Passive Sonar Recognition and Analysis Using Hybrid Neural Networks

B.P. Howell and S. Wood
Department of Marine and Environmental Systems
bhowell@fit.edu,swood@fit.edu

S. Koksal
Department of Math
skoksal@fit.edu

Florida Institute of Technology
150 W. University Blvd
Melbourne, FL 32901, USA

Abstract- The detection, classification, and recognition of underwater acoustic features have always been of the highest importance for scientific, fisheries, and defense interests. Recent efforts in improved passive sonar techniques have only emphasized this interest. In this paper, the authors describe the use of novel, hybrid neural approaches using both unsupervised and supervised network topologies. Results are presented which demonstrate the ability of the network to classify biological, man made, and geological sources. Also included are capabilities of the networks to attack the more difficult problems of identifying the complex vocalizations of several fish and marine mammalian species. Basic structure, processor requirements, training and operational methodologies are described as well as application to autonomous observation and vehicle platforms.

I. Introduction

As researchers attempt to gain more understanding from the ocean environment, limitations on existing methods of exploration have become apparent. One such case is in the acoustic measurement area. In the case of man made objects, active sonar has been traditionally used as well as human based passive sonar methods. In fisheries work, the active returns from schools of fish have also been used to count fish populations, but with limited success in species identification.

In geophysical work, earthquakes and other undersea phenomenon are recorded, but to process the data autonomously has been difficult. Other applications, such as cetacean monitoring and reef monitoring have been problematic, particularly in noisy environments.

Neural networks are a family of structures and methods which provide tools to work with highly nonlinear systems. In such areas as speech recognition, acoustic object recognition, and image processing, neural networks have been successfully applied to solve problems [1, 2, 3]

As early as 1990, researchers from several areas have recognized the potential for neural networks to work in both active sonar and passive acoustic areas. Yet, in these research areas, the emphasis has either been to classify active target returns, or look only at very limited data sets [4]. In addition methods examined have in general been inappropriate for autonomous systems or for embedded applications. However, in work by Lin in 1998, a study was performed on 3 fish species using unsupervised learning which demonstrated the potential for neural networks to attack this problem at least in the area of fish identification [5].

It is the focus of this project to develop a uniform approach to all classes of passive sounds in the marine environment. First, a survey of available sounds was undertaken. Although by no means comprehensive, a sound library was developed from various sources on the Internet and from the University of Rhode Island [6]. These sounds demonstrated the breadth of recording methods, bandwidths, and sound behavior to be encountered.

After examining the sounds, the mission requirements were developed. The desire here has been to develop a generic sound identification and classification system suitable for buoy deployment or vehicle deployment and whose response can trigger both data logging functions as well as navigational or operational cues.

Because of the need for limited supervision, the neural network paradigm to be selected had to entail unsupervised learning or, at least, limited supervised learning during only the initialization process. Also, the system had to be able to adapt to new information and to identify unknown sources and record them for future learning or to report to system operators.

Based on these requirements, self-organizing Kohonen maps (SOM) were selected for initial study. These networks are fast converging, unsupervised, and maintain the topology of the incoming information. Later, multi-level perceptrons (MLP) were added for final data analysis using existing exemplar data. In cases where the sound source is extremely complicated or episodic, such as with whales and dolphins, the MLP layers force recognition to the same source.

II. Source Environment

A: Sound Resources

To determine if the neural network approach would be viable, it was necessary to determine the characteristics of the source signals encountered. The mission planning and hardware also determine the bandwidth to be observed and characterized.

Previously recorded sound sources were acquired through various Internet and physical sources as mentioned. These were received and stored in a variety of formats, specifically as digitized sound in MP3, AU, and WAV format. MP3 and AU files are a compressed format while WAV is not. After reviewing these data sources, it was further discovered that many of these recordings were oversampled when compared with the original bandwidth. For

Funding was provided by the Link Foundation Fellowship program and the Marine Field Project Program of Florida Institute of Technology.

example, many of the URI files were analog recordings which had been filtered to <750 Hz, yet were stored as 44khz sampled streams. Table 1 summarizes the sources and the original data format.

Having completed this review, it was also necessary to examine the means by which future Fla. Tech field data would be recorded. Since the desire was to eventually make this low cost as well as autonomous vehicle deployable, conventional PC sound cards in a PC/104 form factor were selected for the initial study. These cards can sample in stereo to 44 khz sampling rate, thus providing a 22khz bandwidth. For non autonomous work, the fact that DAT recorders are on the decline, forced a decision to use new, hard drive style recorders. After review, it was discovered that only one vendor, Creative Labs, produced a data recorder, the NOMAD 3 system, which could record and store in an uncompressed (WAV) format.

