

**Stereausis: Binaural Processing
Without Neural Delays**

by

**S.A. Shamma, N. Shen, and P.
Gopalaswamy**

Stereausis: Binaural processing without neural delays

Authors: Shihab A. Shamma (a,b), Naiming Shen (c), Preetham Gopalaswamy (c)

**(a) Electrical Engineering Department, Systems Research Center,
& the University of Maryland Institute for Advanced Computer Studies
University of Maryland
College Park, Maryland 20742**

**(b) Mathematical Research Branch
National Institute of Diabetes
Digestive and Kidney Diseases
National Institutes of Health
Bethesda, Maryland 20982**

**(c) Electrical Engineering Department
University of Maryland
College Park, Maryland 20742**

Received:

Abstract

A biologically realistic neural network model is proposed for the binaural processing of interaural-time and level cues that closely resembles computational schemes suggested for stereopsis in vision. 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. The two-dimensional network measures interaural differences by detecting the spatial disparities between the instantaneous outputs of the two ears. 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 cochleae by combining an ipsilateral input at a given CF with contralateral inputs from *locally off-CF* locations. Thus, the network utilizes the delays already present in the travelling waves of the basilar membrane to extract the correlation function. Simulations of the network operation with various signals are presented; Physiological arguments in support of this scheme are also discussed.

I. Introduction

Many attributes of spatial hearing fundamentally require the processing of binaural cues. This is particularly true of *lateralization* and *signal detection and enhancement* - two phenomena which have been the subject of intensive multi-disciplinary 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. A basic operation in all these models involves 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, e.g. 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 and Durlach [1978]) or continuous deterministic functions (Bilsen [1977]) to represent the responses of the auditory-nerve). These differences aside, however, the correlator in all cases computes, at various time-lags, a running cross-correlation measure of the cochlear outputs from corresponding or *equal* characteristic frequency (CF) locations on the two ears. 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 would then reflect the sound spectral and lateralization information along its 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. All this is strictly applicable to interaural-time-differences for low frequency sounds (≤ 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]).

In searching for the neural substrate of the correlation 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. Fig.1a 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 precise topologically ordered range of delays that allows its correlation with the contralateral output *at the same CF* to be computed. Much psychophysical and neurophysiological data have been interpreted in support of this hypothesis. However, there is as yet no direct anatomical or physiological support of the existence of this type of network, or of the essential functional relevance of any neural delays. Furthermore, the available data could be interpreted in the contexts of other models, as evidenced by the multiplicity of the proposed models (Colburn and Durlach [1978]). Therefore, it should be emphasised 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 correlation model) which is coupled to a highly schematic view of cochlear function as merely a bank of bandpass filters. However, as we shall elaborate in this report, detailed features of the spatiotemporal responses of the auditory-nerve encode the binaural cues in a manner which makes further neuronal delays completely redundant. By processing these response features, simple alternative network topologies requiring no such delays may approximate the function of the hypothetical

correlation models.

In the following sections, we shall present a simple, biologically realistic, neural network to account for the psychophysical attributes of two aspects of binaural hearing - lateralization and signal enhancement. Because of the fundamental similarity that emerges between this network and the type of computations proposed for *stereopsis* in vision (Marr and Poggio [1979]), we will refer to it as the *stereausis network*. In section II, we shall outline the topology of this network and the basic principles underlying its operation. In section III, the network outputs are interpreted for a wide range of tasks, and the results are finally discussed in section IV.

II. 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*:

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 network operation should

be robust with respect to parameter variations.

4. It is unclear at present whether separate or identical binaural networks and pathways are involved in the processing of interaural-level and interaural-time-differences. Nevertheless, in view of the similarity of the intended psychoacoustical attributes (e.g. lateralization) it would be desirable to demonstrate the possibility that either one network is capable of gracefully processing both types of cues, or that closely related network architectures underlie their processing. 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, enhancement of noisy signals, and the integration of different source cues.

A simple neural network is proposed that largely 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 travelling wave which is projected relatively intact unto the responses of the spatially ordered array of auditory-nerve-fibers¹. At any instant in time, the central binaural processor receives two spatial *images* (or snap-shots) of the travelling 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 travelling waves differ systematically. Thus, when the tone is phase-shifted (or delayed) in one ear relative to the other, the images appear correspondingly shifted. Since this *spatial* disparity between the travelling 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 which affect the relative amplitudes of the travelling waves. As we illustrate in later sections, many other

¹Because of hair cell transfer characteristics, the travelling wave image undergoes several modifications which include rectification, saturation, and lowpass filtering (see appendix A1). The effect of all these factors are taken into account in the analysis presented in this paper.

possible inequalities in binaural inputs, e.g. in their envelopes, degree of correlation, or bandwidths, can be readily detected and consistently represented via the spatial disparities between the resulting travelling waves.

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. 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]). Fig.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 appendix A1). The stereausis network combines the ipsilateral and contralateral cochlear outputs in a simple ordered matrix of operations as shown in Fig.1b. Thus, at the (i^{th}, j^{th}) node the responses of the i^{th} ipsilateral fiber (x_i) and the j^{th} contralateral fiber (y_j) are combined to produce $c_{ij} = C(x_i, y_j)$. $C(.,.)$ here 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 (e.g. x_i) is systematically correlated with outputs from CF and *off*-CF cochlear fibers of the other ear (e.g. $y_{i-1}, y_i, y_{i+1}, \dots$). The significance of this arrangement for the cross-correlation computations, as we shall elaborate later, is that because of the finite velocity of the travelling 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 (Pfeifer and Kim [1975]; Shamma [1985b]); This possibility seems to have been first proposed by (Schroeder [1977]). An alternate view of these operations is that the stereausis network computes along its different diagonals (parallel to the centered diagonal shown in dashed lines) the correlation of the two cochlear images at different *lateral spatial shifts*. Thus, along the center diagonal (c_{kk} axis in Fig.1b), the cochlear patterns are spatially registered; Off

this diagonal, however, and along axes parallel to it (e.g. 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 axes act as disparity detectors for a characteristic *lateral* spatial disparity.

As we shall discuss in more detail shortly (and in the appendix A2), 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 10 msec, which may also be thought of as the time constant of the output neurons). 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, y). We shall illustrate primarily the results for $c_{ij} = g(x_i - y_j)$ where $g(.)$ is a threshold operation: $g(z) = \max(z, 0)$. Final outputs, however, with $c_{ij} = x_i + y_j$ and $c_{ij} = x_i \cdot y_j$ will also be illustrated in a few cases for comparison.

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 travelling patterns of activity mimicking crudely the cochlear waves. Figs.3a-b show the network *averaged* outputs (computed with $c_{ij} = x_i \cdot y_j$) for a single peak sweeping the spatial axis of the two inputs. In Fig.3a the inputs are identical at all times, and consequently the trajectory of the maximum correlation occurs at units along the diagonal AB. When a peak is delayed in one input relative to the other (e.g. ipsilateral lags in Fig.3b), 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 and evaluate the *temporal* delays between the two inputs by virtue of the accompanying *spatial* disparities they produce.

Figs.3c-d illustrate the outputs for two travelling peaks a distance D apart. As before,

when the inputs are centered (Fig.3c), the units along the diagonal AB become maximally activated. Two parallel, but smaller secondary maxima, also emerge at cells that correlate spatial locations separated by the same distance D. In Fig.3d, a relative temporal delay between the input patterns (ipsilateral pattern leading) causes a displacement of the entire output as expected.

The network outputs with a continuous low frequency tone

A low frequency tone evokes a travelling 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 travelling wave (Figs.4). In addition, The responses exhibit two unique features related to the amplitude and phase characteristics of the cochlear filters (Figs.4b-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. (2) The secondary maxima of the network outputs appear to converge towards the primary maximum (located along the AB diagonal in Fig.4b). This is due to the rapidly increasing slope of the phase function of the travelling wave, and the accompanying decrease of the spatial separation of its peaks, near the point of resonance.

The spatial spread of these outputs reflects the bandwidth of the cochlear filters. The patterns can be readily ‘focused’ with a somewhat elongated *ON-center/OFF-surround* mask (i.e. a two-dimensional lateral inhibitory network) that is applied uniformly to the entire binaural c_{ij} output (see appendix A2). This operation enhances the display by reducing the spread of the patterns and relatively boosting the outputs parallel to the AB diagonal. In all the illustrations that follow, this operation will be applied to the c_{ij} outputs *before* taking the final short time averages. Purely for computational convenience, the mask is applied here in a nonrecursive (feedforward) manner; Qualitatively similar results can be obtained with recursive (feedback) connections (see appendix A2).

When an interaural-time-delay is introduced, the travelling waves of the two ears become relatively delayed, causing a spatial disparity between the binaural inputs to appear (Fig.4a)². 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.4c). The results of these and other manipulations will be discussed in detail in the next section.

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. We will also discuss further transformations of these outputs that lead to more abstract representations of the stimulus parameters.

III. Binaural processing in the Stereausis Network

In this section, we shall illustrate the outputs of the binaural processor for a wide range of typical, psychoacoustically well studied, set of stimuli. All results shown can be computed using $C(x_i, y_j) = g(x_i - y_j)$ and $C(x_i, y_j) = x_i + y_j$. These operations are most likely performed at the LSO and the MSO respectively. Since for the subtraction, c_{ij} exhibits minima (rather than the maxima) at the points of maximum correlations, we shall apply an *OFF-center/ON-surround* mask which, in addition to sharpening the final outputs, will reverse the polarity of the display³. Details of the mask and other computations are given in appendix A2.

²Note, that since the *envelope* of the travelling 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.

³Clearly, reversing the polarity of the display here is not meant to have any physiological significance, but is

The responses to six classes of stimuli are illustrated below. The first five emphasize so-called *ongoing* cues, i.e. binaural processing of continuous sound. The last example deals briefly with transient cues. The processing conditions are:

1. Lateralization of low frequency tones (≤ 1.5 kHz): Interaural-time-delays (ITD).
2. Lateralization of high frequency tones: Interaural-level-differences (ILD).
3. Time-level trading for low frequency tones.
4. Lateralization of complex sound stimuli: Signal envelope cues.
5. Detection and enhancement of tones in noise.
6. Onset cues.

1. 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). Figs.5 and 6 illustrate how the stereausis network represents these percepts for 600 Hz and 1100 Hz tones at $0, \pi/3, 2\pi/3, \pi$ 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 insets⁴. 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 towards the center and becomes now 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]).

done simply for the convenience and consistency of the display throughout the paper.

⁴The disparity plot is always generated by collapsing (unto the bar shown) the local patterns of activity below it (the band of collapsed activity is 12 points wide, basal to the bar). This roughly corresponds to the spatial extent of the cochlear responses due to the stimulus, i.e. the spread of its travelling wave envelope. The spectral axis is generated by collapsing unto the AB diagonal a region of activity, 10 points wide on either side.

There are two basic differences between the network outputs to the 600 Hz and 1100 Hz tones: (1) The spectral peaks (which remain stationary for all plots) are located at different CFs reflecting the frequencies of the tones. (2) The disparity peaks are more spatially separated in the case of the 600 Hz tone. The latter occurs because the spatial phase functions of the travelling wave are steeper for higher frequencies (Pfeifer and Kim [1975]). This brings up the important question of how to interpret the magnitude of the shifts in disparity plots. To answer this question, we first re-iterate that the cells of the stereausis network are *spatial* disparity detectors, i.e. they are tuned to specific horizontal shifts of their binaural inputs. Since, as we argued earlier, an ITD causes a corresponding spatial shift in the fine structure of the travelling wave patterns on the two ears, then different stereausis cells will respond to different ITDs. The mapping of different ITDs to equivalent spatial shifts is independent of frequency only if the velocity of the cochlear waves as a function of space remains the same for all tones (except for a frequency-dependent spatial translation). This is true to a first approximation (Dallos [1973]), and consequently the stereausis cells may display, in addition to their spatial disparity tuning, a sensitivity to characteristic ITDs. It seems, therefore, reasonable to interpret the disparity axis as absolute lateral positions in space (and approximately as different ITDs), and thus associate directly the shifted peaks of the disparity plots with lateralized auditory images. Furthermore, if we associate complete lateralization in space with the positions of the peaks of the π shifted 600 Hz tone (i.e. maximum ITD \approx .7-.8 msec (Colburn and Durlach [1978])), then it becomes immediately apparent why higher frequencies (with their closely spaced disparity peaks in Figs.6) cannot be fully lateralized. This also explains why lower frequencies are completely lateralized at smaller phase-shifts (Sayers and Toole [1964]).

