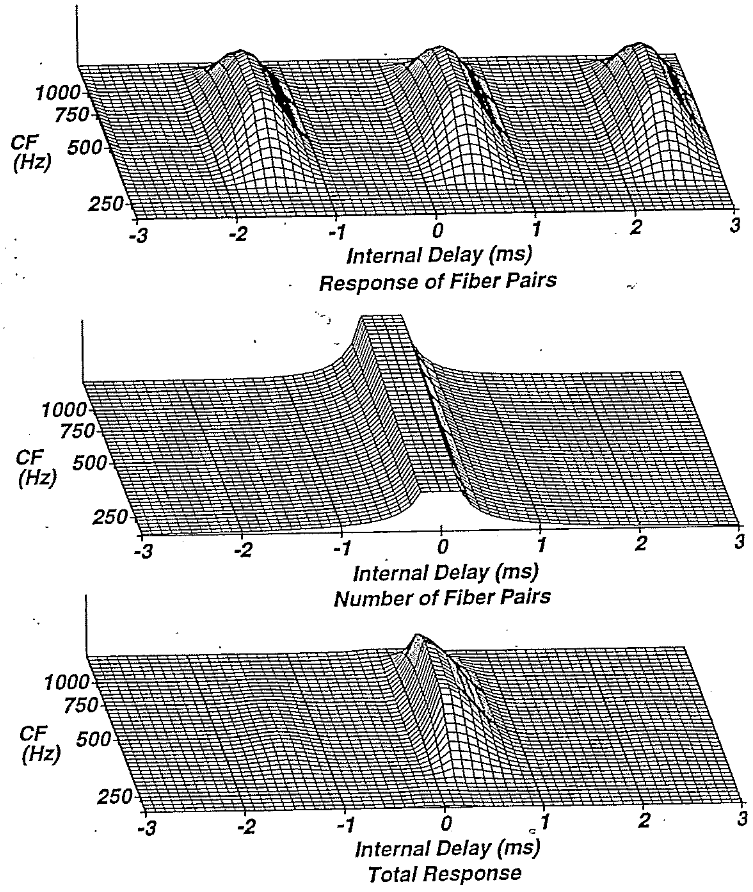


FIGURE 5 Cross-correlation patterns showing the response of an ensemble of binaural fiber pairs to a 500-Hz pure tone with a 0.5-ms ITD. Upper panel: the relative number of coincidences per fiber pair as a function of internal delay τ (in ms) and CF of the auditory-nerve fibers (in Hz). Central panel: the function $p(\tau,f)$, which describes the assumed distribution of internal delays as a function of CF. Lower panel: $E[L_T(\tau,f)]$, the expected total number of coincidences as a function of internal delay and CF, which is the product of the upper and central panels.

ing from finite-bandwidth peripheral filtering produces a synchronized response over a fairly wide range of CFs (cf. Pfeiffer and Kim, 1975). The upper panel shows $E[L(\tau,f)]$, the average number of coincidences per fiber pair. The center panel shows the function $p(\tau,f)$ that describes the distribution of fiber pairs as a function of internal delay and CF. The lower panel displays $E[L_T(\tau,f)]$, the average total number of coincidences at each internal delay and CF, which is the product of the number of counts per fiber pair [$L(\tau,f)$, upper panel] and the number of fiber pairs [$p(\tau,f)$, central panel]. At each CF there is a distinct maximum in the cross-correlation function at a value of internal delay that is close to that of the original interaural delay of the stimulus.

The form of the function $p(\tau,f)$ implies that there are relatively more coincidence-counting units with internal interaural delays of smaller magnitude, which has been confirmed by physiological measurements (e.g., Kuwada et al., 1987). Nevertheless, a substantial fraction of the coincidence counters is assumed to have internal delays that are much greater in magnitude than the largest delays that are physically attainable with free-field stimuli. The existence of very long internal delays is in accord with psychoacoustical as well as physiological data.

As mentioned earlier in connection with the model of Sayers and Cherry (1957), any correlation-based binaural model must include an additional mechanism to incorporate the effects of IIDs on lateralization. The approach taken by Stern and Colburn (1978) was to multiply the function $L_T(\tau,f)$ by a Gaussian-shaped weighting function, referred to as $L_I(\tau,f)$, with location along the τ axis dependent on IID:

$$L_p(\tau,f) = L_T(\tau,f)L_I(\tau,f)$$

Stern and Colburn (1978) proposed that the predicted lateral position of a stimulus, $\hat{P}$, can be obtained by computing the centroid along the τ axis of the position function $L_p(\tau,f)$ while averaging over frequency:

$$\hat{P} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tau L_P(\tau,f)\, d\tau df}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L_P(\tau,f)\, d\tau df}$$

This definition of predicted lateral position was originally adopted by Stern and Colburn for reasons of computational simplicity, and it has been employed by Blauert and his colleagues (e.g., Lindemann, 1986a) as well. It should be noted, however, that models like the position-variable model that predict the intracranial location of only a single image are unable to explain experimental results that suggest the existence of multiple images, such as the studies by Moushegian and Jeffress (1959), Whitworth and Jeffress (1961), and Hafter and Jeffress (1968).

Recently, Stern and Trahiotis (1992) have incorporated an additional modification to the model that is designed to emphasize the modes of the function $E[L(\tau,f)]$ that appear at the same internal delay over a range of CFs. These modes are referred to as the *straight* modes of $E[L(\tau,f)]$, and this weighting mechanism will be discussed in detail in Subsection IV.A.2.

B. Extensions by Blauert, Cobben, Lindemann, and Gaik

Blauert and his colleagues have made important contributions to correlation-based models of binaural hearing over an extended period of time. Their efforts have been directed primarily toward understanding how the binaural system processes more complex sounds in real rooms and have tended to be computationally oriented. This approach is complementary to that of Colburn and his colleagues, who have focused on explaining "classical" psychoacoustical phenomena using stimuli presented through earphones. In recent years Blauert and his colleagues have been applying knowledge gleaned from fundamental research in binaural hearing to help develop a "cocktail party processor" that can identify, separate, and enhance individual sources of sound in the presence of other, interfering sounds.

In the first English-language description of their modeling efforts, Blauert and Cobben (1978) combined the running cross-correlator of Sayers and Cherry (1957) with the model of the auditory periphery suggested by Duifhuis (1973). This model of peripheral processing is functionally similar to the model proposed by Siebert and adopted by Colburn. Blauert and Cobben described the response of the model to single clicks and to pairs of clicks, presented from spatially separated loudspeakers in an anechoic chamber. They found that a characterization of the average response to a sound in a particular frequency-specific auditory-nerve channel was adequate for their purposes (and that a characterization of the individual stochastic neural firing times was unnecessary).

Blauert and his colleagues subsequently developed a series of mechanisms that explicitly introduced the effects of stimulus IIDs into the modeling process. One of the most interesting and best known of these mechanisms is the one proposed by Lindemann (1986a), which may be regarded as an extension and elaboration of an earlier hypothesis of Blauert (1980). Lindemann extended the original Jeffress coincidence-counter model in two ways, adding (1) a mechanism that inhibits outputs of the coincidence counters when activity is produced by coincidence counters at adjacent internal delays, and (2) monaural-processing mechanisms at the "edges" of the display of coincidence-counter output that become active when the intensity of the signal to one of the two ears is extremely small.

The inhibitory mechanisms of the Lindemann model produce a "sharpening" of the peaks of the coincidence-counter outputs along the internal-delay axis. This is illustrated in Figure 6, which compares the temporal evolution of the response of the outputs of coincidence counters with a CF of 500 Hz to a pure tone with a sinusoidally varying ITD with and without the types of inhibition proposed by Lindemann (Palm, 1989). As in Figure 4, the oblique axis represents *running* time, and the horizontal axis represents internal delay. In calculating these responses we included both the static-inhibition and dynamic-inhibition components defined by Lindemann (1986a).

A very interesting property of the Lindemann model is that the interaction of the inhibition mechanism and the monaural processing mechanism causes the locations of peaks of the coincidence-counter outputs along the

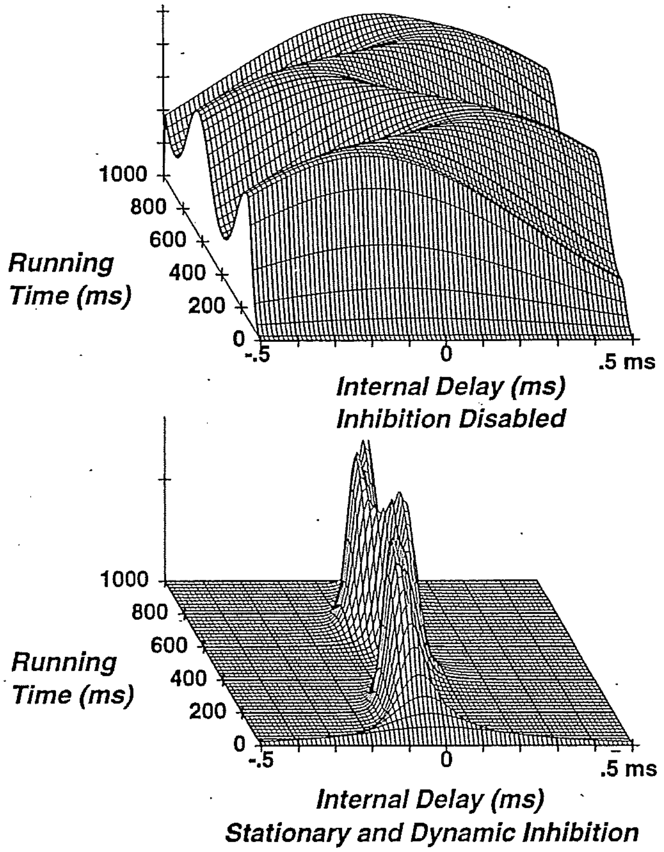


FIGURE 6 Plots of $E[L(t, \tau, f)]$, the expected value of the instantaneous number of coincidences as a simultaneous function of running time t and internal delay τ, to a 500-Hz tone with an ITD that is sinusoidally modulated at a frequency of 2 Hz. Response is shown without the stationary and dynamic inhibition mechanisms of the model of Lindemann (1986a) (upper panel) and with these inhibition mechanisms (lower panel).

internal-delay axis to shift with changes in IID. In other words, this model produces a time–intensity trading mechanism at the level of the coincidence-counter outputs. Although the net effect of IIDs on the patterns of coincidence-counter outputs in the Lindemann model is not unlike the effect of the intensity-weighting function $L_I(\tau,f)$ in the model of Stern and Colburn (1978), the time–intensity interaction of the Lindemann model is more esthetically satisfying to many because it arises naturally from the fundamental assumptions of the model rather than as the result of the imposition of an arbitrary weighting function. It has not yet been possible to determine the physiological plausibility of the Lindemann inhibition mechanism, but this issue is the currently the object of current investigation in the laboratories of Tom C. T. Yin and Shigeyuki Kuwada.

