**CHAPTER 9: Classifying Animal Sounds with Neural Networks**

Eduardo Mercado III

University at Buffalo, The State University of New York

and

Christopher B. Sturdy

University of Alberta

To appear in: Owren, M. J., & Brown, C. H. Research methods in bioacoustics. Bentham Science Publishers, Oak Park, IL.

Corresponding author: Eduardo Mercado III, Department of Psychology, University at Buffalo, SUNY, Buffalo, NY 14260, USA. (voice) +01 716 604 4424, (email) emiii@buffalo.edu

**Abstract**

Humans naturally classify the sounds they hear into different categories, including sounds produced by animals. Bioacousticians have supplemented this type of subjective sorting with quantitative analyses of acoustic features of animal sounds. Using neural networks to classify animal sounds extends this process one step further by not only facilitating objective descriptive analyses of animal sounds, but also by making it possible to simulate auditory classification processes. Critical aspects of developing a neural network include choosing a particular architecture, converting measurements into input representations, and training the network to recognize inputs. When the goal is to sort vocalizations into specific types, supervised learning algorithms make it possible for a neural network to do so with high accuracy and speed. When the goal is to sort vocalizations based on similarities between measured properties, unsupervised learning algorithms can be used to create neural networks that objectively sort sounds or that quantify sequential properties of sequences of sounds. Neural networks can also provide insights into how animals might themselves classify the sounds they hear, and be useful in developing specific testable hypotheses about the functions of different sounds. The current chapter illustrates each of these applications of neural networks in bioacoustics studies of the sounds produced by chickadees (*Poecile atricapillus*), false killer whales (*Pseudoorca crassidens*), and humpback whales (*Megaptera novaeangliae*).

**Key Words: Connectionism, Parallel distributed processing, Perceptron, Self-organizing, Adaptive filter, Learning algorithm, Computational modeling**

**CHAPTER 13: Classifying Animal Sounds with Neural Networks**

**1. Introduction**

Variations in an animal’s vocalizations can provide clues about how the animal uses sound, as well as qualities of the individual that is vocalizing. Bioacoustic research depends heavily on the ability to characterize these variations. Often, variations in animal sounds are salient to human listeners. Some acoustic features are imperceptible, however, such as ultrasonic frequency modulation produced by echolocating bats, and phase shifts in audible sounds. Characterizing sounds typically involves not only describing their properties, but also relating those properties to those of other sounds. This process can involve sorting sounds based on how and when they are produced, the functional contexts within which they occur, and perceptual or acoustic similarities and differences across sounds. The most common approaches to sorting sounds include subjective sorting based on auditory quality or spectrographic features, statistical comparisons of quantified features, and computational methods for assessing how sounds are distributed within a multidimensional “feature space.” The current chapter focuses on the last of these approaches, emphasizing a computational method for sorting patterns known as artificial neural networks, hereafter referred to as *neural networks*.

**1.1 Classifying Sounds Subjectively**

Researchers typically sort animal vocalizations based on how they sound, often by comparing them to other familiar sounds. For example, sounds in the songs of humpback whales (*Megaptera novaengliae*) have been described as *snores*, *moans*, or *cries* ([see also Winn & Winn, 1978](#_ENREF_111)). Some animals even get their names from the aural impressions their sounds produce in humans, including chickadees, cuckoos, and whooping cranes, to name a few. A bioacoustician may attempt to objectively classify sounds by referring to sounds with alphanumeric labels, but even then the basis for differentiating sounds is often subjective.

Subjective classification strategies invariably involve sorting an animal’s sounds by mapping them onto pre-existing perceptual categories acquired through experience. In the case of auditory categorization, it is known that repeated experiences with particular sounds can change how individual neurons respond to those sounds ([Weinberger, 2004](#_ENREF_110)), and that changes in an individual’s neural responses are correlated with changes in the ability to perceive differences between sounds ([Recanzone, Schreiner, & Merzenich, 1993](#_ENREF_90)). Thus, how a researcher aurally classifies animal sounds will depend on how those sounds are processed by large networks of interconnected neurons, which will in turn depend in complex ways on past auditory experiences, which may be culturally determined.How it is that activity within complex networks of neurons give rise to variations in subjective impressions is not yet known.

