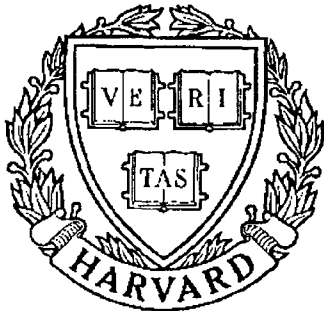


THESIS REPORT
Master's Degree



S Y S T E M S
R E S E A R C H
C E N T E R



*Supported by the
National Science Foundation
Engineering Research Center
Program (NSFD CD 8803012),
Industry and the University*

**Learning Binaural Processing in
Biological Networks**

*by: P. Gopalaswamy
Advisor: S.A. Shamma*

M.S. 89-4
Formerly TR 89-40

ABSTRACT

Title of Thesis: Learning Binaural processing
 in Biological Networks

Name of degree candidate: Preetham Gopalaswamy

Degree and Year: Master of Science, 1988

Thesis directed by: Dr. S. A. Shamma

 Assistant Professor

 Department of Electrical Engineering

It has always intrigued man as to how the human body performs so many complicated functions with the speed and accuracy that it does. One such task is that of sound localization in space, the ability to determine the location of a sound source with considerable accuracy. A biologically realistic neural network is proposed for the binaural processing of interaural time and intensity cues that closely resembles computational schemes suggested for *stereopsis* (depth perception) in vision. The important feature of this network is that it does not use any neural delay lines to generate such attributes of binaural hearing such as lateralization of all audible frequencies and the detection and enhancement of signals in a noisy environment. Temporal shifts between the signals at the ears, arising from sound sources at different locations on the azimuth cause spatial disparities in the corresponding travelling waves set up on the basilar membranes in the two ears. The two dimensional network proposed uses these spatial differences between instantaneous outputs at the two ears to measure interaural differences. The network operation

approximately computes the cross-correlation between the two cochlear outputs by combining the ipsilateral input at a given characteristic frequency(CF) with contralateral inputs from locally Off-CF locations. Some of the results obtained from this network are presented.

Having proposed a network, the next question is whether such a connection is genetically present in the body or whether it is formed over a long period of time by a gradual process of learning. Assuming that the latter solution is more plausible, two learning rules are suggested according to which the network could alter its initial random connectivities. The first learning rule is a supervised technique in which a teaching signal prespecifies the ideal response expected from the network to each input pattern presented. The error between the actual output and the desired response helps to guide the learning process in the desired direction. When the minimum of the error surface is reached, the network is said to have *learned* and the weights do not change any more.

The teaching signal required for the supervised algorithm could be derived from the visual system. However, no physiological evidence exists that links the auditory and visual maps at the level of the olivary complex which is where early binaural processing occurs. To overcome this problem, an unsupervised learning rule is proposed which requires only the cochlear outputs from the two ears. The rule is a competitive learning strategy wherein only one neuron updates its connectivities for a particular input pattern. The neuron chosen to alter its weights is the one which responds maximally to the input. The inherent delays that exist in the neural system are used as guides to form the organized spatial map responses.

Learning Binaural Processing
in
Biological Networks

by
Preetham Gopalaswamy

Thesis submitted to the Faculty of The Graduate School
of The University Of Maryland in partial fulfillment
of the requirements for the degree of
Master of Science
1988

Advisory Committee:

Assistant Professor S. A. Shamma Chairman/Advisor
Professor R. Newcomb
Associate Professor V. Gligor

TABLE OF CONTENTS

1	INTRODUCTION	1
2	OVERVIEW OF BINAURAL INTERACTION MODELS	4
I	Introduction	4
II	Count Comparison Models	5
III	Interaural Difference Detector Models	7
IV	Noise Suppression Models	8
V	Correlation Models	9
VI	Conclusion	11
3	STEREAUSIS : PROCESSING WITHOUT NEURAL DELAYS	13
I	Introduction	13
II	The Network Input Patterns	14
III	The Stereausis Network	19
IV	Binaural processing in the Stereausis network	26
V	Concluding Remarks	30

4	INTRODUCTION TO NEURAL NETWORKS	33
I	Introduction	33
II	The neuron model	33
III	Neural Network Connections and Architectures	37
IV	Learning in neural networks	39
V	Conclusions	42
5	A SUPERVISED LEARNING SOLUTION	44
I	Statement of the Problem	44
II	The Learning algorithm	45
III	Implementation Results	51
IV	Conclusion	54
6	AN UNSUPERVISED LEARNING SOLUTION	56
I	Statement of the Problem	56
II	The Learning Algorithm	57
III	Implementation Results	65
IV	Conclusion	69
7	CONCLUDING REMARKS	70
8	BIBLIOGRAPHY	73

LIST OF FIGURES

2.1	Bekesy's tuned cells	5
2.2	Schematic of Van Berjeik's neuron model	6
2.3	The Jeffress model	7
2.4	Webster's Difference Detector model	9
2.5	Durlach's Noise Suppression model	10
2.6	The Sayers and Cherry model	10
2.7	A Model summarizing common Binaural model aspects	12
3.1	A Neural delay based Binaural model	15
3.2	Envelope of the traveling waves for 600Hz tone	16
3.3	Cochlear Model used in the processing	18
3.4	Traveling waves on the basilar membrane : 600Hz tone	19
3.5	The Stereausis Network	22
3.6	Network response to pulses traveling across the membrane	26
3.7	Network response to Low frequency tone ITDs	28
3.8	Time Intensity trading at low frequencies	31
4.1	A Neuron based system	34

4.2	Feedforward Architecture in neurons	37
4.3	Feedback Architecture in neurons	39
5.1	Travelling peaks on the basilar membranes	48
5.2	Peak amplitudes on the basilar membranes in time	50
5.3	Connectivities obtained when binaural neurons have no time shifts .	51
5.4	Sufficient neural delays are present to compensate the external shifts	53
5.5	Connectivities when neural delays present are not sufficiently large .	55
6.1	Output pattern of Binaural neurons with different delays for different delays	57
6.2	Corresponding peaks on the contralateral and ipsilateral cochlea . .	63
6.3	Connectivity pattern got when every input shift is compensated by the internal delay of a neuron	67
6.4	Weights obtained when more input shifts are presented than output neurons	68

INTRODUCTION

The ability of animals to identify, localize and isolate a sound source in their normally noisy acoustic environment depends significantly on their having two ears. The advantages of two ears are best appreciated in experiments where monaural and binaural recordings of acoustic events are presented to normal hearing subjects through microphones placed in the ear canals. While attempting to recreate the original event, a sharp improvement in the ability of the subject to accurately perceive the event was seen in the binaural presentation over the monaural case. While, in the binaural case, sounds can be easily localized with reasonable accuracy, the monaural experiments cause the sounds to seem fused and as though originating from the stimulated ear. The processing of binaural cues is fundamental to several tasks in spatial hearing of which lateralization and signal detection-enhancement are probably the most important. Extensive work has been done in these areas and several models proposed to explain these attributes of hearing. The most successful of these models in terms of explaining some of the binaural phenomenon have been the correlation models ([19], [32]). Binaural sound localization and enhancement requires the processing of two types of interaural cues, timing differences, and intensity differences. While at lower frequencies, timing differences play an impor-

tant role, level differences come into play as the primary cue at higher frequencies. This report mainly considers the timing differences which are very prominent at lower frequencies. The correlation models consider the issue of lateralization as an engineering problem. The signals from the two ears are cross-correlated in time to determine the temporal disparity between them, which then maps into locations on the azimuth. It is not known if sufficient neural delays exist in the human physiology to be able to perform an accurate correlation of the two signals. In view of this disadvantage with the cross correlation models this report first proposes a binaural processing network that is able to explain several experimentally observed binaural phenomena without using any neural delays. Due to the similarity this network seems to have to the kind of processing in depth perception in studies on vision i.e *stereopsis* ([25]), the proposed network is called the *stereausis* network.

Having proposed a network, the question that immediately arises is whether such a neural network like the one proposed is genetically formed or whether it is acquired by a gradual learning over a long period of time. Assuming an initial random pattern of weights, the second portion of this thesis attempts to propose learning rules by which a network can be formed which can accurately localize sound. The learning rules were implemented only on time shifted low frequency inputs presented at the two ears. First, a supervised learning rule is presented which assumes that during the learning process, a teaching input exists which pre-determines the location of the sound source. If the presence of such a teaching signal cannot be ascertained through physiological experiments, then an unsupervised training rule would be required which does not make any prior assumptions

as to the presence of any additional reinforcing signals to guide the learning process in the correct direction. This report proposes such a learning algorithm where the only external data input available to the network is the basilar membrane outputs from the two ears.

The succeeding chapters in the thesis are arranged as follows :

Chapter 2 is a brief overview of some of the different models that have been proposed till now to explain binaural phenomenon. Chapter 3 explains the *stereausis* network topology and some of the data that was obtained from its implementation. The *stereausis* network was proposed and tested by Naiming Shen as part of his Master's thesis ([36]) and he was assisted by the author in the implementation. Having described the proposed network, the report enters the area of neurons and network learning. Chapter 4 is a description of the neuron model used in later analysis and an introduction to some of the neural network architectures. Chapter 5 elaborates on the supervised learning rule and explains the significance of the results which were obtained when the algorithm was tested on data from the cochlear model. This is followed in Chapter 6 by a different approach to the same problem, where an unsupervised competitive learning algorithm is proposed. The report concludes with a summary of the work that has been done and a look at the scope for future research in the area.

OVERVIEW OF BINAURAL INTERACTION MODELS

I Introduction

Several perceptual attributes of spatial hearing require the processing of binaural cues, particularly signal detection and enhancement and lateralization. These have been the topic of intensive research for several decades ([4], [11], [14]). A number of models have been proposed to explain the various aspects of spatial hearing. The succeeding sections of this chapter will discuss some of the models which have been proposed emphasizing on the methods used to process the input signals from the two ears to highlight and extract the differences between them.

The models that have been proposed till now ([11]) for binaural lateralization and tone enhancement fall under four basic categories :

1. Count-comparison models.
2. Interaural Difference Detector models.
3. Noise Suppression models.
4. Correlation based models.

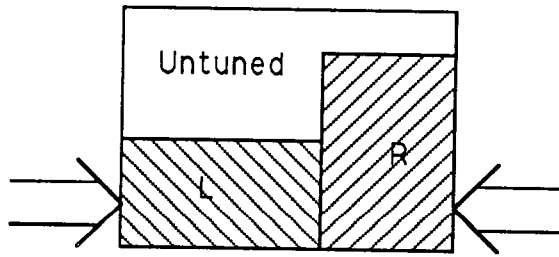


Figure 2.1: Bekesy's tuned cells

II Count Comparison Models

These lateralization models determine the image position based upon a comparison of activity levels in the neurons from the ipsilateral and contralateral ears. Models of this type have been proposed since the early 1900's by several people including Bekesy ([3]) and Bergeijk ([37]). The first model, proposed by Bekesy, describes a group of neurons excited by stimuli from both ears. Depending upon the source of the excitation, the cells are tuned either to the *left* or to the *right*. A comparison of the number of cells tuned to each direction provides the information required for lateralization. If the input stimuli contained an interaural time delay, the excitation from the delayed input would be extinguished before it reaches the center of the neuron population. There would be more cells tuned to the opposite direction and the image would be identified as being on the side whose stimulus arrived earlier. Bekesy proposed that the interaural intensity difference is determined directly from the relative magnitudes of the excitations at each ear. According to his hypothesis, more cells are tuned by a stronger excitation wave, and thus for a given delay the number of cells tuned to either side is proportional to the intensity. Note that the

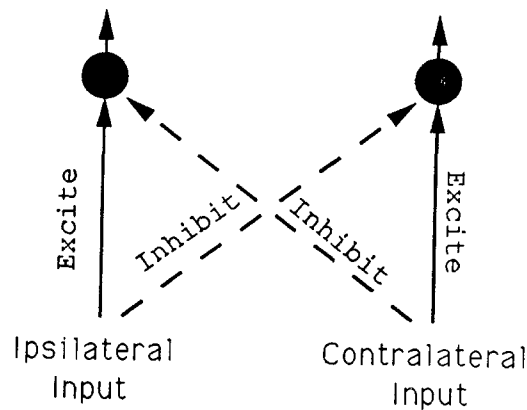


Figure 2.2: Schematic of Van Bergeijk's neuron model

extinction point for a particular delay is still the same. Thus a time delay in a ear can be compensated by an intensity increase in that ear thus explaining the time-intensity trading effect. Figure 2.1 shows a schematic diagram of tuned cells as proposed by Bekesy. A comparison of the areas of the two shaded regions localizes the sound. Van Bergeijk restated Bekesy's model by making it structurally more specific. He assumed that localization occurs at a peripheral pair of nuclei and it is done by a comparison of the number of neural firings in the left nucleus to that in the right nucleus. For each nucleus ipsilateral inputs are inhibitory and contralateral inputs are excitatory with the image being lateralized to the side opposite to the maximally firing nucleus. A schematic of the network he proposed can be seen in figure 2.2. This theory is based on the fact that an inhibitory input arriving before an excitatory input will prevent the neuron from responding to the latter input. Thus identical inputs result in an equal number of neurons firing in each nucleus, creating a centered image; if the left input leads the right, it is immediately obvious that more neurons are excited in the right nucleus than the left, and so the

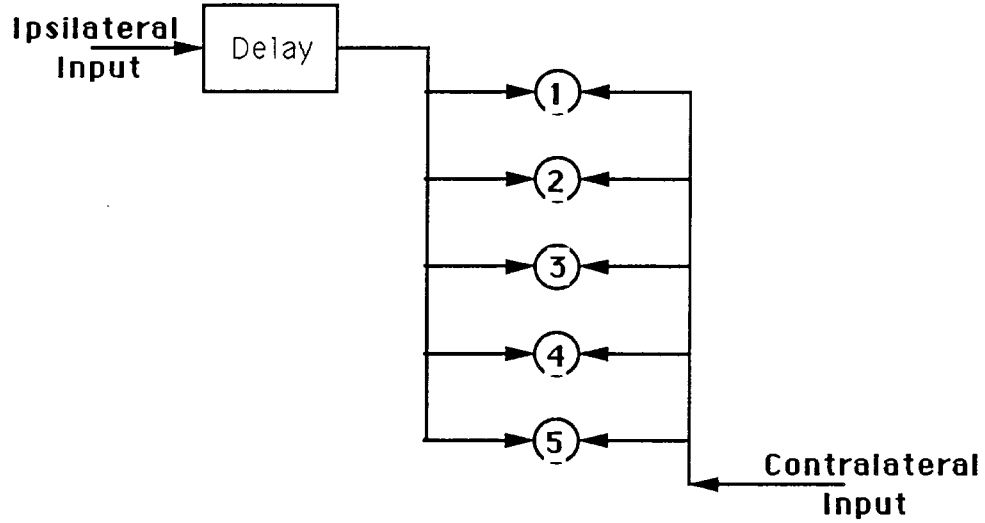


Figure 2.3: The Jeffress model

image is localized on the left. The count-comparison models are qualitative and are restricted to only one binaural phenomenon namely lateralization.

III Interaural Difference Detector Models

The models under this category measure interaural differences in time and/or amplitude. The pioneer in these kind of models was the Jeffress model ([19]). While some of the models in this category explain the lateralization phenomena, others explain detection; but none are able to satisfactorily describe all the aspects of binaural hearing. Jeffress proposed a neural network which coded ITD's into the average rates of firing of neurons. The idea behind the network is shown in figure 2.3. The ipsilateral input is delayed by an amount so that inputs from both ears arrive at corresponding locations A and B at the same time instant. Neurons 1, 2, 3, 4 and 5 are excited maximally when both exciting stimuli arrive simultaneously. Thus it is obvious that depending upon whether the ipsilateral

or contralateral stimulus arrives first, different output neurons are excited. Inputs arriving simultaneously to the two ears would excite neuron 3 while stimuli which lead in the ipsilateral ear would stimulate either neuron 4 or 5 in the figure depending upon the magnitude of the ITD. Jeffress hypothesized that interaural intensity differences are coded into time delays in that a higher intensity stimulus would propagate faster than a lower intensity one. Thus assuming no ITD, if the ipsilateral input were of greater intensity, the impulses would arrive earlier at A than at B. This hypothesis called the *Latency Hypothesis* combined with the topology of the Jeffress network could explain lateralization based on ITD's and ILD's, as well as time-intensity trading effects. To extend the Jeffress model to masking phenomena, Webster ([39]) proposed a variation by introducing a bandpass filter centered at the center frequency of the target signal before the signal is processed in any way, and a monaural processor to explain the case when the interaural differences between the signal and the masker are different and the same respectively. See figure 2.4. Several other variants to the Jeffress model were proposed by Durlach ([9]), Hafter and Carrier ([16]), and Colburn ([5], [7], [6]).