Gaik (1993) extended the Lindemann mechanism by adding a further weighting to the coincidence-counter outputs that reinforces naturally occurring combinations of ITD and IID. This has the effect of causing physically plausible stimuli to produce coincidence outputs with a single prominent peak that is compact along the internal-delay axis and consistent over frequency. Conversely, very unnatural combinations of ITD and IID (which tend to give rise to multiple spatial images) produce response patterns with more than one prominent peak along the internal-delay axis.

C. Mechanisms for Time–Intensity Interaction and Image Formation

1. Time–Intensity Interaction

As noted earlier, all correlation-based models must include some mechanism to describe the effects of IID on subjective lateral position. At one time, it was felt that the effects of IIDs in binaural lateralization could be accounted for by the decrease in latency of the auditory nerve response that occurs as the intensity of the signals is increased. This *peripheral* time–intensity trading mechanism, known as the *latency hypothesis,* was discussed by Jeffress in 1948 and later elaborated by David, Guttman, and van Bergeijk (1958) and Deatherage and Hirsh (1959). This hypothesis was at least qualitatively supported by early lateralization studies that utilized small ITDs and IIDs, but it cannot describe either lateralization data over a wider range of stimulus conditions such as the stimuli presented by Sayers (1964), Domnitz and Colburn (1977), and Bernstein and Trahiotis (1985), or the inability to trade time and intensity differences completely, as first shown by Hafter and Carrier (1972).

We have described three explicit *more central* mechanisms that have been proposed to account for the effects of IID: the multiplicative intensity-weighting pulse of Stern and Colburn (1978), the mechanism involving lateral inhibition of the coincidence-counting response along adjacent delays proposed by Lindemann (1986a), and the separate weighting of the left and right halves of the cross-correlation function of the data as used in the model of Sayers and Cherry (1957). The Stern–Colburn and Lindemann models provide similar predictions for the lateralization of 500-Hz pure tones as a joint function of ITD and IID, which are not as well described by the original Sayers and Cherry model.

In general, no focused attempt has been made to evaluate critically these or other intensity-weighting mechanisms on the basis of their ability to describe the lateralization of stimuli other than 500-Hz pure tones. Nevertheless, the general consensus among contemporary theoreticians is that the time–intensity interaction takes place at the level at which timing information from the signals to the two ears is first compared, if not more centrally, rather than at the level of the auditory nerve.

2. Image Formation

Thus far we have described two specific ways of predicting lateral position from the display of interaural coincidence-counting units: computation of the centroid along the internal-delay axis, as is done by Stern and Colburn (1978) and Lindemann (1986a), and the comparison of activity along the right and left halves of the internal-delay axis, as is done by Sayers and Cherry (1957). Other ways of predicting lateral position exist as well. One plausible alternative is to assume that position can be related to the location of the peaks of the cross-correlation function (as opposed to the centroid). The locations of the peaks of the cross-correlation function allow one to account for the multiple images that can occur for tonal stimuli presented interaurally out of phase (e.g., Sayers, 1964; Yost, 1981), as well as for the secondary "time image" observed for some stimuli presented with conflicting ITDs and IIDs (e.g., Whitworth & Jeffress, 1961; Hafter & Jeffress, 1968). Lateralization mechanisms based on the peaks of functions characterizing the response of the coincidence counters have been discussed by Lindemann (1986a) in conjunction with some of the predictions of his model. More recently, Shackleton et al. (1992) described a model that assumes that the listener computes either the centroid along the τ-axis or the locations of the peaks of the responses of coincidence-counting units, choosing the statistic that more accurately describes the results for a given experiment. Although definitely not parsimonious, this type of approach may be necessary to account for the data in all their complexity.

Few focussed efforts have been made to assess the relative merits of the various ways of generating an estimate of lateral position from the pattern of activity of the coincidence-counting units. For the most part, the specific lateralization mechanism adopted by a given researcher appears to have been selected more for convenience than on the basis of strongly held principles.

IV. ABILITY OF CROSS-CORRELATION MODELS TO DESCRIBE PSYCHOACOUSTICAL DATA

In this section we describe how the patterns of activity of the matrix of coincidence-counting units of the Jeffress–Colburn model and its extensions can describe some of the phenomena that have been important for researchers in binaural perception. In each case, we provide intuitive examples of some of the various ways in which interaural timing information may be utilized in making observations and forming decisions. We then summarize the characteristics and limitations of the ability of the models to describe the phenomena.

A. Subjective Lateral Position

1. Lateralization of Pure Tones

The comparison of theoretical predictions to experimental data concerning lateralization of pure tones is complicated by the fact that the experimental data differ across studies. For example, Domnitz and Colburn (1977) describe a subjective image that returns to the center of the head as the interaural phase difference (IPD) of 500-Hz pure tones approaches ± 180°, while other researchers such as Sayers (1964) and Yost (1981) describe multiple images appearing at each of the two ears at these IPDs. By all accounts, the perceived image (or images) tends to be diffuse and labile under these conditions.

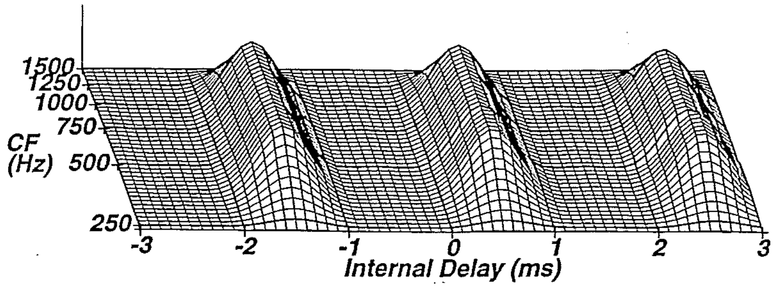
The response of the ensemble of coincidence-counting units to a pure tone has already been discussed in Section III.A, and it is shown in Figure 5. As discussed in Section III.A, models that base their predictions of lateralization on the location of a single centroid along the internal-delay axis cannot describe the perception of multiple images. This shortcoming notwithstanding, models that use the centroid to compute a single image, such as the position-variable model of Stern and Colburn (1978), provide reasonably accurate predictions for a number of fundamental aspects of the lateralization of pure tones based on *ongoing* ITDs and IIDs. These aspects include (1) the periodicity of lateral position with respect to ITD; (2) the joint dependence of the lateralization of low-frequency pure tones on ITD and IID as seen in lateralization studies that describe time–intensity trading, the cue-reversal phenomenon (in which the direction of apparent motion of the image reverses as ITDs approach half the period of the tone) and, as IID increases, the inability to describe the dependence of lateralization on a peripheral conversion of IIDs into equivalent ITDs (e.g., Sayers, 1964; Domnitz & Colburn, 1977); (3) the approximately constant lateral position of a pure tone with a fixed ITD over a range of frequencies (Schiano et al., 1986); and (4) the trajectories of images produced by low-frequency stimuli presented with small interaural frequency differences (i.e., the so-called binaural beats) (e.g., Licklider et al., 1950).

2. Lateralization of Low-Frequency Bandpass Noise

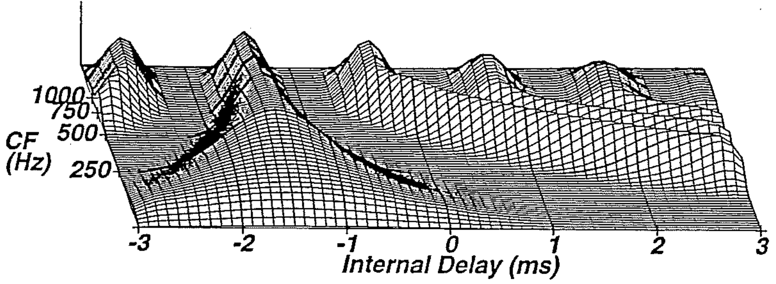
In recent years greater attention has been focused on the lateralization of spectrally and temporally more complex stimuli, such as bandpass noise and amplitude-modulated tones. We discuss the lateralization of bandpass noise and amplitude-modulated tones separately, as different issues arise in understanding the processes by which they are lateralized.

Figure 7 shows the response of the coincidence-counting units to narrowband noise presented with a center frequency of 500 Hz and two different bandwidths, 50 Hz (upper panel) and 800 Hz (lower panel). In both cases the stimuli have an ITD of −1.5 ms. The response pattern for the noise with the bandwidth of 50 Hz looks very similar to the pattern that is observed for 500-Hz tones presented with the same ITD. [For 500 Hz tones, an ITD of −1.5 ms is equivalent to an ITD of +0.5 ms, which is shown in Figure 5 (upper panel).] The dichotic stimulus with the 50-Hz bandwidth is, in fact, lateralized on the "wrong" side of the head (i.e., the right side), whereas for larger bandwidths the sound becomes lateralized toward the left side of the head (Stern, Zeiberg, & Trahiotis, 1988b; Trahiotis and Stern, 1989). This occurs because the response to noise with a bandwidth of 50 Hz exhibits parallel maxima that appear alike (as in the upper panel of Figure 7), and it is not obvious that the true stimulus delay is −1.5 ms in this case (rather than, for example, +0.5 ms). With greater stimulus bandwidths, however, the cross-correlation function exhibits modes at internal delays of −1.5 ms over a broad range of frequencies, and it becomes obvious that this is the true ITD. This effect is illustrated for a noise with an 800-Hz bandwidth in the lower panel of Figure 7.

