Binaural Mechanisms that Emphasize Consistent Interaural Timing Information over Frequency

Richard M. Stern[1] and Constantine Trahiotis[2]

[1]*Department of Electrical and Computer Engineering and Biomedical Engineering Program, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA*

[2]*Department of Surgery (Otolaryngology, Surgical Research Center and Center for Neurosciences, University of Connecticut Health Center, Farmington, Connecticut 06030, USA*

1. Introduction

Recent studies have evaluated the potency of interaural temporal differences (ITD), interaural phase differences (IPD), interaural intensitive differences (IID) and their interaction as a function of frequency. Many of these phenomena have been discussed in excellent reviews by Durlach and Colburn (1978), Grantham (1995), and Hafter and Trahiotis (1996), among others.

Several recent contributions have discussed the importance of consistency, over frequency, of information concerning interaural timing. For example, Jeffress (1972), Stern *et al*. (1988), Trahiotis and Stern (1989), and Buell *et al*. (1994) have shown that the laterality of bandpass noise depends on the bandwidth of the stimulus, as well as its ITD, IPD, and IID.

In order to explain lateralization phenomena, most modern models of binaural interaction utilize an ensemble of units which record interaural coincidences in activity from primary auditory-nerve fibers with matched characteristic frequencies (CFs). At each CF, the collective response of these units is an approximation to the interaural cross-correlation of the stimulus, after processing by the auditory periphery. These models have been described and discussed in recent reviews by Colburn (1995) and Stern and Trahiotis (1995, 1996). Stern *et al*. (1988) have suggested how such cross-correlation-based models can account for the effects of bandwidth on lateralization. Their explanation was based on two, sometimes conflicting, aspects of the internal binaural representation of the stimuli: (1) *straightness*, which represents the extent to which maxima in the interaural cross-correlation of the stimuli appear at the same internal delay over a range of frequencies; and (2) *centrality*, the extent to which maxima of the cross-correlation function occur at internal delays of small magnitude. Stern *et al.* (1988) quantified these concepts using a black-box formulation called the *weighted-image model*. The weighted-image model utilized a simple computational procedure to estimate straightness that could only be applied to a very limited number of types of stimuli. Although the weighted-image model successfully described the phenomena to which it was applied, it was considered an interim formulation because of its *ad hoc* nature and because it could not easily be generalized.

In this paper we will briefly review the quantitative formulation of the extended position-variable model. We will then illustrate how the auditory system appears to exploit interaural temporal coincidences across frequency to achieve lateralization for certain sets of sinusoidally-amplitude-modulated (SAM) tones. An appealing feature of the approach was that such a mechanism was physiologically plausible (*e.g.* Takahashi and Konishi, 1986).

2. Lateralization of Bandpass Noise

Figure 1 illustrates the relative response of the first-level coincidence counters to a band of noise with a center frequency of 500 Hz, a bandwidth of 400 Hz, and an ITD of 1.5 ms. The horizontal axis corresponds to the fixed internal delay of the coincidence-counting units, and the oblique axis refers to the CF of those units.

The upper panel of Fig. 1 represents the expected number of coincidences as a function of internal delay and CF. Note that maxima can be seen at an internal delay of 1.5 ms for all values of CF, and that the resulting pattern appears to be straight and parallel to the CF axis. There are also secondary ridges at interaural delays other than 1.5 ms that are not parallel to the CF axis because the stimulus has a wider bandwidth than the peripheral filters. As a consequence of the broad stimulus bandwidth, the secondary ridges are separated along the internal-delay axis by the reciprocal of the characteristic frequency of the peripheral bandpass filters, and not by the reciprocal of the 500-Hz center frequency of the stimulus.