IV Noise Suppression Models

Models in this category are aimed at detecting narrow band signals in broadband masking noise. The underlying principle is to operate on the two input signals so as to cancel out the masking as is done in the *Equalization and Cancellation* model proposed by Kock ([21]). Details on further modifications to the original model are discussed by Durlach ([10]). The basic idea of the model is to adjust the received

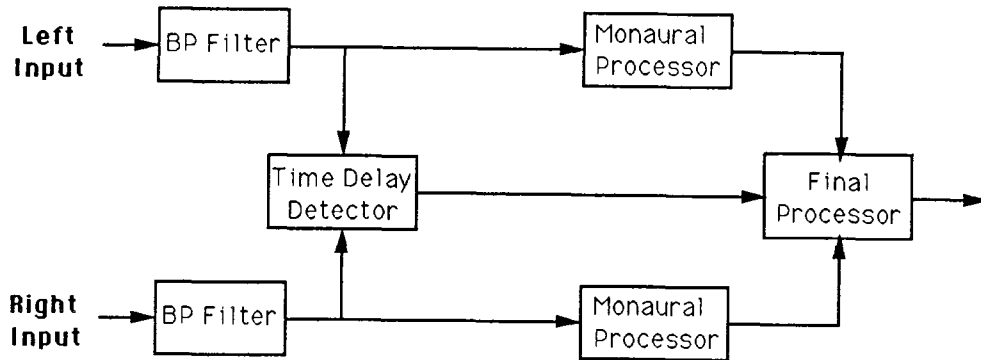


Figure 2.4: Webster's Difference Detector model

signals so that the masking components are equal in the two channels and then to subtract that component out. If the interaural differences of the masker and signal are different, the output would contain only the component of the target signal. The processing is assumed to be corrupted by random jitters which are characterized in terms of two statistically independent random variables, a gain factor $(1 - \epsilon)$ and a time delay δ . A block diagram of the model as proposed by Durlach is shown in figure 2.5. Detection is assumed to be characterized by the channel on which the signal to noise ratio is maximum.

V Correlation Models

The first model to describe binaural phenomena in terms of interaural correlation was presented by Sayers and Cherry ([10]). The models in this category are similar in flavor to the Jeffress model discussed earlier, but are more specific thus allowing comparison to behavioral data. Sayers and Cherry rejected the latency hypothesis used by Jeffress because they noted that an ILD affected the whole function

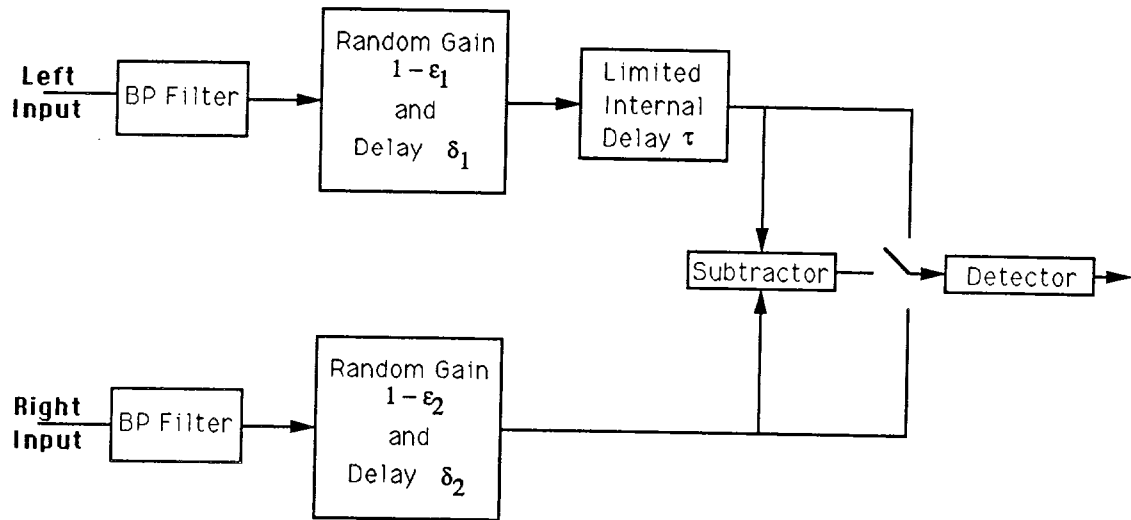


Figure 2.5: Durlach's Noise Suppression model

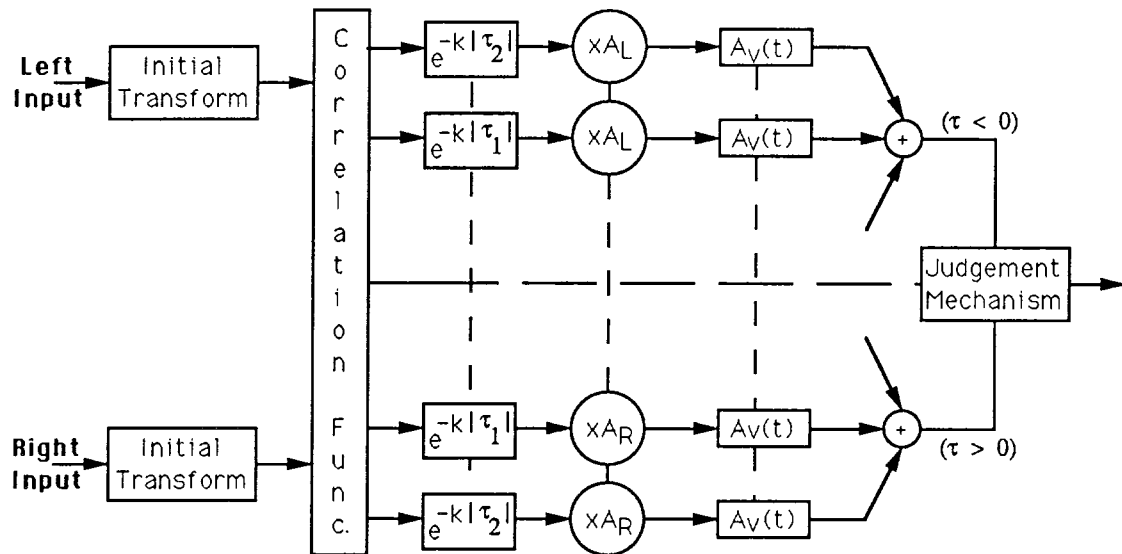


Figure 2.6: The Sayers and Cherry model

describing the dependence of lateralization on time delay towards the ear receiving the stronger intensity signal. The correlation model proposed by Sayers and Cherry is shown in figure 2.6. In this model the input signal is first transformed by adding a dc term proportional to the signal level to it so that the running cross-correlation function of the transformed signals is always positive, and has a dc level proportional to the input signal levels. The cross correlation is then performed and weighted by delay and the other on the polarity of the delay i.e A_L for $\tau < 0$ and A_V for $\tau > 0$. This weighted correlation function is then averaged over time. The position judgement is finally made by comparing the integral of this final over positive values of τ with the value got from the negative values of τ . Dolan and Robinson ([8]) and McFadden ([26]) considered the assumption that binaural unmasking for a fixed target signal is determined by the correlation of the noise and also introduced additive internal noise to explain the dependence of unmasking on the overall and interaural amplitude ratio of the external noise. While correlation based models seem to explain several of the binaural phenomena, they make a fundamental assumption of the presence of neural delays, the presence of which have not been confirmed.

VI Conclusion

All the models that been propose above are unable to explain more than a small portion of all the existing data on binaural interactions. Some of them are unable to relate assumptions or parameters used, to known physiological results. If a model were to be made summarizing the elements common to almost all the models

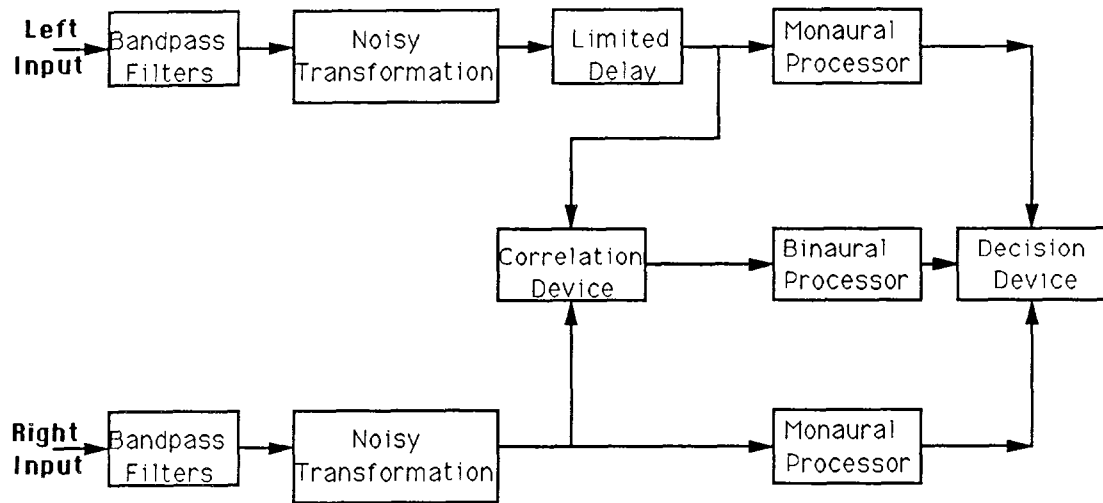


Figure 2.7: A Model summarizing common Binaural model aspects

proposed till now then it would look like the one shown in figure 2.7 ([11]).

STEREAUSIS : PROCESSING WITHOUT NEURAL DELAYS

I Introduction

As was described in the previous chapter, several models have been proposed to explain binaural phenomena, of which the most successful has been the Sayers and Cherry cross correlation model ([32]) which is based on the early Jeffress network ([19]). In an attempt to biologically justify the mathematical processing in the correlation models, the most common assumption has been to associate the various lags in the computations with “*neural delays*” e.g neuronal pathways of different lengths or latency effects. A typical network based on these principles is shown in figure 3.1. It can be seen that following the frequency analysis of the cochlea, each fiber projects to the cross-correlator with a precise topologically ordered range of delays that allows its correlation with the output from the other cochlea of the fiber with the same characteristic frequency(CF). While it may not be tenable to assume a total absence of neural delays in the binaural processing system, it seems incorrect to assume that sufficient precise delays exist to account for the time difference of 750 microseconds difference that can exist between the signals to the two ears especially since there is as yet no direct anatomical or physiological

support as to the existence of such a network or of the essential functional relevance of neural delays. Therefore, it should be emphasized that the idea of a cross-correlation model was not proposed because of compelling experimental evidence but rather to satisfy a literal interpretation of a convenient mathematic function (the correlation function) coupled with a highly schematic view of the cochlea as a bank of extremely sharply tuned bandpass filters which respond to only one frequency. This view of the basilar membrane is inexact and, as shall be elaborated later, several features of the spatio-temporal response are ignored which, if used, could make neural delays totally redundant.

In the following sections, a biologically realistic neural network is proposed which accounts for two important aspects of binaural hearing - lateralization and signal enhancement, without the use of any neural delays. On account of the fundamental similarities that emerge between this network and the type of computations proposed for “stereopsis” in vision ([25]), this network is called the “stereausis” network. Section II will deal with the spatiotemporal responses of the basilar membrane which are the input to the sterausis network; Section III will discuss the topology and details of the network.

II The Network Input Patterns

Auditory stimuli at the external ear set up a series of complex patterns on the basilar membrane which depend upon the spectrum and intensity of the sound stimulus. The stimulus impinges on the tympanic membrane (eardrum) which in turn sets up a vibration in the bones of the middle ear (malleus, incus and stapes).

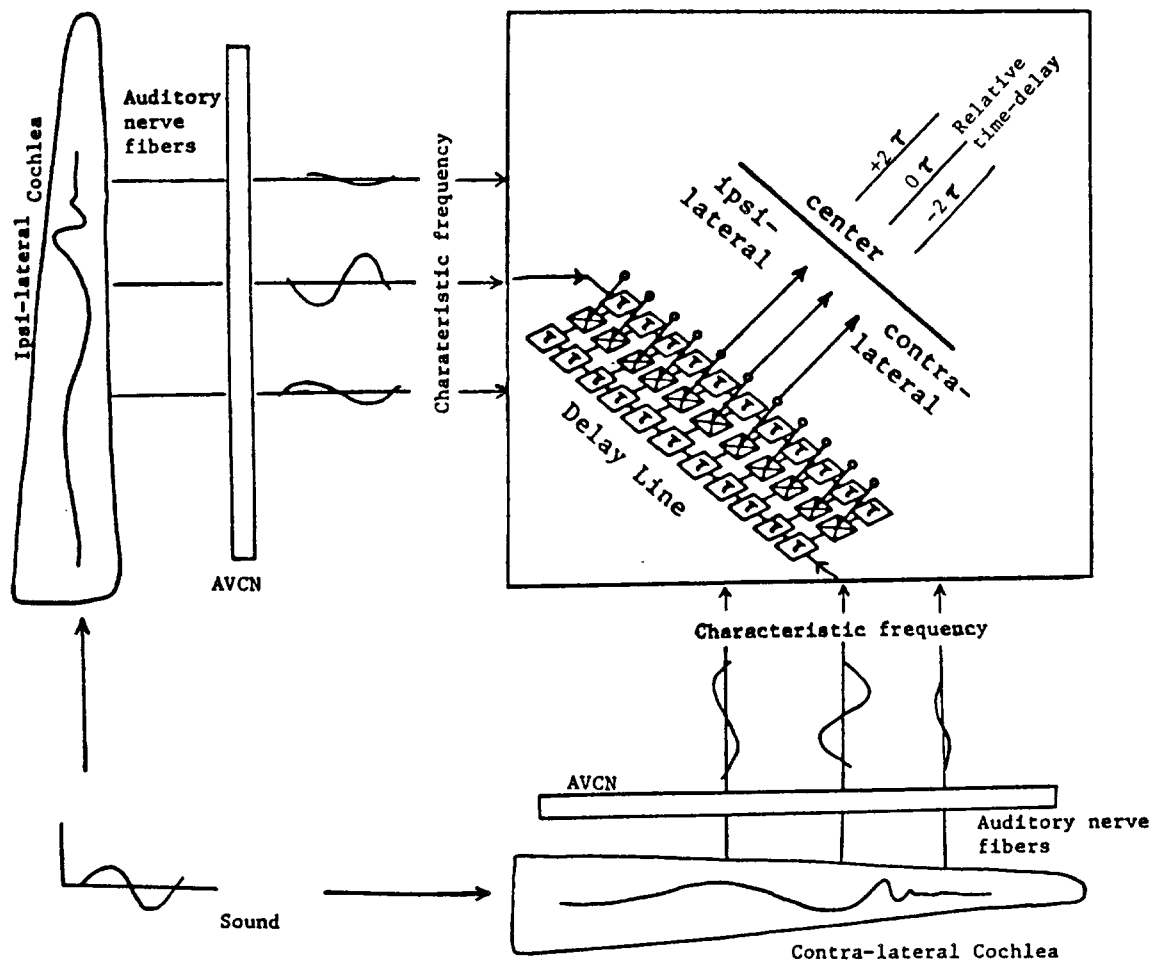


Figure 3.1: A Neural delay based Binaural model

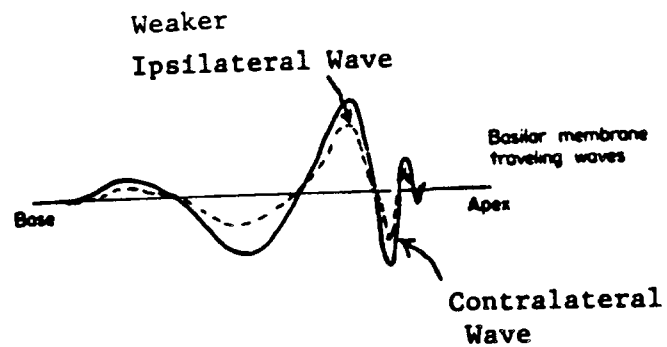
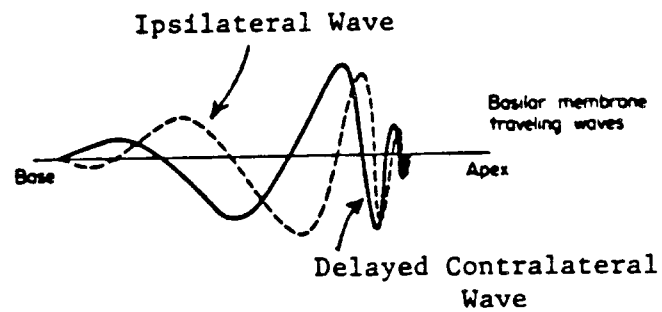
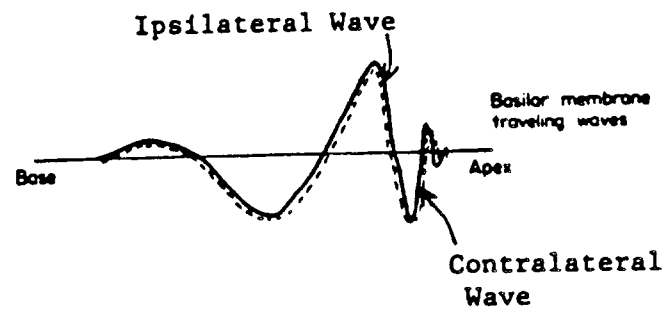


Figure 3.2: Envelope of the traveling waves for 600Hz tone

The displacement of the stapes is transduced into ripples in the cochlear fluid which set up traveling waves on the basilar membrane lining the inner wall of the cochlea, with each point on the basilar membrane vibrating at the frequency of the input sound stimulus. These traveling waves conform to an envelope which depends on the spectrum of the input sound. Corresponding to each frequency there is a characteristic envelope which peaks at a different location on the membrane. The tonotopic arrangement of the membrane is such that the lower the frequency of the stimulus, the closer the peak of the envelope is to the apex of the membrane. The rising edge of the envelope is gradual while the falling edge is extremely steep as can be seen in figure 3.2. In the case of a complex sound input, a peak can be observed in the envelope at locations corresponding to all the main frequencies observed in the spectrum of the stimulus. While the magnitude response of the membrane determines the shape of the envelope, the phase response of the membrane is the source of the traveling waves. A phase shift is observed on the membrane which increases to a maximum around the region of the CF. It must be noted that while at lower frequencies, the fine temporal structure of the traveling waves can be seen clearly and phase locking occurs, at higher frequencies the only distinguishable feature of the spatiotemporal response is the envelope.

