Lateralization of complex binaural stimuli: A weighted-image model

Richard M. Stern and Andrew S. Zeiberg

Department of Electrical and Computer Engineering and Biomedical Engineering Program,
Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Constantine Trahiotis

Center for Neurological Sciences and Department of Surgery (Otolaryngology), University of Connecticut
Health Center, Farmington, Connecticut 06032

This article describes a new model that predicts the subjective lateral position of bandpass stimuli. It is assumed, as in other models, that stimuli are bandpass filtered and rectified, and that the rectified outputs of filters with matching center frequencies undergo interaural cross correlation. The model specifies and utilizes the shape and location of assumed patterns of neural activity that describe the cross-correlation function. Individual modes of this function receive greater weighting if they are *straighter* (describing consistent interaural delay over frequency) and/or more *central* (describing interaural delays of smaller magnitude). This weighting of straightness and centrality is used by the model to predict the perceived laterality of several types of low-frequency bandpass stimuli with interaural time delays and/or phase shifts, including bandpass noise, amplitude-modulated stimuli with time-delayed envelopes, and bandpass-filtered clicks. This model is compared to other theories that describe lateralization in terms of the relative contributions of information in the envelopes and fine structures of binaural stimuli.

INTRODUCTION

The purpose of this article is to describe some attributes of binaural processing that help to describe the subjective lateral position of certain complex stimuli. We present a crude black-box model that expresses these attributes in a quantitative fashion.

Recently, many experimenters have obtained interesting new data concerning the lateralization of high-frequency and complex binaural stimuli. For example, Henning (1974, 1980, 1983), McFadden and Pasanen (1976), and Nuetzel and Hafter (1976, 1981) have all demonstrated, using interaural discrimination experiments and forced-choice laterality judgments, that the lateral position of a high-frequency binaural stimulus is affected by the interaural time delay (ITD) of the envelope of that stimulus, provided that the signals to the two ears occupy comparable spectral regions. More recently, Bernstein and Trahiotis (1985a,b) and Trahiotis and Bernstein (1986) have directly measured the perceived laterality of such stimuli and demonstrated that ITDs of the envelope can affect the laterality of low-frequency as well as high-frequency stimuli.

There now appear to be sufficient new data to enable the characterization of the results in terms of formal models of binaural interaction in an interesting fashion. While a number of models have previously been developed to describe the lateral position of binaural stimuli (e.g., Sayers and Cherry, 1957; Stern and Colburn, 1978; Lindemann, 1986), none directly addresses how information is combined across different frequencies. While attempting to construct a general model that describes the subjective lateral position of such stimuli, we found that obvious extensions across frequency of theories such as the position-variable model (Stern and Colburn, 1978) could not describe all of the observed data. Additional constraints were necessary. Specifically, it became clear that it was necessary to emphasize those internal delays in the model's frequency-time description of interaural stimulus correlation for which the information is most consistent across frequencies. In order to express this concept in a quantitative fashion and to further explore its salience for experimental studies, we have developed a new black-box model, called the weighted-image model. This model is not as detailed as the position-variable model, nor does it directly express its predictions in terms of the physiological signal processing occurring at the auditory periphery. However, the weighted-image model does appear to be useful for developing insight into how complex binaural sounds are processed, and it is much easier to obtain predictions from this model than it is from the position-variable model.

In the next section, we review some recent data that motivated us to reconsider the integration of interaural timing information across frequencies. In Sec. II, we describe the assumptions that specify the weighted-image model. In Sec. III, we compare the predictions of this model to some recent experimental results. In Sec. IV, we discuss these comparisons of predictions and data, and we consider also similarities and differences between predictions in models based on consistency of interaural time differences across frequency and models based on salience of envelope information of complex stimuli.

I. MOTIVATION FOR THE MODEL

Recently, Trahiotis and Bernstein (1986) noted that the lateral position of the image of a low-frequency bandpass noise was dramatically affected by changes in the bandwidth of the stimulus. For example, consider the results in Fig. 1, which shows how bandwidth affects the lateral position of a band of noise presented with a center frequency of 500 Hz and an ITD of $1500 \mu s$. These data, which are part of a larger study that is currently being prepared for publication (Trahiotis and Stern, 1988), were obtained using methods described in detail in Bernstein and Trahiotis (1985a). Position is expressed in terms of the interaural intensity difference (IID) of a second narrow band of noise called the pointer. Listeners adjusted the IID of the pointer so that its position matched that of the stimulus of interest. Note that the sound is perceived toward one side of the head with narrow bandwidths, but moves to the other side of the head as the bandwidth is increased to only 400 Hz.

In retrospect, it is easy to understand why this phenomenon is observed. For very narrow bandwidths, the noise source closely resembles a pure tone of 500 Hz. Consider a 500-Hz tone presented with the signal to the left ear time delayed relative to the signal to the right ear by $1500 \mu s$. This stimulus is equivalent to one in which the signal to the left ear is *leading* the signal to the right ear by $500 \mu s$, because the period of a 500-Hz tone is $2000 \mu s$. Consequently, the image of the binaural sound is perceived toward the left side of the head. As bandwidth is increased, however, the signal becomes less and less tonelike. In the limiting case, one is presented with broadband noise with the signal to the right ear leading the signal to the left ear by $1500 \mu s$, and this signal would certainly be lateralized toward the right (or "leading") side of the head. This raises the question of how the auditory system is able to determine that the signal to the right ear is leading the signal to the left ear by $1500 \mu s$ when the signal is sufficiently broad in bandwidth, even though it is "fooled" when the bandwidth is smaller.

The perception of these stimuli can be easily understood in terms of how the auditory system is likely to process them. Most theories of binaur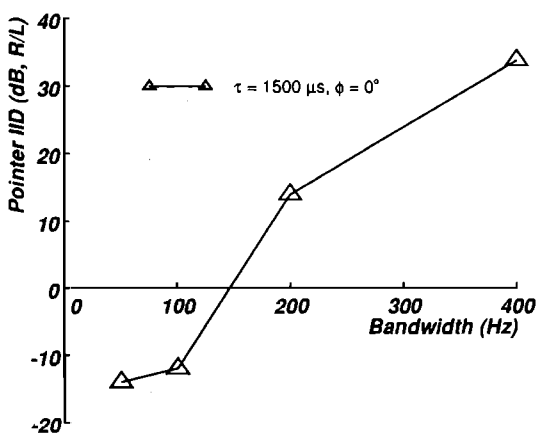al interaction assume that peripheral auditory processing can be modeled (at least in part) by passing signals to each ear through a bank of parallel bandpass filters, each with a slightly different center frequency. It is usually further assumed that the auditory system performs some sort of interaural cross correlation on the outputs of the bandpass filters. Figure 2 is a plot of the locations of the *maxima* of the cross-correlation function of a broadband noise with a 1500-μs ITD, after such bandpass filtering. The horizontal axis represents interaural delay (τ) and is the argument of the cross-correlation function. The vertical axis indicates the center frequencies of the bandpass filter (f). The frequency of 500 Hz is indicated by the horizontal broken line.

