Stereausis: Binaural processing without neural delays

Shihab A. Shamma
Electrical Engineering Department, Systems Research Center, and the University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742 and Mathematical Research Branch, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20982

Naiming Shen and Preetham Gopalaswamy
Electrical Engineering Department, University of Maryland, College Park, Maryland 20742

(Received 4 August 1987; accepted for publication 5 June 1989)

A neural network model is proposed for the binaural processing of interaural-time and level cues. The two-dimensional network measures interaural differences by detecting the spatial disparities between the instantaneous outputs of the two ears. The network requires no neural delay lines to generate such attributes of binaural hearing as the lateralization of all frequencies, and the detection and enhancement of noisy signals. It achieves this by comparing systematically, at various horizontal shifts, the spatiotemporal responses of the tonotopically ordered array of auditory-nerve fibers. An alternative view of the network operation is that it computes approximately the cross correlation between the responses of the two cochleas by combining an ipsilateral input at a given characteristic frequency (CF) with contralateral inputs from *locally off-CF* locations. Thus the network utilizes the delays already present in the traveling waves of the basilar membrane to extract the correlation function. Simulations of the network operation with various signals are presented as are comparisons to computational schemes suggested for stereopsis in vision. Physiological arguments in support of this scheme are also discussed.

PACS numbers: 43.63.Bq, 43.63.Qe, 43.66.Pn

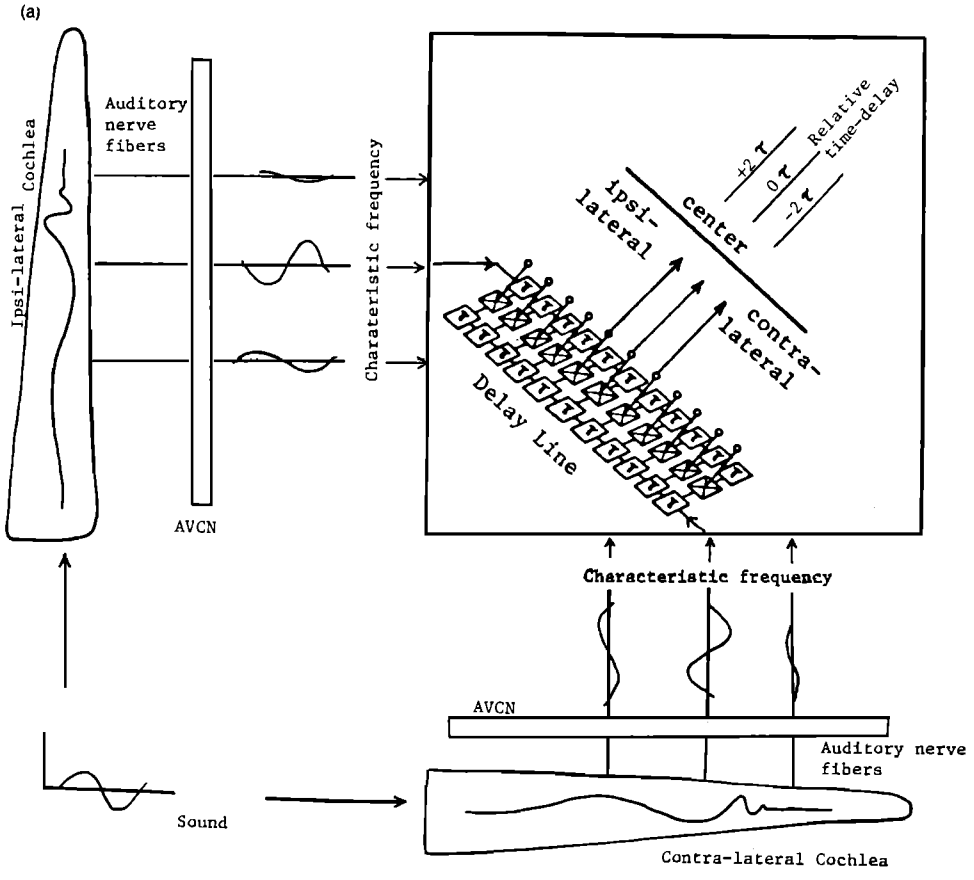
INTRODUCTION

The processing of binaural cues is fundamental to many tasks in spatial hearing. This is particularly true of *lateralization* and *signal detection and enhancement*—two phenomena that have been the subject of intensive multidisciplinary research for several decades (Blauert, 1983; Durlach and Colburn, 1978; Green and Yost, 1975). Numerous computational and phenomenological models have been proposed to account for the experimental, psychophysical, and neurophysiological data and to elucidate the underlying processes generating them. Basic to all these models are the extraction and exploitation of a measure of similarity (or dissimilarity) between the inputs to the two ears. The models differ, however, in the detailed nature of this operation; for instance, there are correlation-based models (Licklider, 1951; Sayers and Cherry, 1957), equalization and cancellation models (Durlach, 1972), and count comparison models (Colburn and Durlach, 1978).

Correlation-based models have been successful in accounting for the widest range of binaural phenomena and in providing a theoretical framework for investigations into the physiological bases and neural networks that can perform these functions. The primary computational structure in these models is the cross-correlator, which generates a measure of the correlation of activities arriving from the two ears. There are many variants of these algorithms, differing primarily in the details of cochlear frequency analysis, and in the nature of the variables at the inputs of the cross correlator [e.g., using stochastic point process models (Colburn, 1973; Colburn and Durlach, 1978) or continuous deterministic functions (Bilsen, 1977) to represent the responses of the peripheral auditory system]. These differences aside, however, the fundamental computation performed in all the proposed models above is a running cross-correlation measure between the cochlear outputs from the two ears at various time delays, i.e., a comparison between the *current* output of one cochlea with progressively delayed or *previous* outputs from the other cochlea. We shall refer to this correlation as a *temporal correlation* to distinguish it from the *spatial correlation* operations described later.

An essential component in implementing *temporal correlations* is the storage element (memory) needed to preserve past cochlear outputs. In searching for the neural substrate of such algorithms, the most common assumption has been to associate the various lags required in the computations with *neural* delays (Jeffress, 1948) (e.g., neuronal pathways of differing lengths or latency effects). Figure 1(a) illustrates a typical network based on these principles. Thus, following the frequency analysis of the cochlea, each output fiber projects to the central cross correlator with topologically ordered range of delays that allows its correlation with the contralateral output at the same characteristic frequency (CF) to be computed. Combining such functions from all output pairs at other CFs, a two-dimensional cross-correlation image results in which one axis represents the CFs of the cochlear outputs and the other represents different lags or delays. Details of these output patterns thus re-

 0001-4966/89/090989-18$00.80



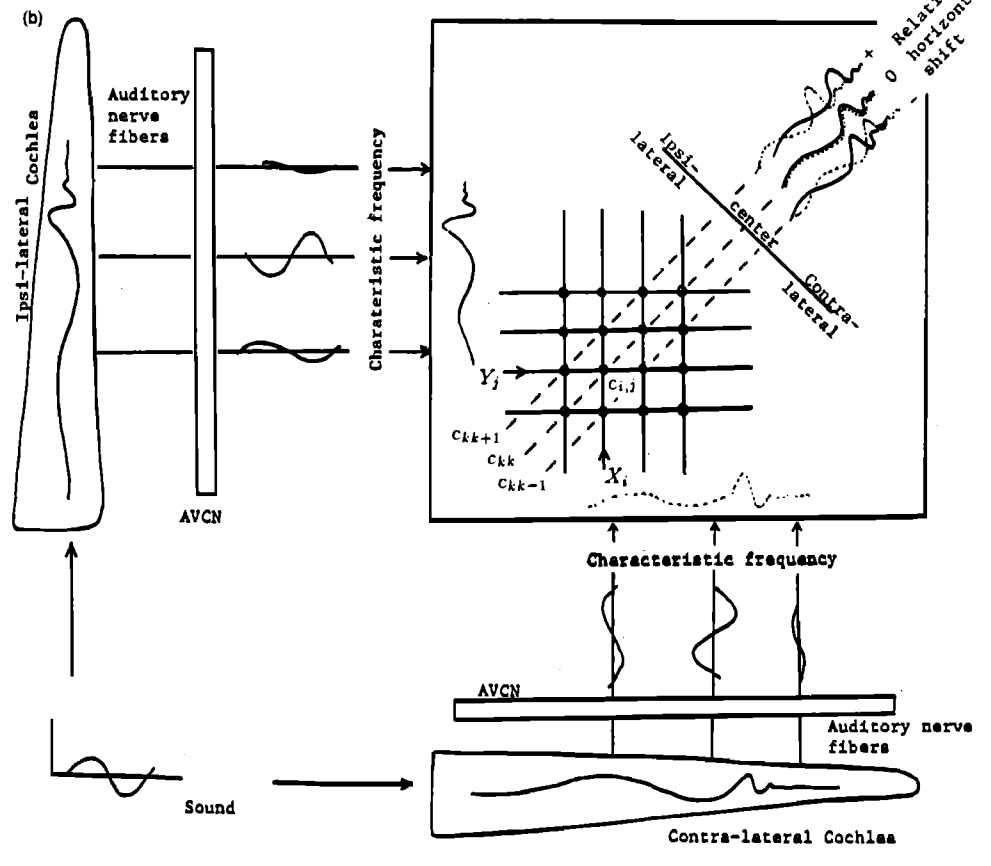


FIG. 1. Schematics of the binaural networks. (a) A schematic of the neural implementation of the correlation-based binaural processor using neuronal delay lines (Jeffress, 1948). The tonotopically ordered fiber arrays of the two auditory nerves project via the AVCNs to the SOC. Each ipsilateral fiber (x_i) synapses with a contralateral fiber (y_i) of the same CF through a series of time delays (τ). Activity along the τ axis indicates the lateralization of the stimulus. (b) A schematic of the sterausis binaural network. Ipsilateral (x_i) and contralateral (y_j) fibers are processed in an ordered matrix of operations C_{ij}.

flect the sound spectral and lateralization information along with two axes (Blauert, 1983; Colburn and Durlach, 1978). For instance, for a binaural tone, a peak of the correlation function appears at the CF corresponding to the frequency of the tone, and at the lag corresponding to the delay or phase shift between the two ears. This is all strictly applicable to interaural-time differences of low-frequency sounds ($\leqslant$1.5 kHz) or, with appropriate modifications, to slowly varying envelopes of high frequencies. For spatial hearing based on interaural-level differences, various mechanisms have been proposed to *augment* the above correlation models, especially for high-frequency sounds (Blauert, 1980).

Much psychophysical and neurophysiological data have been interpreted in support of this hypothesis. However, there is as yet no direct physiological support of the existence or the essential functional relevance of spatially organized neural delays in the mammalian auditory system. It should, therefore, be emphasized that the idea of neuronal delays did not arise out of a compelling experimental evidence, but rather to satisfy a literal interpretation of a convenient mathematical formulation (the *temporal correlation* models), which is coupled to a highly schematic view of cochlear function as merely a bank of bandpass filters. Thus, as we shall elaborate in this report, detailed *spatial* features of the spatiotemporal responses of the auditory nerve encode the binaural cues in a manner that makes performing *temporal correlations*, and, hence, the need for organized neuronal delays, theoretically unnecessary.[1] This possibility seems to have been first proposed by Schroeder (1977) to explain certain aspects of the "equalization and cancellation" theory of binaural hearing, and by Loeb *et al.* (1983) as a mechanism for monaural periodicity detection.

The fundamental objective of this paper is to illustrate how spatial features of the cochlear responses can be used to extract interaural time and level disparities and to generate output representations that are comparable to those obtained with the temporally based algorithms, and that account for most early binaural percepts. Spatially based algorithms, like their temporal counterparts, utilize coincidence detector operations, but they differ significantly in the way their inputs are organized. Furthermore, there are basic differences that emerge in the neural networks which implement these two types of algorithms, with the temporal schemes emphasizing systematic variations in the morphological features of its constituent neurons to produce the delays (e.g., axons or dendrites of regularly changing lengths, diameters, or time constants), while spatial schemes emphasizing various patterns of interneuronal connectivities. As we shall elaborate later, this *temporal–spatial* dichotomy in binaural processing finds its parallels in the algorithms proposed to estimate the spectrum from monaural auditory responses (Evans, 1978; Shamma, 1985a).

In the following sections, we shall first present a simple neural network model that utilizes the spatial responses of the cochlea in order to represent two basic psychophysical attributes of binaural hearing—lateralization and signal enhancement. The discussion of the model and its outputs will emphasize their qualitative nature. This is because the complicated form of the cochlear responses makes detailed quantitative predictions and comparisons to psychophysical data highly arbitrary given the large number of parameters in the cochlear model and the uncertainties that currently exist regarding cochlear nonlinearities and hair cell function.