2. 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 travelling 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 (c_{ij}) which 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 4 kHz tone (using the addition operation, $c_{ij} = x_i + y_j$) are shown in Figs.7. When the binaural inputs are identical (centered tone), a symmetric (with respect to AB) pattern of activity is evoked, whose two edges (one due to each input) are then extracted by the lateral inhibitory mask and displayed in Fig.7a. When projected unto the disparity axis, a centered peak emerges. Fig.7d illustrates the opposite extreme case of a monotonic ipsilateral input. Here, only the weak horizontal edge of activity appears, 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 peak⁵. In between, increasing the input ILDs is systematically reflected in the relative levels of the two extracted edges (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 $c_{ij} = g(x_i - y_j)$, but not with the multiplicative $c_{ij} = x_i \cdot y_j$ correlation. The underlying reason why the subtraction and addition operations work is the consistent *asymmetrical* change of the output c_{ij} around the AB diagonal as a function of the ILDs. In the final display, the lateral inhibitory mask emphasizes the pattern edges and their ILD-induced changes.

Finally, it is evident from the disparity plots of Figs.7 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, i.e. disparities due to lateral shifts in the cochlear patterns. A pure ILD, instead, creates binaural cochlear patterns that approximately

⁵Note that this disparity plot is but one measure of the asymmetry in the 2-dimensional response patterns of the network. Its only (perhaps irrelevant) attribute, as mentioned earlier, is that the collapse is performed over the region of maximum cochlear activity associated with the stimulus (i.e. spatial critical band).

differ *only* in their relative amplitudes. 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 (see example 3 and discussion in section IV).

3. *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. Figs.8 illustrate the effects on the 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; (2) The secondary peak which grows in 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 Figs.8 can be readily understood in the context of the previous two examples. Thus, as seen in the network outputs of Figs.5 and 6, 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 Figs.7). 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 (Blauert [1982])

4. Lateralization of complex sound stimuli

The representation and lateralization of complex signals is 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 similarly by the stereausis network. For instance, consider the cochlear responses to the amplitude modulated (AM) high frequency tone shown in Fig.9a. Here, the carrier frequency (3kHz) is too high to elicit any significant responses phase-locked to its frequency. However, the slow sinusoidal modulation (100 Hz) of the carrier is well reproduced in the modulation of its cochlear responses. As far as the binaural processor is concerned, this AM signal appears roughly as a 100 Hz tone transcribed to the higher CF locations. Consequently, any interaural manipulations of the *envelope*, e.g. 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 frequency (3 kHz) along the AB diagonal. Figs.9b-c illustrate the outputs for a centered and a binaurally delayed envelope (.5 msec, ipsilateral leading). Note the increased width of the peak in the spectral plots reflecting the increased bandwidth of the stimulus (compared to the single tone cases earlier).

In general, the ability to utilize the interaural cues of the envelope of a complex signal (such as speech) depends critically on the details of its spatiotemporal representation in the responses of the auditory-nerve, and hence on the shapes and bandwidth of the cochlear filters. Thus, when the components of a complex sound are closely spaced (and hence unresolved by the analysis filters), a rich temporal structure in the response (due to the inter-modulation product) is expected, and sensitivity to ITDs is likely. The same lines of reasoning can be applied to predict the influences of ITDs on the lateralization of broad-band noise signals.

5. Detection and enhancement of tones in noise

One of the most 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, e.g. the detection of single low frequency tones in noise backgrounds of varying degrees of binaural coherence (Durlach and Colburn [1978]; Hirsch [1948]). In Figs.10 we illustrate the processing and the nature of the outputs of the stereausis network in such stimulus conditions. First, in Fig.10a ($S_o N_o$ case) a centered tone (1100 Hz) in binaurally coherent noise (bandwidth = .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 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 $S_o N_o$ case, around 200 Hz, and 5-10 dB near 1 kHz (Blauert [1983])). The representation of these signals in the binaural network (Figs.10b,c) demonstrates the clear separation of the two components of the complex along both the disparity and spectral axes. Thus, in Fig.10b ($S_\pi N_o$ 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.10c ($S_o N_\pi$ 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, e.g. partially coherent noise (Durlach and Colburn [1978]). The network response to a monaural tone in binaurally coherent noise is depicted in Fig.10d. Here, the monaural stimulus excites (relatively weakly) a band of units off the diagonal AB.

6. Onset cues

A stimulus onset is understood here to mean the short time interval near the beginning of the signal where the stimulus *envelope* rises from a low to a high amplitude rapidly. Viewed as such, the detection of interaural differences at stimulus onsets becomes a special case of the binaural processing of signal envelopes (example 4 above).

Rapid onsets (and offsets) create distinctive spatiotemporal response patterns on the auditory-nerve which, as in the case of amplitude modulation, contain a rich temporal structure reflecting the wide-band nature of the cue. Fig.11a illustrates these patterns for a high frequency tone (3 kHz). The stereausis network processes and represents the ITDs of these patterns as in earlier examples - the major difference being the much shorter window for time averaging (1.5 msec) used to generate the displays (see appendix A2). In Figs.11b-c the signal is binaurally centered and the onset patterns arrive simultaneously to the binaural network. When an ITD is introduced, a mismatch of the input patterns, and the outputs of the network thus reflect the direction and magnitude of the ITD. Note that, because of the absence of significant phase-locking to this high frequency stimulus, *ongoing* ITDs cues are not useful for its lateralization.

IV. Discussion

A simple biologically realistic neural network is proposed as the central binaural processor in the mammalian auditory system. It combines the spatiotemporal responses arriving from the two cochlea 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 (e.g. ITDs and ILDs) by applying exactly the same principle of operation - that is to compute systematically a measure of the correlation of instantaneous activity from CF and off-CF locations on the two cochlea. An important and distinguishing property of this network is that *no* neural delays are required to perform its function.

There are two equivalent perspectives of the fundamental operation of the stereausis network: First, in relation to previously suggested cross-correlation implementations and the extensive literature and data associated with them; Second, in analogy to disparity detection for stereopsis in vision. These two views have close parallels in the monaural processing of cochlear responses by the lateral inhibitory networks (LIN) proposed earlier [spee2], i.e. auto-correlation and edge-detection. Although there is no conclusive evidence in support of any one model, the stereausis network is compatible with much of the physiological data

already available from mammalian preparations and from the barn owl. In the remainder of this article, we shall address these issues in detail.

Relation to correlation-based models

Correlation-based models have for a long time provided a coherent framework to interpret substantial amounts of experimental results (Blauert [1983]; Durlach and Colburn [1978]). Many computational algorithms have been proposed as possible physiological implementations of this operation. The basic components in all include a frequency analysis stage (presumably the cochlea) where the components of a complex sound are separated and transduced unto different channels, followed by a cross-correlator (usually thought of as a coincidence detector). The latter stage computes a running cross-correlation of the responses of fibers originating from *equal* CF locations. If the correlation function is computed at different lags, then such binaural cues as the ITD can be readily estimated as the lag corresponding to the maximum of the correlation function; And, in order to map this function spatially, an ordered array of delays is usually postulated as shown in Fig.1a (Jeffress [1948]). Further extensions of this model have been proposed in order to detect and map ILDs (Blauert [1983]).

The stereausis network is in basic harmony with the philosophy of the correlation models. The primary disagreement is on whether neural delays need to play any functional role in the neural networks that might perform these binaural computations. This is because the basilar membrane already acts as a "delay line". Thus, when a tone excites a travelling wave on the membrane, the phase-locked responses from a small CF region (near the point of resonance of the wave) provide multiple versions of the tone waveform that are systematically delayed relative to each other (Shamma [1985b]). This is a direct manifestation of the phase function of the travelling wave, or equivalently, of the finite velocity of the wave on the basilar membrane (Pfeifer and Kim [1975]; Shamma [1985b]).

Therefore, in order to correlate binaural responses at different lags, we need not insert further delays, but rather correlate with fibers from slightly different CFs on the two ears. This idea leads directly to the stereausis network topology shown in Fig.1b. This possibility, in fact, has already been postulated to explain various aspects of the binaural responses of

cells in the Inferior Colliculus (IC) (Yin and Kuwada [1984]), of binaural unmasking in the "equalization and cancellation" theory (Schroeder [1977]), and as a mechanism for monaural periodicity detection (Loeb, White and Merzenich [1983]).

From a theoretical standpoint, the cross-correlation computed by this method is somewhat different from that obtained with pure (neural) delays in that the spatially delayed responses are also weighted by the envelope of the travelling wave (see section II). This, together with shape of the phase-function, dictate as illustrated in the results earlier, the range of effective delays computable and their interpretation in the binaural spatial patterns. Those are typically introduced in other models as *ad hoc* restrictions or parameters. Finally, as we discuss in the next section, adopting this strategy for processing the ITDs immediately extends to other interaural cues such as ILDs, onset cues, and signal envelope manipulations.

Disparity detection and stereopsis in vision

It has long been recognized that binaural processing of sound is functionally analogous to depth perception in vision, that audition utilizes binaural disparities to perform this task the same way the visual system detects and uses binocular disparities, and that certain auditory specializations to preserve the temporal cues (e.g. the end bulbs of Held) are paralleled by retinal spatial specializations (e.g. the fovea) (Yin and Kuwada [1984]). The stereopsis network takes this correspondence to its logical conclusion - that essentially the *same* network topologies proposed for stereovision be used for binaural processing, e.g. (Marr and Poggio [1979]).

In stereovision, depth information is derived from the disparities of the spatial images of the same object on the retinas. The CNS may derive 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 stereopsis network is very similar. Thus, for each diagonal of cells (parallel to AB), the instantaneous response patterns of the two cochleae are projected

with a specific horizontal shift. Since interaural delays create mismatched cochlear images, the network can detect and encode these temporal cues in exactly the manner of the stereopsis network.

Parallels between monaural and binaural processing

There are striking parallels between the motivations and subsequent developments of the binaural stereopsis network and the lateral inhibitory network (LIN) 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 *auto-correlation* 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, i.e. 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]).

Correlation-based models of binaural processing have a similar history. Thus, in order to compute *absolute* temporal interaural delays, a *cross-correlation* operation between the responses of corresponding (equal CF) fibers from the two ears was invoked and consequently, as was the case for the auto-correlation implementation, a series of delay lines were postulated (Jeffress [1948]). However, just like the LIN, the stereopsis network obviates the need for such delays by relying, instead, on the local spatial correlations of CF and off-CF responses, i.e. detecting spatial disparities between the response images of the two ears regardless of their temporal (ITD) or amplitude (ILD) origins. In fact, for monaural sounds, the computations of the stereopsis network (as formulated in section II) reduce to those of the LIN.

The basic principle that emerges here is that in detecting and estimating the inherently

temporal cues of sound signals (whether spectral or interaural) it is simplest to seek and process the accompanying spatial features that are invariably produced by the cochlea (Shamma [1985b]). In this light, time-to-space conversion appears to be a primary function of the cochlea, allowing the CNS to perform in the *spatial* domain operations that approximate the essentially temporal auto- and cross-correlations, long hypothesized to be the bases of auditory processing. In fact, once reformulated in the framework of spatially distributed patterns and computations, auditory processing at its early central stages appears conceptually very similar to the processing in other more obviously spatial sensory systems, such as the visual and the somatosensory systems⁶.

Anatomical and physiological correlates in mammals and the barn owl

Much of the available physiological data can be interpreted in support of correlation-based models of binaural hearing, though no conclusive evidence of any particular implementation exists (Blauert [1983]). For instance, responses of single units cross-correlating various binaural noise and tonal stimuli (e.g. in the MSO, LSO, and IC (Yin and Kuwada [1984])) can be readily interpreted as the responses of different cells of the stereausis network. A more fundamental question, however, concerns the potential location of this network and its relation to other auditory structures. It should be emphasized here that in searching for biological correlates of this and other networks, we seek correspondences to the fundamental topology and function and *not* the arbitrary mathematical formulation used in this instance (see further comments in Section II).

In mammals, the primary nuclei of the SOC (the MSO, and LSO) are the most likely sites for the matrix of disparity matching operators. The MSO and LSO differ in the nature of their binaural interactions, with most MSO cells being excitable binaurally and most LSO cells being inhibited contralaterally (Yin and Kuwada [1984]). Based on this and other physiological (distribution of best frequencies), anatomical, and ethological evidence, it has been argued that the MSO and the LSO serve different complementary functions - namely

⁶An equivalent view of the role of the cochlea in higher animals is that of a *point source-to-spatially distributed source* transformer. In the fish, the spatial character of the sensory input is explicit (as, for instance, in vision). In mammals, birds, and lizards, the hair cell array is internalized, and it is the basilar membrane which provides the spatially distributed input from the essentially point source excitation of the tympanic membrane.

localization of low and high frequencies (Yin and Kuwada [1984]). However, the above simulations of the stereausis network (section III) have shown that, provided phase-locking is preserved and used by the binaural network, the exact form of the binaural operation is not critical and both the MSO excitatory-excitatory interactions (e.g. $c_{ij} = x_i + y_j$) and the LSO excitatory-inhibitory interactions (e.g. $c_{ij} = g(x_i - y_j)$) can in principle account for all the binaural phenomena examined in section III, the major difference being the polarity of two outputs.

