The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations

Scott O. Murray[a)]
Institute of Theoretical Dynamics, 2201 Academic Surge, 1 Shields Avenue, University of California, Davis, California 95616

Eduardo Mercado
Center for Molecular and Behavioral Neuroscience, Rutgers University-Newark Campus, 197 University Avenue, Newark, New Jersey 07102

Herbert L. Roitblat
Department of Psychology, 2430 Campus Road, University of Hawaii, Honolulu, Hawaii 96822

(Received 16 April 1998; revised 10 July 1998; accepted 5 September 1998)

This study reports the use of unsupervised, self-organizing neural network to categorize the repertoire of false killer whale vocalizations. Self-organizing networks are capable of detecting patterns in their input and partitioning those patterns into categories without requiring that the number or types of categories be predefined. The inputs for the neural networks were two-dimensional characterization of false killer whale vocalizations, where each vocalization was characterized by a sequence of short-time measurements of duty cycle and peak frequency. The first neural network used competitive learning, where units in a competitive layer distributed themselves to recognize frequently presented input vectors. This network resulted in classes representing typical patterns in the vocalizations. The second network was a Kohonen feature map which organized the outputs topologically, providing a graphical organization of pattern relationships. The networks performed well as measured by (1) the average correlation between the input vectors and the weight vectors for each category, and (2) the ability of the networks to classify novel vocalizations. The techniques used in this study could easily be applied to other species and facilitate the development of objective, comprehensive repertoire models.
[S0001-4966(98)03312-8]

PACS numbers: 43.80.Ka [FD]

INTRODUCTION

Quantifying a species' repertoire is a fundamental challenge in the study of animal vocalizations. Many attempts have been made to characterize the various sounds produced by dolphin (*Delphinidae*) species. However, little progress has been made in developing objective, comprehensive repertoire models. The development of such models is important because they can facilitate comparisons both within and between species, aiding in the development of functional models. Currently, the field lacks an objective method capable of classifying the entire vocal repertoire of a dolphin species. Murray *et al.* (1998) describe a method capable of characterizing dolphin vocalizations that can be applied to all signal types (e.g., pulsed and continuous waveforms). This paper extends that work, demonstrating how self-organizing neural networks can classify the repertoire of false killer whale vocalizations.[1]

Techniques that categorize dolphin vocalizations based on objective and quantitative analysis methods have recently been explored (e.g., Buck and Tyack, 1993; Dawson and Thorpe, 1990; McCowan, 1995). A dynamic time-warping method was used by Buck and Tyack (1993) to assess the similarity of bottlenosed dolphin (*Tursiops truncatus*) whistles. The method used an algorithm that first extracted the frequency contour of the whistles through fundamental frequency analysis. The algorithm then performed a nonuniform time dilation to align the contours by minimizing the total square difference between the observed contour and a reference contour. Finally, the algorithm computed a distance measure between the observed contour and a library of reference contours. The observed contour was assigned to the closest reference contour. The technique was derived from speech recognition approaches (e.g., Itakura, 1975), and assumed that two whistles with similar contour shapes were the same, despite any differences in absolute length of the vocalization.

McCowan (1995) made similar assumptions about which features are most important in whistle analysis. In addition to generalizing across whistle length, she assumed that whistles that have been shifted up or down in absolute frequency, while maintaining the same ''shape,'' should be categorized as the same. Twenty measurements of peak frequency were taken to represent each whistle. The frequency measurements were used to generate a correlation matrix, and principal component analysis was conducted using the correlation matrices. The factor scores from each data set of whistles were subjected to *K*-means cluster analysis to group whistles into clusters based on contour similarity. By using correlation matrices, the technique was able to cluster whistles that differed in absolute duration and frequency.

[a)]Electronic mail: smurray@itd.ucdavis.edu

Both of these techniques (Buck and Tyack, 1993 and McCowan, 1995) are vast improvements over subjective judgment in that they ensure reliable classification performance. However, these techniques have only been applied to whistle-type vocalizations and they may, in fact, be limited in their ability to categorize other signal types. Both methods only consider frequency information and do not take into account changes in amplitude characteristics that occur with pulsed vocalizations.

A. Self-organizing neural networks

Neural networks are a promising technique in the analysis of animal vocalizations. Neural networks have been successful at classifying a number of complex signal types, including human speech (e.g., Kohonen, 1988; Huang and Kuh, 1992) and dolphin biosonar echoes (e.g., Au and Nachtigall, 1995; Roitblat *et al.*, 1989). The study reported here employs *unsupervised neural networks*. Unlike supervised neutral networks (e.g., multilayer perceptrons), unsupervised networks require only weak assumptions about the number and type of potential categories. Unsupervised networks are capable of learning to detect regularities and correlations in their input, and adapting their responses to that input (Demuth and Beale, 1993). Unsupervised networks are called self-organizing because the organization is not imposed on them by an outside intelligent agent, but instead is learned as the outcome of the patterns to which they are exposed and the learning algorithm which adjusts their weight structure. Generally, unsupervised neural networks partition a given data set into disjoint subsets (i.e., categories), such that patterns in the same category are as alike as possible, and patterns in different clusters are as dissimilar as possible (Mehrotra *et al.*, 1997). While most unsupervised networks share this similar goal, they may differ in the specifics of their mathematical implementation.

