Across frequency integration in a model of lateralization

Trevor M. Shackleton,[a] Ray Meddis, and Michael J. Hewitt
Speech and Hearing Laboratory, Department of Human Sciences, University of Technology, Loughborough LE11 3TU, United Kingdom

(Received 22 May 1991; accepted for publication 13 December 1991)

A computational model of binaural lateralization is described. An accurate model of the auditory periphery feeds a tonotopically organized multichannel cross-correlation mechanism. Lateralization predictions are made on the basis of the integrated activity across frequency channels. The model explicitly weights cross-correlation peaks closer to the center preferentially, and effectively weights information that is consistent across frequencies more heavily because they have a greater impact in the across frequency integration. This model is complementary to the *weighted-image model* of Stern *et al.* [J. Acoust. Soc. Am. **84**, 156–165 (1988)], although the model described in this paper is physiologically more plausible, is simpler, and is more versatile in the range of input stimuli that are possible.

PACS numbers: 43.64.Bt, 43.66.Ba, 43.66.Qp, 43.66.Pn [WAY]

INTRODUCTION

It is widely accepted that binaural processing occurs in frequency selective bands. However, recent studies (Dye, 1990; Stern *et al.*, 1988; Trahiotis and Stern, 1989) have shown that lateralization decisions are necessarily based upon the information in a frequency region wider than a single critical band, even if this results in nonoptimum performance.

A model for across-frequency integration of binaural information has recently been described as the *weighted-image model* (Stern *et al.*, 1988), based upon the insights of Jeffress (1972). The purpose of this letter is to show that very similar results to Stern's can be obtained using a model which has a simpler binaural mechanism and which also, unlike Stern's, has accurate peripheral processing. Only a restricted range of comparisons between model and experimental data are reported here because cross-correlation models have already been shown to be successful in replicating narrow-band data (e.g., Colburn, 1973, 1977; Colburn and Latimer, 1978; Stern and Colburn, 1978). The across-frequency pooling used has also been shown to be applicable to a large range of pitch phenomena (Meddis and Hewitt, 1991).

I. THE ACROSS-FREQUENCY INTEGRATION MODEL

The model comprises two sections. The first is a representation of the auditory periphery and the second is the binaural analysis mechanism.

The peripheral model has already been described elsewhere (Meddis and Hewitt, 1991; Meddis *et al.*, 1990). All of the peripheral model parameters (excepting the number of filters) are the same as described in these earlier papers. The peripheral model simulates the action of the ear canal and middle ear, the bandpass filtering nature of the basilar membrane, and the complex, nonlinear, properties of the hair-cell/auditory nerve synapse. The inputs to the binaural mechanism from the left and right ears, $l(f,t)$ and $r(f,t)$, are the probabilities of a set of high-spontaneous rate auditory-nerve fibers firing (taking the refractory period into account), and thus represents a scaling of the average instantaneous nerve firing rate. For each side there are a set of 78 frequency channels equally spaced on an ERB-rate scale between 50 and 3000 Hz (Moore and Glasberg, 1986).

The binaural mechanism is an explicit implementation of Jeffress' (1972) model that performs a binaural cross correlation on frequency-selective channels, weights central peaks more heavily than lateral peaks, and combines information across frequencies in making a position judgement. Although cross correlation is a mathematical rather than a physiological operation, it is widely believed that very similar operations are performed in the Medial Superior Olive (Yin and Chan, 1988). The output from the auditory-nerve fibers from each ear which are tuned to the same frequency are processed by a running cross correlation with a time constant Ω of 10 ms. The output from the cross correlator with internal delay τ at time t, is

$$\phi(f,t,\tau) = q(f)p(\tau)\sum_{i=0}^{3\Omega\Delta t} l(f,t-i\Delta t)r(f,t-i\Delta t-\tau)e^{-i\Delta t/\Omega}, \quad (1)$$

where Δt is the inverse of the sampling rate (20 kHz). The choice of time constant is not critical in the simulations reported in this letter. A long time constant (about 100 ms) is suggested by several experiments (Grantham, 1982; Grantham and Wightman, 1978, 1979), but the use of such a long time constant would produce inordinately long run times, so 10 ms was chosen as a compromise.