Subjective visual classification of time-frequency representations is even more problematic. Identical sounds can be decomposed using a wide range of algorithms with various levels of temporal and spectral resolution, each of which can give rise to a variety of distinctive images. Typically, the adequacy of a particular set of visual representations is judged based on how well it corresponds to audible differences between sounds. Thus, visual classifications inherit the subjective biases of listeners while adding further, visually based perceptual subjectivity.

**1.2 Naturalistic Classification of Sounds**

When animals other than humans hear vocalizations, they almost certainly do not classify them by assigning linguistic labels to the sounds. Nevertheless, non-human animals can recognize similarities between novel and familiar sounds, and can respond to novel sounds based on these similarities ([Orduña, Mercado, Gluck, & Merzenich, 2005](#_ENREF_75)). Interestingly, at least some animals seem to be able to sort human speech and other animal sounds in ways that emulate sorting by humans ([Heimbauer, Beran, & Owren, 2011](#_ENREF_36); [Kuhl & Miller, 1975](#_ENREF_45); [Sturdy, Phillmore, Price, & Weisman, 1999](#_ENREF_102); [Sturdy, Phillmore, & Weisman, 2000](#_ENREF_103)). Such overlap in sorting strategies likely reflects similarities in the neural circuits that vertebrates use to process sound, as well as common features of natural acoustic environments. Although researchers have no ability to discern what an animal’s subjective awareness of vocalizations is like, psychophysical methods can be used to evaluate which vocalizations animals can distinguish, as well as how animals categorize sounds.

By combining such behavioral techniques with analyses of the acoustic features of vocal repertoires, bioacousticians have determined that, in some cases, animals respond categorically to sounds. For example, penguins use their vocalizations to identify and locate one another ([Aubin, Jouventin, & Hildebrand, 2000](#_ENREF_3); [Lengagne, Aubin, Jouventin, & Lauga, 2000](#_ENREF_47)). In this case, the categories “my mate” and “my parent” are discrete. In other situations, however, animals perceive sounds as falling along a continuum; graded acoustic changes are perceived as shifts in quality rather than as transitions between categories. For instance, echolocating bats gradually modify their sonar signals as they close in on a target, adjusting frequency, amplitude, and duration as a function of the intervening distance ([Simmons & Stein, 1980](#_ENREF_101)). Humans can subjectively sort bats’ sounds into categories (e.g., *chirps* or *buzzes*) by listening to recordings transformed to make the sounds audible to humans. Nevertheless, changes in the acoustic features of bat sonar sounds are often fluid and graded (i.e., changing gradually and incrementally over time), suggesting that bats process these signals as points along a continuum rather than as discrete categories of sound types ([Moss, Chiu, & Surlykke, 2011](#_ENREF_66)).

**1.3 Classifying Sounds Objectively**

Animal vocalizations can provide information about the identity, location, behavior, and relevance of the vocalizer. An observer’s past experiences (and species) will determine how rapidly and accurately this information can be extracted from a vocalization. As noted earlier, psychophysical evidence supports the idea that non-humans perceive speech sounds in ways that parallel human perception – some birds can even replicate speech, providing strong evidence that they represent features of speech that are salient for humans ([Pepperberg, 2005](#_ENREF_83)). It is clear, however, that humans do not always classify sounds in the same way as the animals that produce those sounds.

A major challenge for bioacousticians is to identify when their subjective impressions of an animal’s vocalizations are incompatible with the animal’s own impressions. For example, a series of *clicks* produced by a dolphin underwater might sound to a human like someone running a stick along a picket fence. For the dolphin producing the clicks, however, the audible click train may be much less perceptually salient than the echoes generated by the clicks. These echoes would typically be imperceptible to a human observer (i.e., the functional component of the dolphin’s signal is not subjectively evident to a human listener). Such ambiguities provide a strong rationale for using objective, computational methods to classify animal vocalizations (for a recent review of objective methods for characterizing animal sounds, see Kershenbaum et al., 2016).