Self-organizing neural networks, similar to the one presented in this paper, have been used previously to classify humpback whale song vocalizations (Walker *et al.*, 1996). Time-frequency representations (spectrograms) of humpback whale song units were used as inputs into a self-organizing feature map. The network classified the song units similarly to human visual and aural impressions and traditional statistical clustering algorithms.

The technique presented by Walker *et al.* (1996) can be applied to the entire song repertoire, but spectrograms may not be the best choice for neural network inputs. The time–bandwidth tradeoff inherent in all spectrogram-like representations can dramatically affect the representation of a signal and how it is classified. Very different spectrograms can result from the same signal following only slight changes in the window size. For example, a signal can appear to be a continuous whistle with a large window size, and appear to be a series of short pulses with a small window. In the absence of information about the animal's integration window, therefore, arbitrary window sizes and shapes may mislead categorization.

The inputs for the neural networks used in the study reported here were two-dimensional characterizations of false killer whale vocalizations. Each vocalization was characterized by its simultaneous modulations in duty cycle and peak frequency (Murray *et al.*, 1998). The short-time duty-cycle measure compares the signal to a continuous sinusoid. As the signal approaches a continuous sinusoid, the duty-cycle measure begins to approach 1.00. Pulses are represented as lower duty-cycle values as a function of pulse repetition rate (Murray *et al.*, 1998). Consequently, high duty-cycle vocalizations are heard as whistles; lower duty-cycle values correspond to pulsed vocalizations. The duty cycle/peak frequency representational scheme presented here is not subject to the same constraints as spectrograms. With spectrograms, the time-frequency tradeoff can qualitatively alter the signal representation (or ''type''). For example, a pulsed signal can appear to be continuous with the proper window size. Here, the time–frequency tradeoff only affects the resolution of the peak frequency measurements and does not affect the type of signal as represented by the duty cycle measurements.

The first neural network used in this study used *competitive learning*, where units in a competitive layer distributed themselves to recognize frequently presented input vectors. The result of this network was a set of classes representing typical patterns in the vocalizations. The second network used was a Kohonen feature map, which is similar to a competitive network in many respects. The additional aspect of a feature map is that the outputs are organized topologically. Similarity among patterns are mapped into closeness relationships on a grid, providing a graphical organization of pattern relationships (Dayhoff, 1990).

The primary advantages to the techniques used in this study are that all emitted vocalizations were characterized using a single method, and the outputs of the characterization were organized into patterns based on the features present in the vocalizations. Both types of neural networks (competitive and Kohonen feature map) require few *a priori* assumptions regarding the categorical structure of the vocalizations. Instead, the networks search for correlational structure in the data and form categories around these centers of correlation. Both networks were used in order to contrast their respective outputs.

I. METHODS

The 500-vocalization data set used in Murray *et al.* (1998) was used for this study. The vocalizations were from two false killer whales, one male and one female, located at Sea Life Park, Oahu, Hawaii. Recordings were made by isolating each animal in a distant portion of its tank while the other animal remained behind a gate in another portion of the tank. The minimum distance between the animal being recorded and the other animal behind the gate was approximately 30 m. While recording, the trainer positioned the animal's melon (forehead region of the animal from which it is believed sound emanates) underwater so that its head was about 1–2 m away from the hydrophone. This procedure ensured very high signal-to-noise ratios, as well as confidence concerning the identity of the animal making the sounds (Murray *et al.*, 1998).

All sounds were recorded with a Sony digital audio-tape recorder (DAT), TCD-D8, which uses a sampling rate of 44.1 kHz, for a frequency bandwidth to 22 kHz. A hydrophone (custom-built by W. Au) with a sensitivity of −185 dB and a bandwidth to 200 kHz was used for all recordings. Using a quasirandom procedure, 500 vocalizations were chosen for analysis and digitized onto a PC using a SoundBlaster-32 sound card.

A single vocalization was defined as an uninterrupted (in time) sound emission and could encompass multiple sound ''types'' (e.g., a whistle and pulse train). The data set included a random sample of 250 vocalizations from each of the two false killer whales. Each vocalization was sequenced into a series of short-duration (512 point—approximately 11.6 ms) nonoverlapping time windows and described along two dimensions: duty cycle and peak frequency. Duty cycle refers to the percentage of time a signal is ''on'' relative to the total length of the signal and in this context is relative to the 512-point window length. The duty cycle algorithm assigned a value between 0.0 (no signal—e.g., an interpulse interval) to 1.00 (a continuous signal—e.g., a whistle). In this way, the duty-cycle measure gave an approximation of the type of waveform (e.g., pulsed versus continuous) within each time window.

