1

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

1

AFIT/GE/ENG/91D-13

Binaural Sound Localization Using Neural Networks

THESIS

Rushby C. Craig

AFIT/GE/ENG/91D-13

Approved for public release; distribution unlimited

AFIT/GE/ENG/91D-13

Binaural Sound Localization Using Neural Networks

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Electrical Engineering

Rushby C. Craig, B.S.E.E.

December 12, 1991

Approved for public release; distribution unlimited

Preface

This research effort was motivated by the desire to make sound localization possible for machines. The experiments of this thesis were designed to investigate the use of Artificial Neural Networks to perform this task. Many of the techniques and the theory used in this effort were based on models of the human auditory system. Consequently, an additional benefit of this research may be increased insight into the mechanisms used by humans in the localization process.

I would like to express my appreciation to several people who have contributed to this thesis effort. First, I thank my advisor, Dr. Rob Williams for his helpful guidance throughout this process. I am also indebted to Dr. Timothy R. Anderson for the inspiration to pursue research in auditory localization, and also for knowledge shared and suggestions given which helped to make this research effort more successful. I am grateful to Capt. Gregory L. Tarr for his help in setting up and using his neural network software. Additionally, I would like to thank Dr. Steven K. Rogers and Dr. Matthew Kabrisky, who provided a wealth of information on pattern-recognition techniques and biological sensory-systems in both classroom lectures and personal conversations. Finally, I would especially like to thank my wife, Cay, and my daughters Melanie and Courtney for their support and understanding during the several months of this endeavor. Without their temporary sacrifice of home and friends for my sake, this tremendous educational experience would not have been possible.

Rushby C. Craig

Table of Contents

List of Figures

Figure Page

Figure Page

Figure Page

List of Tables

Binaural Sound Localization Using Neural Networks

I. Introduction

1.1 Background

Robots are widely used in industry to perform tasks such as movement of materials in warehouses and automobile assembly. In these types of applications, these machines show great promise as a means to increase quality and efficiency in an industrial environment. As engineers look to design more capable and sophisticated robots, it is apparent that these machines could be vastly improved if they were made more human-like. If such human functions as vision, smell, touch, hearing and learning were integrated into a machine it may then be capable of learning and performing tedious and complex tasks that are presently very difficult to program a robot to do.

The Air Force is funding research in robotics and sensory systems for robots. Here at Wright-Patterson AFB, OH, the Bioacoustics and Biocommunications Branch of the Armstrong Laboratory (AL) is involved in investigating the use of sound and hearing in machines.

Much of the auditory research done focuses on the characteristics, capabilities and the possible mechanisms used in human and other mammalian hearing. Some of this research is motivated by the simple desire to understand the human body and its functions. Others are motivated by the desire to understand the principles of perception. Just as early aeronautical scientists studied birds in order to understand the principles of flight, auditory researchers now seek to understand the phenomenon of auditory perception through the study and modelling of biological hearing systems. Modern aircraft do not exactly mimic birds: metal is used in place of featl.ers, propellers and jet engines are used for power rather than flapping the wings. Future machine hearing systems likewise may not exactly follow the method used by biological systems. Nevertheless, living creatures, at this time, remain the best resource for modelling such systems.

One feature of mammalian hearing that appears to be important is the possession of two ears. Colin Cherry, a well known auditory scientist, wrote (7:99):

> I believe it was Epictetus the Stoic who was reported as saying, "God gave man two ears, but only one mouth, that he might hear twice as much as :.e speaks".
>
> This wishful thought may be empty, so far as human wisdom is concerned, but it is curiously near truth on the plane of psychophysics. For the possession of two ears gives us greatly enhanced powers of aural discrimination...

One particular aspect of hearing that is of interest is sound localization. This phenomenon can be demonstrated best by an anecdotal example. Imagine being at a social gathering (such as a cocktail party) in a large room where many people are attending. As you enter, the low noise of music, garbled speech, laughter, and tinkling glasses and dishes seem to fill tl.e room. Moving somewhere toward the center of the room you stop and begin to observantly listen. If you possess normal human auditory capabilities yc u notice that you can recognize the direction and possibly the distance to sounds such as conversations, music, etc. occurring in random radial directions. If you are particularly observant (and maybe a little nosy) you will notice that you have the limited ability to selectively tune into one particular conversation (or other sound) of interest to the exclusion of the others. This example is probably one which nearly everyone has experienced, or else can readily relate to. These abilities provide people with a sense of where sound producing objects are spatially located.

Binaural sound localization will be defined as the process of determining the direction of and perhaps the distance to a sound source using two ears or microphones. The selective tuning capability is another layer of complexity which will not be considered in this thesis effort.

1.2 Problem

A robust method needs to be developed to allow machines to localize auditory signals in near real-time. With this capability, machines such as mobile robots will be more useful and safe in environments where collision hazards exist.

1.3 Summary of Current Knowledge

A great deal of research has been performed over the years on the subject of auditory localization, mainly by medical researchers and psychologists. A result of this research has been several theories about the mechanisms and features which humans use in the localization process. Another result has been a large amount of data collected from experiments documenting the performance of humans in localizing different types of signals in a wide range of circumstances. Many, but not all of the observed characteristics of human localization can be explained by the models and theories which currently exist.

In recent years many researchers have attempted to use pattern recognition techniques such as artificial neural networks (ANNs) to solve problems in related areas such as speech recognition, sonar signal classification and acoustic emission analysis. These successes provided motivation to researchers at Armstrong Laboratory to try using similar techniques on the problem of sound localization. The results of some simple experiments conducted by Anderson (1) using ANNs to localize single tones of known frequency to one of the four quadrants of the horizontal plane have been encouraging. In these experiments the tones were usually correctly localized. In some instances however, ambiguity between sounds in front and to the rear were observed. Because humans have been observed to have this same problem when localizing tones, it was decided that further investigation was warranted.

These previous experiments by Anderson provide the impetus for this thesis effort. Past and current research in the areas of human localization and pattern recognition of acoustical signals will be covered in more detail in chapter 3.

1.4 Objectives

The goals of this thesis are to investigate the use of ANNs for the purpose of performing binaural sound localization on both narrow and wide-band audio signals (i.e. tones and gaussian noise), to implement an ANN to classify the direction of sound sources, and to analyze the performance of these systems based on the results of tests using simulated sound data.

1.5 Scope

As previously stated, the ultimate goal of using ANNs to perform auditory localization is to implement this capability in machines. This research effort however, is restricted to software simulations which will demonstrate the ability of ANNs to recognize the location of sound sources from the features in the signals. The location will be defined as the azimuthal angle to the sound source. Determination of range to the sound source is not considered part of this thesis effort.

1.6 Methodology

The plan of attack for this effort is to investigate the localization of sounds on the horizontal plane. Signals to be used are tones and Gaussian noise.

Tones are narrowband signals, and therefore have a certain frequency associated with them. When attempting to localize tones with ANNs, a question that naturally arises is: what frequency or frequencies of tones should be used in the training and test sets? In general, it seems to be most useful to have a network which can localize tones of any real-valued frequency in the audio range. Cases may exist where tones of only one or perhaps a small set of frequencies need be localized.

For this reason, localization of tones of a single frequency, of discrete frequencies, of random frequencies using discrete training frequencies, and of random frequencies using random training frequencies will be examined.

Different feature sets will be tested in the networks for both tones and noise and the results compared. From these results the feature set providing the best performance will be identified.

1.7 Benefits

Many potential benefits of this type of research exist. As was mentioned, binaural sound localization capability allows machines to sense the direction of objects or events from sounds. This ability helps robots to avoid dangers such as moving in front of a moving airplane on a flightline. Another potential benefit is the discovery of clues about how humans may perform the task of sound localization. Using similar techniques as used in this thesis, aids to assist the hearing impaired with localization are possible. Additionally, it is possible to apply some of the principles used in binaural sound localization to other applications such as radar and sonar (which often use multiple element receiver arrays) in order to reduce the number of receivers used to two.

1.8 Thesis Organization

Chapter 1 introduces the problem being addressed in this thesis effort. Chapter 2 reviews the literature which is important to the theory and techniques used in the research. This review will discuss the types of networks to be used and examples of research which has been performed using these networks. Chapter 3 discusses the methodology used in this thesis effort while chapter 4 reveals the research results. Finally, chapter 5 summarizes the conclusions drawn from the research and makes recommendations for future research.

II. Background

2.1 Introduction

Humans and many other living creatures have the ability to discern the direction from which sounds emanate. The previously mentioned "cocktail party effect" is an example of this.

Some animals have even more highly developed auditory systems than humans. Bats, for example, live in a very dark environment where a visual system is of limited use. Their brains can process the auditory returns from the squeals they produce in order to create a mapping of their surroundings. This enables them to avoid flying into rocks, trees and foliage, and aids them in locating insects to eat (24:60). Dolphins have similar auditory capabilities in their underwater environment (2).

Using these living auditory systems as inspiration, some very useful functions for artificial auditory systems can be imagined. For instance, systems that could calculate the direction and distance to an auditory event would have application in a wide range of military systems as well as in aids for the hearing impaired. Such systems could also provide valuable insight into some of the secrets of how the human brain processes audio information.

2.2 Human Localization Research

Scientists have been studying the subject of human sound localization for many years. Some of the earliest scientific observations of this phenomena were made by Fechner in 1860 (8). Since then, a great deal of research has been conducted in this area by many different people. A result of this research was the discovery that humans probably utilize several different mechanisms in the localization process. This section will review the research in this area, and will emphasize some of the features that are widely believed to be important in human auditory localization.

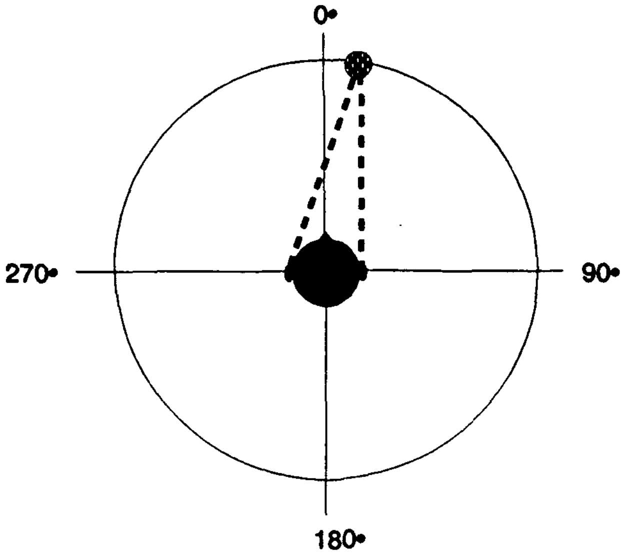

Figure 2.1. Sound source originating at 15 degrees on the horizontal circle.

2.2.1 Interaural Time-Delay. Woodworth (27), in 1938, discovered the importance of interaural time-delay (ITD) in human auditory localization. The ITD results from the difference in the time it takes a sound wave to propagate from an arbitrary location in space to the ears, which are spatially separated by being located on either side of the head. For example, if a sound source is located closer to the left ear, then the left ear will receive the signal first and the right ear will receive the signal at a slightly later moment in time. Other experiments (28:960) have shown that humans are able to detect time delays as short as 20 μsec.

2.2.2 Interaural Amplitude Differences. In 1933 Sivian and White (23) recognized that the path from a sound source to one ear is often occluded by the head. They postulated that this head-shadow effect on the sound signal resulted in varying interaural amplitude differences (IAD) between the sound signals received at the ears, and that this phenomena created important cues for sound localization in humans. Later, evidence was found that seemed to show that the head-shadow effect is important at frequencies at or above 1 kHz (20:448). This threshold feature has been attributed to the relationship between the size of the head and the wavelength of a 1 kHz signal.

2.2.3 Pinna Cues. During the late 1960's Batteau (4) and Blauert (5) contended that the pinna (the external ear structure) performs a very important role in localization. They showed that the pinna acts as a angle-dependent filter on the sound signals as they enter the ear. In 1974, Searle (20:454–455) expanded this idea to suggest that the disparity between the binaural signals may be used as a feature in sound localization, and he went on to show that human subjects were able to detect these disparities. Musicant and Butler (16:1199) published a paper in 1983 containing experimental results which supported these findings. These results showed that frequencies above 4 kHz are necessary for optimal localization of audio signals. The benefit in localization accuracy in the presence of high frequency components was attributed by the authors to spectral cues generated by reflections in the pinna.

2.2.4 Head Movements. Head movements apparently provide additional information for localization. As a person moves his head, he modulates the sound signals entering the ears. Many researchers have recognized that this phenomena improves localization accuracy. This was first observed by Wallach in 1940 (26). Later, in 1974, Lambert (12:166–167) provided support to this observation by showing mathematically how the azimuth and range to an auditory event can be calculated from the rate of change of certain stimulus parameters.

2.2.5 Synthesized Binaural Sounds. Air Force researchers at Armstrong Laboratory (AL), Wright-Patterson AFB OH, have studied the features of interaural time-delay and amplitude differences, pinna cues, and head movements and their use in human sound localization. In 1988, McKinley's master's thesis (sponsored by the AL) (15) dealt with the use of these features to synthesize sounds in headphones which appeared to the listener to be coming from selected directions.

A prototype system was built to carry out the necessary signal processing to synthesize the left and right ear signals from a given sound source signal. The system utilized a pair of high-speed, digital signal processor (DSP) chips, one for each ear signal. A transfer function for each angle in the horizontal circle was calculated using known sound source signals and the corresponding signals received by microphones located in the ears of a manikin. This set of 360 transfer functions were stored in memory, where they could be quickly retrieved and convolved with the source signal to produce the signal received at one of the two ears. This procedure would be carried out for each ear signal in a DSP chip, after which the signals would be converted to analog form, low-pass filtered and output to one of the two audio signal channels leading to the headphones.

The system also utilized a head position measurement device which was attached to the headphones. This enabled the system to determine the position of the head so that the signals being sent to the headphones could be changed as the orientation of the head changed with respect to the direction of the sound source.

2.2.6 Observations on Localization of Different Signal Types. Over the years, researchers selected different types of sound signals for their localization experiments. The types of signals included narrow-band signals such as tones, and wide-band signals like clicks and noise. Because of the different characteristics of these signals, some unique localization characteristics were discovered for different types of source signals.

Wide-band signals (especially noise) have been experimentally shown to be easier to localize than tones (16:1195). The ease of localization for wide- band signals is attributed to the fact that a wide range of frequency components are present, thus allowing all of the human mechanisms (ITD, IAD, and pinna cues) to contribute to the process.

Tones are comprised of one (or a few) discrete frequency components. Localization experiments using tones have helped scientists to understand that localization ability is a function of frequency, and that the frequency range of a signal determines which mechanisms are used in the localization process.

Experiments by Sandel showed that localization of tones by humans was most accurate below the 1500 Hz threshold. From 1500 Hz to 4000 Hz the localization accuracy dropped substantially (19:850). Sandel also observed that time-delays could not be accurately determined on tones above the threshold (19:850). This phenomena can be explained by the periodicity of sinusoids. The maximum interaural time-delay for humans is typically 700 μsec, but varies slightly depending on individual head diameter. A sinusoid with a period of 700 μsec would have a frequency of approximately 1430 Hz. Therefore tones above this frequency would have more than one cycle within the window of the maximum time-delay. With multiple cycles in the window of interest, there is more than one possible time-delay value between the two ear signals. Tones below 1430 Hz would have less than one full cycle, and thus only one possible time-delay value.

Sandel found that above 4000 Hz the localization accuracy on tones began to improve again (19:850). This agrees with the observations of Searle (20:448).

2.3 Artificial Neural Networks

A pattern recognition method which has become a very popular research topic in recent years is that of artificial neural networks (ANNs). ANNs are a mathematical model which attempt to mimic some properties of biological nervous systems. The relationship between ANNs and biological nervous systems is in the manner in which they process information. Both rely on dense interconnections between simple computational elements in order to achieve very high computational rates. There are many tasks that cannot currently be performed well or at all by human engineered systems using traditional sequential computers. Many of these tasks are the very same tasks at which humans and other living creatures excel. For this reason, researchers often look to neural networks as a solution in areas such as speech and image recognition. ANNs are frequently simulated on sequential von Neumann computers, but they can also presently (in some cases) be implemented in hardware which utilizes their parallel nature (13:5).

ANNs are able to perform accurate computation in specific problems only after some training of the network is conducted. Training is accomplished through the use of a learning algorithm which seeks to find an optimal set of interconnection weights for the network based on data provided to the network. This optimal set of weights represents a mapping from the feature space to one of the set of classes.

2.3.1 The Perceptron. One of the first neural network architectures was devised by Rosenblatt in 1959 (17:47) and is known as the perceptron (see figure 2.2). The perceptron has an arbitrary number of inputs and has an output defined by the equation:

$$y = f\left(\sum_{i=0}^{N-1} x_i w_i + \theta\right) \tag{2.1}$$

where

y = output
w_i = i^{th} connection weight
x_i = i^{th} input
θ = threshold
f = nonlinear function
N = number of inputs

The most popular type of non-linear function used in the perceptron has been the sigmoid, although other functions can be used. Examples of other commonly used functions are the hard limiter, and the threshold logic function (13:5). The non-linear transformation on the node inputs is analogous to the non-linear relationship between excitation and neuron firing rates which has been observed in animals (17:49).

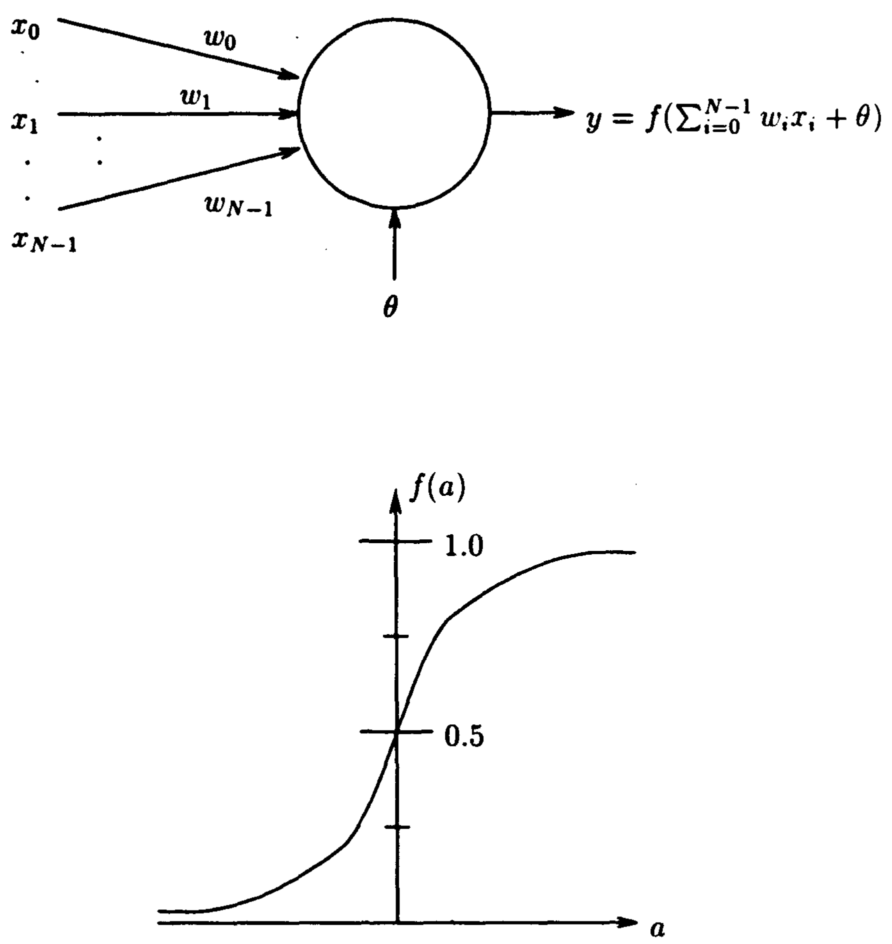

Figure 2.2. The perceptron and the sigmoid function (17:48)

The sigmoid function is defined as:

$$f(\alpha) = \frac{1}{1 + e^{-\alpha}} \tag{2.2}$$

where α is the argument of equation 2.1.

If the argument (α) of equation 2.1 is examined, it is readily seen that this is an equation of a hyper-plane in a multidimensional feature space (or a line in 2 dimensions). Thus the effect of the sigmoidal transformation on the argument (assuming the weights are fixed) is to create a dividing hyper-plane where one side represents the part of the feature space where f(α) has positive values, and another side where f(α) has negative values. The result of this fact is that the single perceptron only has the capability to discriminate between two hyper-plane separable decision regions.

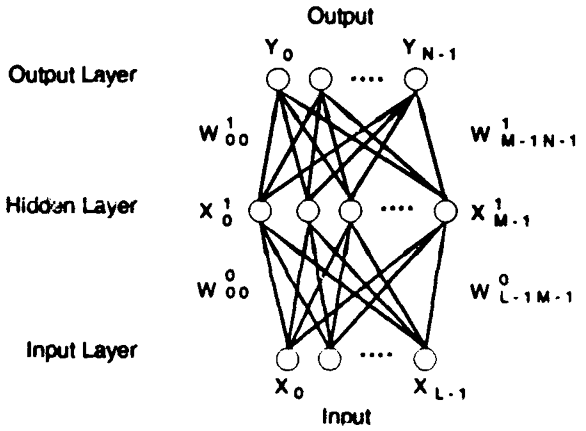

Figure 2.3. Multi-layer perceptron. L = number of input features, M = number of hidden nodes, N = number of output classes.

2.3.2 The Multi-Layer Perceptron. The next logical step is to take the output of multiple perceptrons and to use them as inputs to another row of perceptrons. This architecture can be extended to create multiple layers if desired. The resulting structure is known as the multi-layer perceptron (MLP) (Figure 2.3). Unlike the single perceptron, the multi-layer perceptron has the ability to carve out complex decision regions (17:53–54). Hence pattern recognition problems with classes that are not hyper-plane separable (e.g. the XOR problem) can frequently be solved using this ANN architecture.

The node layers in the MLP which aren't input or output layers are commonly referred to as "hidden layers" because theit output values are not observable when the network is viewed as a "black box". It has been shown in recent years by Cybenko (17:53) that one hidden layer in a MLP is sufficient for any solvable problem, given that enough hidden nodes are used. Despite this proof, additional hidden layers are sometimes used because they may provide an advantage in faster learning. As for the number of hidden nodes required for a given problem, no established "rule of thumb" methods of estimation currently exist (25:57). Consequently, the number of hidden nodes used is usually determined by experimentation.

A popular method of implementing the MLP is to assign one output node to each class in the problem being considered. A decision rule is then used where the ouput node with the highest value is selected as the most likely class for the presented set of inputs (a.k.a. the feature vector).

2.3.3 The Back-Propagation Learning Algorithm The most common learning algorithm used with the MLP is the back-propagation learning method, often called "backprop" for short. The backprop algorithm is a generalization of the LMS algorithm (13:17), often used with adaptive filters. It uses a gradient search method in order to minimize the mean squared error (MSE) between the desired and actual outputs of the network. In terms of this algorithm the MSE is defined as:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(y_n - d_n)^2 \tag{2.3}$$

where

N = number of output nodes or classes
y_n = output value for node n
d_n = desired output value for node n

The desired output values are typically a high value for the node representing the correct class, and a low value for the nodes representing the other classes. Desired output high/low value pairs that are commonly used include 1.0 / 0.0, 0.9 / 0.1, and 0.8 / 0.2.

The steps in the backprop algorithm are (17:56):

1. Initialize weights and thresholds to small random values. Generally values are uniformly distributed between −0.5 and +0.5. The randomness of these

initial values increases the chances that a particular weight will be close to its optimum value, and thus will need minimal training.

2. Present randomly selected input from the training set and its correct classification (desired output).

3. Propagate the input through the ANN to output layer.

4. Compare output node values with desired output values and calculate output errors.

5. Propagate output errors back through the ANN to update the weights via

$$w_{ij}^{+} = w_{ij}^{-} + \eta\delta_j x_i + \alpha(w_{ij}^{-} - w_{ij}^{--}) \qquad (2.4)$$

where:

w_{ij} = the weight from node i to node j in the next layer
x_i = the output of node i.
δ_i = the error associated with node j.
η = the learning rate constant (typically 0.3)
α = the momentum constant (typically 0.7 when used)
w_{ij}^{+} = the new weight value
w_{ij}^{-} = the old weight value
w_{ij}^{--} = the value of the weight before the last update

Thresholds are adapted similarly where x_i is replaced by +1 if the threshold is *added* to the weighted sum and -1 if it is *subtracted.* The δ_j are defined as follows:

$$\delta_j = y_j(1 - y_j)(d_j - y_j) \qquad (2.5)$$

for the output node j, while:

$$\delta_j = x_j(1 - x_j)\Sigma_k \delta_k w_{jk} \qquad (2.6)$$

for hidden node j, where

d_j = the desired value of output node j
y_j = the actual value of output node j.
x_j = the output value of node i
δ_k = the errors for the layers above

2.3.4 Determining the Number of Training Exemplars Required. One of the first questions which comes to mind when attempting to solve a given problem using a neural network is: "How many training exemplars are needed?". Foley showed that the size of the training set should be linked to the size of the number of features used in the network. His work concluded that the number of vectors in the training set should be at least three times the number of features per output class in order to accurately predict the error rate on a random test set. Thus the total number of training vectors required would be:

$$N = 3FC \tag{2.7}$$

where:

N = total number of training vectors required
F = number of features used
C = number of classes in problem

Another guideline was derived by Baum and Haussler which specifically applies to ANNs which use a hard-limiter non-linearity. This rule says that the number of training vectors needed is on the order of W/ϵ, where W is the number of weights in the network, and ϵ is the amount of error that is acceptable to the designer. Using this equation, Bernard Widrow derived an equation sometimes known as "Uncle Bernie's Rule":

$$10W \approx N \tag{2.8}$$

The assumption used in this derivation is that the acceptable error is 10%.