Because of the fundamental similarity that emerges between the proposed network and the type of computations used for *stereopsis* in vision (Marr and Poggio, 1979), we refer to it as the *stereausis network*. In Sec. I, we shall outline the topology of this network and the basic principles underlying its operation. In Sec. II, the network outputs are interpreted for a wide range of tasks, and the results are finally discussed in Sec. III.

I. THE STEREAUSIS NETWORK

It is difficult at present to find conclusive evidence in support of any neural network model for binaural processing. At best, one may show that relevant physiological, anatomical, and psychophysical data are consistent, with various aspects of the model and, furthermore, that the basic design criteria are in harmony with the fundamental principles of organization in the auditory system. The following are a few of the guiding principles that led to the *stereausis network*, most of which are also satisfied by the Jeffress model (Jeffress, 1948) above.

(1) The primary pathways of the auditory system maintain their tonotopic (and, hence, topographic) order through several central nuclei and up to the cortex. This emphasizes the importance of the *spatial* dimension in auditory processing at all levels (Keidel and Neff, 1975).

(2) The fine temporal structure of the responses on the auditory nerve is crucial in binaural processing. It is largely preserved in the responses of the *Bushy cells* of the anteroventral cochlear nucleus (AVCN) which in turn project, partially via the nucleus of the trapezoid body (NTB), to the nuclei of the superior olivary complex (SOC), where significant binaural interactions are first recorded. The binaural networks that utilize this temporal information are presumably located at this level (Yin and Kuwada, 1984).

(3) Given the inherent variabilities in nerve cell properties, the binaural network operation should be robust with respect to small variations in cell thresholds and time constants, and in axonal lengths and targets of projecting fibers.

(4) It is unclear at present whether separate or identical binaural networks and pathways are involved in the processing of interaural-level (ILD) and interaural-time differences (ITD), and of different low- and high-frequency signals (Irvine, 1986). To clarify these issues, it is desirable to isolate the essential elements needed in the detection of the different cues in the models, and to address the possibility that one network is capable of gracefully processing both types of cues.[2] This applies both to continuous (ongoing) and to onset interaural differences.

(5) Finally, the network should be able to encode "naturally" other more complex attributes of spatial hearing such as diffuseness and compactness of sound, and perform such tasks as the enhancement of noisy signals and the integration of different source cues.

A simple neural network is proposed that adheres to the above principles, and does not require any neuronal delay

lines to perform its processing. The basic functional principle underlying the network operation is that binaural cues can be derived solely from the *spatial disparities* in the traveling waves of the two ears. For instance, a low-frequency tone produces in each cochlea a spatially distributed traveling wave that is projected relatively intact onto the responses of the spatially ordered array of auditory-nerve fibers.[3] At any instant in time, the central binaural processor receives two spatial *images* (or snapshots) of the traveling waves, one from each ear, via the pathways of the AVCN. When the tone is centered, the images are identical. For binaurally unequal signals, however, the traveling waves differ systematically. Thus, when the tone is phase shifted (or delayed) in one ear relative to the other, the instantaneous images appear correspondingly shifted in space. Since this *spatial* disparity between the traveling waves is proportional to the *temporal* delays between the two ears, the binaural processing of all interaural-time differences can be reduced to purely spatial operations. The same arguments apply to spatial disparities due to interaural-level differences that affect the relative amplitudes of the traveling waves. As we illustrate in later sections, many other possible inequalities in binaural inputs, for instance in their envelopes, degree of correlation, or bandwidths, can be readily detected and consistently represented via the spatial disparities between the resulting traveling waves.

A. The topology of the stereausis network

The binaural processing network is presumably located in the SOC (e.g., the MSO, the LSO, or a combination of both). It receives tonotopically ordered inputs via the AVCN. For ITD processing, it is crucial that this pathway preserves the fine temporal structure of the auditory-nerve responses. In the simulation results shown in this paper, the input patterns are generated using a simplified biophysical model of the basilar membrane and inner hair cells (Shamma *et al.*, 1986). The nerve responses are represented by the instantaneous probability of firing computed from the cochlear model (Shamma, 1985b; Shamma *et al.*, 1986). Figure 2 shows an example of the responses of this model to a 600-Hz tone. For details of the model parameters and computations, see Shamma *et al.* (1986) and the Appendix, Sec. 1. The stereausis network combines the ipsilateral and contralateral cochlear outputs in a simple ordered matrix of operations, as shown in Fig. 1(b). Thus, at the (ith, jth) node, the responses of the ith ipsilateral fiber (x_i) and the jth contralateral fiber (y_j) are combined to produce $c_{ij} = C(x_i, y_j)$. Here, $C(.,.)$ computes a measure of the *correlation* between the instantaneous activity of its two coincident input fibers. In this manner, the cochlear responses at a given CF location in one ear (x_i) is systematically correlated with outputs from CF and *off*-CF cochlear fibers of the other ear ($y_{i-1}, y_i, y_{i+1}, \ldots$). The significance of this arrangement for the cross-correlation computations is that, because of the finite velocity of the traveling waves, delayed versions of the responses at a given CF can be obtained from off-CF fibers in the local neighborhood of the CF, and not necessarily through further neuronal delays (Pfeiffer and Kim, 1975; Shamma, 1985b). An alternate view of these operations is that the stereausis network computes along its different diagonals (parallel to the center diagonal illustrated by the dashed line AB) the correlation of the two cochlear images at different *lateral* (*or horizontal*) *spatial* shifts. Thus, along the center diagonal [c_{kk} axis in Fig. 1(b)], the cochlear patterns are spatially aligned. Off this diagonal, however, and along axes parallel to it (c_{kk+1} and c_{kk-1} axes), the output is computed from inputs that are horizontally shifted relative to each other. In this sense, cells along each of these

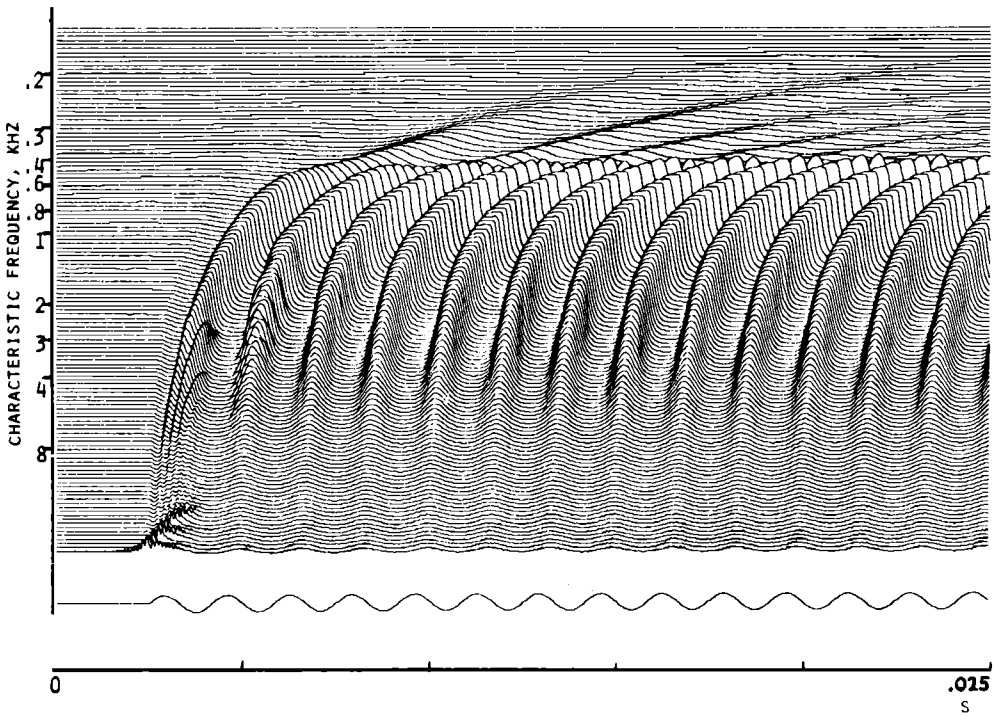


FIG. 2. Responses of the cochlear model. The spatiotemporal response patterns of the cochlear model (Appendix, Sec. 1) to a 600-Hz tone. The spatial (ordinate) axis represents the basal-to-apical (bottom-to-top) spread of the cochlear partition; it is labeled by the CF of each output channel (see method of CF labeling in the Appendix Sec. 1). The input waveform of the tone is also shown. In the auditory system, these patterns represent the probability of firing on the auditory nerve.

axes act as disparity detectors for a characteristic *lateral* spatial disparity.

As we shall discuss in more detail shortly (and in the Appendix, Sec. 2), the network outputs (c_{ij}) are further processed through simple spatial lateral inhibitory and excitatory interactions to enhance the final patterns. In order to facilitate the presentation of the results, the output patterns are finally averaged over a short time interval (typically 12.5 ms). Note that the $C(.,.)$ function reflects not only the operations expected to take place in the appropriate physiological structures (e.g., MSO or LSO) but also incorporates other possible transformations along the way (e.g., a change of sign in NTB). The exact form of this function is unimportant provided it generates a consistent correlation measure of its half-wave-rectified inputs ($x_i(t), y_j(t)$).

All the outputs illustrated in this paper are generated using the following form of $C(.,.)$: $c_{ij} = (x_i + y_j)^2$. Besides its simple and computationally efficient form, this operation was chosen because it approximately mimics two response-properties most commonly seen in binaural MSO cells: (1) Both inputs are excitatory, and (2) the *binaural* dynamic range is larger than that of monaural inputs, reflecting facilitatory binaural interactions. This latter property can also be modeled by more complex mathematical forms involving both excitatory and inhibitory interactions or other nonlinearities (Yin and Kuwada, 1984). All the results of this paper were qualitatively reproduced (although not shown) using other forms of $C(.,.)$ such as: $c_{ij} = x_i + y_j$, $c_{ij} = g(x_i - y_j)$, where $g(.) = \max(.,0)$ is a threshold operation, and $c_{ij} = x_i y_j$ (this operation does not work for ILD detection; see Sec. II B). A summary of all the model computations is given at the end of this section.

B. Examples of the network outputs for synthetic patterns

In order to highlight the major features of the network operations and the nature of its output representation, we shall first illustrate its responses to two simple traveling patterns of activity mimicking crudely the cochlear waves. Figure 3(a) and (b) shows the network-*averaged* outputs for single peaks sweeping the spatial axis of the two inputs. In Fig. 3(a), the inputs are identical at all times, and, consequently, the trajectory of the maximum output occurs at units along the diagonal AB. When a peak is delayed in one input relative to the other [ipsilateral leads in Fig. 3(b)], the location of the maximum shifts proportionately, reflecting the magnitude and direction of the instantaneous disparity of the input patterns. In this way, the network effectively operates as an ordered array of disparity detectors that sense

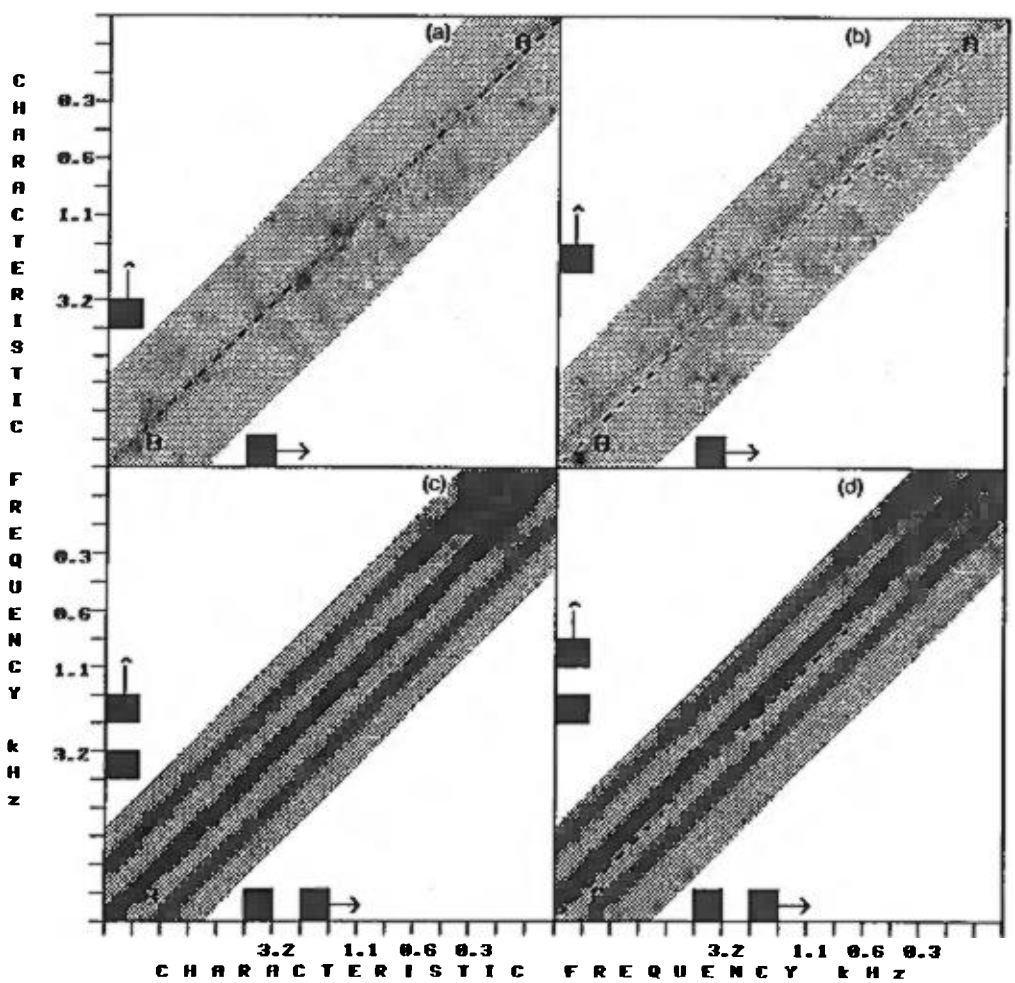


FIG. 3. Stereausis processing with synthetic patterns. (a) The output of the stereausis network with a synthetic pattern (a binaural traveling pulse). The centered pattern is represented by two pulses sweeping the cochlear partition in synchrony. Dark shading indicates areas of maximal activity. The ipsilateral CF axis is the ordinate; the contralateral axis is the abscissa. (b) The output of the stereausis network with the synthetic pattern binaurally delayed. The ipsilateral pulse leads the contralateral pulse. (c) The network output with a pair of centered traveling pulses. (d) Same as (c), but with pulses binaurally delayed. Ipsilateral leads are contralateral.

and evaluate the *temporal* delays between the two inputs by virtue of the accompanying *spatial* disparities they produce.