The curves in the upper panel of Fig. 1 do not take into account the fact that there are more coincidence-counting units with internal delays of smaller magnitude than there are with internal delays of larger magnitude (*e.g.* Kuwada *et al*., 1987). The functional significance of having a greater number of internal delays of smaller magnitude is referred to as centrality. It is manifested in binaural models by a greater weighting of the internal delays of smaller magnitude. Such a weighting has been described and discussed by several researchers (*e.g.* Blauert and Cobben, 1978; Colburn, 1977, Shackleton *et al*., 1992; Stern and Shear, 1996). The curves in the central panel of Fig. 1 represent the relative number of coincidences after centrality weighting by a function approximating the relative number of fiber pairs. The resulting function is a representation of the total actual number of coincidences observed for all units in the matrix. The function we use to describe the distribution of internal delays was developed by Colburn (1973, 1977) and refined by Stern and Shear (1996). Note that the consequence of applying the weighting function is that the straight ridge at 1.5 ms (the internal delay that represents the actual ITD of the stimulus), is greatly attenuated because there are not many coincidence-counting units with an internal delay of 1.5 ms.

Nonetheless, listeners actually lateralize according to the true stimulus ITD. In order to see why this is true, we now focus on straightness , which refers to the e xtent to which maxima in the interaural cross-correlation of the stimuli occur at the same internal delay over a range of frequencies. The lower panel of Fig. 1 shows the dramatic effects of applying the second level of coincidence weighting. Note that such weighting provides much greater emphasis to the straight ridge at 1.5 ms. The emphasis of the straight ridge at an internal delay of 1.5 ms occurs because, for that ridge, all of the first-level coincidence counters are responding essentially simultaneously at or near their maximum output. In contrast, the ridge closer to the midline (*i.e.* at an internal delay of approximately 0 ms) becomes attenuated because of the relatively reduced responses at CFs below approximately 600 Hz at that internal delay.

In addition to providing the weighting of straightness, this manner of combining coincidence information across frequency has the serendipitous advantage of sharpening the ridges of the two-dimensional cross-correlation function along the internal-

delay axis. For example, the ridges in the lower panel of Fig. 1 exhibit a smaller width (along the internal-delay axis) than the ridges in the upper panel of Fig. 1. The sharpening of the ridges along the internal-delay axis occurs because the rate functions of the outputs of the second-level coincidence counters are approximately proportional to the products of the rate functions of the (first-level) coincidence counters that comprise their inputs. For straight ridges, this has the effect of enhancing the peaks and suppressing the valleys in the patterns of second-level coincidence output. Recently Stanford and his colleagues showed that ITD-tuning is sharper for single units in the medial geniculate body as compared to units within the inferior colliculus of awake rabbits (Stanford *et al.*, 1992). As shown above, such sharpening along the internal-delay axis can occur without the explicit lateral-inhibition network proposed by Lindemann (1986).

We believe that the actual location of a perceived image results from a competition between straightness and centrality. If all other factors are equal, we believe that images will be lateralized toward the side of the head with the straightest ridge of coincidences. On the other hand, if all ridges are equally straight, lateralization would be dominated by centrality. This makes sense intuitively: real sound sources produce ITDs that are consistent over frequency. By weighting more heavily those values of internal delay in the model that appear to be responding to stimulus ITDs that are consistent over frequency, we are weighting more heavily the response to possible sound sources that are more physically plausible.

3. Extending the Position-Variable Model to Incorporate Straightness Weighting

In this section we describe how the straightness-weighting factor can be used to extend the position-variable model (Stern and Colburn, 1978) so that it can account for the behavioral data.