The characterization vectors (duty cycle and peak frequency) for each vocalization served as the inputs into a self-organizing, competitive neural network and a two-dimensional, self-organizing feature map. The input vectors for the networks must have the same number of elements; therefore, the characterization vectors for each vocalization were sampled 30 times at regular intervals. The average duration of each vocalization was 506 ms $(\text{s.d.}=761\,\text{ms})$, meaning that most vocalizations had approximately 40 windows. The use of 30 samples was arbitrary, but preliminary analysis demonstrated it to be sufficient to capture the dynamics of most signals. When pulse trains were analyzed, the zero elements (representing interpulse intervals) and nonzero elements (representing individual pulses) were sampled separately. This was done to maintain the same relative spacing of zero and nonzero elements in the vectors.

Before presentation to the neural network, the input vectors were scaled to z scores, using the grand mean and standard deviation over all signals. The mean for the duty-cycle values was 0.46 $(\text{s.d.}=0.34)$ and the mean for peak frequency was 7122 Hz $(\text{s.d.}=2687\,\text{Hz})$. Signal vectors were then normalized to unit length by dividing each vector by its length, meaning that the input vectors lay on a unit hypersphere. The input values were normalized because the neural network algorithm used maximum dot product as a similarity measure. If two vectors are of unit length, the dot product is equal to the cosine of the angle between the two vectors (i.e., a ''meaningful'' measure of similarity). Normalizing to unit length removes magnitude information from the inputs and is important to consider when interpreting the results. For example, after normalization of z-scored feature vectors, a window which was 0.1 standard deviations above the mean in frequency and of mean duty cycle, [0.1 0], would be treated as equivalent to a window that was 3 s.d. above the mean in frequency and of mean duty cycle, [3 0]. (After normalization both vectors equal [1 0].) What is preserved after normalization is the dynamics, or the change of the signal across time relative to the mean.

The inputs for the neural networks were a combination of duty cycle and peak frequency values. Input vectors were constructed by concatenating the two 30-element vectors into a single 60-element vector. To test the reliability of the categories formed by the network, a subset of 250 input vectors was chosen randomly from the set of 500 to serve as a training set. These vectors served as inputs to train the network. The remaining 250 vectors served as a test set. The performance of the network developed with the training set was compared with the test set. The neural networks were implemented using custom script-code accessing functions in MATLAB's Neural Network Toolbox (The MathWorks, Inc.).

II. COMPETITIVE NETWORK

The units in the competitive network were initialized to random weight vectors with the number of elements in each weight vector equal to the number of elements in the input vectors (i.e., 60 elements). An input vector was presented to the network and the angle between the input vector and each of the unit's weight vectors was computed. The unit with the smallest angular difference from the input vector was the ''winner.'' The weights of the winning unit were adjusted in the direction of the input vector. The size of the adjustment was controlled by a learning-rate parameter. Therefore, when the same input vector was presented again, the winning unit was more likely to win and its values were adjusted closer to the input vector. The weight vectors of each of the units, at the end of training, represented prototypes or category ''centroids.''

To summarize, the competitive network worked as follows:

(1) Apply an input vector $\mathbf{X}$.

(2) Calculate the angular distance $\mathbf{D}_j$ between $\mathbf{X}$ and the weight vectors $\mathbf{W}_j$ of each unit. Since normalized inputs and weight vectors were used, the cosine of the angle between $\mathbf{X}$ and $\mathbf{W}$ equals the dot product:

$$\mathbf{D}_j=\mathbf{X}\cdot\mathbf{W}_j$$

(3) The unit that has the weight vector closest to $\mathbf{X}$ (i.e., the largest dot product) is declared the winner. The winner's weights are adjusted in the direction of $\mathbf{X}$ by the formula:

$$\mathbf{W}_j[n+1]=\mathbf{W}_j[n]+\alpha(\mathbf{X}-\mathbf{W}_j[n]),$$

where n indicates the iteration number, and α the learning rate.

(4) Perform steps (1) through (3), cycling through each input vector.

After training, each of the input vectors was assigned to the unit (category) whose weight vector (category centroid) was closest. The performance of the network was assessed by calculating the average cosine of the angle between each unit's weight vector and the input vectors assigned to it. In other words, the degree to which each input vector was related to its respective unit's weight vector was measured.

A. Results

The number of units in a winner-take-all network determines the maximum number of potential categories. The number of units ultimately used in training the network was arrived at through a trial-and-error procedure by first starting with a large number of units—i.e., many more than reasonably suspected categories—and then reducing the number. Forty units (possible categories) were first used. Presentation of the training vectors was iterated 20 000 times at learning rates of 0.05 and repeated with a rate of 0.10. In both cases, only five of the units learned—i.e., showed adjustments in their values. The number of units was subsequently reduced by one-half (from 20 to 10). In all cases, only five units adjusted their weights.