A simplified schematic of the biophysical cochlear model is shown in figure 3.3 ([35]). It consists of a linear 2-dimensional hydroelastic basilar membrane model, followed by the velocity fluid-cilia coupling stage and finally the biophysical model of the inner hair cell. The hair cell model consists of a sigmoidal non-linearity due to the threshold and saturation characteristics of the hair cell transduction

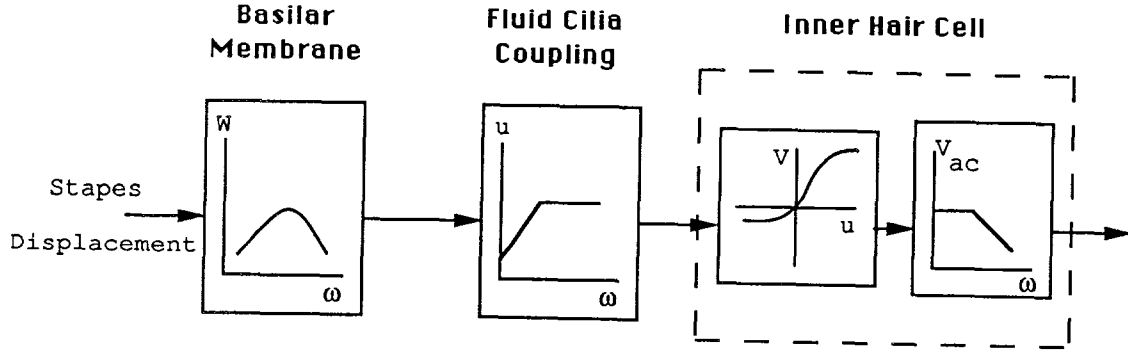


Figure 3.3: Cochlear Model used in the processing

mechanism. The non-linearity leaves the spatial features of the spatio-temporal patterns unaffected over wide ranges of stimulus intensity. The cochlear model used, ([18], [35]) assumes the basilar membrane to be divided into 128 channels, each channel corresponding to a nerve fiber where the frequency scale of the membrane is approximately logarithmic. The characteristic frequency of the fiber is defined as that frequency for which the fiber shows maximal activity. The cochlear spatiotemporal patterns are formed using digital filters and the responses to the stimulus are generated using an FFT based overlap-and-add method. The output is then highpass filtered ($w_n = u_n - 0.8u_{n-1}$) modeling the outer ear and fluid-cilia coupling stages; compressed by a sigmoidal function of the form

$$x = \frac{M}{(1 + be^{-aw})}$$

where a , b and M are parameters of the non-linearity; finally lowpass filtered with time constant 0.1 to 1 msec to smooth the output. The response of the model to a 600 Hz. pure sinusoid is shown in figure 3.4 along with the input stimulus. The 3-dimensional output shown is also called the “waterfall” output and plots the

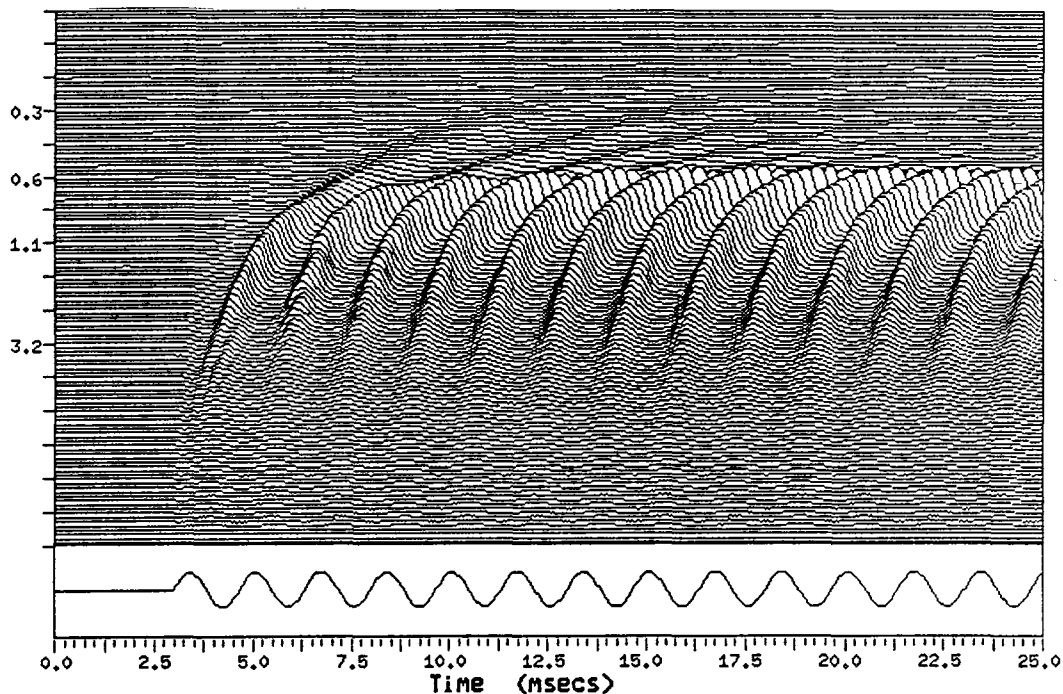


Figure 3.4: Traveling waves on the basilar membrane : 600Hz tone

cochlear excitations in time. The markings on the vertical axis represent frequencies on the membrane. The maximal phase shift of the response at the frequency of the input is clearly seen in the sharp bending of the pattern around the region of the CF.

III The Stereausis Network

While it is difficult to find conclusive evidence in support of any binaural processing neural network model, at best, one can show that the relevant existing physiological, anatomical and psychophysical data are consistent with the various aspects of the proposed model and that the basic model design principles are in harmony with the fundamental principles of organization of the mammalian auditory system. The

following are a few of the basic assumptions of the stereausis network :

1. The primary pathways of the auditory system maintain their tonotopic order from the basilar membrane through several central nuclei up to the cortex ([20]). This emphasizes the importance of the spatial dimension in auditory processing at all the levels.
2. The fine temporal structure of the responses on the auditory nerve, crucial to binaural processing is preserved in the responses of the Bushy cells of the anteroventral central nucleus (AVCN) which project partially via the Nucleus of the Trapezoidal Body (NTB) to the nuclei of the superior olivary complex (SOC), where significant binaural interactions are recorded. It can thus be concluded that binaural networks that utilize this temporal information are presumably located at this level ([41]).
3. While it is known that binaural processing at lower frequencies primarily depends upon the Interaural Time Differences (ITD) and processing at higher frequencies on the Interaural Level Differences (ILD), it is unclear at present whether separate or identical binaural networks process these differences. Nevertheless, considering the similarity of the intended psychoacoustical attributes, it would be desirable to propose one network that can gracefully process both types of cues or at least show that closely related networks underlie their processing. The same would hold true for onset and ongoing (continuous) interaural cues.

The input patterns used by the binaural processing network are generated using a simplified biophysical model of the basilar membrane where the nerve responses are represented by their instantaneous probability of firing. The stereausis network combines the ipsilateral and contralateral cochlear outputs in a simple ordered matrix of operations. The activity of node (i, j) of the matrix is the output of combining the responses of the i^{th} ipsilateral fiber x_i and the j^{th} contralateral fiber y_j . The inputs are combined to produce an output $o_{ij} = C(x_i, y_j)$ where C is a correlation measure between inputs. Thus the cochlear response at a given CF location in one ear is correlated with outputs from the same CF and locally off-CF fibers from the other ear such as x_i with y_{i-1} , y_i and y_{i-2} . See figure 3.5. As shall be shown shortly, this spatial cross-correlation is proportional to a normal cross-correlation in time because of the finite velocity of the traveling waves. This possibility seems to have been first proposed by Schroeder ([33]). Thus nodes along the diagonal c_{kk} receive inputs from corresponding CF locations from both ears while nodes on axes parallel to the diagonal receive systematically spatially delayed inputs. Next the outputs are further processed by a spatial lateral inhibitory-excitatory mask to enhance the results. Finally the output patterns are averaged over time for about 12 msec(which may be thought of as the time constant of the neurons). The exact form of the correlation is not crucial to the network as long as it generates a consistent correlation measure of its half wave rectified inputs. All the data that is shown in this report has been done using the function $C(x, y) = (x + y)^2$.

The output patterns obtained using other functions such as

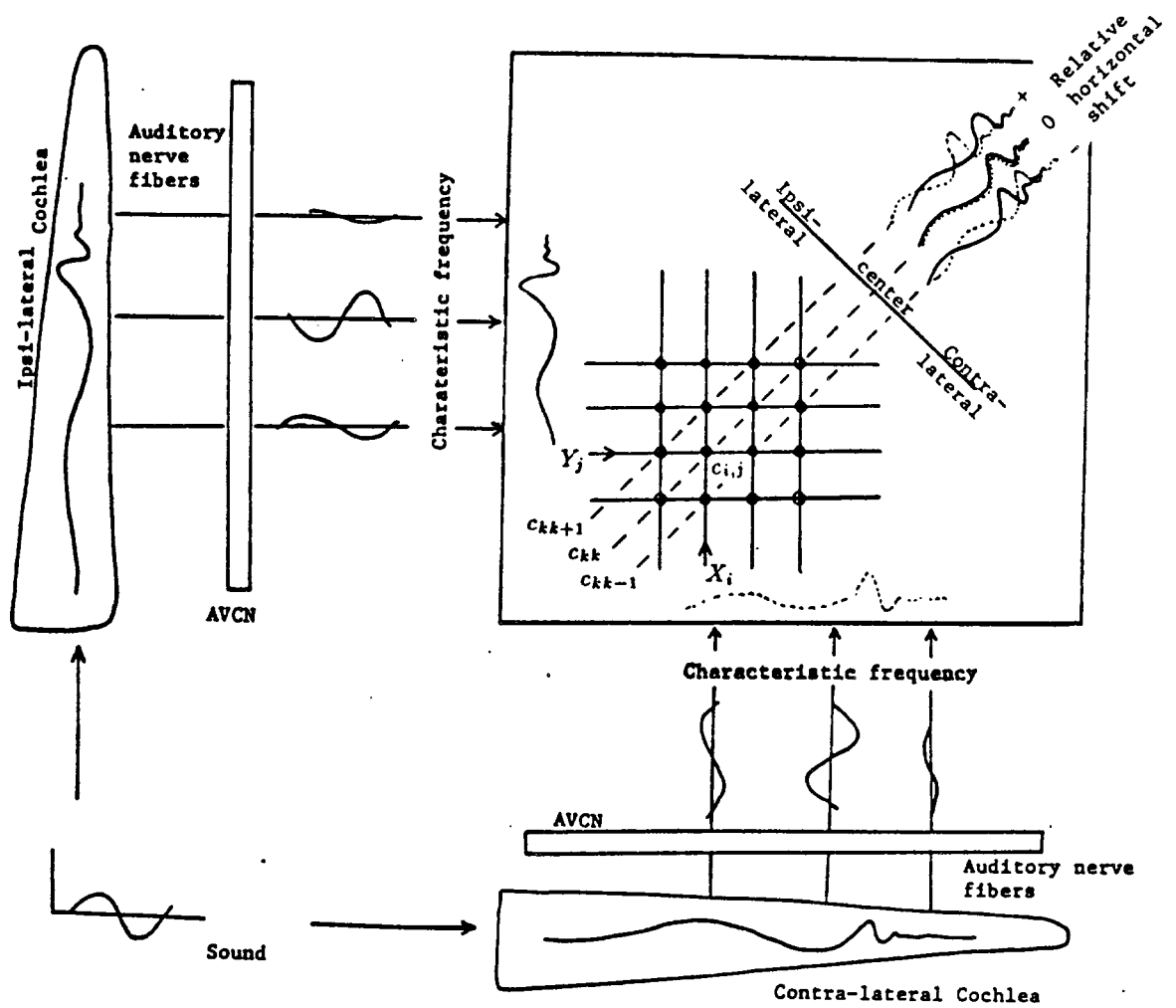


Figure 3.5: The Stereausis Network

$C(x, y) = x - y$, $C(x, y) = x * y$, and $C(x, y) = x + y$

though not shown in this report, are essentially similar to the results shown.

Now to show that the spatial correlation is equivalent to a temporal correlation, consider the output of the network to be specified as :

$$\begin{aligned} o_{ij} &= C(x_i, y_j) \\ &= \int_T x_i(t) \oplus y_j(t) dt \end{aligned} \quad (3.1)$$

where \oplus is some mathematical correlation operator

$x_i(t)$ is the response of the i^{th} ipsilateral fiber

$y_j(t)$ is the response of the j^{th} contralateral fiber

T is the time period over which the output patterns are averaged

Let

$$x_i(t) = A_i(\omega) \sin(\omega t + \theta_i(\omega)) \quad (3.2)$$

$$y_j(t) = A_j(\omega) \sin(\omega t + \theta_j(\omega)) \quad (3.3)$$

where $A_i(\omega)$, $A_j(\omega)$, $\theta_i(\omega)$ and $\theta_j(\omega)$ are the amplitudes and the phases of the traveling waves at the i^{th} and j^{th} locations of the two cochlea for a frequency ω .

If i and j have close CF locations on the membrane,

$$\begin{aligned} A_i(\omega) &\approx A_j(\omega) \\ \theta_j(\omega) &= \theta_i(\omega) - \Delta\theta(\omega) \\ y_j(t) &= A_i(\omega) \sin(\omega t + \theta_i(\omega) - \Delta\theta(\omega)) \end{aligned} \quad (3.4)$$

If the velocity of the traveling wave is assumed to be a constant v over the small distance Δs between the locations i and j on the membrane, the spatial frequency

of the wave ω_s can be expressed by the relation :

$$\omega_s = \frac{\omega}{v} \quad (3.5)$$

$$\begin{aligned} \Delta\theta(\omega) &\approx \omega_s * \Delta s \\ &= \frac{\omega}{v} \Delta s \\ &= \omega \tau_s \end{aligned} \quad (3.6)$$

where τ_s is the time taken by the wave to travel the distance Δs at location i on the membrane.

Substituting in equation 3.4

$$\begin{aligned} y_j(t) &= A_i(\omega) \sin(\omega(t - \tau_s) + \theta_i(\omega)) \\ &= y_i(t - \tau_s) \end{aligned} \quad (3.7)$$

Thus, substituting in equation 3.1

$$o_{ij} = \int_T x_i(t) \oplus y_i(t - \tau_s) dt \quad (3.8)$$

It can be seen from the above equation that a spatial cross correlation along the length of the basilar membranes of the two ears over small areas around the CF is equivalent to a temporal correlation of the inputs at identical locations on the two membranes.