Nevertheless, the possibility of a functional division between the MSO and LSO may be re-expressed in slightly different terms - that of *accurate* ITD and ILD detection, *regardless* of frequency. Such a distinction has significant implications for the architecture of the possible underlying networks. Thus, as we argued earlier (example 2, section III), the specific form of the stereausis network proposed here is fundamentally suited for ITD detection. This is because the detector cells are so innervated as to be highly sensitive to *horizontal* disparities in the binaural inputs. Such disparities are primarily generated by ITDs, and must be conveyed to the network through phase-locked pathways.

In contrast to ITD shifts, ILD-induced disparities are *vertical* in that they primarily change the relative amplitudes (envelopes) of the cochlear waves on the two ears. Such disparities, while not localized precisely by the stereausis network, do influence its outputs in a manner consistent with the effects of other interaural differences; And this may well be sufficient to account for ILD perception in many animals, including humans (Blauert [1983]). For the *accurate* detection of ILDs, however, the binaural network should instead directly sense a range of vertical disparities. A simple network would achieve this by having its input fiber arrays registered (i.e. no lateral shifts of the cochlear images) and by progressively attenuating (and hence vertically shifting) one of its inputs down the network (for instance, through a gradual change in thresholds or in the efficacy of the contralateral inhibitory or the ipsilateral excitatory inputs)⁷. Each row of cells thus becomes sensitive to a particular ILD. Note that phase-locking plays no role in the intended function here. While such a network is one of

⁷The topology of this network is identical to the Jeffress model (Fig.1a or 12a) except that delays are replaced by a graded threshold or contralateral inhibition.

many possible schemes suggested earlier for ILD encoding (Colburn and Durlach [1978]), it especially complements the design of the stereausis network, functionally (ITD vs. ILD detection), conceptually (horizontal vs. vertical disparities), and topologically (unregistered vs. registered inputs).

Recently, the first example of an organized map of responses to different ITDs was identified in the Nucleus Laminaris (NL) of the barn owl (Sullivan and Konishi [1986]). Sullivan and Konishi interpreted their physiological and anatomical data in support of the schematic structure of the NL binaural processor shown in Fig.12a - i.e. the cross-correlator model of (Jeffress [1948]). Specifically, they have found that ipsilateral fibers enter the NL, bifurcating and synapsing with NL cells, along roughly the isofrequency planes. Furthermore, they postulated that, based on the distances that the ipsilateral and contralateral fibers have to course past one another across the NL, and based on nerve conduction velocities, this arrangement may well be the neural substrate of the hypothetical delay lines. These arguments are reasonable and it is, in fact, quite possible that different species might employ different processing strategies. Nevertheless, *if* the stereausis network is the underlying neural substrate in the NL, then only minor topological changes are necessary in Fig.12b - namely that each ipsilateral fiber becomes slightly tilted so as to synapse with *several* fibers derived from the corresponding contralateral local region. Because of its limited (local) spatial extent, this overlap perturbs only slightly the overall isofrequency planes of the NL.

There are basic conceptual differences, however, between the two figures, particularly regarding the *relevance and functionality* of the neural delays. Neural delays are an inevitable occurrence in any network where information is transmitted from one point to another. In Fig.12a, these delays are at the essence of the network operation, while in Fig.12b they are primarily an epi-phenomenon. They may, however, have a secondary beneficial effect in improving the resolution along the disparity axis. The essential aspect of the stereausis network topology is therefore not the neural delays, but rather the correlation of CF and off-CF responses.

It should be emphasized, however, that although we have contrasted as two extremes the networks of Figs.12, a continuum between the two exists. That is, the amount of CF-overlap

(or tilt of the ipsilateral fibers) in stereausis-like networks may vary, reflecting the balance between using neuronal or basilar membrane originated delays. In fact, it is likely that, in a biological realization of the stereausis network, such compromises occur to compensate for the delays inherent in neuronal transmission or to anatomical imbalances between the binaural pathways or cochleae. Experience during development would presumably achieve the fine tuning of the final map.

At higher levels of the auditory system of the barn owl (ICX), detailed physiological response maps have been described, with cells tuned to different combinations of the two binaural cues (ILDs and ITDs), and hence to well defined points in the external space (Knudsen [1984]). It is yet unclear how the NL map is transformed further at higher structures, and particularly how sensitivity to ILDs at a given ITD arises and later organized. The most direct hypothesis is a point-to-point combination of the outputs of independent ITD and ILD networks as described earlier (Sullivan and Konishi [1984]; Takahashi, Moiseff and Konishi [1984]). A similar map, but less accurate on its ILD axis, can be based solely on the results of the time-intensity trading experiments discussed earlier in section III. There, the two interaural cues had separate and stable simultaneous representations - the ITD in the location of the central narrow peak, and the ILD in the location (along the same axis) of the center of gravity of the overall pattern. Those two measures can be readily extracted from the output patterns (via lateral inhibitory and excitatory networks) to form the bases of a more abstract map of space. Finally, we suggest that using similar neural network topologies and representations for the detection of horizontal and vertical disparities of acoustic and visual images would facilitate the formation of higher audio-visual maps (as found in the Optic Tectum of the barn owl (Knudsen [1984])).

Acknowledgment: This work is partially funded through an NSF Initiation Award (ECE-85-05581) and NSF (CDR-85-00108).

Appendix

A1. The cochlear model:

The cochlear spatiotemporal patterns are computed using digital algorithms based on a detailed multi-stage biophysical model of the 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 highpass filtered ($w_n = u_n - .8u_{n-1}$; modelling both outer ear and fluid-cilia coupling stages) and compressed by a sigmoidal function of the form: $x = M \cdot 1 / (1 + b \cdot e^{-a \cdot w})$, where a , b , and M are parameters of the nonlinearity, and x , w are the output and input respectively. Finally, a lowpass filter smooths the output (time constant = .1 msec). The parameters of the compressive nonlinearity should be such that approximately 30 dB of linear gain is available between threshold and saturation (defined as .1 - .9 of maximum output level, M) and that the output is saturated at moderate sound levels (approximately 60 dB SPL). Each fiber (channel) is labeled by a characteristic frequency (CF). 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.

A2. The computations of the stereausis network:

The computations in the stereausis network are based on a matrix of correlation-like 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 appendix A1. 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:

$$c_{ij} = C(x_i, y_j) = \int_T x_i(t) \cdot y_j(t) dt \quad (1)$$

where $x_i(t)$ and $y_j(t)$ are the response of the i^{th} ipsilateral and j^{th} contralateral fibers, respectively, and T is the period 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 travelling 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.:

$$x_i(t) = A_i(\omega) \sin(\omega t + \theta_i(\omega)) \quad (2)$$

and

$$y_j(t) = A_j(\omega) \sin(\omega t + \theta_j(\omega)) \quad (3)$$

where $A_i(\omega)$, $A_j(\omega)$ and $\theta_i(\omega)$, $\theta_j(\omega)$ are the amplitudes and phases of the travelling waves at the i^{th} and j^{th} locations of the two cochleae. 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:

$$y_j(t) \approx A_i(\omega) \sin(\omega t + \theta_i(\omega) - \delta\theta(\omega)) \quad (4)$$

If we assume now that the velocity of the travelling wave (v) over the small distance (δs) between the i^{th} and j^{th} locations is approximately constant, then the spatial frequency (ω_s) of the travelling wave at i can be related to the frequency of the tone as:

$$\omega_s = \frac{\omega}{v} \quad (5)$$

and hence,

$$\delta\theta(\omega) \approx \omega_s \cdot \delta s = \frac{\omega}{v} \cdot \delta s = \omega \tau_s \quad (6)$$

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:

$$y_j(t) \approx A_i(\omega) \sin(\omega t + \theta_i - \omega \tau_s) = y_i(t - \tau_s) \quad (7)$$

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

$$c_{ij} = \int_T x_i(t) \cdot y_i(t - \tau_s) dt \quad (8)$$

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

The multiplicative form of $C(x_i, y_j)$ is often invoked for mathematical convenience. Two other operations produce qualitatively similar results, and are probably more biologically

plausible. They are the addition ($C(x_i, y_j) = x_i + y_j$) and subtraction with threshold ($C(x_i, y_j) = g(x_i - y_j)$) operations. In all the examples shown in section III, the outputs c_{ij} are computed within a stripe surrounding the diagonal AB ($|i - j| \leq 30$). At each time instant, a two dimensional image (c_{ij}) is produced and sharpened by an elongated ON-center/OFF-surround lateral inhibitory network to generate (o_{ij}):

$$o_{ij} = \sum_{ijkl} w_{ijkl} \cdot c_{kl} \quad (9)$$

where the network connectivities w_{ijkl} around each (i, j) neuron are identical (i.e. a uniform network). The specific mask shown below is used with the addition operation; Its inverse is used with the subtraction operation in order to invert the polarity of the final display. The

$$\begin{array}{ccccccc} +1 & +1 & +1 & \dots\dots\dots & c_{kk+2} \\ -3 & -3 & -3 & \dots\dots & \\ +4 & +4 & +4 & \dots\dots\dots & c_{kk} \\ -3 & -3 & -3 & \dots\dots & \\ +1 & +1 & +1 & \dots\dots\dots & c_{kk-2} \end{array}$$

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.5a and 6a). 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. For the purposes of this paper, both forms produce similar results; The latter topology, however, is more commonly reported in the biological literature (Hartline [1974]; Marr and Poggio [1979]; Shamma [1985a]). For the final displays of examples 1-5 in section III, the outputs o_{ij} for 250 time samples (10.5 msec) are averaged over the entire duration. For the onset cues, the averaging was restricted to 1.5 msec around the onset of the stimulus.

Figure Legends

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 stereausis binaural network. Ipsilateral (X_i) and contralateral (Y_j) fibers are processed in an ordered matrix of operations c_{ij} ; (shown in more detail to the right).

Fig.2: Responses of the cochlear model.

The spatiotemporal response patterns of the cochlear model (appendix A1) 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 Characteristic Frequency (CF) of each output channel (see method of CF labeling in appendix A1). The input waveform of the tone is also shown. In the auditory system, these patterns represent the probability of firing on the auditory-nerve.

Fig.3: Stereausis processing with synthetic patterns.

(3a) The output of the stereausis network with a synthetic pattern (a binaural travelling 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.

(3b) The output of the stereausis network with the synthetic pattern binaurally delayed. The ipsilateral pulse leads the contralateral pulse.

(3c) The network output with a pair of centered travelling pulses.

(3d) Same as c, but with pulses binaurally delayed. Ipsilateral leads contralateral.

Fig.4: Stereausis processing with a low frequency tone.

(4a) A schematic of the binaurally shifted travelling waves on the basilar membrane.

(4b) 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 patterns of activity computed near the bar (see text). The spectral plot displays the activity near and along the diagonal AB (see text).

(4c) Same as (b) but with the 600 Hz tone binaurally delayed (ipsilateral side leads contralateral by $\frac{2\pi}{3}$ phase-shift). The disparity plot shows clearly the resulting shift of the patterns.

Fig.5: Representation of ITDs with low frequency tones (600 Hz).

(5a) Stereausis network outputs of a centered 600 Hz tone. Same as (Fig.4a) but patterns are sharpened with a lateral inhibitory mask.

(5b) Same as (a) but with tone $\frac{\pi}{3}$ phase-shifted (Ipsilateral leading).

(5c) Same as (a) but with tone $\frac{2\pi}{3}$ phase-shifted (Ipsilateral leading).

(5d) Same as (a) but with tone π phase-shifted (Ipsilateral leading).

Fig.6: Representation of ITDs with low frequency tones (1100 Hz).

(6a) Stereausis network outputs of a centered 1100 Hz tone.

(6b) Same as (a) but with tone $\frac{\pi}{3}$ phase-shifted (Ipsilateral leading).

(6c) Same as (a) but with tone $\frac{2\pi}{3}$ phase-shifted (Ipsilateral leading).

(6d) Same as (a) but with tone π phase-shifted (Ipsilateral leading).

Fig.7: Representation of ILDs with high frequency tones (3 kHz).

(7a) Stereausis network outputs of a centered 3 kHz tone (ILD = 0 dB).

(7b) Same as (a) but with ILD = 3 dB (Ipsilateral higher).

(7c) Same as (a) but with ILD = 6 dB (Ipsilateral higher).

(7d) Same as (a) but with ILD = 12 dB (Ipsilateral higher).

Fig.8: Effect of ILDs on low frequency tones.

(8a) Stereausis network outputs with a centered 600 Hz tone (ILD = 0 dB).

(8b) Same as (a) but with ILD = 3 dB (Ipsilateral higher).

(8c) Same as (a) but with ILD = 6 dB (Ipsilateral higher).

(8d) Same as (a) but with ILD = 12 dB (Ipsilateral higher).

Fig.9: Lateralization of an AM tone.

(9a) Cochlear model outputs with an AM tone (Carrier frequency = 3 kHz; Modulation frequency = 100 Hz).

(9b) Stereausis network outputs with the centered AM stimulus.

(9c) Same as (b) but with AM tone binaurally delayed by .5 msec (Ipsilateral leading).

Fig.10: Detection and enhancement of tones in noise.