Figure 3(c) and (d) illustrates the outputs for two traveling peaks a distance D apart. As before, when the inputs are centered [Fig. 3(c)], the units along the diagonal AB become maximally activated. Two parallel but smaller secondary maxima also emerge at cells that correlate outputs from spatial locations separated by the same distance D. In Fig. 3(d), a relative temporal delay between the input patterns (ipsilateral pattern leading) causes a displacement of the entire output as expected. Note also that, for a given delay, the spatial shift of the stereausis patterns in either example is proportional to the *velocity* of the input peaks.

C. The network outputs with a continuous low-frequency tone

A low-frequency tone evokes a traveling wave that is conveyed to the binaural network via the phased-locked responses of the input pathways. The responses of the binaural network with this input are qualitatively similar to those discussed above in that a primary correlation maximum emerges, accompanied by several secondary peaks due to the multiple peaks within the envelope of the traveling wave [Fig. 4(a), middle]. In addition, the responses exhibit two unique features related to the amplitude and phase characteristics of the cochlear filters [Fig. 4(b) and (c)]. (1) The output activity is concentrated around the CF location specific to the tone (CF = 600 Hz in this case). Varying the frequency of the tone would cause the active region to move along the *spectral axis* of the network, i.e., along and parallel to the AB diagonal. The *spectral plot* in the upper-left inset figure samples the activity along and adjacent to this spectral axis. (2) The secondary maxima of the network outputs appear to converge toward the primary maximum [located along the AB diagonal in Fig. 4(b)]. This is due to the rapidly increasing slope of the spatial phase function of the traveling wave, and the accompanying decrease of the spatial separation of its peaks at and beyond the point of resonance.

When an interaural-time delay is introduced, the traveling waves of the two ears become relatively delayed, causing a spatial disparity between the binaural inputs to appear [Fig. 4(a)] (see footnote 4). As in the case of the synthetic inputs earlier, the ITD causes the binaural patterns to shift off the AB axis. This shift is clarified in the *disparity plot*, which samples the network outputs near and along the bar drawn in the figures.[5] Disparities may also be caused by interaural-level differences. In either case, the binaural network reflects the presence of these cues by changes in the output patterns [Fig. 4(c)]. The results of these and other manipulations will be discussed in detail in the next section.

Finally, the spatial spread of these outputs reflects both the bandwidth of the cochlear filters and the saturation of the auditory-nerve-fiber responses due to their limited dynamic range (30–40 dB; see the Appendix, Sec. 1). At higher sound levels, the spread of the binaural patterns increases significantly. The patterns can be readily "focused" with a somewhat elongated *on-center/off-surround* mask, i.e., a two-dimensional lateral inhibitory network (LIN) that is applied uniformly to the entire binaural c_{ij} output (see the Appendix, Sec. 2). This LIN mask enhances the cochlear outputs in a manner very similar to that of the LIN used in monaural spectral estimation (Shamma, 1985a), in that it spatially highpasses the patterns of binaural coincidence, emphasizing the activity in regions of rapid spatial change (near the point of resonance) while suppressing outputs from relatively flat regions of activity.[6] In all the illustrations that follow, this operation is applied to the c_{ij} outputs *before* taking the final short-time averages. Applying the LIN mask *after* averaging produces similar results in most cases, the main difference being in the case of large ILD disparities where the spectral sharpening of the stronger input is worse for post-averaging LIN.[7] Purely for computational convenience, the LIN mask is applied here in a nonrecursive (feedforward) manner; qualitatively similar results can be obtained with recursive (feedback) connections (see the Appendix, Sec. 2).

D. Summary

To summarize, interaural-time and interaural-level differences result in spatial disparities between the instantaneous responses of the two ears. A simple binaural network of cells correlating the cochlear responses of different locations along the tonotopic axis can detect and compute the binaural cues of the signal. There are two major axes of information in the network: (1) the *disparity or lateralization axis:* activity projected against this axis encodes the perceived lateralization of the auditory event; (2) the *spectral (CF) axis:* activity on this axis reflects the spectrum of the stimulus. Therefore, in an approximate sense, each cell in the network responds maximally to a characteristic spatial shift, at a characteristic frequency. In the following sections, we shall interpret the two-dimensional response patterns relative to these two axes.

The computations performed to generate the binaural outputs shown in this paper are summarized as follows.

(1) The stimulus is processed by a cochlear model to generate the spatiotemporal response patterns of the auditory nerve (Fig. 2). All responses are expressed in terms of the instantaneous firing rates (rather than the stochastic firings) of the cells. The patterns are projected to the central binaural network in a spatially organized way, as shown in Fig. 1(b).

(2) Each (i,j)th node of the binaural network performs the following coincidence (correlation) operation:

$$c_{ij}(n) = [x_i(n) + y_j(n)]^2, \tag{1}$$

where $x_i(n)$ and $y_j(n)$ are the ipsilateral and contralateral inputs at time n. Thus, at each time instant (n), a two-dimensional plane (frame) of activities $[c_{ij}(n)]$ is computed.

(3) Each frame is then processed by the LIN nonrecursive mask (given in the Appendix, Sec. 2). Thus, at each (ij), the output (o_{ij}) is computed from neighboring c_{ij}'s as follows:

$$o_{ij}(n) = g\left(\sum_{kl} w_{ijkl} c_{kl}(n)\right), \tag{2}$$

where w_{ijkl} is the two-dimensional LIN mask centered around (i,j), and $g(.) = \max(.,0)$ represents a thresholding

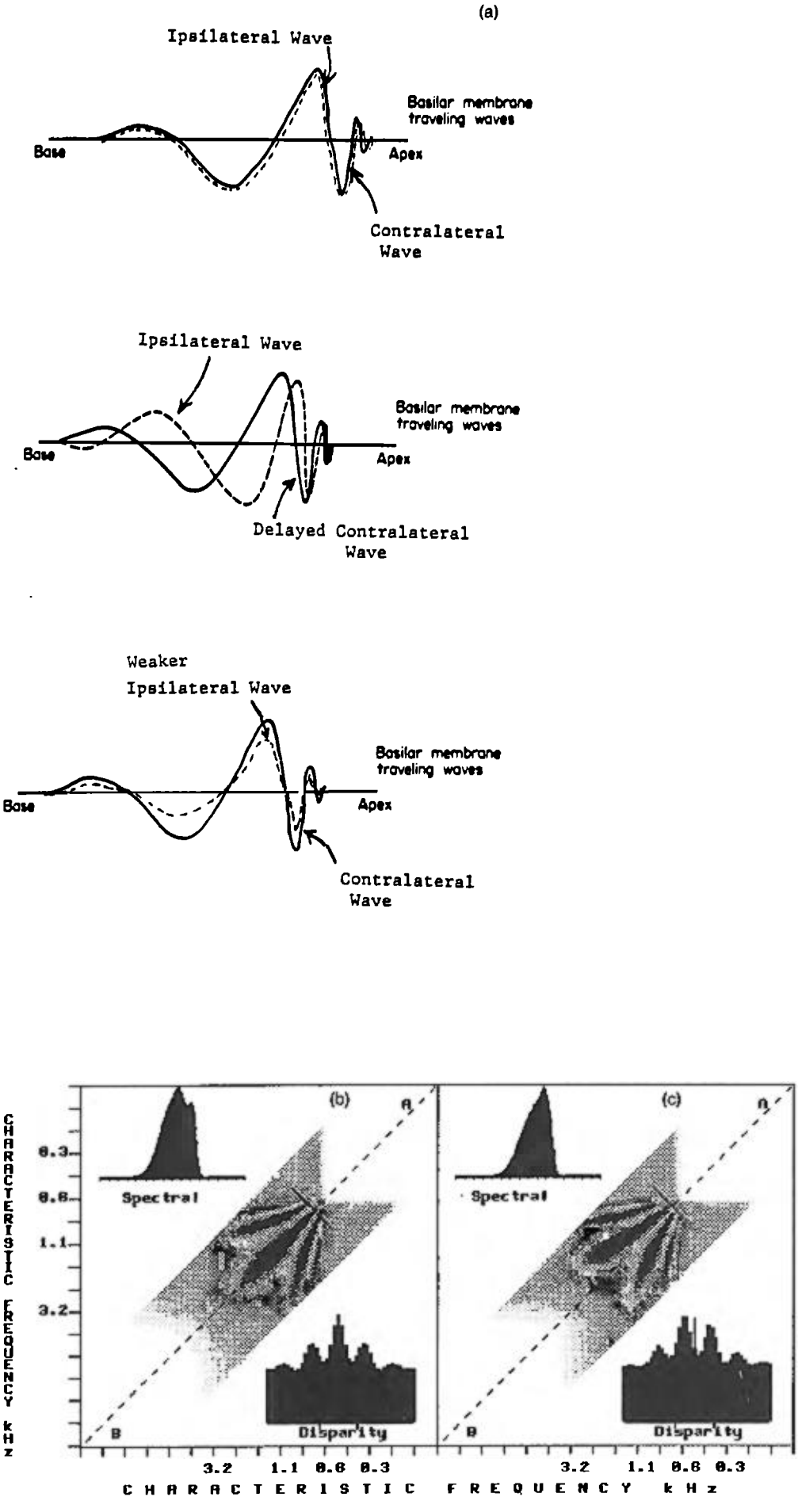


FIG. 4. Stereausis processing with a low-frequency tone. (a) A schematic of the traveling waves due to a centered tone (top), a binaurally delayed tone (middle), and a tone of binaurally unequal amplitudes (bottom). (b) The output of the stereausis network with a centered 600-Hz tone. The input patterns are shown in Fig. 2. The *disparity plot* shows a cross section of the two-dimensional patterns of activity computed near and along the bar shown. The total length of the cross section (or the x axis of the disparity plot) is 15 points (i.e., $i \quad j \leqslant 14$). When measured against the two cochlear axes, the CF disparity at either end of the bar is approximately 1 oct (please see also the Appendix, Sec. 1). Note that the top and bottom ends of the cross bar correspond to the left and right ends of the disparity plot, respectively. The activity represented in the disparity plot is computed by collapsing (diagonally) onto the bar a narrow band (12 points wide) of the activity below it. The *spectral plot* displays the activity near and along the diagonal AB (shown dashed). The left and right ends of the spectral plot correspond to the end points B and A on the diagonal, respectively (i.e., CFs increase from right to left). The divisions on the spectral plot are identical to those on the CF axes. (c) Same as (b) but with the 600-Hz tone binaurally delayed (ipsilateral side leads contralateral by $2\pi/3$ phase shift). The disparity plot shows clearly the resulting shift of the patterns.

operation to remove the negative outputs. The orientation and actual values of the mask weights used are given in the Appendix, Sec. 2.

(4) All the simulations discussed in this paper involve stationary signals (i.e., ongoing ITD and ILD cues). Thus, for the final display, $o_{ij}(n)$ frames over a 12.5-ms time interval (250 frames) are averaged and plotted together with the resulting disparity and spectral plots (in the insets).[8]

II. BINAURAL PROCESSING IN THE STEREAUSIS NETWORK

The responses to six classes of continuous sound stimuli (using ongoing cues) are illustrated below. The first five simulations emphasize lateralization tasks, while the last one deals with signal detection and enhancement in noisy environments.