Table 1
Sources of Sound Recordings

Agency	Program	Data Type	Sample Rate	Status
NOAA	VENTS	Ship Ship at Hphone Quake	8khz/16bit speedup 10x mono WAV	Conv. 8kbps 16kbps
Arrtec	DolphinEar	Boat Humpback Whale Dolphin Boat	88-176kbps WAV	Conv. 8kbps 16kbps
Cetacean Research	1997,1998 survey	Orcas Bivalves Boat	MP3 32kbps	Conv. 8kbps 16kbps
Acoustical Society of America	Demo.	Humpback	.AU 64kbps	Conv. 8kbps 16kbps
Submarine Project	Steel in the Deep	Sub Sounds C-Launch	Mp3 64kbps	Conv. 8kbps 16kbps
SOEST Hawaii	HUGO	Volcanic Submersible	.AIFF	Not Conv. yet

Since the emphasis of this work was on the neural network processing as opposed to high frequency recognition, it was decided to limit the bandwidth for field recordings to 24khz (48 ksps), the maximum allowed by the NOMAD. Any additional bandwidth reduction would be performed just prior to application of the neural network. In addition, the sound source files would be resampled to the lowest data rate of the set. Thus, the data sets were all made uniform to 8ksps and 16ksps data rates, 16 bit PCM encoding, monaural.

B. Spectral Characteristics

The nature of the sounds encountered is variable depending on the sound source. Man-made sources such as vehicles comprise a class of sounds characterized by distinct frequency content and continuous output during an event as in Figure 1. Geological activity involves components down into the infrasound range. Cetacean sounds are highly complex as in Figure 3, with a rich variety of frequency content and time domain variation. Fish vertebrate sound is highly episodic, though within each pulse, the spectral content is similar and unique for that species as can be seen in Fig 2. Lastly, invertebrate noise, rainfall, and surf noise provide broadband background noise on a semi-continuous basis over a sample episode.

The question of whether band limiting reduces the information in the signal is one which needs to be addressed. From general studies of underwater sounds, the attenuation of sound increases with frequency [7]. This low attenuation at low frequencies has been used to advantage for long range studies such as SOSUS [8]. In general, much of the information content of the acoustic signal is below 3 Khz. For example Figure 1 and 2 show spectrograms for a diesel submarine and fish sound. As can be seen from these spectrograms, the choice of examining only the lower frequencies should not adversely impact the recognition process. Note also that the episodic behavior of the fish is on a different time scale (1 sec.) than that of the submarine or the whale. It was observed that the longer time intervals were required for fish samples than other sound sources.

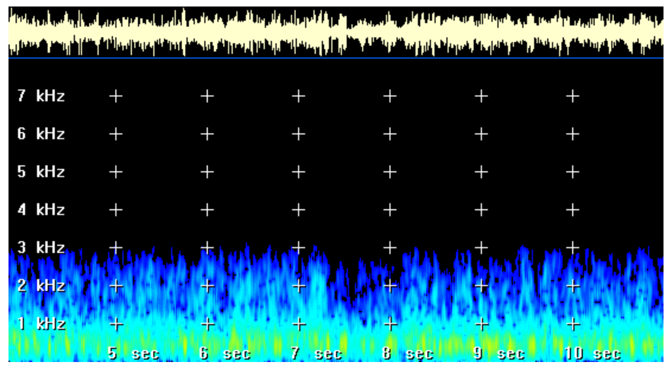

Fig 1. Spectrogram of Submarine Signal
(16ksps, 16 bit,1024 pt FFT)

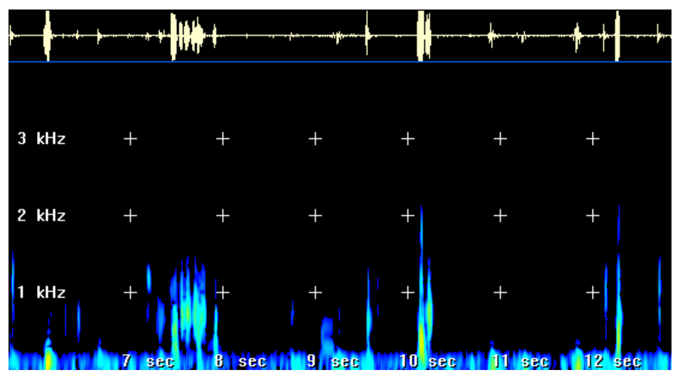

Fig 2. Spectrogram of Sea Catfish
(16ksps, 16 bit,1024 pt FFT)

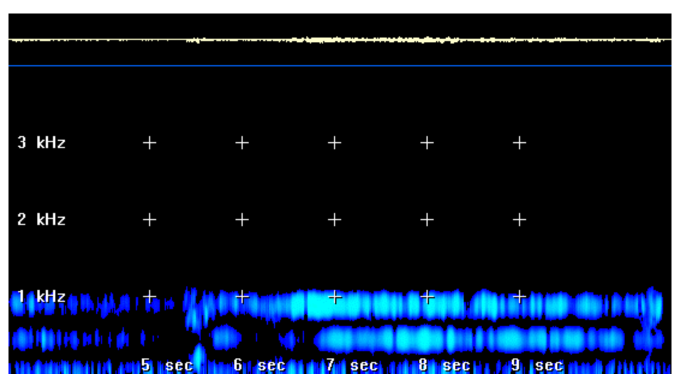

Fig 3. Spectrogram of Whale
(16 ksps, 16 bit, 1024 pt FFT)

C. Creation of Data Sets

The range of the data allowed a wide variety of possible training sets to be applied to the network problem. First, the networks needed to be tested to see if they could separate biological, man-made, and geophysical sound sources. To this end, a data set for training and testing was created from several sources in each category. During this activity, certain data types produced effects requiring further analysis for the neural networks. To evaluate these data types, an additional data set was created. These datasets will be referred to as dataset 1 and 2 (DS1, DS2). For the remainder of this initial study, the network structure and functionality will be evaluated against these two standard datasets. It should be noted that most of the data from PMEL was stored at 10X normal speed and was resampled back to 1X speed and 16ksps before training.