The network was trained with five units at a learning rate of 0.10 for 10 000 iterations. The weight vectors from each of the five units after training are shown in Fig. 1. The *x* axis represents each of the 60 elements of the vectors and the *y* axis represents normed *z*-score values. Zero represents the mean; values above and below zero represent deviations from the mean. The first 30 elements representing duty-cycle values are shown above the second 30 elements representing peak frequency.

Looking at weight vector 1 (W1) in Fig. 1, the first 30 elements (representing duty cycle) are constant and of relatively high value. The representation of peak frequency (dashed line) is ascending. This vector represents ascending whistle vocalizations. The false killer whales used in this study frequently made short-duration ascending whistles; these vocalizations are one of the most salient vocalizations when listening to the animals in almost any behavioral context. Because these vocalizations were so commonly observed, the observance of a weight vector that represented these vocalizations gave validity to the performance of the neural network.

Looking at weight vector 2 (W2) in Fig. 1, the first 30 elements (representing duty cycle) begin at relatively high values, then approximately halfway through (element 13 or 14) drop in value. This weight vector seems to represent the whistle→pulse-train vocalizations. These vocalizations begin as whistles, then switch to what sounds to us like a click train or a rapidly pulsed vocalization (Murray *et al.*, 1998). Looking at elements 31 to 60 (representing peak frequency), it appears that during the high duty-cycle portion (i.e., the whistle), peak frequency is ascending. During the pulse-train portion, the peak frequency of the end of the whistle is maintained at a relatively constant level throughout the duration of the pulse train, similar to the examples presented in Murray *et al.* (1998).

Weight vector 3 is straightforward to interpret. Both the duty cycle (first 30 elements) and the peak frequency (second 30 elements) are relatively constant and at low values. This vector is the result of low-frequency pulse trains. Weight vector 4 has a similarly straightforward interpretation. It has intermediate duty-cycle values and low peak-frequency values, and is likely the result of lower frequency, rapidly pulsed vocalizations. Weight vector 5 has low duty-cycle values and high peak-frequency values and seems to represent high-frequency pulse trains.

The performance of the neural network was evaluated by first calculating how much of the input space was accounted for by each weight vector. The cosine of the angle between each input vector used for training and each unit's weight vectors were calculated. The input vectors were assigned to the category represented by the unit with the closest weight vector. Thus, there were five clusters of input vectors corresponding to the five units. A total of 94 of the training vectors clustered with the weight vector of unit 1 (W1), 25 with W2, 43 with W3, 27 with W4, and 61 with W5.

The average similarity (as measured by angular distance) across all training vectors and their respective category's weight vector was 0.72. Additionally, the average similarity was calculated between the members of each category and the four ''unassociated'' units. The average between-category similarity was -0.11. The results reflect the general goal of an unsupervised network—to partition a data set into disjoint subsets (i.e., categories), such that patterns in the same category are as alike as possible, and patterns in different clusters are as dissimilar as possible. Referring to Table I, the average within-category similarity is shown in the main diagonal. The other cells in the table show between-category similarity.

The 250 vectors not used to train the network served as a novel test set. Each member of the test set was clustered with the nearest weight vector from each of the five units. A total of 93 of the test vectors clustered with W1, 20 with W2, 41 with W3, 28 with W4, and 68 with W5. The distribution of vocalizations among the weight vectors in the test set is closely aligned with the training set, suggesting representative samples for both the training and testing data set. The average similarity between each test input vector and its unit's respective weight vector was 0.69. Performance of the network with the test set (0.69) is comparable to that of the network using the training set (0.72). The average between-category similarity for the test set was -0.11 (see Table II).

The competitive neural network recognized five major categories in the false killer whale vocalizations analyzed. The within-category similarity was high, with an average correlation of 0.72. Additionally, the categories learned with the training set were able to be generalized to the novel test set, suggesting that the categories are reliable. The competitive network approximates the minimum number of categories present in the input patterns. The next network used, a Kohonen feature map, provides a different representation of the vocalizations.