We refer to the consistency over frequency of the maxima of the coincidence-count response that indicates the true ITD as *straightness*. By independently manipulating ITD, IPD, and bandwidth, we have found that the binaural system appears to apply special emphasis to the straight components of the response to bandpass-noise stimuli such as those shown in Figure 7 (Stern et al., 1988; Trahiotis and Stern, 1989). We have developed two extensions of the general cross-correlation model to describe this concept in a quantitative fashion. The first model was a black-box formulation called the *weighted-image model* (Stern et al., 1988b). In this model, each ridge of maxima of the two-dimensional function $L_T(\tau,f)$ is weighted by an *ad hoc* function that is approximately proportional to the reciprocal of the variance of the internal delay of the mode over frequency. This function serves as an empirical estimate of the straightness of the mode. The weighted-image model successfully described the phenomena to which it



Bandwidth 50 Hz

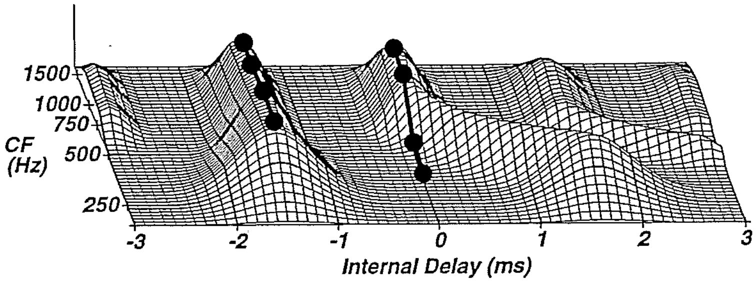


Bandwidth 800 Hz

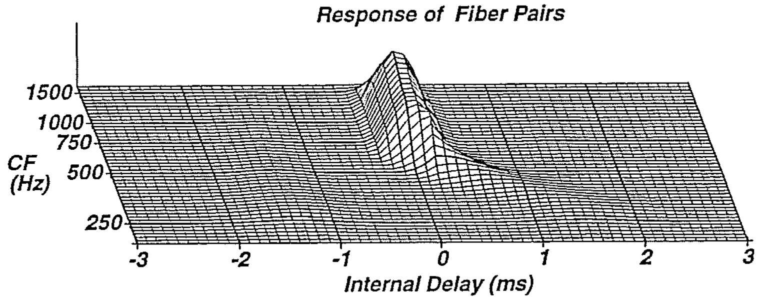
FIGURE 7 The response of an ensemble of coincidence-counting units to low-frequency bandpass noise with a center frequency of 500 Hz and an ITD of −1.5 ms. Upper panel: response to bandpass noise with a bandwidth of 50 Hz. Lower panel: response to bandpass noise with a bandwidth of 800 Hz.

has been applied. Nevertheless, it has always been considered to be an interim formulation because it cannot easily be generalized to enable predictions for many interesting stimuli that do not produce isolated ridges of maxima of $L_T(\tau,f)$.

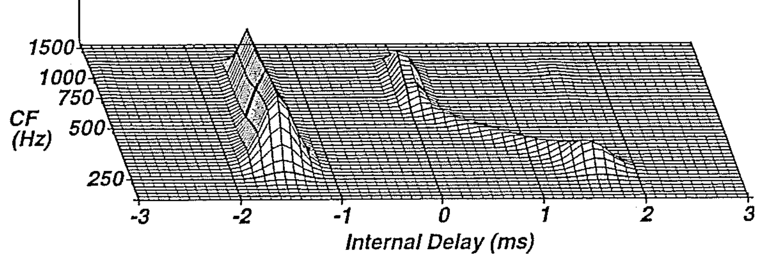
A more satisfying explanation for the straightness-weighting phenomenon was more recently proposed by Stern and Trahiotis (1992). This model, referred to as the *extended position-variable model,* assumes that the outputs of the coincidence-counting units are passed through a *second* level of coincidence-counting units. Each set of inputs to this second layer of temporal processing is assumed to come from first-level coincidence counters representing a range of CFs, but with a common internal delay. The effect of this type of processing is illustrated in Figure 8, which compares the response of the original model (without any additional straightness weighting) and the response of the extended model. The stimulus in this figure is bandpass noise centered at 500 Hz with an ITD of −1.5 ms and a bandwidth of 400 Hz. The sets of points denoted by the filled circles in the upper panel of Figure 8 are examples of combinations of CF and internal delay that would constitute inputs to the second-level coincidence counters. The center panel of Figure 8 shows the effect of weighting by the relative number of fiber pairs, which suppresses the effects of the responses at the true ITD of −1.5 ms. The lower panel of Figure 8 shows the dramatic effects of applying the second level of coincidences, which provides much greater emphasis to the straight ridge at −1.5 ms. This occurs because, for that ridge, all of the first-level coincidence counters are firing at rates at or near their maximum output. In contrast, the ridge closer to the midline (i.e., at an ITD of approximately 0) is attenuated because of the minimal response at characteristic frequencies below approximately 600 Hz at that ITD. In addition, this manner of weighting straightness also sharpens the ridges of the two-dimensional cross-correlation function along the internal-delay axis. It is



Response of Fiber Pairs



Total Response before
Second-Level Coincidence Counters



Total Response after
Second-Level Coincidence Counters

FIGURE 8 The effect of the putative secondary level of coincidence-counting units that produce "straightness weighting." Upper panel: cross-correlation patterns showing the response of an ensemble of binaural fiber pairs to noise with an ITD of −1.5 ms, a center frequency of 500 Hz, and a bandwidth of 400 Hz. Locations of constant internal delay but different CFs are identified by filled circles joined by lines. Central panel: same as upper panel, but incorporating the effects of the relative number of fiber pairs, as specified by the function $p(\tau, f)$. Lower panel: same as central panel, but after further processing by the second-level units that compute coincidences over frequency of the outputs of the original coincidence counters with the same internal delay.

important to note that "sharpening" along the internal-delay axis can occur without the explicit lateral-inhibition network proposed by Lindemann (1986a).

Shackleton et al. (1992) provide a different point of view, arguing that these data can be predicted by simply averaging the response of the coincidence-counting units over frequency, *without* any explicit mechanism that weights more heavily the straighter modes of the two-dimensional cross-correlation functions. There are at least two possible reasons why Shackleton et al. and the present authors differ in their conclusions concerning the modeling of these data. First, the model of Shackleton et al. lacks an explicit function like $p(\tau,f)$ to specify the distributions of fiber pairs with respect to internal delay and CF. It is not obvious that it will predict the much wider range of phenomena addressed by the position-variable model. Second, Shackleton et al. make use of the more detailed computational description of the auditory-nerve response to the stimuli based on the work of Meddis et al. (1990), rather than the analytical characterization used by Colburn, Stern, and their colleagues, which is simpler but not as descriptive of the physiological data. Setting aside these distinctions, we believe that the experimental data of Stern et al. (1988b) are more accurately described by the predictions of the extended position-variable model that explicitly includes straightness weighting (Stern and Trahiotis, 1992, Fig. 6) than by the predictions of the model of Shackleton et al. (1992, Figs. 2a and 2c). In our opinion, the predictions obtained by Stern and Trahiotis without straightness weighting (1992, Fig. 6) are not unlike those generated by the model of Shackleton et al.

The position-variable model as extended by Stern and Trahiotis (1992) appears to be able to describe most experimental results on the lateralization of low-frequency bandpass noise as a joint function of ITD, IPD, and bandwidth, when the signals are presented with equal amplitude to the two ears. Nevertheless, it does not describe the some of the results of Buell, Trahiotis, and Bernstein (1994), which describe the joint dependence of the lateral position of bandpass noise on ITD, IID, IPD, and bandwidth. (For example, the lateralization of stimuli with an ITD of 0 ms and an IPD of 270° is not affected by bandwidth for any IID, whereas the lateralization of stimuli with an ITD of 1.5 ms and an IPD of 0° is greatly affected by bandwidth for many IIDs.) Despite a concerted effort by Tao (Tao, 1992; Tao & Stern, 1992), which included the development of an alternative additive combination of interaural timing and intensity information, these data of Buell et al. (1994) appear to pose a continuing challenge for all models of binaural interaction.

3. Lateralization of Low-Frequency Amplitude-Modulated Tones

It has been known since the mid-1970s that the binaural system can lateralize high-frequency stimuli on the basis of the ITDs of their low-frequency envelopes (e.g., Henning, 1974; McFadden & Pasanen, 1976). In contrast, it had been believed until more recently that the lateralization of low-frequency stimuli was based solely on the ITD of their fine structure and that the ITD of the envelope had no impact on subjective lateralization. Using amplitude-modulated (AM) 500-Hz tones, Bernstein and Trahiotis (1985) demonstrated that the lateral position of low-frequency AM stimuli was affected, albeit by a small amount, by the ITD of the envelope of the stimulus, as well as by the ITD of its fine structure.