The specific ON-center OFF-surround mask used to sharpen the output in the

processing was as below :

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

Although the mask was applied in a non recursive (feedforward) manner to speed up the computations, similar results can also be got in a recursive (feedback) configuration, a topology that is commonly reported in physiological literature ([17], [25], [34]). Figure 3.6 shows the output of the network to two simple input patterns crudely mimicking the actual cochlear inputs : a single pulse sweeping across the spatial axis of the two inputs and a pair of pulses with a fixed separation moving across the input space. In figure 3.6a when the input patterns are identical, maximum activity occurs along the diagonal corresponding to maximum correlation between corresponding locations on the two membranes. Since the pulse height is fixed, the activity occurs uniformly all along the diagonal. When the inputs are now relatively phase shifted as is the case in figure 3.6b, the line of activity moves away from and parallel to the diagonal towards the side corresponding to the leading input. The shift is parallel to the diagonal since the pulse moves at a constant velocity and maintains a constant spatial shift with respect to the pulse on the other membrane. While the output patterns in figures 3.6c and 3.6d are similar to the cases already cited, the output feature to be noted is the presence of two secondary peaks of smaller amplitude corresponding to the correlation between the

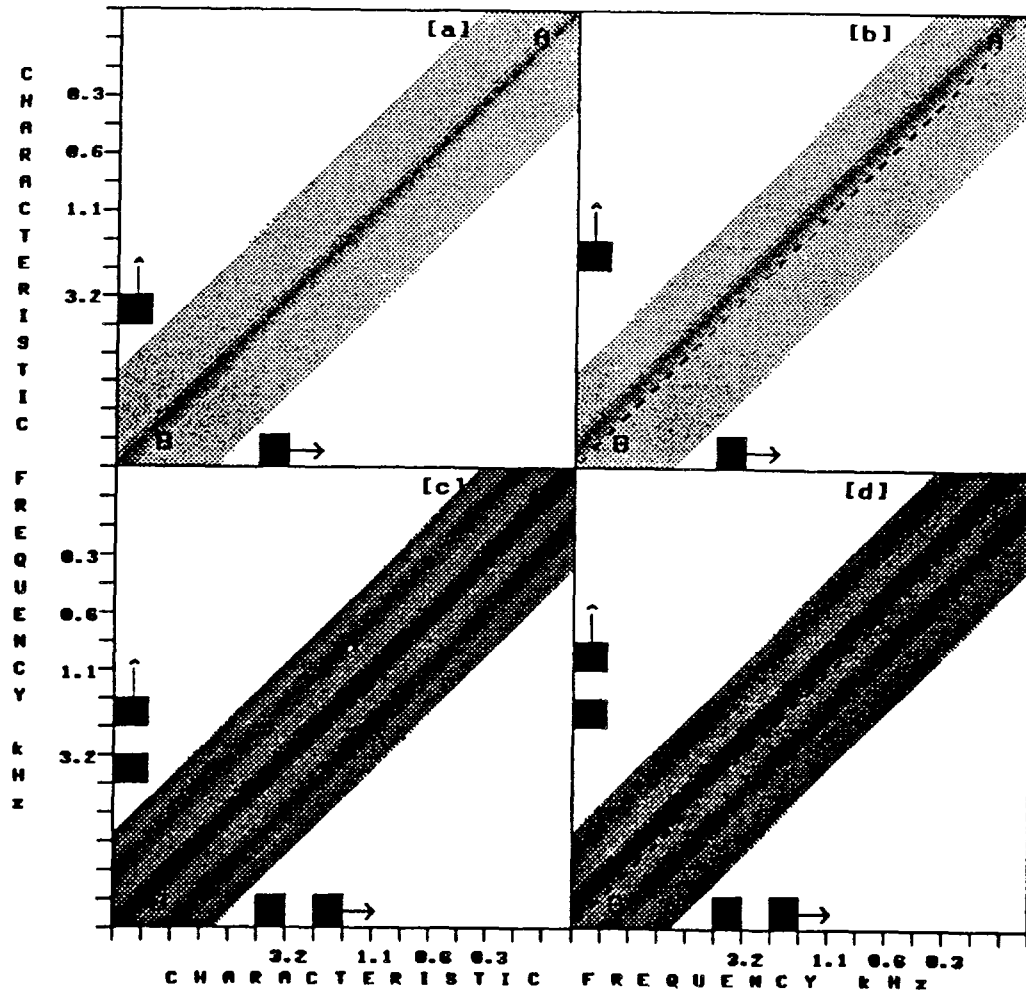


Figure 3.6: Network response to pulses traveling across the membrane

non-corresponding pulses on the two membranes. It should be noted that maximal activity is still in the central peak.

IV Binaural processing in the Stereausis network

The stereausis network was tested on several classes of inputs to show its ability to explain several binaural phenomena. There were six classes of which the first dealt with lateralization and the last with signal detection and enhancement in noisy environments :

1. Lateralization of low frequency tones ($\leq 1.5\text{kHz}$.) using interaural time delays (ITDs).
2. Lateralization of high frequency tones – Interaural intensity differences (IIDs).
3. Time-intensity trading for low frequency tones.
4. Lateralization of speech and harmonic complex stimuli.
5. Lateralization of broadband noise.
6. Detection and enhancement of tones in noise.

The computations performed in the network on the inputs are as follows :

1. The stimulus is processed by the cochlear model to generate the spatiotemporal response patterns of the auditory nerve (similar to those in figure 3.4). All responses are expressed in terms of the instantaneous firing rates of the cells.
2. Each $(i, j)^{th}$ node of the binaural network performs the following coincidence operation :

$$c_{ij} = (x_i(n) + y_j(n))^2 \quad (3.9)$$

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 matrix of activities c_{ij} is computed.

3. Each frame is then processed by the LIN nonrecursive mask explained earlier and then half wave rectified to remove negative outputs.

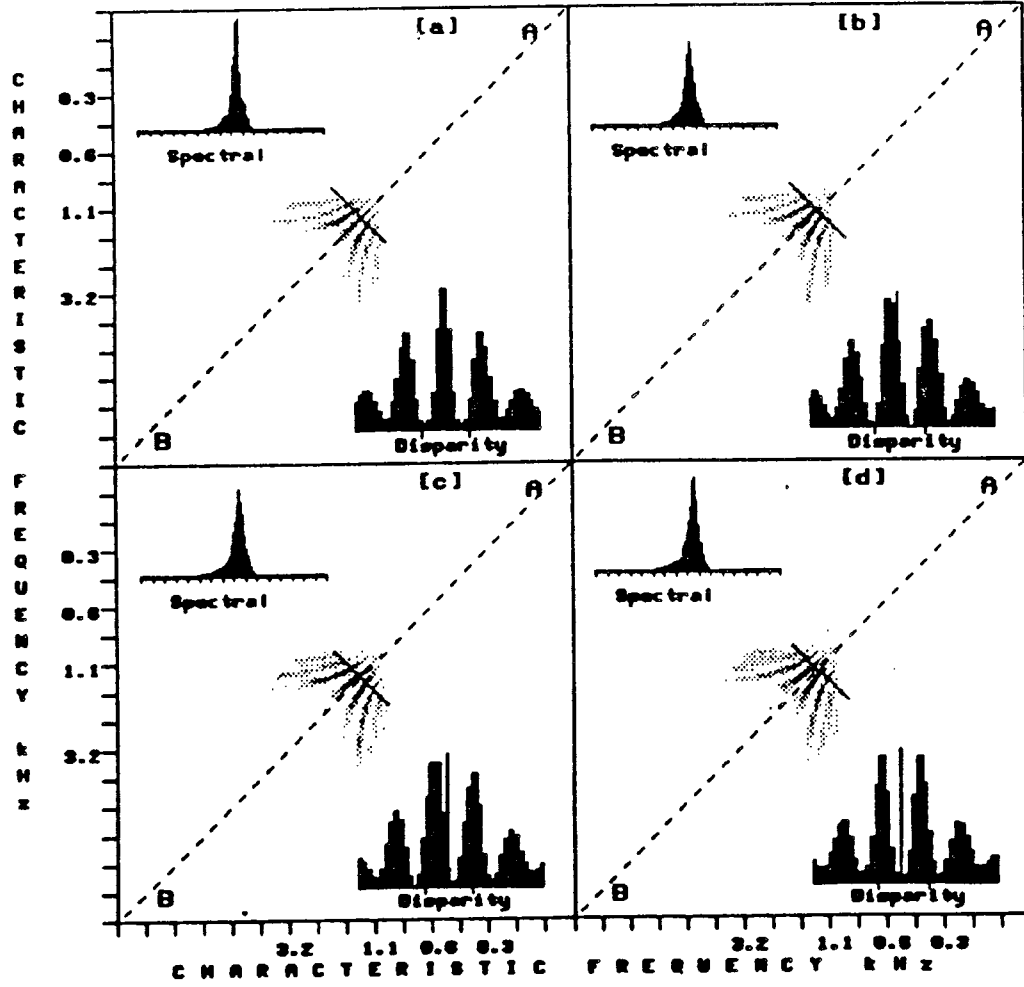


Figure 3.7: Network output for time shifted 600Hz. tone

a) zero shift b) $\frac{\pi}{3}$ shift c) $\frac{2\pi}{3}$ shift d) π shift

In this thesis, only two of the above cases are discussed, the lateralization of low frequency tones and the time-intensity trading effects. The results for the other cases can be found in the thesis work of Naiming Shen ([36]).

Lateralization of a 600Hz. tone :

A lower frequency tone evokes a traveling wave that is conveyed to the binaural network via the phase locked responses of the input pathways. The responses of the network to phase shifts of 0 , $\frac{\pi}{3}$, $\frac{2\pi}{3}$ and π are shown in figure 3.7. They

are qualitatively very similar to the responses of the network to the simulated inputs (figure 3.6) in that a primary correlation maximum emerges, accompanied by several secondary peaks arising from the multiple peaks within the traveling wave envelope. It is observed that almost all the output activity is concentrated around the CF location specific to the input tone (600Hz. in this case). Varying the frequency of the input excitation would cause the active region to move along the spectral axis i.e. along and parallel to the diagonal AB. The spectral plot in the upper left corner of each output samples activity along and adjacent to this spectral axis. The other interesting feature of the outputs which is different from the simulated case is that the secondary maxima converge towards the primary peak rather move parallel to it. 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 the peaks beyond the point of resonance. 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 around the region of the band drawn in the figure. For the centered tone, a dominant peak of activity appears along the AB diagonal; 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 secondary peak is so large that there are now two equal peaks on either side of the midline. With further shifts, the previously secondary image moves further towards 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 ([11], [31]).

Time-Intensity trading for low frequency tones

For low frequency tones, both phase as well as amplitude disparities can be preserved in the responses of the auditory nerve and hence detected by the stereausis network. Figure 3.8 illustrates the effect on the network output of increasing the ILD of a centered 600Hz. tone. Two regions of activity emerge along the disparity axis in this experiment: (1) the centered primary peak whose location remains relatively fixed reflecting the zero phase shift between the two inputs and (2) the secondary peak which grows relatively in height and broadens with the increase in the ILD. The result can either be viewed as a single broad auditory event with a center of gravity that is lateralized as the relative height of the secondary peak increases or alternatively, the two peaks can be viewed as two distinct auditory percepts - one remaining in the middle corresponding to the zero phase shift and the other migrating to the side of the head corresponding to the ILD becoming more spatially blurred as the level difference increases. The second interpretation was most often reported in similar psychoacoustical experiments ([40]).

V Concluding Remarks

The fundamental difference between the stereausis network and the Jeffress model is in the use of spatial versus temporal correlations to extract the binaural cues. The major implication of this distinction is the functional role of neural delays, one that both relates and distinguishes the two models. Neural delays are an inevitable occurrence in any biological network where information is transmitted from one

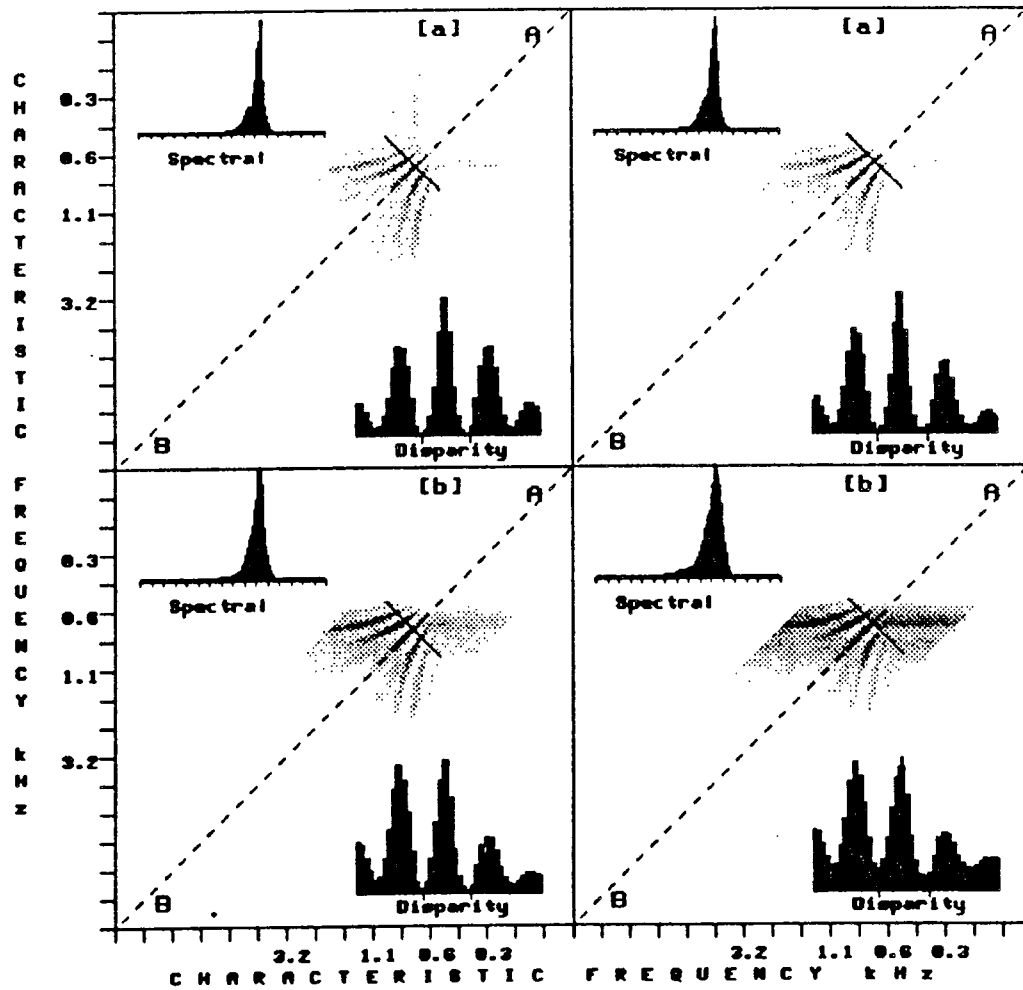


Figure 3.8: Network output showing time-intensity trading at low frequency

a] centered tone b] same as (a) with ILD = 3db.

c] same as (a) with ILD = 6db d] same as (a) with ILD = 12db.

point to another. The two models represent two extremes on the relevance of such delays. While they are pivotal to the Jeffress model, they are completely ignored by the stereausis network. In reality, it would seem likely that an intermediate view exists and the two models simply represent two ideologies through which the functioning of the biological network can be understood.

INTRODUCTION TO NEURAL NETWORKS

I Introduction

A neuron is the basic unit of processing and memory storage in the human body. A set of neurons connected together with different connectivities forms a network which is a strong tool for pattern recognition and associative memory. Networks are formed due to the synapses between neural axons and dendritic trees. The strength of these synapses can be varied based on several learning rules. Each neuron can be modeled by a transfer function combining the inputs to generate a train of spikes which is proportional to the output. Section II in this chapter will discuss the neuronal model used in later computations. Section III is an overview of the different kinds of network connections possible and their functional differences. The final section of the chapter deals with the various kinds of learning rules used for connection formations in neural networks.

II The neuron model

A neuron consists of a cell body, receiving inputs from a dendritic arborization. The cell body terminates in a long axon which transmits the output signal to

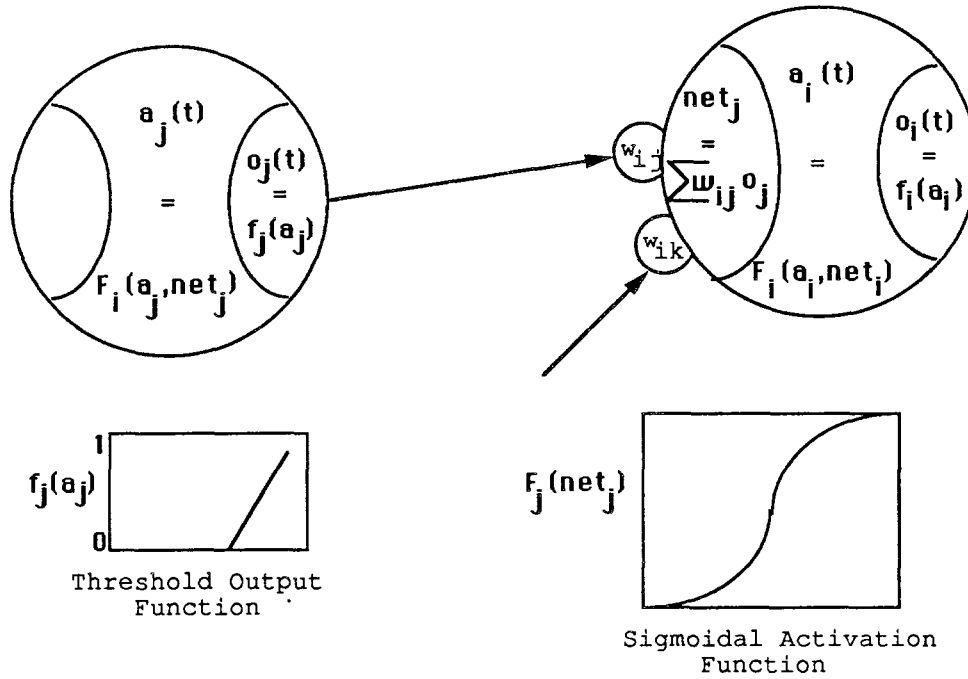


Figure 4.1: A Neuron based system

one or more neurons and in some recursive networks to the same neuron itself. Input and output signals in neurons are in the form of electrical pulses also termed as “spikes” where the number of spikes per time instant in an axon or dendrite is proportional to the level of activity in that process. Axons connect to other neurons by converging on the dendritic input trees and making a connection at a “synapse”. The synapse regulates the amount of current flowing from the axon into the dendrite and thus the influence of the source neuron(from which the axon emanated) on the destination neuron (the neuron associated with the dendritic tree). A neuron would thus receive inputs from various axons each weighted by the strength of the corresponding synapse. These input signals combine to generate a potential in the cell, which, if above a threshold causes the neuron to fire. Depending upon the type of neuron and its output it could either saturate or maintain a steady value.