(10a) Stereausis network outputs with binaurally identical tone (1100 Hz) in noise background (.1-10 kHz) and S/N= -10 dB.

(10b) Same as (a) but with signal reversed in amplitude ($S_{\pi}N_o$).

(10c) Same as (a) but with noise reversed in amplitude (S_oN_{π}).

(10d) Same as (a) but with signal monaurally presented on the ipsilateral side (S_mN_o).

Fig.11: Detection of onset delays.

(11a) The cochlear response patterns to the onset of a 3 kHz tone.

(11b) Stereausis network outputs with identical tone onsets.

(11c) Same as (b) but with onset binaurally delayed (.25 msec).

Fig.12: Schematics of two binaural networks for ITD detection.

(12a) Schematic diagram of the NL in the barn owl as suggested in Sullivan and Konishi (1986). It is functionally equivalent to the model by Jeffress (1948). (12b) Schematic of the NL if the stereausis topology is the underlying network.

References

- F. BILSEN, *Pitch of noise signals: Evidence for a "central spectrum"*, J. Acoust. Soc. Am., 61 (1977), pp. 150–161.
- J. BLAUERT, *Modelling of interaural time and intensity difference discrimination*, in *Psychophysical, physiological, and behavioral studies in hearing*, G. vanden Brink and F. Bilsen, eds., Delft University Press, 1980, pp. 421–424.
- J. BLAUERT, *Binaural localization*, Scand. Audiol. Suppl., 15 (1982).
- J. BLAUERT, *Spatial Hearing*, MIT Press, Cambridge, Massachusetts, 1983.
- E. CHERRY, *Some experiments on the recognition of speech with one or two ears*, J. Acoust. Soc. Am., 25 (1953), pp. 975–979.
- S. COLBURN AND N. I. DURLACH, *Models of binaural interactions*, in *Handbook of Perception*, E. C. Carterette and M. P. Friedman, eds., IV, 1978.
- P. DALLOS, *The Auditory Periphery: Biophysics and Physiology*, Academic Press, New York, 1973.
- B. DELGUTTE, *Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds*, J. Acoust. Soc. Am., 75 (1984), pp. 879–886.
- N. DURLACH, *Binaural signal detection: Equalization and cancellation theory*, in *Foundations of modern auditory theory*, J. Tobias, ed., 2, Academic Press, New York, 1972, pp. 369–462.
- N. DURLACH AND S. COLBURN, *Binaural phenomena*, in *Handbook of Perception*, E. C. Carterette and M. P. Friedman, eds., IV, 1978, pp. 365–466.
- D. M. GREEN AND W. A. YOST, *Binaural analysis*, in *Handbook of Sensory Physiology*, W. D. Keidel and W. D. Neff, eds., 2, Berlin, 1975.
- H. K. HARTLINE, *Studies on Excitation and Inhibition in the Retina*, Rockefeller University Press, New York, 1974.
- H. HIRSCH, *The influence of interaural phase on interaural summation and inhibition*, J. Acoust. Soc. Am., 20 (1948), pp. 536–544.

- M. H. HOLMES AND J. D. COLE, *Cochlear mechanics: analysis for a pure tone*, J. Acoust. Soc. Am., 76 (1984), pp. 767-778.
- L. JEFFRESS, *A place theory of sound localization*, J. Comp. Physiol. Psych., 61 (1948), pp. 468-486.
- D. H. JOHNSON, *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 (1974).
- W. D. KEIDEL AND W. D. NEFF, *Handbook of Sensory Physiology*, Berlin (1975).
- E. KNUDSEN, *Synthesis of a neural map of auditory space in the owl*, in Dynamic aspects of neocortical function, G. Edelman, W. Gall and W. Cowan, eds., A Neurosciences Institute Publication, John Wiley and Sons, New York, 1984, pp. 375-396.
- J. LICKLEIDER, *A duplex theory of pitch perception*, Experientia, 7 (1951), pp. 128-134.
- G. LOEB, M. WHITE AND M. MERZENICH, *Spatial cross-correlation: A proposed mechanism for acoustic pitch perception*, Biol. Cybern., 47 (1983), pp. 149-163.
- D. MARR AND T. POGGIO, *A computational theory of human stereo vision*, Proc. R. Soc. Lond., 204 (1979), pp. 301-328.
- R. R. PFEIFER AND D. O. KIM, *Cochlear nerve fiber responses: Distribution along the cochlear partition*, J. Acoust. Soc. Am., 58 (1975), pp. 867-869.
- B. SAYERS, *Acoustic-image lateralization judgement with binaural tones*, J. Acoust. Soc. Am., 36 (1964), pp. 923-926.
- B. SAYERS AND E. CHERRY, *Mechanisms of binaural fusion in the hearing of speech*, J. Acoust. Soc. Am., 29 (1957), pp. 973-987.
- B. SAYERS AND F. TOOLE, *Acoustic-image lateralization judgments with binaural transients*, J. Acoust. Soc. Am., 36 (1964), pp. 1199-1205.
- M. R. SCHROEDER, *New Viewpoints in binaural interactions*, in Psychophysics and Physiology of Hearing, E. F. Evans and J. P. Wilson, eds., Academic Press, 1977, pp. 455-467.
- S. SENEFF, *Pitch and spectral estimation of speech based on auditory synchrony model*, MIT, Working Papers on Linguistics (1984).

- S. A. SHAMMA, *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 (1985a), pp. 1622–1632.
- S. A. SHAMMA, *Speech processing in the auditory system. I: Representation of speech sounds in the responses of the auditory-nerve*, J. Acoust. Soc. Am., 78 (1985b), pp. 1612–1621.
- S. A. SHAMMA, R. CHADWICK, J. WILBUR, J. RINZEL AND K. MOORISH, *A biophysical model of cochlear processing: intensity dependence of pure tone responses*, J. Acoust. Soc. Am., 1986.
- D. G. SINEX AND C. D. GEISLER, *Responses of auditory-nerve fibers to consonant-vowel syllables*, J. Acoust. Soc. Am., 73 (1983), pp. 602–615.
- W. SULLIVAN AND M. KONISHI, *Neural map of interaural phase difference in the owl's brainstem*, Proc Nat. Acad. Sci., 83 (1986), pp. 8400–8404.
- W. E. SULLIVAN AND M. KONISHI, *Segregation of Stimulus Phase and Intensity Coding in the Cochlear Nucleus of the Barn Owl*, The Journal of Neuroscience, 4 (7) (1984), pp. 1787–1799.
- T. TAKAHASHI, A. MOISEFF AND M. KONISHI, *Time and Intensity Cues are Processed Independently in the Auditory System of the Owl*, The Journal of Neuroscience, 4 (7) (1984), pp. 1781–1786.
- R. WHITWORTH AND L. JEFFRESS, *Time versus intensity in the lateralization of tones*, J. Acoust Soc. Am., 33 (1961), pp. 925–929.
- T. YIN AND S. KUWADA, *Neuronal mechanisms of binaural interactions*, in Dynamic aspects of neocortical function, G. Edelman, W. Gall and W. Cowan, eds., A Neurosciences Institute Publication, John Wiley and Sons, New York, 1984, pp. 263–314.
- E. D. YOUNG AND M. B. SACHS, *Representation of steady state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers*, J. Acoust. Soc. Am., 66 (1979), pp. 1381–1403.

TABLE I: COMPARISON OF MONAURAL AND BINAURAL PROCESSING

<u>Objectives</u>	<u>Monaural</u>	<u>Binaural</u>
Operation	<ul style="list-style-type: none"> • Stimulus spectrum estimation • Sound recognition 	<ul style="list-style-type: none"> • ITD and ILD detection • Localization and signal enhancement
Function		
<u>Temporally-based Processing</u>		
Algorithm	<ul style="list-style-type: none"> • Auto-correlation (Fourier analysis) • Delay lines 	<ul style="list-style-type: none"> • Cross-correlation • Delay lines
Neural implementation		
<u>Spatially-based Processing</u>		
Algorithm	<ul style="list-style-type: none"> • Local cross-fiber correlations (Detection of discontinuities in monaural responses) • Lateral inhibitory network (LIN) 	<ul style="list-style-type: none"> • Local cross-fiber correlations (Detection of disparities between binaural responses) • Stereausis network
Neural implementation		
<u>Corresponding Operation in vision</u>	<ul style="list-style-type: none"> • Edge detection (retinal ON-center/OFF-surround) 	<ul style="list-style-type: none"> • Depth perception (stereopsis networks)