If the bandwidth of a 500-Hz stimulus were sufficiently broad, the ITD would be readily identifiable because the locus of the peaks is a straight and vertical line at only the value of the interaural delay parameter τ corresponding to that ITD (1500 μs in this case). This "straightness" could very well be the cue used to recognize that the interaural delay of the stimulus is exactly 1500 μs. Since real sounds emitted by point sources produce ITDs that are consistent over a range of frequencies, the "straighter" a particular line or trajectory of maxima, the more that trajectory is likely to represent the actual ITD of the stimulus. On the other hand, if the signal is very narrow in bandwidth, the auditory system is unable to distinguish between the various peaks because there is not enough range of frequencies to estimate adequately the "straightness" of any trajectory. Under these circumstances, we believe that the auditory system tends to lateralize the sound toward the side of the head of the basis of the trajectory that is closest to the f axis, i.e., the most "central" locus of maxima. We believe that the more central trajectories are weighted more heavily because there are more binaural coincidence-counting units with small internal interaural delays than there are neural units with larger internal delays. This is a common assumption in binaural models (e.g., Sayers and Cherry, 1957; Colburn, 1977) and was used by Jeffress in a remarkably insightful qualitative attempt to explain essentially identical auditory phenomena in 1972 (Jeffress, 1972, pp. 357–358). Jeffress focused upon centrality and did not discuss straightness nor how both centrality

FIG. 1. Pointer IID needed to match the position of bandpass targets with center frequency 500 Hz and ITD 1500 μs as a function of the bandwidth of the target (unpublished data of Trahiotis and Bernstein).
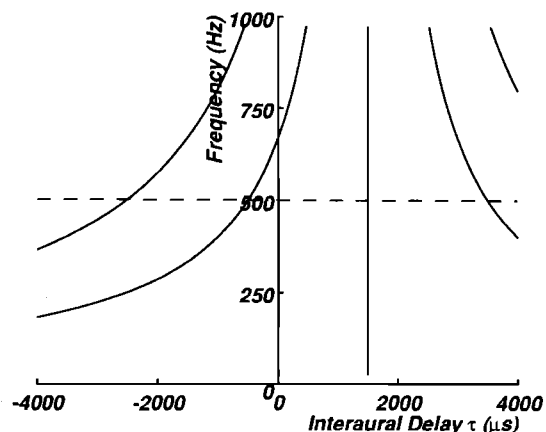
FIG. 2. Location of the peaks of the cross-correlation function following peripheral bandpass filtering for broadband noise with an ITD of 1500 μs. The vertical axis indicates the center frequency of the bandpass filters, while the horizontal axis indicates the argument of the cross-correlation function.

and straightness could be manipulated so that their respective effects and interactions could be discovered and quantified. Nevertheless, Jeffress' insights are impressive, and it seems remarkable to us that his lucid statements do not appear to have stimulated studies of lateralization as a function of the bandwidth of stimulation.

In many cases (including the 500-Hz noise with a 1500-μs ITD), either straightness or centrality considered in isolation would cause the sound to be lateralized toward different sides of the head. We believe that, in these cases, lateral position is the result of some type of resolution of these conflicting pieces of information (perhaps not unlike the "trade" of interaural timing and intensity information that characterizes the lateralization of simple stimuli with small interaural differences).

In order to examine the extent to which straightness plays a role in the lateralization of complex stimuli, it is useful to consider some new data that describe the joint dependence of the lateral position of 500-Hz bandpass noise on ITD and interaural phase difference (IPD). These stimuli are particularly useful because appropriate combinations of ITD and IPD enable the experimenter to manipulate the interaural correlation of a binaural bandpass noise such that straightness and centrality can be independently specified at a particular frequency of interest. Specifically, consider a bandpass noise presented with an ITD of T_s and an IPD of ϕ_s. After the signals undergo peripheral bandpass filtering, the maxima of the cross-correlation function at frequency f will occur at

$$\tau = T_s + (\phi_s/2\pi f)(k/f),$$

where k is an integer. Hence, a maximum of the cross-correlation function will appear at τ equals some arbitrary T_d at frequency f if values of T_s and ϕ_s are chosen such that

$$T_s = T_d - (\phi_s/2\pi f) - (k/f). \qquad (1)$$

The parameter ϕ_s controls the straightness of the trajectory with k equal to 0, as ϕ_s equals 0 rad produces a trajectory that is perfectly straight for τ equals $T_d = T_s$ (as in Fig. 2). If, on the other hand, ϕ_s equals π rad, and if T_s is selected according to Eq. (1), the correlation function will include a trajectory of maxima that also passes through τ equals T_d at frequency f, but that trajectory will be less "straight."

Using again the pointing procedure of Bernstein and Trahiotis (1985a), we measured the subjective lateral position of bandpass noises with a center frequency of 500 Hz as a function of the stimulus bandwidth for combinations of ITD and IPD. Measurements were obtained for four different values of IPD: 0°, 90°, 180°, and 270°. The ITD in each case was equal to $[1500 \mu s - (\phi_s/2\pi\, 1000 f)]$ (with ϕ_s in rad and f in kHz), so that maxima of the cross-correlation function at 500 Hz would always occur at values of τ equal to 1500, $-500, 3500, -2500\mu s$, etc. Within a narrow region around 500 Hz, the trajectories of maxima of the cross-correlation function exhibit equal centrality but different straightness for the four combinations of ITD and IPD considered. Trajectories of the maxima of the correlation functions for these stimuli are sketched in Fig. 3 for the four values of IPD used in the experiment. Note that all trajectories intersect at 500 Hz and that the straightness of the trajec-
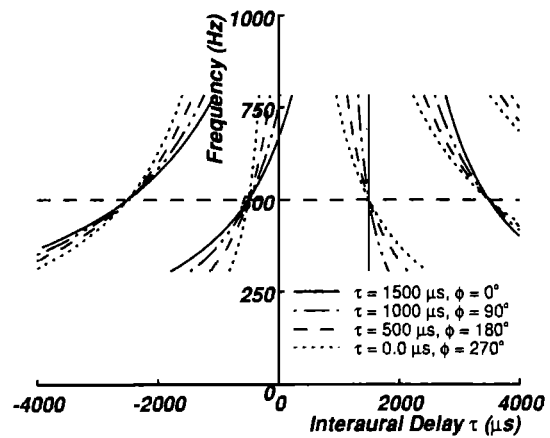

FIG. 3. Loci of the peaks of the cross-correlation function for noise with several combinations of ITD and IPD chosen to produce maxima at 1500 μs and 500 Hz in each case.

tories at 1500 μs decreases as the IPD is increased from 0° to 270°.