Figure 9 shows the response of the coincidence-detecting units to a 500-Hz low-frequency tone presented without amplitude modulation (upper panel), and with 100% sinusoidal amplitude modulation at a rate of 50 Hz (lower panel). The ongoing interaural delay is −1.5 ms in both cases. Bernstein and Trahiotis (1985) have shown that, although both stimuli are lateralized toward the right side of the head, the perceived location of the signal with the amplitude modulation (producing the response in the lower panel) is slightly to the left of that of the pure tone that produces the response curves in the upper panel. There are several possible causes for this. For example, Bernstein and Trahiotis suggested that the AM tone is perceived to the left of the pure tone because of the salience of low-frequency envelope cues. Stern et al. (1988b) suggested that the AM tone is

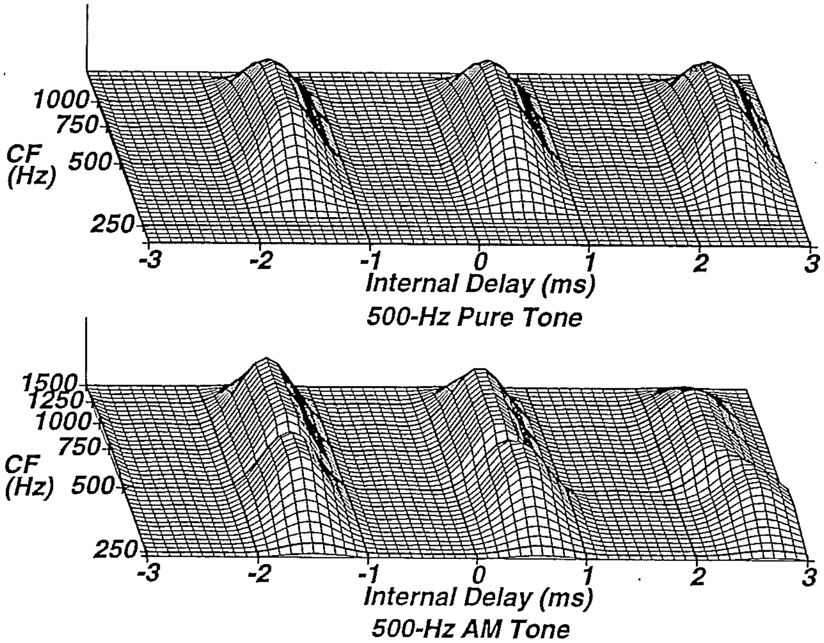


FIGURE 9 Patterns of coincidence-counting activity showing the effects of amplitude modulation on low-frequency tones. Upper panel: the response to a 500-Hz tone with a waveform ITD of −1.5 ms. Lower panel: the response to a 500-Hz tone with the same waveform delay and amplitude modulated with a modulation frequency of 50 Hz.

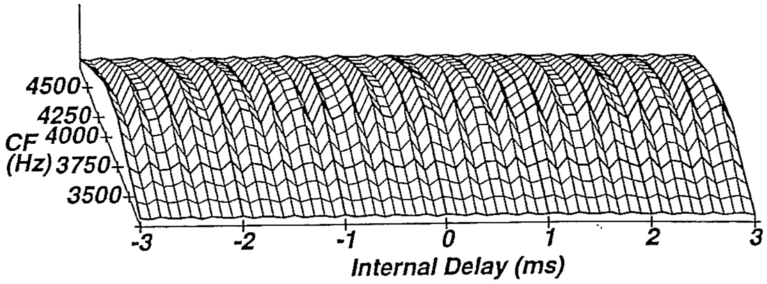
perceived to the left of the pure tone because the ridge at +0.5 ms is not as straight as the ridge at −1.5 ms. Nevertheless, we now believe that the most likely reason for the AM tone to be perceived to the left of the pure tone is simply because the peaks in the response to the AM stimulus are unequal in amplitude, with the peak (in the lower panel) at the true ITD (−1.5 ms) being greatest in size. In contrast, the peaks of the response to the pure tone (depicted in the upper panel) are all of equal amplitude. This would cause the centroid along the internal-delay axis of the response to the AM tone to be "pulled" farther toward the left side.

Stern et al. (1988a) and Stern and Shear (1995) have reported that the extended position-variable model correctly predicts the dependence of lateral position for 500-Hz AM tones on ongoing ITD and modulation frequency measured by Bernstein and Trahiotis (1985). The extended model also describes other aspects of the lateralization of low-frequency stimuli with complex envelopes, including the dependence of lateral position on pure modulator delay measured by Bernstein and Trahiotis (1985) (Stern, Zeppenfeld, & Shear, unpublished).

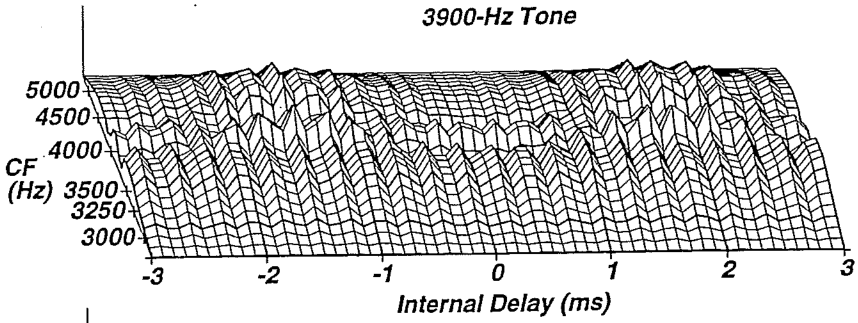
4. Lateralization of High-Frequency Amplitude-Modulated Tones and Bandpass Noise

As noted in the preceding section, the lateral position of high-frequency binaural stimuli with low-frequency envelopes such as AM tones and bandpass noise can be affected by the ITD of the envelope. Figure 10 illustrates how such stimuli are represented by the ensemble of coincidence-counting units. These plots were produced without the use of any additional explicit envelope extraction mechanism other than the low-pass filtering in the model for auditory-nerve activity. The low-pass filter has a frequency response that decreases linearly from 1200 Hz to 5200 Hz, as suggested by the physiological data of Johnson (1980), and the minor ripples in the plots show the effects of the residual energy at the relatively high carrier frequency after processing by the low-pass filter. The upper panel of Figure 10 shows the relative number of coincidences observed in response to a pure tone of frequency 3900 Hz. The central panel depicts the response to an AM tone with a carrier frequency of 3900 Hz and a modulation frequency of 300 Hz. The lower panel of the same figure shows the response to a bandpass noise with a center frequency of 3900 Hz and a bandwidth of 600 Hz. Each stimulus has an ITD of −1.5 ms. Lateralization of the AM tones and bandpass noise is dominated by the mode of the *envelope* of the cross-correlation functions, which in each of these two examples occurs at an internal delay of approximately −1.5 ms.

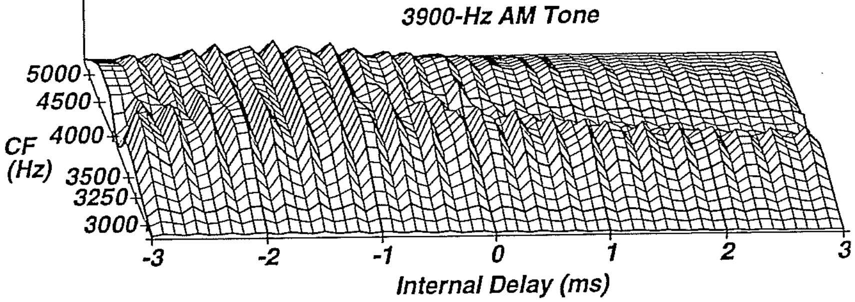
These observations are in accord with the conclusions of Colburn and Esquissaud (1976), who first suggested that cross-correlation-based models could be used to predict high-frequency binaural processing based on only



3900-Hz Tone



3900-Hz AM Tone



3900-Hz Bandpass Noise

FIGURE 10 The response of an ensemble of coincidence-counting units to several types of high-frequency stimuli with an ITD of −1.5 ms. Upper panel: response to pure tones with a frequency of 3900 Hz. Central panel: response to AM tones with a modulation frequency of 300 Hz. Lower panel: response to bandpass noise with a center frequency of 3900 Hz and a bandwidth of 600 Hz.

the implicit envelope-extraction properties of the peripheral auditory system. For example, the response at the carrier frequency (3900 Hz) in Figure 10 is greatly attenuated by the implicit envelope-detecting effects of the cascaded combination of bandpass filter, nonlinear rectifier, and lowpass filter in the model that characterizes the processing of the peripheral auditory system. Although models such as the extended position-variable model should in principle be able to describe most high-frequency lateralization data based on envelope delays, few detailed attempts have been made to compare the predictions of these models to the experimental data. Stern et al. (1988a) did observe that the extended position-variable model in its present form fails to predict the unexpected observation by Trahiotis and Bernstein (1986) that bandpass noise with a given ITD tends to be lateralized further from the center of the head than AM tones of similar carrier frequency, modulation frequency, and effective bandwidth.

5. Other Lateralization Phenomena

The discussions of lateralization mechanisms in the preceding sections have all concerned simple stimuli that have been used in "classical" psychoaccoustical experiments. Several recent studies have shown that direct application of the cross-correlation-based binaural processing models described in this chapter can describe more complex phenomena as well. For example, Hafter and Shelton (Hafter, Shelton, & Green, 1980; Hafter & Shelton, 1991) described the lateralization of diotic bandpass noise gated by brief rectangular pulses that themselves had an ITD. Surprisingly, these stimuli are frequently lateralized toward the ear receiving the gating signal that is *lagging* in time. It was later shown (Stern et al., 1991) that these counterintuitive lateralization effects are predicted quite elegantly by the models described in Section III of this chapter. Similarly, Bilsen and Raatgever (1973) have described a "dominant region" effect in which frequency components in the neighborhood of about 700 Hz appeared to be more salient than higher-frequency and lower-frequency components in the lateralization of broadband noise. Again, it was later shown (Stern et al., 1988a; Stern & Shear, 1995) that this phenomenon can be naturally accounted for by the shape of the density function for internal delays of the fiber pairs, $p(\tau,f)$. Finally, Lindemann (1986b) has shown that his extensions to the Jeffress–Colburn model can describe (at least qualitatively) the minimum temporal separation between the onsets of pairs of bandpass-filtered binaural clicks that is needed for echo perception, as well as the laterality of the fused image of binaural click pairs with short temporal separations, and the laterality of the echoes produced by binaural click pairs with longer temporal separations.