A central weighting function $p(\tau)$ is applied to each cross-correlation function to emphasize those delays corresponding to central stimuli. This weighting function is Gaussian in form and has a standard deviation (s.d.) of 600 μs. A spectral weighting function $q(f)$ is then applied to each cross-correlation channel. This weighting is the same as de-

[a] New address: Laboratory of Experimental Psychology, Sussex University, Brighton BN1 9QG, E. Sussex, United Kingdom.

scribed in Stern *et al.* (1988, p. 160), based upon the binaural dominance measurements of Raatgever (1980).

The running cross correlation is sampled at the end of the input signal, and this sample is used for further analysis. This set of functions is shown in the upper portion of Fig. 1. All 78 channels are summed across frequency to form a summary cross correlogram $s(t,\tau)$ (shown at the bottom of Fig. 1) which is used to generate predictions from the model

$$s(t,\tau) = \sum_{i=1}^{78} \phi[f(i),t,\tau], \qquad (2)$$

where $f(i)$ are equally spaced on an ERB-rate scale between 50 and 3000 Hz (Moore and Glasberg, 1986).

We use one of two metrics to indicate lateral position. One position estimator P_{peak} is the position of the largest peak in the summary cross correlogram. The other estimator P_{av} is the average position of all the peaks in the summary cross correlogram weighted by their height. The choice of whether to use P_{av} or P_{peak} is, to a certain degree, *ad hoc* but we prefer to use P_{peak} unless there is evidence that subjects are using judgement averaging (cf. Sayers and Lynn, 1968) in which case P_{peak} is preferred. These metrices are more fully discussed in Sec. III.

II. EVALUATION OF THE MODEL

A. Delayed, phase-shifted noise

Trahiotis and Stern (1989; Stern *et al.*, 1989) have designed a stimulus which combines interaural delays with interaural phase differences (IPDs). If a signal is time delayed to one ear relative to the other, then an ITD which is constant across frequency is obtained, and all cross-correlation peaks will be coincident. This is described as producing a "straight" cross-correlation track. If, however, an IPD is imposed, then the ITD will vary across frequency and the cross-correlation peaks will be in different positions in each frequency channel. This is called a "sloping" cross-correlation track.

In their experiments Trahiotis and Stern (1989; Stern *et al.*, 1988) combined interaural delays with IPDs so that they could fix the ITD at 500 Hz, and alter the slope of the cross-correlation track around 500 Hz. Subjects were required to match the position of a narrow-band noise with a variable interaural level difference (ILD) to the position of the test stimulus. Their results, normalized and averaged across subjects, are shown in Fig. 2(a) as a function of stimulus bandwidth (Stern *et al.*, 1988). These data are adequately matched by the model. Predictions based upon the average peak position P_{av} are shown in Fig. 2(c). P_{av} is used here because subjects were required only to make one judgement per condition, and the stimulus was repeated until the subjects made their lateralization judgement. Subjects were also specifically requested to estimate the centroid of stimuli which sounded diffuse. These conditions would favor judgement averaging (Sayers and Lynn, 1968).

Additional experimental data and modeling results are shown in Fig. 2(b) and (d), where the interaural delay was fixed at 1500 μs and IPD varied as the parameter (Trahiotis and Stern, 1989). The agreement between the model and experimental data is as good as the agreement between subjects (Trahiotis and Stern, 1989, Fig. 4). The 270° condition especially shows considerable divergence between subjects' responses. Part of this divergence can be explained by assuming that subjects' criteria vary between reporting the position of the highest peak (P_{peak}) and the average peak position (P_{av}). In this condition, this variability is particularly important because the combined ITD at 500 Hz corresponds to