Results of a lateralization experimental using these stimuli for one subject are shown in Fig. 4. Note that the data for the IPD of 0° are similar to those in Fig. 1 (although the image appears to traverse the midline at a narrower bandwidth), and that, as the IPD is increased from 0° to 270°, there is an increasing tendency for the images of the noises with broader bandwidths to remain on the same side of the head as the images of the narrow-band noise. It is easy to understand these data in terms of the competition between straightness and centrality as discussed above. Specifically, all narrow-band stimuli are lateralized on the left side of the head (which corresponds to negative values of the vertical coordinate in Fig. 4) because the trajectory at $-500\mu s$ (at 500 Hz) is much more central than the trajectory at 1500 μs. With an IPD of 0°, as bandwidth is increased, we believe that the image of the noise tends to move toward the right side of the head because it becomes increasingly clear that there is a straight trajectory at 1500 μs which represent the "true" ITD. As the IPD is increased to 270° (and the ITD is reduced appropriately), the trajectory that passes through 1500 μs at 500 Hz becomes increasingly less effective in
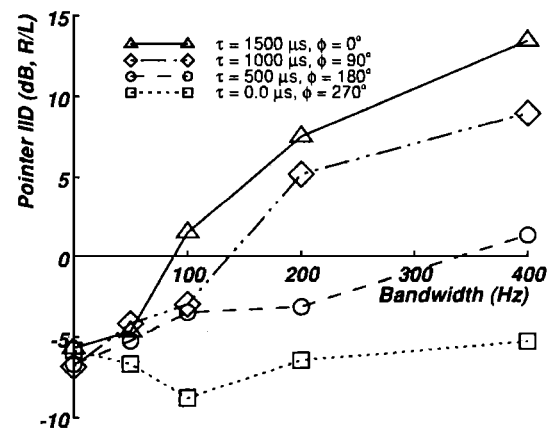

FIG. 4. Perceived position of bandpass targets with center frequency 500 Hz, presented with combinations of ITD and IPD producing different straightness but similar centrality at 500 Hz (see text).

"pulling" the image of the noise toward the right side because it becomes increasingly less straight, while simultaneously the trajectory around -500 μs becomes more straight (Fig. 3). Note that the stimulus with an IPD of 270° and an ITD of 0 μs produces a pattern of cross correlation for which both straightness and centrality would tend to cause the image to be lateralized toward the left side of the head (as is seen in the data of Fig. 4).

Before beginning our work on the model described in the next section, we first compared predictions of the position-variable model of Stern and Colburn (1978) to these data. That model assumes that the lateral position of binaural stimuli is predicted from the centroid of a function related to the interaural cross correlation of the stimuli after filtering and rectification assumed in almost all models of the auditory periphery. (The specific function that describes interaural timing information is expressed in terms of the response of hypothetical auditory-nerve fibers to the stimuli.) The position-variable model includes a function that implicitly weights components of the cross-correlation function that exhibit greater "centrality." However, there is no feature of that model that provides greater weighting for components of the function that are straighter over frequency.

Using the position-variable model, we obtained predictions for the narrow-band stimuli of Fig. 4 by computing the interaural cross-correlation function of the stimuli after peripheral processing. Predictions were obtained utilizing several implementations of the model; each version was motivated by practical as well as theoretically based factors. We explored the effects of varying the nature of three features of the model: the characteristics of the bandpass filters and nonlinear rectifier in the model of peripheral processing, and the more central function that describes the distribution of units that reflect interaural coincidences of neural activity. Predictions do appear to be affected by the particular functions that are used to characterize such elements of the position-variable model, and these issues will be discussed at length in a subsequent article (Shear and Stern, 1988).

The above efforts notwithstanding, we feel that the complexity of the position-variable model and its implementation can obscure some of the fundamental aspects of the stimuli and how they affect the overall trends of predictions of any model. Consequently, we focused upon a "weighted-image model" that formalizes the notions of "straightness" and "centrality" and uses them to more easily obtain quantitative predictions of the laterality of complex stimuli. We now describe this model in detail.

II. THE WEIGHTED-IMAGE MODEL

The weighted-image model was developed to refine our understanding of the concepts of straightness and centrality that were qualitatively described above, and to explore the extent to which they could account for several binaural phenomena. Because we wished to explore a number of phenomena for which the auditory-nerve response is not well specified, this model was formulated in a way that would provide rough predictions quickly, with less concern for the details of physiological processing. Toward this end, predictions of the weighted-image model are based only to the maxima of the

frequency-dependent cross-correlation function (such as those shown in Fig. 2) and how these maxima vary as a function of frequency. [In contrast, the models of Sayers and Cherry (1957); Stern and Colburn (1978); and Lindemann (1986) all describe lateralization in terms of operations on the entire frequency-dependent cross correlation of the stimulus.]

Predictions of the weighted-image model are derived from a weighted linear combination of the centroids along the τ axis of the maxima of the internal interaural cross-correlation function of the binaural stimuli. For simple time-delayed noise, the loci of these maxima form hyperbolic trajectories in the τ-f plane that are spaced apart by $\tau = 1/f$ for a given f. The relative weight of each trajectory is assumed to depend on its straightness and centrality. The straightness component of the weighting function is designed to weigh more heavily those trajectories that are more straight (parallel to the f axis) and, hence, are more likely to represent a true ITD of a real stimulus that is consistent over frequency. The centrality component of the weighting factors enhances the contribution of trajectories located along values of internal delay (the τ axis) for which there are presumed to be a greater number of neural units available. This emphasizes the contributions of trajectories of values of τ that are smaller in magnitude. Note that, if the stimulus is a pure tone, *all* of the trajectories of maxima of the cross-correlation function are straight and vertical, and, consequently, the weighting of the trajectories is only affected by their relative centrality. [Tonal stimuli produce trajectories with nonzero extent along the f axis because of the imperfect frequency resolution of the filters. The spacing of the maxima is determined by the frequency of the stimulus (rather than the center frequency of the filter) as long as the bandwidth of the stimulus is less than the filter bandwidth.] Therefore, the predictions of the weighted-image model for tones are similar to the predictions of the position-variable model, except that the weighted-image model computes the centroid of the *maxima* of the cross-correlation function, while the position-variable model computes the centroid of the *entire* cross-correlation function.