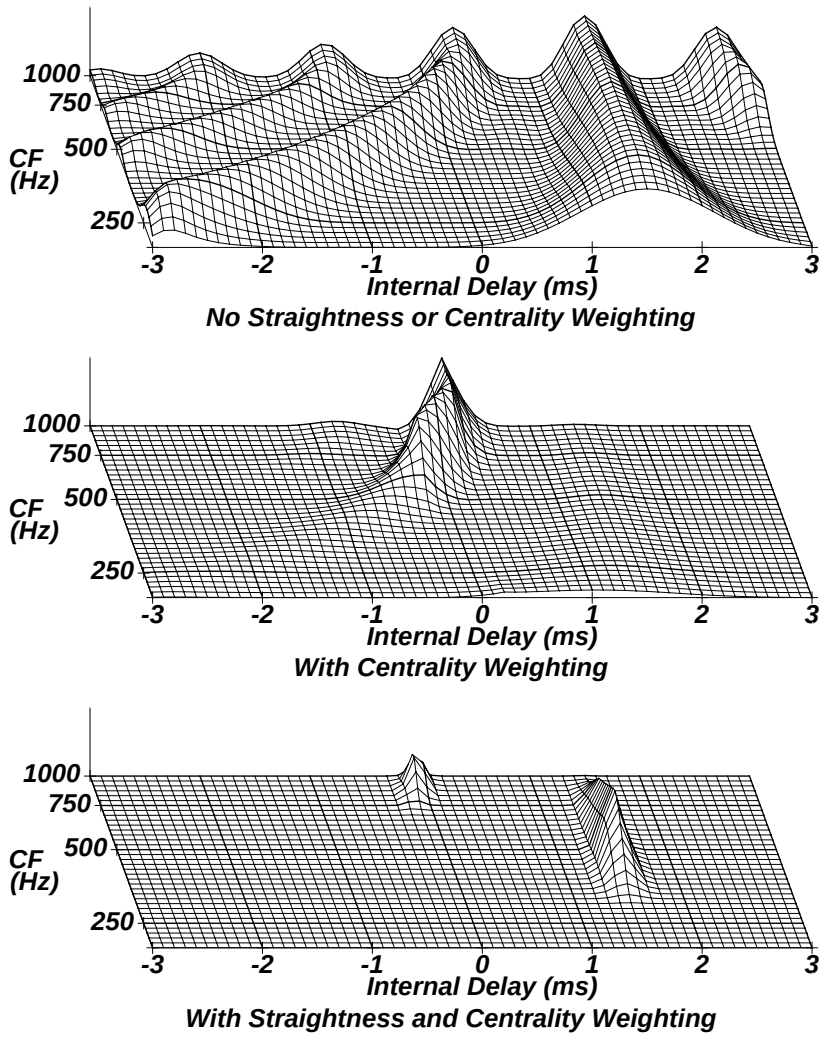

Figure 1. Relative response of an ensemble of binaural coincidence-counting units to a broadband noise with a center frequency of 500 Hz and a bandwidth of 800 Hz, presented with a 1.5-ms ITD. Upper panel: The relative number of coincidences per fiber pair as a function of internal delay τ (in ms) and CF of the auditory-nerve fibers (in Hz). Central panel: The average total number of coincidences as a function of internal delay and CF, taking into account the assumed density of internal delays. Lower panel: Relative response of the layer of second-level units that compute coincidences over frequency of the outputs of the original coincidence counters with the same internal delay.

The position-variable model provides predictions for these stimuli by computing the centroid along the internal-delay axis of cross-correlation functions such as those shown in Fig 1. Specifically, if $L_m(\tau_m, f_m)$ represents the number of coincidences recorded by a coincidence-counting unit of internal delay τ_m and CF f_m, the predicted position of the model is

$$\hat{P} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tau_m \, L_m(\tau_m, f_m)\, p(\tau_m, f_m)\, df_m \, d\tau_m}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L_m(\tau_m, f_m)\, p(\tau_m, f_m)\, df_m \, d\tau_m} \tag{1}$$

where $p(\tau_m, f_m)$ is the function that specifies the distribution of fiber pairs. Further details about the position-variable model are available in Stern and Colburn (1978) and Stern and Shear (1996).

The straightness-weighting stage is implemented by extending the position-variable model to include the *second* layer of coincidence-counting units referred to in Fig. 1. Each second-level unit is assumed to inputs from *N* first-level units that have the same characteristic delay, but come from first-level coincidence counters with logarithmically-spaced CFs. Mathematically, the expected number of coincidences of the second-level units can be written as

$$E[S_m(\tau_m, f_m)] = E\left[\prod_{i=-(n-1)/2}^{i=(N-1)/2} L_m(\tau_m, f_m\delta^i)\right] \tag{2}$$

where $L_m(\tau_m, f_m\delta^i)$ is the number of counts produced by a first-level coincidence counting unit with internal delay τ_m and characteristic frequency $f_m\delta^i$, and $S_m(\tau_m, f_m)$ is the number of counts produced by a second-level coincidence counting unit with internal delay τ_m over a range of characteristic frequencies centered at f_m. *N* is the number of outputs from first-level fiber pairs that are assumed to converge on a second-level coincidence-counting unit, which we refer to as the order of the straightness weighting. The parameter δ determines the spacing of the characteristic frequencies of the first-level cells, which we refer to as the straightness step . The operation *E*[Æ] denotes statistical expectation over the coincidence-counting units.