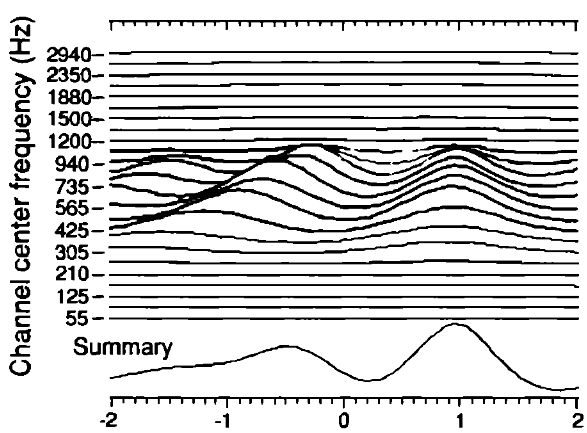

FIG. 1. Multichannel cross-correlation function and summary cross correlation for a noise stimulus centered on 500 Hz with a bandwidth of 200 Hz with an ITD of 1000 μs (IPD = 180° at 500 Hz). The tonotopically arranged cross-correlation functions $\phi(f,t,\tau)$ are shown in the top part of the figure, and the summary cross correlation $s(t,\tau)$ is shown in the bottom part. The functions are shown at the end of the stimulus $t = 200$ ms.

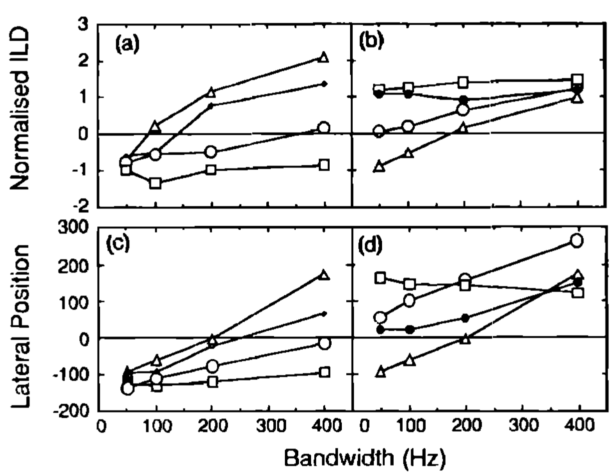

FIG. 2. Comparisons of experimental data and model results for delayed, phase-shifted noise. Experimental data are shown in (a) (Stern *et al.*, 1988) and (b) (Trahiotis and Stern, 1989). The results were normalized so that the lateralization of the 0 μs/270° stimulus at 50-Hz bandwidth was assigned a lateralization of − 1. Model results are shown in (c) and (d). The conditions in (a) and (c) are; triangles, 1500-μs delay, 0° IPD; diamonds, 1000-μs delay, 90° IPD; circles, 500 μs, 180°; squares, 0 μs, 270°. The conditions in (b) and (d) are a delay of 1500 μs combined with an IPD of; triangles, 0°; open circles, 90°; squares, 80°; filled circles, 270°.

180°, leading to cross-correlation peaks that are equidistant from the center.

B. The role of spectral dominance in lateralization

The spectral weighting function $q(f)$ was introduced to account for the binaural spectral dominance region (Bilsen and Raatgever, 1973; Raatgever, 1980). Similar results were demonstrated by Henning (1983). An 800-Hz bandwidth click was generated at a number of center frequencies with a group delay of 200 μs and zero phase delay. Subjects were instructed to judge the side on which the click was lateralized. The click was lateralized on the leading side for center frequencies below 700 Hz, but on the lagging side for center frequencies above 800 Hz [Fig. 3(a)]. Henning interpreted these data in terms of the lateralization being determined by the IPD in the region closest to 700 Hz. The position of the largest peak P_{peak} in the summary cross correlogram is shown in Fig. 3(b) as a function of center frequency both with, and without spectral weighting. Henning's data are better described when spectral weighting is included.