2.4 Related Research Using ANNs

Much research has been done in subjects which are closely related to auditory localization such as underwater electronic sonar systems, and acoustic emissions analysis. This section will review some of the research dealing with the use of ANNs to localize and to classify acoustic signals.

2.4.1 Sonar Signal Classification. Several researchers have examined classification of various types of sonar signals. Their research focused on classification of sonar returns of signals generated by both electronic and biological systems.

2.4.1.1 Underwater Electronic Sonar Classification. The experiments by Gorman were an attempt to classify between two different objects using a two layer backprop network with various numbers of nodes in the hidden layer. The two objects used were a rock and a metal cylinder. The feature vectors were made by calculating the normalized spectral envelope of sonar return signals (10).

The experiments showed that the networks were able to converge using the training set of feature vectors. The optimum number of hidden nodes was found to be 12. Incrementally increasing the number of hidden nodes up to this point improved the classification performance each time, but no improvement was observed for networks with more than 12 hidden nodes.

2.4.1.2 Comparison Between ANN and Dolphin Sonar Classification. Roitblatt performed very similar experiments to those of Gorman, with a few differences. Roitblatt compared the echolocation classification performance of a counter-propagation network to that of a dolphin (18).

The blindfolded dolphin was trained to emit echolocation clicks and to indicate to the researchers which of four different objects was present in the test pool. The objects were: a steel ball, an aluminum cone, a large PVC tube, and a small PVC tube. The ANN was trained and tested using return signals from simulated dolphin clicks. The feature vectors presented to the network were the FFTs of the return signals encoded into 20 frequency bins.

The results of the tests showed that the dolphin correctly classified the object approximately 94.5% of the time. Comparatively, the ANN able to correctly classify 100% of the test vectors presented after training was performed (18).

2.4.1.3 ANN Model of Sonar-Based Target Recognition in the Echolocating Brown Bat. This study, by Brennan, involved the implementation of a neural network to model the ability of a bat to discriminate between a mealworm and an inedible object. Two different inedible objects were considered, namely disks and spheres. This capability had previously been demonstrated in experiments with the big brown bat *Eptesicus Fuscus*, by Griffin (1958) (6:1). These bats use ultrasonic, frequency-modulated and constant frequency sonar signals to detect, locate, identify and capture airborne prey (6:2).

The sonar returns were collected from the mealworms, spheres and disks at various rotations (90 to - 90 degrees) using electronically synthesized impulse signals similar to those used by a bat. The signals were used to train the network using backprop learning. The same set of signals were used for test, and 100% accuracy was achieved in discriminating between the edible and inedible objects (6:7).

2.4.2 Acoustic Emissions Analysis. Acoustic emissions are defined as "the transient elastic waves accompanying the sudden, localized change in the stress or strain in a material" (11:1226). These signals appear as a result of cracks, deformations, or corrosion in the material. The purpose of analyzing acoustic emissions is to determine the location and/or the nature of the defect which exists in the material. This section discusses research involving the use of ANNs to process acoustic emissions.

2.4.2.1 Localization of Acoustic Emissions. Experiments by Grabec involved the use of an ANN-like structure to localize acoustic events occurring on a metal plate (11:1226). The ANN structure used is known as an autoassociative recall system. The autoassociative recall system resembles a perceptron with the exception that it does not use a nonlinear function in the processing node. Piezoelectric sensors were mounted on the plate in two different configurations. The first was with 2 sensors, one at each end of a 20 inch line parallel to one edge of the plate (11:1230). The second was with 4 sensors, each located at a corner of a 12 inch square array (11:1232). Feature vectors for the experiments were made up of samples of the signals received by the sensors, concatenated together.

Using the first configuration, a steel ball of diameter 8 mm was dropped from a height of 5 mm at 22 different discrete points on the line. For the second configuration the stimulus was the fracture of a pencil lead of diameter 0.3 mm and length 2.5 mm at discrete locations in the array. For both configurations plots of the input vectors and representative output vectors were presented to show the apparent similarity between them. This particular neural implementation method was judged by the authors to imperfect because of the susceptibility to noise and the similarity of the vectors in the learning set (11:1233).

2.4.2.2 Source Location of Atmospheric Leaks. Barga tried a multilayered perceptron using back-propagation learning to localize atmospheric leaks from the acoustic emmisions produced. This research was performed in order to investigate ways of quickly finding such leaks in the future NASA Space Station Freedom. According to the authors, no reliable method of performing this task currently exists (3:602).

The experiments used a test plate similar to the material to be used on the skin of the space craft (see figure 2.4). Piezoelectric sensors were placed on two diagonal corners of a 6 by 7 grid where each vertical and horizontal element was spaced 5

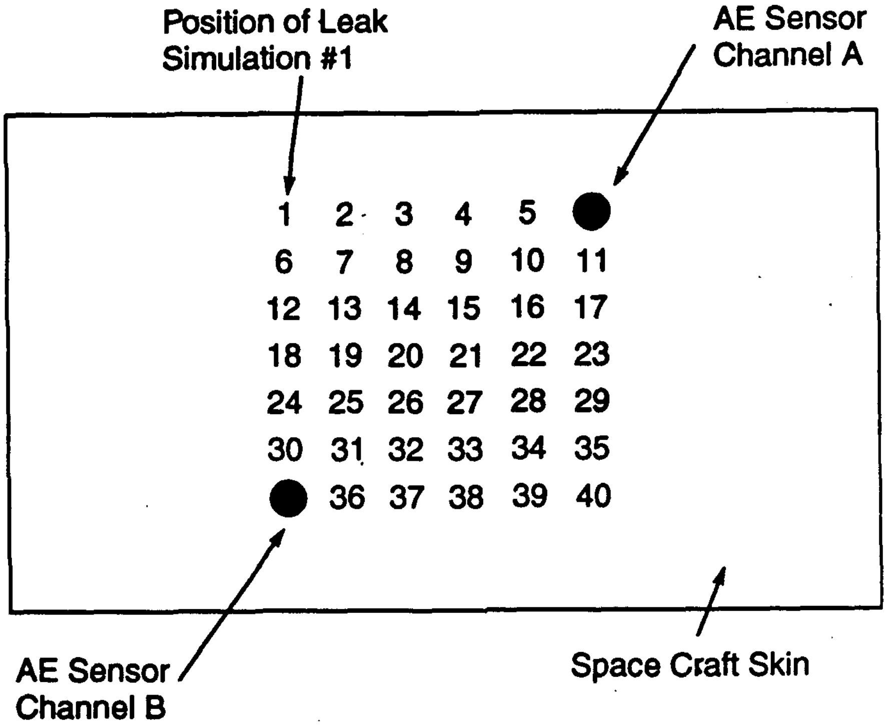

Figure 2.4. Setup of experiment by Barga to localize atmospheric leaks. The numbers show the locations represented by each class (3).

inches apart. The remaining 40 points were used as discrete leak locations in the experiments. Training of the network was performed with samples of the signals received by the two sensors.

The MLP used in these experiments received two sets of 512 inputs: one for the signal received by each sensor (figure 2.5). Each set of inputs were interconnected with a separate set of 85 hidden nodes on the next layer of the network. The first layer of hidden nodes were connected to a second layer of 38 hidden nodes. The output layer consisted of 40 nodes, each one representing one of the 40 discrete locations on the grid.

The input and first hidden layers of this ANN acted as a feature extraction network on the sensor signals. This part of the network was created by implementing an identity network (figure 2.6). An identity network has an equal number of input

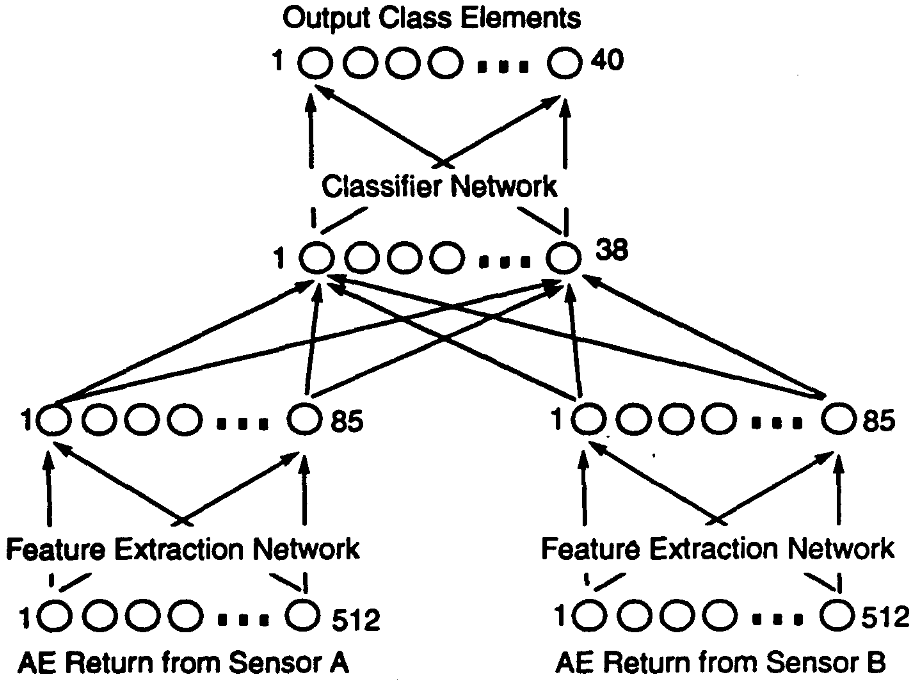

Figure 2.5. Network configuration for atmospheric leak localization. This type of network is often referred to as a "Cottrell" network, named after its inventor (9) (3).

and ouput nodes, and a reduced number of hidden nodes. Exemplars are input and the identity network is trained to reproduce the exemplars at the output. At the conclusion of this training, the values of the hidden nodes can be viewed as a set of coefficients representing the input values in a very compressed form. This method has been shown to produce compression ratios of 8 to 1 on images (9:68). The weights were then fixed, and the output layer of nodes stripped away. Two identical copies of this structure were connected to a classifier represented by the second hidden layer and the output layer. The classifier was next trained using the compressed features produced by the feature extraction portion of the network.

A set of four returns were obtained from each of the forty simulated leak locations, resulting in 160 total training vectors. Two tests were run, the first using the training vector set as the test set (location dependent), and the second using

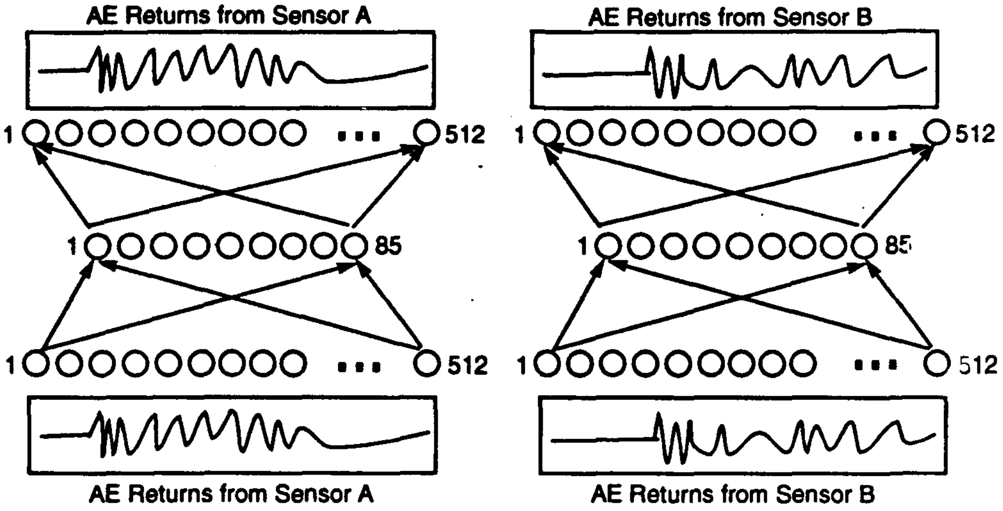

Figure 2.6. Identity network. This type of network learns to reproduce the input sequence at the output nodes (3).

simulated leaks at unique locations in between the 40 locations used in training (location independent).

The results of the tests showed that the network was able to generally classify the correct area of the simulated leak location, but not with great precision. For the location dependent test the mean localization error was 4.54 inches, while the mean error was 5.98 inches for the location independent test.

2.4.3 Binaural Sound Localization. Anderson has done some preliminary work showing that neural networks can perform binaural localization, in azimuth, of single tones of known frequency. Raw samples (32 samples, 20 kHz sampling rate) of the simulated tone signals received by each ear were input to feature extraction networks which identified the peak value in each ear signal (figure 2.7). The outputs of these feature extraction networks were the peak value at the sample position where the peak occurred. Two parallel classifier networks used the outputs of the feature extraction networks. One network classified either front or back, while the other classified either right or left. Using this technique, the quadrant from which the sound

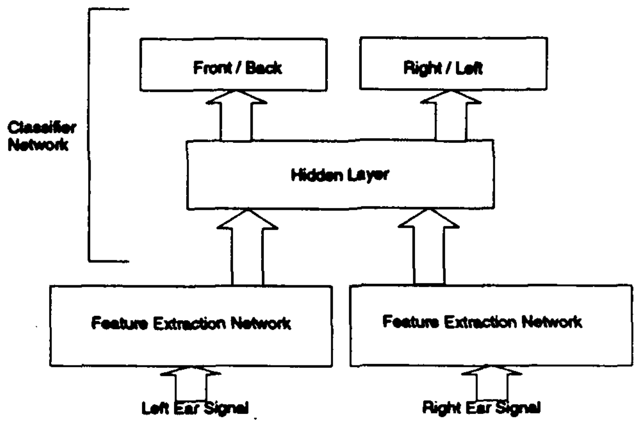

Figure 2.7. ANN configuration used by Anderson for tone localization.

source originated could be classified. For example, if the network outputs indicated front and right, then the sound source would be classified as having originated in the region from 0° to 90° (1).

This network configuration performed well on tones where the period was matched to the 32 sample window of the feature extraction network. This was because there was only one peak value in the window, and the peak value of the window was known to be the peak value of the sinusoid. However, this method would not be reliable on tones of other frequencies. If the period of the tone were shorter than the window length, then more than one peak value may appear in the window. In this case the time delay could not be accurately determined. If the period of the tone were longer than the width of the window then the peak value of the sinusoid may not be present in window. In either case, the classifier network would not be likely to receive the right features to use for localization. Thus, a more robust method of feature extraction must be used if localization is to be done on tones of variable or random frequency.

2.5 Conclusion

It is clear that there may be several different features which may be used by humans to perform auditory localization. Scientists have presented evidence that interaural time-delays and amplitude differences, pinna cues, and head movements may be among the important features used. It is believed that the mechanisms used in human sound localization are frequency dependent. If the frequency is below 1000 Hz, ITD is the predominant feature. Between 1000 Hz and 4000 Hz IAD becomes an additional factor. Above 4000 Hz pinna cues also become important (20:448).

Despite this knowledge, there are difficulties in obtaining some of these features for use in a machine that performs localization. Pinna cues are a difficult thing to quantify in a small set of numbers. Also, head movements in a machine require mechanical complexity.

Because of the impressive localization capabilities of biological auditory systems, and the successes by researchers in related areas, neural networks appear to be a very promising method for performing auditory localization. Despite these signs of promise, no one has yet demonstrated a neural network system which can robustly localize narrow and wideband auditory sources.

III. Methodology

3.1 Introduction

This chapter will present the specifics of how the thesis experiments were performed. First, the resources used in this effort will be discussed. Next, the special terms and notation to be used in the remainder of this thesis paper will be defined. And finally, the design of the experiments and the procedures used will be presented.

3.2 Resources

This section will detail the resources which were used in order to carry out the aforementioned experiments. The resources used include the binaural localized sound source model, the ANN simulator software, and software used to implement the model and generate data for input to the network.

3.2.1 Binaural Localized Sound Source Model. A model which may be used to produce the left and right ear signals received by a human from a given source signal was developed by the Bioacoustics and Biocommunications Branch of the Armstrong Laboratory. The model was designed from the experimentally measured transfer functions of each ear at one-degree increments in a circle with a seven foot radius, in the horizontal plane. The transfer functions were averaged over the range 100 Hz to 20 kHz to obtain an average gain constant for each ear. Additionally, the measured relative time delay between signals received at each ear was obtained at each one-degree increment. This data was collected and tabulated into a database. The binaural signals are generated from this database by multiplying the source signal by the gain factors and by shifting one of the two signals in time by the appropriate number of samples.

This model has been used to generate artificial binaural sound signals in headphones that create the illusion of a sound originating from a selected angle in the horizontal plane (1).

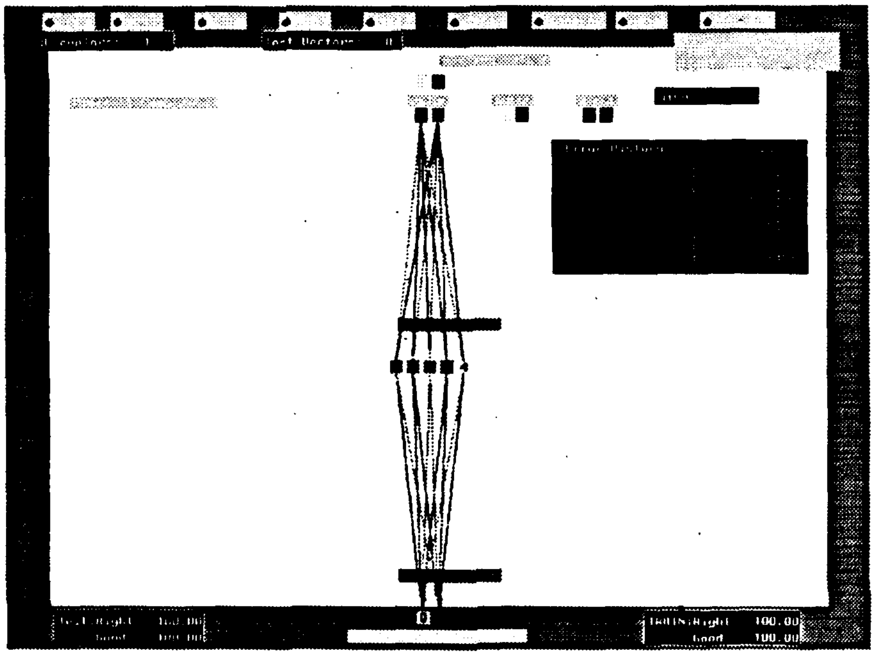

Figure 3.1. NeuralGraphics Display

3.2.2 ANN Simulator Software. The ANN simulation software, called NeuralGraphics, was developed at AFIT by G. Tarr. This simulation package is written in the C language and may be run on any Unix-based machine, but in order to take advantage of the graphical displays and interface, a Silicon Graphics IRIS 3130 or 4D machine should be used if available.

NeuralGraphics is an interactive program that allows the user to examine node and weight values, turn training on and off, and remove or replace selected nodes. The graphical interface provides information about the values of nodes and weights via colo: coding, and displays information about the progress of the learning process by an error plot and by the percentages of correct classifications on the training and test sets (see figure 3.1). The program supports the use of several different learning paradigms including back propagation, radial basis functions, Kohonen networks and hybrid networks.

A data file for the program consists of a header line which contains the number of training vectors, the number of test vectors, the number of elements in an input vector and the number of classes. The remainder of the file consists of the training and test vectors. The vectors are identified by an index number (beginning at one) at the beginning of each vector, and the class to which the vector belongs at the trailing edge of the vector.

When running the program, the user is initially greeted by a menu where he or she selects a learning paradigm and the type of normalization desired (statistical, spread, energy or none). Additionally the user must specify a data filename, an initial weight file or random initial weights, the number of layers and the number of nodes in each layer, the number of iteraticns to be run before the program stops, and the number of iterations between updates of the graphics displays.

The training and test history of the network is recorded in output files named "data-stats" and "item-report" respectively. In order to gain added insight, the program was altered for this thesis to output the values of all of the output nodes of the network every thousand iterations to a file named "test".

3.2.3 Construction of Data Files. In order to construct the data files to be presented to the ANNs in the experiments, a means was needed to extract information from the model to produce left and right ear signals from a given source signal and angle of origin. These tasks were accomplished using a program written in Turbo C and run on an IBM-compatible 386 personal computer. This program has several different versions, depending on the number of classes, the number of feature vectors to be generated, and the type of sound source signal which was being used. Nevertheless, the general structure of each of these versions of the program remained the same.

The program constructed the binaural signals using the model database and a file which contained random angles, uniformly distributed between 0 and 359 degrees. For some experiments, time-samples of the binaural signals were used as features, while for other experiments the features were computationally extracted, from the binaural signals. After generation of the features was completed, the feature vectors and the class to which each vector belonged was output to an ASCII file which was formatted properly for use with the ANN simulation software.

3.3 Definitions and Notation

This section lists the definitions and notation to be used in the remainder of this document.

MSE: The mean squared error. Previously, the MSE was defined in terms of a set of desired outputs, and a set of the actual outputs resulting from the presentation of a given feature vector to an ANN. When using the MSE over time as a measure of the learning occuring in an ANN, the MSE must be defined slightly differently. In this new context the MSE is defined as:

$$MSE = \frac{1}{N}\frac{1}{M}\sum_{n=1}^{N}\sum_{m=1}^{M}(y_{nm} - d_{nm})^2 \qquad (3.1)$$

where

N = number of output nodes or classes
M = number of training vectors
y_{nm} = the network output for node n, test vector m
d_{nm} = the desired output for node n, test vector m

front/back classification error: This error is defined as when the ANN misclassifies the vector presented to it in such a manner as that the called class and the actual class are symmetrical spatially to the front and back on a single hemisphere of the horizontal circle. For example, in an 18 class problem (figure 3.10), if the called class were class 2 and the actual class were class 7 (or vice-versa), then a front/back classification error has occured.

"Good" classification: This classification criteria is based on the idea that the output node of network which yields the highest output should be the class chosen for the input feature vector. Thus, if the class 1 output node were 0.4 and the class 2 output node were 0.5 for a given input, class 2 would be chosen as the predicted correct class. This criteria is used as the correct classification metric in this thesis.

"Right" classification: This classification criteria is more stringent than the "Good" criteria. The NeuralGraphics program defines "right" as being the case when all of the actual values of the output nodes of an ANN are within 5% of the desired values. The "right" metric is useful as a means of determining how well a network has learned the training set. The 5% value is an arbitrary measure of being very close to the desired output values.

P(Good): One performance measure used in this thesis is the probability of "Good" classification on the test vectors. As an example, if 83 out of 100 test vectors were assigned a "Good" classification, then P(Good) = 0.83.

P(W1C): The probability of the called class being "within 1 class" of the correct class, that is either the correct class or an adjacent class (see figure 3.10). An error of only 1 class provided some useful information. If the called class were one of the adjacent classes, then at least a generalized direction to the sound source would have been recognized.

P(HHN): The probability of the correct class being either the highest output node value or the highest valued neighbor output node where "neighbor" is defined as an output node which is on either side of the high node (see figure 3.2). This statistic provides some insight into the level of confidence that can be attained that a given source signal can be localized to within the angular width of two classes on the horizontal circle. For example, if the highest output node value was node number

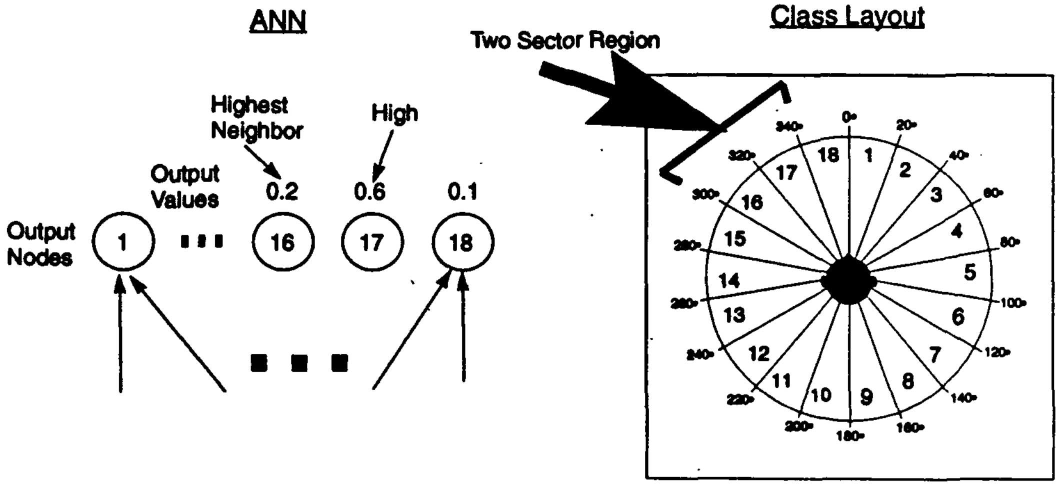

Figure 3.2. Definition of the "High and Highest Neighbor" rule.