A general neuron based system is shown in figure 4.1.

For computational and analytical purposes, a neuron can be represented as a processing unit and a whole set of such units would form a parallel distributed processing system ([29]). Each unit is relatively very simple, performing the rudimentary task of receiving inputs from its neighbors, computing an output value which is a function of its inputs and sending that value to its neighbors. A unit can compute at the same time as some other unit leading to a massively parallel system. In a general system, units can be divided into three categories : input, hidden and output units. Input units receive signals from sources which are external to the framework under consideration, but their outputs propagate to other units within the system. Such units are usually sensory units or interfaces to other processing systems. Output units obtain their inputs from within the system but they send output signals out of the system under consideration either to other processors or to the motor system. The hidden units, as the name suggests are completely imbedded inside a system and are not visible at the interface levels. They receive signals from and transmit signals to other units within the system. Associated with each neuron is an activation state which is a measure of the activity of that unit. It is a function of the inputs signals received by the unit and could either be continuous or discrete. The simplest case would be when the state of activation is a weighted combination of all the inputs to a neuron. A discrete activation function would have a value at one of two or more state values. (The most common case would be a binary one where the unit could either ON or OFF.)

The output of the unit at any instant in time is a function of the state of

activation of the unit at that moment. The output signal would serve as an input to other processing units or to an external system and the strength of this signal would affect the functioning of the succeeding neuron. The function transforming the state of activation to the output signal could be a simple step function, a sigmoidal or a signum function. Depending upon the type of system, any neuron could excite or inhibit all other units, or their patterns of connectivity could be restricted by some constraints imposed by the system under consideration. If strong output activity of neuron X reinforces the output of neuron Y, X is said to “excite” Y or alternatively the input to Y from X is called “excitatory”. If on the other hand, strong activity on neuron X causes neuron Y to be shut off, or its output to diminish, the input to Y is said to be “inhibitory” or neuron X inhibits neuron Y. Each link between any two neurons is associated with a weight which is a measure of the impact that one neuron has on the other. Mathematically, a positive weight implies an excitatory link and a negative weight an inhibitory connection. A weight of zero would imply that there is no direct effect of one neuron on the other. The pattern of connectivity and the weights of the links are very important since these represent what the system knows or what features the system can identify in the input. For every system, a connectivity matrix can be formed which is a square matrix with as many rows as there are processing units. Elements in the matrix represent the weight of the link between the neuron corresponding to its row and the neuron corresponding to its column. The nature of the matrix determines the nature of the system and the kind of processing it performs. Other information such as the fan-in (number of elements that excite or inhibit a unit) and fan-out

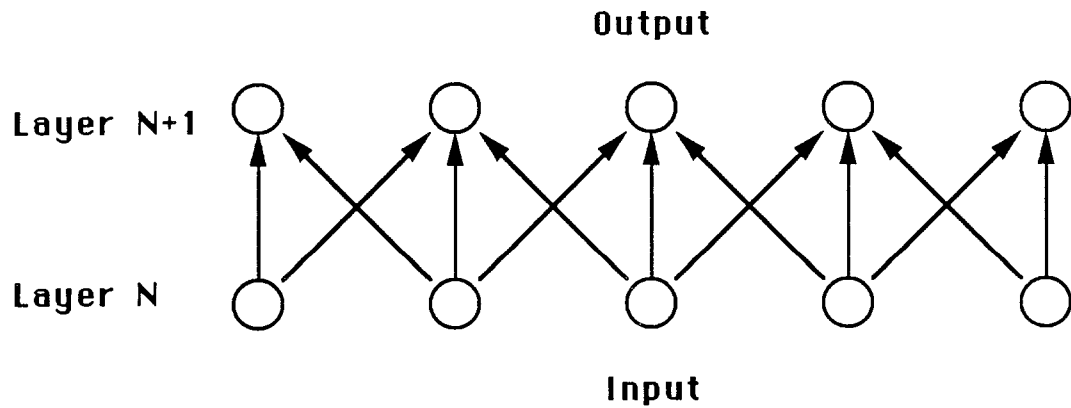


Figure 4.2: Feedforward Architecture in neurons

(number of units directly affected by a unit) are also embedded in this matrix.

III Neural Network Connections and Architectures

Neural networks can be connected by two kinds of architectures, a feed-forward fashion or a recursive connection. If neurons are assumed to be stored in layers, all connections are only in one direction in feed-forward connections. Neurons in a layer are affected only by neurons in previous layers. An example is shown in the figure 4.2. A recursive or feedback architecture on the other hand imposes no restriction on the direction of the connections. See figure 4.3. A neuron can feed its outputs to previous layers or even to itself. It is obvious that feedforward networks are just a subset of recursive networks. A recursive architecture can be modeled by a feedforward network although the representation would most often require more neurons, more connections or both.

A neural network can represent two basic kinds of knowledge: an associative memory or a regularity/pattern detector. An associative memory is a setup which

learns to adjust its connectivities so that a particular input or pattern of activation to a system will always result in another distinct pattern at the output. It would usually be required that an associative memory map an orthogonal set of inputs into an orthogonal output set. Such a network is employed when patterns must be stored to recall them at some future time. Regularity detectors are networks that update their connectivities so that one or more characteristic features in a pattern of input activity can be detected or locked onto. They would be useful in detecting distinct classifying features in input signals. In several cases, however, a combination of both these modes of representations may be required to achieve the desired output. It may be required to detect a regularity in the input which must later be mapped into a certain specific output pattern. An example of this is the problem of sound localization of sound to which these learning networks have been applied in succeeding chapters; a regular systematic difference in the cochlear input from the two ears must be detected and depending upon the spatial separation and the frequency of the tone (determined by its location on the basilar membrane), it should be mapped into a specific location/region on the azimuth.

An associative learning paradigm can either be a pattern association paradigm or an auto-associative one, the former being a subset of the latter. A simple pattern associator learns to build connections between patterns defined over one subset of units with a pattern in another disjoint subset of units. Most often, a teacher is provided at the latter set of units in order to guide the direction of pattern matching. An auto-association paradigm is one where the same pattern serves both as input as well as the output. A goal of such a system would be pattern

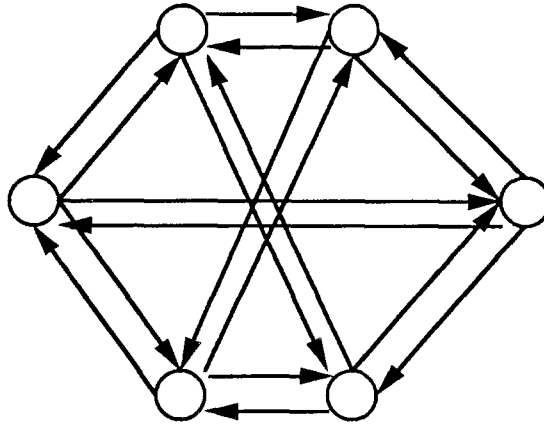


Figure 4.3: Feedback Architecture in neurons

completion wherein if a portion of the input is presented, it can complete the rest of that pattern, the correctness of response depending upon the fraction of the entire pattern being presented. Since the input and output units are indistinguishable, connections can exist from any unit in the system to any other. It can be surmised that while a pattern associator is a feedforward network, the auto associator is recursive in nature. Due to the presence of a teacher in the associative learning process, it is also called supervised learning. A regularity detector on the other hand has no explicitly specified teacher. The system, based on its setup determines its own teaching function. The nature of the teaching function and the input patterns determine the feature that the network learns to extract.

IV Learning in neural networks

For any neural network, the connectivity matrix is determined by a learning rule which is pre-specified either externally or is implicit in the network architecture. The learning rule gathers the information it requires from the inputs, from the

network arrangement, and from the teacher signal if it is known, and iteratively generates a connectivity matrix which will generate a desirable output. All learning rules can be categorized as either supervised or unsupervised learning depending upon the presence or absence respectively of a teacher input. In the absence of an explicit teacher, unsupervised learning algorithms rely solely upon a teaching function which must be determined or fixed earlier. Depending upon the nature of this function, unsupervised learning can be further subdivided into reinforcement learning and competitive learning.

1. *Supervised Learning :-*

As the name suggests, a supervised learning algorithm requires a teacher or some desired output corresponding to each input pattern/signal. Assuming initial (random) connectivities and an output function for each unit, the system updates the weights of the links in an attempt to minimize the error between the teacher and the actual output i.e. update the matrix vector W such that an error function $f(t_p, o_p)$ is a minimum where t_p is the teacher value and o_p the output of the neuron for a pattern p . The exact form of the error function f varies for different algorithms. In the special case when the outputs are linearly separable, the supervised learning rule is reduced to an algorithm to determine a hyperplane that separates the input patterns into two different categories. The most commonly used learning rule in this class of algorithms is the perceptron learning rule where the output of a unit is a linear combination of its inputs and the error function to be minimized

is the mean square error between the teacher and the output. While the perceptron rule is useful only for single layer networks, Rumelhart extended it to the backpropagation algorithm for multilayer structures.

2. *Competitive Learning* :-

Being a type of unsupervised learning, in this class of learning rules, there is no guiding or teaching signal which ensures at every step that the network update is progressing in the correct direction. Every unit in the network *competes* with all the other units for the right to update their weights. The unit which has the maximum output to a particular input pattern is the one closest tuned to that input and it wins the competition. This chosen unit alone now adjusts its connectivities so that it is even better tuned to this input. In this manner, each output unit will capture a certain subset of the input patterns, and respond to all the inputs in that subset. Several learning rules based on or similar to the competitive learning scheme have been proposed by von der Malsburg ([38]), Fukushima ([13]), Grossberg ([15]) and Kohonen ([22], [24], [23]). This method has been used in later analysis of sound lateralization and is discussed in detail in chapter 6. A disadvantage of this learning algorithm is that in the absence of a teacher, it is impossible to be sure that the learning is progressing in the desired direction thus making it difficult to guarantee convergence to a solution.

3. *Re-inforcement Learning* :-

This scheme is similar to supervised learning in that there is a teacher; how-

ever the teaching signal is not a deterministic output as in the supervised case, but rather a less informative teaching or reinforcing signal, which evaluates the response of a unit to an input pattern. The reinforcing signal is a fuzzy measure of whether the output is good, bad or indifferent to the input. This signal is very often noisy and inconsistent making it unreliable, unlike the definite hints given by the teacher in a supervised learning algorithm. The learning rule associates input patterns with different output patterns in an attempt to find the association which maximizes the positive reinforcement. Such a learning method has been test-implemented in various areas such as reinforcement learning control ([12]; [27]), learning automata theory ([28]), and associative searching ([2]; [1]). Finding a reliable reinforcement signal is difficult making this learning procedure a complex one. In trying to explain the biological phenomenon of sound localization, it is difficult to identify, or propose and biologically justify a reinforcing signal and hence reinforcement learning was rejected as a possible solution to the network learning problem discussed later.

V Conclusions

This chapter is an introduction to the models and ideas used in succeeding sections of this report. It began with a discussion on the neural model used in all further computations, followed by a simplistic look at the different architectures present in neural networks. A cursory study of the two main types of learning algorithms supervised and unsupervised was presented with a closer look at two particular

unsupervised learning procedures viz. competitive learning and re-inforcement learning. With particular reference to the problem being studied in this report sound localization on the azimuth, two possible solutions have been proposed; the first based on a supervised learning algorithm is discussed in chapter 5; the other is an unsupervised, competitive learning rule explained in chapter 6.

A SUPERVISED LEARNING SOLUTION**I Statement of the Problem**

As has been explained in the previous chapter, the process of supervised learning requires the presentation of pairs of input and output patterns. By repeated presentation of these fixed pairs, the network must learn to align its connectivities so that it will always associate a particular input pattern with a specific output pattern. Once the network connectivities can produce the correct output corresponding to each input it was trained on, the network is said to have *learned*. For the problem of sound lateralization being discussed, the input to the network is the responses of the basilar membrane from the two ears; the output is the activity of a layer of neurons, in which only one is ON for the different location of the sound source on the azimuth. Since, each location of the source corresponds to a different spatial separation between the traveling waves on the two basilar membranes, a set of input patterns along with a corresponding orthogonal set of output patterns can be presented to the network as the data on which it must train. The basilar membrane responses are obtained from a cochlear model described earlier in chapter 3. The network obtains its input not from the entire membrane but rather from

the small region around the characteristic frequency of the input. For the learning process presented in the next section, inputs from a set of 20 points around the location of the CF were used to train the connectivities. The network topology consists of two layers, a layer of input neurons and a layer of output neurons. Each input neuron is initially connected to every output neuron. It is assumed that the initial connectivities are such that maximum weight occurs between an output neuron and the fiber corresponding to the CF while the weights to the other input neurons taper away on either side.

II The Learning algorithm

The algorithm used is derived from the gradient descent rule. According to this rule, the change in connectivity of a link between two neurons is proportional to the rate of change of output error (defined as the difference between the teacher input and the actual output observed at the neuron) with respect to the connectivity. Let the error surface be

$$E = \sum_p E_p$$

where E_p is the error associated with a particular output pattern p . If t_{pi} is the value of the teacher at a neuron i for a pattern p , and o_{pi} the actual observed output at the same neuron for the same pattern, E_p can be defined as

$$E_p = \frac{1}{2} \sum_i (t_{pi} - o_{pi})^2 \quad (5.1)$$

$$o_{pi} = o_{pLi} * o_{pRi} \quad (5.2)$$

$$= \sum_j w_{ij} x_{jL} * \sum_j w_{ij} x_{jR} \quad (5.3)$$

where w_{ij} is the weight between neuron i and neuron j

o_{pLi} is the output from the left cochlea

o_{pRi} is the output from the right cochlea

The output at a neuron is defined as the product of the output values from the two cochlea; the output of each cochlea is the sum of all the inputs from that ear in the region of the stimulus CF, each weighted by the connectivity of the link between the input fiber and the output neuron. The product term in the output ensures that the neuron is excited only if there is an input component from both ears, not just from one of them. Thus in the case of a monaural input, these neurons are never excited. Hence such neurons are called “*binaural neurons*”.

$$\begin{aligned} \frac{\partial E_p}{\partial w_{ij}} &= \frac{\partial E_p}{\partial o_{pi}} * \frac{\partial o_{pi}}{\partial w_{ij}} \\ &= -(t_{pi} - o_{pi}) * \frac{\partial o_{pi}}{\partial w_{ij}} \end{aligned} \quad (5.4)$$

$$\frac{\partial o_{pi}}{\partial w_{ij}} = o_{pRi} * x_{jL} \quad \text{for the left ear} \quad (5.5)$$

$$\frac{\partial o_{pi}}{\partial w_{ij}} = o_{pLi} * x_{jR} \quad \text{for the right ear} \quad (5.6)$$

$$\begin{aligned} -\frac{\partial E_p}{\partial w_{ij}} &= (t_{pi} - o_{pi}) o_{piop} x_j \\ \Delta w_{ij} &\propto -\frac{\partial E_p}{\partial w_{ij}} \quad \text{by the gradient descent rule} \end{aligned} \quad (5.7)$$

$$\Delta w_{ij} = \eta (t_{pi} - o_{pi}) o_{piop} x_j \quad (5.8)$$

where η is the learning rate and is usually a constant.

o_{piop} is the output of the cochlea opposite to the one being updated

It can be seen from the learning rule above that the weights are fashioned in a way so that the output neurons approximate the excitation profile specified by the teacher. The presence of the input term x_j shows that if a highly excitatory signal appears at an input neuron when the teacher is also high at the output fiber, the connection between the input and output is strengthened proportional to the level of excitation. In contrast, should the teacher be low, the weight of the connection is correspondingly reduced. Update of the weights occurs only when an input is present at both ears. Let us assume that the sound stimulus being presented is monaural at the left ear, the update terms for the right ear would be zero since the input term in the update equation would be zero always; at the left ear, the o_{piop} term would be the output of the right cochlea which is always zero since there is no excitation there. Each connection also undergoes a decay at every iteration. Thus the complete learning rule is

$$w_{ij} = \eta(t_{pi} - o_{pi})o_{piop}x_j + (1 - \beta)w_{ij} \quad (5.9)$$

where β is the decay rate.