Taking advantage of the statistical independence of the firing rates of the fiber pairs, Eq. (2) can be rewritten as

$$E[S_m(\tau_m, f_m)] = \prod_{i=-(n-1)/2}^{i=(N-1)/2} E[L_m(\tau_m, f_m\delta^i)] \tag{3}$$

Equation (3) states that each second-level straightness unit fires at a rate proportional to the product of the firing rates of the *N* first-level units that provide its input. If there is a high rate of output from all of the *N* first-level units, the second-level (straightness) unit will also fire at a high rate. If the outputs of any of the *N* first-level units are low, then the second-level unit will fire at a relatively low rate. Intermediate values of first-level firing-rates result in a second-level firing rate between the two extremes. This type of across-frequency interaction serves to emphasize those regions in the $\tau_m - f_m$ plane that exhibit consistent peaks at a single internal delay over a range of frequencies. Increasing *N* provides greater straightness weighting by including more first-level coincidence-counting units as inputs to the second-level units. Increasing δ accomplishes a similar effect by increasing the separation in CF of the first-level units that are input to a given second-level unit. Compared to the outputs $L_m(\tau_m, f_m\delta^i)$ of the first-level units, the outputs $S_m(\tau_m, f_m)$ of the network of second-level units exhibit place emphases at values of τ that are consistent over a range of CFs.

4. Lateralization of SAM Tones with Temporally-Coherent Envelopes

In order to evaluate whether the mechanism for straightness requires synchronous activity across a range of CFs, we employed sets of stimuli for which the intracranial images would be the same when simply averaged over frequency but would be dramatically different if straightness weighting were accomplished by the second level of coincidences proposed in Sec.3.

We made use of ensembles of three sinusoidally amplitude-modulated (SAM) tones with identical modulation frequencies but three different carrier frequencies. A typical modulation frequency was 20-25 Hz. Typical carrier frequencies included the harmonically-related set of 300, 500, and 700 Hz, as well as several inharmonically-related sets of frequencies containing similar frequencies. We refer to these carrier frequencies as f_1, f_2, and f_3, respectively. The first set of SAM tones (which we call Set 1 stimuli) were presented with the modulation monaurally in phase for all three carrier frequencies. The second set of SAM tones (which we call Set 2 stimuli) were presented with the modulation of the components with the outer two carrier frequencies monaurally in phase with respect to each other, but out of phase with respect to the component with the central carrier frequency.

We compared the binaural images produced by these two sets of SAM tones. We chose values of the stimulus parameters for which the amplitude modulation was sufficiently slow to ensure that the components at each of the three carrier frequencies remained unresolved.. The three carrier frequencies were chosen to be sufficiently separated to preclude interaction of their respective sidebands, yet be close enough to become fused into a single binaural image when presented as in Set 1. The interaural delay for stimuli comprising both Set 1 and Set 2 was always 1.5 ms, three-quarters of a period for the carrier of the 500-Hz SAM tones.

The crux of the argument concerned whether simple across-frequency averaging would be sufficient to produce lateralization toward the leading side (as suggested by Jeffress, 1972, and Shackleton *et al.,* 1992), or whether it would be necessary to have temporally-coincident neural activity across a range of frequencies within the binaural representation. The following discussion follows closely Trahiotis and Stern (1994).

For our purposes it was important that the two sets of SAM tones differ only in terms of the phase of amplitude modulation applied to the carrier frequencies f_1 and f_3, and that all three components of both sets of stimuli have exactly the same interaural time difference, T_S. Specifically, the stimuli of Set 1 have the property that the peaks of the amplitude modulation at each of the three carrier frequencies occur simultaneously. Hence
the responses of the interaural coincidence-counting units at each of the three frequencies would be expected to occur more or less simultaneously, and the second layer of coincidence-counting units which signal consistency of interaural delay across frequency would be activated. On the other hand, the stimuli of Set 2 would be expected to produce approximately-simultaneous responses of the coincidence-counting units only for the two outer carrier frequencies that are modulated in phase (f_1 and f_3). The peaks of modulation of the remaining carrier frequency (f_2) occur at the valleys of modulation of f_1 and f_3 and *vice-versa*. As a result, there should be substantially less activity at the second level of coincidence.