4 and the highest valued neighbor were node number 5, then the correct class must be either class 4 or 5 in order to meet the criteria. If 75 out of 100 test vectors met the criteria, then P(HHN) = 0.75.

P(FBE | E): The probability that a classification error was a front/back error. This statistic provides a metric of the amount of front/back miscalculation which occurred.

P(CBCC | W1C): The probability that the sound source is close to the boundary of the called class given that a 1 class error occurred. Because the class boundaries were artificially laid out on a continuous circle, it was expected that there would be a significant number of classification errors in the boundary regions. This statistic provides insight into this phenomena. Defining "close" to the boundary is a subjective call, but for the experiments of this thesis this was defined as being +/- 4 degrees.

η: This parameter is the learning rate used in all weight and bias update equations of the backprop algorithm (equation 2.4). The NeuralGraphics program uses a default value for η of 0.7.

Iterations: This parameter is used to control the number of times that the training data set will be presented to the network.

Statistical Normalization: Normalization is often performed on the data which is input to neural networks. One purpose of normalization is to keep the values which are being multiplied in the network small and manageable. Sometimes normalization can help to improve the classification accuracy of the network. Statistical normalization is designed to both scale normalize and to discount the influence of feature values which are far from the mean. This type of normalization is defined as follows

$$n_i^j = \frac{u_i^j - \mu_i}{\sigma} \tag{3.2}$$

where:

i	=	feature index
j	=	feature vector index
n_i^j	=	the statistically normalized feature value
u_i^j	=	the un-normalized feature value
μ_i	=	the mean of the ith feature over all j
σ	=	the standard deviation of the ith feature over all j

$R_{xx}(0)$: The autocorrelation at lag value $\tau = 0$. In general the autocorrelation is defined as:

$$R_{xx}(\tau) = \frac{1}{N}\sum_{n=1}^{N}(x_n x_{n-\tau}). \tag{3.3}$$

where:

N	=	number of data points in the sequence
n	=	index of data points
τ	=	lag value index

$R_{xx}(\tau)$ shows the similarity of the sequence x_n with itself as a function of time shift. For the case of $\tau = 0$, the above equation reduces to:

$$R_{xx}(0) = \frac{1}{N}\sum_{n=1}^{N}(x_n^2) \tag{3.4}$$

$R_{xy}(\tau)$: The cross-correlation function. The definition of this function is:

$$R_{xy}(\tau) = \frac{1}{N-\tau} \sum_{n=1}^{N-\tau} (x_n y_{n-\tau}) \tag{3.5}$$

where the parameters are defined as with the autocorrelation. The crosscorrelation is a measure of the similarity of the sequence x_n with another sequence y_n as a function of the time shift τ.

Both the auto-correlation and cross-correlation functions here defined are time-average representations and are valid under the assumption of ergodicity (22:179).

3.4 Presentation of Network Results

This section will describe the methods which were used for reporting the performance of the networks used in the thesis experiments. First, the presentation of training performance results is discussed. Next, the presentation of test results is covered.

3.4.1 Training Performance. The training performance was reported through the use of curves which demonstrated the average values of the P(Good) and the mean squared error (MSE) of the output nodes (with respect to the training set) over the multiple training runs for the data sets used. These values were recorded at 1,000 iteration intervals in data files created by the ANN simulation software. The training performance curves were used to demonstrate the convergence of the P(Good) and the MSE statistics to optimum values for the given configuration of the network. A sample run on the data was made to estimate where learning ceased. This set the number of iterations at which learning was terminated.

3.4.2 Test Performance. The average classification performance of each network with respect to the set of test vectors was summarized in a table of statistics.

The statistics used in calculating the averages were calculated at the termination of learning for each network. The statistics reported in each experiment differed somewhat on different experiments due to the differences in focus and the results obtained. These statistics included P(Good), P(W1C), P(HHN), P(FBE | E), and P(CBCC | W1C). Additionally, plots were generated, showing the angular distribution of the errors in the horizontal plane.

3.5 Experiment Design

This section will present the experiments which were carried out as part of this thesis effort. This section will describe the purpose of the experiments, the features used in each experiment, the architecture of the each network used, and the size, configuration, and manipulations made on the training and test sets.

The similarities in the experiments are listed below.

1. First, the main goal of all of the experiments was to localize the angular position of a sound source originating somewhere on a circle (radius 7 feet) on the horizontal plane, assuming that the head is centered in the circle. The circle was divided like slices in a pie to form the regions which represented the different classes. Experiments were conducted with 4 and 18 classes (figures 3.3 and 3.10).
2. The multi-layer perceptron architecture with one hidden layer of nodes was used in all of the following experiments.
3. Learning was done in all cases via the back propagation algorithm. The parameters of the algorithm were set as follows: the learning rate $\eta = 0.3$, the momentum term $\alpha = 0$, and the desired output values $d_n = 0.9$ for the correct class and $d_n = 0.1$ for the incorrect classes. The reason for the selection of the parameters was simply that experimentation showed that these values seemed to consistently work the best in terms of training. Although having $\alpha > 0$

usually makes learning converge faster, it was found that the convergence and accuracy of the network were adversely affected by some combinations of η and α. Also it was observed that using $d_n = 0.1/0.9$ resulted in better classification results than when $d_n = 0/1$.

4. The angle and frequency (tones only) values used to construct the sound source signals were placed in ASCII data files to allow for manipulation of the ordering and pairing in different training runs.

5. Fifth, vectors for each experiment were presented to the networks in a random order. This function was performed by the NeuralGraphics program.

6. Some trial runs were made on each configuration with various numbers of hidden nodes. The results of these trials determined the number of hidden nodes used in the tests.

7. The number of training vectors in each experiment generally was chosen in order to satisfy Foley's rule (equation 2.7). Exceptions were made in the preliminary experiments because in some cases it was not possible, and in one case because only a concept was being tested where the "real world" accuracy was not required.

8. Statistical normalization (as defined in equation 3.2) was performed on all of the data sets prior to presentation to the network. It was observed in sample runs of the preliminary experiments that approximately 5% improvement was achieved using statistical rather than just scaling normalization.

9. The point at which termination of training occurred was judged by making a training run with each configuration and (after many training iterations)making a judgment where the P(Good) and the MSE appeared to have converged.

10. The results of the latest classification tests using the test vector set and the output values of each output node of the network, and the MSE and P(Good)

training data were obtained and reviewed from each training run and used to tabulate statistics for the experiments.

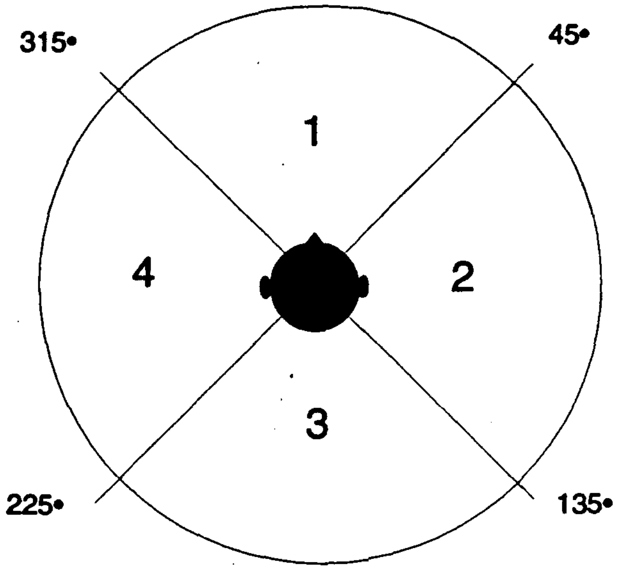

Figure 3.3. Top view of the horizontal circle: 4 classes

3.5.1 Preliminary Experiments. This series of experiments was performed with the goal of learning something about the ability of an ANN to localize tones to one of 4 classes: front, back, left, or right (figure 3.3) using normalized raw samples of the signals received by each ear. The following method of encoding the time-samples of both ear signals into a feature vector was devised.

The feature vectors were constructed as follows: fifty samples from each ear signal were concatenated, forming a one-hundred element input vector to the network (see figure 3.6). These signals were assumed to be sampled at a rate of 20 kHz, producing a lower frequency limit (due to the fifty sample window) of 400 Hz and an upper frequency limit (due to the sampling rate) of 10 kHz. These upper and

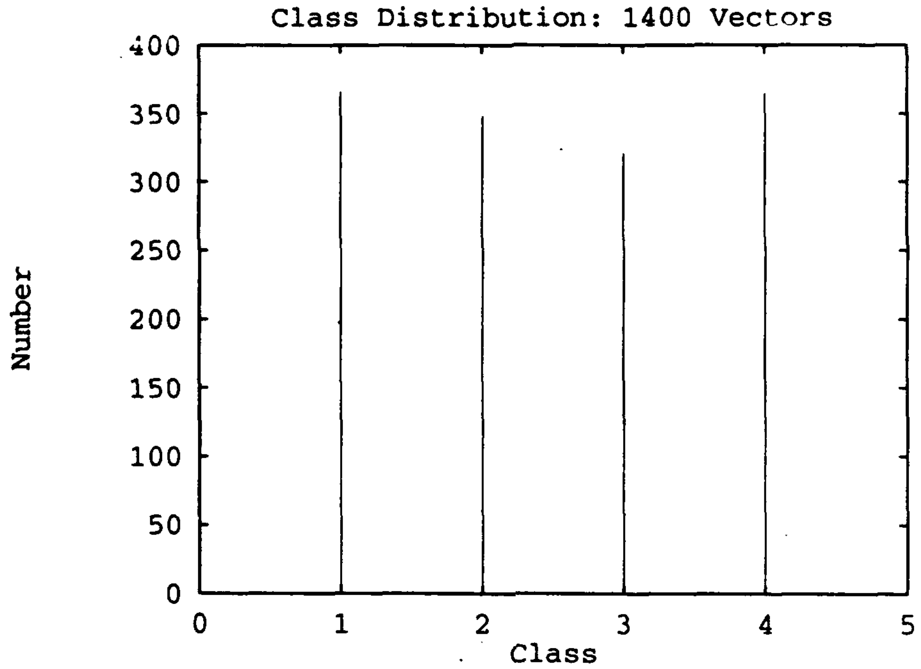

Figure 3.4. Histogram of the random angles for 4 classes.

lower frequency limitations provided a reasonable frequency range for use with audio signals.

Experimentation with the data sets showed that the networks in this series of experiments performed optimally with 10 hidden layer nodes.

3.5.1.1 Preliminary Experiment 1: Localization of 1kHz Tones. This was the simplest of the experiments. The purpose was to verify that the direction of a fixed frequency tone could be localized by a neural network using time samples as features. Only 360 unique vectors were possible in this experiment (one for each of the 360 degrees), so the training and test sets were identical with all of the vectors represented in each set. Four training runs were made on the data so that averages could be calculated.

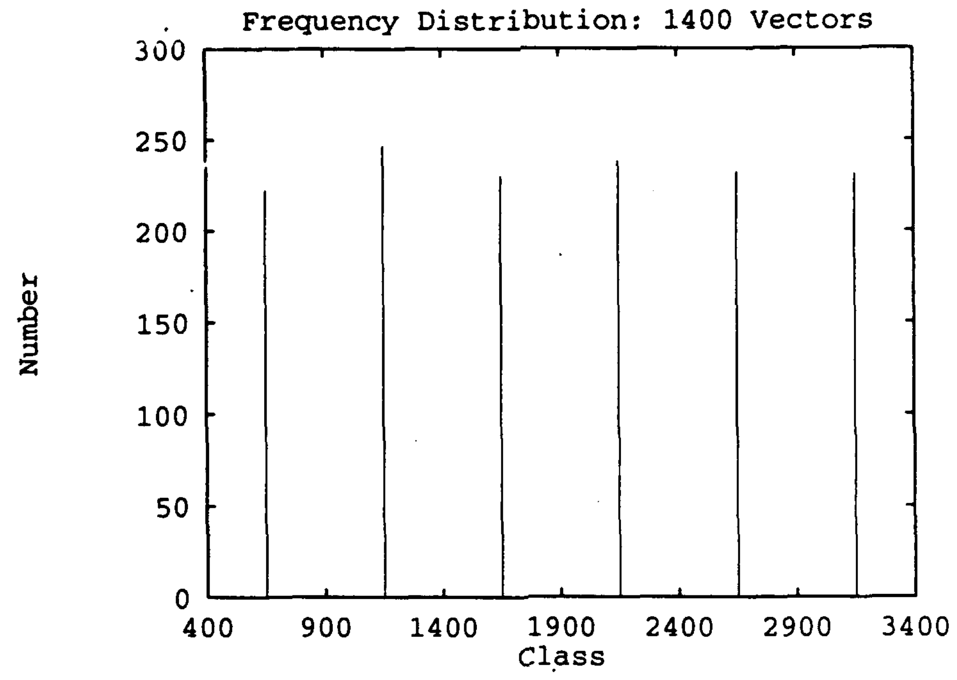

Figure 3.5. Histogram of the random frequenciess for 4 classes.

3.5.1.2 Preliminary Experiment 2: Localization of Tones at 5 Discrete Frequencies. The purpose of this experiment was to verify that the neural network could localize tones of five different discrete frequencies. The frequency values used for the tones were 500 Hz, 1 kHz, 1.5 kHz, 2 kHz and 3 kHz. In this experiment, all 360 angles were represented for all five frequencies thus creating a total of 1440 training vectors. The test set consisted of 90 vectors at each frequency, divided as equally as possible between the four classes. As in the prior experiment, four training runs were made.

3.5.1.3 Preliminary Experiment 3: Localization of Tones at Frequencies Other Than Those Used to Train the Network. This experiment was performed for the purpose of examining the ability of the network to localize tones across the spectral range of 400 Hz to 4 kHz when trained with tones at 500 Hz, 1 kHz, 1.5 kHz, 2 kHz and 3 kHz. The same training set was used in this experiment as in the

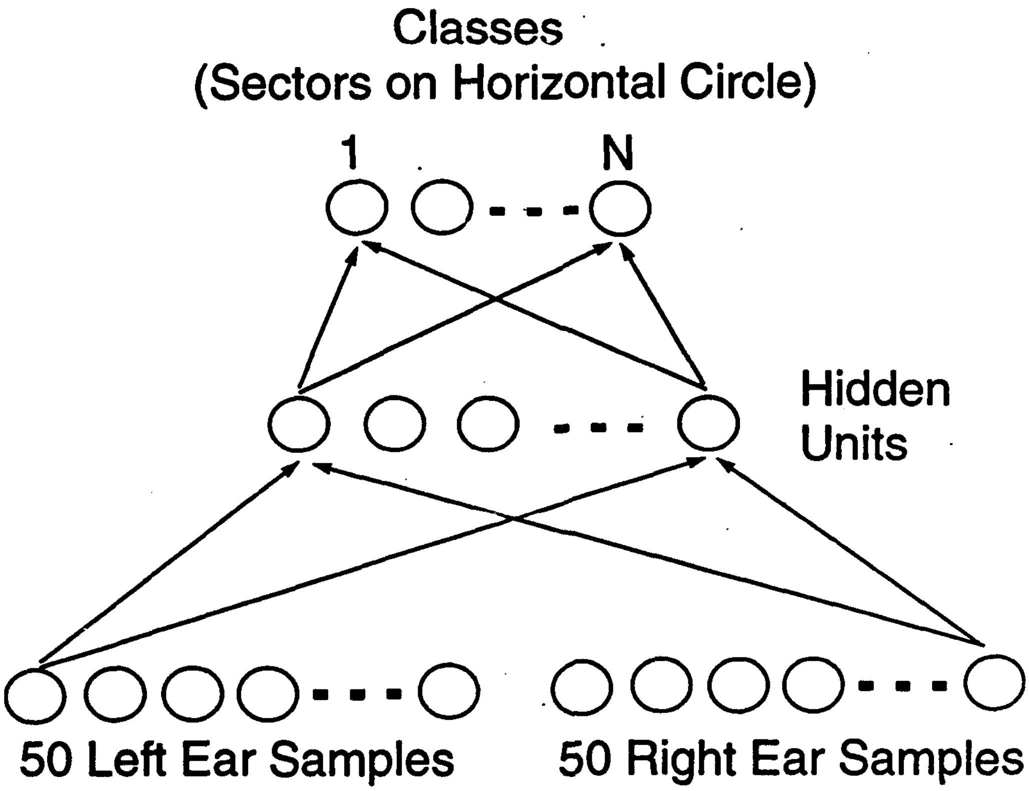

Figure 3.6. ANN Configuration for 4-Class experiments.

prior experiment where localization of 5 discrete tones was investigated. Because the goals of this experiment were not specifically pointed toward overall classification accuracy, but rather toward the classification accuracy as a function of frequency, it was decided that one training session would be sufficient.

The test set consisted of 312 vectors: one in each of the four classes at frequencies distributed unevenly in the frequency range mentioned. The uneven distribution of the frequencies was set up to provide resolution of 20 Hz in the frequency regions close to the 5 training frequencies, and 50 or 100 Hz in the regions farther away from the training frequencies.

3.5.1.4 Preliminary Experiment 4: Localization of Tones Using Random Frequencies. This experiment was designed to explore the performance of a tone-

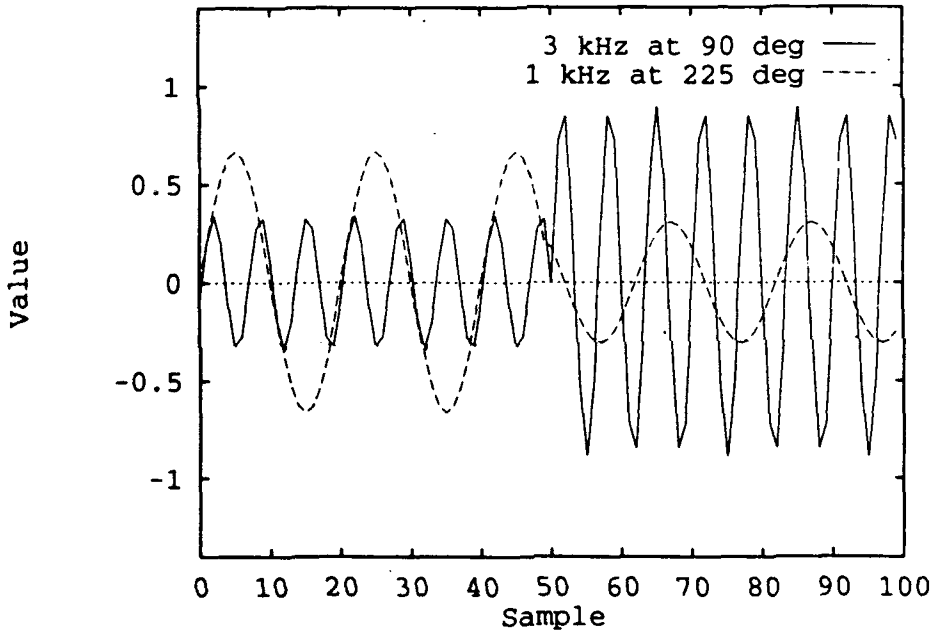

Figure 3.7. Feature vectors created by concatenation of two received ear signals. Tonal sound sources located at 90 and 225 degrees

localization network using random frequencies for training, for comparison with fixed frequency training. Following Foley's Rule (equation 2.7), 1200 training vectors (4 classes, 100 features), and 200 test vectors were used in each run. For this experiment, four training runs were made.

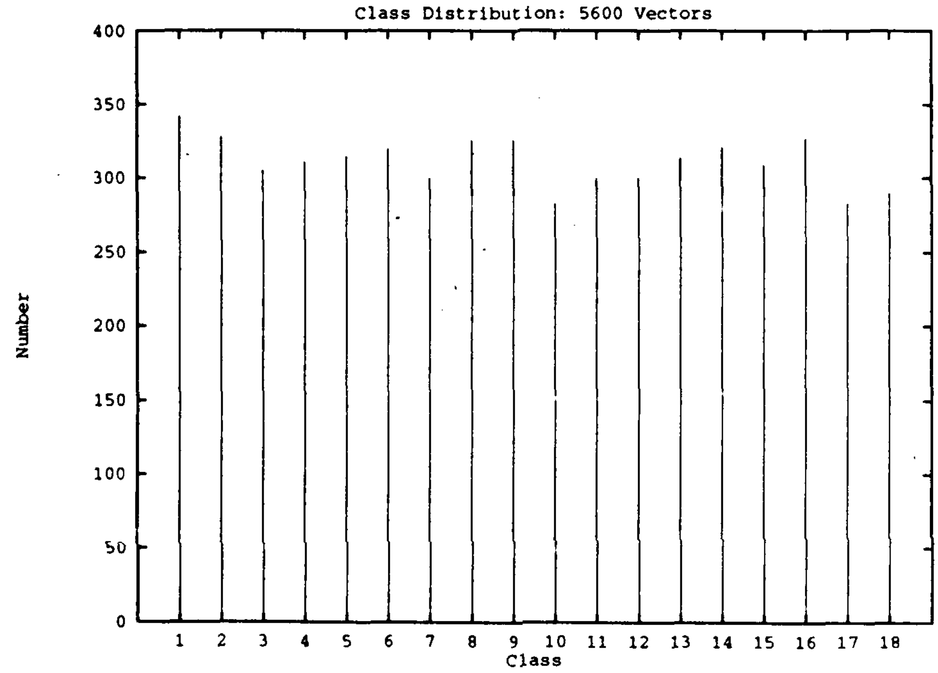

Figure 3.8. Histogram of the random angles for 18 classes.

3.5.2 18-Class Localization Experiments Using Time-Samples as Features. One experiment used tones, and another used gaussian noise as the sound source signal. A file containing integer-valued angles from which the sounds originated for each feature vector was constructed using randomly selected numbers in the range 0 to 359 degrees. The same anglefile was used for both the tone and gaussian noise signal experiments. The histogram of the angle set (figure 3.8) shows that the distribution of the angles was approximately uniform.

These experiments both used training sets containing 5400 vectors, and test sets containing 200 vectors. Eight training runs were conducted in each experiment.

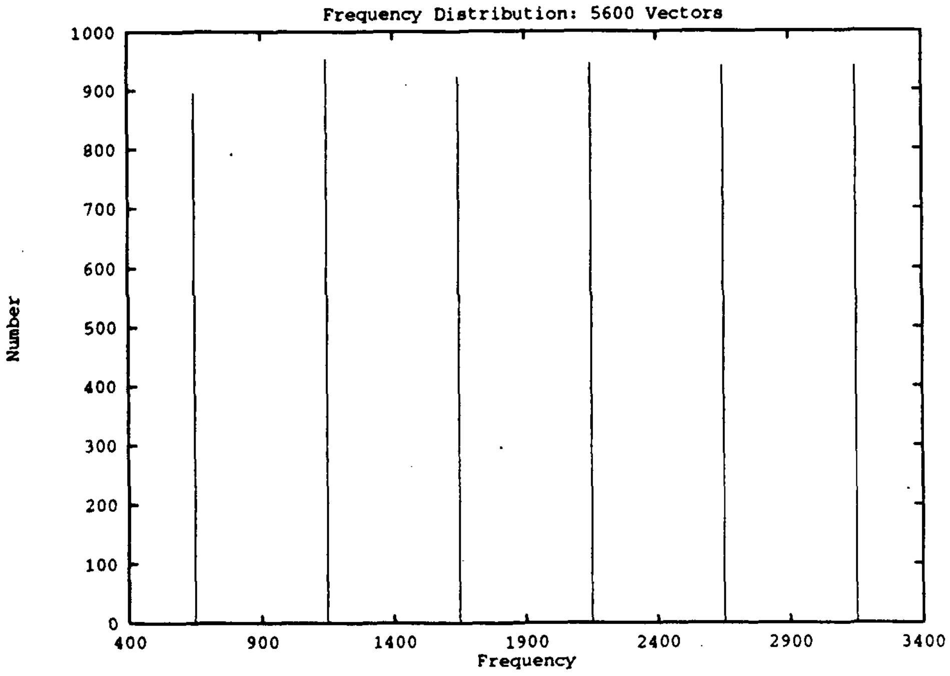

Figure 3.9. Histogram of the random frequencies for 18 classes.

In each training run, a unique test set was used. This was accomplished by adding the 200 test vectors after each training run to the top of the training set and removing the last 200 vectors from the training set to use as test vectors for the next training run.

Experimentation with the network using the data set showed that the optimum performance was achieved with 20 hidden layer nodes.

3.5.2.1 18-Class Experiment 1.1. For the tone localization experiment, the tonal source signals were assigned real valued random frequencies in the range 400 Hz to 3400 Hz. A histogram of the frequency file shows (figure 3.9) that the distribution of the frequencies used was practically uniform. Examples of vectors used in this experiment are plotted in figure 3.7.

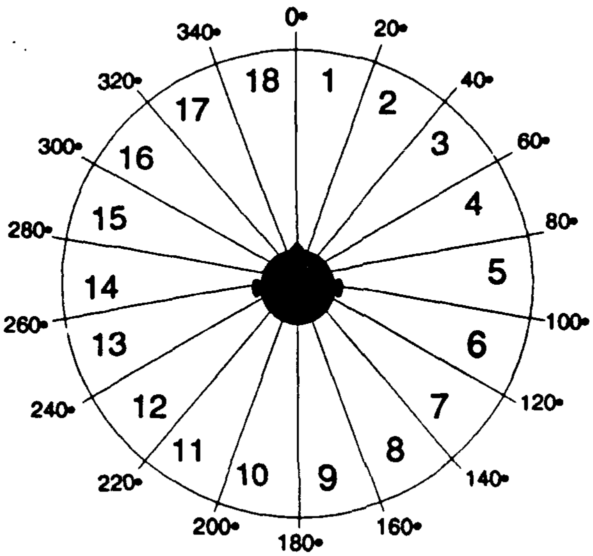

Figure 3.10. Top view of the horizontal circle: 18 classes

3.5.2.2 18-Class Experiment 1.2. The sound source signal in this experiment is gaussian noise with zero mean and unity variance. An example of a feature vector corresponding to a source signal originating at an angle of 90^o is shown in figure 3.11.