Until now, the application of binaural models to more complex stimuli has been limited by the difficulty in developing analytical expressions to characterize the response to these stimuli at the levels of the auditory nerve and the ensemble of coincidence-counting units. It is expected that, as the use of realistic computational models of the peripheral auditory response to sound becomes more widespread, and as the cost of computational resources decreases, the breadth of phenomena that are successfully predicted by the cross-correlation-based models will continue to increase.

B. Interaural Discrimination Phenomena Related to Subjective Lateral Position

The perceptual cue used by subjects in many interaural discrimination experiments is a change in subjective lateral position of the stimuli. We summarize in this section some of the ways in which several of the models of binaural lateralization have used lateral position to predict results of interaural discrimination experiments.

Models that describe the lateral position of binaural stimuli can be directly applied to discrimination experiments by computing or estimating the variance as well as the mean values of the predicted lateral positions of the stimuli, using optimal decision theory to estimate the best possible discrimination performance (cf. Van Trees, 1968). Most of the early black-box binaural models (e.g., Jeffress et al., 1956; Hafter, 1971) implicitly assumed that the variance of the position estimate is constant, which is likely to be valid if the changes in the ITD and IID of the stimuli are of sufficiently small magnitude. For example, models that assume that position is a linear combination of ITD and IID and that position variance is constant can predict the results of many tone-on-tone and noise-on-noise experiments typified by the data of Jeffress and McFadden (1971) and Yost, Nielsen, Tanis, and Bergert (1974).

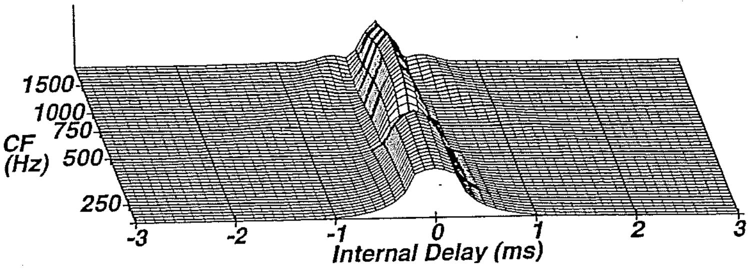
Colburn (1973) and Stern and Colburn (1985) have provided predictions for interaural discrimination experiments using expressions for the variance of predicted position that were derived from the Poisson variability inherent in the auditory-nerve model used to describe the response to the stimuli. Colburn (1973) based his predictions on the amount of information in the ensemble of coincidence-counting units (without making any assumptions about the perceptual cue used by the subjects) and predicted the dependence of just-noticeable differences (jnds) in ITD and IID on baseline ITD, IID, and overall level (Hershkowitz & Durlach, 1969). Stern and Colburn (1985) derived an analytical expression for the variance of the predicted position variable $\hat{P}$. Calculating predictions on the basis of the mean and variance of $\hat{P}$ for the stimuli of each experiment, Stern and Colburn (1985) found that the original position-variable model correctly predicted many of the trends of interaural time and amplitude jnds (e.g., Domnitz & Colburn, 1977), and studies of masking using correlated targets and maskers (e.g., Yost et al., 1974; Jeffress & McFadden, 1971) at 500 Hz. To date, no attempts have been made to generate predictions for similar data at other frequencies. As noted in Subsection III.C.2, the model is unable to account for the results of certain other discrimination experiments concerning time–intensity tradability (e.g., Hafter & Carrier, 1972; Gilliom & Sorkin, 1972) because the data imply the use of multiple perceptual images and the theoretical predictions are based only on the dominant time–intensity traded image of the stimuli.

Stern et al. (1988) have developed a small number of predictions for the extended position-variable model involving the discrimination of high-frequency stimuli with low-frequency envelopes. They observed that this model correctly describes the general dependence of discrimination performance on modulation frequency for high-frequency AM tones (Henning, 1974), and it describes the dependence of the sensitivity to ITD on the interaural frequency difference of the carrier frequency (Nuetzel & Hafter, 1981). Because it is more difficult to calculate the variance of the position estimate for the extended position-variable model than for the original position-variable model, these predictions were obtained by assuming constant position variance.

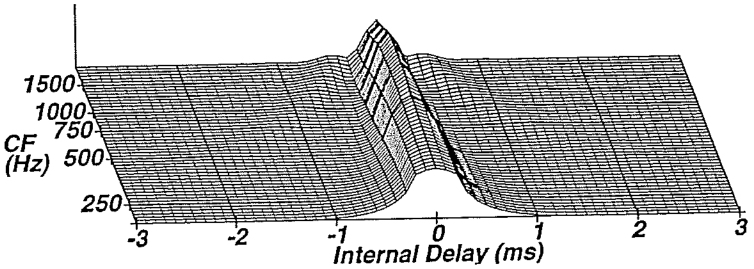
C. Binaural Masking-Level Differences

The binaural masking-level difference (or MLD) is an extremely well known and robust binaural phenomenon, discussed in some detail in Chapter 9. Figure 11 illustrates how the ensemble of coincidence-counting units accounts for this phenomenon. The figure shows the cross-correlation patterns that result when a 500-Hz tonal target and a broadband masking noise are presented in the N_0S_π (masker interaurally in phase, target interaurally out of phase) and N_0S_0 (masker and target both interaurally in phase) configuration. The plots in Figure 11 include the effects of the relative number of fiber pairs, as specified by the function $p(\tau,f)$. Note that when the N_0 masker is presented alone (Figure 11, lower panel), the ridge of maxima at 0 internal delay has approximately constant amplitude over a broad range of frequencies. The addition of an in-phase (S_0) target to the masker at a target-to-masker intensity ratio of -20 dB has virtually no effect on the pattern of coincidence-counting activity, because the interaural time differences of the combined target and masker are unchanged (Figure 11, central panel). On the other hand, the addition of the 500-Hz out-of-phase (S_π) target to the in-phase masker cancels masker components at that frequency, causing a "dimple" to appear in the central ridge for CFs near the target frequency (Figure 11, upper panel). The target in the N_0S_π configuration is easily detected at -20 dB SNR because the pattern of responses in the upper panel of Figure 11 is easily discriminated from that in the lower panel. The N_0S_0 stimulus is not detected because the response of the binaural system is unaffected by whether the target is present or absent (central and lower panels of Figure 11).

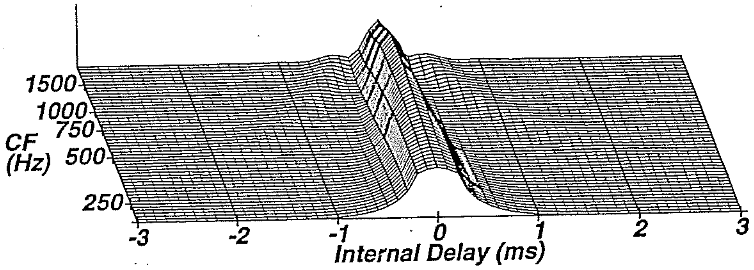
Colburn (1977) was able to describe virtually all of the "classical" data obtained in experiments measuring binaural masking-level differences on the basis of the predicted outputs of the coincidence counters. His predictions were developed using the simplifying assumption that experimental performance is limited by the variability of the auditory-nerve response to the signals, as opposed to the intrinsic variability of the masker components. This assumption has since been shown to be invalid for some stimuli by Siegel and Colburn (1983). More recently, Gilkey and his colleagues (e.g., Gilkey, Robinson, & Hanna, 1985; Hanna & Robinson, 1985; Gilkey & Robinson, 1986) have presented a number of results using "frozen-noise"



SNR −20 dB, N_0S_π



SNR −20 dB, N_0S_0



Masker Alone

FIGURE 11 Patterns of coincidence-counting activity showing the response to stimuli used in N_0S_π and N_0S_0 binaural masking-level difference experiments. The target is presented at 500 Hz, either interaurally in phase or out of phase, as indicated, and the masker is broadband diotic noise. These plots include the effects of the relative number of fiber pairs, as specified by the function $p(\tau,f)$.

maskers in which the actual variability of the masker component of the stimulus can be experimentally controlled. To date no binaural model has been able to account for differences of detectability associated with the individual masker waveforms used in these studies.

The outputs of the coincidence-counting units are used to obtain predictions for all experiments. Nevertheless, we believe that binaural detection phenomena are mediated by a reading of the information from the display of coincidence-counting units that is different from that used for subjective lateral position and interaural discrimination. Specifically, the subjective lateral position of binaural stimuli and the ability to perform certain interaural discrimination tasks based on changes in lateral position both appear to depend on the *locations* of the ridges of the cross-correlation function along the τ axis. In contrast, successful predictions for binaural detection tasks can be obtained by quantifying the *decrease in amplitude* of these ridges at the target frequency produced by the addition of the target to the masker.

D. Dichotic Pitch Phenomena

Many dichotic broadband stimuli can produce a clear sensation of pitch when presented simultaneously to the two ears even though no pitch is perceived when the respective signals to the two ears are presented monaurally (e.g., Cramer & Huggins, 1958; Bilsen & Goldstein, 1974; Bilsen, 1976). This phenomenon, referred to as *dichotic pitch,* can be created by IPDs that change as a function of frequency, as discussed in Chapter 8.