A. Lateralization of low-frequency tones (≤1.5 kHz): Interaural-time delays (ITD)

Low-frequency tones can be lateralized through pure interaural delays (or phase shifts). Figure 5 illustrates how the stereausis network represents these percepts for an 1100-Hz tone at 0, $\pi/3$, $2\pi/3$, π phase shifts. In order to highlight the changes in the output patterns for different delays, cross sections along the disparity and spectral axes are also shown in the insets. For the centered tones, a dominant peak of activity appears along the AB diagonal (zero disparity). When a tone is binaurally delayed, the pattern shifts accordingly, and the relative height of the primary to secondary peaks decreases gradually. At π shift, the two peaks are equal and on either side of the midline. With further shifts, the previously secondary image moves further toward the center and now becomes the dominant peak. The periodic behavior of these patterns and the appearance of multiple *confusing* images at π phase shifts correspond closely to the lateralization of continuous low-frequency tones performed by human and animal subjects (Durlach and Colburn, 1978; Sayers, 1964).

The excursion of the primary peak along the *disparity plot* axis provides for an estimate of the maximum CF disparities needed in the tuning of the coincidence detecting cells of such a binaural network. Thus, from the π shift condition, the primary peak shifts approximately 1/5 of the entire axis on either side of the midline. Since the end points represent octave CF disparities [see Fig. 4(b) caption], then the cells needed to encode the primary peak should display at most 0.2-oct CF disparity, i.e., around 100–200 Hz at the CF location of 1 kHz. Similar estimates can be made based on considerations of the velocities of the traveling wave (as discussed in the last section).

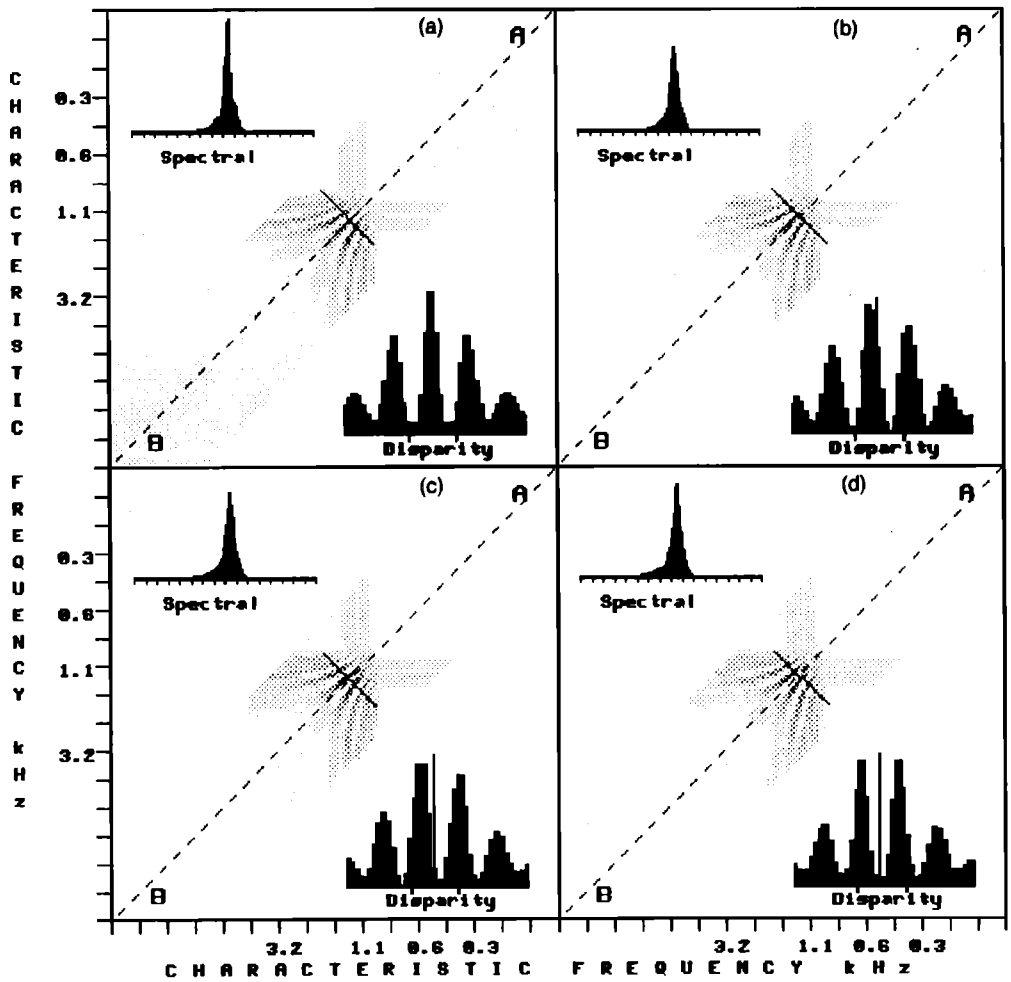


FIG. 5. Representation of ITDs with low-frequency tones (1100 Hz). (a) stereausis network outputs of a centered 1100-Hz tone. Same as Fig. 4(a) but patterns are sharpened with a lateral inhibitory mask. (b) Same as (a) but with tone $\pi/3$ phase-shifted (ipsilateral leading). (c) Same as (a) but with tone $2\pi/3$ phase-shifted (ipsilateral leading). (d) Same as (a) but with tone π phase-shifted (ipsilateral leading).

A basic difference between these results and those obtained from an equivalent simulation of the Jeffress model is the way in which the size of the secondary (ambiguous) peaks is reduced. In the stereausis network, this arises primarily from the narrow bandwidth of the cochlear filters. The filters limit the spatial extent (the envelope) of the spatially correlated instantaneous traveling wave patterns, which, in turn, causes the output (c_{ij}) to decay gradually away from the diagonal. In the Jeffress model [Fig. 1(a)], the secondary peaks are reduced by assuming (arbitrarily) that correlations away from the midline (i.e., with progressively increasing leads or lags) are less effective.

The stereausis network outputs exhibit two basic changes with higher frequency tones: (1) The spectral peaks (which remain stationary for different ITDs) move toward more basal CFs. (2) The disparity peaks become more closely spaced for higher frequency tones because of the steeper spatial phase functions of their traveling waves (Pfeiffer and Kim, 1975), or, equivalently, the smaller spacing between the peaks of their waves. The latter observation is clearly illustrated in the binaural outputs discussed later in Sec. II D. It can also be readily understood by considering the effects of reducing the distance D between the peaks of the synthetic patterns of Fig. 3(c) and (d).

This brings up the important question of how to interpret the magnitude of the shifts in the disparity plots in relation to the ITD. To answer this question, we first note that the amount of spatial shift that a given ITD induces in the binaural patterns of a single tone is proportional to the *velocity* of the underlying traveling wave (cf. similar remarks regarding the synthetic patterns in Sec. II B). Thus, for a particular ITD to induce the same spatial shifts at all CFs, *independent of frequency*, the velocity of the cochlear waves as a function of space must remain unchanged for all tones (except for a frequency-dependent spatial translation). Physiological data (Pfeiffer and Kim, 1975) and model simulations (see Sec. II D) suggest that this function (while remaining relatively unaltered in shape) exhibits a gradual elevation toward the base of the cochlea (Greenwood, 1988), with higher frequency tones having faster traveling waves near their CF. Consequently, a single ITD will produce larger spatial shifts for higher frequency tones. This point will be clarified further by the outputs of Sec. II D and E.

B. Lateralization of high-frequency tones: Interaural-level differences (ILD)

Phase locking in the responses of the mammalian auditory nerve deteriorates for frequencies beyond 1.5–2 kHz and little is preserved above about 3–4 kHz (Johnson, 1974). A high-frequency tone, therefore, evokes a response with a spatial profile reflecting the *envelope* of the traveling wave, but *not* its phase. Consequently, the stereausis network is insensitive to ITDs at these frequencies. An ILD, however, does create disparities between the amplitudes of the profiles of activity from the two ears. This, in turn, evokes a response pattern (o_{ij}) that is asymmetric with respect to the diagonal AB. The final network representation of this imbalance (following the lateral inhibitory mask) is somewhat different from, but still consistent with that due to the ITDs. Examples of the network responses with a 3-kHz tone are shown in Fig. 6. When the binaural inputs are identical (centered tone), a symmetric pattern of activity (with respect to AB) is evoked [Fig. 6(a)]. When projected onto the disparity axis, a centered peak emerges. Figure 6(d) illustrates the opposite extreme case of a nearly monotonic ipsilateral input. Here, only the weak horizontal ridge of activity (due to the ipsilateral tone) remains, intersecting the AB diagonal at the same CF as before. The disparity plot reflects this asymmetry by a general broadening and a lateral shift of the peak. In between, increasing the input ILDs is systematically reflected in the relative levels of the two ridges of activity (and, hence, in the broadening and shifting of the disparity peak). This broadening of the network outputs is reminiscent of the increased width of the perceived auditory event with increased ILDs reported in most psychoacoustical studies (Sayers, 1964).

The stereausis network can process ILD cues similarly with many other correlation operations such as $c_{ij} = g(x_i - y_j)$ and $c_{ij} = (x_i + x_0)(y_j + y_0)$ (x_0 and y_0 represent spontaneous input firing rates), but not with the pure multiplicative $c_{ij} = x_i y_j$ correlation. The underlying reason why most correlation operations work is the consistent *asymmetrical* change of the *stereausis* network outputs c_{ij} around the AB diagonal as a function of the ILDs (see footnote 9). In contrast to the *stereausis* network layout, the computations along the disparity axis in the *Jeffress* model are inherently symmetric with respect to the diagonal. This is because all coincidence detectors at a given CF [e.g., like those shown in Fig. 1(a)] receive exactly the same input pair (except for time delays which are irrelevant here). Thus, in order to produce an ILD-dependent asymmetry in the Jeffress outputs, additional asymmetries are necessary, such as a graded threshold or inhibition along the disparity axis from one or both inputs.

Finally, it is evident from the disparity plots of Fig. 6 that the ILD lateralization derived in this network is less accurate than that due to the ITDs. This is because the stereausis network architecture, with its regular arrays of *horizontal* disparity detectors, is fundamentally suited to detect and display accurately ITD disparities, that is, disparities due to lateral shifts in the cochlear patterns. A pure ILD, instead, creates binaural cochlear patterns that approximately differ *only* in their relative amplitudes [Fig. 4(a), bottom]. Consequently, such *vertical* disparities between the input patterns are not sharply detected by any one cell in the network; rather, their influence on the outputs is more broadly distributed. This may be sufficient to account for human ILD detection, although slightly different networks can be designed to be specially sensitive to vertical disparities (Sullivan and Konishi, 1984).

C. Time/level trading for low-frequency tones

For low-frequency tones, both phase and amplitude disparities can be preserved in the responses of the auditory nerve, and, hence, detected by the binaural network. Therefore, both ITD and ILD cues can influence the lateralization of the stimulus. Figure 7 illustrates the effects on the