If across-frequency coincidence of neural activity of the interaural timing information is salient, these two sets of stimuli should produce *different* binaural images. Alternatively, if the binaural lateralization mechanism simply averages over running time and frequency (as had been assumed, for example, by Jeffress, 1972, and Shackleton *et al.*, 1992), then the stimuli of Set 1 and Set 2 should have the same spatial qualities.

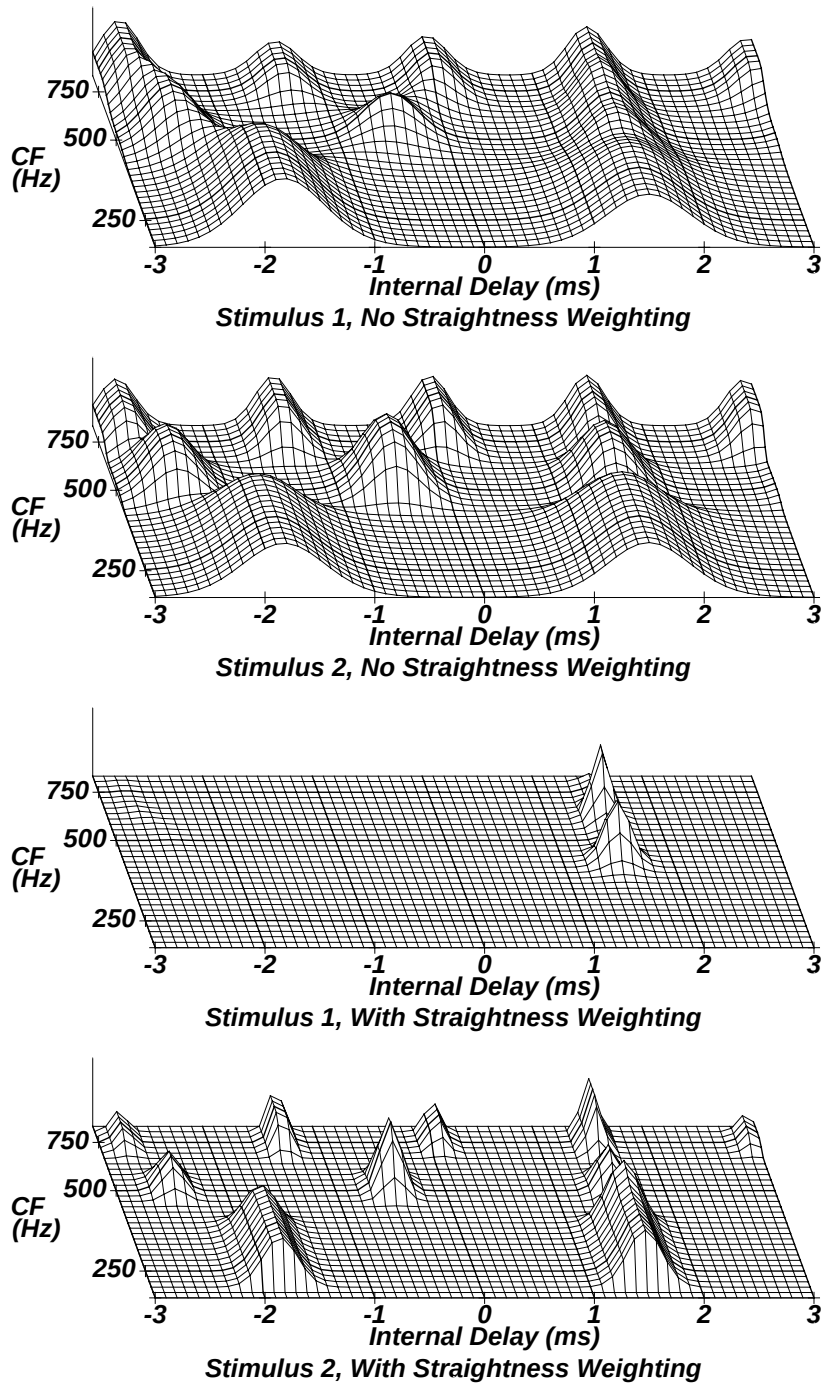

Figure 2. Responses to the two sets of triple-SAM stimuli as specified in the text, with and without the effects of straightness weighting by the second level of coincidence-counting units. Predictions were obtained using a value of N equal to 5, and a value of δ *equal to .01.*