Fig. 1a

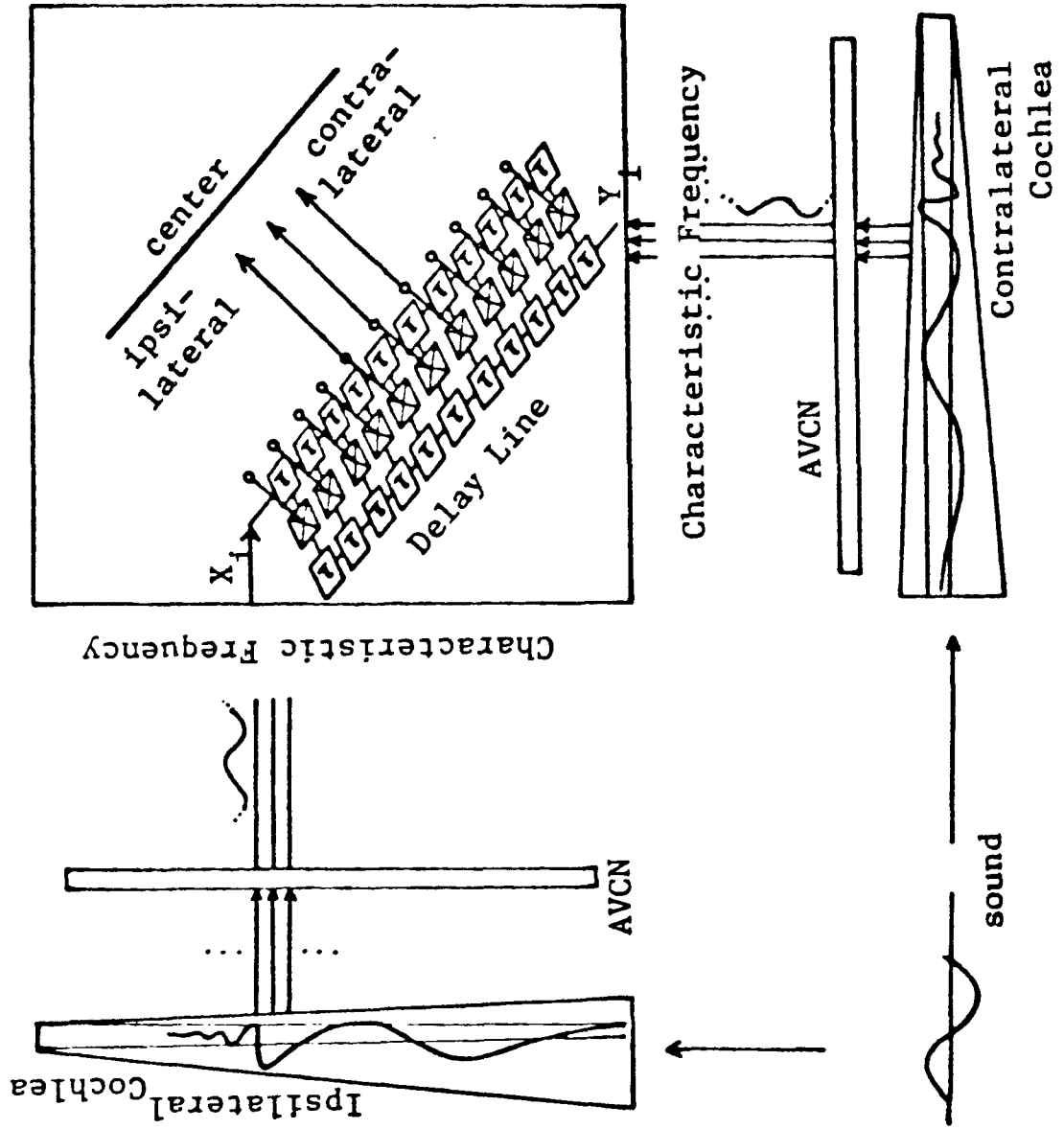


Fig. 1b

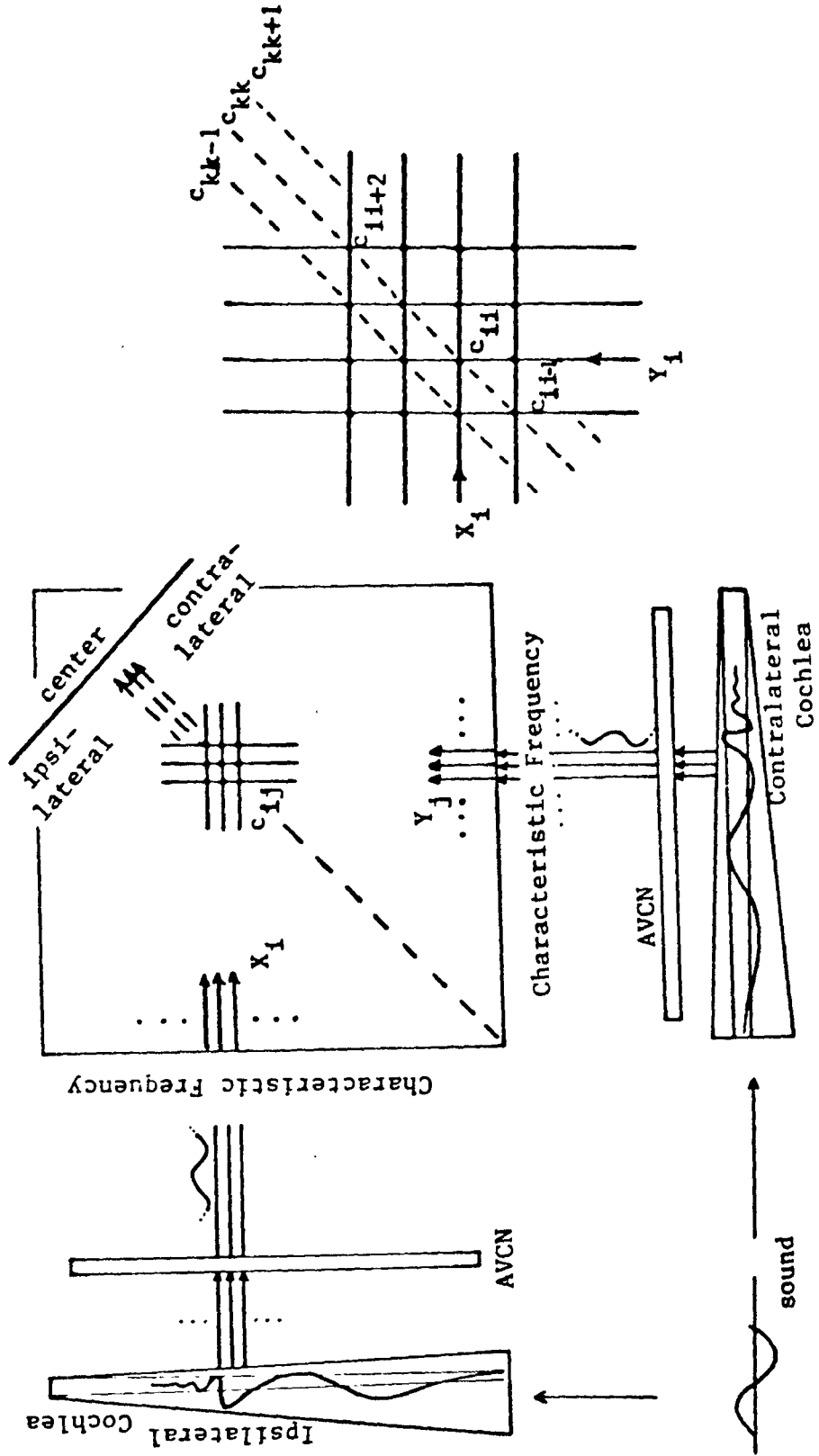


Fig. 2

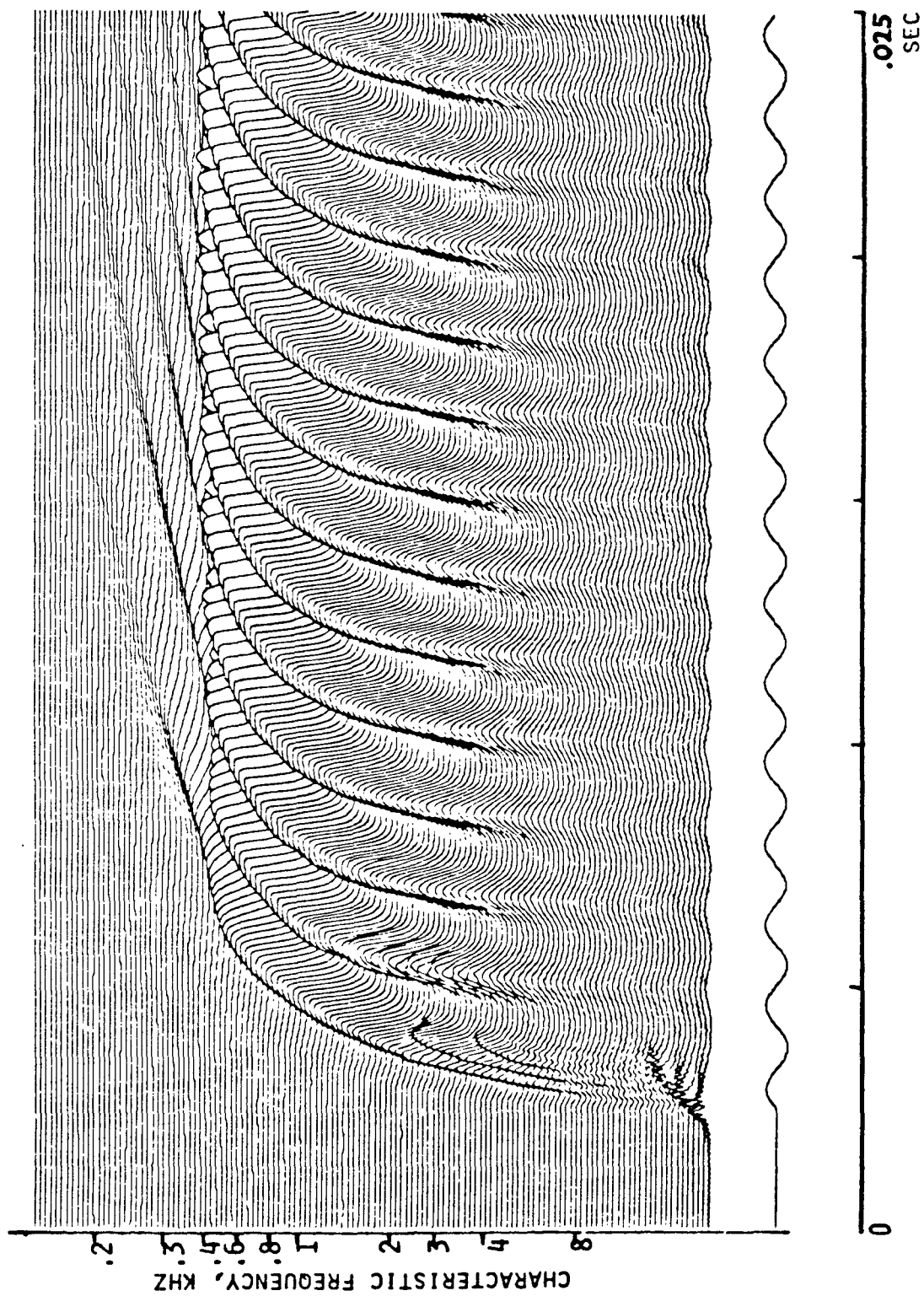


Fig. 3a, b, c, d

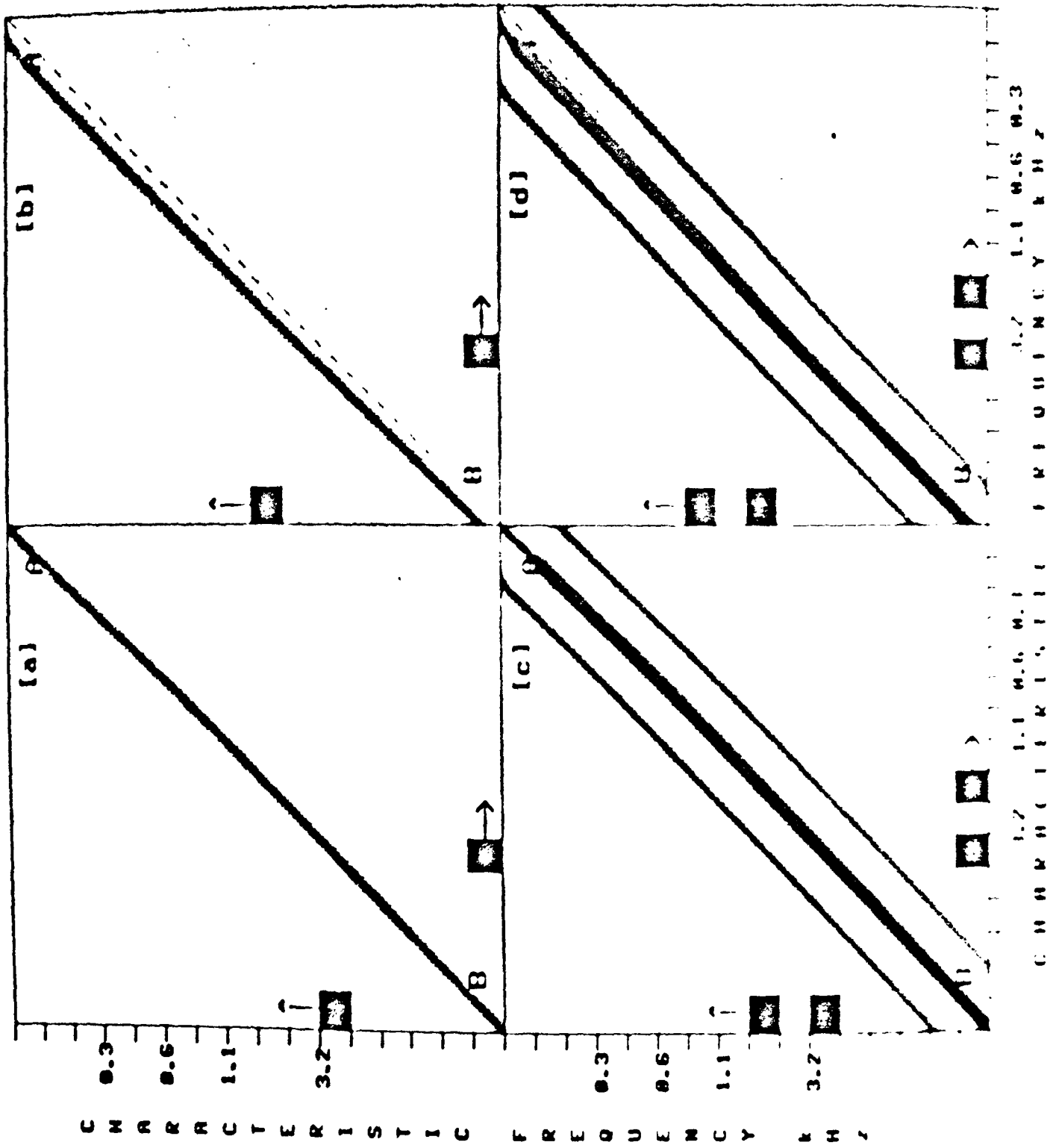