Bilsen and his colleagues have argued convincingly that such pitch phenomena can also be explained in terms of the outputs of the coincidence counters (e.g., Bilsen, 1977; Raatgever & Bilsen, 1986; Frijns, Raatgever, & Bilsen, 1986). To illustrate, the upper panel of Figure 12 shows the pattern of coincidence-counter outputs produced by a "Huggins-pitch" stimulus, which has a rapid transition in IPD from $-\pi$ to π radii in a narrow range of

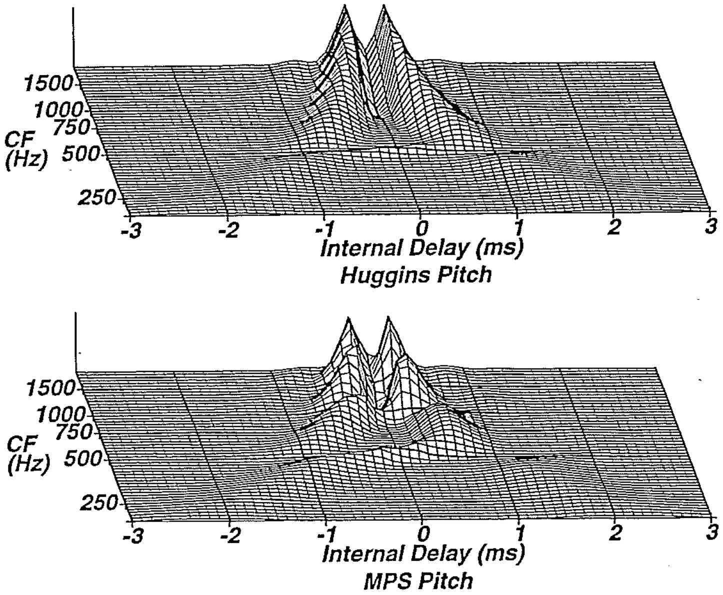


FIGURE 12 Upper panel: response of an ensemble of coincidence-counting units to a "Huggins-pitch" stimulus with an interaural phase transition from $-\pi$ to π rad, occurring in a narrow range of frequencies about 500 Hz. Lower panel: response of an ensemble of coincidence-counting units to a multiple-phase-shift stimulus producing dichotic pitch perceptions at 500 Hz. These plots include the effects of the relative number of fiber pairs, as specified by the function $p(\tau,f)$.

frequencies about 500 Hz. The lower panel of the same figure illustrates responses for a typical "multiple-phase shift" (MPS) stimulus, which contain an IPD that undergoes rapid transitions at integer multiples of 500 Hz. The plots in Figure 12 also include the effects of the relative number of fiber pairs, as specified by the function $p(\tau,f)$. The MPS stimuli produce a particularly strong pitch (Bilsen, 1976). The pitch sensation is presumably caused by the peaks of activity observed along the f-axis (at 0 internal delay) that appear at 500 Hz for the Huggins-pitch stimulus and at integer multiples of 500 Hz for the MPS-pitch stimulus. These figures closely resemble similar plots produced by the "central spectrum model" of Bilsen, Raatgever, and their colleagues.

In previous sections of this chapter we described two ways in which the auditory system appeared to use the activity from the ensemble of outputs of coincidence-counting units: The locations of peaks along the internal-delay axis appear to provide information needed to estimate auditory lateralization, and the decreases in activity can signal the presence of the target in a binaural MLD experiment. Bilsen, Raatgever, and their colleagues suggest that the information from the coincidence counters can also be used to estimate the pitch of dichotic-noise stimuli, by considering the patterns of activity of coincidence counters with 0 internal delay.

Raatgever and Bilsen (1986) have also measured the lateral position of the dichotic-pitch image in stimuli such as Huggins pitch and MPS pitch and found that the lateral position of this image is relatively unaffected by IID. This result reinforces the hypothesis that the interaction between ITD and IID in the lateralization process takes place centrally, combining information from the coincidence-counting units in a fashion that is different from that used in producing dichotic pitch. Raatgever and Bilsen (1986) and Frijns et al. (1986) have also estimated the pattern of activity of the ensemble of coincidence counters in response to dichotic-pitch stimuli by measuring the binaural MLDs that they produce for tonal targets. A good correspondence was observed between their experimental data and theoretical predictions.

The development of a theoretical framework for interpreting dichotic-pitch stimuli in the context of cross-correlation-type models represents a major step forward in developing a unified view of the way we process complex signals to the two ears.

E. Temporal Effects

In the previous sections we have considered primarily stimuli containing interaurally static binaural cues. In recent years, there has been increasing interest in the perception of stimuli with dynamically changing ITDs, IIDs, and interaural correlation. Many of the primary data are summarized in

Chapter 9. In this section we will summarize the theoretical formulations that have been used to explain the data within the context of the binaural display of coincidence-counting units.

1. Binaural "Sluggishness"

As noted in Section II.A, the binaural system is somewhat "sluggish" in its response to stimuli with time-varying interaural differences. For example, subjects are unable to track the instantaneous values of ITD or interaural correlation if they are varied with a frequency of more than a few Hz.

For the general model described in Section II.D, the instantaneous outputs of the coincidence-counting units undergo temporal integration using the temporal weighting function $w(t)$. This type of temporal integration inevitably causes temporal "sluggishness" because the duration of the integration window limits the resolution with which one can observe time-varying interaural differences of complex stimuli.

On the basis of some initial studies (Bachorski, 1983; Stern & Bachorski, 1983), we believe that many, if not all, of the sluggishness phenomena can be explained in terms of simple temporal integration of the coincidence-counter outputs, provided that the time constants for processing ITDs and for IIDs are allowed to differ (Grantham & Wightman, 1978; Grantham, 1984). Gabriel (1983) developed a black-box model that incorporated separate time constants for processing ITD and IID. The type of temporal averaging that is likely to mediate binaural sluggishness also provides at least a qualitative explanation for the disappearance of binaural beats at high beat frequencies (Licklider et al., 1950).

2. The Precedence Effect

The precedence effect refers to the dominant role that early-arriving, direct components of a sound (as opposed to later, reflected wavefronts) play in determining the location of that sound (e.g., Wallach, Newman, & Rosenzweig, 1949; Haas, 1951; Gardner, 1968). This phenomenon has motivated a large number of contemporary experimental studies, many of which are summarized in Zurek (1987) and in Chapter 9 of this volume. Although this is also a "temporal" effect, it is likely to be mediated by a different aspect of the binaural system from the mechanisms producing binaural sluggishness.

Lindemann's extension of the cross-correlation model is able to describe at least qualitatively several observations related to the precedence effect and other dynamic binaural phenomena including summing localization, the law of the first wavefront, and echo suppression (Lindemann, 1986b). McFadden (1973) also provided an insightful discussion of localization of sound and suppression of echoes, including a consideration of how each may be accomplished via inhibitory interactions of the outputs of the coincidence counters.

A number of contemporary researchers have conducted headphone studies to measure sensitivity to binaural cues using transient stimuli that are believed to give rise to the precedence effect (e.g., Yost and Soderquist, 1984; Zurek, 1987; Clifton, 1987). Nevertheless, there have been relatively few efforts to date to develop quantitative correlation-based models that can predict the results of such experiments. Progress has been hampered by the difficulty in extending analytical approaches such as those used by Colburn (1973, 1977) and Stern and Colburn (1978) to include time-varying interaural differences and by the lack of an explicit source of internal noise in computational models such as those of Blauert and colleagues (e.g., Lindemann, 1986a; Gaik, 1993). These limitations may be overcome in the future by combining a computational model that simulates the stochastic response of the auditory nerve to arbitrary stimuli (e.g., Payton, 1988; Meddis et al., 1990; Carney, 1993) with central processing as formulated either along the lines suggested by Lindemann model or according to one of the computational models of binaural processing in the brainstem (as reviewed by Colburn, 1994).

Setting these issues aside, it should be mentioned that Litovsky and Yin (1993) have noted that neural responses measured in the inferior colliculus to the second of a pair of binaural clicks may be moderately or severely suppressed depending on the temporal relations of the stimuli. Identifying these neural responses with the precedence effect would be premature at this time, and it is always possible that some of the "precedence-effect" phenomena are mediated by more central processes. Nevertheless, the temporal relations of the physiological stimuli used by Yin and his colleagues parallel those of the psychophysical studies, and it seems probable that some type of neural inhibition of the type postulated by Lindemann is actually occurring within the nervous system.

V. SUMMARY AND CONCLUSIONS

In this chapter we have reviewed the evolution and development of current theories of binaural interaction. We have noted that most current models now include as an intermediate display the interaural cross-correlation as a function of CF after processing by the peripheral auditory system. We believe that this display of information can be used as a powerful tool for understanding the perception of several quite different types of binaural phenomena, and we have attempted to facilitate this understanding by providing examples of these responses to a number of classical stimuli used in binaural experiments.

It is likely that different types of psychophysical tasks are mediated by different ways of interpreting the information contained in the binaural display. Lateralization phenomena can be predicted by considering the locations of the modes of the cross-correlation function along the internal-delay axis. Binaural detection thresholds can be predicted by consideration of the depth of notches of the ridges of the cross-correlation function at or near the target frequency. Many dichotic-pitch phenomena can be described at least qualitatively by examining the locations of modes of the cross-correlation function along the frequency axis at 0 internal delay. Consideration of the implicit temporal integration of the cross-correlation function of the stimuli goes a long way toward describing many binaural "sluggishness" phenomena, and the cross-correlation display provides a jumping-off point for more detailed efforts to characterize mechanisms underlying the perception of binaural stimuli with dynamically changing interaural time and intensity differences.

Over the next several years we expect to see cross-correlation-based models being applied to a wider variety of stimuli, with widespread use of computational simulations of the auditory-nerve response to stimuli and of neural interactions at more central sites. Similarly, we expect to see increasing use of signal processing schemes motivated by our knowledge of how the human binaural system functions in many diverse areas, including sound reproduction, environment simulation, and speech enhancement and recognition.

Acknowledgments

Preparation of this manuscript has been supported by NSF Grant IBN 90-22080 to Richard Stern and by NIH Grant DC-00234 and AFOSR Grant 89-0030 to Constantine Trahiotis. We also thank Les Bernstein and Steve Colburn for their comments on previous drafts of this manuscript and other valuable contributions to this work. Preparation of the figures has been facilitated by the efforts of Carl Block, Wonseok Lee, Steve Palm, Glenn Shear, Sammy Tao, Xaohong Xu, Andreas Yankopolus, and Torsten Zeppenfeld.