III. FEATURE MAP

The two-dimensional feature map used in this study was similar to the competitive network described above. However, the competitive units were ordered topologically in a two-dimensional square grid. Each unit had neighbors on the grid where a neighborhood of diameter 1 included a specified unit and its immediately adjacent neighbors. A neighborhood

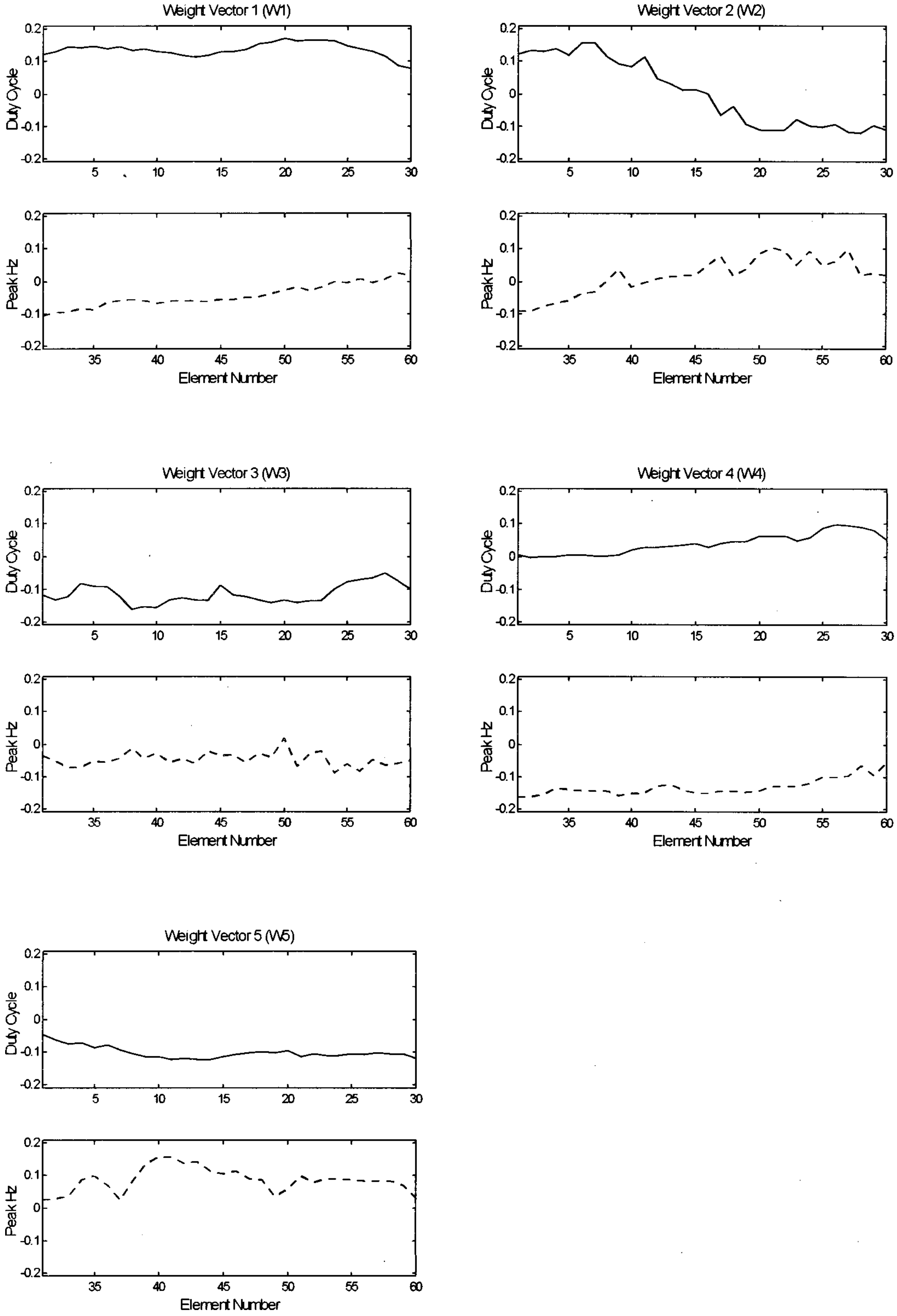


FIG. 1. The weight vectors of each of the five units after training with duty-cycle and peak-frequency inputs. The 60-element weight vectors are shown with the first 30 elements (duty cycle) plotted above the second 30 elements (peak frequency). Zero (*y* axis) represents the grand mean for each dimension (normed *z* scores). The weight vector from unit 1 (W1) has relatively high duty-cycle values and ascending peak frequency. This unit represents ascending whistle vocalizations. The other units can be interpreted similarly.

of diameter 2 included the diameter 1 units and their immediately adjacent neighbors. The feature map in this study used a 5×5 grid.

The feature map differed from the competitive network in terms of which units had their weights updated. In addition to updating the winner, the feature map updated the winner's neighbors. The result was that neighboring units tended to have similar weight vectors (i.e., represent similar portions of the input space). During the initial stage of training, the neighborhood size encompassed the entire 5×5 grid (i.e., each unit adjusted its weights in response to each input vector) and was decreased linearly so that it reached a mini-

TABLE I. The mean similarity/correlation between each cluster of the training set and each of the five unit's weight vector. Within-category similarities are in the main diagonal; all other cells show between-category similarities. This table highlights the ability of the competitive neural network to form maximally distinct categories.