The decay rate must be slow enough to make certain that the connectivities do not decay to zero; if this should occur, the output from both ears o_{pLi} and o_{pRi} would be zero and no update would occur. The learning rate is also critical for the stability of the system. The learning rule proposed is trying to find the minimum on the error surface. A low η would make learning sluggish and at the same time may cause the system to settle at a local minimum rather than the global minimum of the error surface which is the actual solution of the problem. A high value of η on the

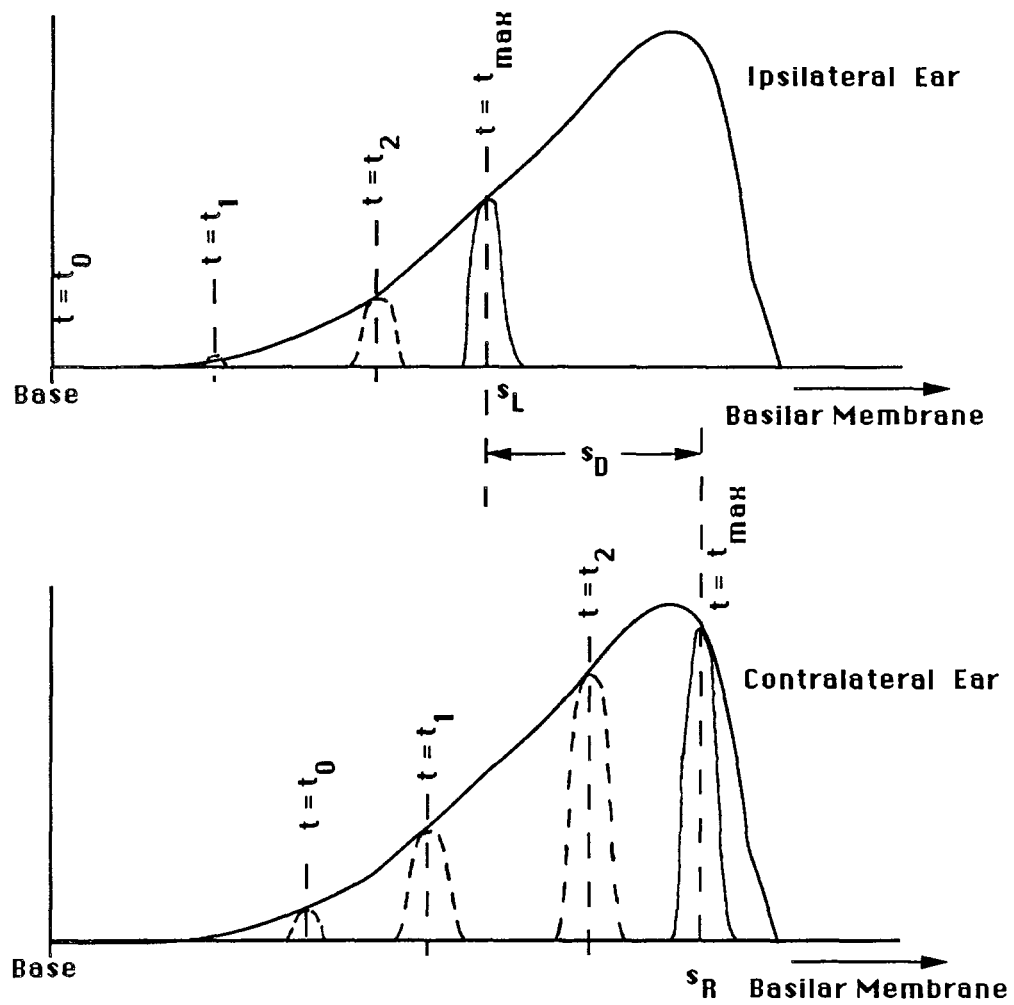


Figure 5.1: Travelling peaks on the basilar membranes

other hand would make the error value oscillate across the minimum value never actually reaching it; the connectivities therefore would not stabilize to a steady optimal value, but instead would switch back and forth around it.

Having derived an equation for a supervised learning rule, an intuitive understanding of how the rule should work is presented next. It was discussed earlier in chapter 3 as to how the traveling waves on the basilar membrane conform to an envelope which is characteristic to the frequency of the input sound stimulus. While interaural time disparities do not cause any difference in the envelopes of the traveling waves at the two cochlea, the details of the traveling waves within the envelopes will show that the peaks in one of the ears will lead those in the other ear. Assuming that the sound stimulus leads at the contralateral ear and that time $t = t_0$ corresponds to the instant the stimulus is applied to the ipsilateral ear, figure 5.1 plots the spatial view of a traveling peak on the basilar membrane in each ear. The details of a particular peak in the contralateral ear is shown at different time instants as it moves along the membrane with the positions of the corresponding peak in the ipsilateral ear. It can be seen that at any time instant t_n , the peak in the contralateral cochlea always leads its counterpart in the ipsilateral cochlea.

In figure 5.2 the instantaneous activity of the same pair of peaks is plotted against the time axis. It is obvious from figure 5.1 that the activity profile of the peak in time, will be identical to its envelope on the basilar membrane. Since the envelopes of both peaks are identical, the input profile for the peak in the contralateral ear is identical to the activity of the peak in the ipsilateral ear, but it is shifted in

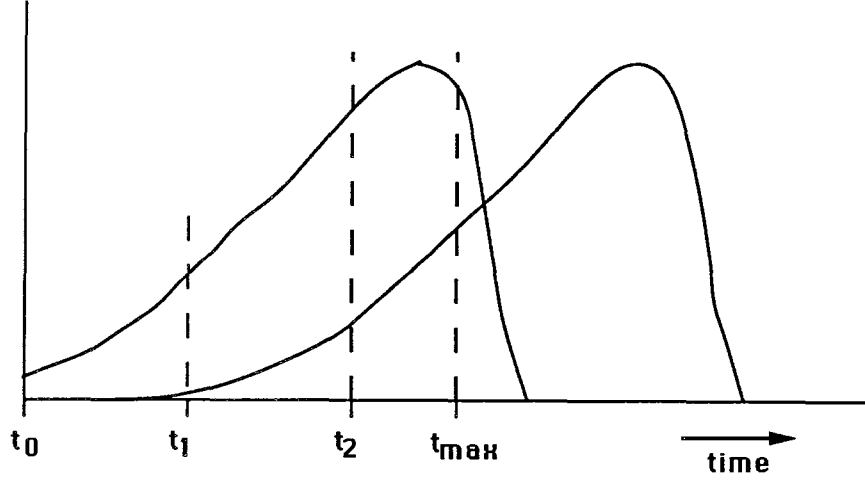


Figure 5.2: Peak amplitudes on the basilar membranes in time

time because of the interaural time disparity. The connectivity update at each time instant is proportional to the output of the binaural(output) neuron at that instant which is in turn proportional to the product of the inputs from each ear. Hence maximum update occurs at that time instant when the product of the inputs from the two ears is a maximum. This would occur approximately at time t_{max} as shown in figure 5.2. It should be emphasized that all disparities shown in the figures have been considerably amplified for visual clarity. At time t_{max} , the peak in the contralateral cochlea is at location s_L and at location s_R on the ipsilateral cochlea. Hence, the strongest connectivity for this shift forms between the output neuron, and the neuron at s_L in the contralateral ear, and the neuron at s_R in the ipsilateral ear. Thus, the temporal disparity between the traveling waves in the two ears causes a spatial disparity of s_D between the connectivities to the inputs from the two ears. As the temporal disparity increases (decreases), it is obvious that the spatial disparity also correspondingly increases (decreases).

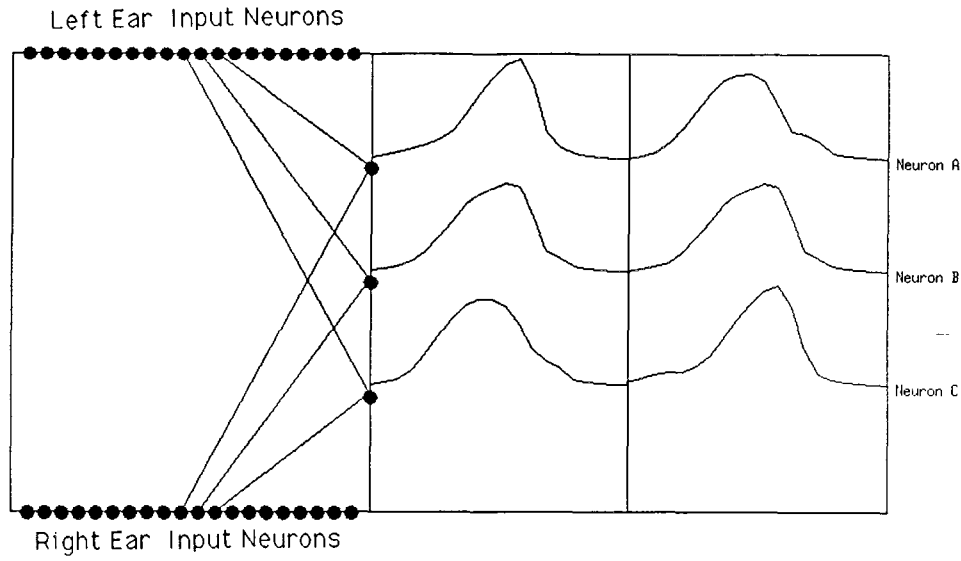


Figure 5.3: Connectivities obtained when binaural neurons have no time shifts

III Implementation Results

The supervised learning rule solution was implemented for three cases :

- The output neurons have no time shifts and the inputs from both ears reach the neuron at the same time.
- The neurons have enough time shifts in them to compensate the shifts between the input patterns from the two ears.
- Each neuron has an inherent time shift but the shifts are not enough to compensate for the external time delays in the signal.

In all the cases, the network was trained on 21 shifts for a 600 Hz. input tone. The shifts ranged from a phase shift of $\frac{2\pi}{3}$ left ear leading, to a shift of $\frac{2\pi}{3}$ right ear leading, in steps of $\frac{\pi}{15}$. There were 3 output neurons receiving inputs from

20 input neurons centered around the CF(600 Hz. in this case) from each of the basilar membranes. Assuming that the output neurons are named A, B and C. The teacher input at neuron A was fixed as ON for a phase shift of $\frac{\pi}{3}$ where the left ear is leading. Neuron B is ON for a zero phase shift between the two ears while Neuron C is ON for a phase shift of $\frac{\pi}{3}$ with the right ear leading. For each case, the output is graphically illustrated in two parts : the right side of the figure will show the pattern of connectivities of each output binaural neuron to all the input neurons; the left side of the figure is a representation of the point of maximal connectivities for each of the neurons. A line is drawn connecting the input neuron associated with the maximum and the output neuron, so that any disparities between points of maximal connectivity are immediately apparent.

In the first case (See fig 5.3), the connectivities of neuron B are symmetric since the inputs from both ears are identical. Maximal connectivity is to the input fiber corresponding to the characteristic frequency of the sound on the basilar membrane and occurs at the same point on both ears. Neuron A which received a leading input from the left ear shows an asymmetry in its connections. The input to which maximum weight is assigned has now moved away from the CF towards the left on one membrane and towards the right on the other ear. The exact reverse occurs at neuron C. Thus it is seen that when there are no time delays in the output neurons to compensate for the external time shifts between the two inputs, the connectivities align themselves in a stereausis network fashion, i.e., a spatial disparity in the point of maximal connectivity along the membrane is observed for neurons which are active for a non-centered input. A comparison of the connectivity

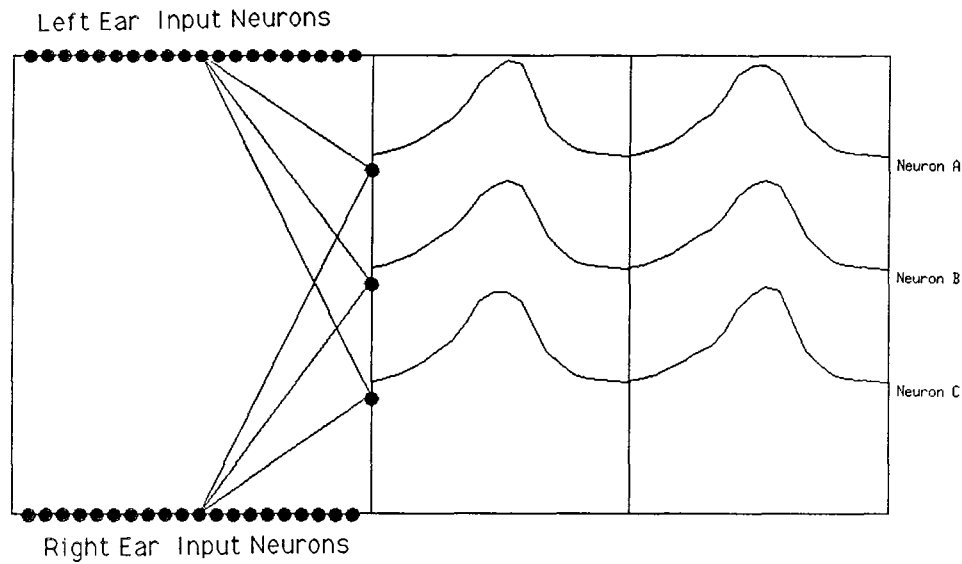


Figure 5.4: Sufficient neural delays are present to compensate the external shifts

pattern obtained by this rule and the one proposed in chapter 3 would show that they are identical.

In case 2, there is a time shift associated with each output neuron, which is equal to the time delay between the two inputs for which that neuron is active. The intrinsic time lag of the neuron would compensate completely for any external delay in the sound caused by the source not being centered. Thus, although the sound at the two ears is not centered, at the neurons A, B and C it would appear so since the input from the ear with the leading signal would arrive later at the neuron by a time equal to the amount by which it was leading. In such a case, the distribution of weights at both ears would be identical and maximal connection would be to the same input fiber on both membranes. This means that if an internal delay between the left and right inputs to a neuron could compensate for an external

time shift between the signals, corresponding locations on the basilar membrane with the same CF would be correlated. Thus a Jeffress-like ([19]) connectivity pattern would emerge. See figure 5.4.

The most probable scenario that exists in the human body however must be a combination of the two extreme cases discussed above. While small displacements on the azimuth would cause shifts between the ears small enough to be compensated by internal shifts, larger shifts may not be compensated in a similar fashion. So, for smaller shifts, the neurons would connect themselves in a Jeffress fashion where no spatial asymmetries would be required. As the shifts increase however, the neurons would look for an alternate way to incorporate the external shifts since sufficient internal delays do not exist and would achieve this by spatial separations between the connections on the membrane. This can be seen in figure 5.5 where the spatial movement of the point of maximal connectivity is not as much as in the first case since the neuron now relies partially on the internal delay between the paths from the two ears to it. For the same phase shift of $\frac{\pi}{3}$, while the spatial disparity of the points of maximum weights was 2 points on the basilar membrane in case 1, it has reduced to just 1 neuron in case 3.

IV Conclusion

The learning rule proposed in this chapter relies largely on the assumption that there is some sort of teacher which is able to distinguish between various temporal shifts and accordingly set the response of the binaural neurons for each input. The most obvious choice for such a teacher would be a visual map. The location of the

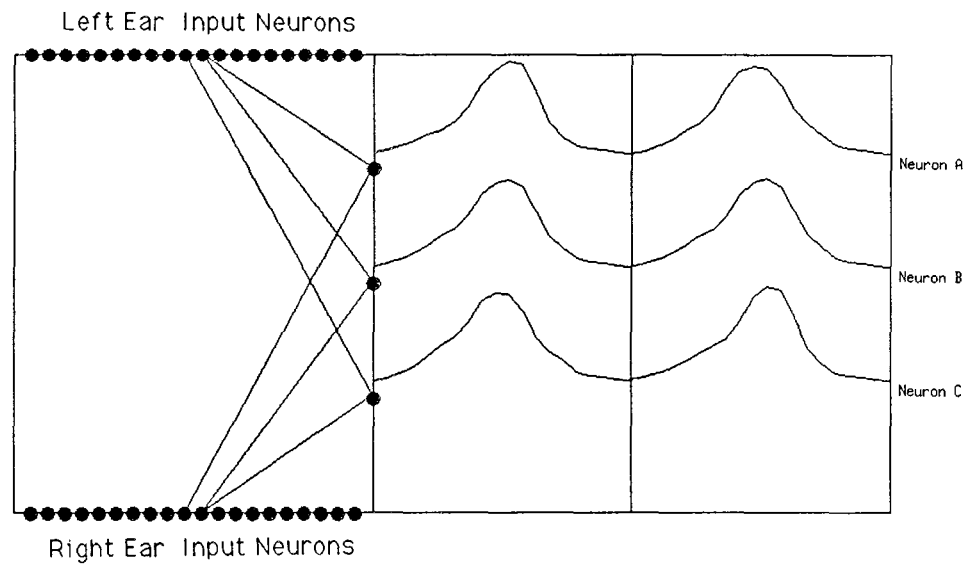


Figure 5.5: Connectivities when neural delays present are not sufficiently large

sound as identified visually, would decide which neuron should be favorable to the input. The demerits of such an assumption are discussed in the next chapter. Yet, if such a teaching input were identified, the delta rule based supervised learning rule proposed would be a very possible solution to the sound lateralization learning process.

AN UNSUPERVISED LEARNING SOLUTION

I Statement of the Problem

The problem requiring solution is the same one for which a supervised learning rule (based on the delta rule) was proposed in chapter 5 : sound lateralization. There is a serious drawback to the supervised update process : the requirement of a teacher input which fixes the ideal output of a binaural neuron for each input pattern presented. If the model were to be biologically realistic, the only source for such a teacher would be the visual map. Based upon the visual location of a sound, the temporal shift between the excitations at the two ears is associated with a displacement on the azimuth. However no synapses between the auditory and visual maps have been found at or before the level of the olivary complexes, which is where most of the early binaural processing occurs. This also poses a problem in explaining the ability of visually impaired people to localize sound accurately. In view of the lack of sufficient physiological evidence, some other learning rule is required which is unsupervised in nature, one which does not rely on any teaching signal/input. The only input available to the network is the cochlear outputs from the two ears identical to those used in the supervised learning case.