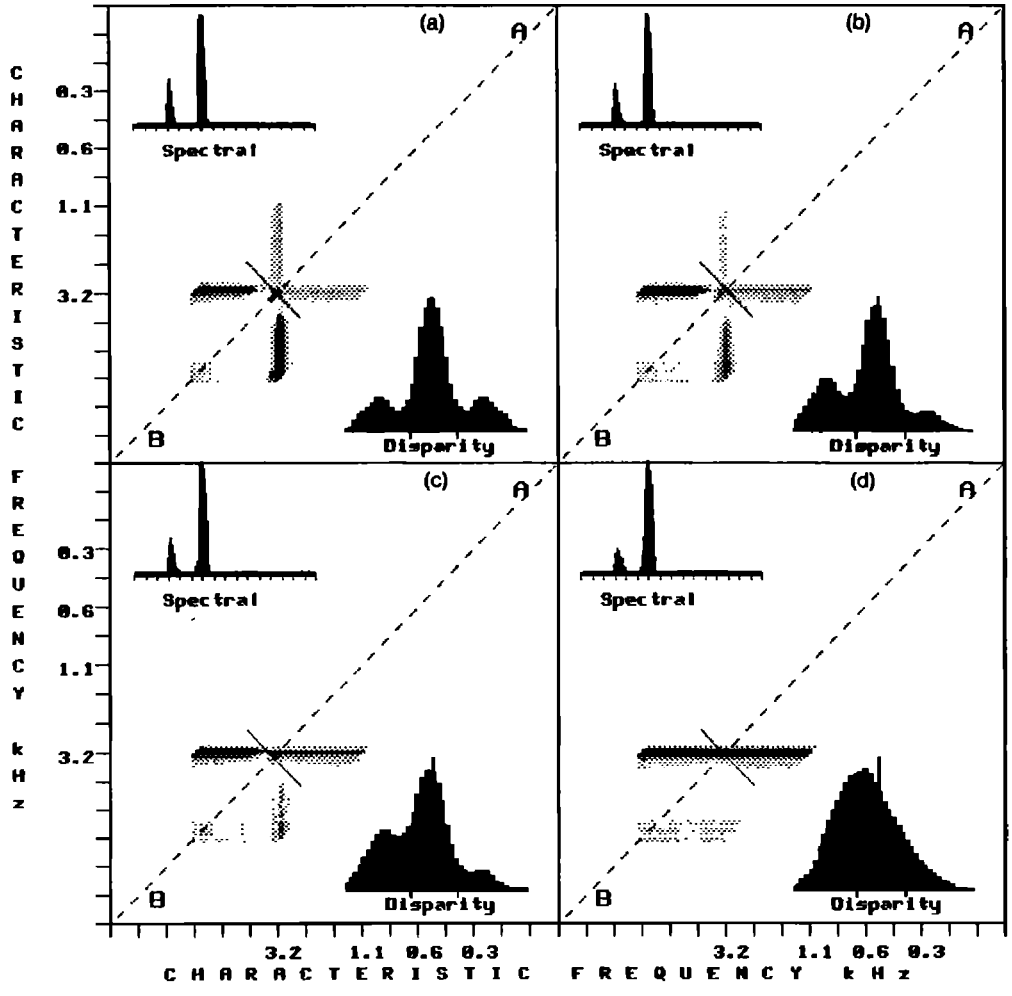


FIG. 6. Representation of ILDs with high-frequency tones (3 kHz). (a) Stereausis network outputs of a centered 3-kHz tone (ILD = 0 dB). (b) Same as (a) but with ILD = 3 dB (ipsilateral higher). (c) Same as (a) but with ILD = 6 dB (ipsilateral higher). (d) Same as (a) but with ILD = 12 dB (ipsilateral higher).

network output of increasing the ILD of a centered tone (600 Hz). There are two important regions of activity along the disparity axis: (1) the centered primary peak whose location remains relatively fixed reflecting the zero ITD of the stimulus and (2) the secondary peak, which grows relatively in height and broadens with the increase in the ILD. Two interpretations of these results are possible. The first is to view the entire pattern as a single broad auditory event with a center of gravity that is gradually lateralized as the relative height of the secondary peak increases. The second interpretation, which is most often reported by experienced subjects in similar psychoacoustical experiments (Whitworth and Jeffress, 1961), is that of *two* auditory percepts—one remains in the middle of the head (presumably associated with the primary peak), while the other migrates to the side becoming more spatially blurred (reflecting the center of gravity).

The reason for the stability of the primary disparity peak and the sequence of ILD-induced changes in Fig. 7 can be readily understood in the context of the previous two examples. Thus, as seen in the network outputs of Fig. 5, a particular ITD establishes a characteristic shift (disparity) and relative peak sizes in the patterns. An added ILD primarily causes an imbalance in the outputs *surrounding* the primary peak, which now acts as the anchor point (see Fig. 6). The secondary (side) peaks, consequently, undergo asymmetrical growth and broadening. This behavior of the disparity plots is very similar to the effects produced ILD-dependent weighting of cross correlograms often proposed to augment the ITD sensitivity of correlation-based binaural models (Stern and Colburn, 1978).

D. Lateralization of speech and harmonic complex sounds

The representation and lateralization of complex signals are essentially an extension of the case of single tones above in that the disparities in the phase and amplitude of the complex cochlear responses are processed and represented similarly by the stereausis network. Consider first the responses to the in-phase harmonic series: $\Sigma_{n=1}^{10} \sin(2\pi 300\, nt)$ (Fig. 8). Two spectral regions can be qualitatively distinguished: (1) the region of lower frequency resolved harmonics (300, 600 Hz), and (2) the higher spectrally unresolved harmonics. In the first region, the responses to each harmonic are separated from those of its neighbors and resemble closely the patterns seen earlier for single tones. Thus, for a centered complex, all the spectral peaks lie along the AB diagonal [Fig. 8(a)]. When an ITD (250 μs) is introduced, the patterns due to the different harmonics shift (in the same direction) by an amount that slightly increases with the harmonic number [Fig. 8(b)]. The increasing shift simply reflects the

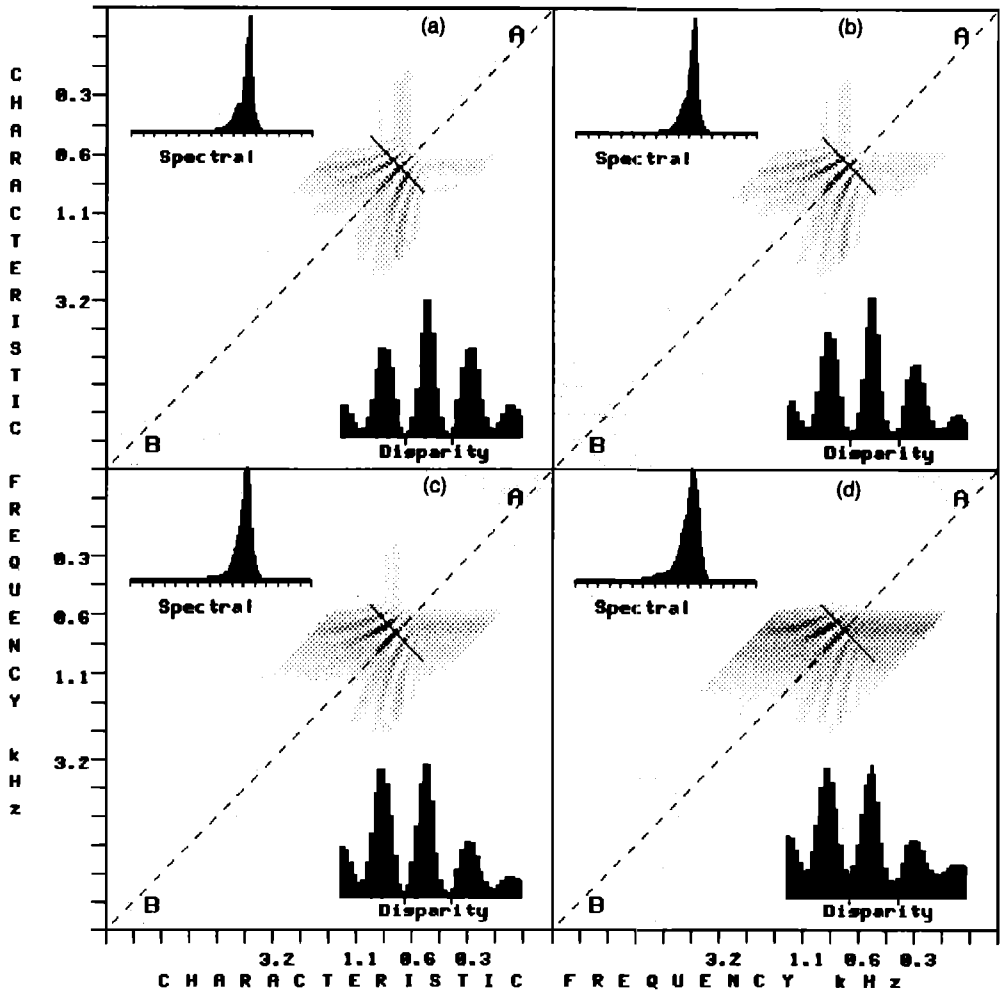


FIG. 7. Effect of ILDs on low-frequency tones. (a) Stereausis network outputs with a centered 600-Hz tone (ILD = 0 dB). (b) Same as (a) but with ILD = 3 dB (ipsilateral higher). (c) Same as (a) but with ILD = 6 dB (ipsilateral higher). (d) Same as (a) but with ILD = 12 dB (ipsilateral higher). The figures here are normalized individually.

higher velocities of the more basal traveling waves. Note, also, the decreasing separation between the primary and secondary peaks of the higher harmonics in this region. In the second (basal) region, the temporal structure of the responses to the unresolved harmonics produces binaural patterns with similarly increasing spatial shifts.

Figure 8(c) and (d) illustrates the responses to a centered and a lateralized naturally spoken speech stimulus (the vowel /a/). They display similar shifts around the most prominent and partially resolved harmonics (near the vowel's first and second formants at approximately 900 and 2000 Hz). *For didactic purposes*, the disparity plots in all Fig. 8 insets are formed by collapsing the entire stereausis plane. The plots illustrate the overall lateral shifts of the patterns.

The details of the lateralized binaural outputs of Fig. 8 illustrate an important point regarding the effect of the changing traveling wave velocities (and hence varying spatial shifts) on the representation of a given ITD in the stereausis network. For any given ITD [for example, ITD = 250-μs ipsilateral leading in Fig. 8(b)], there is an *arc* of coincident detector neurons that are maximally activated. Regardless of the exact nature of the stimulus [single tones, speech, harmonic series, or noise (Sec. II E)], members of this arc of cells are activated whenever the *characteristic* ITD (250 μs) is detected at their location (or frequency). Similarly, adjacent arcs correspond to different ITDs [e.g., the center arc (diagonal) corresponds to ITD = 0 μs]. In this manner, an organized spatial map corresponding to absolute lateral positions in space (different ITDs) can be formed, presumably through experience during development.[10] Therefore, the only apparent effect of the changing velocities is to bend the *iso–ITD* planes away from the diagonal at the higher CFs.

Finally, the stereausis outputs for the higher frequency, unresolved harmonics, region of Fig. 8(a) and (b) implicitly demonstrate the lateralization of the general class of amplitude-modulated high-frequency signals (> 1.5 kHz). In this case, temporal phase locking in the cochlear spatiotemporal responses mostly reflects the envelope of the signal rather than the individual (unresolved) *carrier* components. Consequently, any interaural manipulations of the *envelope*, such as inserting ITDs or ILDs, will generate at the stereausis network outputs similar to those observed for single tones earlier, except for being located near the CF of the carrier frequencies. These remarks apply equally to the special case of *transient sounds or cues*, where the useful temporal structure exists mostly near the rapidly changing portions of the stimulus (for instance, at onsets and offsets).[11]

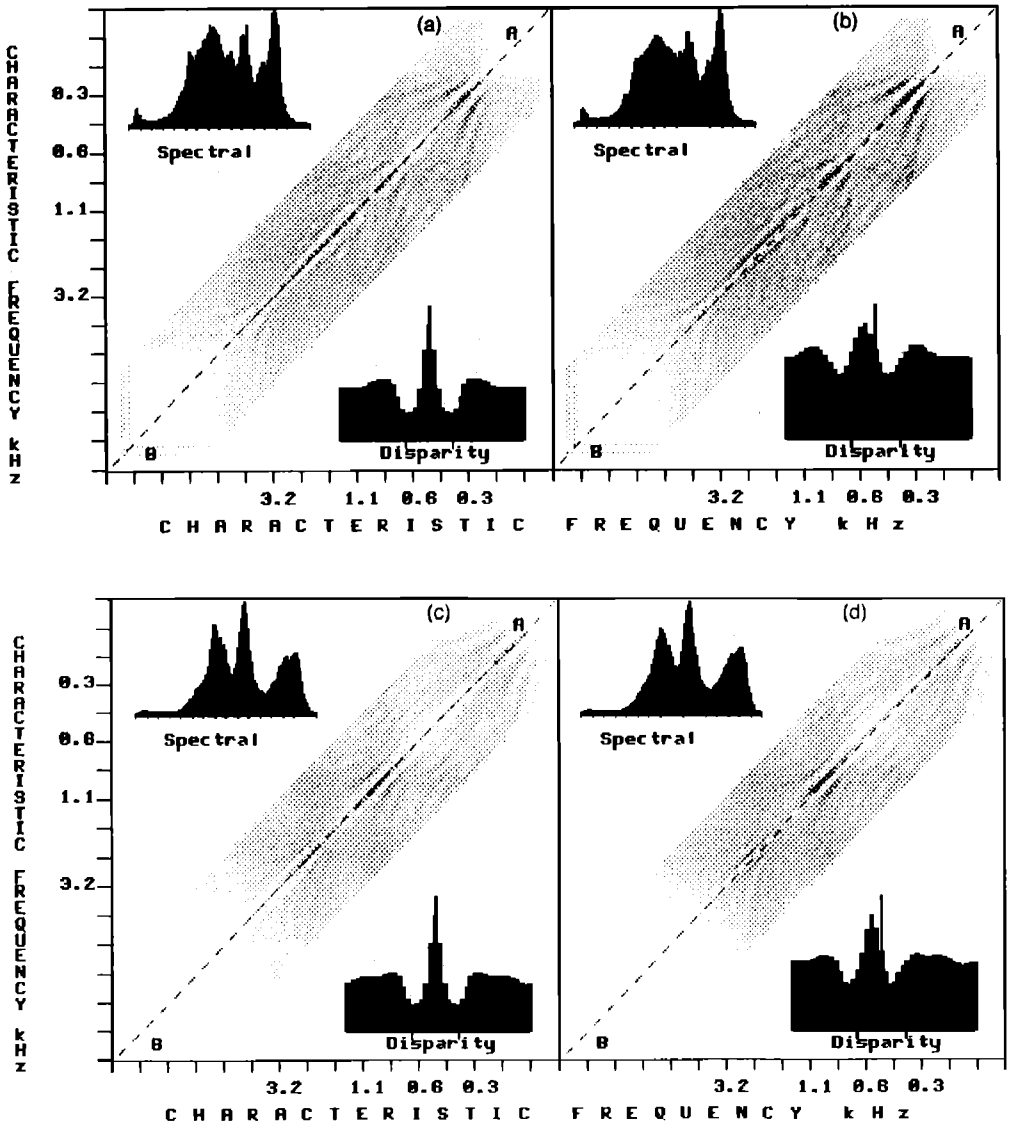


FIG. 8. Lateralization of speech and complex harmonic series. (a) Stereausis outputs for a centered harmonic series: $\Sigma_{n=1}^{10}\sin(2\pi \times 300nt)$. Disparity plot represents the collapse of the entire plane (see text). (b) Same as (a) but with ITD = 250 μs. (c) Stereausis outputs for the vowel |a| centered. (d) Same as (c) but with ITD = 250 μs.