	Weight 1	Weight 2	Weight 3	Weight 4	Weight 5
Cluster 1	**0.81**	0.06	−0.58	0.45	−0.70
Cluster 2	0.06	**0.63**	−0.08	−0.11	0.05
Cluster 3	−0.53	−0.08	**0.70**	0.07	0.33
Cluster 4	0.42	−0.22	0.07	**0.74**	−0.59
Cluster 5	−0.67	0.07	0.33	−0.61	**0.74**

mum of 1 after one-quarter of the training cycles and remained there for the rest of training. This allowed the entire feature map to move initially in the direction of the input space, then, as the neighborhood size decreased to 1, the map ordered itself topologically over the presented input vectors.

The first three steps outlined above for the competitive network apply to the feature map. The following are the additional properties of the feature map.

(1) The winning unit, $\mathbf{W}_c$, is designated as the center of a group of units (i.e., a neighborhood) that lie within a distance D (neighborhood size) from $\mathbf{W}_c$.

(2) Train this group of units according to the formula:

$$\mathbf{W}_j[n+1]=\mathbf{W}_j[n]+\alpha(\mathbf{X}-\mathbf{W}_j[n])$$

for all weight vectors within a distance D of $\mathbf{W}_c$.

As the training progresses, the values of D and α (the learning rate) are gradually reduced.

By assessing the number of input vectors that activated (clustered with) each unit, it was possible to examine the distribution of the input space across the topology represented by the network. Similar to the competitive network, a set of 250 vectors, randomly selected from the total pool of 500, were used as input vectors for training. The remaining 250 were used to test reliability. Reliability was measured by correlating the distributions across the topological map of the training vectors and the test vectors.

A. Results

The feature map was trained for 15 000 iterations at a learning rate of 0.15. The weight vectors after training are presented in Fig. 2. The first 30 elements in each plot represent duty-cycle values, and elements 31 through 60 represent peak frequency. Row and column notation (row, column), will be used to refer to specific units on the grid in Fig. 2.

TABLE II. The mean similarity/correlation between each cluster of the test set and each of the five unit's weight vector. When the trained network is presented with novel vocalizations, very similar patterns of similarity/dissimilarity are found as compared to the training set (Table I).

	Weight 1	Weight 2	Weight 3	Weight 4	Weight 5
Cluster 1	**0.75**	0.09	−0.60	0.35	−0.62
Cluster 2	0.07	**0.59**	−0.05	−0.07	0.01
Cluster 3	−0.52	−0.09	**0.72**	0.12	0.29
Cluster 4	0.41	−0.11	0.04	**0.68**	−0.58
Cluster 5	−0.66	0.07	0.31	−0.61	**0.73**

Referring to Fig. 2, units (2,4), (2,5), (3,4), and (3,5) (i.e., middle/right of grid) represent relatively high and constant duty cycles and gradually ascending peak frequencies. These weight vectors are similar to the weight vector of unit 1 (W1 in Fig. 1) in the competitive network and correspond to ascending whistle vocalizations. Similarly, units (4,1), (4,2), (5,1), and (5,2) (i.e., lower-left portion of the grid) represent low duty-cycle, high-frequency vocalizations. In the upper-left portion of the grid, units (1,1), (1,2), (2,1), and (2,2), represent whistle→pulse-train vocalizations. These units are similar to unit 2 (W2 in Fig. 1) of the competitive network.

The distribution of the training-set input space across the topology of the network is shown in Fig. 3 as ''training set.'' The input space is heavily distributed in the lower left (corresponding to high-frequency pulse trains) and middle right (corresponding to ascending whistles) of the topology. The distribution was also calculated for the novel test inputs and is depicted in Fig. 3 as ''test set.'' The two distributions are similar with a correlation of 0.89.

The categories developed by the feature map were consistent with those of the competitive neural network. Many of the patterns in the weight vectors that were seen in the competitive networks were evident in the feature map. Additionally, the input spaces in both the competitive and feature map networks seemed to distribute themselves similarly. For example, units representing constant/high duty cycle and ascending peak frequency (ascending whistles) attracted a large percentage of the input space in both the competitive network and the feature map.

IV. DISCUSSION

Two types of neural networks were used to classify the vocalizations: a competitive network and a two-dimensional feature map. Both networks were trained with a combination of duty-cycle and peak-frequency input values. The competitive network learned five different categories. The fact that the network learned the two obvious categories—whistles and click trains, reflected by both high and low duty-cycle weight vectors, respectively, attests to the validity of the network.

Based on interpretation of the five weight vectors from the competitive network, the main categories seem to be ascending whistles, low-frequency pulse trains, and high-frequency pulse trains. The network also recognized the whistle→pulse-train transitions as a significant category (see Fig. 1, W2). The peak frequency of the high duty-cycle portion (i.e., the whistle) of this category was ascending. The pulse-train component (low duty cycle) seemed to maintain the peak frequency of the end of the whistle.