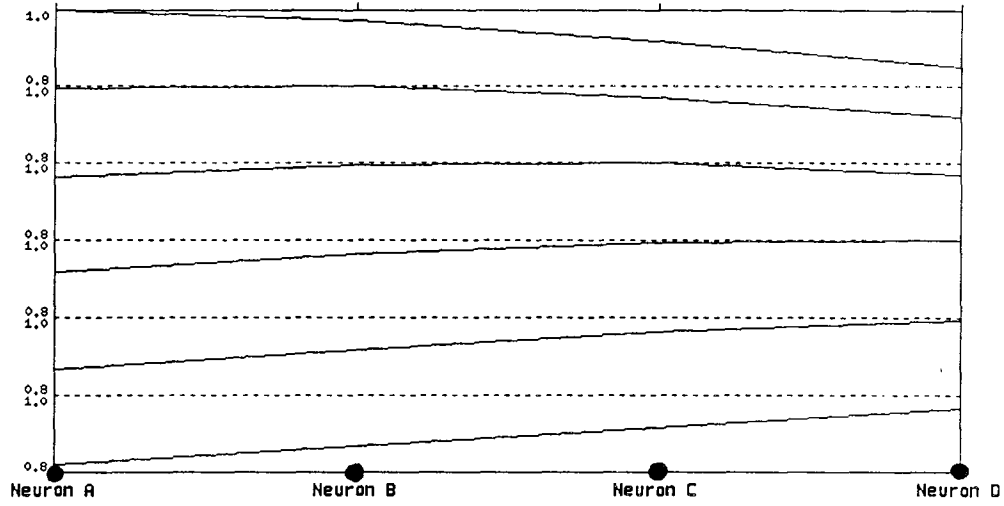


Figure 6.1: Output pattern of Binaural neurons with different delays
for different shifts

The network consists of a layer of input neurons which are excited directly by the basilar membrane. The input neurons are all connected to a layer of output binaural neurons. Each output neuron has an interaural delay associated with it i.e., the input from one ear is delayed relative to the signal from the other ear by a fixed time interval. There are no other signals or inputs in the system.

II The Learning Algorithm

The unsupervised rule proposed relies upon some ordered intrinsic time shifts to exist along the output binaural layer. Assume that there is an array of N binaural neurons such that they are in order of increasing internal time delay, contralateral ear always leading. The first neuron in the row receives its inputs from both the ears at the same time while the i^{th} neuron sees a delay of iT between the signals

from the two ears. If two signals relatively delayed by an amount xT (where x is an integer) were presented to the two ears such that the leading signal is at the ipsilateral ear, the network will perform an approximate cross correlation of the two signals and hence neuron x shows maximal activity. However, for any shift xT larger than the maximum interaural delay NT , neuron N would always respond maximally. Figure 6.1 shows the output activity for an array of 4 neurons named A, B, C and D whose internal delays are 0.0 msec to 0.3 msec in steps of 0.1 msec in response to input signals from the two ears relatively delayed by 0.0 to 0.5 msec in steps of 0.1 msec. It can be seen that for a shift of 0.0 msec, neuron A shows maximal activity, neuron B for a shift of 0.1 msec, neuron C for a shift of 0.2 msec and neuron D for all shifts equal to or greater than 0.3 msec as was expected from the preceding discussion.

The learning rule being proposed uses a competitive learning strategy wherein only the maximally activated neuron updates its connectivities. The output activity profile of the neurons for different shifts depends on the varying internal delays of the neurons and hence in the absence of internal delays, this learning rule would not produce an ordered network. The weight update rule can be derived based on the delta rule as follows :

Assume a network with M output neurons and N input neurons where w_{ij} is the connectivity between output neuron i and input neuron j and W_{ik} is the connectivity between output neuron i and output neuron k . Note $W_{ii} = 0$ for all i .

The error E_i at an output neuron i is defined as the difference between the final output at the neuron after inhibition from the other output neurons and the output

due to the inputs from the two ears alone.

$$E_i = \frac{1}{2} (o_i - (\sum_{l=1}^N w_{il} x_l * \sum_{r=1}^N w_{ir} x_r))^2 \quad (6.1)$$

x_l is the input from the left cochlea and

x_r is the input from the right cochlea

o_i is the final output at neuron i and is defined as follows

$$o_i = f(\sum_{l=1}^N w_{il} x_l * \sum_{r=1}^N w_{ir} x_r + \sum_{k=1}^M W_{ik} o_k) \quad (6.2)$$

where $f(x)$ is the step function,

$$f(x) = \begin{cases} 1 & x > \theta \\ 0 & x \leq \theta \end{cases}$$

θ is the threshold and W_{ik} are such that the output layer forms a winner take-all network with unit vectors constituting the only stable states of the network.

We use gradient descent to determine the update rule :

$$\frac{\partial E_i}{\partial w_{ij}} = (o_i - \sum_{l=1}^N w_{il} x_l * \sum_{r=1}^N w_{ir} x_r) * (o'_i - x_j \sum_{r=1}^N w_{ir} x_r) \quad (6.3)$$

$$o'_i = f'(\sum_{l=1}^N w_{il} x_l * \sum_{r=1}^N w_{ir} x_r + \sum_{k=1}^M W_{ik} o_k)$$

The value of o_i will always be either 0 or 1 and hence, $f'_{(x)}$ is always 0. Hence

$$o'_i = 0$$

Thus

$$\frac{\partial E_i}{\partial w_{ij}} = -x_j (o_i - \sum_{l=1}^N w_{il} x_l * \sum_{r=1}^N w_{ir} x_r) \sum_{r=1}^N w_{ir} x_r \quad (6.4)$$

By the gradient descent rule,

$$\begin{aligned}
\Delta w_{ij} &\propto -\frac{\partial E_i}{\partial w_{ij}} \\
\Delta w_{ij} &= \eta x_j (o_i - \sum_{l=1}^N w_{il} x_l * \sum_{r=1}^N w_{ir} x_r) \sum_{r=1}^N w_{ir} x_r \\
\Delta w_{ij} &= \eta x_j o_i \sum_{r=1}^N w_{ir} x_r - \eta x_j \sum_{l=1}^N w_{il} x_l * \sum_{r=1}^N w_{ir} x_r \sum_{r=1}^N w_{ir} x_r \quad (6.5)
\end{aligned}$$

If the weight w_{ij} is restricted to being between 0 and 1, and since the input is normalized to a maximum of 1, Δw_{ij} is less than zero if o_i is 0 and is always positive if $o_i = 1$. In equation 6.5 above, the second term on the right hand side is a balancing term which prevents the connectivity from becoming too large. If the weights were constrained by some other method, the balancing term could be eliminated from the equation. Also since the factor $\sum_{r=1}^N w_{ir} x_r$ is always positive it can be merged into the the constant η thus simplifying equation 6.5 to the following update equation

$$\Delta w_{ij} = \eta * x_j * o_i \quad (6.6)$$

where η is the learning rate.

As mentioned earlier, some sort of constraint is required to prevent the weights from becoming too large. Hence, a restriction is imposed on the length of the connections to each neuron. The length of a neuron's connectivities is defined as the square root of the sum of the squares of the weights of all the links to that neuron. So after every update, the weights are adjusted so that the length of the connections to each neuron is a constant.

Let us assume that there are N links to a neuron i each having a weight w_{ij} associated with it. The length of the connectivity for neuron i is L_i defined by the

expression

$$L_i = \sqrt{\sum_{j=1}^N w_{ij}^2} \quad (6.7)$$

If the weights are now updated to v_{ij} , each weight must be altered by a factor α to maintain the constraint. Thus

$$L_i = \sqrt{\sum_{j=1}^N [\alpha v_{ij}]^2}$$

Hence

$$\alpha = \frac{L_i}{\sqrt{\sum_{j=1}^N v_{ij}^2}} \quad (6.8)$$

In the absence of this constraint, if a particular neuron wins the competition more often than the others, its connectivities would increase continuously until they become so high that, irrespective of the input pattern presented, this neuron would always win, and this would increase weights further ([30]). It can be seen that restricting the length of the connection to a constant has a similar effect to decaying the weights at each update (as was done in the supervised training rule), the important difference being that the decay rate here is not fixed, but changes depending upon the existing connectivities. This constraint is more effective than merely saturating a connectivity beyond a certain value, since it maintains the shape of the connectivity pattern while restricting the values of the weights at the same time.

An intuitive description of the working of the rule described above in equation 6.6 is now presented. Assume an array of neurons 1, 2,, N arranged in order of increasing axonal delays. At lower interaural time delays, since there are sufficient internal neural delays to compensate the external delay, there is one

particular neuron for which the axonal delay completely cancels out the interaural time difference. Hence for any input shift, there is one neuron which sees the inputs from both ears as arriving at the same time. Thus for that shift this is the neuron which will respond maximally. As per the learning rule described earlier, the connectivities of that neuron are tuned to respond even better to that particular shift. Thus each neuron in the array is tuned to recognize a distinct interaural time disparity. However the connectivity profile for all of these neurons is identical since they all see identical inputs from both ears.

For larger interaural shifts, which cannot be compensated for through axonal delays, the neuron with the largest axonal delay (neuron N at the edge) initially responds best and hence its connectivities get updated for all such shifts. As is explained below, this neuron gradually becomes selectively responsive only to the largest one or two interaural shifts. The $N - 1^{st}$ neuron then begins to respond to the remaining shifts, and in turn becomes selective to the largest remaining shifts. This propagating process continues until all interaural shifts are assigned to the output neurons. This process is explained with the aid of figure 6.2, where we assume that the input is leading in the ipsilateral ear. Two peaks are shown on the ipsilateral membrane representing two different interaural shifts; peaks $Q1$ and $Q2$ are the ipsilateral counterparts of peak $P1$ for two different shifts, $Q2$ obviously corresponding to the larger shift. Let us assume that the input to the binaural output neurons is initially derived from the region $I1$ to $I4$ on the ipsilateral cochlea and from $C1$ to $C4$ on the contralateral cochlea. When peak $Q1$ moves from $I1$ to $I2$, there is no corresponding peak from the contralateral ear since the peak lags in

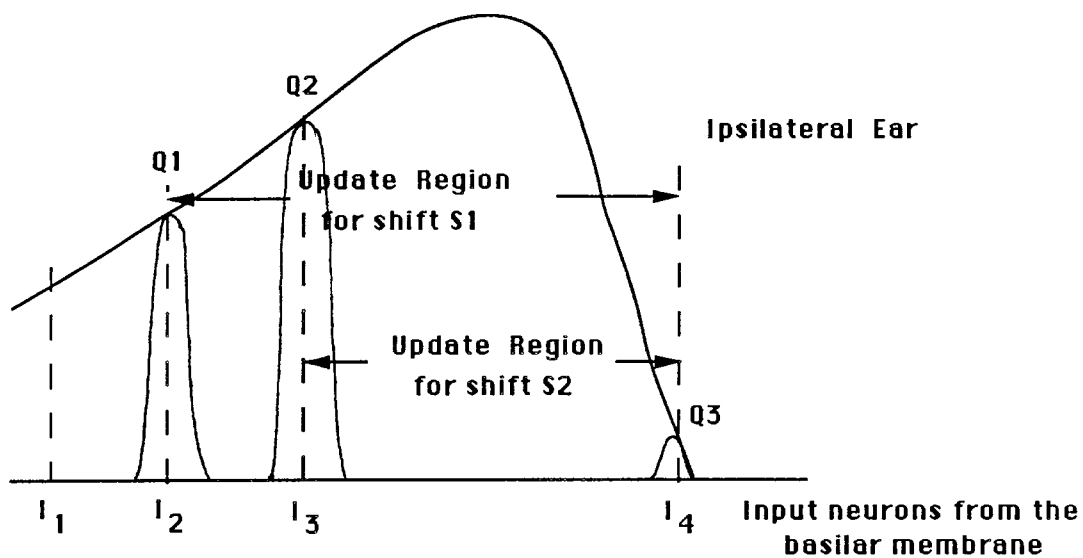
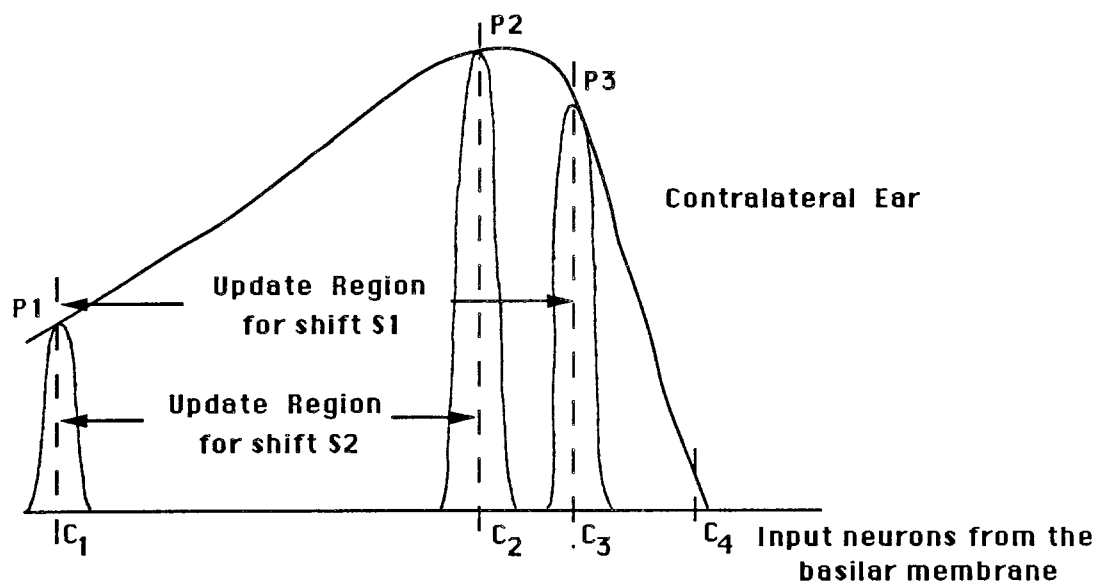


Figure 6.2: Corresponding peaks on the contralateral and ipsilateral cochlea

that ear. Since incremental update of the connectivities only occurs when input is present in both cochlea, the positive update region on the ipsilateral ear for shift $S1$ is restricted only to the inputs from $I2$ to $I4$. Similarly, on the contralateral ear, the update region for $S1$ is from $C1$ to $C3$. If the shift is increased to $S2$, the peak $Q2$ has to travel to $I3$ before any input appears in the contralateral ear. Thus the update region reduces to the portion between $I3$ and $I4$ on the ipsilateral membrane and $C1$ to $C2$ on the contralateral membrane. Thus it is seen that depending upon the time disparity between the inputs, only some of the connectivities get incremented while the others just decay during the normalization. This results in two characteristics of the connectivity pattern :

1. The connectivities between the output neuron and the input neurons from the two ears become asymmetric.
2. As the interaural time disparity becomes larger, the update regions become smaller.

Now consider what happens to an output neuron that initially responds to both shifts $S1$ and $S2$. From figure 6.2, the connectivities to this neuron from the region $I3$ to $I4$ are updated for shifts $S1$ and $S2$. However, the inputs $I2$ to $I3$ are updated only for shift $S1$ and decay for shift $S2$. Consequently the connectivities to the region $I2 - I3$ begin to weaken, and with it, the response of this neuron to the $S1$ shift inputs, i.e., it becomes selectively tuned to the largest shift at its input. Simultaneously, its output response to the smaller shift keeps decreasing until another output neuron's response becomes relatively stronger, and the same

process occurs at this new neuron. In general, then, selectivity to the largest shifts is propagated from the N^{th} edge neuron towards the center neurons in a process in which higher axonal delay neurons relinquish their maximal response to a particular interaural time disparity to neurons with smaller axonal delays. The extent of this translation depends upon the number of uncompensated (by axonal delays) disparities which are presented to the network. Thus, for higher interaural shifts, the same neuron may respond to more than one interaural delays, while at lower shifts there is a one to one correspondence between the delays and the responding neurons, i.e., the resolution of the resulting “spatial” map is best near the center and gradually decreases towards the edges.

III Implementation Results

In this section of the chapter, the results obtained by implementing the unsupervised competitive learning rule and the significance and reasoning behind the kind of results obtained are outlined. The kind of processing done in the course of the learning process was as follows :

1. The time shifted input waveforms are processed in the cochlear model discussed in chapter 3 to obtain the traveling waveforms on the basilar membrane.
2. A pair of waveforms which are shifted by a randomly chosen time period are then presented to the network for approximately 4 periods of the input signal (6.4 msec for a 600Hz waveform). During this period, the outputs of the

neurons are averaged (the time can be considered as the time constant of the neurons) and the neuron with the maximum output is selected for update. In this step however, none of the neuron connectivities are updated.