Table 2
Data Set 1 for Network Evaluation

Sound	Source
Tarpon	URI
Boat	NOAA/PMEL
Boat Close to Hydrophone	NOAA/PMEL
ORCA	Cetacean Research
Quake	NOAA/PMEL
Diesel Sub	Sub Project
Tremor	NOAA/PMEL
Catfish	URI
Blue Whale	NOAA/PMEL
Submerged Cold Launch	Sub Project
Grunt	URI
Humpback whale(Long Cry)	ASA

Table 3
Data Set 2 for Network Evaluation

Sound	Source
Humpback(Haunting Cry)	ASA
Boat	NOAA/PMEL
North Pacific Blue whale	NOAA/PMEL
ORCA	Cetacean Research
Quake	NOAA/PMEL
Diesel Sub	Sub Project
Tremor	NOAA/PMEL
Humpback Whale(Whistle)	ASA
Blue Whale	NOAA/PMEL
C-Launch	Sub Project
Grunt	URI
Humpback whale(Long Cry)	ASA

III. Neural Nets

A. Neural Net Introduction

As can be seen from the spectrograms, the variety of the incoming information is such that conventional modeling methods are challenged. The signals received from hydrophone data are episodic in nature and vary considerably from sample to sample. Even conventional stochastic methods cannot adequately describe the significant features of the sound sources.

As early as 1991, researchers in SONAR recognized that these neural networks had potential for analyzing underwater acoustic data [5,9]. The majority of this work utilized the multi-level perceptron neural paradigm. Work also has been done in speech processing using self organizing maps (SOM) neural networks with great success[9]. Later, self organizing maps (SOM) were examined on limited data sets of fish sounds with success by Lin [6]. In this work, the sounds from three different fish species were analyzed and sorted using the SOM, then used to locate feeding areas in the Indian River Lagoon in Florida.

In general, the neural network is a model of biological neuron systems. The model of a basic neuron is shown in Fig 2. It consists of several inputs tied to a summing junction by various weights. The result of this operation for input vector x(i) i=1:n inputs and w(i) for the weights [w(i)] connecting the ith input to the neuron is:

$$y=\mathrm{sum}(x(i)*w(i)) \qquad \text{for } i = 1 \text{ to } n \text{ inputs} \qquad (1.1)$$

After the weighted inputs are summed, the signal may be passed through a non-linear element, usually modeled on the TANH function or SIGMOID operator. This allows the neuron to behave in a nonlinear manner.

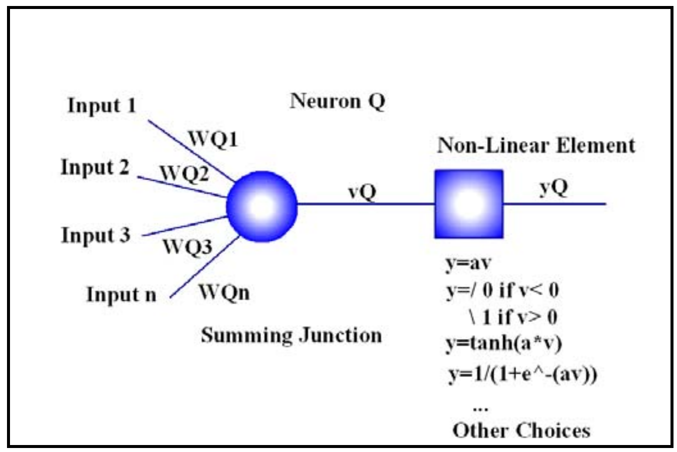

Fig. 2. Single Neuron Model

The power of the neural network approach comes in connecting individual neurons together. The way in which the neurons are connected combined with the training method used defines the paradigm of the neural network. A typical multi-level perceptron is shown in Figure 3.