It is important to point out that the ability of the neural network to learn these ''combination'' categories (categories with both continuous-wave and pulsed components) was facilitated by the use of a measure of waveform shape. Because of the aural and spectral distinctiveness of many pulsed versus continuous sounds, vocalizations that possess combinations may be arbitrarily separated into different components. Therefore, it is unlikely that the combination categories would have been arrived at through subjective

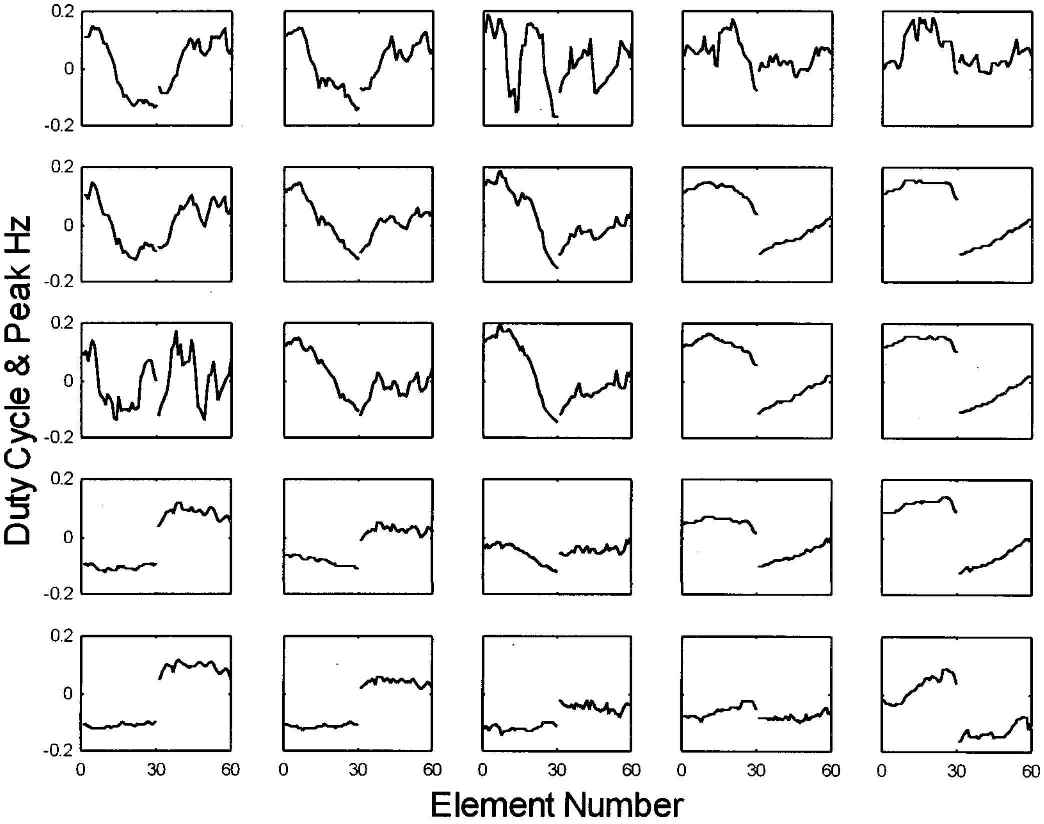


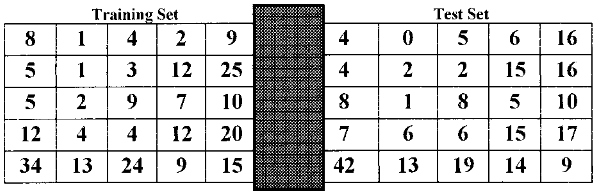
FIG. 2. The topology of the input space using duty-cycle and peak-frequency values. The *x* axis in each plot represents the element number of each weight vector. The first 30 elements represent duty cycle and the elements 31 to 60 represent peak frequency. The *y* axis represents scaled and normalized duty-cycle and peak-frequency values, where zero represents the grand mean. Units on the right/middle side of the topology have relatively high duty cycle and ascending peak frequency (ascending whistles). Units in the lower-left portion have low duty cycle and relatively high peak frequency (high-frequency pulse trains). Also, units that are close to each other have similar-looking weight vectors.

classification techniques such as aural analysis of visual analysis of spectrograms. For example, if aural analysis were used, a whistle that suddenly changed into a pulse train (see Murray *et al.*, 1998, for examples) might be classified into two separate vocalizations: a whistle immediately followed by a pulse train. However, with the short-time duty-cycle measure, the continuity of the vocalizations was preserved. The use of an objective measure of signal type allows for a different definition of a single vocalization: an uninterrupted (in time) sound emission as opposed to a certain subjective class of vocalization (e.g., whistle or pulse train). Such a definition is likely more functionally relevant because it is defined by the vocalizing animal and not by the subjective judgment of a human listener.