E. Lateralization of broadband noise

The principles underlying the lateralization of noise in the stereausis network (Fig. 9) are essentially identical to those of other complex sounds (Sec. II D). This is somewhat harder to see initially because of the subtlety of the *spatial* coherence in the temporal structure of the responses to noise. The coherence (or correlation) among the responses of adjacent (local) fibers exists because of the finite (relatively broad) bandwidth and extensive spatial overlap of the cochlear filters, and is demonstrated in Fig. 9(a) by the relatively gradual change in the responses from one channel to another.

The stereausis outputs for a centered noise stimulus are shown in Fig. 9(b). The peak correlation occurs as before along the center diagonal. The correlation decreases monotonically away from the diagonal, and no secondary peaks emerge because of the absence of a pseudoperiodic spatial fine structure, as seen in the cochlear patterns of single tones [Fig. 5(a)]. When an ITD is introduced, the peak of correlation shifts accordingly [Fig. 9(c)], and the amount of shift increases at the higher CFs just as in earlier simulations, and for exactly the same reasons. This peak persists, but gradually diminishes in size, with increasing ITDs. It will vanish at a particular CF when the spatial shift exceeds the width of the cochlear filter at that CF. This is, of course, the same reason for the smaller secondary peaks in the case of single tones (Fig. 5).

F. Detection and enhancement of tones in noise

One of the well-recognized functions of binaural hearing is the vast improvements it affords in perceiving a particular signal in complex acoustic environments of many sound sources [the so-called "cocktail party" effect (Cherry, 1953)]. Extensive psychoacoustical investigations of this problem have been carried out using simple stimuli with well-defined auditory tasks, for example in the detection of single low-frequency tones in noise backgrounds of varying degrees of binaural coherence (Durlach and Colburn, 1978; Hirsch, 1948). In Fig. 10, we illustrate the processing and

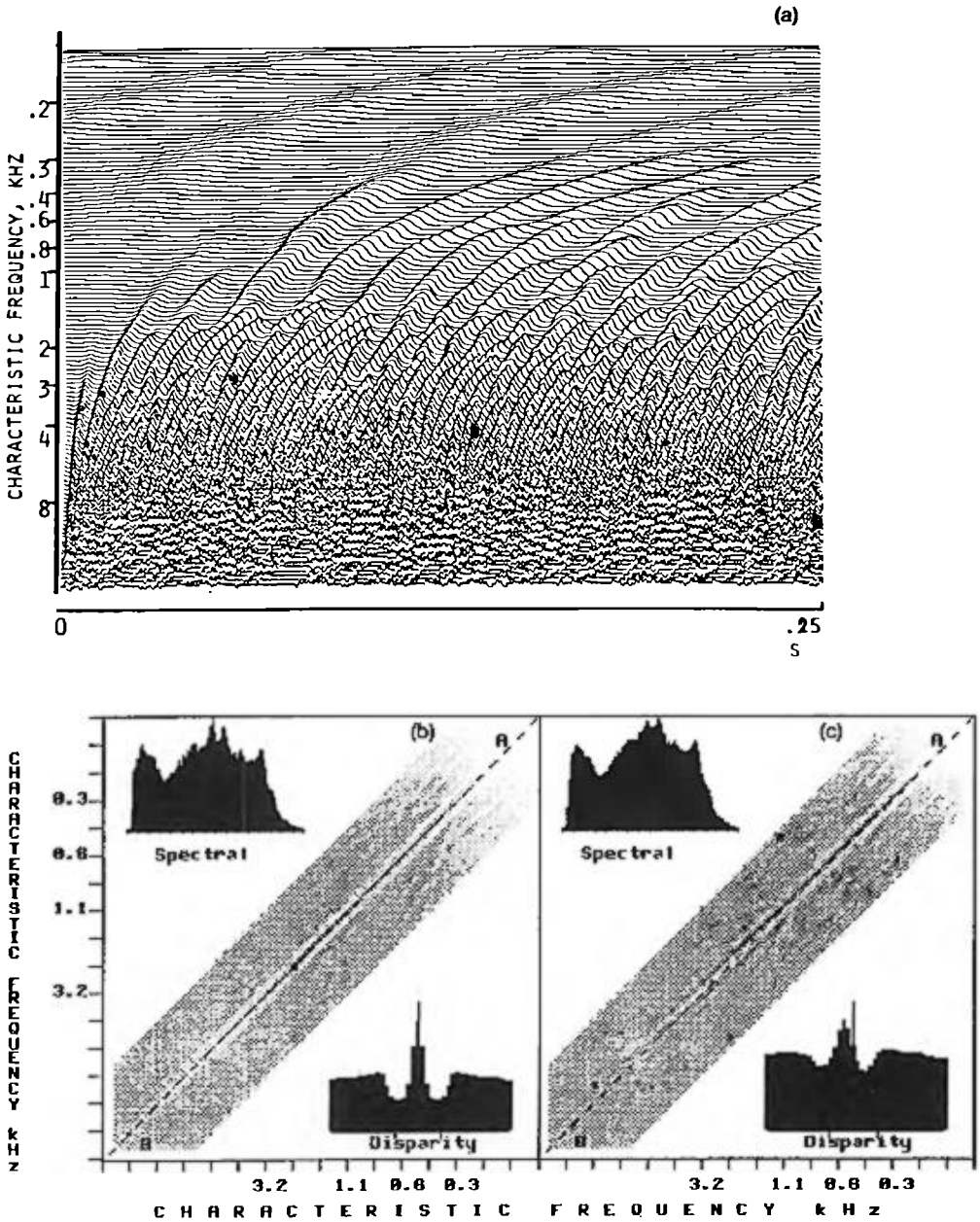


FIG. 9. Lateralization of broadband noise. (a) The cochlear response patterns to the broadband noise stimulus (0.1 10 kHz). (b) Stereausis network outputs with the broadband noise centered. (c) Same as (b) but with noise binaurally delayed (250 μs).

the nature of the outputs of the stereausis network in such stimulus conditions. First, in Fig. 10(a) (S_0N_0 case), a centered tone (1100 Hz) in binaurally coherent noise (bandwidth $= 0.1$–10 kHz; $S/N = -10$ dB) evokes, as expected of binaurally identical signals, its maximal responses along the AB diagonal. Because of the low signal-to-noise (S/N) ratio, considerable noise background accompanies the peak due to the tone in the diagonal spectral slice shown in the inset. When the tone *or* the noise is π reversed interaurally, a dramatic improvement in the detection of the signal occurs psychoacoustically [10–15 dB relative to S_0N_0 case, around 200 Hz, and 5–10 dB near 1 kHz (Blauert, 1983)]. The representation of these signals in the binaural network [Fig. 10(b) and (c)] demonstrates the clear separation of the two components of the complex along both the disparity and spectral axes. Thus, in Fig. 10(b) ($S_\pi N_0$ case), reversing the tone (interaural π phase shift) causes its binaural output to shift off the noisy diagonal AB, and, hence, to stand out as a separate component. The opposite situation is shown in Fig. 10(c) (S_0N_π case) where reversing the noise reduces drastically its output activity along the AB diagonal, distributing it instead, in part, randomly over the network, and, in part, along the π phase-shift diagonals. Therefore, the representation of the centered tone is dramatically enhanced on the diagonal spectral slice (AB). These results can be readily extended to explain many other binaural stimulus conditions, such as partially coherent noise (Durlach and Colburn, 1978).

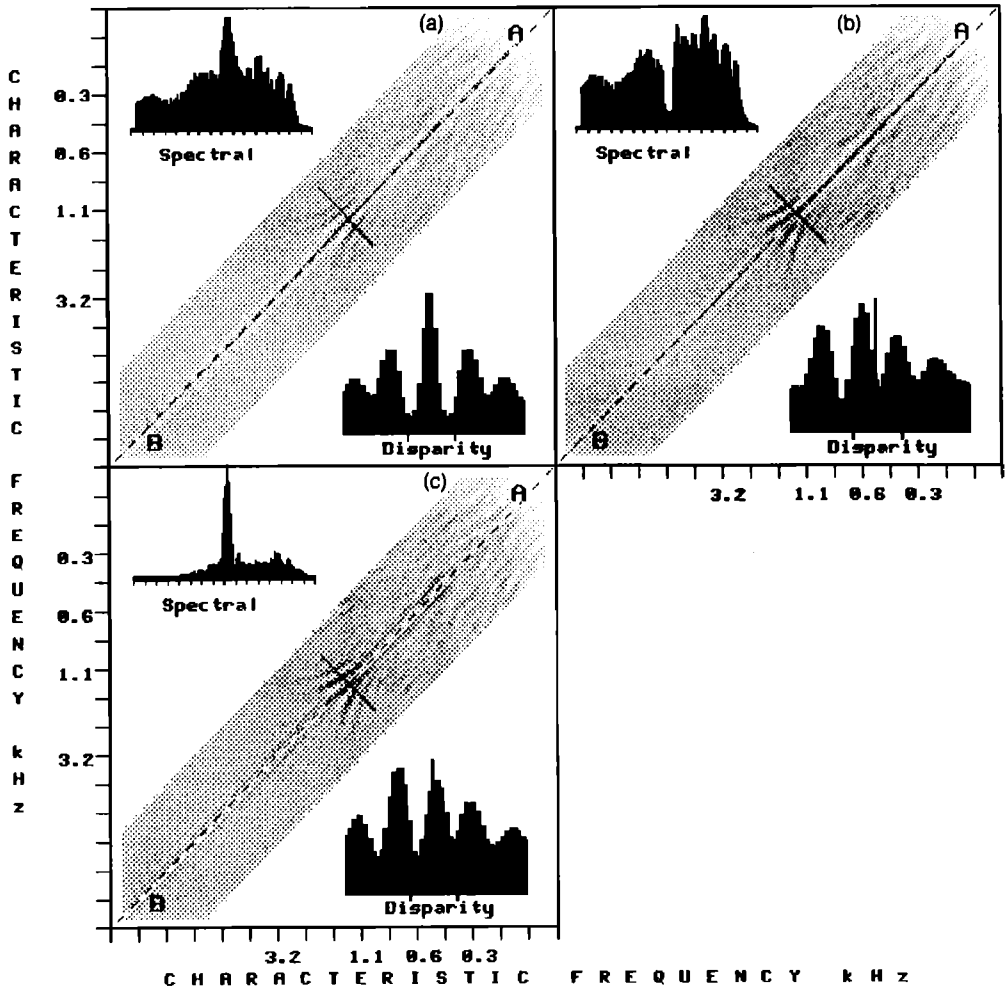


FIG. 10. Detection and enhancement of tones in noise. (a) Stereausis network outputs with binaurally identical tone (1100 Hz) in noise background (0.1–10 kHz) and S/N = − 10 dB. (b) Same as (a) but with signal reversed in amplitude ($S_\pi N_0$). (c) Same as (a) but with noise reversed in amplitude ($S_0 N_\pi$).

III. DISCUSSION

A simple neural network is proposed as the central binaural processor in the mammalian auditory system. It combines the spatiotemporal responses arriving from the two cochleas via the tonotopically organized pathway of the auditory nerve and AVCN and generates output patterns that reflect various attributes of spatial hearing. The network detects and uses all perceptually significant binaural cues (ITDs and ILDs) by applying exactly the same principle of operation—that is, to compute systematically a measure of the *spatial* correlation (or disparity) between the instantaneous patterns of activity from the two cochlea. An important and distinguishing property of this network is that *no* neural delays are required to perform its function.