Fig. 4.a

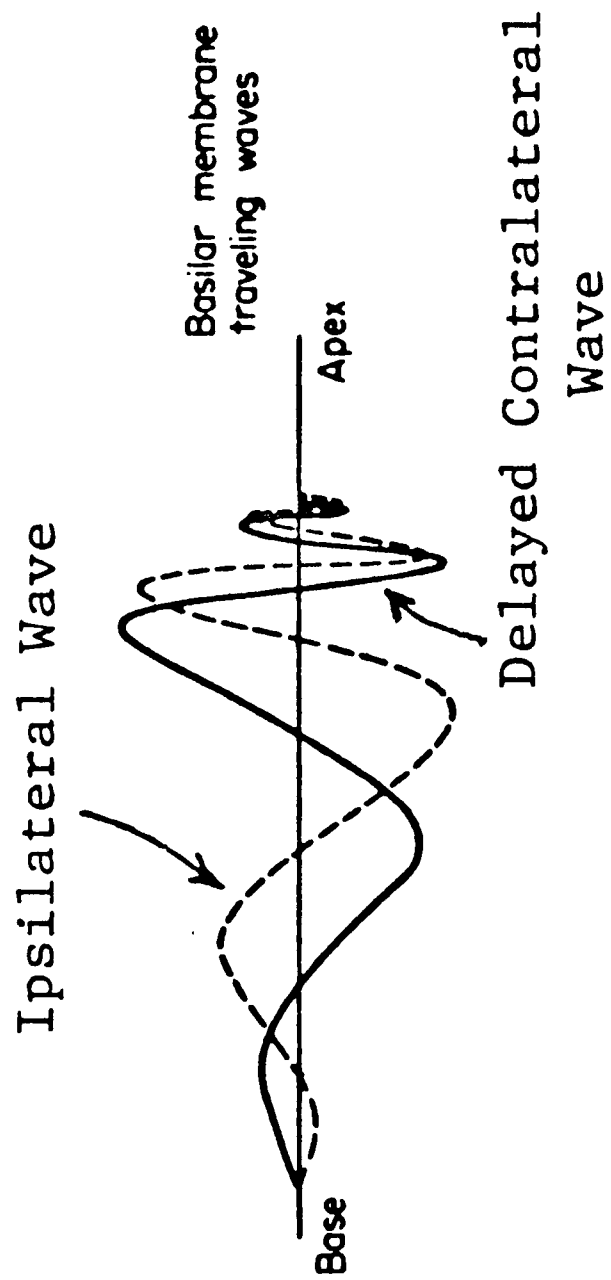


Fig 4b,c

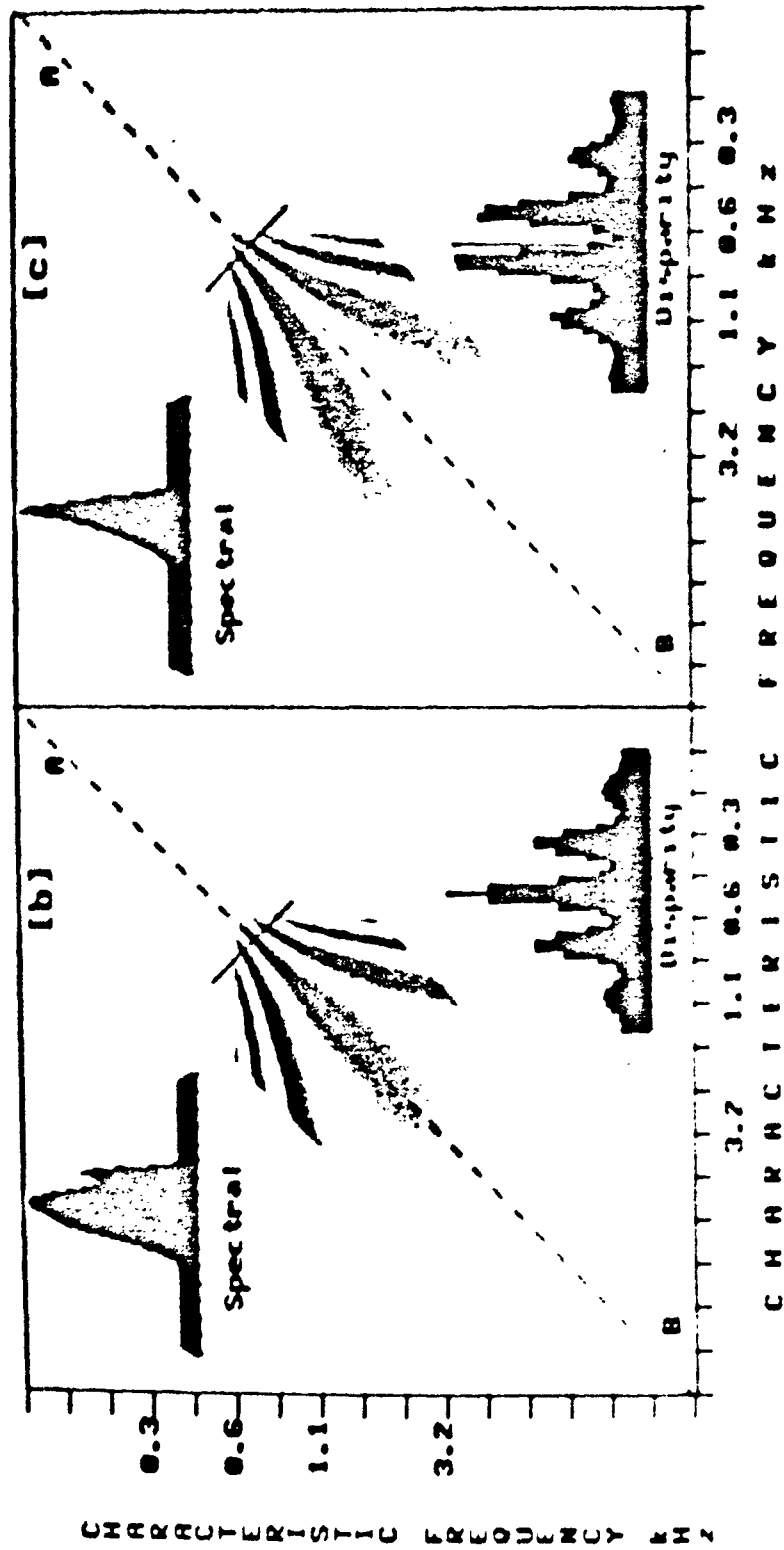


Fig. 5a, b, c, d

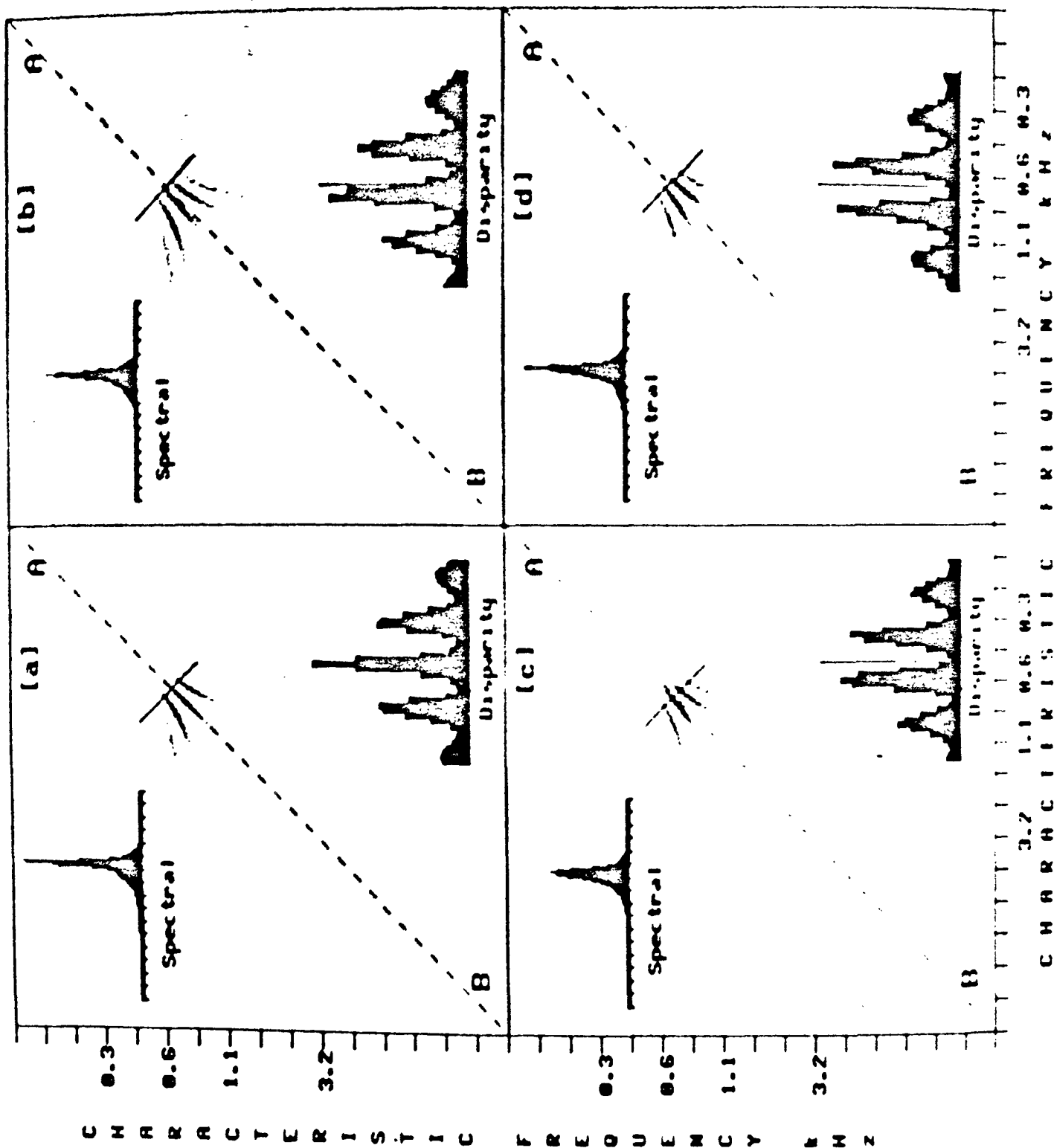


Fig. 6. a, b, c, d

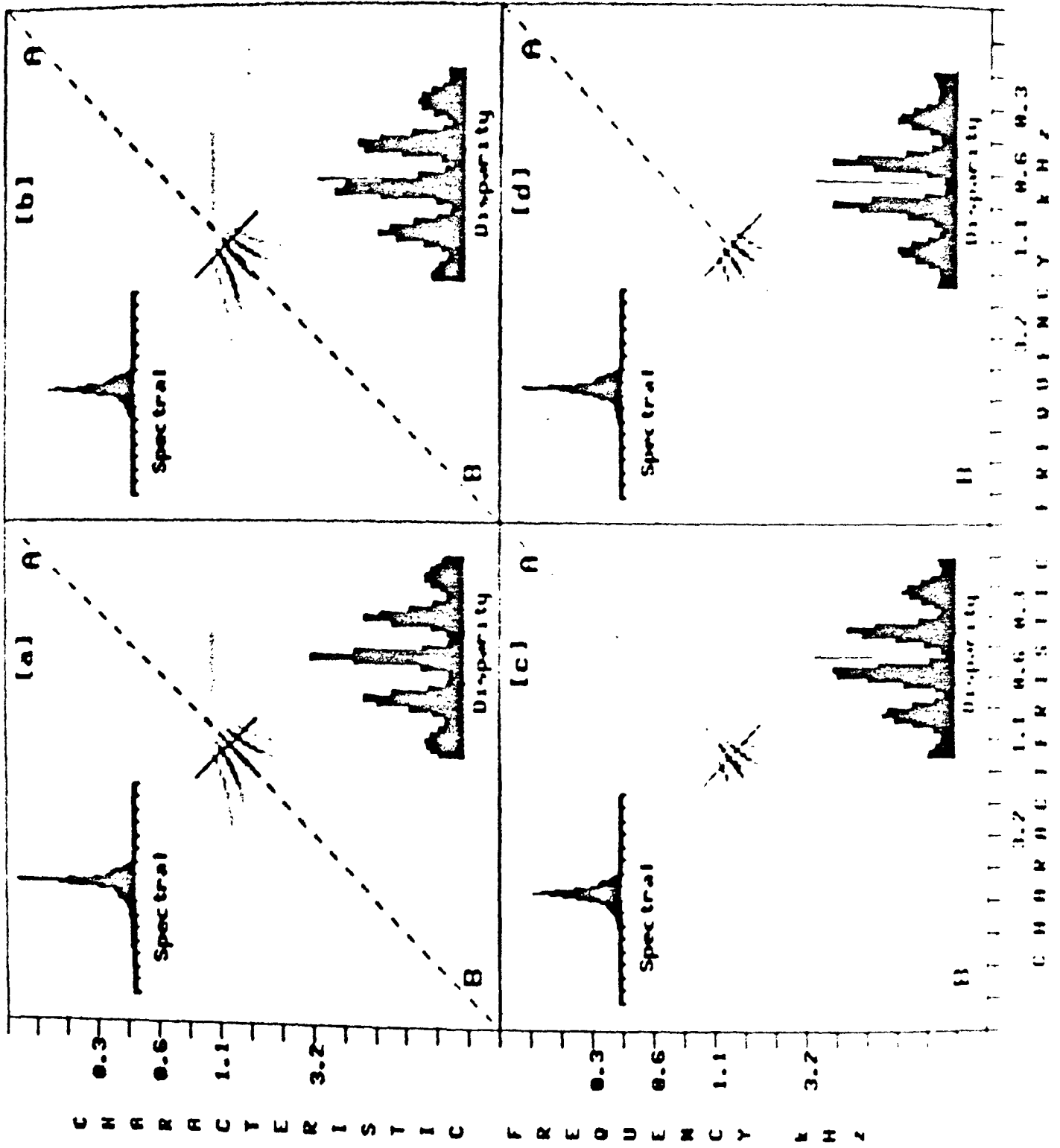