More specifically, the weighted-image model generates a variable designated by $\hat{P}$, which is obtained by a linear combination of individual images μ_i:

$$\hat{P} = \sum_i \omega_i \mu_i \bigg/ \sum_i \omega_i. \qquad (2)$$

In the above expression μ_i is the location along the τ axis, averaged over frequency, of the ith trajectory of maxima of the cross-correlation function

$$\mu_i = \sum_k \tau(i, f_k) p(\tau(i, f_k), f_k) q(f_k)$$

$$\times \left(\sum_k p(\tau(i, f_k), f_k) q(f_k) \right)^{-1}, \qquad (3)$$

where $\tau(i, f_k)$ is the location along the τ axis of the ith maximum of the cross-correlation function at characteristic frequency f_k. The frequency-dependent function $p(\tau(i, f_k), f_k)$ indicates the relative number of neural units at interaural delay $\tau(i, f_k)$. The function $q(f_k)$, which is sketched in Fig. 5, is a frequency-weighting function evaluated at frequency f_k and as obtained by a polynomial ap-
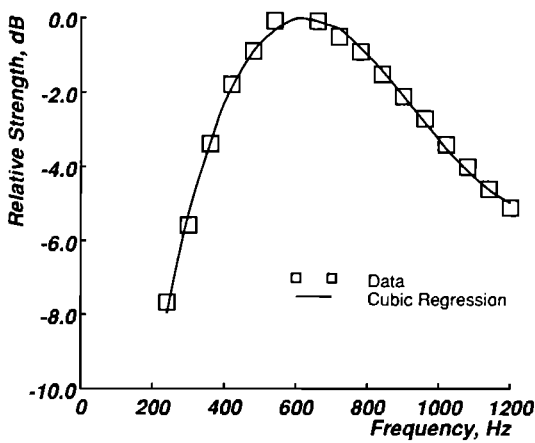
FIG. 5. An empirical frequency weighting function for the salience of components of binaural stimuli. Data were obtained by Raatgever (1980) and were fit by a cubic polynomial approximation. The frequency weighting function is assumed to be a constant for frequencies above 1200 Hz.

proximation to empirical data of Raatgever (1980). [The functions $p(\tau,f)$ and $q(f)$ are specified below.] The computation of μ_i weighs more heavily the contribution of frequencies around the dominant region of 600 Hz [through the contribution of $q(f)$], and it weighs more heavily values of τ that are smaller in magnitude [through the contribution of $p(\tau,f)$]. The summations in Eq. (3) are evaluated over logarithmically spaced values of f_k.

The weighting coefficient ω_i for the ith trajectory is specified by the equation

$$\omega_i = \left[\sigma_i^2 + K_0\left(\sum_k p(\tau(i,f_k),f_k)q(f_k)\log\frac{f_U}{f_L}\right)\right]^{-1}, \quad (4)$$

where $p(\tau,f)$ and $q(f)$ are defined as above, and the constants f_U and f_L refer to the upper and lower frequencies, respectively, over which the computation is evaluated. The parameter σ_i^2 describes the straightness or curvature of the ith trajectory and is obtained using the formula

$$\sigma_i^2 = \sum_k (\tau(i,f_k) - \mu_i)^2 p(\tau(i,f_k),f_k)q(f_k)$$

$$\times\left(\sum_k p(\tau(i,f_k),f_k)q(f_k)\right)^{-1}. \quad (5)$$

Here, σ_i^2 can be thought of as the frequency-weighted "variance" about μ_i along the axis of the points that comprise the ith trajectory of maxima of the cross-correlation function in the τ–f plane. Again, we note that, if the stimuli are pure tones, the straightness parameter σ_i^2 equals 0, and the various trajectories contribute to $\hat{P}$ [through $p(\tau(i,f_k),f)$] exactly as they do in the postion-variable model.

We used the following function to describe the greater weighting of more central maxima of the cross-correlation function imposed by $p(\tau,f)$:

$$p(\tau,f) = \begin{cases} \gamma(f), & |\gamma(f)\tau| \leqslant 0.15 \text{ ms}, \\ \gamma(f)\exp\{-[|\gamma(f)\tau| - 0.15]/0.6\}, \\ & \text{otherwise}, \end{cases}$$

where the frequency-dependent parameter $\gamma(f)$ is equal to

$$\gamma(f) = \begin{cases} 1, & f \leqslant 0.3 \\ (f/0.3)^2, & f > 0.3, \end{cases}$$

where f is in kHz.

The form of this function is based on that of the $p(\tau)$ used by Colburn (1977) and Stern and Colburn (1978), except that it becomes a narrower function of τ for frequencies above 300 Hz. (This frequency dependence was included for reasons related to predictions for higher frequency stimuli that are not discussed in the present article.) The amplitude of the function is scaled to maintain constant area in the tails of $p(\tau,f)$ along the τ axis at all frequencies.

The function $q(f)$ used in the weighted-image model was a best-fit third-order polynomial approximation to the frequency-weighting function implied by the data of Raatgever (1980, p. 48). For frequencies at or below 1200 Hz, this parametric approximation is

$$q(f) = 10^{-(b_1(1000f) + b_2(1000f)^2 + b_3(1000f)^3)/10},$$

where f is in kHz and

$$b_1 = (-9.383)(10^{-2}),$$
$$b_2 = (1.126)(10^{-4}),$$
$$b_3 = (-3.992)(10^{-8}).$$

Raatgever's results do not extend to frequencies above 1200 Hz, so, at these frequencies, we assume that $q(f)$ is a constant equal to

$$q(f) = 10^{-(b_1(1200) + b_2(1200)^2 + b_3(1200)^3)/10}.$$

The parameter K_0 in Eq. (4) determines the relative weighting of straightness and centrality in the theoretical predictions with straightness contributing to the predictions more if K_0 is smaller in magnitude. The predictions of this article were obtained using a value of 0.035 for K_0, which seems to best describe the data of Fig. 4. In general, we expect that the best value of K_0 would vary from subject to subject, since the relative salience of the straightness and centrality cues are likely to differ for different individuals. We are currently examining the extent to which subject-to-subject variability in a number of additional low-frequency lateralization results can be accounted for by varying the parameter K_0 in the weighted-image model.

It is difficult to succinctly characterize the sensitivity of the predictions to the exact form of the functions $p(\tau,f)$ and $q(f)$ because the effects of the two functions interact with each other in a complex fashion. Nevertheless, we are convinced that, even though the exact numerical predictions of the model depend on the specific form that is chosen for these two functions, the form of these predictions (and therefore the extent to which they qualitatively describe the form of the observed data) will remain the same for any reasonable pulse shapes.

III. COMPARISONS OF THE MODEL TO EXPERIMENTAL DATA

In this section, we present and discuss comparisons of the predictions of the weighted-image model to the results of several experimental studies. At present, these comparisons are limited to low frequencies because we believe that lateralization above 1500–2000 Hz is affected by the well-known inability of the peripheral auditory system to track the fine structure of these higher frequency stimuli. We believe that the trade-off between straightness and centrality proposed

above operates in the same fashion at higher frequencies, even though processing must be based only on the ITDs of the envelopes of these stimuli. However, quantitative predictions for these stimuli would require a much more detailed specification of how they are processed physiologically than the very simple description of the peaks of the cross-correlation function that is the basis for the weighted-image model.