In the following two sections, we shall elaborate on two fundamental issues arising from the stereausis hypothesis: (1) the functional consequences of any neural delays in the binaural auditory system, and (2) the role of space and time in auditory and other sensory processing.

A. Distinguishing between the stereausis network and the Jeffress model: Physiological considerations

The fundamental conceptual difference between the stereausis network and the Jeffress model lies in the use of *spatial* versus *temporal* correlations to extract the binaural cues. The major implication of this distinction—one that simultaneously relates and distinguishes the two models apart—concerns the functional role of *neural delays*. Neural delays are an inevitable occurrence in any biological network where information is transmitted from one point to another. The relevance of such delays in the two models represents two extremes: It is pivotal to the Jeffress model while completely ignored in the stereausis network. In reality, however, it is most likely that an intermediate view exists, and that the above models represent two basic idealizations through which the function of the biological network can be understood.

To elaborate on these statements, consider the two schematic networks of Fig. 1. In both models, the basic computational element is the binaural "coincidence detector" cell whose responses approximate a "cross correlation" of its input fibers (in agreement with much of the available neurophysiological evidence [see Irvine (1986), Sullivan and Konishi (1986), Yin and Kuwada (1984) for a review]. An obvious property of the stereausis network [Fig. 1(b)] is the misalignment in the CFs of its inputs. Such misalignment must provide for the maximum interaural delays detectable by the species (e.g., $\approx 700\,\mu$s in man, $\approx 400\,\mu$s in cats, and $\approx 200\,\mu$s in the barn owl). The required misalignments,

however, are rather small since the traveling waves are slow near their point of resonance and can, therefore, provide large delays over short distances. For instance, considering a typical velocity of 3 mm/ms for the traveling wave of a 1.2-kHz tone,[12] a spatial shift of 1 mm can provide for more than 300 μs of basilar membrane-originated delays. Consequently, the spatial mismatch necessary for the inputs to the stereausis network cells encoding the largest ITDs corresponds to CF disparities of the order of 200 Hz around 1 kHz in the cat (see, also, estimates derived in Sec. II A). Such maximal shifts are well within the kind of mismatches typically observed in experiments [see Fig. 13 in Kuwada *et al.* (1984)]. It should be emphasized that these mismatches represent only a minority of the binaural cells at the extremes of the networks; most cells are likely to exhibit smaller CF disparities being closer to the midline. These estimates also suggest that for a strictly Jeffress-like model to avoid these mismatches, the spatial (CF) alignments of the inputs need to be very tightly controlled to within a few tens of micrometers; otherwise, basilar membrane delays will overwhelm the relatively smaller neural delays.

The opposite side of these arguments concerns the influence of any neural delays on the operation of the stereausis network. There are conceptually two kinds of delays: The first kinds are due to the unequal effective lengths of the two input pathways leading to the network (e.g., synaptic and transmission delays in the auditory nerve and AVCN), and the second kinds are due to distributed axonal delays *within* the binaural network—the kind of delays invoked in the Jeffress-type models. In the first case, mismatched transmission delays cause an overall translation of the binaural response patterns away from the center diagonal. In the second case, introducing distributed delays along the lines (axons) between the coincidence cells of the stereausis network [Fig. 1(b)] simply "stretches" the patterns along their disparity axis, i.e., providing more spatial resolution.[13] This can be most readily seen in the case of the simple synthetic patterns of Fig. 3.

The conclusion that arises from the above arguments is that the effects of axonal delays and basilar membrane-originated delays are complementary or approximately additive, and that it is likely that a biological realization of the stereausis network, the amount of CF overlap reflects the balance between the use of these sources of delays. The fine tuning of the final map would presumably be achieved through experience during development. Futhermore, it seems possible to evaluate experimentally the relative dominance of these different types of delays (i.e., whether the MSO is more like a Jeffress model performing *temporal* correlations or a stereausis network performing *spatial* correlations) by observing the effects on the binaural map [e.g., the spatial maps of the barn owl (Sullivan and Konishi, 1986)] of eliminating one source or the other, for instance by electrically stimulating the auditory nerve. It will be difficult, however, to distinguish the two networks psychoacoustically since the temporal and spatial dimensions of the cochlear responses are intimately coupled, and, hence, manipulating one almost always influences the other.

B. The role of space and time in auditory processing

Unlike the visual and somatosensory systems, the role of the spatial axis in the representation and processing of sound in the auditory system has always been elusive. This is primarily due to the temporal character of the stimulus, the apparent temporal specializations of many auditory synapses and structures, and the success of numerous temporally based algorithms in accounting for most auditory percepts. As mentioned earlier, this has led to the conclusion that profound conceptual differences must exist between the neural networks that underly auditory and visual processing. Instead, we shall argue below that a unified spatially oriented framework for the auditory and visual systems does exist, and that the basic functions of early auditory processing of sound can be achieved with computational algorithms and neural networks that are essentially similar to those commonly used in early vision. Specifically, two fundamental principles of visual processing can be invoked for binaural and monaural processing: disparity detection (for depth perception) and lateral inhibition (for edge detection).

It has long been appreciated that binaural hearing is analogous to binocular vision in endowing perception with an extra spatial dimension based on disparity measures in

TABLE I. Comparison of monaural and binaural processing.

	Monaural	Binaural
Objectives		
operation	stimulus spectrum estimation	ITD and ILD detection
function	sound recognition	localization and signal enhancement
Temporally based processing		
algorithm	auto-correlation (Fourier analysis)	cross-correlation
neural implementation	delay lines	delay lines
Spatially based processing		
algorithm	local cross-fiber correlations (detection of discontinuities in monaural responses)	local cross-fiber correlations (detection of disparities between binaural responses)
neural implementation	lateral inhibitory network (LIN)	stereausis network
Corresponding operation in vision	edge detection (retinal on-center/off-surround)	depth perception (stereopsis networks)

the stimulus projection upon the sensory organs (Yin and Kuwada, 1984). In most stereopsis algorithms (Marr and Poggio, 1979), depth information is derived from the disparities of the spatial images of the same object on the retinas. The network derives these cues using a simple ordered array of disparity detector cells that correlate the binocular images at various horizontal shifts. Depending on the original mismatch of the input images, correspondingly different cells will fire maximally, thus extracting and spatially encoding the depth cue. Binaural processing in the stereausis network is very similar in that the network detects the interaural delays based on the disparities between the two instantaneous spatial cochlear images.

There are striking parallels between the motivations and subsequent developments of the binaural stereausis network and the lateral inhibitory network proposed earlier for the monaural processing of auditory-nerve responses (Shamma, 1985a) (Table I). Briefly, the LIN was first proposed to provide a biologically realistic network capable of utilizing the phase-locked (temporal) information on the auditory nerve. A basic operation of most earlier algorithms was to measure the *absolute* periodicity of the responses on a given fiber using some form of Fourier analysis (Seneff, 1984; Sinex and Geisler, 1983; Young and Sachs, 1979), or by its temporal equivalent operation—estimating the *autocorrelation* function. In order to implement the latter operation biologically, a series of delay lines was often postulated (Delgutte, 1984). The LIN, instead, derived its spectral estimates by a simple local comparison (correlation) of the responses across different fibers, hence, detecting spatial disparities in the responses. In effect, the LIN detected edges in the response patterns regardless of their temporal or amplitude (average rate) origin, a well-understood operation in the vision literature (Hartline, 1974).

The auditory and visual systems, however, differ significantly in the "means" for expressing the spatial features that their monaural (monocular) and binaural (binocular) networks might detect and process. In the auditory system, temporal phase locking in the responses of the auditory nerve and the AVCN fundamentally serves as the *carrier* of the spatial cues to the CNS. Without phase locking, the detailed structure of the basilar membrane traveling wave, and, hence, the edge and the relative disparity cues, will not be preserved and conveyed to the central monaural and binaural processors. It is in this light that one may interpret the significance of the "temporal" specializations that abound in the early pathways and nuclei of the auditory system (e.g., the extremely rapid synapses of the *bushy cells* of the AVCN).

The experimental data available from the auditory system are consistent with the presence of *spatially based* networks, but still fall far short of an unequivocal evidence. Spatial algorithms are *not* necessarily more accurate than temporal algorithms in describing stimulus–percept relations in the auditory system (especially if sufficient number of free parameters are allowed). Furthermore, they are often more awkward to formulate mathematically compared to the convenient forms usually chosen for the temporal algorithms, such as the correlation and Fourier analysis operations. Nevertheless, the appeal of auditory spatial algorithms stems from their immediate interpretation as well understood biological networks of the CNS, such as the lateral inhibitory networks of the Limulus (Hartline, 1974), and the disparity detectors in the cat (Poggio, 1984).

ACKNOWLEDGMENTS

This work is partially funded through an NSF Grant (EET-8716099), NSF (CDR-85-00108), the University of Maryland Institute for Advanced Computer Studies, and by the Air Force Office of Scientific Research. The authors wish to thank Dr. Steve Colburn, Dr. Mark Konishi, and an anonymous reviewer for their reviews and criticisms.

APPENDIX

1. The cochlear model

The cochlear spatiotemporal patterns are computed using digital algorithms based on a detailed multistage biophysical model of the guinea pig cochlea (Holmes and Cole, 1984; Shamma *et al.*, 1986). At each of 128 locations along the cochlear partition, the transfer function of the basilar membrane is computed and used in an FFT-based overlap-and-add method to generate the membrane's response to the stimulus. This output is then high-pass filtered ($w_n = u_n - 0.8\, u_{n-1}$; modeling both outer ear and fluid–cilia coupling stages) and compressed by a sigmoidal function of the form: $x = M \quad 1/(1 + be^{-aw})$, where $a(= 0.00044)$, $b(= 0.111)$, and M are parameters of the nonlinearity, and x, w are the output and input, respectively. Finally, a low-pass filter smooths the output (time constant $= 0.1$ ms). The parameters of the compressive nonlinearity are chosen such that approximately 30 dB of linear gain is available between threshold and saturation (defined as 0.1–0.9 of maximum output level M) and that the output is saturated at moderate sound levels (approximately 60 dB SPL). For all simulations shown in this paper, the input signals were such that the cochlear outputs were just below saturation at the maximally activated channels. Each fiber (channel) is labeled by a characteristic frequency. The CF of the fiber is defined here as the frequency of the tone whose *peak* activity at the output of the stereausis network is located spatially at this fiber.

2. The computations of the stereausis network

The computations in the stereausis network are based on a matrix of correlationlike operations applied to the responses of the two ordered arrays of cochlear fibers. The spatiotemporal outputs of the cochlea are generated using the cochlear model described above in Sec. I 1. In order to relate the processing of the stereausis network to other correlation-based models, we consider first the following correlation operation $[C(x_i, y_j)]$ performed at each node of the network:

$$c_{ij} = C(x_i, y_j) = \int_T x_i(t) y_j(t)\, dt, \qquad \text{(A1)}$$

where $x_i(t)$ and $y_j(t)$ are the response of the ith ipsilateral and jth contralateral fibers, respectively, and T is the period

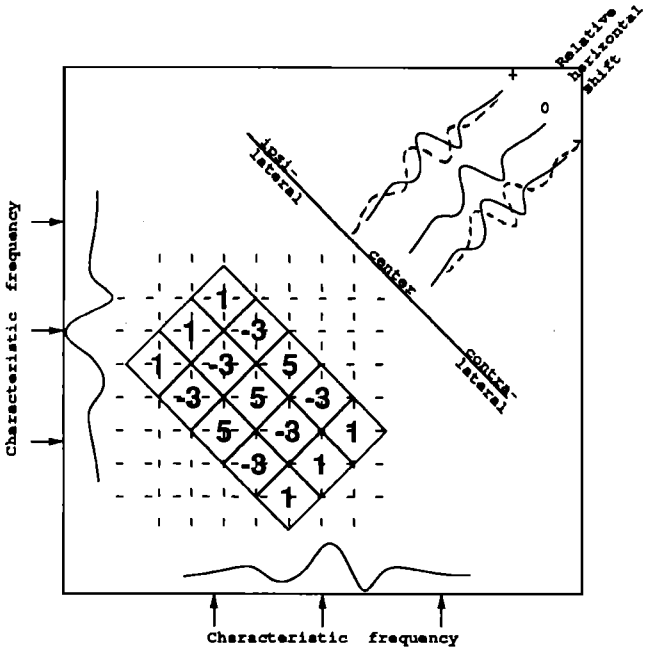


FIG. A1. The two-dimensional mask of the elongated on-center/off-surround network.

of integration. No neuronal delays are inserted in the two pathways here; however, since they originate at different CF locations along the cochlear partition, delays due to the finite velocity of the basilar membrane traveling wave are already incorporated in their responses. For instance, let x_i and y_j be the basilar membrane responses to a centered low-frequency (ω) tone; i.e.,

$$x_i(t) = A_i(\omega)\sin[\omega t + \theta_i(\omega)] \tag{A2}$$

and

$$y_j(t) = A_j(\omega)\sin[\omega t + \theta_j(\omega)], \tag{A3}$$

where $A_i(\omega)$, $A_j(\omega)$ and $\theta_i(\omega)$, $\theta_j(\omega)$ are the amplitudes and phases of the traveling waves at the ith and jth locations of the two cochleas. If i and j have close CFs [i.e., $A_i(\omega) \approx A_j(\omega)$], and if we define $\theta_j(\omega) = \theta_i(\omega) - \delta\theta(\omega)$, then,

$$y_j(t) \approx A_i(\omega)\sin[\omega t + \theta_i(\omega) - \delta\theta(\omega)]. \tag{A4}$$

If we assume now that the velocity of the traveling wave (v) over the small distance (δs) between the ith and jth locations is approximately constant, then the spatial frequency (ω_s) of the traveling wave at i can be related to the frequency of the tone as

$$\omega_s = \omega/v \tag{A5}$$

and, hence,

$$\delta\theta(\omega) \approx \omega_s\,\delta s = (\omega/v)\delta s = \omega\tau_s \tag{A6}$$

where τ_s is the time it takes the wave to travel the small distance (δs) between i and j. Therefore, we may rewrite $y_j(t)$ as

$$y_j(t) \approx A_i(\omega)\sin(\omega t + \theta_i - \omega\tau_s) = y_i(t - \tau_s), \tag{A7}$$

i.e., a delayed version of $y_i(t)$. The correlation operation $C(x_i, y_j)$ defined above therefore becomes

$$c_{ij} = \int_T x_i(t)y_i(t - \tau_s)dt, \tag{A8}$$

which is exactly equivalent to the usual correlation operations hypothesized using neuronal delay lines. This result can be generalized for other stimuli.