The categories developed by the self-organizing feature map complemented the results obtained with the competitive network. The types of weight vectors observed in the competitive network were also seen in the feature map. Additionally, the input distribution patterns (i.e., input clustering) were very similar in both the competitive and feature map networks. Both types of networks demonstrated that the self-organizing approach, using the two types of inputs (duty cycle and peak frequency), is a very effective way of categorizing dolphin vocalizations.

Though the results of the two types of networks were complementary, each has its own advantages. The competitive network was effective in finding the minimum number of potential categories in the data set. Additional characteristics were revealed by the feature map, such as the relative distribution of the input space (through category redundancy) and the topological relationships between categories.

The neural network classification scheme presented here is easily amenable to different types of acoustic signal representations. For example, there may be other relevant dimensions, such as signal duration, which may be important to include in future implementations of these networks. Likewise, some investigators may only want to consider a certain duty-cycle category (e.g., whistles) in their analysis. Such networks could limit their inputs to relevant spectral features



Training Set						Test Set				
8	1	4	2	9		4	0	5	6	16
5	1	3	12	25		4	2	2	15	16
5	2	9	7	10		8	1	8	5	10
12	4	4	12	20		7	6	6	15	17
34	13	24	9	15		42	13	19	14	9

FIG. 3. The distribution of the duty-cycle and peak-frequency input space over the topology depicted in Fig. 2. Most of the input vectors of both the training set and test set clustered in the right-middle (ascending whistles) portion of the topology and in the lower-left portion (pulse trains). The two distributions have a Pearson's correlation of 0.89, meaning that the distribution of the input space in the trained network is generalizable to novel vocalizations.

(e.g., fundamental frequency). Overall, the networks are very flexible, and it is ultimately up to the investigator to determine which inputs are most relevant to his or her particular classification task.

In summary, the techniques used in this study provide a unique and objective method for classifying cetacean vocalizations. Forming simple categories using self-organizing networks can facilitate comparisons between different species and different behavioral contexts, as well as aid in the development of functional models of cetacean vocalizations. The ability of self-organizing networks to ''search'' for inherent relationships in the data and form categories based on those relationships makes them well-suited for classifying animal vocalizations.

ACKNOWLEDGMENTS

This study was conducted as part of an M.A. program at the University of Hawaii, and was made possible through the support of Marlee Breese, formerly of Sea Life Park, Oahu.

[1] The use of the term ''vocalization'' in this paper is not meant to imply that vocal folds are necessarily the mechanism producing the sounds. It is used as a general term for dolphin sounds that are internally generated via airflow in the head region.

Au, W. W. L., and Nachtigall, P. E. (**1995**). ''Artificial neural network modeling of dolphin echolocation,'' in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and J. A. Thomas (De Spil, The Netherlands), pp. 183–199.

Buck, J., and Tyack, P. (**1993**). ''A quantitative measure of similarity for *Tursiops truncatus* signature whistles,'' J. Acoust. Soc. Am. **94**, 2497–2506.

Dawson, S., and Thorpe, C. (**1990**). ''A quantitative analysis of the sounds of Hector's Dolphin,'' Ethology **86**, 131–145.

Dayhoff, J. E. (**1990**). *Neural Network Architectures: An Introduction* (Van Nostrand Reinhold, New York).

Demuth, H., and Beale, M. (**1993**). *Neural Network Toolbox* (The MathWorks, Inc., Natick, Massachusetts).

Huang, Z., and Kuh, A. (**1992**). ''A combined self-organizing feature map and multilayer perceptron for isolated word recognition,'' IEEE Trans. Signal Process. **11**, 2651–2675.

Itakura, F. (**1975**). ''Minimum prediction residual principle applied to speech recognition,'' IEEE Trans. Acoust., Speech, Signal Process. **23**, 67–72.

Kohonen, T. (**1988**). ''The 'neural' phonetic typewriter,'' Computer **21**, 11–22.

McCowan, B. (**1995**). ''A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (*Delphinidae, Tursiops truncatus*),'' Ethology **100**, 177–193.

Mehrotra, K., Chilukuri, K. M., and Sanjay, R. (**1997**). *Elements of Artificial Neural Networks* (MIT, Cambridge, MA).

Murray, S. O., Mercado, E., and Roitblat, H. L. (**1998**). ''Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations,'' J. Acoust. Soc. Am. **104**, 1679–1688.

Roitblat, H. L., Moore, P. W. B., Nachtigall, P. E., Penner, R. H., and Au, W. W. L. (**1989**). ''Natural echolocation with an artificial neural network,'' Int. J. Neural Syst. **1**, 239–247.

Walker, A., Fisher, R. B., and Mitsakakis, N. (**1996**). ''Singing maps: classification of whale-song units using a self-organizing feature mapping algorithm,'' DAI Research Paper No. 833.