3.5.3 18-Class Localization Experiments Using the Mean FFT Magnitudes and the Cross-Correlation as Features. In many cases, pattern recognition problems are simplified by preprocessing the available data and obtaining a reduced number of features for input into the recognition system. Based on the concepts used to design the binaural hearing model used in this thesis, the three most important features appeared to be the intensity of the sound signal received at the left and right ears, and the relative time delay between these two signals. A means of quantifying these three features was needed. For the sound intensity measure, it was decided to use the mean of the FFT magnitude. The mean of the FFT magnitude was defined as

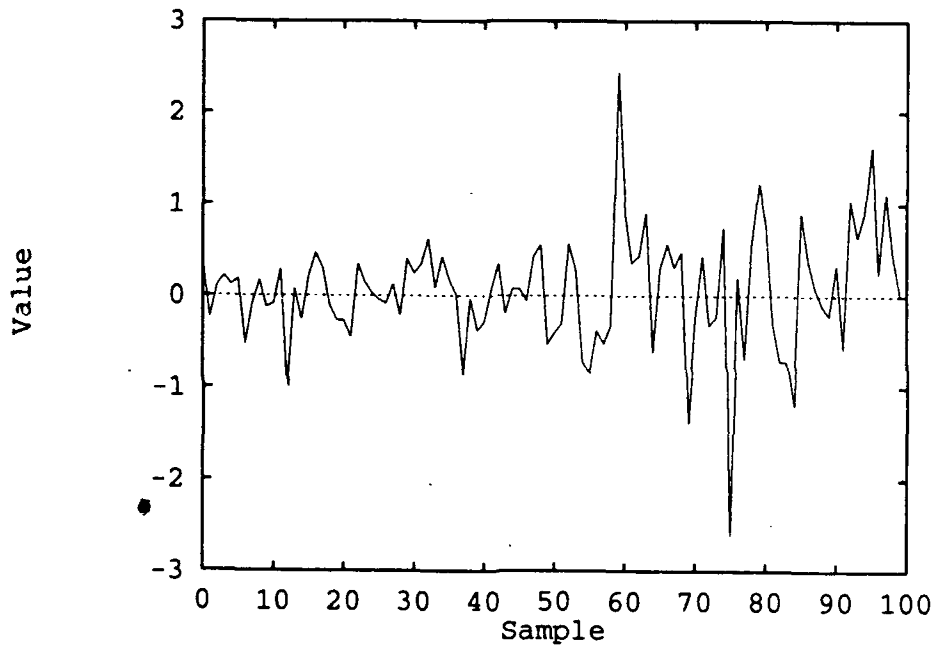

Figure 3.11. Feature vector created by concatenation of two received ear signals. Noise sound source located at 225 degrees. The second half of the plot is a scaled and shifted version of the first half.

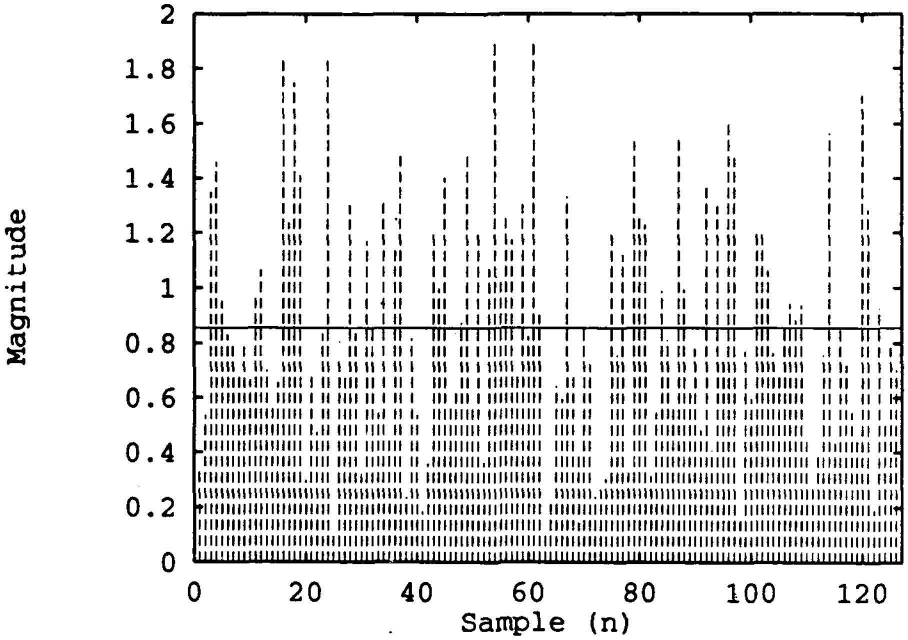

Figure 3.12 Definition of mean of the FFT magnitude. The impulses represent the FFT magnitude sequence of gaussian noise time-samples and the line shows the mean of these values.

illustrated in figure 3.12.

And for the time delay estimation, the position of the peak value in the cross-correlation function was calculated. After finding the lag value at which the peak occured, these values were normalized between -1 and +1. The negative range signifying that the left ear signal arrives first, and the positive range signifying the prior arrival of the right ear signal.

In these experiments, 360 training vectors and 360 test vectors were used. Both the training and test sets contained one vector obtained from a sound source originating at each of the 360 one degree increments around the horizontal circle. Eight training runs were made in each experiment. Again, the networks were put through a few trial runs where 20 hidden-layer nodes was found to provide optimal performance.

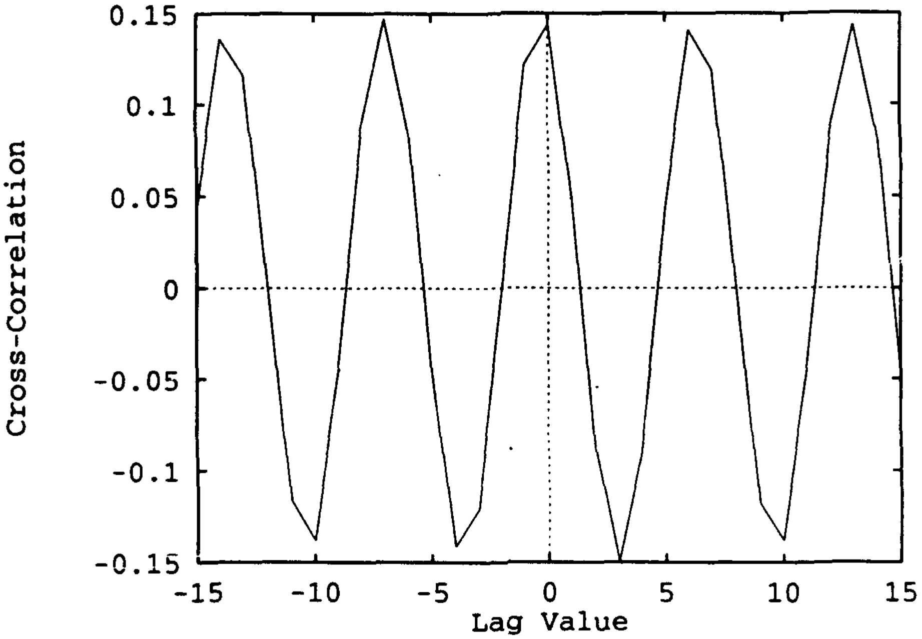

Figure 3.13. Crosscorrelation of two received ear signals. 3 kHz tonal source located at 90°. Note periodic nature of the curve.

3.5.3.1 18-Class Experiment 2.1. In this experiment, random frequencies were selected for each of the 720 tone sound sources. A histogram of the frequency file shows the distribution to be approximately uniform (figure 3.9). Each training run used a different value from the frequency file for each of the 720 values in the angle file. This shuffling of the frequency file was performed by rotating the list of frequencies in a wrap-around circular manner, while the angle file remained static. The amount of rotation after each training run was 90 list items. This effectively divided the frequency file into 8 equally sized segments. Thus, after 8 training runs each 90 item segment in the angle file had been paired with each 90 item segment in the frequency file.

3.5.3.2 18-Class Experiment 2.2. The gaussian noise used in this experiment was generated in exactly the same manner as in the previous experiment

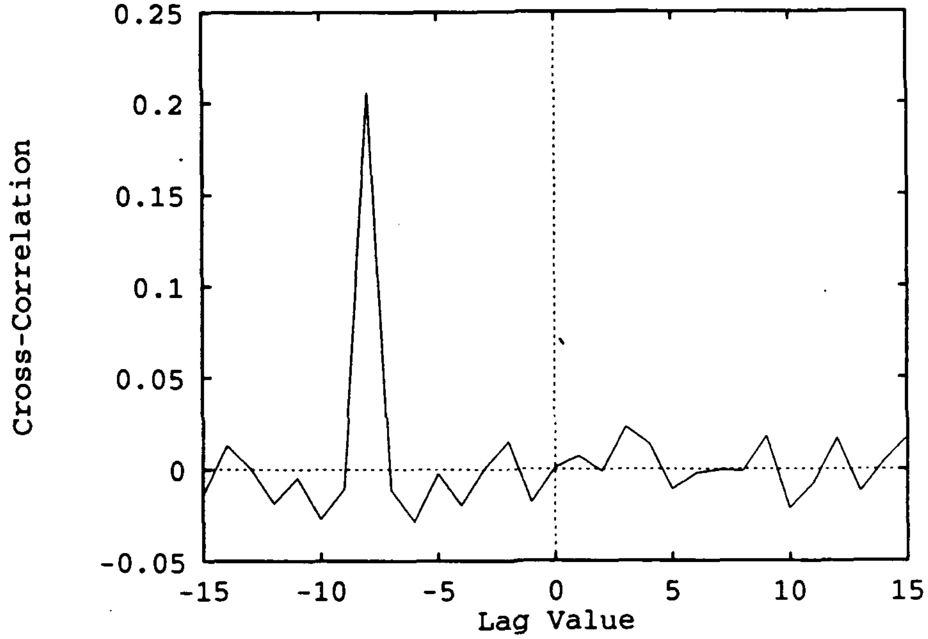

Figure 3.14. Crosscorrelation of two received ear signals. Gaussian noise source located at 225°. The maximum value of this function is apparent for the case of gaussian noise.

(section 3.5.2.2). For each training run, a completely new set of training and test vectors was generated from unique noise sound-source signals.

3.5.4 Localization Experiments Using the Auto-Correlations and the Cross-Correlation as Features. As in the previous set of experiments, three features were used as input features to the ANN. This time, a different measure of the intensity of the signal received at each ear was used: the average power $R_{xx}(0)$. Once again, the position of the peak value in $R_{xy}(\tau)$ will provide an estimate of the relative time delay between the received signals at the two ears. This feature set is related in theory to the spatial hearing model proposed by Shamma (21:989) which uses correlation information derived from the received ear signals.

Again, the two classes of signals used in these experiments were random-frequency tones and gaussian noise. Localization of tones using this feature set were examined in 18-Class Experiment 3.1, while the localization of the noise sources were performed in 18-Class Experiment 3.2. The procedures and parameters used in these two experiments were identical to those used in the experiments in 18-Class Experiments 2.1 and 2.2.

3.6 Conclusion

The results of the experiments described in this chapter are summarized in chapter 4. These results will include the statistics gathered from each experiment as well as their interpretations.

IV. Results

4.1 Introduction

This chapter contains the results of the experiments which were described in chapter 3. The training and test performances are discussed for each of the experiments. In addition, the experimental results are analyzed. The organization of the experimental results will follow that of the *Experiment Design* section of the previous chapter.

4.2 Preliminary Experiments

As was explained previously, the preliminary experiments were designed to explore the ability of a MLP to localize tones using samples of the signals received by the ears as features. In these experiments, localization was made to one of the four angular quadrants in the horizontal plane (see figure 3.3). These experiments were also designed to examine the results of training a neural network with fixed, discrete and random frequency tones.

4.2.1 Preliminary Experiment 1: Localization of 1 kHz Tones. The feature vectors used in this experiment were samples of the received signals from tonal sound sources with a fixed frequency of 1 kHz.

4.2.1.1 Training Performance. The P(Good) and MSE function values converged to maximum and minimum values respectively, after less than 10,000 iterations (see figures 4.1 and 4.2). Learning was terminated at 60,000 iterations. The P(Good) and the MSE curves were flat from approximately 10,000 to 60,000 iterations.

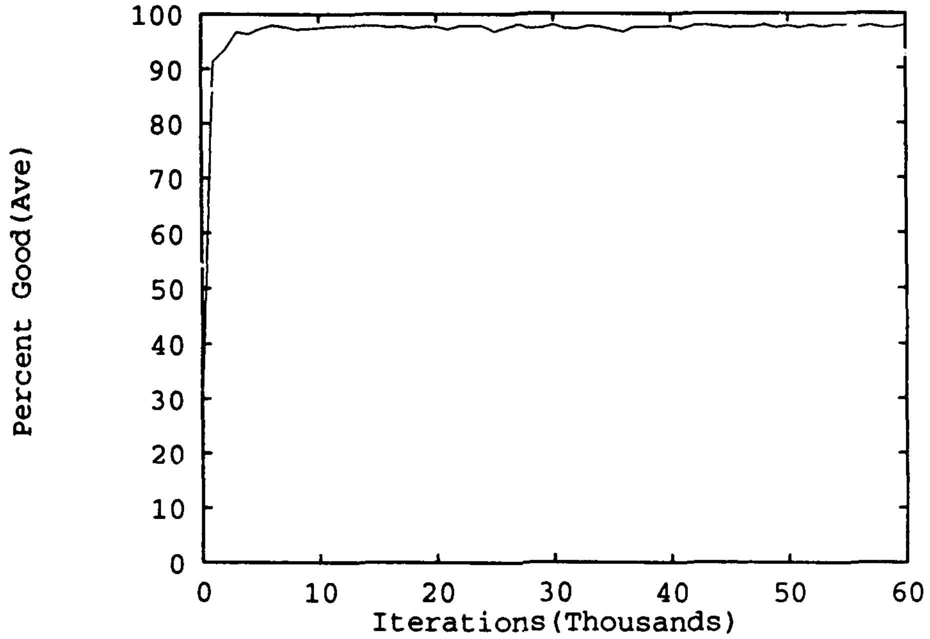

Figure 4.1. Learning plot: Percent Good, Preliminary Experiment 1.

Table 4.1. Test Statistics: Preliminary Experiment 1.

Run	$P(Good)$	$P(W1C)$	$P(HHN)$	$P(FBE \mid E)$	$P(CBCC \mid W1C)$
1	0.977	1.000	1.000	0.000	1.000
2	0.977	1.000	1.000	0.000	1.000
3	0.977	1.000	1.000	0.000	1.000
4	0.977	1.000	1.000	0.000	1.000
Ave	0.977	1.000	1.000	0.000	1.000

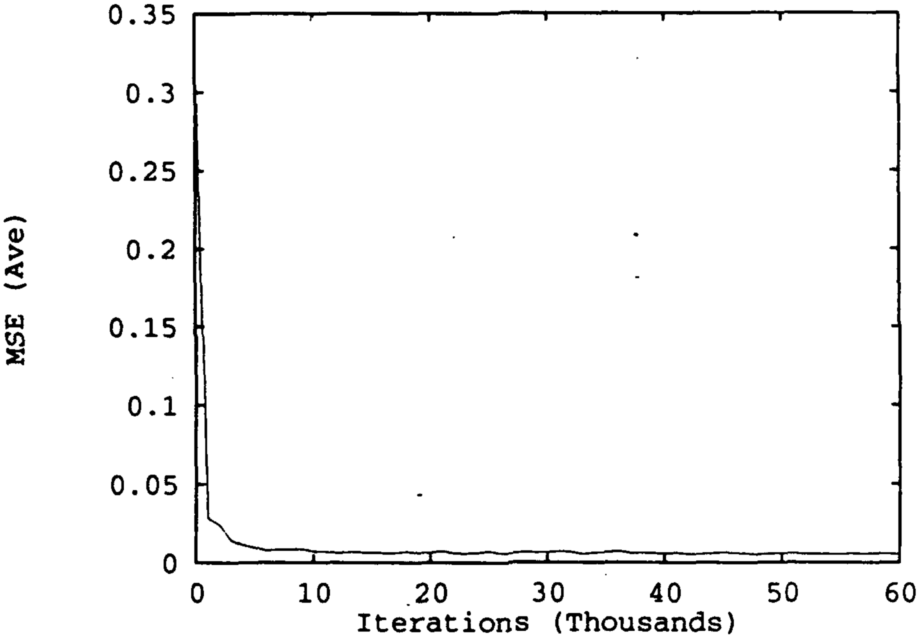

Figure 4.2. Learning plot: Preliminary Experiment 1.

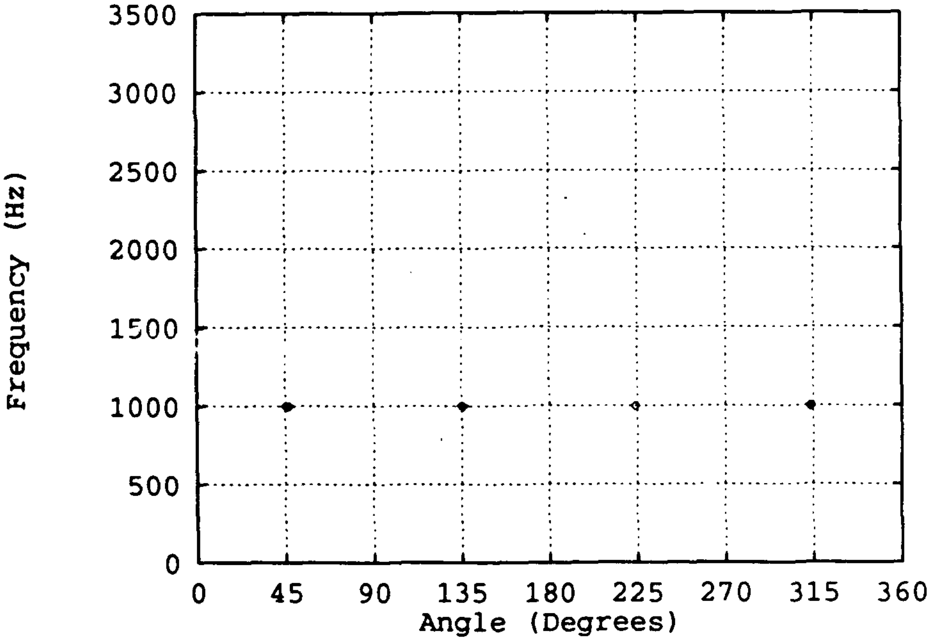

Figure 4.3. Distribution of errors (1 run): Preliminary Experiment 1. Each point represents a classification error in the experiment. The errors are plotted in terms of angular position of the sound-source and the frequency of the tone.

4.2.1.2 Analysis of Test Results. In each of the four runs made in this experiment, the P(Good) achieved was 0.977 (table 4.1). Likewise, all of the other statistics tabulated were identical from run to run. The difference between the runs was only in the particular test vectors which were incorrectly classified in each. As can be seen from the table, P(W1C) = 1.000, P(HHN) = 1.000, and P(CBCC | W1C) = 1.000. These statistics indicate that all errors were one class away from the correct class, that they all were close to the class boundaries, and that the highest neighbor node to the winning output node of the network corresponded to the correct class in every case. The error distribution plot (figure 4.3) clearly supports the statistical conclusion that the errors are clustered at the class boundaries (see figure 3.3). In figure 4.3 it can be seen (upon close inspection) that multiple errors occurred near the boundaries at 45°, 135° and 315°. No front-back classification errors occurred in this experiment, as shown by the P(FBE | E) values in the table.

4.2.2 Preliminary Experiment 2: Localization of Discrete-Frequency Tones. In this experiment the feature vectors used were samples of the received signals at the ears from a tonal sound source at one of five discrete frequencies (500, 1000, 1500, 2000 and 3000 Hz).

4.2.2.1 Training Performance The average learning plots for this experiment (figures 4.4 and 4.5) show that the networks did converge to a solution. In this case, the P(Good) and the MSE curves flattened out after 20,000 iterations.

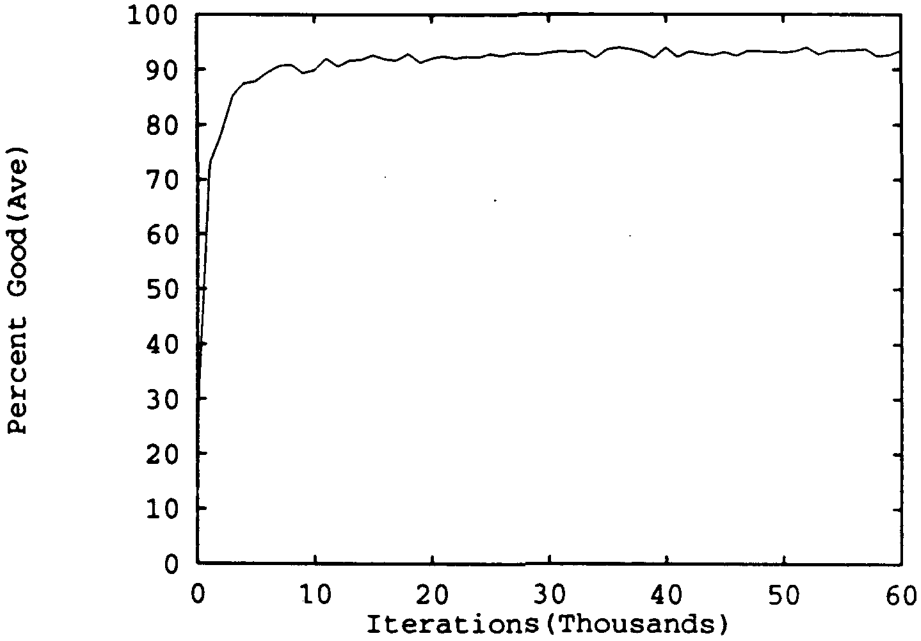

Figure 4.4. Learning plot: Percent Good, Preliminary Experiment 2.

Table 4.2. Test Statistics: Preliminary Experiment 2.

Run	$P(Good)$	$P(W1C)$	$P(HHN)$	$P(FBE \mid E)$	$P(CBCC \mid W1C)$
1	0.945	1.000	1.000	0.000	0.455
2	0.915	1.000	1.000	0.000	0.529
3	0.925	1.000	1.000	0.000	0.333
4	0.950	1.000	1.000	0.000	0.300
Ave	0.934	1.000	1.000	0.000	0.404

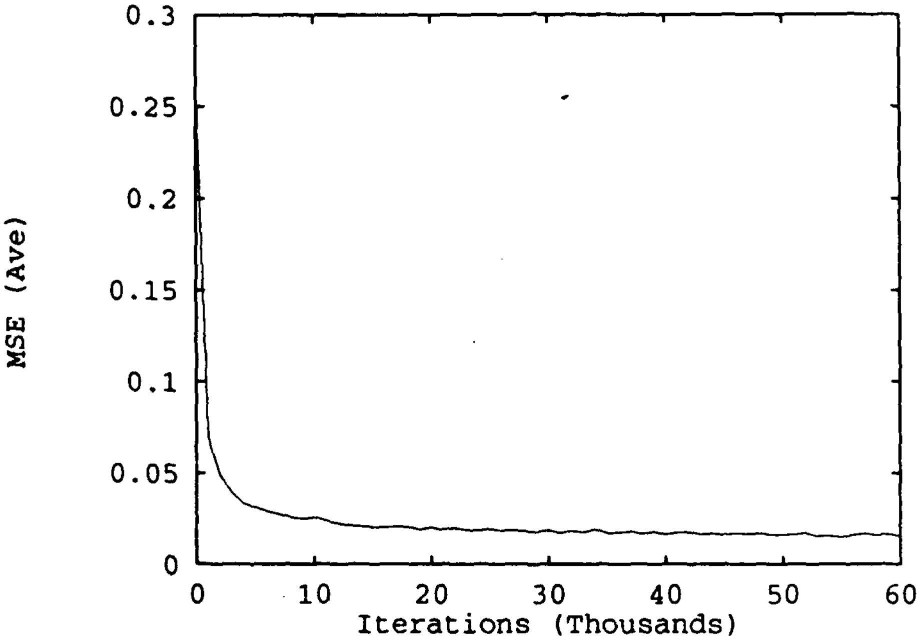

Figure 4.5. Learning plot: MSE, Preliminary Experiment 2.