We listened to a number of such stimuli. All of the stimuli from Set 1, for which the peaks of the amplitude modulation at each carrier frequency were always in phase, were heard as a single, compact, binaural image that was well-lateralized toward the ear that received the signal that was leading in time. The compactness and location of the binaural image indicates that some type of across-frequency integration of the internal response according to the actual interaural time delays was taking place. The binaural spatial properties of the stimuli from Set 2 were quite different and can be described in two manners. For some combinations of carrier frequency and frequency of modulation, stimuli from Set 2 were heard as two distinct images. The images corresponding to the outer two carrier frequencies (f_1 and f_3) were heard

toward the ear receiving the leading signal. *At the same time*, however, the image corresponding to f_2 was heard toward the opposite ear (which received the signal lagging in time). For other combinations, stimuli from Set 2 produced a diffuse hollow-sounding image that frequently filled the head .

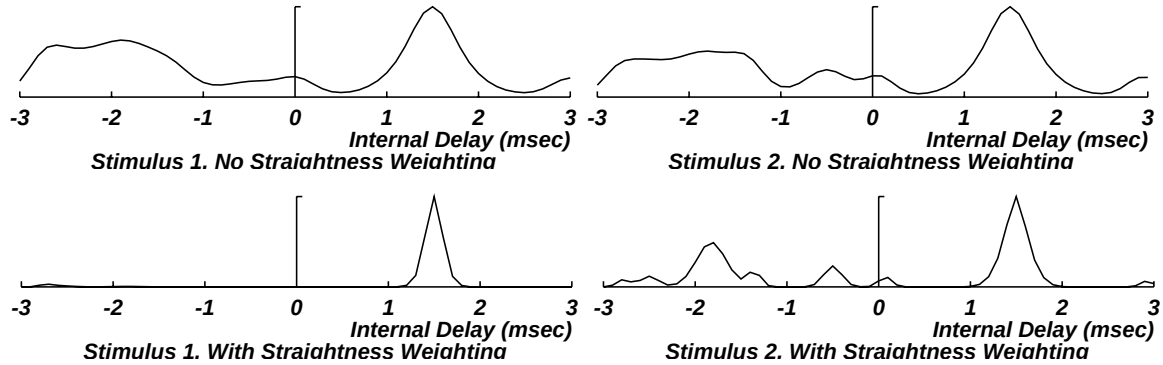

Figure 3. Responses to the two sets of triple-SAM stimuli as in Fig. 2, but summed over CF.

Figures 2 and 3 display the model s responses to the triplet of SAM tones, using a modulation frequency of 25 Hz and carrier frequencies of 300, 500, and 700 Hz. The upper two panels of each figure show responses obtained by the first level of coincidence-counting units, without straightness weighting. The lower two panels of each figure show responses for the same stimuli after straightness weighting. The responses after straightness weighting depend on the specific choice of values for the parameters N and δ used in the calculations. The functions in Figs. 2 and 3 were plotted using N equal to 5 and δ equal to .01 and .05 in Figs. 2 and 3, respectively. These parameter values are in the range of parameters that accurately describe the lateralization of noise as a function of bandwidth, ITD, and IPD, and the pattern of responses in these plots is typical of what we observed for other reasonable values of N and δ as well. Figure 4 depicts the response patterns shown in Fig. 2 summed across CF, as suggested by Jeffress (1972) and Shackleton *et al.* (1992).