In general, predictions were obtained by weighting the contributions of 21 trajectories of maxima of the cross-correlation function. This large number of trajectories is required by the gradual decrease in the magnitude of $p(\tau, f)$ as $|\tau|$ increases. Computations were evaluated at 80 frequencies f_k in each trajectory, which were logarithmically spaced over the range of interest.

A. Joint dependence of laterality of bandpass noise on interaural time delay and interaural phase difference

Predictions of the weighted-image model for the data of Fig. 4 are provided in Fig. 6, plotted with an arbitrary linear scaling of the vertical axis. As discussed in Sec. I, these predictions and data are easily understood in terms of the straightness and centrality of the maxima of the interaural cross-correlation function. For very narrow bandwidths, the stimuli are similar to 500-Hz tones with an ITD of $-500\,\mu s$ and they are lateralized toward the left side of the head. As the bandwidth of the stimuli is increased, a rivalry becomes established between the trajectory at $\tau = +1500\,\mu s$ (which is straighter), and the trajectory passing through $\tau = -500$ μs (which is more central, or closer to the f axis). When the IPD equals $0°$, the straighter trajectory at $1500\,\mu s$ is much more dominant. The stimuli with IPDs of $90°$, $180°$, and $270°$ produce cross-correlation functions with trajectories that pass through $\tau = 1500\,\mu s$ at 500 Hz with progressively decreasing straightness. As a result, the model predicts that the trajectory at $+1500\,\mu s$ is weighted less heavily (for these other values of IPD), decreasing the extent to which the image is lateralized toward the right side of the head for a given bandwidth. These trends, of course, are also observed in the data, although there are some quantitative differences between predictions and data (such as the frequency at which the 1500-μs curve crosses the horizontal axis).
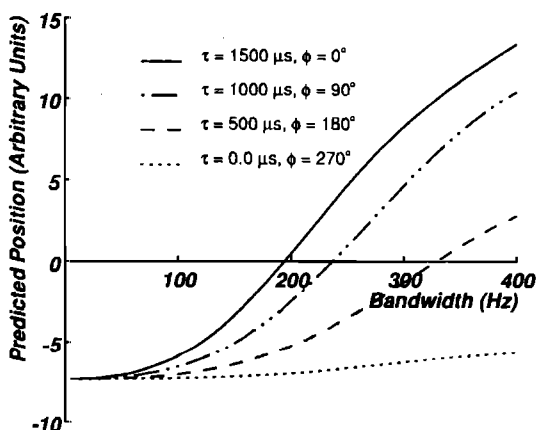

FIG. 6. Predictions of the weighted-image model for the data of Fig. 4.

B. Laterality of amplitude-modulated narrow-band stimuli

Bernstein and Trahiotis (1985a,b) recently measured the laterality of sinusoidally amplitude-modulated (SAM) low-frequency tones. The signals to the two ears in these experiments were

$$S_L(t) = [1 + \cos(\omega_m t)]\cos(\omega_c t) \tag{6a}$$

and

$$S_R(t) = [1 + \cos(\omega_m(t - \tau_m))]\cos(\omega_c(t - \tau_c)) \tag{6b}$$

or, equivalently,

$$S_L(t) = \cos(\omega_c t) + \tfrac{1}{2}\cos[(\omega_c + \omega_m)t] + \tfrac{1}{2}\cos[(\omega_c - \omega_m)t] \tag{7a}$$

and

$$\begin{aligned} S_R(t) = \cos[\omega_c(t - \tau_c)] + \tfrac{1}{2}\cos\{(\omega_c + \omega_m) \\ \times[t - ((\omega_c\tau_c + \omega_m\tau_m)/(\omega_c + \omega_m))]\} \\ + \tfrac{1}{2}\cos\{(\omega_c - \omega_m) \\ \times[t - ((\omega_c\tau_c - \omega_m\tau_m)/(\omega_c - \omega_m))]\}, \end{aligned} \tag{7b}$$

where ω_c is the carrier frequency and ω_m is the modulator frequency. Bernstein and Trahiotis considered two types of ITD: *waveform delay* (in which both the fine structure and low-frequency envelope were delayed equally; i.e., τ_c equaled τ_m) and *modulator delay* (in which only the envelope contained an ITD and the carrier remained interaurally in phase, i.e., τ_c equaled 0). Figure 7 shows an average (over three subjects) of the lateral postion of 500-Hz tones, and SAM tones with modulation frequencies of 25 and 50 Hz. For a small number of ITDs, data points are available for only two of the three subjects. Missing data points from a given subject were estimated by linearly interpolating between adjacent data points for that subject, weighted according to differences of the ITDs.

As seen in Eqs. (7a) and (7b), sinusoidally amplitude-modulated tones consist of three sinusoids (one at the carrier frequency and one each at the carrier frequency plus and minus the modulator frequency). Although these stimuli are easy to describe mathematically, it is difficult to predict accurately the physiological response that they produce. Specifically, the precise interaural cross-correlation function that would be observed for SAM tones depends on the exact bandwidth of the peripheral auditory filters, the amount of two-tone suppression produced by the stimuli, and other details of physiological processing that are not specified explicitly in the simple weighted-image model. In order to obtain a basic understanding of how the factors of straightness and centrality come into play for the data of Bernstein and Trahiotis without further complicating the model, we obtained predictions for a simple type of stimulus that is closely related to the SAM tone. Specifically, we converted the line spectrum of the SAM tones to a continuous bandpass spectrum with interaural phase shifts that were equal at the frequencies of the carrier and of the two sidebands and that were linearly interpolated between these frequencies. In other words, the IPD of the stimuli at any frequency between the
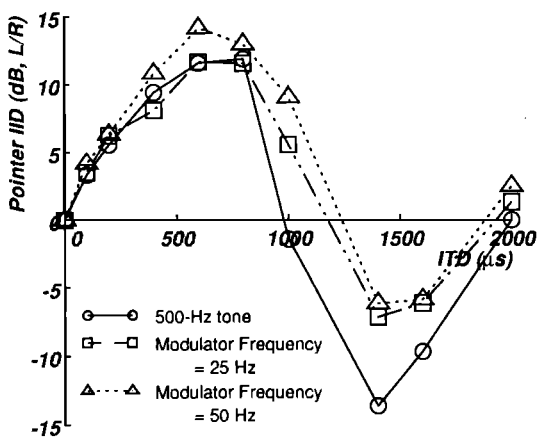
FIG. 7. Experimental measurements of position as a function of waveform ITD for 500-Hz tones with and without amplitude modulation, averaged over three subjects. Data are replotted from Bernstein and Trahiotis (1985a).

two sidebands was

$$\phi_{EQ} = \omega_c (\tau_c - \tau_m) + \omega\tau_m.$$

This means that maxima of the cross-correlation function at frequency f will appear at

$$\tau = \tau_m + (f_c/f)(\tau_c - \tau_m) + (k/f_c), \tag{8}$$

where f is in Hz, f_c equals $\omega_c/2\pi$, and k is an integer.