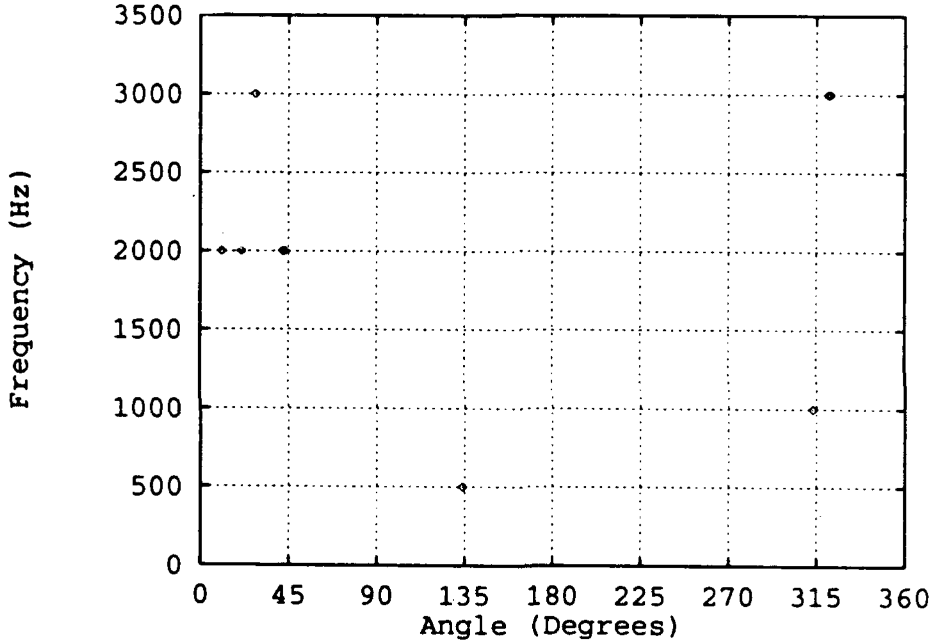

Figure 4.6. Distribution of errors (1 run): Preliminary Experiment 2. Each point represents a classification error in the experiment. The errors are plotted in terms of angular position of the sound-source and the frequency of the tone.

4.2.2.2 Analysis of Test Results. The classification accuracy achieved by the networks was 0.934 on average (table 4.2). As in the first experiment, the errors were all classified in neighbors of the correct class and the highest neighbor of the high output node for each misclassified vector indicated the correct class (P(W1C) & P(HHN) = 1.000). The P(CBCC | W1C) statistic indicates that approximately 40 % of the errors were close to the boundaries. Figure 4.6 indicates that some of the other errors may have been just slightly outside of the +/- 4° range used to define "close"(just above 315°). No front-back errors occurred in any runs.

4.2.3 Preliminary Experiment 3: Localization of Tones at Frequencies Other Than Used for Training. A table of statistics is not included for this experiment because the focus was on which vectors were correctly classified. The training vectors represented tones with frequencies at the discrete values of 500, 1000, 1500, 2000 and 3000 Hz. The tones used to create the test vectors were at discrete frequencies distributed between 400 Hz and 3400 Hz. The angle of origin for the sound sources was one of four discrete values (20°, 110°, 200° and 290°). Each of these angles were represented at each discrete frequency value.

4.2.3.1 Training Performance. The network in this experiment achieved a maximum value of P(Good) at 3000 – 4000 iterations (figure 4.7). From that point on, the P(Good) declined slightly, converging to a value of roughly 0.550. The slight decrease in classification accuracy may be a result of the network memorizing the training vectors. Evidently, the network's ability to generalize to the vectors in the test set was better at 4000 than at 60,000 iterations. The MSE plot (figure 4.8) shows steady decline to a convergence of approximately 0.020 at approximately 15,000 iterations.

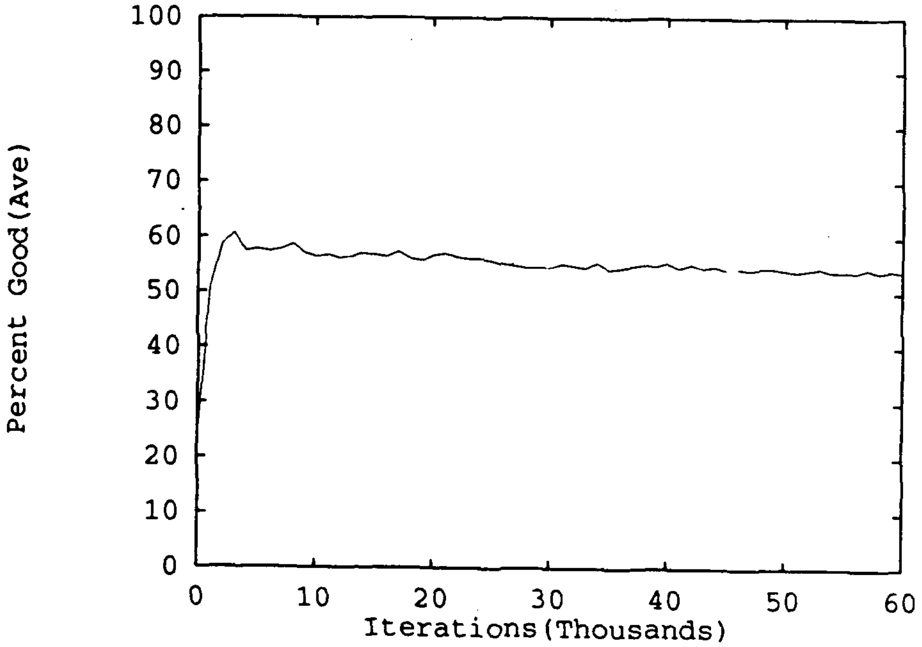

Figure 4.7. Learning plot: Percent Good, Preliminary Experiment 3.

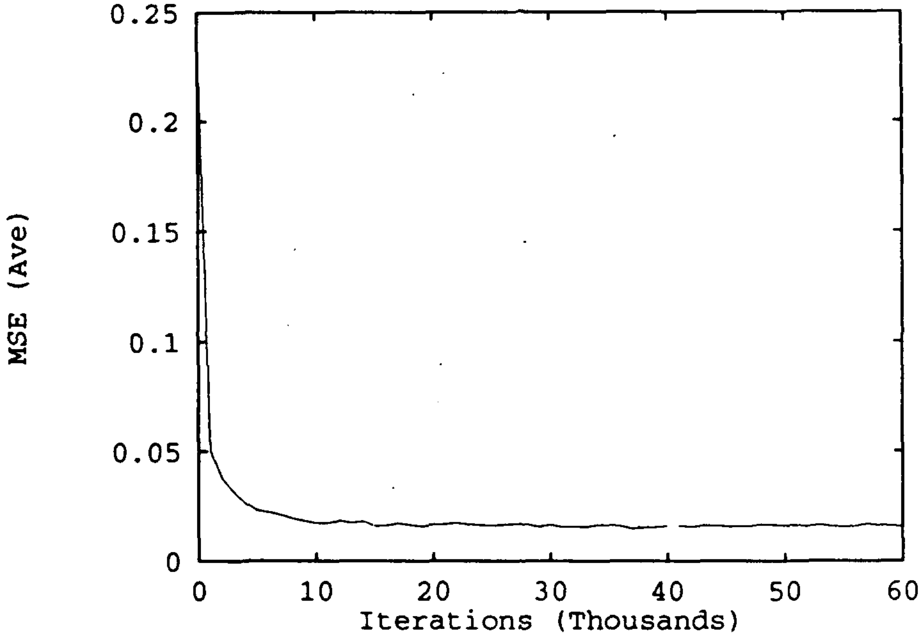

Figure 4.8. Learning plot: MSE, Preliminary Experiment 3.

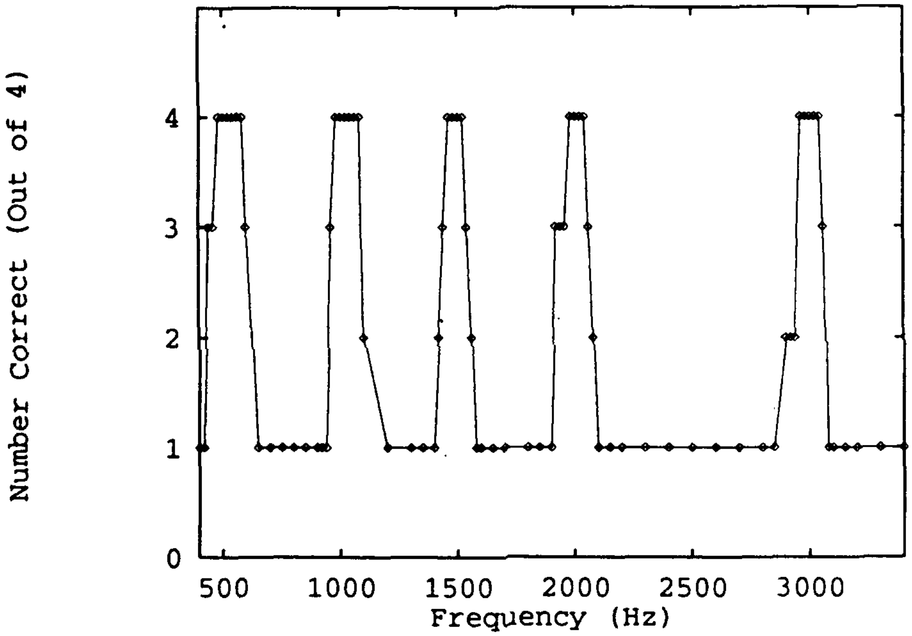

Figure 4.9. Accuracy plot: Preliminary Experiment 3. The plot shows the number of correctly classified vs frequency (points are at frequencies tested).

4.2.3.2 Analysis of Test Results. Figure 4.9 shows how many of the four vectors (one from each class) for each frequency in the range 400 Hz – 3400 Hz were correctly classified. The plot shows that at least one vector was correctly classified at each frequency. This was a result of the fact that all class 3 vectors (sounds originating from the rear) were correctly chosen in the test. Without exception, the vectors of the other classes were correctly classified only when the frequency of the tone was within +/- 50 Hz of the frequencies of the tones used to train the network. This phenomena leads to the conclusion that the network can only correctly classify vectors which are similar to those it is trained with. If the test-tone has a frequency which is not in the same range as the training-tone, then the network will not perform well.

4.2.4 Preliminary Experiment 4: Localization of Random-Frequency Tones to 4 Classes.

The feature vectors in this experiment are samples of the received ear signals from random-frequency tonal sound sources.

4.2.4.1 Training Performance. The networks in this experiment successfully converged. Examination of figures 4.10 and 4.11 reveals that convergence occurred in terms of both P(Good) and MSE after approximately 25,000 iterations.

Table 4.3. Test Statistics: Preliminary Experiment 4.

Run	$P(Good)$	$P(W1C)$	$P(HHN)$	$P(FBE \mid E)$	$P(CBCC \mid W1C)$
1	0.815	0.995	0.980	0.027	0.278
2	0.805	0.995	0.960	0.026	0.184
3	0.810	0.995	0.990	0.026	0.162
4	0.865	0.990	0.990	0.000	0.185
Ave	0.824	0.994	0.980	0.020	0.202

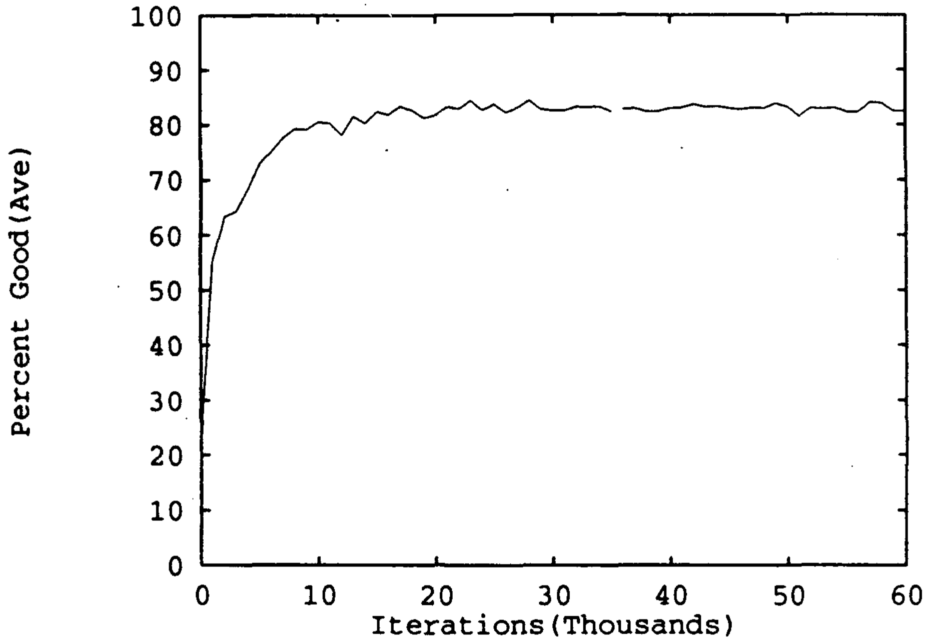

Figure 4.10. Learning plot: Percent Good, Preliminary Experiment 4.

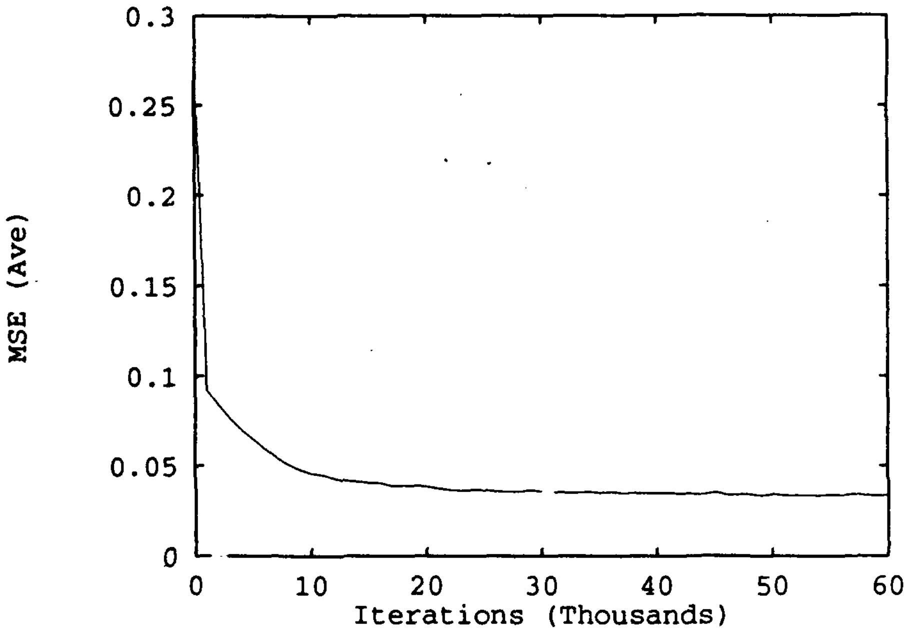

Figure 4.11. Learning plot: MSE, Preliminary Experiment 4.

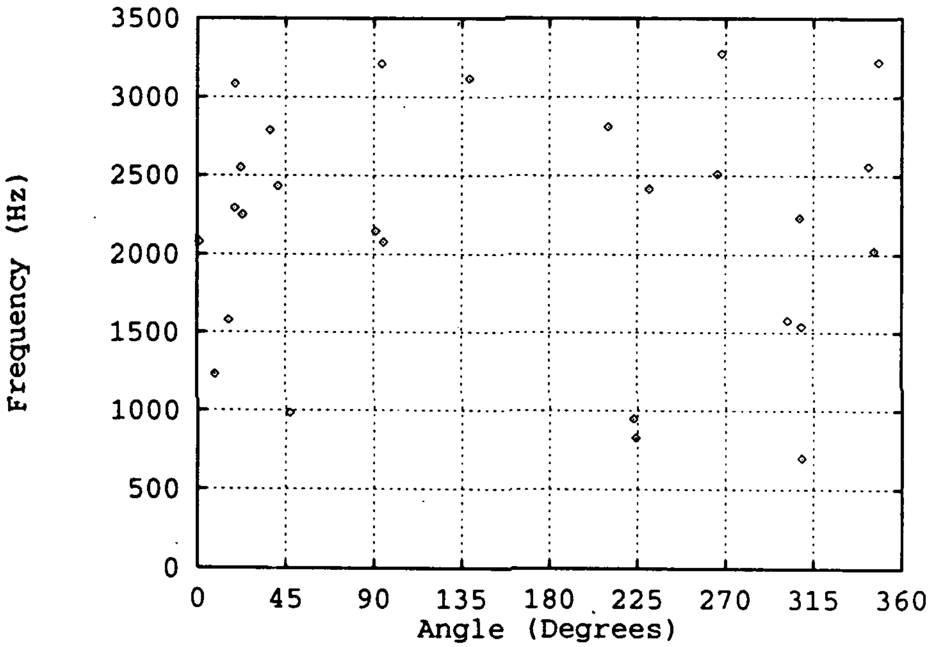

Figure 4.12. Distribution of errors (1 run): Preliminary Experiment 4. Each point represents a classification error in the experiment. The errors are plotted in terms of angular position of the sound-source and the frequency of the tone.

4.2.4.2 Analysis of Test Results. The statistics in table 4.3 show that the average classification accuracy obtained by the networks in this experiment was 82.4 %. The percentage of vectors which were classified in neighbors of the correct class and the percentage where the correct class could be predicted by the highest valued neighbor to the high output node were 99.4 % and 98.0 % respectively. This means that nearly all of the vectors could be localized to an area of two classes, which in this case would be half of the horizontal plane.

The value of the P(CBCC | W1C) statistic indicates that 20.2 % of the one-class errors were close to the boundaries. Figure 4.12 shows that errors were clustered around the class boundaries (45°, 135°, 225°, and 315°), however errors also occurred in right-left symmetrical positions (90° and 270°, 20° and 340°). These results indicate that a significant number of the errors are due to proximity of the sound source to a class boundary, and that the errors occur in a pattern. The angular symmetry of the error locations could be explained by the geometric symmetry of the head and the horizontal circle (see figure 3.3).

It is also noted that no errors occurred in the region straight to the rear (180°). This is consistent with the results of preliminary experiment 3 (see section 4.2.3.2). For an unknown reason, the sound sources to the rear (class 3) appear to be the easiest for the ANN to localize. This is counter to observations made by Makous (14:2198) which indicated that localization of rear stimuli was inferior to localization of frontal stimuli. Front-back errors only made up 2.0 % of the errors in this experiment.

4.3 18-Class Localization Experiments Using Time-Samples as Features

This set of experiments used the same features as in preliminary experiment 4. In this case however, the horizontal plane is divided into 18 rather than 4 classes.

4.3.1 18-Class Experiment 1.1. In this experiment random-frequency tones were used as the sound sources.

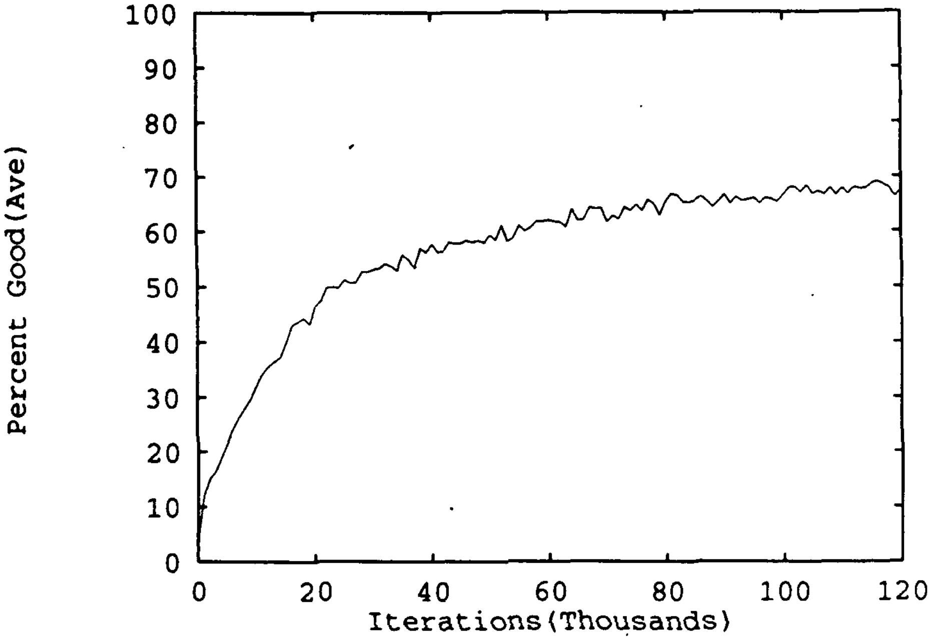

Figure 4.13. Learning plot: Percent Good, 18-Class Experiment 1.1.

4.3.1.1 Training Performance. Figures 4.13 and 4.14 show the average learning progress made by the networks in this experiment. The MSE dropped precipitously in the first few thousand iterations, while the P(Good) showed more gradual convergence. The reason for the mismatch in the rates of convergence for the two curves is unknown.

The point at which the P(Good) converged appears to have been just above 100,000 iterations.

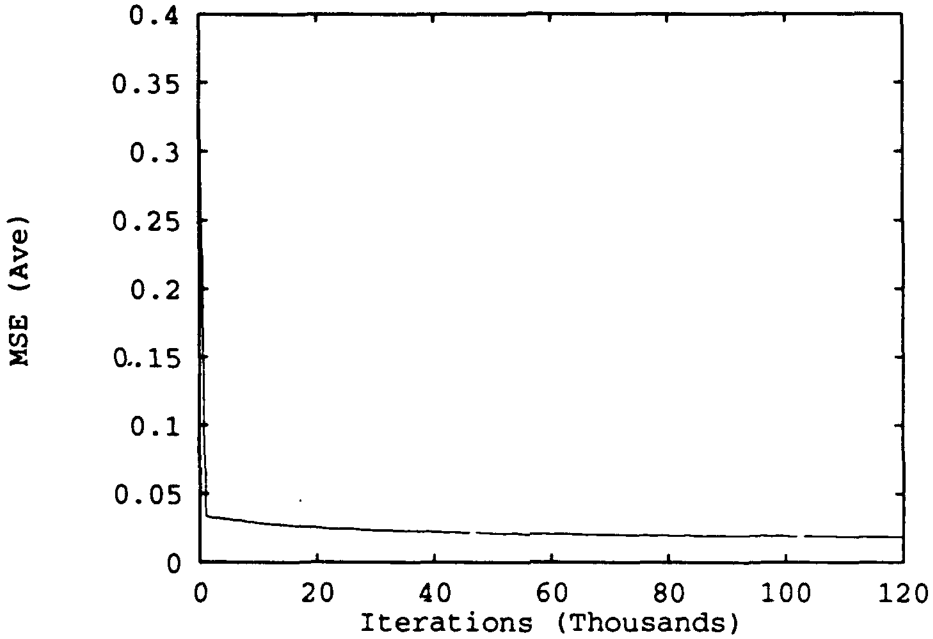

Figure 4.14. Learning plot: MSE, 18-Class Experiment 1.1.

Table 4.4. Test Statistics: 18-Class Experiment 1.1.

Run	$P(Good)$	$P(W1C)$	$P(HHN)$	$P(FBE \mid E)$	$P(CBCC \mid W1C)$
1	0.685	0.845	0.795	0.175	0.222
2	0.695	0.880	0.855	0.033	0.361
3	0.680	0.855	0.830	0.109	0.328
4	0.665	0.885	0.825	0.075	0.299
5	0.725	0.875	0.850	0.055	0.400
6	0.625	0.840	0.800	0.040	0.320
7	0.710	0.900	0.865	0.017	0.379
8	0.615	0.855	0.810	0.026	0.260
Ave	0.675	0.867	0.829	0.066	0.321

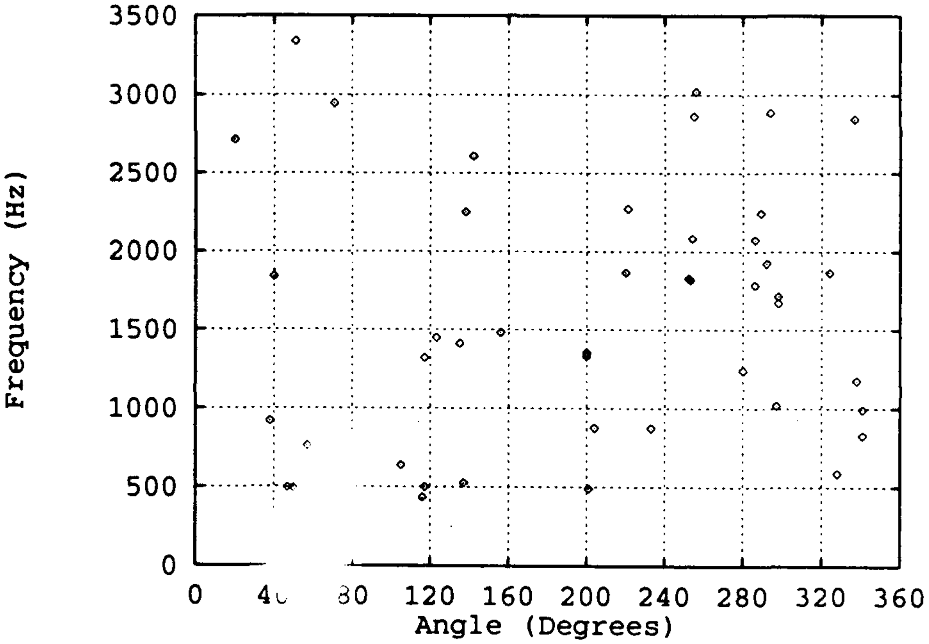

Figure 4.15. Distribution of errors (1 run): 18-Class Experiment 1.1. Each point represents a classification error in the experiment. The errors are plotted in terms of angular position of the sound-source and the frequency of the tone.