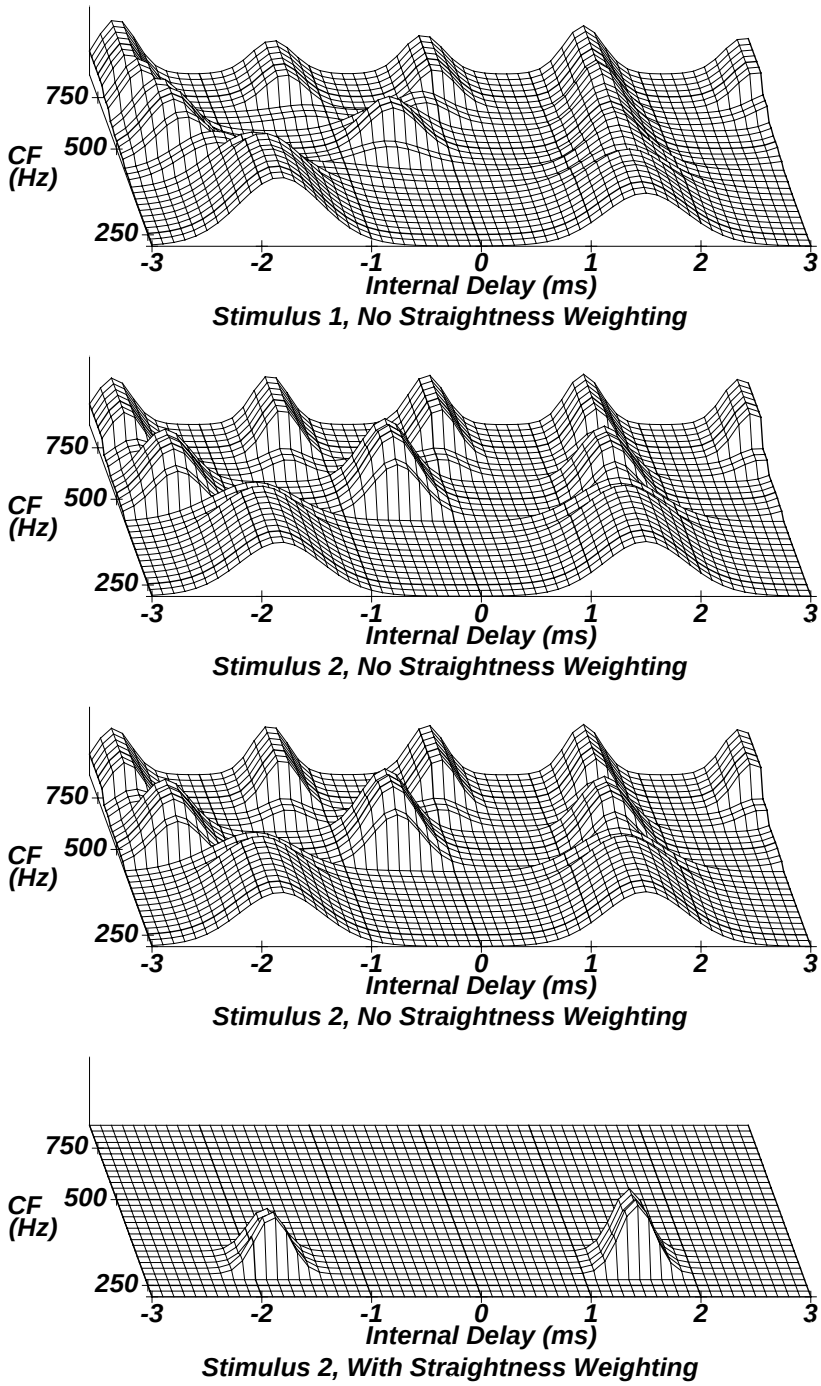

Figure 4. Same as Fig. 2, except that predictions were obtained using a value of N equal to 5, and a value of δ *equal to .05.*

As can be seen in the upper two panels of Figs. 3 and 4, without straightness weighting, the responses to the stimuli of Set 1 and Set 2 are virtually identical, even though they are perceived very differently. As can be seen in the lower two panels of Figs. 2 and 3, the effect of straightness weighting is to greatly attenuate responses at internal delays other than the actual ITD of the stimulus. These effects are also reflected in the response patterns when summed across CF, as seen in Fig. 5: without straightness weighting the responses to the triple-SAM stimuli are almost identical, but with straightness weighting they clearly differ. The stimuli from Set 1 were predicted to exhibit a compact image lateralized far toward the side of the head receiving the leading signal, as in fact they were perceived. The stimuli from Set 2, on the other hand, were predicted to produce either multiple or diffuse images, as again they were perceived. Although the predictions of the model for the stimuli in Set 2 depend on the assumed values of δ, in each case the predicted response exhibits multiple modes over a range of internal delays.