References

Bachorski, S. J. (1983). Dynamic cues in binaural perception. M.S. thesis, Carnegie Mellon University.

Bernstein, L. R., & Trahiotis, C. (1985). Lateralization of low-frequency complex waveforms: The use of envelope-based temporal disparities. *Journal of the Acoustical Society of America, 77,* 1868–1880.

Bilsen, F. A. (1976). Pronounced binaural pitch phenomenon. *Journal of the Acoustical Society of America, 59,* 467–468.

Bilsen, F. A. (1977). Pitch of noise signals: Evidence for a "central spectrum." *Journal of the Acoustical Society of America, 61,* 150–161.

Bilsen, F. A., & Goldstein, J. L. (1974). Pitch of dichotically delayed noise and its possible spectral basis. *Journal of the Acoustical Society of America, 55,* 292–296.

Bilsen, F. A., & Raatgever, J. (1973). Spectral dominance in binaural lateralization. *Acustica, 28,* 131–132.

Blauert, J. (1980). Modelling of interaural time and intensity difference discrimination. In G. van den Brink & F. A. Bilsen (Eds.), *Psychophysical, physiological, and behavioural studies in hearing* (pp. 421–424). Delft: Delft University Press.

Blauert, J. & Cobben, W. (1978). Some consideration of binaural cross-correlation analysis. *Acustica, 39,* 96–103.

Bodden, M. (1993). Modeling human sound source localization and the cocktail-party-effect. *Acta Acustica, 1,* 43–55.

Boudreau, J. C., & Tsuchitani, C. (1968). Binaural interaction in the cat superior olive S segment. *Journal of Neurophysiology, 31,* 442–454.

Brugge, J. F., Anderson, D. J., & Aitkin, L. M. (1970). Response of neurons in the dorsal nucleus of the lateral lemniscus of the cat to binaural stimuli. *Journal of Neurophysiology, 33,* 441–458.

Buell, T. N., Trahiotis, C., & Bernstein, L. R. (1994). Lateralization of bands of noise as a function of combinations of interaural intensitive differences, interaural temporal differences, and bandwidth. *Journal of the Acoustical Society of America, 95,* 1482–1489.

Carney, L. H. (1993). A model for the responses of low-frequency auditory nerve fibers in cat. *Journal of the Acoustical Society of America, 93,* 401–417.

Carr, C. E., & Konishi, M. (1990). A circuit for detection of interaural time differences in the brainstem of the barn owl. *Journal of Neuroscience, 10,* 3227–3246.

Clifton, R. K. (1987). Breakdown of echo suppression in the precedence effect. *Journal of the Acoustical Society of America, 82,* 1834–1835 (L).

Colburn, H. S. (1969). Some physiological limitations on binaural performance. Ph.D. dissertation, MIT.

Colburn, H. S. (1973). Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination. *Journal of the Acoustical Society of America, 54,* 1458–1470.

Colburn, H. S. (1977). Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. *Journal of the Acoustical Society of America, 61,* 525–533.

Colburn, H. S. (1995). Computational models of binaural processing. In H. Hawkins & T. McMullin (Eds.), *Auditory computation.* New York: Springer-Verlag.

Colburn, H. S., & Durlach, N. I. (1978). Models of binaural interaction. In E. C. Carterette & M. P. Friedman, (Eds.) *Handbook ofperception* (pp. 467–518). New York: Academic Press.

Colburn, H. S., & Esquissaud, P. (1976). An auditory-nerve model for interaural time discrimination of high-frequency complex stimuli. *Journal of the Acoustical Society of America, 59,* S23(A).

Cramer, E. M., & Huggins, W. H. (1958). Creation of pitch through binaural interaction. *Journal of the Acoustical Society of America, 30,* 413–417.

Crow, G., Rupert, A. L., Moushegian, G. (1978). Phase-locking in monaural and binaural medulary neurons: Implications for binaural phenomena. *Journal of the Acoustical Society of America, 64,* 493–501.

David, E. E., Guttman, N., & van Bergeijk, W. A. (1958). On the mechanism of binaural fusion. *Journal of the Acoustical Society of America, 30,* 801–802.

Deatherage, B. H., & Hirsh, I. J. (1959). Auditory localization of clicks. *Journal of the Acoustical Society of America, 31,* 486–492.

DeSimio, M. P., & Anderson, T. R. (1993). Phoneme recognition with binaural cochlear models and the stereoausis representation. Proceedings of the IEEE international conference on acoustics, speech, and signal processing, *1,* 521–524.

Domnitz, R. H., and Colburn, H. S. (1977). Lateral position and interaural discrimination. *Journal of the Acoustical Society of America, 61,* 1586–1598.

Duifhuis, H. (1973). Consequences of peripheral frequency selectivity for nonsimultaneous masking. *Journal of the Acoustical Society of America, 54,* 1471–1488.

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *Journal of the Acoustical Society of America, 35,* 1206–1218.

Durlach, N. I. (1972). Binaural signal detection: Equalization and cancellation theory. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (pp. 369–462). New York: Academic Press.

Durlach, N. I. & Colburn, H. S. (1978). Binaural phenomena. In E. C. Carterette & M. P. Friedman, (Eds.) *Handbook of perception* (pp. 365–466). New York: Academic Press.

Frijns, H. M., Raatgever, J., & Bilsen, F. A. (1986). A central spectrum theory of binaural processing: The binaural edge pitch revisited. *Journal of the Acoustical Society of America, 80,* 442–451.

Gabriel, K. J. (1983). Binaural interaction in hearing impaired listeners. Ph.D. dissertation, MIT.

Gaik, W. (1993). Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling. *Journal of the Acoustical Society of America, 94,* 98–110.

Gardner, M. B. (1968). Historical background of the Haas and/or precedence effect. *Journal of the Acoustical Society of America, 43,* 1243–1248.

Gilkey, R. H., & Robinson, D. E. (1986). Models of auditory masking: A molecular psychophysical approach. *Journal of the Acoustical Society of America, 79,* 1499–1510.

Gilkey, R. H., Robinson, D. E., & Hanna, T. E. (1985). Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise. *Journal of the Acoustical Society of America, 78,* 1207–1219.

Gilliom, J. D., & Sorkin, R. D. (1972). Discrimination of interaural time and intensity. *Journal of the Acoustical Society of America, 52,* 1635–1644.

Goldberg, J. M., & Brown, P. B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tone stimuli: Some physiological mechanisms of sound localization. *Journal of Neurophysiology, 32,* 613–636.

Grantham, D. W. (1984). Discrimination of dynamic interaural intensity differences. *Journal of the Acoustical Society of America, 76,* 71–76.

Grantham, D. W., & Wightman, F. L. (1978). Detectability of varying interaural temporal differences. *Journal of the Acoustical Society of America, 63,* 511–523.

Haas, H. (1951). Uber den Einfluss eines Einfachechos auf die Horsamkeit von Sprache. *Acustica, 1,* 49–58. English translation in H. Haas (1972), The influence of a single echo on the audibility of speech, *Journal of the Audio Engineering Society, 20,* 146–159.

Hafter, E. R. (1971). Quantitative evaluation of a lateralization model of masking-level differences. *Journal of the Acoustical Society of America, 50,* 1116–1122.

Hafter, E. R., & Carrier, S. C. (1972). Binaural interaction in low-frequency stimuli: The inability to trade time and intensity completely. *Journal of the Acoustical Society of America, 51,* 1852–1862.

Hafter, E. R., & Jeffress, L. A. (1968). Two-image lateralization of tones and clicks. *Journal of the Acoustical Society of America, 44,* 563–569.

Hafter, E. R., Shelton, B. R., & Green, D. M. (1980). A reversal in lateralization, with images appearing on the side of the delay. *Journal of the Acoustical Society of America, 68,* S16 (A).

Hafter, E. R. & Shelton, B. R. (1991). Counterintuitive reversals in lateralization using rectangularly modulated noise. *Journal of the Acoustical Society of America, 90,* 1901–1907.

Hafter, E. R., & Trahiotis, C. (1994). Functions of the binaural system. In M. J. Crocker (Ed.), *Handbook of acoustics.* New York: Wiley.

Hanna, T. E., & Robinson, D. E. (1985). Phase effects for a sine wave masked by reproducible noise. *Journal of the Acoustical Society of America, 77,* 1129–1140.

Henning, G. B. (1974). Detectability of interaural delay in high-frequency complex waveforms. *Journal of the Acoustical Society of America, 55,* 84–90.

Hershkowitz, R. M., & Durlach, N. I. (1969). Interaural time and amplitude jnds for a 500-Hz tone. *Journal of the Acoustical Society of America, 46,* 1464–1467.

Hirsch, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *Journal of the Acoustical Society of America, 29,* 536–544.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology, 41,* 35–39.

Jeffress, L. A., Blodgett, H. C., Sandel, T. T., & Wood, C. L., III. (1956). Masking of tonal signals. *Journal of the Acoustical Society of America, 28,* 416–426.

Jeffress, L. A., & McFadden, D. (1971). Differences of interaural phase and level in detection and lateralization. *Journal of the Acoustical Society of America, 49,* 1169–1179.

Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *Journal of the Acoustical Society of America, 68,* 1115–1122.

Klumpp, R. G., & Eady, H. R. (1956). Some measurements of interaural time difference thresholds. *Journal of the Acoustical Society of America, 28,* 859–860.

Kuwada, S., Stanford, T. R., & Batra, R. (1987). Interaural phase-sensitive units in the inferior colliculus of the unanesthetized rabbit: effects of changing frequency. *Journal of Neurophysiology, 57,* 1338–1360.