4.3.1.2 Analysis of Test Results. This experiments showed the network to have the ability to correctly classify 67.5 % of the test vectors (table 4.4). This compares with 82.4 % accuracy achieved earlier using the same features with 4 classes (preliminary experiment 4). The value of P(HHN) indicates that the accuracy of the present network would be 82.9 % when localizing to within 40°. Thus, greater localization accuracy was achieved to a resolution of 40° using the "high and highest neighbor" criteria on an 18-class network, than was obtained to a resolution of 90° on the 4-class network. This indicates a large advantage in favor of using the 18-class network.

The percentage of 1-class errors appearing "close" to the boundaries is 32.1 %. This is a larger percentage than was observed in preliminary experiment 4 (20.2 %). It would be expected however, that more errors would occur near boundaries if more boundary lines exist on the horizontal plane. This is what happens when the number of classes is increased. As indicated by the P(CBCC | W1C) statistic, some errors were clustered about the class boundaries (0°, 20°, 40° etc.). Also as before (preliminary experiment 4), some symmetry is apparent in the angular location of the errors about the 180° point.

6.6 % of the errors in this experiment were identified as front-back errors. This is greater than the 2.0 % figure found in preliminary experiment 4 (for 4 classes). The occurrence of front-back errors has also been noted in human localization experiments (14:2188).

4.3.2 18-Class Experiment 1.2. Gaussian noise sound sources were used in this experiment.

4.3.2.1 Training Performance. Figures 4.16 and 4.17 illustrate the learning of the networks in this experiment. As can be seen, the "Percent Good" function increased slowly as compared to the corresponding plot where the same features were used with random-tone sound sources (figures 4.16 and 4.13). Convergence of

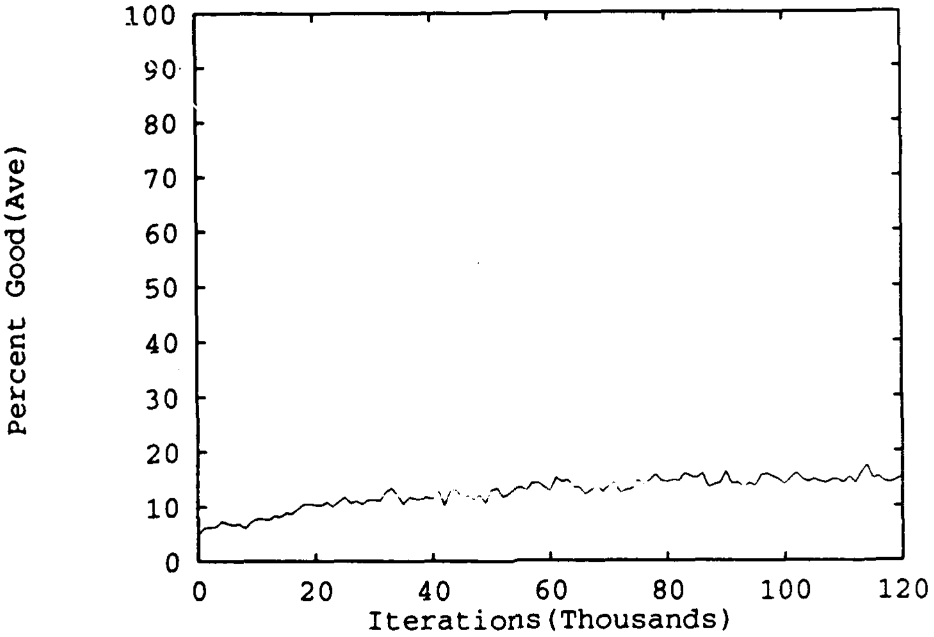

Figure 4.16. Learning plot: Percent Good: 18-Class Experiment 1.2.

the "Percent Good" function appears to have taken place at approximately 100,000 iterations.

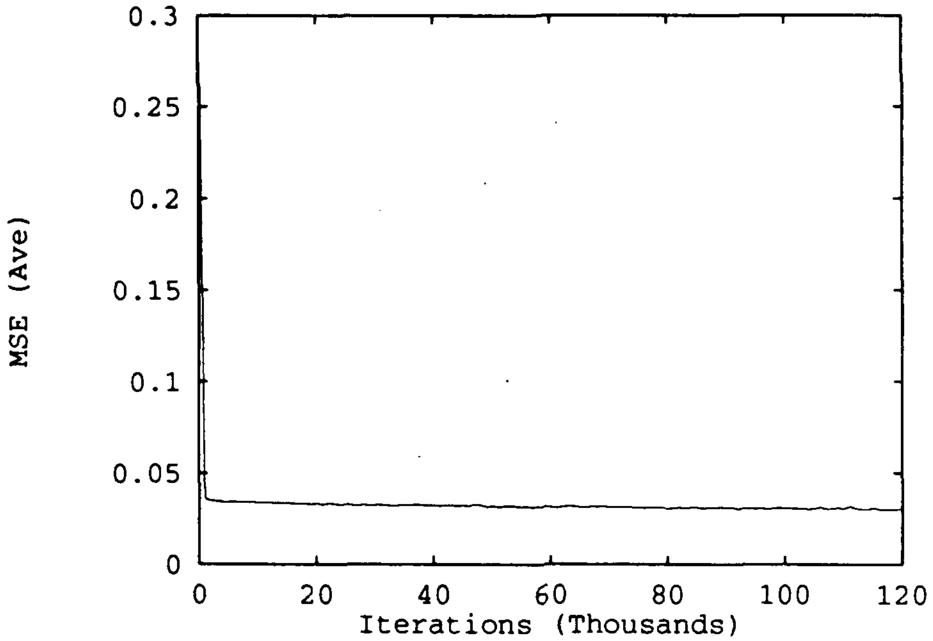

Figure 4.17. Learning plot: MSE, 18-Class Experiment 1.2.

Table 4.5. Test Statistics: 18-Class Experiment 1.2

Run	$P(Good)$	$P(W1C)$	$P(HHN)$	$P(FBE \mid E)$	$P(CBCC \mid W1C)$
1	0.140	0.300	0.230	0.052	0.375
2	0.150	0.295	0.245	0.012	0.345
3	0.120	0.305	0.255	0.028	0.270
4	0.160	0.330	0.275	0.065	0.382
5	0.145	0.300	0.230	0.041	0.387
6	0.130	0.305	0.250	0.034	0.257
7	0.210	0.450	0.340	0.044	0.313
8	0.170	0.385	0.270	0.030	0.279
Ave	0.153	0.334	0.262	0.038	0.326

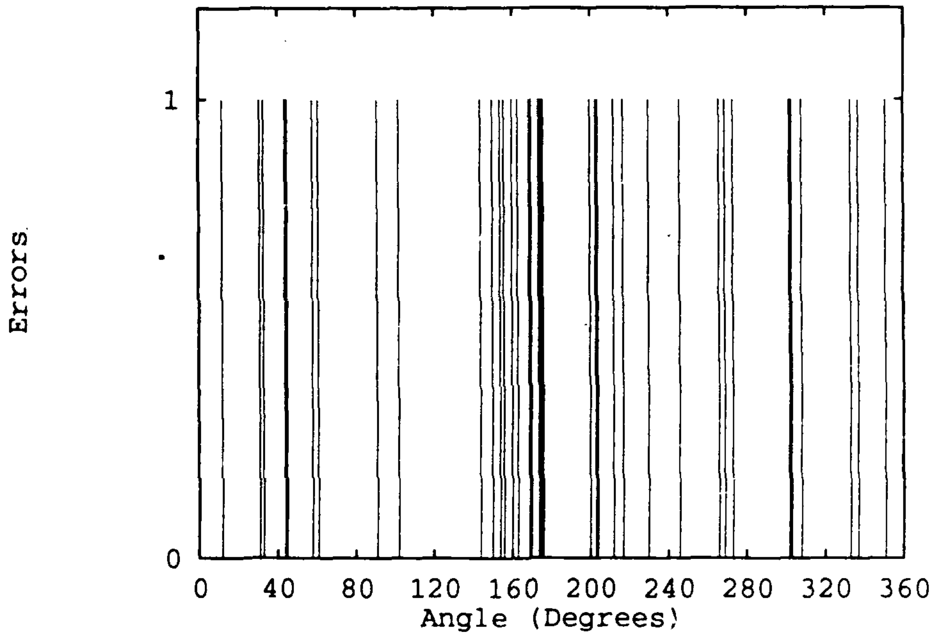

Figure 4.18. Distribution of correct classification (1 run): 18-Class Experiment 1.2. The position of the impulses represents the angular location of the correctly classified vectors. The apparent differences in line widths is due to clusters of correct classifications.

4.3.2.2 Analysis of Test Results. The P(Good) obtained in this experiment was an average of 0.153 (figure 4.5). This value is much lower than the 0.675 accuracy (table 4.4) obtained on the same number of classes, using the same features, with random-frequency tones (18-class experiment 1.1). P(HHN) averaged 0.262, meaning that the location of the sound source could be localized to a region of two adjacent classes approximately 26 % of the time. The random nature of the samples of the gaussian noise may be the cause of the poor performance in classification accuracy.

Front-back errors did not occur frequently, making up just 3.8 % of the classification errors in this experiment. Errors close to the boundaries were a significant portion (32.6 %) of the "one-class" errors.

In this experiment, very few of the vectors were correctly classified. It was therefore much simpler to keep track of the correct classifications rather than the errors as was done in the other experiments. The accuracy distribution plot (figure 4.18) shows that the correct classifications occurred in what appears to be a random pattern.

4.4 18-Class Localization Experiments Using the Mean FFT Magnitudes and the Cross-Correlation as Features

4.4.1 18-Class Experiment 2.1. The sound sources used in this experiment were random-frequency tones.

4.4.1.1 Training Performance. As can be seen in figures 4.19 and 4.20, convergence of the average Percent Good and MSE values was accomplished during training. For unknown reasons, the MSE function leveled out earlier than that of the Percent Good. The Percent Good function appears to have converged at roughly 100,000 iterations.

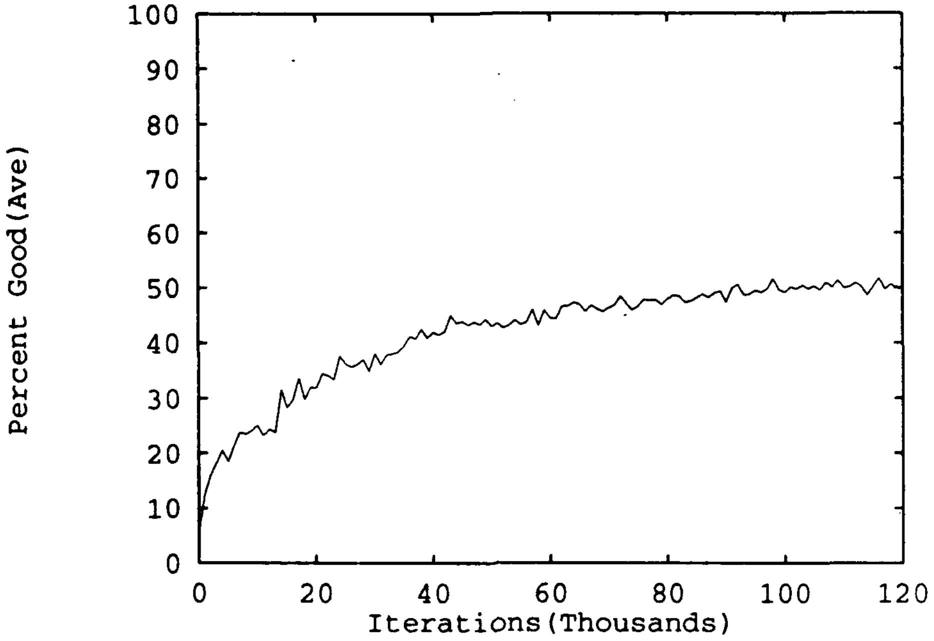

Figure 4.19. Learning plot: Percent Good, 18-Class Experiment 2.1

Table 4.6. Test Statistics: 18-Class Experiment 2.1.

Run	$P(\mathit{Good})$	$P(W1C)$	$P(HHN)$	$P(FBE \mid E)$	$P(CBCC \mid W1C)$
1	0.519	0.725	0.661	0.277	0.392
2	0.492	0.683	0.647	0.268	0.377
3	0.458	0.708	0.669	0.282	0.433
4	0.472	0.686	0.622	0.263	0.403
5	0.486	0.725	0.653	0.232	0.337
6	0.564	0.756	0.686	0.331	0.464
7	0.489	0.678	0.631	0.304	0.515
8	0.505	0.728	0.639	0.360	0.388
Ave	0.498	0.711	0.651	0.290	0.414

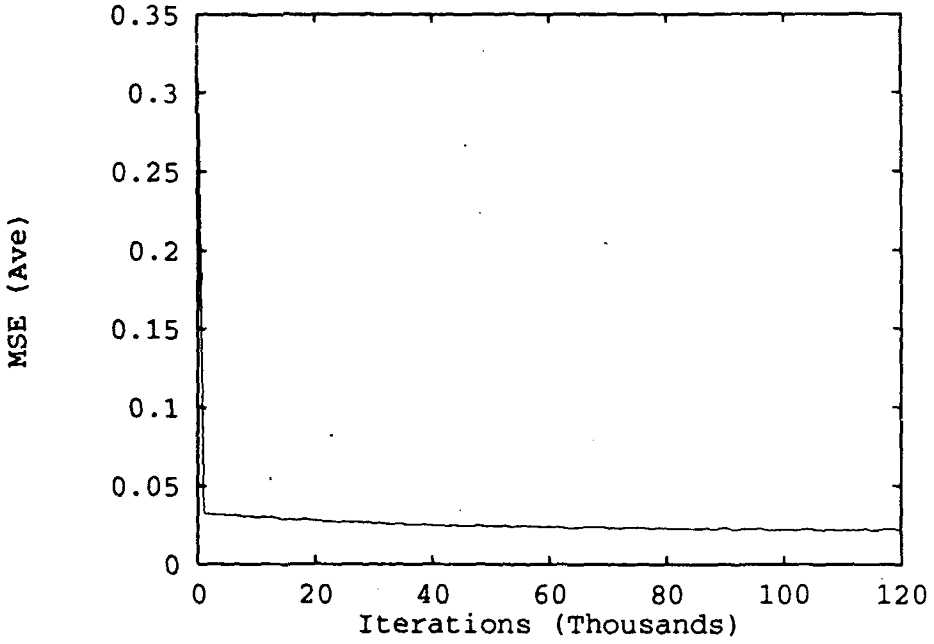

Figure 4.20. Learning plot: MSE, 18-Class Experiment 2.1.

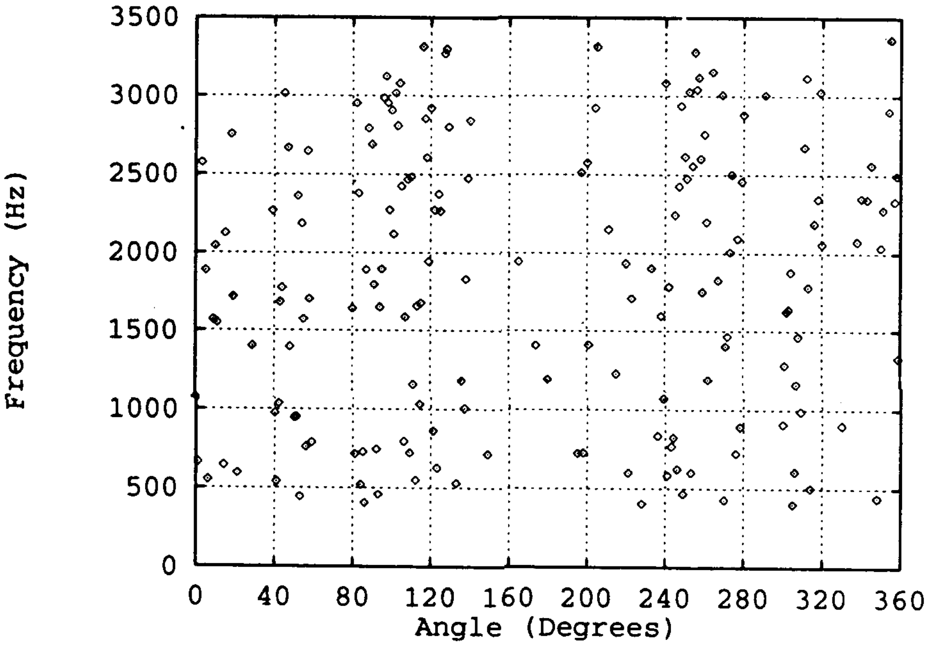

Figure 4.21. Distribution of errors (1 run): 18-Class Experiment 2.1. Each point represents a classification error in the experiment. The errors are plotted in terms of angular position of the sound-source and the frequency of the tone.

4.4.1.2 Analysis of Test Results. In this experiment an average classification accuracy (P(Good)) of 49.8 % was achieved (table 4.6), while 71.1 % of the vectors were either in the correct class or one of the two bordering classes. Using the high and highest neighbor criteria (defined in section 3.3), 65.1 % of the vectors could be localized to a region consisting of two adjacent classes. All of these statistics indicate that the features used in this experiment do not perform as well as the time-samples used as features in 18- class experiment 1.1.

In this experiment a significant number of front-back errors were observed. The front-back misclassifications accounted for 29.0 % of the errors. The high percentage of front-back errors is probably due to variability in the calculated feature values. Observations of the values calculated for the mean FFT magnitudes of the ear signals showed a significant amount of variance in these values for tones of different frequencies from a given sound source direction. The variance in these values appeared to depend on how close the frequency of the tone was to one of the frequency bins in the FFT sequence. Because the ITD (estimated by the cross-correlation) would be the same for front-back mirror positions (about the axis through the ears), lack of consistency in the other features could possibly "muddy the waters" between front-back positions.

41.4 % of the one-class errors were from sound sources located close to the boundaries but many of the errors were not of the one-class variety (P(Good) = 0.498, P(W1C) = 0.711, 1 - P(W1C) = errors not within one class = 0.289). Figure 4.21 shows that the angular position of the errors in this experiment was somewhat symmetrical about 180°. The errors appeared to be at random frequencies.

4.4.2 18-Class Experiment 2.2 Gaussian noise was used as the sound source in this experiment.

4.4.2.1 Training Performance Figures 4.22 and 4.23 show the average training progress made by the networks in this experiment. The Percent Good and

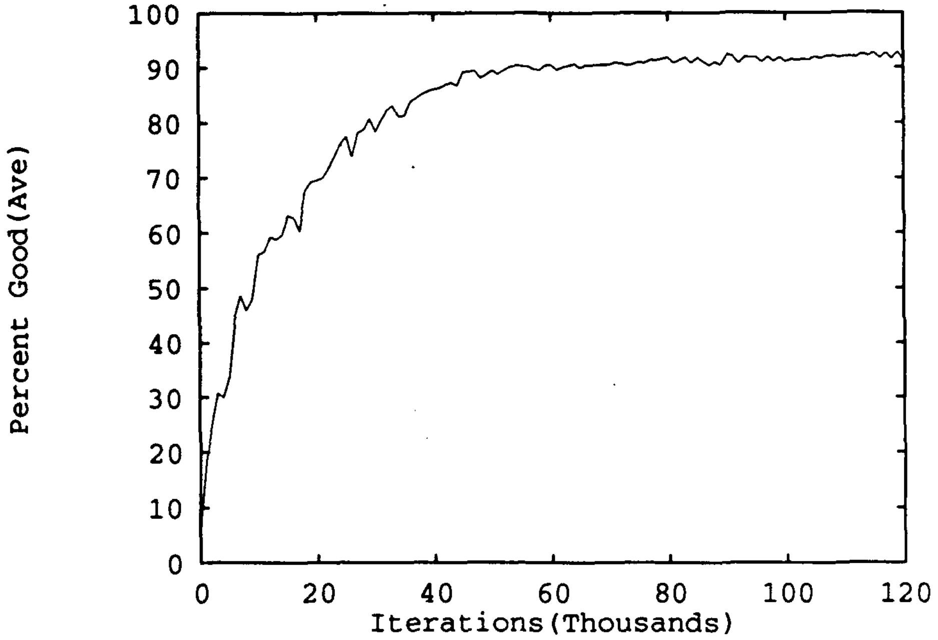

Figure 4.22. Learning plot: Percent Good, 18-Class Experiment 2.2.

MSE curves show that convergence occurred for these functions at approximately 100,000 iterations.

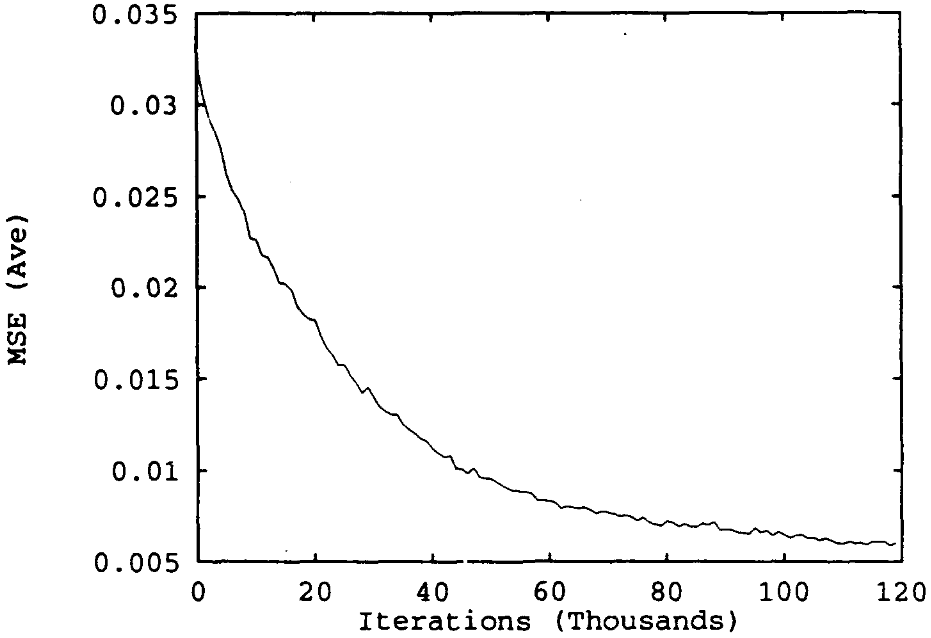

Figure 4.23. Learning plot: MSE, 18-Class Experiment 2.2.

Table 4.7. Test Statistics: 18-Class Experiment 2.2.

Run	$P(Good)$	$P(W1C)$	$P(HHN)$	$P(FBE \mid E)$	$P(CBCC \mid W1C)$
1	0.906	1.000	1.000	0.000	0.971
2	0.894	1.000	1.000	0.000	0.895
3	0.919	1.000	1.000	0.000	0.897
4	0.928	1.000	1.000	0.000	0.962
5	0.919	1.000	1.000	0.000	0.931
6	0.908	1.000	1.000	0.000	0.909
7	0.919	1.000	1.000	0.000	0.966
8	0.914	1.000	1.000	0.000	1.000
Ave	0.913	1.000	1.000	0.000	0.941

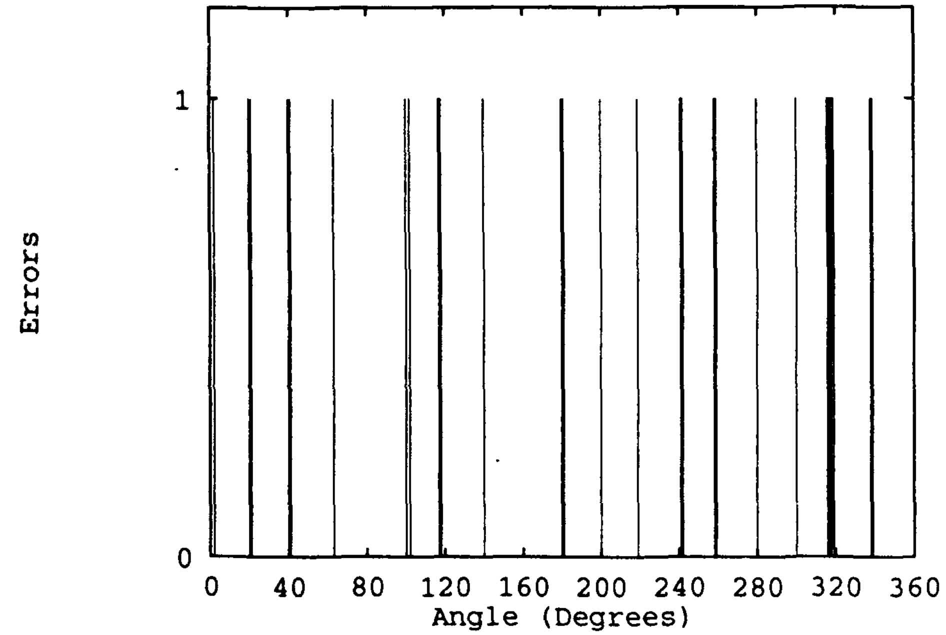

Figure 4.24. Distribution of errors (1 run): 18-Class Experiment 2.2. The position of the impulses represents the angular location of the classification errors. Clusters of errors are evidenced by the impulses which appear to be thicker.