III. DISCUSSION

In this letter we have shown that a simple, across-frequency integration, model can simulate some lateralization experiments that present varying ITDs across frequency. The model is similar in some respects to the weighted-image model of Stern *et al.* (1988). The differences are that (i) we use an accurate peripheral model, (ii) we use the principle of summation across frequency rather than calculate track variance, and (iii) the central weighting function $p(\tau)$ is different. The two models appear to give similar results, which is reasonable since they possess a similar theoretical background.

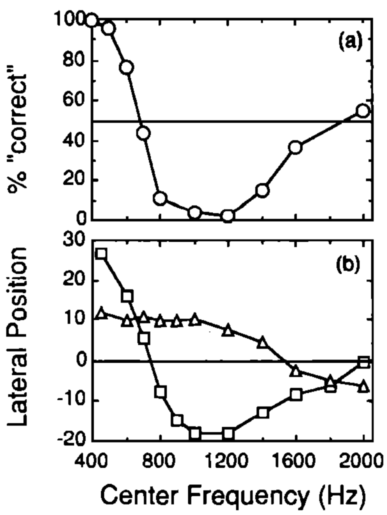

FIG. 3. A comparison of experimental data (a) and model predictions (b) for click stimuli of bandwidth 800 Hz with zero phase-delay and 200-μs group delay (Henning, 1983). The experimental results are expressed in terms of the percentage of responses that correspond to the sign of the group delay. The model predictions are expressed in terms of the predicted extent of lateralization. Predictions with spectral weighting included are shown by squares, predictions without spectral weighting are shown by triangles.

The use of a peripheral model is obviously an advantage since a great deal of nonlinear processing occurs in the periphery, however it is not expected that the peripheral model significantly alters the results obtained in the simulations reported in this letter.

The fundamental difference between the weighted-image model and our own is in how we choose to combine the information in different frequency regions. The weighted-image model concentrates upon individual cross-correlation peak tracks, finds their mean position, and then weights them according to their distance from the center and their spread along the delay axis, corrected for the length of the track [Stern *et al.*, 1988, Eqs. (4), (5); Stern *et al.*, 1991b]. The average of the weighted positions is then found.

In our model, we simply summate the cross-correlation functions across frequency and find a weighted average of peak positions. This can be shown to give a qualitatively similar result to the variance in the weighted-image model by considering individual cross-correlation peak tracks (Fig. 1). If the track is straight (on the right of Fig. 1), then the variance will be zero, and the across-frequency summation will yield a narrow peak. The track variance increases with the slope of the track, similarly the across-frequency summation yields a peak that becomes wider and lower as the slope increases (on the left in Fig. 1). In other words, the variance of the cross-correlation track is reflected by the height of the summary cross-correlogram peak. Although these procedures are likely to be quantitatively different, they share similar qualitative properties. In this respect we would expect the two models to behave similarly.

Early attempts by Stern to modify his *position-variable model* (Stern and Colburn, 1978) using a principle similar to our across-frequency summation were not successful (e.g., Stern *et al.*, 1991a). We feel that the reason for this is that he used a central-weighting function $p(\tau)$ that was much narrower than ours. This results in a track that is straight, but distant from the center, being greatly attenuated relative to a sloping track that is closer to the center. In this case, the model's predictions are biased toward the side opposite the straight track, whereas human listeners give responses closer to the straight track. To compensate for this a "straightness" factor was introduced in the weighted-image model to effectively "amplify" the straighter tracks and thus move the model predictions toward the straighter track. By using a wider central-weighting function we avoid the need to use such a "straightness" factor because straight tracks that are distant from the center are not attenuated as much. This results in predictions that are far closer to human performance than those obtained using the position-variable model (Stern and Colburn, 1978) without inclusion of a straightness factor. A discussion of the implications of this wider central-weighting function for other binaural phenomena is beyond the scope of this short letter.