We believe that these predictions and observations using triple-SAM tone complexes provide independent support for our contention that the across-frequency integration of binaural information is not due to a simple averaging of the internal response to binaural stimuli. Instead, it appears that the mechanism of integration depends upon the extent to which responses occurring within each frequency channel are temporally proximate. That is the defining feature of the second level of coincidence postulated by Stern and Trahiotis (1992).

Acknowledgments

This work was supported by Research Grant Number NIH DC-00234 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We also thank Samuel H. Tao and Gerald Ng for their contributions to this work.

References

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

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (1994) Lateralization of bands of noise as a function of combinations of interaural intensitive differences, interaural temporal differences, and bandwidth. J. Acoust. Soc. Am. 95, 1482-1489.

Colburn, H. S. (1973) Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination. J. Acoust. Soc. Am. 54, 1458-1470.

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. (1995) Computational models of binaural processing. In: H. Hawkins and T. McMullin (Eds.), Auditory Computation. Springer-Verlag, New York, pp. 332-400.

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

Grantham, D. W. (1995) Spatial hearing and related phenomena. In: B. C. J. Moore (Ed.), Hearing. Academic Press, New York, pp. 297-345.

Hafter, E. R., and Trahiotis, C. (1997) Functions of the Binaural System. In M. J. Crocker (Ed.), Handbook of Acoustics, Wiley, New York, in press.

Kuwada, S., Stanford, T. R., and Batra, R. (1987) Interaural phase-sensitive units in the inferior colliculus of the unanesthetized rabbit: effects of changing frequency. J. Neurophysiol. 57, 1338-1360.

Jeffress, L. A. (1972). Binaural Signal Detection: Vector Theory. In: J. V. Tobias (Ed.), Foundations of Modern Auditory Theory, Vol. II, Academic Press, 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.

Shackleton, T. M., Meddis, R., and Hewett, M. J. (1992). Across frequency integration in a model of lateralization. J. Acoust. Soc. Am. 91, 2276-2279 (L).

Shear, G. D. (1987), Modeling the Dependence of Auditory Lateralization on Frequency and Bandwidth. M. S. Thesis, Electrical and Computer Engineering Department, Carnegie Mellon University.

Stanford, T. R., Kuwada, S., and Batra, R. (1992) A comparison of the interaural time sensitivity of neurons in the inferior colliculus and thalamus of the unanesthetized rabbit, J. Neurosci. 12, 3200-3216.

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.

Stern, R. M., and Shear, G. D. (1996) Lateralization and detection of low-frequency binaural stimuli: effects of distribution of internal delay. J. Acoust. Soc. Am. 100, 2278-2288.

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

Stern, R. M., and Trahiotis, C. (1995) Models of Binaural Interaction. In: B. C. J. Moore, (Ed.), Handbook of Perception and Cognition, Volume 6: Hearing. Academic Press, New York, pp. 347-386.

Stern, R. M., and Trahiotis, C. (1996) Models of Binaural Perception. In: R. H. Gilkey and T. R. Anderson (Eds.), Binaural and Spatial Hearing in Real and Virtual Environments. Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 499-531.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (1988b) Lateralization of complex binaural stimuli: a weighted image model. J. Acoust. Soc. Am. 84, 156-165.

Takahashi, T. A., and Konishi, M. (1986) Selectivity for interaural time differences in the owl s midbrain. J. Neurosci., 6, 3413-3422.

Trahiotis, C., and Stern, R. M. (1989) Lateralization of bands of noise: effects of bandwidth and differences of interaural time and intensity. J. Acoust. Soc. Am. 86, 1285-1293 (L).