Figure 8 contains predictions of the weighted-image model for the data of Fig. 7. Both the observed and predicted positions are roughly sinusoidal functions of the waveform delay. More interestingly, the frequency of modulation has a much greater effect on perceived lateral position at ITDs of approximately $1300 \mu s$ than it does at ITDs of approximately $700 \mu s$, and a similar effect is seen in the predictions as well.

Once more, consideration of the loci of the maxima of the interaural cross correlation of the stimuli provides an intuitively appealing explanation for the greater dependence of position on modulation frequency for ITDs near $1300 \mu s$ compared to $700 \mu s$. Figures 9 and 10 show the locations of the peaks of the cross correlation for the idealized stimuli that were used to obtain the theoretical predictions, presented with waveform delays of 700 and $1300 \mu s$, respectively. When the frequency of modulation is small, the trajectories
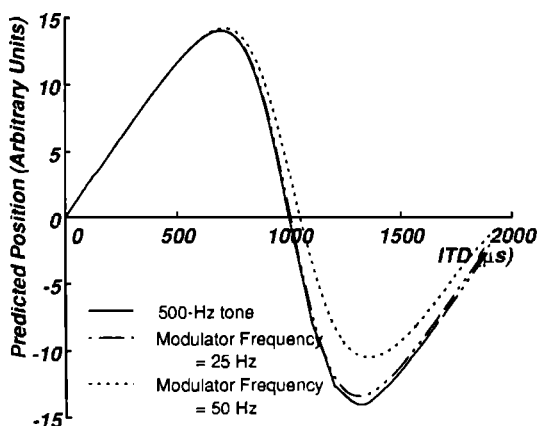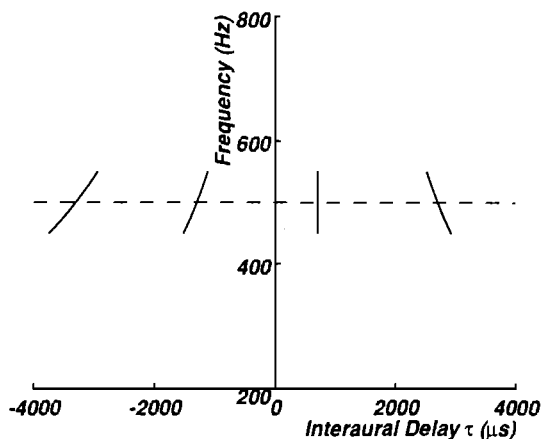

FIG. 9. Loci of the peaks of the cross-correlation function for 500-Hz amplitude-modulated narrow-band noise with a bandwidth of 50 Hz and a waveform ITD of 700 μs.

of maxima span only a very narrow bandwidth, and lateralization must be dominated by the relative centrality of the trajectories. The most central trajectory is at $\tau = 700 \mu s$ for an ITD of 700 μs (Fig. 9), and it passes through $\tau = -700$ μs for an ITD of $1300 \mu s$ (Fig. 10). As the frequency of modulation of the SAM tones increases, the trajectories of maxima of the cross-correlation function span an increasingly wider range of frequency, and the weighted-image model predicts that the trajectories that are straighter should be weighted increasingly more heavily. For the stimulus with the 700-μs waveform delay, the more central trajectory at 700 μs is also the straighter one, and hence lateral position should remain relatively unchanged as the frequency of modulation increases. When the waveform ITD is $1300 \mu s$, however, the trajectory at $1300 \mu s$ is the straightest. Therefore, as the frequency of modulation increases, the lateral position is predicted to change as the trajectory at $1300 \mu s$ becomes weighted increasingly more heavily because it is straighter than the more central trajectory at $-700 \mu s$.

Application of the straightness and centrality concepts for the original SAM tone data would also produce predictions that exhibit the major trends seen in Fig. 8 if an appropriate description of the physiological processing of SAM tones were available. This can be asserted with confidence

FIG. 8. Predictions of weighted-image model for the data of Fig. 7. The vertical axis has been linearly scaled by an arbitrary constant.
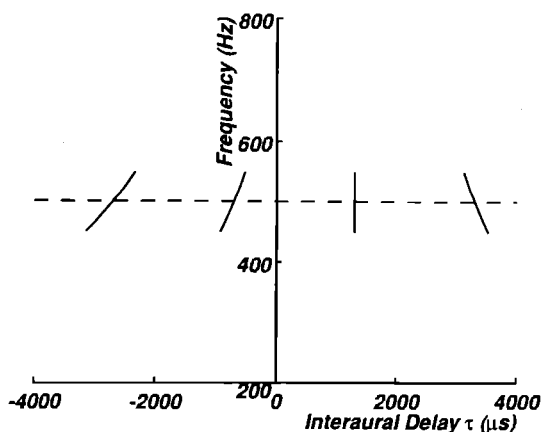

FIG. 10. Loci of the peaks of the cross-correlation function for 500-Hz amplitude-modulated narrow-band noise with a bandwidth of 50 Hz and a waveform ITD of 1300 μs.

because SAM stimuli with waveform delays (like all other stimuli) *always* produce a straight trajectory at the value of τ corresponding to the ITD, regardless of the specifics of the model. The other trajectories would take on a form that is more complex than the loci shown in Figs. 9 and 10, but they will be relatively less straight in all cases. In summary, predictions will be dominated by centrality considerations at small modulation frequencies, but straightness becomes an increasingly important factor as the modulation frequency increases.

The predictions in Fig. 8 for 500-Hz tones (the solid curve) can also be compared to earlier data obtained by Sayers (1964), Domnitz and Colburn (1977), and Yost (1981), which describe the position of tonal stimuli as a function of ITD (or IPD) using different experimental procedures. These predictions are an excellent description of the data of Domnitz and Colburn. They also provide a good description of the data of Sayers and Yost except for stimuli with ITDs of about half the period of the tone (which correspond to 180° IPDs). The data for these stimuli from the various experiments differ: Sayers' results indicate that subjects hear these stimuli at each of the two ears and occasionally at the center of the head, Yost's observers tended to lateralize these stimuli consistently to one side of the head, and Domnitz and Colburn's subjects described images with centers that were at the center of the head. While many of these differences in the data are due, in part, to differences in experimental methods and response procedures, it is clear that tonal stimuli with 180° IPDs are quite labile and difficult to lateralize in a consistent fashion. The weighted-image model (like the position-variable model) predicts that IPDs of 180° will produce an image that is at the center of the head, because its predictions are based on a centroid computation. At other ITDs, the data from these three studies are in reasonably good agreement with each other, with the data of Bernstein and Trahiotis (1985a,b), and with the predictions of the weighted-image model.