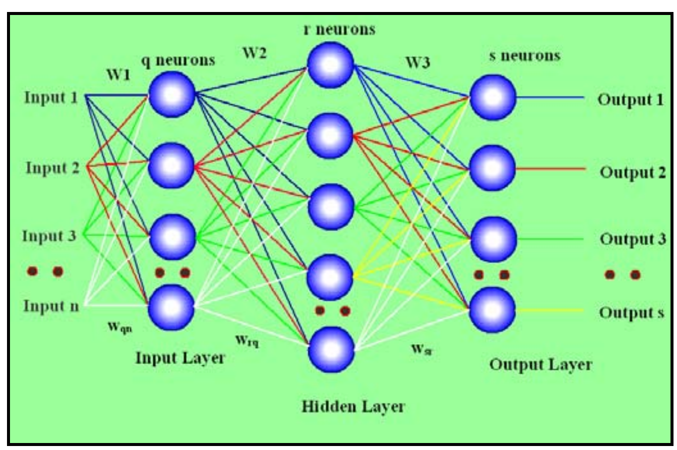

Fig. 3. Multi Level Perceptron

In addition to this highly parallel processing, the problem is solved via iterative training of the network to minimize some defined error signal or cost function. In this way, the network learns the problem "space" as well as cross correlations between data connections. This truly defines one of the key differences between neural network methods and traditional methods. In traditional methods, the information space must either be modeled by deterministic equations or stochastic probability mappings. Indeed, many neural networks are designed to emulate the latter. However, both of these traditional methods imply that the information space can be completely known or known within certain limitations. In the case of the neural network, the constraint is limited to the condition that the statistical information need be wide-sense stationary during the course of the sample, but that the information space need NOT be known entirely a priori. This condition is easily met in the ocean environment. In the context of this work, it implies that the sound sources must be only slowly changing over the course of the measurement or not at all. In other words, a whale needs to sound like a whale. With the fact that the network learning can be updated as additional information is made available, the network can indeed adapt to statistical changes in the sound field sources.

Information is stored throughout the network as weight values. During iterative training, the weights of the network are changed through a variety of methods until some termination condition is met. At that point, the weights are held constant and the network can then be used to solve the problem it was designed and trained for.

If additional information is made available, the network weights are unfrozen and the network can begin to learn again.

B. Training Methods

Neural networks can be divided into two basic categories based on the training method: *supervised* and *unsupervised*. Most neural networks are based on *supervised* learning where during a training cycle the output of the neural network is compared to some desired exemplar. The difference between the two responses is measured and the weights are adjusted accordingly. After several iterations the minimum error is reached. The multi-level perceptron (MLP) when trained via backpropagation is such a network paradigm.

In unsupervised learning, the network uses no exemplar, but instead patterns it's weight matrix to the features of the incoming space. In this manner, the network remembers the topology of the space and can then classify new inputs based on the remembered space. Networks like SOM are usually unsupervised networks, and have the drawback, that the space classification must be interpreted by a user as it will not always learn the same way. This is due to the randomness needed in the weight initialization.

For example, let an SOM network map several bit patterns into one of three categories as in Fig. 4. On each run, the net will correctly separate the bit patterns into the three categories. But on each run, the number of that category may be different.

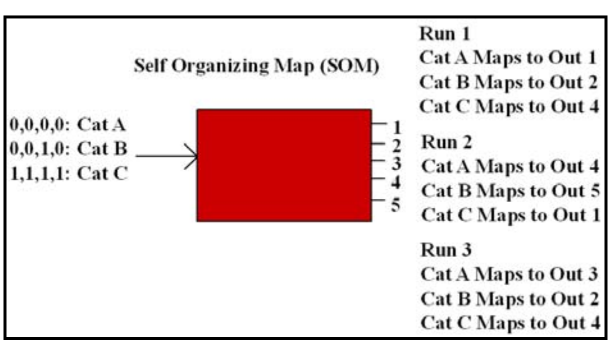

Fig 4. Mapping Variability

This ambiguity can be resolved with a variety of methods, one of which is a focus of this paper.

C. Self Organizing Map (SOM)

As mentioned above, the goal of this work is to develop strategies for autonomous or semiautonomous analysis of incoming acoustic signals. Since this implies no supervision, it made sense to begin this work with an unsupervised training network paradigm. SOM was selected in tis most basic form, the Kohonen map as the initial choice for analysis.

The Kohonen map simply is an array of weighted connections between an input plane and an output plane. As seen in Fig. 5, the network consists of an input array, where the information to be processed is stored, a weight matrix, whose weights are to be modified during operation, and an output array which describes the mapping taking place. In addition, it should be noted that the SOM has no non-linear element. It is therefore a linear map of the input data to some output space. In the case of our work, it was desired to compress the data from a two dimensional information stream to a one dimensional category output. In some literature, this network would be described as a learning vector quantizer (LVQ) network, however in the case of this work, the network is trained by competitive learning, not by supervised learning. The output categories are interpreted by either the mapping supervisory algorithm, or a secondary network. LVQ has been demonstrated to work well for fish sounds as mentioned in Lin's work [5].

The SOM is trained via a method known as competitive learning. In this method, an input is applied to the network and the outputs are examined. The output with the greatest response is "rewarded" by having it's weights strengthened to the inputs. From here, there are several variations possible as to additional weight updating. These include leaving other weights alone, reducing connection strength, etc. For a better understanding of these issues, several excellent references are available [10, 11, 12]. For this work, it was decided to try the simplest case, that of no change to losing weights, and reward only the element and its nearest neighbors equally for the first 75 iterations. After this, only the winning neuron is strengthened in it's connections.