4.4.2.2 Analysis of Test Results. The statistics for this experiment are contained in table 4.7. The average P(Good) achieved was 0.913. As indicated by the P(W1C) statistic, all of the errors were "within one class". Additionally, the correct class could be identified in all of the misclassifications by using the "high and highest neighbor" criteria (as defined in section 3.3). These statistics indicate a high level of classification accuracy of this network on gaussian noise sound sources.

No front-back errors were observed in any of the network runs. Also, nearly all of the errors in this experiment resulted from sound sources which were located close to the class boundaries (P(CBCC | W1C) = 0.941). This observation is supported by the error distribution plot in figure 4.24. The mean FFT magnitudes and the cross-correlation appear to be much more effective features for use on noise than on tones. This fact is likely to be a result of the wideband nature of the noise signal, which theoretically would produce a constant FFT magnitude sequence.

4.5 18-Class Localization Experiments Using the Auto-Correlations and the Cross-Correlation as Features

4.5.1 18-Class Experiment 3.1. This experiment uses random-frequency tones as the sound sources.

4.5.1.1 Training Performance. The average training progress made by the networks in this experiment is shown in figures 4.25 and 4.26. Convergence in these two plots appears to have occurred after approximately 100,000 iterations.

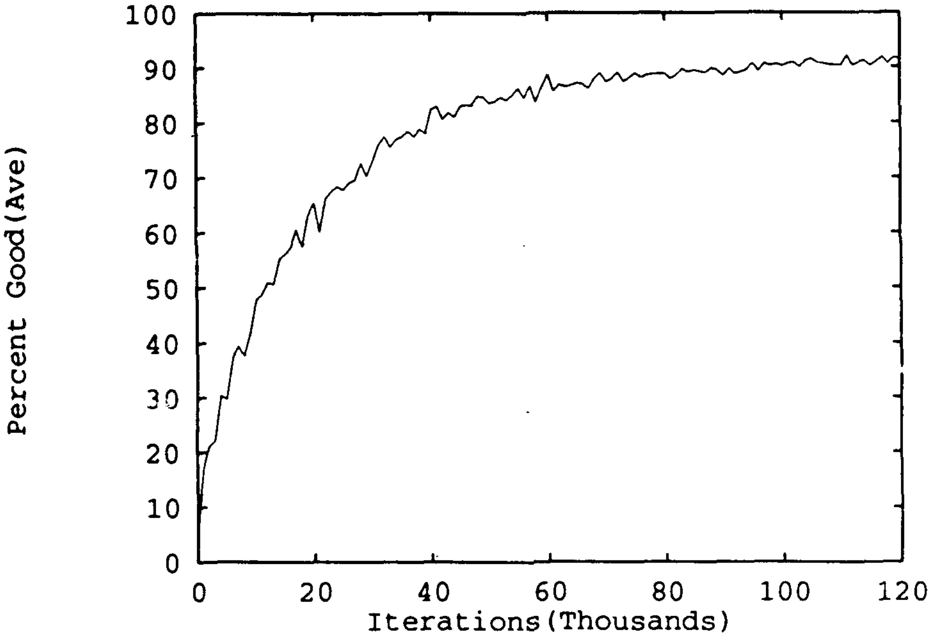

Figure 4.25. Learning plot: Percent Good, 18-Class Experiment 3.1.

Table 4.8. Test Statistics: 18-Class Experiment 3.1.

Run	$P(Good)$	$P(W1C)$	$P(HHN)$	$P(FBE \mid E)$	$P(CBCC \mid W1C)$
1	0.903	0.989	0.983	0.000	0.774
2	0.900	0.994	0.992	0.000	0.853
3	0.944	1.000	0.994	0.000	0.900
4	0.914	0.997	0.978	0.000	0.700
5	0.942	1.000	0.997	0.000	0.905
6	0.925	0.992	0.992	0.000	0.889
7	0.886	0.989	0.986	0.000	0.730
8	0.936	0.997	0.994	0.000	0.995
Ave	0.919	0.995	0.983	0.000	0.838

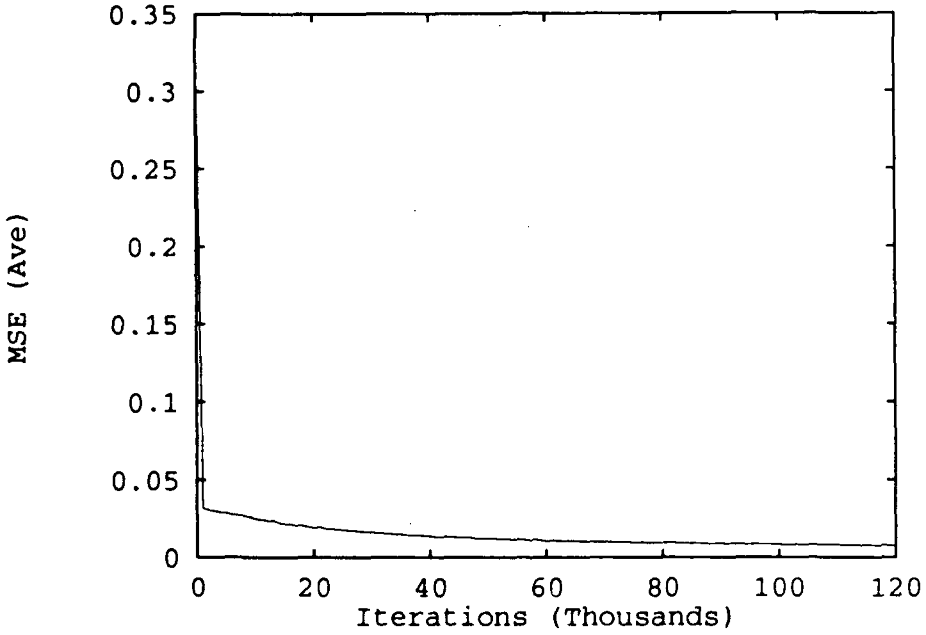

Figure 4.26. Learning plot: MSE, 18-Class Experiment 3.1.

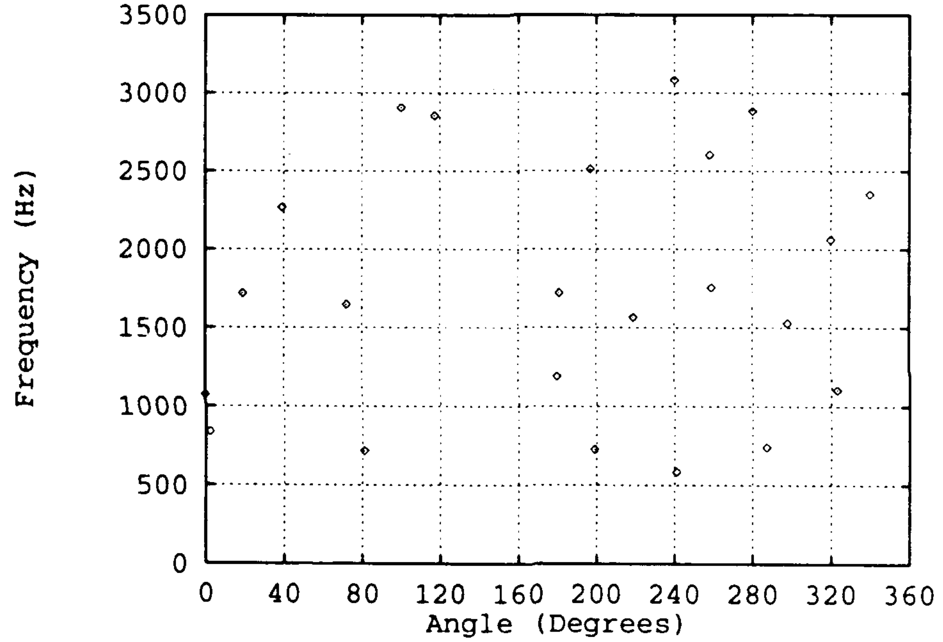

Figure 4.27. Distribution of errors (1 run): 18-Class Experiment 3.1. Each point represents a classification error in the experiment. The errors are plotted in terms of angular position of the sound-source and the frequency of the tone.

4.5.1.2 Analysis of Test Results. The statistics tabulated in this experiment are contained in table 4.8. The classification accuracy of the networks in this experiment averaged 91.9 %, while 99.5 % of the vectors were either classified in the correct class or a neighbor of the correct class. The P(HHN) statistic implies that the position of the sound source could be correctly localized to within two adjacent classes 98.3 % of the time using the "high and highest neighbor" criteria (see section 3.3). Thus, the accuracy of this network is much greater on the tonal sound sources than the network of 18-class experiment 2.1. The difference in the features used in this experiment was the use of $R_{xx}(0)$ rather than the mean of the FFT magnitude for the received ear signals. The test results show that $R_{xx}(0)$ was the better feature.

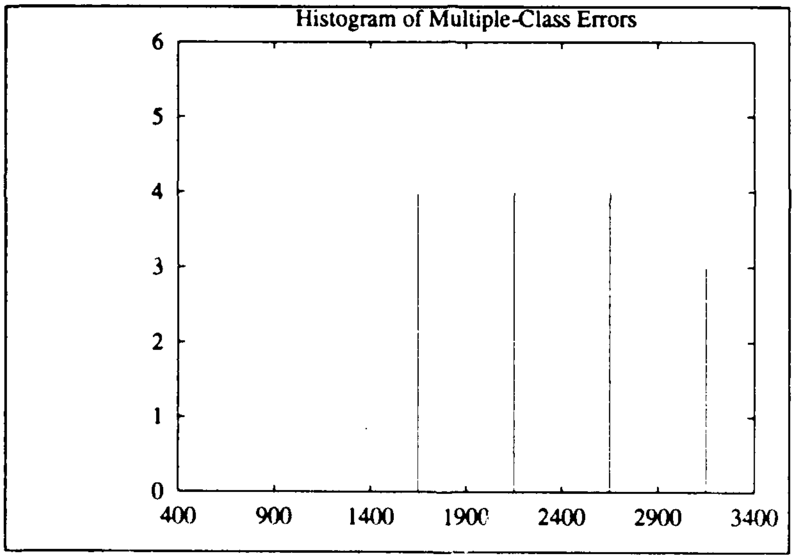

Figure 4.28. Histogram of Multiple-Class Errors: 18-Class Experiment 3.1.

Most of the one-class errors were the result of sound sources being near the class boundaries (P(CBCC | W1C)). Because of the high percentage of one-class errors, the conclusion is that most of all types of errors occurred near the class boundaries. The error distribution plot in figure 4.27 shows that the errors are indeed clustered at the class boundaries. No front-back errors occurred in this experiment.

A small percentage of errors in this experiment were not "one-class" misclassifications and thus may be called "multiple-class" errors. An examination of these errors produced an interesting result. A histogram in terms of frequency of the multiple-class errors is shown in figure 4.28. The histogram shows that all of these errors were above 1400 Hz and were uniformly distributed. This occurrance agrees with observations made on humans in psychological experiments where a 1400 – 1500 Hz threshold was observed, above which localization became more difficult.

The theoretical explanation for this phenomenon was lies in the fact that the ears receive tonal signals which differ in amplitude, but are the same frequency. If the period of the sinusoidal signal falls below the maximum amount of time-delay possible between the two ears (approximately 700 μsec) then confusion about the real interaural time-delay may occur because of multiple periods in the time-window of interest. The cross-correlation function ($R_{xy}(\tau)$) also illustrates the problem. In figures 3.13 and 3.14 it is seen that for a tonal sound source with frequency above approximately 1430 Hz, $R_{xy}(\tau)$ contains multiple peaks with similar magnitudes, while the noise signals produce a very distinct peak in $R_{xy}(\tau)$. The implication is thus: that time-delay estimation is inherently more difficult on tones than on noise, and because of the potential for error in time-delay estimation the possibility for multiple-class errors with tonal sound-sources is greater.

4.5.2 18-Class Experiment 3.2. Gaussian noise was used as the sound source in this experiment.

4.5.2.1 Training Performance. The average training history of the networks in this experiment are displayed in figures 4.29 and 4.30. These plots are both very similar to the plots obtained in 18-class experiment 3.1, where the same features were used with random-frequency tones as the sound source (figures 4.25 and 4.26). Once again the convergence of these curves appears to occur after approximately 100,000 iterations.

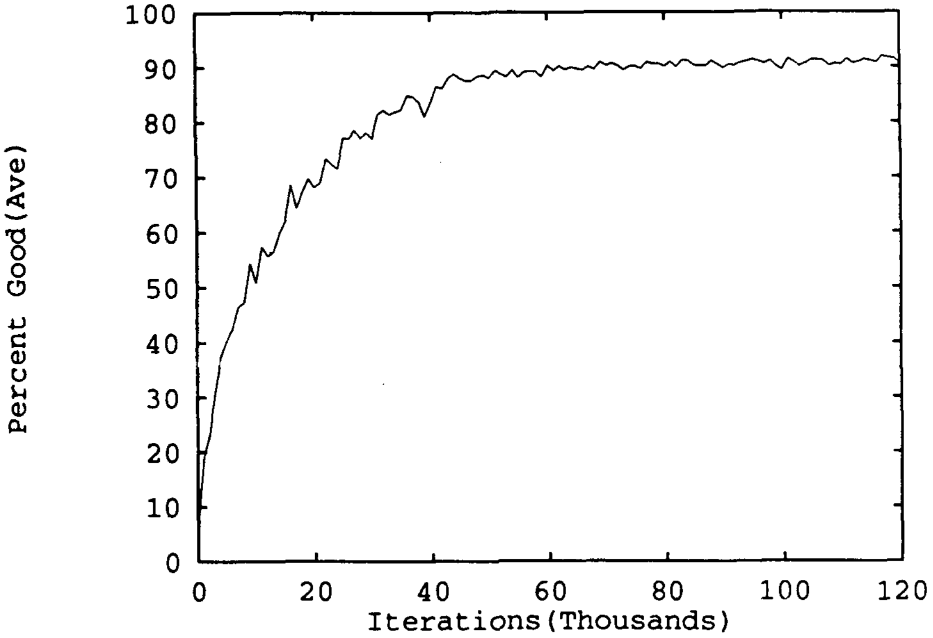

Figure 4.29. Learning plot: Percent Good, 18-Class Experiment 3.2.

Table 4.9. Test Statistics: 18-Class Experiment 3.2.

Run	$P(Good)$	$P(W1C)$	$P(HHN)$	$P(FBE \mid E)$	$P(CBCC \mid W1C)$
1	0.911	1.000	1.000	0.000	0.969
2	0.931	1.000	1.000	0.000	1.000
3	0.908	1.000	1.000	0.000	0.970
4	0.878	1.000	1.000	0.000	0.909
5	0.933	1.000	1.000	0.000	0.875
6	0.931	1.000	1.000	0.000	1.000
7	0.892	1.000	1.000	0.000	0.949
8	0.908	1.000	1.000	0.000	0.909
Ave	0.912	1.000	1.000	0.000	0.947

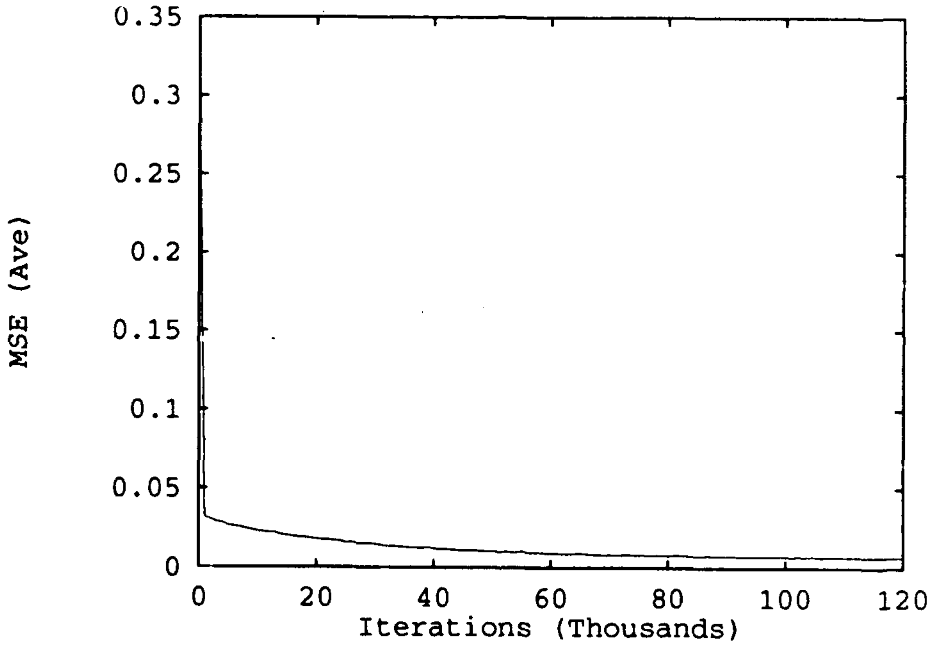

Figure 4.30. Learning plot: MSE, 18-Class Experiment 3.2.

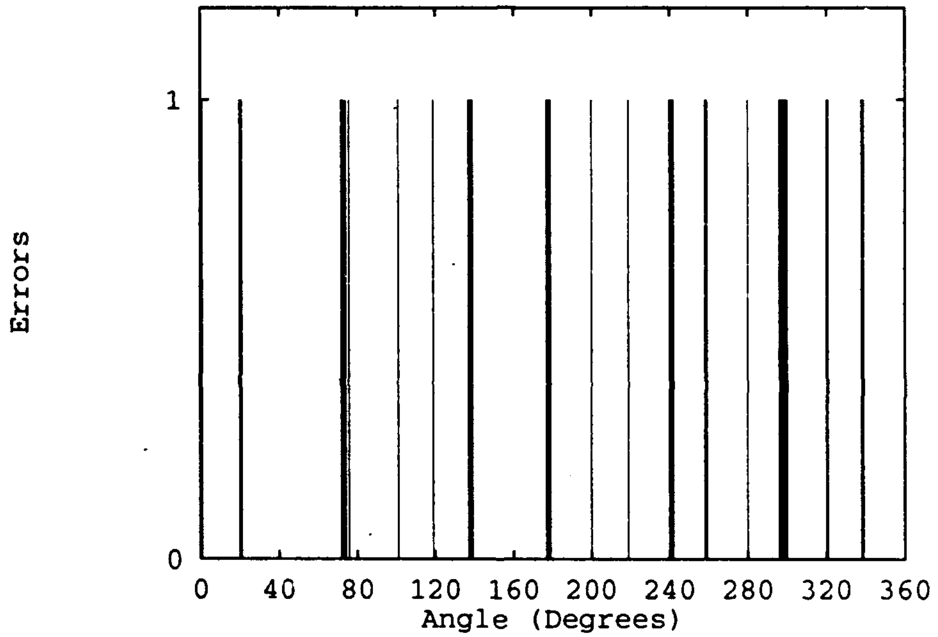

Figure 4.31. Distribution of errors (1 run): 18-Class Experiment 3.2. The position of the impulses represents the angular location of the classification errors. Clusters of errors are evidenced by the impulses which appear to be thicker.

4.5.2.2 Analysis of Test Results. The statistics which summarize the results of this experiment are contained in table 4.9. The average classification accuracy obtained was 91.2 %. This value is very close to the accuracy obtained in the prior experiment which used the same features on random-frequency tones. The values of P(W1C) in the table show that all of the classification errors were one class-errors. The P(HHN) statistic shows that all of the errors could have been correctly localized to a region of two adjacent classes.

As before (in 18-class experiment 3.1) the vast majority of errors were due to close proximity of the sound source to the class boundaries (94.7 %). This is supported by the error distribution plot (figure 4.31), which shows the errors clustered about angle values which are multiples of 20. Again, no front-back errors occurred.

4.6 Conclusion

In this chapter, the results of the experiments performed in this thesis effort were discussed. Chapter 5 will state the conclusions which may be drawn from the results contained in this chapter, and will contain recommendations for further research which may be performed in this area.

V. Conclusions and Recommendations

5.1 Conclusions

5.1.1 Preliminary Experiments. The preliminary experiments show that tonal sound sources can be localized using an ANN. These experiments demonstrated localization of tones of a single frequency, of five discrete frequencies and of random frequencies (with varying degrees of success). The classification was to 1 of 4 classes (front, back, right, or left) using time-samples of the signals received by the ears as features.

It was also shown that the networks could consistently correctly classify tones only when the test and training sets were constructed using tonal sound sources in the same frequency range.

Table 5.1. Summary of average P(Good) results: 18-class Experiments

Features	Time-Samples	Mean FFT Mag. & Cross-Corr.	Auto-corr. & Cross-Corr.
Random-Freq. Tones	0.675	0.498	0.919
Gaussian Noise	0.153	0.913	0.912

5.1.2 18-Class Networks.

5.1.2.1 Classification Accuracy. The P(Good) results of the 18-class experiments are summarized in table 5.1. It can be seen that the only feature set which produced consistently high rates of classification accuracy on both tones and noise was the auto-correlations and cross-correlation. With the gaussian noise as the sound source, essentially no difference in accuracy was noted between the network where the mean FFT magnitudes and the cross-correlation were the features used, and the network where the auto-correlations and the cross-correlation were used as the features. The mean of the FFT magnitudes and the cross-correlation did not work well on tones as a feature set on tones (in terms of classification accuracy) compared with the other two feature sets. Time-samples, used as features, were less successful in classifying tones in 18 classes than in 4 classes (down to 0.675 from 0.824)(see table 4.3). The networks which used time-samples as features to localize gaussian noise sound sources had the worst performance.

The reasons for the success or lack of success of each feature set were not proven, but educated guesses can be made. As for time-samples, the ANN had some success in localizing tones, but very little in localizing noise. This result may stem from the random nature of the samples in noise, and the more predictable nature of the sinusoidal function.

The success of the FFT magnitude and the cross-correlation features in localizing noise probably results from the fact that noise has a relatively constant FFT magnitude across the spectrum. The Fourier Transform of a sinusoid would ideally be a single impulse at the frequency at which it oscillates. Because a Discrete Fourier Transform (DFT) was used, the impulse can be sharp if the frequency of the tone happens to closely correspond to one of the frequency bins in the DFT sequence. On the other hand, if the frequency of the tone is more centered between two bins in the DFT sequence, then the impulse will be of lower amplitude and will be smeared into more than one bin. The mean FFT magnitude values calculated from these sequences showed that the ratio of the values for left and right ear signals were relatively constant, but the scale of the values varied up to 25 %. The variability of these feature values as a function of the frequency accounts for the relatively low accuracy results obtained on tones.

The success of the auto-correlation and cross-correlation features on both tones and noise is a result of the fact that these functions reveal *statistical* information about the signals. This makes these features less susceptible to randomness and variabilities in the signals. In other words, this feature set is more *robust*.

Table 5.2. Summary of average P(HHN) results: 18-class experiments

Features	Time-Samples	Mean FFT Mag. & Cross-Corr.	Auto-Corr. & Cross-Corr.
Random-Freq. Tones	0.829	0.651	0.983
Gaussian Noise	0.262	1.000	1.000

5.1.2.2 High and Highest Neighbor Classification. Comparing the P(Good) results in table 5.1 and the P(HHN) results in table 5.2, it can be seen that the probability of localizing a sound source to a region within two adjacent classes is significantly higher (in all cases) than the probability of localization to the single correct class. This means that when a one-class error occurred, the correct class could often be predicted from the output node values of the network. Thus when a localization error happened, the called class was frequently in the same general direction as the correct class. This implies some error tolerance in the ANN system. When the features used in the network were the auto-correlations and the cross-correlation, nearly every test vector could be localized to an angular slice of the horizontal plane 40° wide.

Table 5.3. Summary of average P(CBCC | W1C) results: 18-class experiments

Features	Time-Samples	Mean FFT Mag. & Cross-Corr.	Auto-corr. & Cross-Corr.
Random-Freq. Tones	0.321	0.414	0.838
Gaussian Noise	0.326	0.914	0.947

5.1.2.3 Classification Error Distributions. The angular distribution of the errors in the 18-class experiments varied. In every experiment at least 30 % of the one-class errors were from sound sources located within 4° of a class boundary (see table 5.3). In general, the percentage of the total errors occurring near the boundaries increased as the classification accuracy increased (see auto-correlation and cross-correlation column in table 5.3). The one-class errors near boundaries are probably due to a slight misplacement of the boundaries in the decision region created by the neural net during training.

Another observation was that the error distribution plots tended to have a symmetrical appearance about the 180° point. This result was to be expected because of the geometric symmetry of the head and the horizontal circle (see figure 3.10).

5.2 Recommendations

1. For future research involving binaural localization using neural networks, the localization should be extended to include more than just the angle in the horizontal plane.

   Localization in three-space, including distance to the sound source should be attempted.

2. Additional research should involve the addition of noise to the signals received by each ear. The addition of noise would provide a more realistic test of the ANNs performance in "real-world" conditions.

3. Attempts should be made to localize other types of sounds in addition to single tones and gaussian noise. These may include multiple tones, clicks etc.

4. Further research should be done to investigate the use of other ANN architectures (in additıon to backpropagation) for the purpose of sound localization. One possible archite.ιure which should be considered is the "Cottrell network" (3:603).

5. Future research should look at the use of more sophisticated models of the binaural hearing process to create simulated ear signals. These models may

involve the use of filters to represent the transfer function of the human head and ear structure at each possible sound source direction.

6. Finally, real sound signals collected in a laboratory, using a manikin with microphones in the ears should be used in ANN sound localization experiments. This experiment would be a real "acid-test" for any ANN localization system.