C. Lateralization of low-frequency transients presented with group delay

Because of difficulties in characterizing the auditory processing of SAM tones (including those cited above), Henning (1983) performed a series of lateralization experiments using brief bandpass transient signals for which ITDs in the "envelope" and fine structure could be separately manipulated. Using Henning's notation, the signal to one ear in these experiments is

$$s_L(t) = \frac{2A_0}{\pi} \frac{\sin[(\omega_{BW}/2)(t)]}{t} \sin[\omega_c(t)], \quad (9a)$$

and the signal to the other ear is

$$s_R(t) = \frac{2A_0}{\pi} \frac{\sin[(\omega_{BW}/2)(t - \Delta t_{gr})]}{t - \Delta t_{gr}}$$

$$\times \sin[\omega_c(t - \Delta t_{gr}) - \Delta t_c \omega_c], \quad (9b)$$

where ω_{BW} is the bandwidth of the signal and ω_c is the carrier frequency. When $\Delta t_{gr} = -\Delta t_c$, the binaural signal is said to exhibit group delay but not phase delay (a condition that is equivalent to the modulator ITD of the Bernstein and
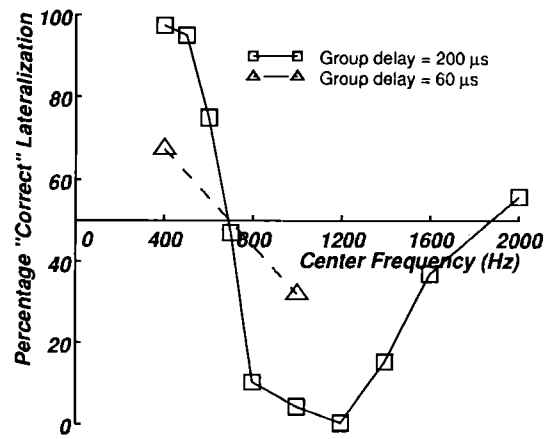

FIG. 11. Lateralization performance with 60- and 200-μs group delays as a function of the center frequency of an 800-Hz-wide transient signal.

Trahiotis experiments). Therefore, the maxima of the cross-correlation function for these stimuli are also described by Eq. (8) with the parameter τ_c set equal to zero.

Figure 11 contains results obtained by Henning (1983) for 800-Hz-wide narrow-band transients as a function of center frequency. Henning's subjects reported whether the stimuli were perceived to the left or to the right of the subjective midline, and results are plotted in terms of the percentage of "correct" or consistent lateralizations. If the perceptual image is unimodal, then this percentage should be monotonically related to the expected position predicted by the weighted-image model. The most outstanding feature of these data is that the transients are clearly lateralized to one side of the head for low frequencies, to the other side of the head at somewhat higher frequencies, and at the middle of the head at the highest frequencies examined.

Figure 12 shows theoretical predictions of the weighted-image model for the data plotted in Fig. 11. The predictions exhibit many of the significant features of the data, including the movement of the image from one side of the head to the other as center frequency increases from 400 to 1000 Hz. An examination of the trajectories of the maxima of the cross-correlation functions of Henning's stimuli indicated that these results cannot be explained in terms of competition between straightness and centrality. Henning (1983) suggested that the data shown in Fig. 11 reflect the dominance
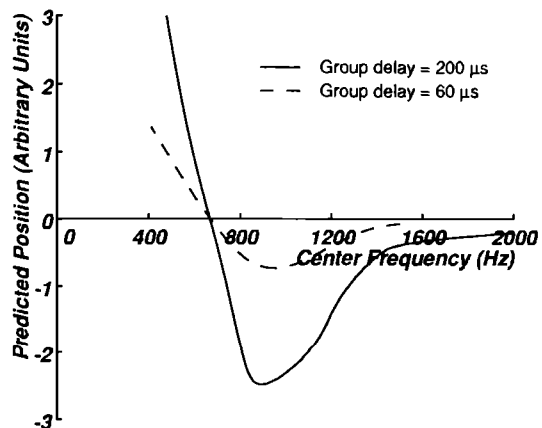

FIG. 12. Predictions of the weighted-image model for the data of Fig. 11.

of stimulus components near about 700 Hz for the lateralization of complex stimuli. This "dominant region" effect has been explored by Raatgever (1980) and others, and it is reflected in the model by the frequency-weighting function $q(f)$ (Fig. 5). In order to explore further the importance of straightness and the dominant-frequency effect in the lateralization data of Fig. 11, we obtained two additional sets of predictions of the model for these stimuli. First, we removed the straightness weighting from the predictions by setting the straightness parameter σ_i^2 equal to zero. Second, we removed the effects of the dominant frequency region by setting the frequency-weighting function to a constant independent of frequency. These predictions confirm Henning's hypothesis in that the shape of the predictions was affected by removing the frequency-weighting function but not by removing the straightness-weighting parameter.

IV. DISCUSSION

A. Limitations of the weighted-image model

Although we believe that the basic conceptual framework of the weighted-image model is valid and useful, it is clear to us that many details need to be worked out more carefully. For example, while we estimated the straightness of each trajectory for the theoretical predictions by computing its second central moment along the τ axis (i.e., its "variance"), there are a number of other equally plausible measures that need to be explored. Presently, we are especially concerned with trying to understand how the relative weighting of the centroids of the maxima should be determined.

In addition, the weighted-image model in its present form has some fundamental weaknesses. The most significant structural weakness of the model is that it describes mechanisms of lateralization in terms of operations on trajectories of maxima of the cross-correlation function without specifying how the system is able to identify these maxima and cluster them into discrete trajectories. While some stimuli (such as broadband noise presented with an ITD of the entire waveform) exhibit cross-correlation functions with clearly identified trajectories of maxima in the τ–f plane, other stimuli (such as SAM tones with modulator delays equal to half the period of the carrier component) produce cross-correlation functions that appear to split and merge at various frequencies. The weighted-image model cannot be easily applied to such stimuli. Other stimuli such as SAM tones with high carrier frequencies produce cross-correlation functions with many small local maxima and minima, and some criteria must be established for identifying which are the "important" maxima. Finally, we again note that the auditory system becomes unable to track the fine structure of stimuli as frequency increases about 1600 Hz. As a result, many of the maxima and minima of the interaural cross-correlation functions will become less pronounced as timing information about the fine structure diminishes. In order for the model to describe the perception of these stimuli, it must be modified so that it can specify when these peaks become too small to affect perception.