In operation, a signal is applied to the SOM and one or more outputs is stimulated, thus indicating what the sound is "like". Thus the SOM maps sounds into various classes based on the nature of the sound. Instead of being a one to one mapping, it would be more like a person describing the type of sound heard. For example, a boat could be described as a "Swish" type sound, a parrot fish as "Scrape" type sound, a blue whale a "Low Reverberation" type of sound. In this manner, the network groups these sounds by type. Sufficient size is given to the network to encompass the possible number of types of sounds. If the network is made to big, the unused categories have their weights reduced to zero during training. If the network is made with too few categories, the network attempts to fit the data to the allowed categories, sometimes with mixed success. It is therefore better to make the network larger rather than smaller for a given decision space.

D. Multi-Level Perceptron (MLP)

The MLP network is trained via backpropagation of the errors to the inputs. In typical operation, a data pair, consisting of a data input and a desired output is presented to the network. The input cascades in a feedforward mode to the outputs, being summed and processed through three network layers. Each layer consists of the summing junction of weighted inputs from the previous layer and then as discussed, an array of nonlinear elements connected to each summing junction. At the end, the output is compared with the desired output, and an error is calculated. This error is processed and the weights are updated via a training rule such as the least mean squares algorithm, whereby the effects of the error are "backpropagated" to the inputs of each layer and ultimately, to the global network input[10]. Then, the next training pair is applied. The process is repeated until the error is smaller than some threshold level.

It should be mentioned here that this training process is repeated thousands of times usually. It is not uncommon to use 10-20000 iterations to achieve network convergence. However, in computer processing time, the process time only takes a few minutes.

In operation, the weights are again held constant and the network results in the desired output.

E. Hybridnet

The need to translate the date from the SOM into a more deterministic form drove the creation of the hybrid neural network, HYBRIDNET. In this topology the output classes of the SOM are input to a fairly small backpropagation network the goal of which is to translate the sound classes into repeatable sound identification. This is shown in Fig 5. The combination of unsupervised and supervised methods has resulted in a robust network system which can rapidly identify many sound samples.

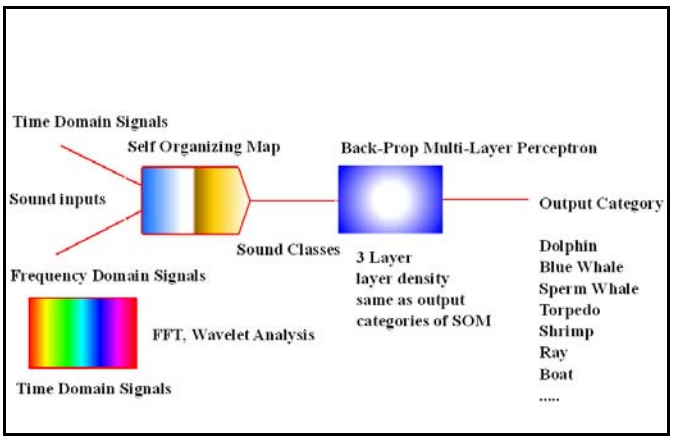

Fig. 5. Hybridnet Architecture

F. Experimental Focus

The area which this paper examines is the difference in functionality of HYBRIDNET versus SOM for two basic characteristics of the input data: (1) Data structure and (2) Spectral resolution of the data. Other parameters which dramatically impact the behavior of the networks and which have been examined are included in Table 4.

Table 4
Parameters Affecting Network Performance

# Neurons	Resolution of Data
# Layers	# Exemplars
Activation Function	Data Presentation
Activation Function Parameter	Frequency resolution
Learning Algorithm	Termination Condition
Learning Rate	SOM Neighborhood function
Weight Initialization	Learning rate Adjustment
# Classes of SOM	# Backprop Categories
Sample window length	Sample window overlap

G. Data Presentation

As can be seen from Fig. 5, the data can be presented either as time domain or frequency domain information. However, based on existing successes in sonar and in speech processing work, the data here was presented as two dimensional arrays of spectra.

Initially, all data was converted to 16 bit PCM coded .WAV files, sampled at 16 ksps. All networks were encoded in MATLAB from scratch. After validation of the networks on standard data sets such as the XOR problem, the networks were configured to accept 2 dimensional arrays in which one dimension is the FFT spectrum, and the other is time. Each sound file was processed with MATLAB using the standard FFT algorithm resident in MATLAB, based on the the Cooley-Tukey algorithm [13]. The window and overlap of each sample was held constant during each test. In the case of the supervised learning, a category number was assigned to each sample. For example, a blue whale might have two file sources in the training data, but both would be mapped to sound type "4".

In the program, the width of the sound sample window, the time overlap between spectra, and the offset from the beginning of the file are moveable depending on the data sample and experiment desired.

H. Error Evaluation

In typical neural network applications, and for general training of the neural networks, the 2norm is regularly used to evaluate the error resident in the system. As new inputs are presented to the network, the individual sample error will vary considerably, sample to sample. A network is said to converge however if the average error over one presentation of all inputs is reduced. This is referred to as the epoch error. In evaluating the performance of the neural network however, this may not be the most useful measurement. Particularly in the case of SOM maps, at training completion, the average error is well above 1.