Appendix A. *Human Auditory Model Database*

This appendix chapter contains the database which defines the human auditory model for sound sources on a circle of radius 7 ft. on the horizontal plane (see figure 2.1). The database was furnished by the Armstrong Aero-Medical Research Laboratory at Wright-Patterson AFB, OH.

The database is structured in five columns. Column one is the angle in degrees (0^o is straight ahead, subsequent values are in clockwise direction), columns 2 and 3 are the time delays (in μ seconds) of the left and right ear signals respectively, and columns 4 and 5 are the gain values of the left and right ears respectively.

The left and right ear signals can be constructed from these values by multiplying the source signal by the respective gains and by implementing the respective time delays.

```
0, 0, 0, 5.552964, 5.637564
1, 0, 0, 5.466848, 5.729240
2, 25, 0, 5.831856, 5.818728
3, 25, 0, 5.690928, 5.914016
4, 25, 0, 5.650952, 5.995988
5, 50, 0, 5.614664, 6.083680
6, 50, 0, 5.475244, 6.159912
7, 50, 0, 5.415984, 6.237700
8, 75, 0, 5.316788, 6.319736
9, 75, 0, 5.307604, 6.418592
10, 75, 0, 5.302032, 6.499112
11, 100, 0, 5.135428, 6.588016
12, 100, 0, 4.977536, 6.669932
13, 100, 0, 4.888280, 6.773584
14, 125, 0, 4.759460, 6.851312
15, 125, 0, 4.568592, 6.941240
16, 125, 0, 4.598932, 7.043492
17, 150, 0, 4.544716, 7.109396
```

```
18, 150, 0, 4.672460, 7.195512
19, 150, 0, 4.517416, 7.250484
20, 175, 0, 4.467736, 7.312544
21, 175, 0, 4.725068, 7.384020
22, 175, 0, 4.539552, 7.437540
23, 200, 0, 4.472584, 7.506164
24, 200, 0, 4.255724, 7.562764
25, 200, 0, 4.277008, 7.617720
26, 225, 0, 4.160640, 7.674004
27, 225, 0, 4.051704, 7.750872
28, 225, 0, 4.013268, 7.812868
29, 250, 0, 4.027256, 7.865592
30, 250, 0, 3.914474, 7.923224
31, 250, 0, 3.909542, 7.976716
32, 250, 0, 3.782914, 8.033032
33, 275, 0, 3.739046, 8.082424
34, 275, 0, 3.727286, 8.113996
35, 275, 0, 3.672456, 8.157300
36, 300, 0, 3.619198, 8.202340
37, 300, 0, 3.675187, 8.240452
38, 300, 0, 3.581888, 8.287060
39, 325, 0, 3.483670, 8.337300
40, 325, 0, 3.441559, 8.367740
41, 325, 0, 3.418649, 8.405936
42, 350, 0, 3.358647, 8.462968
43, 350, 0, 3.327244, 8.487164
44, 350, 0, 3.277642, 8.180708
45, 375, 0, 3.300774, 8.185448
46, 375, 0, 3.228144, 8.209256
47, 375, 0, 3.175962, 8.225552
48, 400, 0, 3.171965, 8.218484
49, 400, 0, 3.171018, 8.228576
50, 400, 0, 3.123951, 8.238840
51, 425, 0, 3.103996, 8.237784
52, 425, 0, 3.092157, 8.232688
53, 450, 0, 3.151834, 8.234384
54, 450, 0, 2.876366, 8.238460
55, 450, 0, 2.829086, 8.240080
56, 450, 0, 2.837682, 8.233744
57, 475, 0, 2.791164, 8.224380
58, 475, 0, 2.760864, 8.220552
59, 475, 0, 2.747756, 8.210196
```

```
60, 500, 0, 2.700184, 8.189724
61, 500, 0, 2.688874, 8.174752
62, 500, 0, 2.654749, 8.152092
63, 525, 0, 2.628304, 8.144632
64, 525, 0, 2.611488, 8.127852
65, 525, 0, 2.598259, 8.105776
66, 525, 0, 2.581795, 8.096836
67, 550, 0, 2.576359, 8.085232
68, 550, 0, 2.557136, 8.066396
69, 550, 0, 2.541984, 8.060728
70, 575, 0, 2.519886, 8.044200
71, 575, 0, 2.535217, 8.031860
72, 575, 0, 2.523599, 8.025428
73, 600, 0, 2.537045, 8.009712
74, 600, 0, 2.540524, 8.009452
75, 600, 0, 2.554521, 7.997960
76, 625, 0, 2.592806, 7.997792
77, 625, 0, 2.645846, 7.982292
78, 625, 0, 2.651768, 7.982320
79, 625, 0, 2.696706, 7.859252
80, 650, 0, 2.728064, 7.839132
81, 650, 0, 2.733716, 7.815660
82, 650, 0, 2.745462, 7.802444
83, 650, 0, 2.792922, 7.769484
84, 650, 0, 2.813598, 7.748712
85, 675, 0, 2.836232, 7.731272
86, 675, 0, 2.842786, 7.701472
87, 675, 0, 2.875874, 7.639936
88, 675, 0, 2.901025, 7.617880
89, 675, 0, 2.887189, 7.583016
90, 675, 0, 2.886130, 7.571488
91, 675, 0, 2.904686, 7.532060
92, 700, 0, 2.898504, 7.503752
93, 700, 0, 2.881643, 7.470800
94, 700, 0, 2.863865, 7.440328
95, 700, 0, 2.870499, 7.419972
96, 700, 0, 2.806318, 7.380544
97, 700, 0, 2.785066, 7.344288
98, 700, 0, 2.751054, 7.305492
99, 675, 0, 2.724502, 7.265724
100, 675, 0, 2.722563, 7.235092
101, 675, 0, 2.651449, 7.207468
```

```
102, 650, 0, 2.637735, 7.160308
103, 625, 0, 2.566055, 7.128776
104, 625, 0, 2.535838, 7.086480
105, 600, 0, 2.480426, 7.058700
106, 600, 0, 2.459834, 7.009072
107, 600, 0, 2.477028, 6.978492
108, 575, 0, 2.345260, 6.944556
109, 575, 0, 2.334450, 6.907284
110, 575, 0, 2.361668, 6.871016
111, 575, 0, 2.288361, 6.823744
112, 550, 0, 2.266786, 6.793352
113, 550, 0, 2.246805, 6.752948
114, 550, 0, 2.235258, 6.736640
115, 550, 0, 2.251026, 6.705084
116, 525, 0, 2.231540, 6.662936
117, 525, 0, 2.240995, 6.382380
118, 525, 0, 2.396150, 6.343304
119, 525, 0, 2.406093, 6.295648
120, 525, 0, 2.404210, 6.238772
121, 500, 0, 2.433176, 6.196632
122, 500, 0, 2.413351, 6.152568
123, 500, 0, 2.415896, 6.122280
124, 475, 0, 2.410379, 6.082920
125, 475, 0, 2.409891, 6.021268
126, 475, 0, 2.419057, 5.991648
127, 450, 0, 2.435342, 5.957396
128, 450, 0, 2.461739, 5.905224
129, 450, 0, 2.493918, 5.861300
130, 450, 0, 2.528020, 5.809160
131, 425, 0, 2.532243, 5.770604
132, 425, 0, 2.567832, 5.711836
133, 425, 0, 2.587606, 5.673896
134, 400, 0, 2.598868, 5.643912
135, 400, 0, 2.621326, 5.600536
136, 400, 0, 2.643590, 5.547456
137, 375, 0, 2.662481, 5.513992
138, 375, 0, 2.697966, 5.476204
139, 375, 0, 2.720183, 5.440868
140, 350, 0, 2.742795, 5.409516
141, 350, 0, 2.654250, 5.344908
142, 350, 0, 2.693499, 5.296588
143, 325, 0, 2.764798, 5.264776
```

```
144, 325, 0, 2.770946, 5.222300
145, 325, 0, 2.806067, 5.202400
146, 300, 0, 2.829093, 5.165864
147, 300, 0, 2.872999, 5.131288
148, 300, 0, 2.890527, 5.097712
149, 275, 0, 2.926472, 5.073180
150, 275, 0, 2.964096, 5.043496
151, 275, 0, 3.008576, 4.981708
152, 250, 0, 3.055311, 4.954900
153, 250, 0, 3.088986, 4.910120
154, 250, 0, 3.141163, 4.900752
155, 225, 0, 3.182173, 4.834936
156, 225, 0, 3.224582, 4.808848
157, 225, 0, 3.274808, 4.768068
158, 200, 0, 3.337780, 4.742540
159, 200, 0, 3.357604, 4.706736
160, 175, 0, 3.395194, 4.689272
161, 175, 0, 3.404588, 4.662820
162, 175, 0, 3.444669, 4.639144
163, 150, 0, 3.430962, 4.609744
164, 150, 0, 3.463110, 4.588956
165, 150, 0, 3.475979, 4.554544
166, 125, 0, 3.537652, 4.528664
167, 125, 0, 3.520124, 4.517756
168, 100, 0, 3.564934, 4.494520
169, 100, 0, 3.605374, 4.463604
170, 100, 0, 3.632558, 4.425644
171, 75, 0, 3.690662, 4.401452
172, 75, 0, 3.694052, 4.369636
173, 75, 0, 3.754096, 4.351896
174, 50, 0, 3.763649, 4.311844
175, 50, 0, 3.768475, 4.279004
176, 50, 0, 3.995740, 4.231996
177, 25, 0, 4.035560, 4.187368
178, 25, 0, 4.070468, 4.132200
179, 0, 0, 4.092256, 4.119872
180, 0, 0, 4.119872, 4.092256
181, 0, 0, 4.132200, 4.070468
182, 0, 25, 4.187368, 4.035560
183, 0, 25, 4.231996, 3.995740
184, 0, 50, 4.279004, 3.768475
185, 0, 50, 4.311844, 3.763649
```

```
186, 0, 50, 4.351896, 3.754096
187, 0, 75, 4.369636, 3.694052
188, 0, 75, 4.401452, 3.690662
189, 0, 75, 4.425644, 3.632558
190, 0, 100, 4.463604, 3.605374
191, 0, 100, 4.494520, 3.564934
192, 0, 100, 4.517756, 3.520124
193, 0, 125, 4.528664, 3.537652
194, 0, 125, 4.554544, 3.475979
195, 0, 150, 4.588956, 3.463110
196, 0, 150, 4.609744, 3.430962
197, 0, 150, 4.639144, 3.444669
198, 0, 175, 4.662820, 3.404588
199, 0, 175, 4.689272, 3.395194
200, 0, 175, 4.706736, 3.357604
201, 0, 200, 4.742540, 3.337780
202, 0, 200, 4.768068, 3.274808
203, 0, 225, 4.808848, 3.224582
204, 0, 225, 4.834936, 3.182173
205, 0, 225, 4.900752, 3.141163
206, 0, 250, 4.910120, 3.088986
207, 0, 250, 4.954900, 3.055311
208, 0, 250, 4.981708, 3.008576
209, 0, 275, 5.043496, 2.964096
210, 0, 275, 5.073180, 2.926472
211, 0, 275, 5.097712, 2.890527
212, 0, 300, 5.131288, 2.872999
213, 0, 300, 5.165864, 2.829093
214, 0, 300, 5.202400, 2.806067
215, 0, 325, 5.222300, 2.770946
216, 0, 325, 5.264776, 2.764798
217, 0, 325, 5.296588, 2.693499
218, 0, 350, 5.344908, 2.654250
219, 0, 350, 5.409516, 2.742795
220, 0, 350, 5.440868, 2.720183
221, 0, 375, 5.476204, 2.697966
222, 0, 375, 5.513992, 2.662481
223, 0, 375, 5.547456, 2.643590
224, 0, 400, 5.600536, 2.621326
225, 0, 400, 5.643912, 2.598868
226, 0, 400, 5.673896, 2.587606
227, 0, 425, 5.711836, 2.567832
```

```
228, 0, 425, 5.770604, 2.532243
229, 0, 425, 5.809160, 2.528020
230, 0, 450, 5.861300, 2.493918
231, 0, 450, 5.905224, 2.461739
232, 0, 450, 5.957396, 2.435342
233, 0, 450, 5.991648, 2.419057
234, 0, 475, 6.021268, 2.409891
235, 0, 475, 6.082920, 2.410379
236, 0, 475, 6.122280, 2.415896
237, 0, 500, 6.152568, 2.413351
238, 0, 500, 6.196632, 2.433176
239, 0, 500, 6.238772, 2.404210
240, 0, 525, 6.295648, 2.406093
241, 0, 525, 6.343304, 2.396150
242, 0, 525, 6.382380, 2.240995
243, 0, 525, 6.662936, 2.231540
244, 0, 525, 6.705084, 2.251026
245, 0, 550, 6.736640, 2.235258
246, 0, 550, 6.752948, 2.246805
247, 0, 550, 6.793352, 2.266786
248, 0, 550, 6.823744, 2.288361
249, 0, 575, 6.871016, 2.361668
250, 0, 575, 6.907284, 2.334450
251, 0, 575, 6.944556, 2.345260
252, 0, 575, 6.978492, 2.477028
253, 0, 600, 7.009072, 2.459834
254, 0, 600, 7.058700, 2.480426
255, 0, 600, 7.086480, 2.535838
256, 0, 625, 7.128776, 2.566055
257, 0, 625, 7.160308, 2.637735
258, 0, 650, 7.207468, 2.651449
259, 0, 675, 7.235092, 2.722563
260, 0, 675, 7.265724, 2.724502
261, 0, 675, 7.305492, 2.751054
262, 0, 700, 7.344288, 2.785066
263, 0, 700, 7.380544, 2.806318
264, 0, 700, 7.419972, 2.870499
265, 0, 700, 7.440328, 2.863865
266, 0, 700, 7.470800, 2.881643
267, 0, 700, 7.503752, 2.898504
268, 0, 700, 7.532060, 2.904686
269, 0, 675, 7.571488, 2.886130
```

```
270, 0, 675, 7.583016, 2.887189
271, 0, 675, 7.617880, 2.901025
272, 0, 675, 7.639936, 2.875874
273, 0, 675, 7.701472, 2.842786
274, 0, 675, 7.731272, 2.836232
275, 0, 675, 7.748712, 2.813598
276, 0, 650, 7.769484, 2.792922
277, 0, 650, 7.802444, 2.745462
278, 0, 650, 7.815660, 2.733716
279, 0, 650, 7.839132, 2.728064
280, 0, 650, 7.859252, 2.696706
281, 0, 625, 7.982320, 2.651768
282, 0, 625, 7.982292, 2.645846
283, 0, 625, 7.997792, 2.592806
284, 0, 625, 7.997960, 2.554521
285, 0, 600, 8.009452, 2.540524
286, 0, 600, 8.009712, 2.537045
287, 0, 600, 8.025428, 2.523599
288, 0, 575, 8.031860, 2.535217
289, 0, 575, 8.044200, 2.519886
290, 0, 575, 8.060728, 2.541984
291, 0, 550, 8.066396, 2.557136
292, 0, 550, 8.085232, 2.576359
293, 0, 550, 8.096836, 2.581795
294, 0, 525, 8.105776, 2.598259
295, 0, 525, 8.127852, 2.611488
296, 0, 525, 8.144632, 2.628304
297, 0, 525, 8.152092, 2.654749
298, 0, 500, 8.174752, 2.688874
299, 0, 500, 8.189724, 2.700184
300, 0, 500, 8.210196, 2.747756
301, 0, 475, 8.220552, 2.760864
302, 0, 475, 8.224380, 2.791164
303, 0, 475, 8.233744, 2.837682
304, 0, 450, 8.240080, 2.829086
305, 0, 450, 8.238460, 2.876366
306, 0, 450, 8.234384, 3.151834
307, 0, 450, 8.232688, 3.092157
308, 0, 425, 8.237784, 3.103996
309, 0, 425, 8.238840, 3.123951
310, 0, 400, 8.228576, 3.171018
311, 0, 400, 8.218484, 3.171965
```

```
312, 0, 400, 8.225552, 3.175962
313, 0, 375, 8.209256, 3.228144
314, 0, 375, 8.185448, 3.300774
315, 0, 375, 8.180708, 3.277642
316, 0, 350, 8.487164, 3.327244
317, 0, 350, 8.462968, 3.358647
318, 0, 350, 8.405936, 3.418649
319, 0, 325, 8.367740, 3.441559
320, 0, 325, 8.337300, 3.483670
321, 0, 325, 8.287060, 3.581888
322, 0, 300, 8.240452, 3.675187
323, 0, 300, 8.202340, 3.619198
324, 0, 300, 8.157300, 3.672456
325, 0, 275, 8.113996, 3.727286
326, 0, 275, 8.082424, 3.739046
327, 0, 275, 8.033032, 3.782914
328, 0, 250, 7.976716, 3.909542
329, 0, 250, 7.923224, 3.914474
330, 0, 250, 7.865592, 4.027256
331, 0, 250, 7.812868, 4.013268
332, 0, 225, 7.750872, 4.051704
333, 0, 225, 7.674004, 4.160640
334, 0, 225, 7.617720, 4.277008
335, 0, 200, 7.562764, 4.255724
336, 0, 200, 7.506164, 4.472584
337, 0, 200, 7.437540, 4.539552
338, 0, 175, 7.384020, 4.725068
339, 0, 175, 7.312544, 4.467736
340, 0, 175, 7.250484, 4.517416
341, 0, 150, 7.195512, 4.672460
342, 0, 150, 7.109396, 4.544716
343, 0, 150, 7.043492, 4.598932
344, 0, 125, 6.941240, 4.568592
345, 0, 125, 6.851312, 4.759460
346, 0, 125, 6.773584, 4.888280
347, 0, 100, 6.669932, 4.977536
348, 0, 100, 6.588016, 5.135428
349, 0, 100, 6.499112, 5.302032
350, 0, 75, 6.418592, 5.307604
351, 0, 75, 6.319736, 5.316788
352, 0, 75, 6.237700, 5.415984
353, 0, 50, 6.159912, 5.475244
```

```
354, 0, 50, 6.083680, 5.614664
355, 0, 50, 5.995988, 5.650952
356, 0, 25, 5.914016, 5.690928
357, 0, 25, 5.818728, 5.831856
358, 0, 25, 5.729240, 5.466848
359, 0, 0, 5.637564, 5.552964
```

Bibliography

1. Anderson, T. R., Acoustics Researcher. Personal Communication. AL/CFBA, Bioacoustics and Biocommunication Branch, Armstrong Laboratory, Wright-Patterson AFB OH 45433-6573, 20 December 1990.

2. Au, W. W. L. "Echolocation in Dolphins." In Berkely, M. and W. Stebbins, editors, *Comparative Perception*, New York: Wiley, 1989.

3. Barga, R. S. and others. "Source Location of Acoustic Emissions from Atmospheric Leakage Using Neural Networks." In *SPIE Proceedings: Applications of Artificial Neural Networks II*, Volume 1469, pages 602–611, 1991.

4. Batteau, D. W. "The Role of the Pinna in Human Localization." In *Proceedings of the Royal Society, London, Series B*, Volume 165, pages 158–180, 1967.

5. Blauert, J. "Sound Localization in the Median Plane," *Acustica*, *22*:205–213 (1969).

6. Brennan, C. and L. Chen. *Preliminary to a Neural Network Model of Sonar-Based Target Discrimination in the Echolocating Bat.* Technical Report ONR-88-2, Providence RI: Brown University, May 1988. Research performed for the Office of Naval Research (AD-A205681).

7. Cherry, C. "Two Ears But One World." In Rosenblith, W. A., editor, *Sensory Communication*, Cambridge, Mass: MIT Press, 1961.

8. Fechner, G. T. *Elemente der Psychophysics.* Leipzig: Breitkopk und Hartel, 1860.

9. Flemming, M. K. and G. W. Cottrell. "Categorization of Faces Using Unsupervised Feature Extraction," *IEEE International Joint Conference on Neural Networks*, pages 65–70 (1990).

10. Gorman, R. P. and T. J. Sejnowski. "Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Signals," *Neural Networks*, *1*:75–89 (1988).

11. Grabec, I. and W. Sachse. "Application of an Intelligent Signal Processing System to Acoustic Emission Analysis," *Journal of the Acoustical Society of America*, *85*:1226–1235 (March 1989).

12. Lambert, R. M. "Dynamic Theory of Sound-Source Localization," *Journal of the Acoustical Society of America*, *56*:165–171 (1974).

13. Lippmann, R. P. "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, pages 4–22 (April 1987).

14. Makous, J. C. and J. C. Middlebrooks. "Two-Dimensional Sound Localization by Human Listeners," *Journal of the Acoustical Society of America*, *87*:2188–2200 (May 1990).

15. McKinley, R. L. *Concept and Design of an Auditory Localization Cue Synthesizer*. MS thesis, AFIT/GE/ENG/88D-29, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1988.

16. Musicant, A. D. and others. "The Influence of Pinnae-based Spectral Cues on Sound Localization," *Journal of the Acoustical Society of America*, *75*:1195–1200 (1984).

17. Rogers, S. K. and others. *An Introduction to Biological and Artificial Neural Networks*. Wright-Patterson AFB OH: Air Force Institute of Technology, 1990.

18. Roitblat, H. L. and P. W. B. Moore. *Dolphin Echolocation: Identification of Returning Echos Using a Counterpropagation Network*. Technical Report DN308-262, Kailua HI: Naval Ocean Systems Center, Hawaii Laboratory, August 1989. (AD-A211805).

19. Sandel, T. T. and others. "Localization of Sound from Single and Paired Sources," *Journal of the Acoustical Society of America*, *27*:842–852 (September 1955).

20. Searle, C. L. and others. "Binaural Pinna Disparity: Another Auditory Localization Cue," *Journal of the Acoustical Society of America*, *57*:448–455 (February 1975).

21. Shamma, S. A. "Stereausis: Binaural Processing Without Neural Delays," *Journal of the Acoustical Society of America*, *86*:989–1006 (September 1989).

22. Shanmugan, K. S. and A. M. Breipohl. *Random Signals: Detection, Estimation and Data Analysis*. New York: Wiley, 1988.

23. Sivian, L. J. and S. D. White. "On Minimum Audible Fields," *Journal of the Acoustical Society of America*, *4*:288–321 (1933).

24. Suga, N. "Biosonar and Neural Computation in Bats," *Scientific American*, pages 60–68 (June 1990).

25. Tarr, G. L. *Dynamic Analysis of Feedforward Neural Networks Using Simulated and Measured Data*. MS thesis, AFIT/GE/ENG/88D-54, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1988.

26. Wallach, H. "The Role of Head Movements and Vestibular and Visual Cues in Sound Localization," *Journal of Experimental Psychology*, *27*:239–368 (1940).

27. Woodworth, R. S. *Experimental Psychology*. New York: Holt, 1938.

28. Wright, D. and others. "Pinna Reflections as Cues for Localization," *Journal of the Acoustical Society of America*, *56*:957–962 (September 1974).

Fxd. from [illegible]

Abstract

The purpose of this study was to investigate the use of Artificial Neural Networks to localize sound sources from simulated, human binaural signals. Only sound sources originating from a circle on the horizontal plane were considered. Experiments were performed to examine the ability of the networks to localize using three different feature sets. The feature sets used were: time-samples of the signals, mean FFT magnitude and cross-correlation data, and auto-correlation and cross-correlation data. The two different types of sound source signals considered were tones and gaussian noise.

The feature set which yielded the best results in terms of classification accuracy (over 91 %) for both tones and noise was the auto-correlation and cross-correlation data. These results were achieved using 18 classes (20° per class). The other two feature sets did not produce accuracy results as high or as consistent between the two signal types.

When using time-samples of the signals as features it was observed that in order to accurately classify tones of random-frequency, it was necessary to train with random-frequency tones rather than with tones of one, or a few discrete frequencies.

December 1991 Master's Thesis

Binaural Sound Localization Using Neural Networks

Rushby C. Craig

Air Force Institute of Technology, WPAFB OH 45433-6583

AFIT/GE/ENG/91D-13

Tim Anderson
AL/CFBA
Wright-Patterson AFB OH 45433

Approved for public release; distribution unlimited

The purpose of this study was to investigate the use of Artificial Neural Networks to localize sound sources from simulated, human binaural signals. Only sound sources originating from a circle on the horizontal plane were considered. Experiments were performed to examine the ability of the networks to localize using three different feature sets. The feature sets used were: time-samples of the signals, mean FFT magnitude and cross-correlation data, and auto-correlation and cross-correlation data. The two different types of sound source signals considered were tones and gaussian noise.

The feature set which yielded the best results in terms of classification accuracy (over 91 %) for both tones and noise was the auto-correlation and cross-correlation data. These results were achieved using 18 classes (20^o per class). The other two feature sets did not produce accuracy results as high or as consistent between the two signal types.

When using time-samples of the signals as features it was observed that in order to accurately classify tones of random-frequency, it was necessary to train with random-frequency tones rather than with tones of one, or a few discrete frequencies.

Neural Nets, Bioacoustics, Acoustic Signals, Sound Signals 119

Unclassified Unclassified Unclassified UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to **stay within the lines to meet optical scanning requirements.**

Block 1. Agency Use Only (Leave Blank)

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Names(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency. Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of ..., To be published in .... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availablity Statement. Denote public availability or limitation. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR)

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - DOD - Leave blank
DOE - DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports
NASA - NASA - Leave blank
NTIS - NTIS - Leave blank.

Block 13. Abstract. Include a brief (Maximum 200 words) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (NTIS only).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.