3. The input patterns are then presented again for one period of the input signal (1.6 msec for 600Hz. input) during which the connectivity to the neuron which won the competition in step 2 are updated based on the learning rule of equation 6.6.
4. Each time the connectivities are updated, they are adjusted so that the length of the connection to each binaural neuron is a constant. The adjustment is done according to the procedure discussed in the previous section

The rule was tested for two different conditions :

- The internal delays of the output binaural neurons were sufficient to compensate for the maximum shift between the signals presented at the two ears during the training process. Here there are as many output neurons as there are shifts presented to the network, i.e., there is a neuron with an equal and opposite internal delay corresponding to every input shift.
- The maximum temporal delay between the binaural excitations is greater than the maximum internal delay in the binaural neurons.

The results are displayed in the same way as in previous chapters, the distribution of the weights on the right side and the points of maximal connectivities on the left.

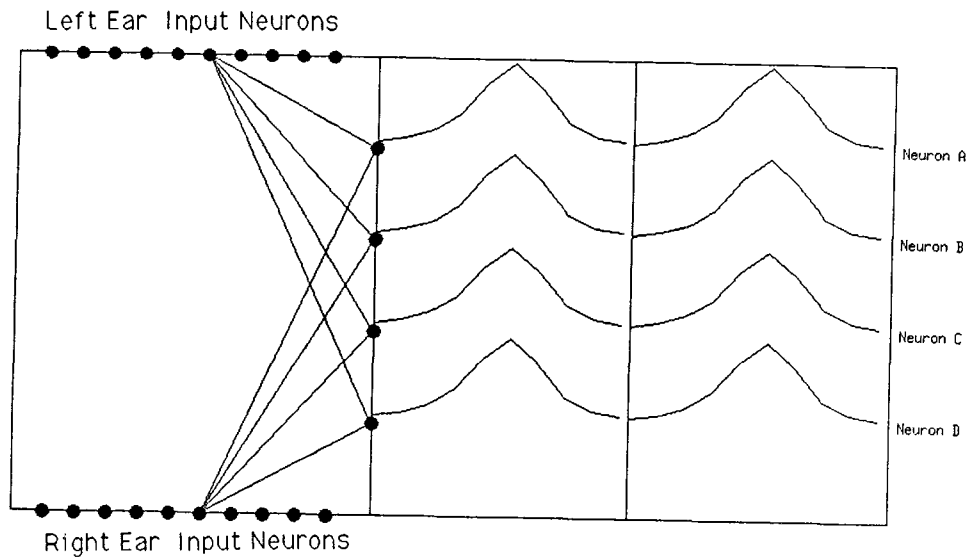


Figure 6.3: Connectivity pattern got when every input shift is compensated
by the internal delay of a neuron

In the first case, the results obtained are identical to what could be expected based on the results obtained in the supervised case. Due to the output activity profile already explained, each of the output neurons would respond maximally to one of the shifts being presented. Since there is a one-to-one correspondence between the shifts and the output neurons, a different neuron would respond favorably to each of the shifts presented. That neuron for which the external delay is compensated completely by its intrinsic delay will respond to that pair of shifted input patterns. Hence a different neuron will update its connectivities for each of the input patterns and that neuron will see the pair of patterns as arriving at the same time. The neurons would change their weights so that they become better tuned to a particular pattern. However the connectivities at all the neurons remains symmetric since

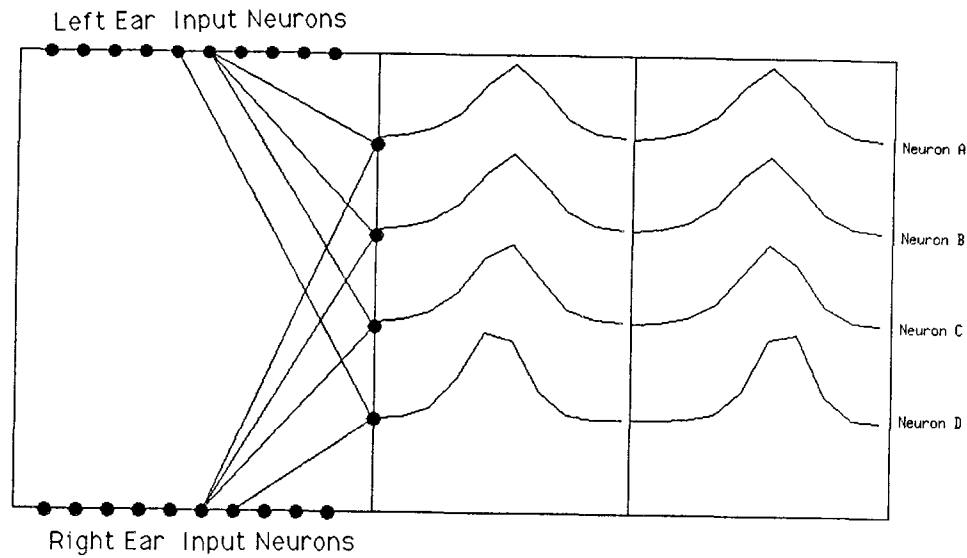


Figure 6.4: Weights obtained when more input shifts are presented than output neurons they are all seeing signals arriving at the same time from the two ears. Thus inputs from corresponding CF locations on the two basilar membranes will be correlated thus resulting in a Jeffress model like network. The results obtained are shown in figure 6.3.

In the second case, more shifts were presented than there were output neurons. The difference between the delays was 0.15 msec for the neurons and the input patterns. However while there were only 4 output neurons, the minimum internal delay being zero milliseconds and the maximum being 0.45 msec, there were 6 input pattern pairs presented. The pair with the minimum delay was centered (in phase) while the maximum temporal delay between the two signals was 0.75 msec. Let us call the neurons in order of increasing internal delays as neurons A to D respectively and the input pattern pairs as $P1$ through $P6$ again in order

of increasing delay. Initially neuron A responds to pattern $P1$, neuron B to $P2$, neuron C to $P3$ and neuron D would update its connectivities for all the remaining patterns. By a update process similar to the one described earlier, at the end of the training phase, neuron D becomes selectively tuned to the highest two shifts, viz., $P6$ and $P5$. Neuron C responds to patterns $P4$ and $P3$, neuron B to $P2$ and neuron A to $P1$. Thus, while the localization is accurate for smaller phase shifts (smaller displacements from the centered case on the azimuth), as the shifts get larger, the sound sources seem to be more fused and localization becomes more blurred. This is a commonly experienced phenomenon in binaural sound lateralization.

IV Conclusion

A learning rule was proposed which did not rely upon any external signal or teacher and learned to localize sounds accurately based solely upon the inputs received from the two ears. The learning rule proposed uses axonal delays to establish an initial rudimentary order of the input patterns into groups, each such group associated with a different output neuron and then tune each of the neurons further. Other algorithms for self organizing maps have been presented by several others such as Kohonen ([23]). The Kohonen algorithm was implemented to solve the problem of sound lateralization. Although no encouraging results were obtained, the algorithm cannot be discounted. The solution most probably requires the determining of the optimal parameters so that some meaningful results can be obtained. The algorithm proposed in this report is more robust and the learning rate, which is the only free parameter can be chosen from a fairly wide range of values.

CONCLUDING REMARKS

This thesis covered two major areas, the proposal of a network topology to perform sound lateralization, and the implementing of learning schemes to form such a network starting with arbitrary connectivities, the latter topic being dealt with in greater depth. Two learning rules were proposed, a supervised rule and an unsupervised one. The premise of the stereausis network is the total absence of neural delays which is perhaps a biologically unrealistic view. However, the aim of this network is to emphasize that it is possible to determine temporal shifts between the signals at the two ears by using spatial disparities on the basilar membrane alone. It is not a long step to incorporate delay lines into the stereausis topology and to show how the temporal delays intrinsically present in the human body and the spatial shifts can work complementarily or even in conjunction to compensate the time shifts between the signals at the two ears. Thus a realistic network emerges which efficiently uses all the tools that it has available, viz., neural delays and spatial shifts, to achieve its goals. It must be emphasized here that in order to lateralize a sound, the primary peak alone is sufficient, thus restricting the spatial disparities to a small region only. Thus, fibers with widely differing CFs are not correlated at all.

The learning rules attempt to explain the formation of a network which can localize sound on the azimuth. The update rules used are general and are not biased towards any particular kind of topology. The only inputs to the network are the travelling waves from the two ears and the desired goal, viz., lateralization of sound. In the supervised case, the goal is in the form of an explicitly specified output for each input pattern pair. Using the error at the binaural neurons and the Widrow-Hopf rule, the weights were updated. In the unsupervised case, the presence of delays in the neurons is used to generate an output profile which resolves the competition between the neurons as to which neuron shall update its connectivities. The interesting feature of the results obtained in both cases was that they encompassed the previously proposed Jeffress network as well as the stereausis network proposed in this thesis. In the case when sufficient delays are present in the binaural neurons to compensate the maximum temporal shift at the two ears, for both the learning rules, a Jeffress like network emerges where inputs from corresponding CF locations on the two cochlea are cross correlated to determine the shift. In the other extreme when no delays exist at all, the unsupervised rule breaks down since there is no criterion by which the competing neurons can resolve which one will update its weights. While Kohonen ([23]) does suggest that an automatic sorting of the inputs into groups would occur, implementation of that technique for this problem did not result in any ordering at all. In the supervised case, however, a lack of neural delays resulted in a network which depended on spatial disparities alone, similar to the stereausis topology. Finally, in a realistic case where smaller delays were available (though not enough to entirely compen-

sate the external delays), these delays in conjunction with small spatial disparities on the membrane (required only for higher phase shifts) were able to compensate the external delays and thus localize the source. In the supervised case again, an output neuron exists corresponding to every distinct input pattern pair presented and hence each location in space can be uniquely identified. In sharp contrast, a one-to-one correspondence is not present between the binaural neurons and the input pairs in the excitation set in the unsupervised case. It is observed that while small phase shifts uniquely correspond to a certain neuron, larger shifts are more grouped in that, a single binaural neuron can respond maximally to more than one input pattern.

The network learning rules proposed have been tested successfully for low frequency time shifted inputs. The next step would be to determine the effect of such rules on signals with level differences. At higher frequencies, phase locking does not occur and hence time shifts are of no use at all. It has been shown that the stereausis network can detect level differences ([36]); but the training rules have not been implemented at higher frequencies. Even at lower frequencies, level differences show a time intensity trading where one can compensate the other. Similar processing has been proposed for visual depth perception or binocular vision. This would imply that such a network is a general one which could possibly be implemented in other kinds of processing as well.

BIBLIOGRAPHY

- [1] A. Barto, C.W. Anderson, and R. Sutton. Synthesis of nonlinear control surfaces by a layered associative net. *Biological Cybernetics*, 43:175–185, 1982.
- [2] A. Barto and R. Sutton. Landmark learning: an illustration of associative search. *Biological Cybernetics*, 42:1–8, 1981.
- [3] G.von. Bekesy. Zur theorie des horens. *Physikalische Zeitschrift*, 31:857–868, 1930. Translation in chapter 8 of Experiments in hearing, G.V.Bekesy (E.G.Weaver, Ed. and Trans.) New York, McGraw-Hill, 1960.
- [4] J. Blauert. *Spatial Hearing*. M.I.T. Press, Cambridge, Massachusetts, 1983.
- [5] H.S Colburn. Some physiological limitations on binaural performance. 1969. Unpublished dissertation, Ph.D, Electrical Engg. Dept., MIT.
- [6] H.S Colburn. Theory of binaural interaction based on auditory nerve data i. detection of tones in noise. *Journal of the Acoustical Society of America*, 61:525–533, 1977.

- [7] H.S. Colburn. Theory of binaural interaction based on auditory nerve data i. general strategy and preliminary results on interaural discrimination. *Journal of the Acoustical Society of America*, 54:1454–1470, 1973.
- [8] T.R. Dolan and D.E. Robinson. Explanation of masking level differences that result from interaural intensive disparities of noise. *Journal of the Acoustical Society of America*, 42:977–981, 1967.
- [9] N.I. Durlach. Note on binaural masking level differences as a function of the interaural correlation of the masking noise. *Journal of the Acoustical Society of America*, 23:452–462, 1964.
- [10] N.I. Durlach. On the application of the ec model to interaural jnd's. In J.V. Tobias, editor, *Foundations of modern auditory theory*, New York: Academic Press, 1972.
- [11] N.I. Durlach and H.S. Colburn. Models of binaural interaction. In E.C. Carterette and M.P. Friedman, editors, *Handbook of Perception*, pages 467–518, Academic Press Inc., 1978.
- [12] K. Fu. Learning control systems - review and outlook. *IEEE transactions on Automatic Control*, 1970.
- [13] K. Fukushima. Cognitron: a self-organizing multilayered neural network. *Biological Cybernetics*, 20:121–136, 1975.
- [14] D.M. Green and W.A. Yost. Binaural analysis. In W.D. Keidel and W.D. Neff, editors, *Handbook of Sensory Physiology*, Berlin, New York, 1975.

- [15] S. Grossberg. Adaptive pattern classification and universal recoding: part 1. parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23:121–134, 1976.
- [16] E.R. Hafter and S.C. Carrier. Masking level differences obtained with pulsed tonal markers. *Journal of the Acoustical Society of America*, 47:1041–1047, 1970.
- [17] H.K. Hartline. *Studies on excitation and inhibition in the retina*. Rockefeller University Press, New York, 1974.
- [18] M.H. Holmes and J.D. Cole. Cochlear mechanics: analysis for a pure tone. *Journal of the Acoustical Society of America*, 767–778, 1984.
- [19] L.A. Jeffress. A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41:35–39, 1948.
- [20] W.D. Keidel and W.D. Neff. *Handbook of Sensory Physiology*. Berlin, New York, 1975.
- [21] W.E Kock. Binaural localization and masking. *Journal of the Acoustical Society of America*, 22:801–804, 1950.
- [22] T. Kohonen. *Associative Memory: A Systems Theoretical Approach*. Springer Verlag, Berlin, 1977.
- [23] T. Kohonen. *Self Organization and Associative Memory*. Springer Verlag, Berlin, 1984.

- [24] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.
- [25] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of the Royal Society of London*, 204:301–328, 1970.
- [26] D. McFadden. Masking level differences determined with and without interaural disparities in masker intensity. *Journal of the Acoustical Society of America*, 44:212–223, 1968.
- [27] J.M. Mendel and R.W. McClaren. Reinforcement learning control and pattern recognition systems. In E. Dale and D. Michie, editors, *Adaptive Learning and Pattern Recognition Systems: Theory and Applications*, pages 137–152, Edinburg: Oliver and Boyd, 1970.
- [28] K.S. Narendra and M.A.L. Thathachar. Learning automata - a survey. *IEEE Transactions on Systems, Man and Cybernetics*, 4:323–334, 1974.
- [29] D.E. Rumelhart, G.E. Hinton, and J.L. McClelland. A general framework for parallel distributed processing. In D.E. Rumelhart and J.L. McClelland, editors, *Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Volume 1: Foundations*, pages 45–76, The MIT Press, Cambridge, Massachusetts, 1986.
- [30] D.E. Rumelhart and D. Zipser. Feature discovery by competitive learning. In D.E. Rumelhart and J.L. McClelland, editors, *Parallel Distributed Process-*

- ing, *Explorations in the Microstructure of Cognition, Volume 1: Foundations*, pages 151–193, The MIT Press, Cambridge, Massachusetts, 1986.
- [31] B. Sayers. Acoustic-image lateralization judgement with binaural tones. *Journal of the Acoustical Society of America*, 36:923–926, 1964.
 - [32] B. Sayers and E. C. Cherry. Mechanism of binaural fusion in the hearing of speech. *Journal of the Acoustical Society of America*, 29:973–987, 1957.
 - [33] M.R. Schroeder. New viewpoints in binaural interaction. In E.F. Evans and J.P. Wilson, editors, *Psychophysics and Physiology of Hearing*, pages 455–467, Academic Press, 1977.
 - [34] S.A. Shamma. Speech processing in the auditory system. ii. lateral inhibition and the processing of speech evoked activity in the auditory nerve. *Journal of the Acoustical Society of America*, 78:1622–1632, 1985.
 - [35] S.A. Shamma, R. Chadwick, Wilbur J., J. Rinzel, and K. Moorish. A biophysical model of cochlear processing. *Journal of the Acoustical Society of America*, 80:133–145, 1986.
 - [36] Naiming Shen. *A Binaural processing Network without Neural Delay Lines*. Master’s thesis, Univ. Of Md., College Park, MD., December 1987.
 - [37] W.A. van Bergeijk. Variation of a theme of von bekesy : a model of binaural interaction. *Journal of the Acoustical Society of America*, 34:1431–1437, 1962.

- [38] C. von der Malsberg. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100, 1973.
- [39] F.A. Webster. The influence of interaural phase on masked thresholds. i. the role of interaural time deviation. *Journal of the Acoustical Society of America*, 23:452–462, 1951.
- [40] R. Whitworth and L. Jeffress. Time versus intensity in the lateralization of tones. *Journal of the Acoustical Society of America*, 33:925–929, 1961.
- [41] T. Yin and S. Kuwada. Neuronal mechanisms of binaural interactions. In G. Edelman, W. Gall, , and W. Cowan, editors, *Dynamic aspects of neocortical function*, pages 263–314, John Wiley and Sons, New York, 1984. A Neuro-Sciences Institute Publication.