The repeated trial method was used to evaluate the network performance. Using this method, a separate testing data set is created. Because of limited sound files with single sources, these were often made from the sound file used for training, but in a portion of the file not used in training. Then repeated trials were performed with several different samples over the testing set. Outputs were binned by category, characteristic groupings of sounds by source. To date, the only value looked at has been the correct positive number, which for man made objects has approached 100% in several trial runs. This has included munitions deployment, diesel submarines and boats, even with hydrophone saturation. This method also illuminated early problems with the SOM in characterizing cetaceans which was resolved later by HYBRIDNET. However, in future work with more robust source data, the system is configured to examine statistics on false alarm (FA) and other figures of merit.

IV. SOM Results

A. Network Topology

Table 5 lists the parameters used for these tests to demonstrate the capability of the networks. These were selected based on the incoming data characteristics on overall utility for a variety of data sets.

The network used for these runs was a basic Kohonen map. The input plane is 2 dimensional, frequency versus time. The weight matrix is thus a three dimensional matrix, mapping to the output. The output plane is a one dimensional array of sound classes.

Table 5
Variables for SOM Study

#Specs	Frame Size	#FFT
10	.5 sec	256
#SOM Cat	Frame Overlay	1024
20	25%	4096

A neighborhood function was included, but the network converged equally well with no neighborhood function, and reward distances of 0, 1, and 2.

B. Data Set 1 Results

The number of FFT points and thus the spectral resolution was changed. Table 6 reports the results on data set 1.

Table 6
Data Set 1 SOM Mappings

	FFT:	256	1024	4096
Sound:				
Tarpon		12 (10,6)	4	6
Ship		5	15	7
Ship_Close		5	15	7
ORCA		16(10)	3(4)	11(17,15)
Quake		12	14(4)	3
Sub		17	16(17)	17
Tremor		12	14	3
Catfish		12(10)	4(3,14)	6(11)
Blue Whale		16	3	14
C-Launch		5(16)	3(15)	7(11)
Grunt		10(16)	4(3)	6(7,11,14,16)
Humpback Long Cry		8(12,15)	18	8(3,13)
Net Parameters				
2000 iter	20 cat. 10 spectra	.5 width, .5 off	.25 overlap	Lrn Rate=.99, R=1

The primary numbers represent which category the sound was mapped to over repeated trails of 10 samples. If only one number is shown, then the sound was mapped to that category number all 10 times.

In the 256 point fft, it is observed that the boat is mapped to a unique category on all occasions as is the earthquake and the submarine. In others, there is some overlap. For the grunt, there is a unique primary category but with some overlap with the Orca. The humpback whale is even more variable, mapping to the tremor and two unique categories. Type 16 sound for this run may in fact reflect background noise or may be a shared sound component, as it is seen in 4 of the source tests. Lastly, category 12 is used by two of the fish and the earthquake.

By expanding the FFT to 1024 and 4096 points, the results improve. The boat and sub map to unique categories, although the cold-launch still maps to the boat category. The earthquake and tremor now map to the same category, distinct from the biological sources. All three fish now map to the same category which is a persistent problem when mixing fish data with other data. And the blue whale and Orca map to unique sources.

What is observed is that for most of these sounds, there are several component sounds in the more complex acoustic ensemble. This is particularly important in the biological sources.

C. Data Set 2 Results

To investigate the category splitting more, several sound files were added to the set, while removing the fish as it was determined from other experiments that the fish can be treated separately. Table 7 describes the results of this data set interacting with the SOM.

Table 7
Data Set 2 SOM Mappings

	FFT:	256	1024	4096
Sound:				
Tarpon		19(14,13)	17(16,13,1,2)	10(12,14,15)
Ship		12	9	19
Pacific Blue Whale		14	2	18
ORCA		5	20(3)	20(16,7)
Quake		14(5)	2	18
Sub		13(15)	17	16
Tremor		14	2	10
Humpback Whistle		2(14,1,5)	2(1,5,13,14)	10
Blue Whale		5	8	7
C-Launch		5(12)	12(9,3)	19(20,16)
Grunt		5	20(12)	7(16,19)
Humpback Long Cry		2(14,7)	1(2,4,6)	10(13,5)
Net Parameters				
2000 iter	20 cat. 10 spectra	.5 width, .5 off	.25 overlap	Lrn Rate=.99, R=1

It is interesting to note that the added humpback whale sounds map sometimes to unique categories, sometimes to fish type sounds, and sometimes to other whales such as the blue whale. This is an example of how complex the vocalization patterns are in the data set. One item during repeated trials is that the category selected directly tracks where in the sound file the frame set came from. In other words, when the SOM is classifying a sound one way and then at other times another way, it is because of actual differences in the structure of the sound. Notice lastly, that the category numbers mapped to are different for each run, but the patterns remain. For example, the quake and tremor map typically to the same class of sounds and the blue whale maps to the earthquake typically.