Fig. 7 a, b, c, d

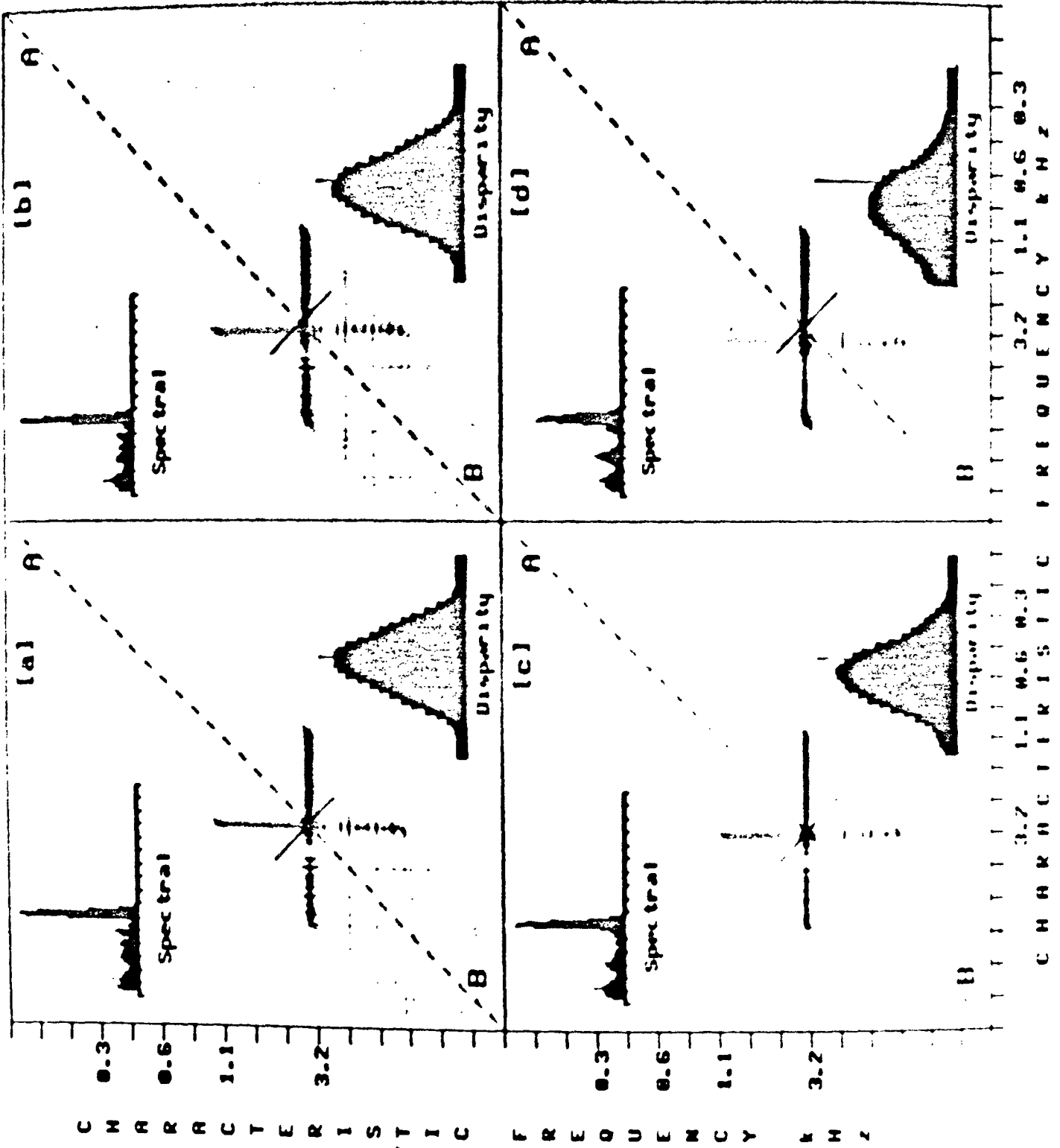


Fig 8. a, b, c, d

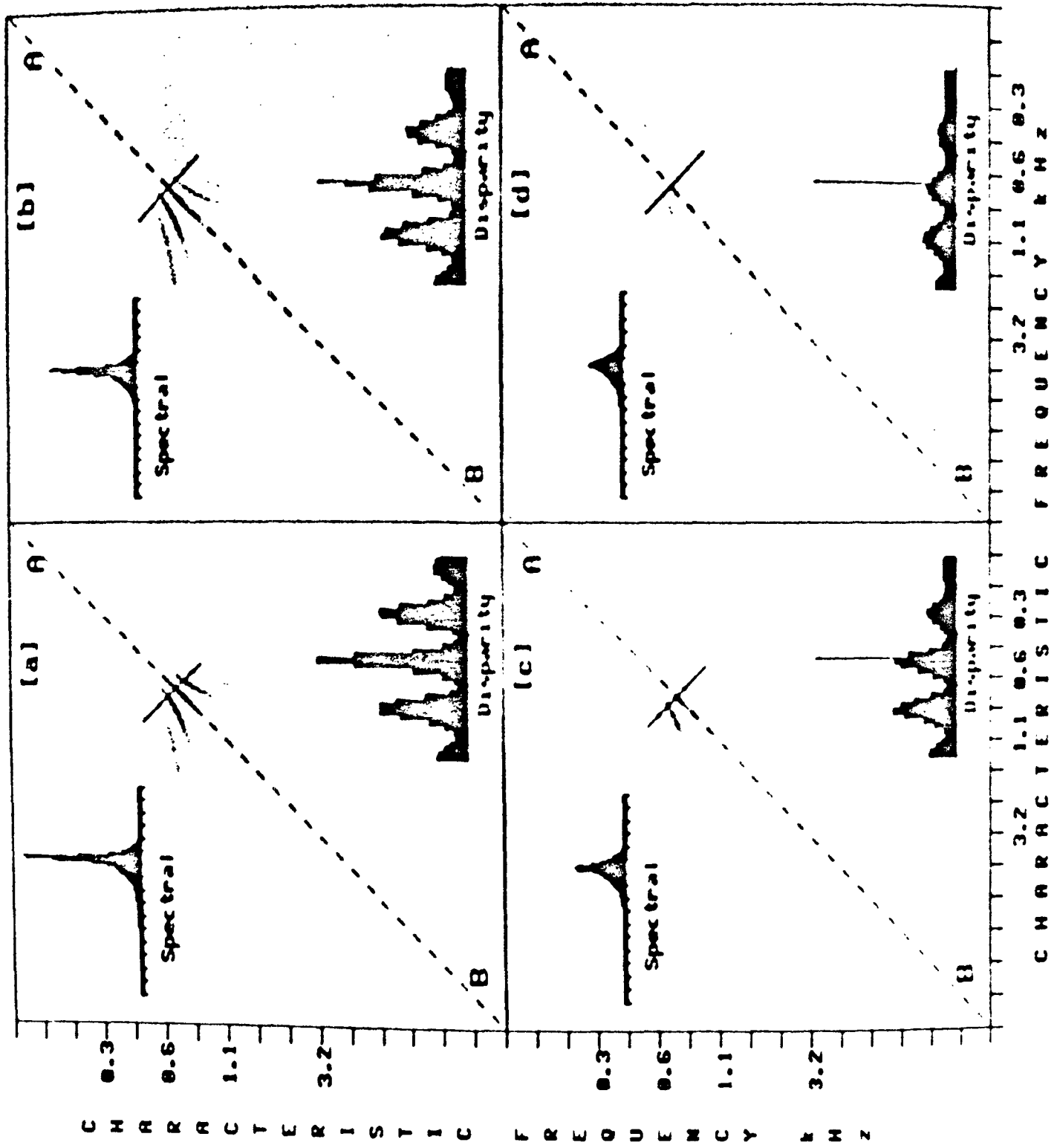


Fig. 9a

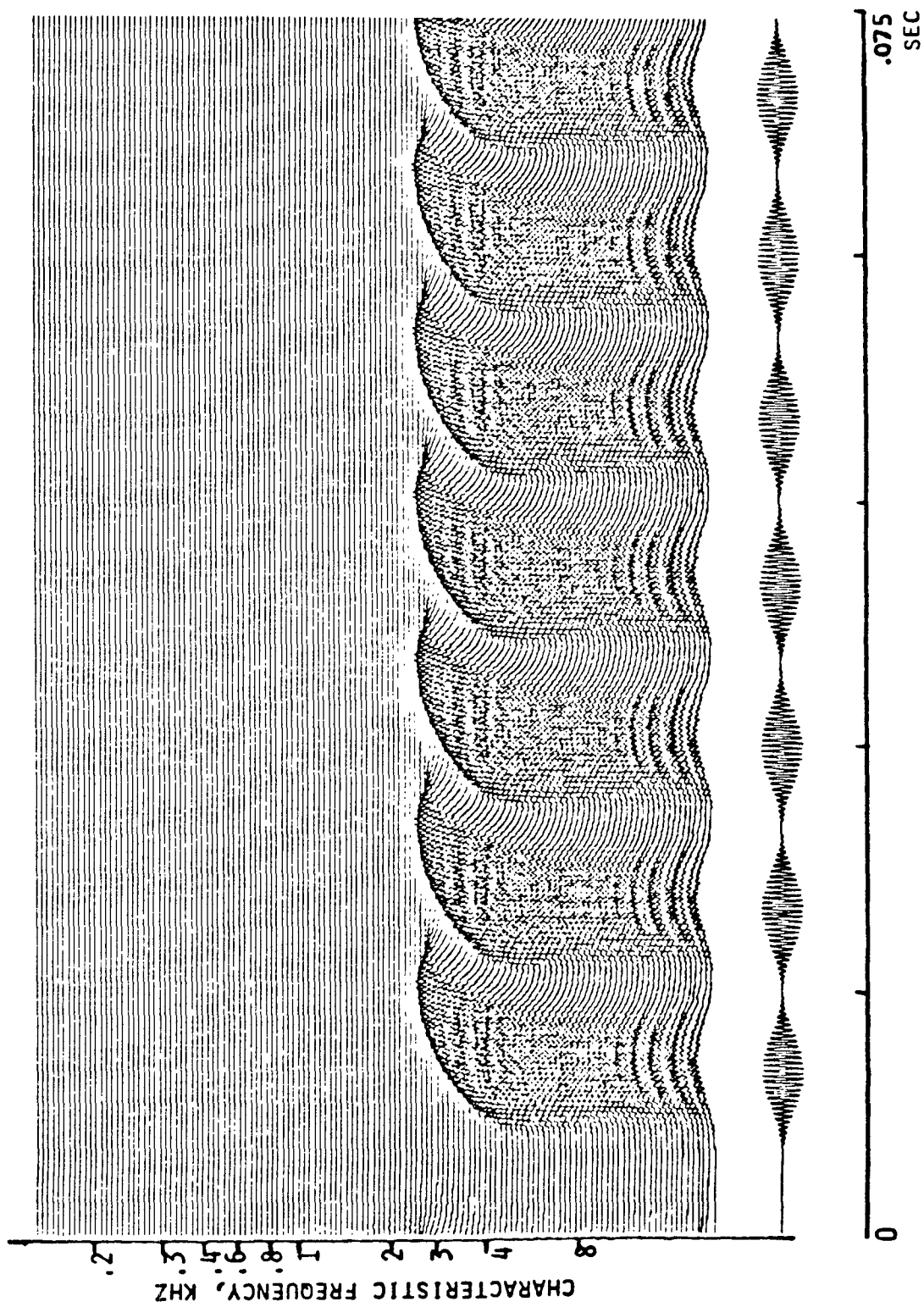


Fig 9 b.c