All of the models share the same spectral weighting function, and weight central peaks more heavily. They also combine point estimates of peak positions into a centroidlike measure for lateralization. It is debatable whether this last feature is completely justifiable. In an experiment that encouraged subjects to report multiple *simultaneous* images, Shackleton *et al.* (1991) found that when an IPD of 180° is imposed on narrow-band stimuli then images are often heard simultaneously on both sides of the head, and that when only a single image was heard in the center of the head it was reported as being very diffuse. This complements experimental data summarized by Yost and Hafter (1987) that shows that some experiments produce reports of images extremely lateralized on both sides of the head, and other experiments produce reports of centralized images. This would suggest to us that the basic cross-correlation mechanism must be capable of producing multiple images to explain the bimodal data, but also have the capacity to combine these into a single, average, lateralization. This is the reason why our model has two possible lateralization measures, the centroidlike average of peak position P_{av} and the peak positions themselves. It requires further, carefully controlled, experimentation to determine whether there are any principles upon which the choice of which of these metrics is used can be made on anything other than a *post hoc* basis. Most stimuli will produce similar results for both metrics, however differences will arise when there are two significant peaks in the cross-correlation window (e.g., Fig. 1).

ACKNOWLEDGMENTS

This work was supported by the Science and Engineering Research Council (UK) Image Interpretation Initiative (Grant No. GR/E 88240). We thank Richard Stern for providing us with the computer program for his weighted-image model. We also thank Steve Colburn and an anonymous reviewer for their useful comments on an earlier version of this letter.

Bilsen, F. A., and Raatgever, J. (1973). "Spectral dominance in lateralization," Acustica 28, 131–132.

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., and Latimer, J. S. (1978). "Theory of binaural interaction based on auditory-nerve data. III. Joint dependence on interaural time and amplitude differences in discrimination and detection," J. Acoust. Soc. Am. 64, 95–106.

Dye, R. H. (1990). "The combination of interaural information across frequencies: Lateralization on the basis of interaural delay," J. Acoust. Soc. Am. 88, 2159–2170.

Grantham, D. W. (1982). "Detectability of time-varying interaural correlation in narrow-band noise stimuli," J. Acoust. Soc. Am. 72, 1178–1184.

Grantham, D. W., and Wightman, F. L. (1978). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. 63, 511–523.

Grantham, D. W., and Wightman, F. L. (1979). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. 65, 1509–1517.

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. 2*, edited by J. V. Tobias (Academic, New York), pp. 351–368.

Meddis, R., and Hewitt, M. J. (1991). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I pitch identification," J. Acoust. Soc. Am. 89, 2866–2822.

Meddis, R., Hewitt, M. J., and Shackleton, T. M. (1990). "Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse," J. Acoust. Soc. Am. 87, 1813–1818.

Moore, B. C. J., and Glasberg, B. R. (1986). "The role of frequency selectivity in the perception of loudness, pitch and time," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 251–308.

Raatgever, J. (1980) "On the binaural processing of stimuli with different interaural phase relations," Doctoral dissertation, Technische Hogeschool Delft, The Netherlands.

Sayers, B. M., and Lynn, P. A. (1968). "Interaural amplitude effects in binaural hearing," J. Acoust. Soc. Am. 44, 973–978.

Shackleton, T. M., Bowsher, J. M., and Meddis, R. (1991). "Lateralization of very short duration tone-pulses of low- and high-frequencies," Q. J. Exp. Psychol. 43A, 503–516.

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

Stern, R. M., Xu, X., and Tao, S. (1991a). "A coincidence-based model that describes straightness weighting in binaural perception," presented at the 14th Midwinter meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL.

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

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (1991b). "Erratum: Lateralization of complex binaural stimuli: A weighted-image model," J. Acoust. Soc. Am. 90, 2202.

Trahiotis, C., and Stern, R. M. (1989). "Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase," J. Acoust. Soc. Am. 86, 1285–1293.

Yin, T. C. T., and Chan, J. C. K. (1988). "Neural mechanisms underlying interaural time sensitivity to tones and noise," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 385–430.

Yost, W. A., and Hafter, E. R. (1987). "Lateralization," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 49–84.