When looking at the output of the SOM, it is observed that sounds produce some output at categories other than that of the primary map. When looking at the distribution of mapping, it becomes clearer that this pattern of multiple class selection could be used as the input to another network for further processing.

V. Hybridnet

A. Network Structure

The structure of the HYBRIDNET is displayed in Fig. 5. The network was initially trained by cascading the output of the SOM into the MLP and applying the reference category to the output of the MLP for each individual sample. This worked reasonably well, but after comparison with another approach, batch processing, the batch method was chosen for this work. The batch method works as follows in Table 8.

Table 8:
Batch Method for Hybridnet Training

Step	Description
1	Apply Dataset to SOM until termination is met
2	Freeze weights of SOM
3	Apply Dataset to SOM and cascade SOM output to MLP
4	Apply reference vector to output of MLP and train via backprop
5	Repeat for all data set until error is minimized
6(Operation)	Apply test data and perform repeated trails

Two main differences exist in training and interpretation of the HYBRIDNET system. First, it is a supervised learning approach, requiring a desired output category in order to determine error for training. Secondly, there is a number value associated with the magnitude of the neuron output. The neurons in this system were scaled between 0 and .8, a binary approach, though bipolar neurons have been evaluated. Any output less than .5, even if the largest output, was labeled indeterminate. In the tables of results, particularly with only 1000 training iterations, several outputs, though correct, were still indeterminate. As the number of cycles increased, most numbers increased indicating a better trained network. At a certain point, the numbers actually go down indicating the limit of training, and the need for additional training data. Any additional training iterations will only result in overfit of the data by the network, giving spurious results.

B. Data Set 1 Results

Data set one was applied to the network. The results are in Table 9. The 12 patterns are mapped into 8 categories.

Table 9
Data Set 1 Hybridnet Mappings

	Iter.	1000	10000	20000
Sound:	Categ			
Tarpon	1	1(.21)	4(.20)	3(.19)
Ship	2	2(.63)	2(.73)	2(.65)
ShipClose	2	2(.63)	2(.73)	2(.63)
ORCA	3	5(.46)	4(.46)	5(.39)
Quake	4	4(.59)	4(.62)	4(.60)
Sub	5	5(.77)	5(.71)	5(.79)
Tremor	4	4(.75)	4(.77)	4(.60)
Catfish	1	6(.23)	4(.26)	2(.25)
Blue Whale	7	9(.34)	9(.32)	9(.32)
C-Launch	8	2(.28)	2(.37)	6(.18)
Grunt	1	9(.263)	2(.34)	2(.26)
Humpk Long Cry	10	10(.641)	10(.77)	10(.75)
Net Params	Epoch 12			
1024 FFT	20 cat. 10 spec.	.5 width, .5 offset	.25 overlap	Mu=.99, R=1

As can be seen there is some confusion particularly with the fish. Even at 20000 iterations, the network is changing its mapping for Tarpon, Catfish, and Grunt. Again the boat, sub, earthquake, and humpback correctly map. Also the system recognizes that both the small tremor and the earthquake are the same type of event.

One feature unique and indicating that some interference and confusion is taking place is that the blue whale consistently maps to the wrong category but it is a unique category, distinct from the others. In earlier runs, even with fish mapped to separate categories by species, after 20000 iterations, the fish converged to one category, albeit incorrect.

C. Data Set 2 Results

When the fish are removed, and replaced with a variety of sounds, the results are much better. Even at 1000 iterations, the network shows several confidence values above .5. By 10000 cycles, the system identifies 11 out of 12 correct categories. By 20000 samples, the confidence values are dropping as well as only 10 of 12 categories identifying correctly, indicating an overfit situation. The cold launch and the grunt appear to be the elements most difficult to the network for this training set.

Table 10
Data Set 2 Hybridnet Mappings

	Iter.	1000	10000	20000
Sound:	Map Cat			
Humpb HauntCry	1	1(.94)	1(.91)	1(.83)
Ship	2	2(.81)	2(.80)	2(.78)
NEP Blue Whale	7	4(.40)	7(.44)	7(.55)
ORCA	3	3(.73)	3(.64)	3(.66)
Quake	4	4(.48)	4(.65)	4(.70)
Sub	5	5(.62)	5(.64)	5(.69)
Tremor	4	4(.46)	4(.64)	4(.82)
Humpbk Whistle	1	1(.85)	1(.89)	1(.74)
Blue Whale	7	7(.46)	7(.65)	7(.76)
C-Launch	8	3(.54)	2(.48)	3(.44)
Grunt	10	1(.51)	10(.41)	4(.39)
Humpk Long Cry	1	1(.76)	1(.65)	1(.72)
Net Params	Epoch 12			
1024 FFT 1000 iter SOM	20 cat. 10 spectra	.5 width, .5 off	.25 overlap	Lrn=.99, R=1

VI. Conclusions

A. Network Performance

The results are encouraging. For tasks such as man made object identification and tracking, the passive sonar processed by the HYBRIDNET gave high recognition and deserves further analysis. The networks appear to be suffering from overfit due to either too many neurons or too little training data. A normal robust training set should be on the order of 200-300 samples per epoch.