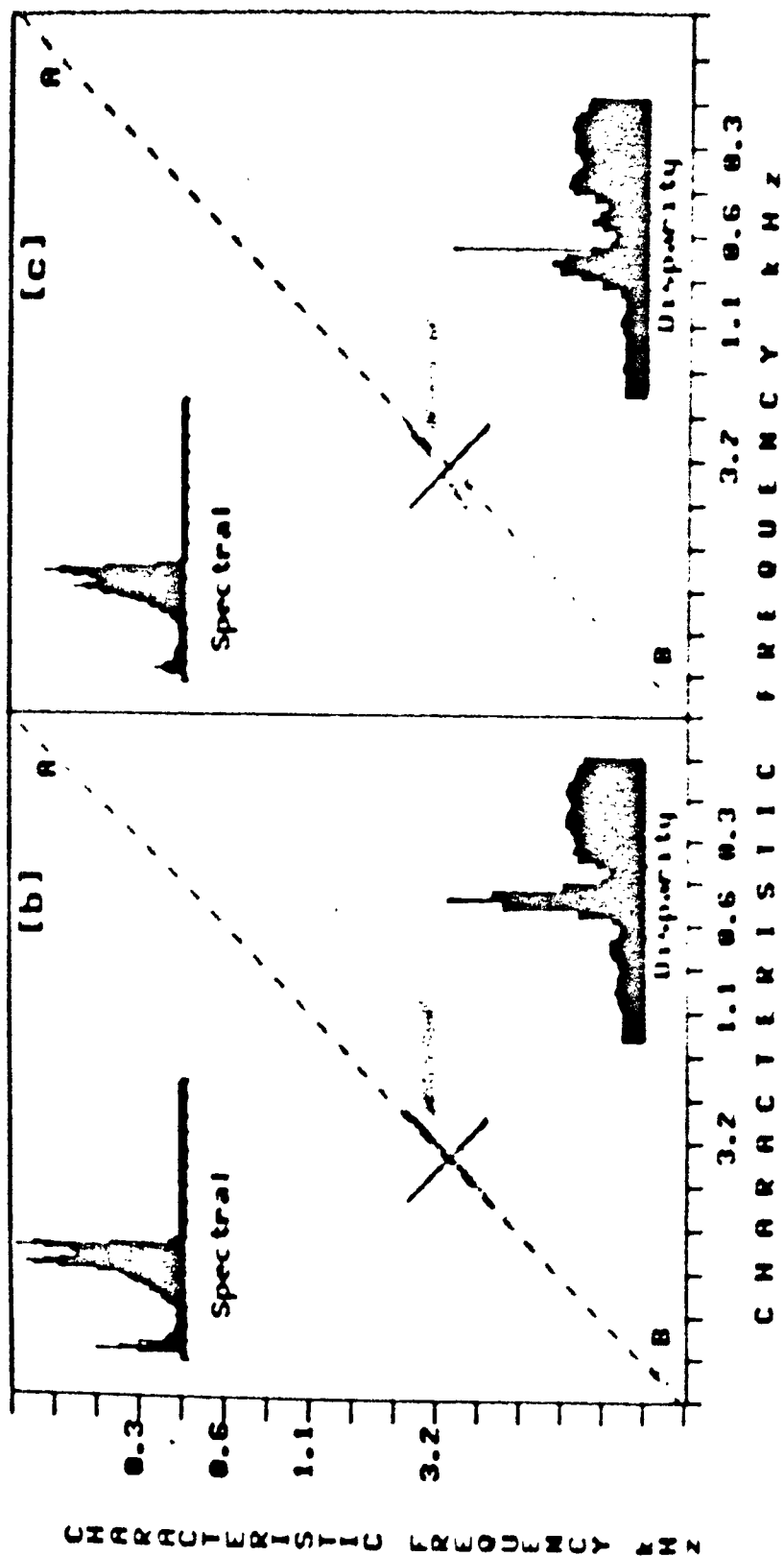


Fig. 10. a, b, c, d

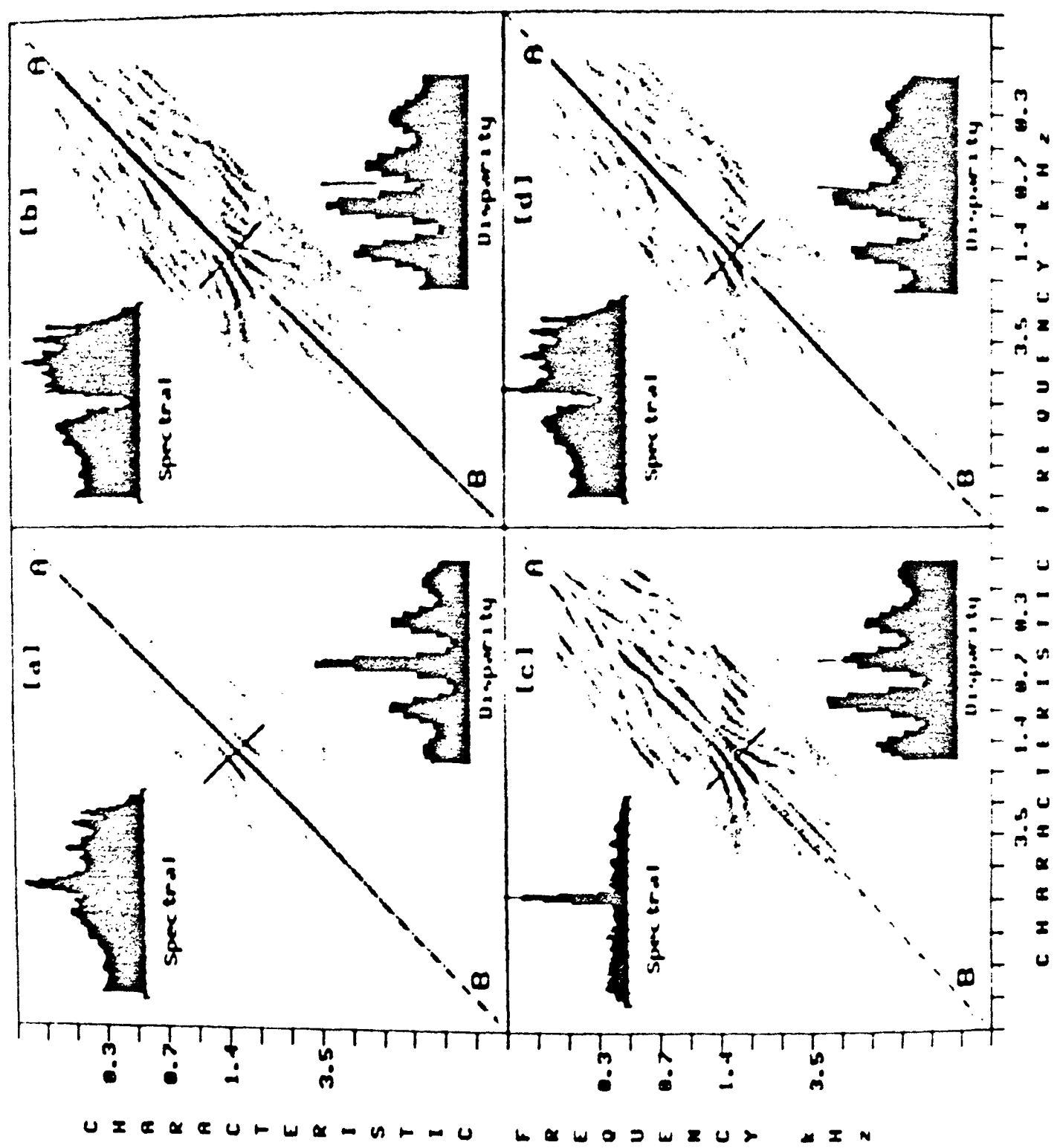


Fig. 11. a

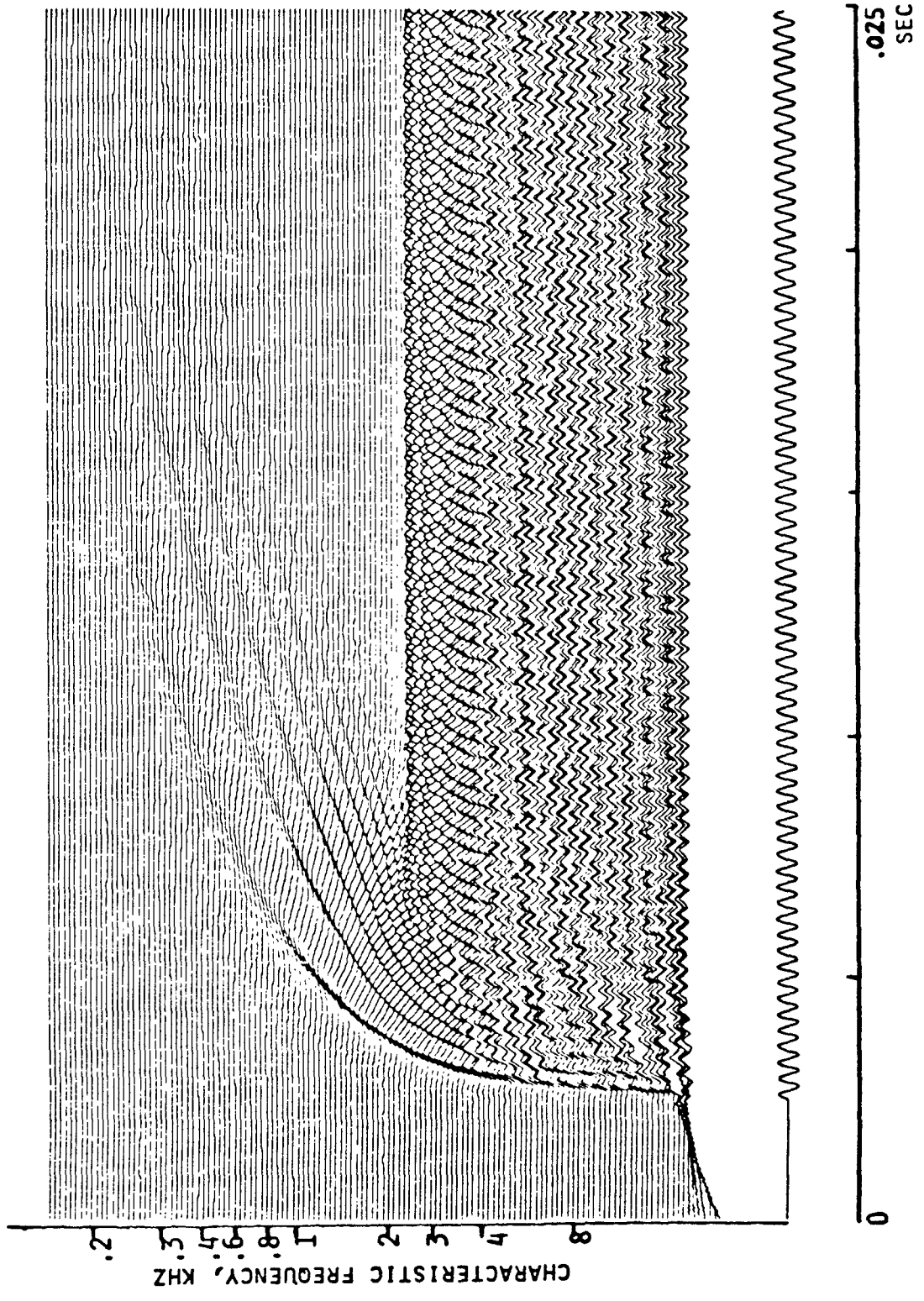


Fig. 11 b, c

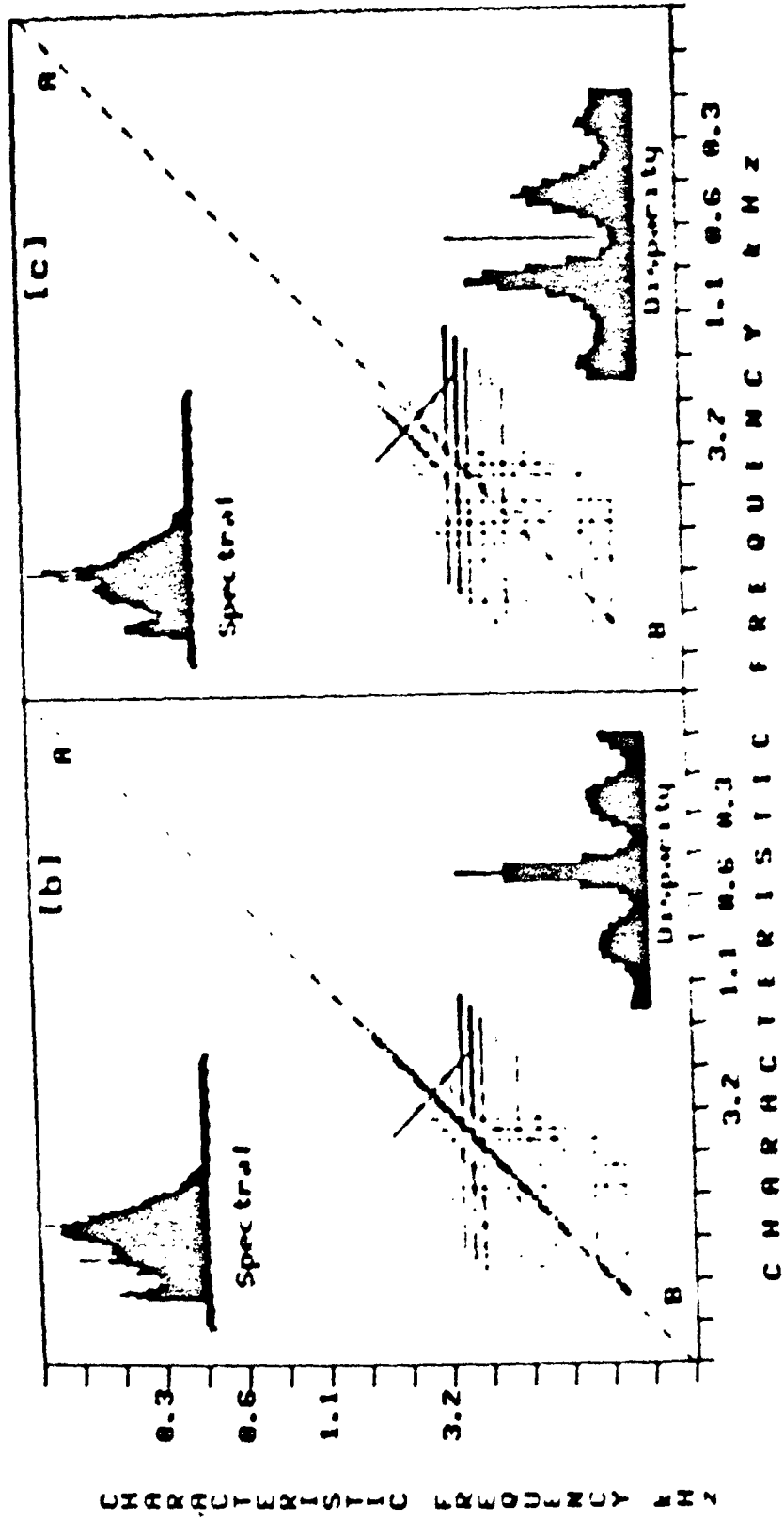


Fig. 12 a, b