Kuwada, S., & Yin, T. C. T. (1987). Physiological studies of directional hearing. In W. A. Yost, & G. Gourevitch (Eds.), *Directional hearing* (pp. 146–176). New York: Springer-Verlag.

Lazzaro, J. (1991). A silicon model of an auditory neural representation of spectral shape. *IEEE Journal of Solid-State Circuits, 26,* 772–777.

Licklider, J. C. R. (1948). The influence of interaural phase relations upon the masking of speech by white noise. *Journal of the Acoustical Society of America, 20,* 150–159.

Licklider, J. C. R. (1959). Three auditory theories. In S. Koch (Ed.), *Psychology: A study of a science* (pp. 41–144). New York: McGraw-Hill.

Licklider, J. C. R., Webster, J. C., & Hedlun, J. M. (1950). On the frequency limits of binaural beats. *Journal of the Acoustical Society of America, 22,* 468–473.

Lindemann, W. (1986a). Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *Journal of the Acoustical Society of America, 80,* 1608–1622.

Lindemann, W. (1986b). Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wavefront. *Journal of the Acoustical Society of America, 80,* 1623–1630.

Litovsky, R. Y., & Yin, T. C. T. (1993). Single-unit responses to stimuli that mimic the precedence effect in the inferior colliculus of the cat. Abstracts of the 16th midwinter research meeting of the Association for Research in Otolaryngology, St. Petersburg, FL, p. 128(A).

Lyon, R. F. (1983). A computational model of binaural localization & separation. Proceedings of the IEEE international conference on acoustics, speech, and signal processing, pp. 1148–1151.

McFadden, D. (1973). Precedence effects and auditory cells with long characteristic delays. *Journal of the Acoustical Society of America, 54,* 538 (L).

McFadden, D., & Pasanen, E. G. (1976). Lateralization at high frequencies based on interaural time differences. *Journal of the Acoustical Society of America, 59,* 634–639.

Mead, C. A., Arreguit, X., & Lazzaro, J. (1991). Analog VLSI model of binaural hearing. *IEEE Transactions on Neural Networks,* NN-*2,* 230–236.

Meddis, R., Hewitt, M. J., & Shackleton, T. M. (1990). Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse. *Journal of the Acoustical Society of America, 87,* 2866–2882.

Moushegian, G., & Jeffress, L. A. (1959). Role of interaural time and intensity differences in the lateralization of low-frequency tones. *Journal of the Acoustical Society of America, 31,* 1441–1445.

Nuetzel, J. M. & Hafter, E. R. (1981). Discrimination of interaural delays in complex waveforms: spectral effects. *Journal of the Acoustical Society of America,* 69, 1112–1118.

Palm, S. (1989). Enhancement of reverberated speech using models of the human binaural system. M.S. thesis, Carnegie Mellon University.

Payton, K. L. (1988). Vowel processing by a model of the auditory periphery: A comparison to eighth-nerve responses. *Journal of the Acoustical Society of America, 83,* 145–162.

Pfeiffer, R. R., & Kim, D. O. (1975). Cochlear nerve fiber responses: Distribution along the cochlear partition. *Journal of the Acoustical Society of America, 58,* 867–869.

Raatgever, J. & Bilsen, F. A. (1986). A central spectrum theory of binaural processing. Evidence from dichotic pitch. *Journal of the Acoustical Society of America, 80,* 429–441.

Rose, J. E., Geisler, C. D., & Hind, J. E. (1966). Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source. *Journal of Neurophysiology, 29,* 288–314.

Sayers, B. M. (1964). Acoustic-image lateralization judgments with binaural tones. *Journal of the Acoustical Society of America, 36,* 923–926.

Sayers, B. M., & Cherry, E. C. (1957). Mechanism of binaural fusion in the hearing of speech. *Journal of the Acoustical Society of America, 29,* 973–987.

Schiano, J. L., Trahiotis, C., & Bernstein, L. R. (1986). Lateralization of low-frequency tones and narrow bands of noise. *Journal of the Acoustical Society of America, 79,* 1563–1570.

Schroeder, M. R. (1977). New viewpoints in binaural interactions. In E. F. Evans and J. P. Wilson (Eds.), *Psychophysics and physiology of hearing* (pp. 455–467). London: Academic Press.

Shackleton, T. M., Meddis, R., & Hewitt, M. J. (1992). Across frequency integration in a model of lateralization. *Journal of the Acoustical Society of America, 91,* 2276–2279 (L).

Shamma, S. A., Shen, N., & Gopalaswamy, P. (1989). Binaural processing without neural delays. *Journal of the Acoustical Society of America, 86,* 987–1006.

Siebert, W. M. (1970). Frequency discrimination in the auditory system: place or periodicity mechanisms. *Proceedings of the IEEE, 58,* 723–730.

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

Smith, P. H., Joris, P. X., & Yin, T. C. T. (1993). Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: Evidence for delay lines to the medial superior olive. *Journal of Computational Neurology, 331,* 245–260.

Stanford, T. R., Kuwada, S., & Batra, R. A. (1992). A Comparison of the interaural time sensitivity of neurons in the inferior colliculus and thalamus of the unanesthetized rabbit. *Journal of Neuroscience, 12,* 3200–3216.

Stern, R. M., Jr., & Bachorski, S. J. (1983). Dynamic cues in binaural perception. In R. Klinke & R. Hartmann (Eds.), *Hearing—Physiological bases and psychophysics* (pp. 209–215). Berlin: Springer-Verlag.

Stern, R. M., Jr., & Colburn, H. S. (1978). Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position. *Journal of the Acoustical Society of America, 64,* 127–140.

Stern, R. M., & Colburn, H. S. (1985). Lateral-position-based models of interaural discrimination. *Journal of the Acoustical Societv of America, 77,* 753–755.

Stern, R. M., & Shear, G. D. (1995). Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay. *Journal of the Acoustical Society of America* (in revision).

Stern, R. M., Shear, G. D., & Zeppenfeld, T. (1988a). High-frequency predictions of the position-variable model. *Journal of the Acoustical Society of America, 84,* S60 (A).

Stern, R. M., & Trahiotis, C. (1992). The role of consistency of interaural timing over frequency in binaural lateralization. In Y. Cazals, K. Horner, & L. Demany (Eds.), *Auditory physiology and perception* (pp. 547–554). Oxford: Pergamon Press.

Stern, R. M., Zeiberg, A. S., & Trahiotis, C. (1988b). Lateralization of complex binaural stimuli: A weighted image model. *Journal of the Acoustical Society of America, 84,* 156–165.

Stern, R. M., Zeppenfeld, T., & Shear, G. D. (1991). Lateralization of rectangularly modulated noise: Explanations for counterintuitive reversals. *Journal of the Acoustical Society of America, 90,* 1901–1907.

Sullivan, T. M., & Stern, R. M. (1993). Multi-microphone correlation-based processing for robust speech recognition. Proceedings of the IEEE international conference on acoustics, speech, and signal processing, 2, 91–94.

Tao, S. H. (1992). Additive versus multiplicative combination of differences of interaural time and intensity. M.S. thesis, Carnegie Mellon University.

Tao, S. H., & Stern, R. M. (1992). Additive versus multiplicative combination of differences of interaural time and intensity. *Journal of the Acoustical Society of America, 91,* 2414(A).

Trahiotis, C., & Bernstein, L. R. (1986). Lateralization of bands of noise and sinusoidally amplitude-modulated tones: Effects of spectral locus and bandwidth. *Journal of the Aoustical Society of America, 79,* 1950–1957.

Trahiotis, C., & Stern, R. M. (1989). Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and intensity. *Journal of the Acoustical Society of America, 86,* 1285–1293.

Van Trees, H. L. (1968). *Detection, estimation, and modulation theory, Part I.* New York: Wiley.

Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). The precedence effect in sound localization. *Journal of Psychology, 52,* 315–336.

Webster, F. A. (1951). The influence of interaural phase on masked thresholds. I. The role of interaural time-deviation. *Journal of the Acoustical Society of America,* 23, 452–462.

Wenzel, E. M. (1992). Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments, 1,* 80–107.

Whitworth, R. H., & Jeffress, L. A. (1961). Time versus intensity in the localization of tones. *Journal of the Acoustical Society of America, 33,* 925–929.

Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. I. stimulus synthesis. *Journal of the Acoustical Society of America,* 85, 858–867.

Yin, T. C. T., & Chan, J. C. L. (1990). Interaural time sensitivity in medial superior olive of cat. *Journal of Neurophysiology, 65,* 465–488.

Yin, T. C. T., & Kuwada, S. (1984). Neuronal mechanisms of binaural interation. In G. M. Edelman, W. C. Cowan, & W. E. Gall (Eds.), *Dynamic Aspects of Neocortical Function* (pp. 263–313). New York: Wiley.

Yost, W. A. (1972). Tone-on-tone masking for three listening conditions. *Journal of the Acoustical Society of America, 52,* 1234–1237.

Yost, W. A. (1981). Lateral position of sinusoids presented with intensive and temporal differences. *Journal of the Acoustical Society of America, 70,* 397–409.

Yost, W. A., Nielsen, D. W., Tanis, D. C., & Bergert, B. (1974). Tone-on-tone binaural masking with an antiphasic masker. *Perception and Psychophysics,* 15, 233–237.

Yost, W. A., & Soderquist, D. R. (1984). The precedence effect revisited. *Journal of the Acoustical Society of America, 76,* 1377–1383.

Young, S. R., and Rubel, E. W. (1983). Frequency-specific projections of individual neurons in chick brainstem auditory nuclei. *Journal of Neuroscience, 3,* 1373–1378.

Zurek, P. (1987). The precedence effect. In W. A. Yost & G. Gourevitch (Eds.), *Directional hearing* (pp. 85–105), New York: Springer-Verlag.

Zwislocki, J. & Feldman, R. S. (1956). Just noticeable differences in dichotic phase. *Journal of the Acoustical Society of America, 28,* 860–864.