In all the examples shown in Sec. II, the outputs c_{ij} are computed within a stripe surrounding the diagonal AB($/i - j/ \leqslant 30$). At each time instant, a two-dimensional image [$c_{ij} = (x_i + y_j)^2$] is produced and sharpened by an elongated on-center/off-surround lateral inhibitory network to generate (o_{ij}):

$$o_{ij} = g\left(\sum_{kl} w_{ijkl}c_{kl}\right), \tag{A9}$$

where the network connectivities w_{ijkl} around each (i, j) neuron are identical (a uniform network), and $g(.) = \max(.,0)$ represents a thresholding operation to remove the negative outputs. The specific mask used in the computations of this paper is schematically illustrated in Fig. A1 [cf. Fig. 1(b)]. The parameters of this mask are not critical for the results discussed in this report. Its purpose is simply to sharpen the c_{ij} outputs by attenuating the low spatial frequencies at the tails of the binaural patterns (compare Figs. 4 and 5). Although the mask is applied here in a nonrecursive (feedforward) manner (to speed up the computations), it may also be applied in a recursive (feedback) configuration [see Shamma (1988) for an example]. For the purposes of this paper, both forms produce similar results. For the final displays in Sec. II, the outputs o_{ij} for 250 time samples (12.5 ms) are averaged over the entire duration.

[1]Of course, neuronal delays inevitably occur in the transmission of signals from one point to another in the nervous system. The influence of such delays on the function of the binaural networks is considered in Sec. III.

[2]Significant additions and assumptions are usually needed to adapt the Jeffress model to the detection and representation of ILD cues (Blauert, 1983).

[3]Because of hair cell transfer characteristics, the traveling wave image undergoes several modifications, which include rectification, saturation, and low-pass filtering (see part 1 of the Appendix). The effect of all these factors is taken into account in the analysis presented in this paper.

[4]Note that since the *envelope* of the traveling wave remains unaltered when a pure ITD is introduced, the corresponding spatial disparity does not precisely correspond to a relative horizontal shift in the waves (as was the case earlier for the synthetic patterns), but rather to a slightly distorted shift in the fine structure (within the envelope) of the wave.

[5]Clearly, the *disparity plot* and the *spectral plot* are used here primarily for didactic purposes and are not meant to reflect any actual physiological operations. Further central processing to *interpret* these binaural patterns is not addressed in this paper.

[6]The *monaural* LIN discussed in (Shamma, 1985a) is likely to exist in the AVCN and, hence, is not related to the binaural LIN mask used here. The binaural network receives essentially "unprocessed" input from the AVCN (e.g., via the bushy cells) and the function of the binaural LIN would most likely be accomplished within or beyond the MSO.

[7]The conceptual difference between the post- and pre-averaging LIN is that the former operation can only sharpen those spatial features generated directly by the binaural interactions (e.g., regions of high spatial frequency), whereas the latter can also extract these features when only monaural patterns are present or dominant.

[8]The length of this window is critical only when dealing with nonstationary signals and phenomena (e.g., onset cues). This final averaging step is performed here only to improve the quality of the displays (given the limited sampling rate and the high computational costs of using faster rates); otherwise, there are no theoretical or physiological reasons for not using universally shorter windows.

[9]In order to see this point more clearly, consider the activity of two symmetrically placed nodes of the network, c_{ij} and c_{ji}. Assume, for a given ILD, that the two input patterns (x and y) are similar except for a scaling factor; i.e., $y_i = ax_i$ and $y_j = ax_j$, where a is dependent on the ILD. In this case, c_{ij} and c_{ji} are given by

for multiplication

$$c_{ij} = x_i y_j = a x_i x_j,$$

$$c_{ji} = x_j y_i = a x_j x_i,$$

for addition

$$c_{ij} = x_i + y_i = x_i + a x_j,$$

$$c_{ji} = x_j + y_i = x_j + a x_i,$$

Since x_i and x_j are in general unequal (representing the activities at different points along the cochlear partition), then it follows that *addition*, unlike *multiplication*, produces asymmetric patterns around the AB diagonal. Furthermore, the output patterns with *multiplication* vanish with increasing ILD (as $a \rightarrow 0$).

[10]We have confirmed in computer simulations that both supervised and unsupervised learning algorithms can produce Jeffress-like or stereausis binaural networks depending on whether sufficient neuronal delays exist or not (Gopalaswamy, 1989).

[11]Since the shifts caused by ITDs of transient sounds exist only over a few frames of the stereausis outputs, it is critical that the size of the averaging window be kept to a minimum in the simulations.

[12]This estimate is derived from the population data of (Pfeiffer and Kim, 1975) by measuring the wavelength of the responses to the 1.2-kHz tone (= 2.5 mm) and multiplying by the frequency of the tone.

[13]There are three other possible topological arrangements of the stereausis network corresponding to the relative orientation of the two CF axes in Fig. 1(b) (for instance, reversing the axes so that the low CFs constitute the lower left corner of the figure). In all cases, the added neural delays improve the spatial resolution of the output disparity plots.

Bilsen, F. (1977). "Pitch of noise signals: Evidence for a central spectrum," J. Acoust. Soc. Am. **61**, 150–161.

Blauert, J. (1980). "Modeling of interaural time and intensity difference discrimination," in *Psychophysical, Physiological, and Behavioral Studies in Hearing*, edited by G. vanden Brink and F. Bilsen (Delft U. P., Delft, The Netherlands), pp. 421–424.

Blauert, J. (1983). *Spatial Hearing* (MIT, Cambridge, MA).

Cherry, E. (1953). "Some experiments on the recognition of speech with one or two ears," J. Acoust Soc. Am. **25**, 975–979.

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, S. and Durlach, N. I. (1978) Models of binaural interactions," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman, (Academic, New York), Vol. IV.

Delgutte, B. (1984). "Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds," J. Acoust. Soc. Am. **75**, 879–886.

Durlach, N. (1972). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. Tobias (Academic, New York), Vol. 2, pp. 369–462.

Durlach, N., and Colburn S. (1978). "Binaural phenomena," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. IV, pp. 365–466.

Evans, E. F. (1978). "Place and time coding of frequency in the peripheral auditory system: Some physiology pros and cons," Audiology **17**, 369–420.

Gopalaswamy, P. (1989). "Learning binaural processing in biological networks," M. S. thesis, Department of Electrical Engineering, University of Maryland, College Park, MD.

Green, D. M., and Yost, W. A. (1975). "Binaural analysis," in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer, Berlin), Vol. 2.

Greenwood, D. D. (1988). Personal communication.

Hartline, H. K. (1974). *Studies on Excitation and Inhibition in the Retina* (Rockefeller U. P., New York).

Hirsch, H. (1948). "The influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.

Holmes, M. H., and Cole, J. D. (1984). "Cochlear mechanics: Analysis for a pure tone," J. Acoust. Soc. Am. **76**, 767–778.

Irvine, D. F. (1986). "The auditory brainstem," in *Sensory Physiology* (Springer-Verlag, Berlin), Vol. 7.

Jeffress, L. (1948). "A place theory of sound localization," J. Comp. Physiol. Psych. **61**, 468–486.

Johnson, D. H. (1974). "The response of single auditory-nerve fibers in the cat to single tones: synchrony and average discharge rate," Ph.D. thesis, Department of Electrical Engineering, MIT, Cambridge, MA.

Keidel, W. D., and Neff, W. D. (1975). *Handbook of Sensory Physiology* (Springer, Berlin).

Kuwada, S., Yin, T., Syka, J., Buunen, T., and Wickesberg, R. (1984). "Binaural interactions in low-frequency neurons in Inferior Colliculus of the cat. IV. Comparison of monaural and binaural response properties," J. Neurophys. **51**(6), 1306–1325.

Licklider, J. (1951). "A duplex theory of pitch perception," Experientia **7**,128–134.

Loeb, G., White, M., and Merzenich, M. (1983). "Spatial cross-correlation: A proposed mechanism for acoustic pitch perception," Biol. Cybern. 149–163.

Marr, D., and Poggio, T. (1979). "A computational theory of human stereo vision," Proc. R. Soc. London Ser. **204**, 301–328.

Pfeiffer, R. R., and Kim, D. O. (1975). "Cochlear nerve fiber responses: Distribution along the cochlear partition," J. Acoust. Soc. Am. **58**, 867–869.

Poggio, G. (1984). "Processing of stereoscopic information in primate visual cortex," in *Dynamic Aspects of Neocortical Function*, edited by G. Edelman, W. Gall, and W. Cowan (Neurosciences Institute Publication, Wiley, New York), pp. 613–636.

Sayers, B. (1964). "Acoustic-image lateralization judgment with binaural tones," J. Acoust. Soc. Am. **36**, 923–926.

Sayers, B., and Cherry, E. (1957). "Mechanisms of binaural fusion in the hearing of speech," J. Acoust. Soc. Am. **29**, 973–987.

Schroeder, M. R. (1977). "New Viewpoints in binaural interactions," in *Pyschophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York), pp. 455–467.

Seneff, S. (1984). "Pitch and spectral estimation of speech based on auditory synchrony model," MIT, Work Papers Linguist., Cambridge, MA.

Shamma, S. A. (1985a). "Speech processing in the auditory system. II: Lateral inhibition and the processing of speech evoked activity in the auditory-nerve," J. Acoust. Soc. Am. **78**, 1622–1632.

Shamma, S. A. (1985b). "Speech processing in the auditory system. I: Representation of speech sounds in the responses of the auditory-nerve," J. Acoust. Soc. Am. **78**, 1612–1621.

Shamma, S. A. (1988). The acoustic features of speech phonemes in a model of auditory processing: Vowels and unvoiced fricatives," J. Phon. **16**, 77–91.

Shamma, S. A., Chadwick, R., Wilbur, J., Moorish, K., and Rinzel, J. (1986). "A biophysical model of cochlear processing: intensity dependence of pure tone responses," J. Acoust. Soc. Am. **80**, 133–145.

Sinex, D. G., and Geisler, C. D. (1983). "Responses of auditory-nerve fibers to consonent–vowel syllables," J. Acoust. Soc. Am. **73**, 602–615.

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

Sullivan, W., and Konishi, M. (1986). "Neural map of interaural phase difference in the owl's brain-stem," Proc. Nat. Acad. Sci. **83**, 8400–8404.

Sullivan, W. E., and Konishi, M. (1984). "Segregation of stimulus phase and intensity coding in the cochlear nucleus of the barn owl," J. Neurosci. **4**(7), 1787–1799.

Whitworth, R., and Jeffress, L. (1961). "Time versus intensity in the lateralization of tones," J. Acoust. Soc. Am. **33**, 925–929.

Yin, T., and Kuwada, S. (1984). "Neuronal mechanisms of binaural interactions," in *Dynamic Aspects of Neocortical Function*, edited by G. Edelman, W. Gall, and W. Cowan (Neurosciences Institute Publication, Wiley, New York), pp. 263–314.

Young, E. D., and Sachs, M. B. (1979). "Representation of steady state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," J. Acoust. Soc. Am. **66**, 1381–1403.