Yet even with this extremely small set, the network can distinguish between whale species, yet correctly group the complex vocalizations of humpback, orca, blue, and in other runs, sperm whales, by species. The ability to distinguish between cetacean sounds is important for long term counting studies and to identify the behavior being performed at the time of the vocalization (mating, social, feeding, defense). For harbor vehicle counting, the ability to discriminate the vehicle type is very encouraging and the network performs well separating the boat from the submarine. It is anticipated that with cold-launch data, likewise good results will be achieved.

The same system could be achieved in MLP only networks, however, the network would be on the order of 6 Mbytes in size. By reducing the size of the MLP by the SOM, these networks were achieved in fewer than 50k bytes of storage, suitable fro embedded implementation. Further reduction can be achieved by using fixed point math in the networks. This opens the possibility of smart acoustic sensors which can wake from sleep on not just raw acoustic signal pressure, but rather on specific signature cues.

B. Performance Limits

The limits of the networks at this time are due to limited training data. The manner in which the system can be programmed is large. For example, the system could be trained to map non-ship traffic to a category "other" while focusing most of its attention and storage on vehicle signatures. Similarly the network can be optimized for fish only. However after reviewing the URI data set, it is felt that this system will be able only to break fish into 6-8 broad categories based on noise mechanism, periodicity, feeding strategy, mating strategy, etc. One of the major problems with the fish sounds in this data set was that vocalizations were highly episodic and transient. In other words, any source signal was 90% background noise and only one frame in ten might contain noise generated from the fish. This is being investigated further.

MLP convergence was an issue, but the implementation of Nguyen-Widrow weight initialization and momentum learning has sped up training until as is seen here, convergence no longer is the limiting factor. By 10000 training iterations, the network is at a minimum error.

VII. Summary

A novel neural network paradigm has been designed and characterized in an early form. Test results using MATLAB indicate that the network strategy can easily detect and classify man made noises, geological sounds, and some bioacoustics. Previous work has indicated that fish sounds can also be characterized. However this was not demonstrated in these data sets. Future work involves acquiring more extensive data sets, characterizing multi-source performance, and implementing autonomous operation.

Acknowledgements

Field deployment resources were provided by Marine Resources Development Foundation of Key Largo, Fl.

References

[1] T. Kohonen, K. Torkolla, M. Shozakai, J. Kangas, and O. Venta, "Microprocessor implementation of a large vocabulary speech recognizer and phonetic typewriter for finnish and japanese," European Conference on Speech Technology, Edinburgh, Scotland: 1987, vol 2, pp. 377-380.
[2] W. W. Streilein, P. Gaudiano and G. A. Carpenter, " A neural network for object recognition through sonar on a mobile robot," *Proceedings of the 1998 IEEE/ISIC/CIRA/ISAS Joint Conference*, Gaithersburg, Md: September, 1998, pp. 271-276.
[3] J. D. Paola and R. A. Schowengerdt, "A review and analysis of neural networks for classification of remotely sensed multi-spectral imagery," *RIACS Technical Report 93.05, NASA-CR-194291*, 1993.
[4] R. Baran and J.P. Coughlin, "A neural network for target classification using passive sonar," *Proceedings of the Conference on Analysis of Neural Network Applications*, Fairfax, Virginia, 1991, pp. 188-198.
[5] Y.D. Lin, Real-time recognition of fish species using neural network analyis of acoustic signature," Dissertation, Melbourne, Fl: Florida Institute of Technology, 1998.
[6] M. P. Fish and W. H. Mowbray, *Sounds of western north atlantic fishes*, Baltimore, Md: Johns Hopkins Press, 1970.
[7] R. J. Urick, " Principles of Underwater Sound,', Los Altos, Ca: Peninsula, 1983, p 104.
[8] C.G. Fox et al, "Acoustic detection of a seafloor spreading episode on the Juan de Fuca Ridge using military hydrophone arrays," Geophysical Research Letters, vol. 22, no. 2, pp. 131-134.
[9] J. Ghosh, K. Turner, S. Beck, and L. Deuser, "Integration of neural classifiers for passive sonar signals," Control and Dynamic Systems—Advances in Theory and Applications, volume 77, C. T. Leondes, ed. New York, NY: Academic Press, 1996, pp. 301-338.
[10] L. Fausett, *Fundamentals of neural networks*, Upper Saddle River, NJ: Prentice Hall, 1994. pp. 156-192.
[11] S. Haykin, *Neural networks, a comprehensive foundation*, Upper Saddle River, NJ: Prentice Hall, 1999. pp. 443-477.
[12] F. M. Ham and I Kostanic, *Principles of neurocomputing for science and engineering*, New York: 2001, pp 96-119.
[13] J.W. Cooley and J. W. Tukey, " An algorithm for the machine computation of the complex fourier series," *Mathematics of Computation*, Vol. 19, April 1965, pp. 297-301.