We believe we have gained a great deal of insight by analyzing predictions of the weighted-image model even though it has the shortcomings cited above. For example, we found that the trade between straightness and centrality in the model plays a very important role in understanding the lateralization of bandpass noise as a function of bandwidth (Fig. 4) and low-frequency SAM tones as a function of modulation frequency (Fig. 7). On the other hand, straightness does not greatly affect lateralization in other cases such as experiments using tonal stimuli, and the laterality of low-frequency transients presented with pure group delays (Fig. 11).

Presently, we believe that the best research strategy is to incorporate the best features of the weighted-image model into a more physiologically plausible model in order to adequately describe these complex stimuli. We are currently attempting to modify the position-variable model (Stern and Colburn, 1978) to incorporate an attribute similar to the straightness of the trajectories in the weighted-image model.

B. Envelope detection or consistency over frequency?

Many recent discussions of the processing of complex and high-frequency binaural stimuli have focused on the relative salience of timing information that can be provided by the envelopes as compared to the fine structure of the stimulus. Some researchers discuss information obtained from the envelope in terms of group delay, or the local slope of the phase spectrum. (Variations of the fine structure can be discussed in terms of phase delay, which, at a given frequency, is equal to the IPD divided by the frequency.) It is reasonable, then, to think of the salience of timing information available in the envelopes of complex stimuli in terms of the extent to which group delay is consistent over a range of frequencies. If a particular stimulus has a group delay that is consistent over frequency, that group delay will correspond to an interaural phase difference that increases linearly with frequency. This, of course, is exactly the condition encountered if one has a stimulus with an ITD that is consistent over frequency, which produces a stimulus with at least one perfectly straight trajectory. In other words, those stimuli that produce trajectories that are straight in the τ–f plane are also exactly those stimuli that tend to exhibit pronounced envelopes with interaural timing information that can be used for localization. As a stimulus condition is manipulated in such a way as to decrease the amount of straightness, the salience of the envelope is also likely to decrease. As a result, one cannot distinguish on the basis of these data and predictions along whether the phenomena are mediated by processing of interaural delays of the envelopes of the stimuli (in the strict sense), or by the competition between the straightness of the trajectories of the cross-correlation function versus the centrality of the various modes in that function.

We generally believe that consideration of the straightness versus the centrality of the modes of the interaural cross-correlation function is a more intuitively pleasing way to think of how we process binaural sounds than comparisons of the ITDs of the envelope versus the fine structure of the stimuli, or comparisons between interaural group delay versus phase delay. First, theories based on emphasis of the straight components of the cross-correlation function are

based on a pattern recognition concept, that of searching for the ITD that is most likely to be generated by a physical sound source. Second, we believe that the straightness–centrality trade describes the lateralization phenomena equally well at high frequencies (at which timing information in the envelopes is generally assumed to mediate lateralization) as it does for stimuli at low frequencies (at which timing information in the fine structure of a sound had been previously assumed to mediate lateralization). As discussed above, envelopes of high-frequency stimuli possess ITDs that are consistent over a range of frequencies. The weighted-image model is not very well suited to describe the lateralization of these higher frequency stimuli because it cannot neglect the cycle-by-cycle information that they contain. We are confident that a more physiologically plausible model that incorporates the concepts of straightness and centrality should be able to describe these high-frequency data.

V. SUMMARY AND CONCLUSIONS

In this article we examined the results of several studies of the subjective laterality of low-frequency complex binaural sounds. We found that many of the phenomena considered can be predicted by a very simple model that is based on the centroids of the trajectories of the maxima of the interaural cross-correlation function of the stimuli after peripheral bandpass filtering. In order for the model to predict the data, it must weigh more heavily the contributions of those regions of the cross-correlation function for which the ITD of the stimulus is consistent over a range of frequencies. We believe that this property of the model and the data are a consequence of a natural tendency of the auditory system to attempt to form perceptual images under assumptions that the sound source is physiologically realizable. Because of this we believe that it is more appropriate to model the lateralization of complex binaural stimuli in terms of the consistency of the interaural cues presented over a range of frequencies than in terms of physical attributes such as phase delay versus group delay.

ACKNOWLEDGMENTS

Bernstein, L. R., and Trahiotis, C. (1985a). "Lateralization of Low-frequency, Complex Waveforms: The Use of Envelope-based Temporal Disparities," J. Acoust. Soc. Am. 77, 1868–1880.

Bernstein, L. R., and Trahiotis, C. (1985b). "Lateralization of Sinusoidally Amplitude-modulated Tones: Effects of Spectral Locus and Temporal Variation," J. Acoust. Soc. Am. 78, 514–523.

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

Colburn, H. S., and Durlach, N. I. (1978). "Models of Binaural Interaction," in Handbook of Perception, Vol. IV, Hearing, edited by E. C. Carterette and M. P. Friedman (Academic, New York).

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

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

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

Henning, G. B. (1983). "Lateralization of Low-frequency Transients," Hear. Res. 9, 153–172.

Jeffress, L. A. (1972). "Binaural Signal Detection: Vector Theory," in Foundations of Modern Auditory Theory, Vol. II, edited by J. V. Tobias (Academic, New York).

Lindemann, W. (1986). "Extension of a Binaural Cross-correlation Model by Contralateral Inhibition. I. Simulation of Lateralization for Stationary Signals," J. Acoust. Soc. Am. 80, 1608–1622.

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

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

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

Raatgever, J. (1980). "On the Binaural Processing of Stimuli with Different Interaural Phase Relations," Doctoral dissertation, Technische Hogeschool Delft, The Netherlands.

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

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

Schiano, J. L., Trahiotis, C., and Bernstein, L. R. (1986). "Lateralization of Low-frequency Tones and Narrow Bands of Noise," J. Acoust. Soc. Am. 79, 1563–1570.

Shear, G. D., and Stern, R. M. (1988). "Lateralization of Broadband and Complex Stimuli: Predictions of the Position-Variable Model," J. Acoust. Soc. Am. (in preparation).

Stern, R. M., Jr., and Colburn, H. S. (1978). "Theory of Binaural Interaction Based on Auditory-Nerve Data. IV. A Model for Subjective Lateral Position," J. Acoust. Soc. Am. 64, 127–140.

Trahiotis, C., and Bernstein, L. R. (1986). "Lateralization of Bands of Noise and Sinusoidally Amplitude-Modulated Tones: Effects of Spectral Locus and Bandwidth," J. Acoust. Soc. Am. 79, 1950–1957.

Trahiotis, C., and Stern, R. M., Jr. (1988). "Lateralization of Bands of Noise: Effects of Bandwidth and Differences of Interaural Time and Phase," J. Acoust. Soc. Am. (in preparation).

Yost, W. A. (1981). "Lateral Position of Sinusoids Presented with Interaural Temporal and Intensive Differences," J. Acoust. Soc. Am. 70, 397–409.