In scientific research, a synonym for classification is taxonomy. Early taxonomic classifications of animals, such as the Linnean taxonomy, were based on subjective anatomical similarities ([Ereshefsky, 2000](#_ENREF_24)). Modern day taxonomies of species, however, rely more extensively on quantitative computerized analyses of DNA sequence data. Traditional subjective classification of animal vocalizations into categories can be viewed as analogous to the older Linnaean biological taxonomies in that broad domains of animal sounds are identified based on perceived similarities in structure (e.g., calls versus songs), which are then broken down hierarchically using more specific descriptors of particular sound types. Like modern DNA-based taxonomies of species, recent quantitative strategies for classifying sounds based on statistical or computational analyses use algorithms to organize clusters of similar taxonomic units rather than subjective similarities.

**1.4 Using Connectionist Models to Classify Sounds**

In the 1990s, bioacousticians began exploring new computational techniques for classifying vocalizations that went beyond basic statistical analyses. Interest in using neural networks, in particular, likely stemmed from studies of connectionist models that occurred in the late 1980s ([Rumelhart & McClelland, 1986](#_ENREF_98)). *Connectionist models* (described in more detail below) consist of sets of simple computational units that collectively perform mathematical operations on numerical inputs. These models were inspired by the structure of neural circuits and theories of neural plasticity mechanisms ([for review, see Mercado & Henderson, 2012](#_ENREF_58)). Neural networks have been used primarily to classify mammalian vocalizations, with a few researchers using them to analyze sounds of birds and frogs (see Table 1). Bioacousticians typically have used neural networks to replicate subjective classification by humans. Occasionally, neural networks have also been used to simulate classification by the animals under study. For example, Roitblat and colleagues developed neural networks that could use echoes to classify objects with performance levels comparable to those of echolocating dolphins ([Au, Andersen, Rasmussen, Roitblat, & Nachtigall, 1995](#_ENREF_2); [Roitblat, Au, Nachtigall, Shizumura, & Moons, 1995](#_ENREF_94)). Use of neural networks to classify animal sounds has expanded considerably in the last twenty years, both in terms of the vocalizing species being analyzed and in the ways in which neural networks have been applied.

**2. Neural Network Basics**

As Table 1 illustrates, researchers have used many different kinds of neural networks to objectively classify animal vocalizations. Although neural networks can vary considerably in complexity and computational capacities, they typically share some basic processing principles. In particular, all of the neural networks used by bioacousticians consist of computer programs that instantiate connectionist models in which simple processing units interact to recognize patterns.

**2.1 Simulating Neural Processing**

The computational units that comprise a neural network were developed to emulate signal processing by prototypical neurons. In neurons, multi-branched dendrites often respond to the activity of thousands of other neurons, integrate the electrical effects of this activity at the cell body, and then produce a discrete output (an action potential) when a minimum electrical threshold is reached. As noted earlier, neural networks are often referred to as artificial neural networks to distinguish them from actual circuits within brains. While some neural networks do attempt to capture what is known about brain circuits, most are only inspired by the kinds of processes that occur within brains. Nevertheless, artificial neural networks do perform computations in ways that are closer to neural processing than most statistical analysis techniques.

An output from one unit of a neural network can serve as the input of another unit. In this case, the link between units (referred to as a *connection*) is associated with a number (called a *weight*) that amplifies or attenuates the output value before it is processed by the receiving unit (Figure 1). The use of weights in neural networks simulates variations in synaptic strengths within neural circuits. Processing units within neural networks typically are also associated with a number called the *bias*. This parameter modulates the calculation that determines the output of a unit. For example, the bias value is often added to the weighted sum of all inputs before the output of the unit is computed. The bias of a unit can be viewed as analogous to the threshold that determines how much incoming activity a neuron must receive before it fires.

As in neural circuits, the inputs processed by each computational unit within a neural network often consist of outputs from other units. Typically, every unit within a neural network performs identical computations (described below). One of the fundamental computational properties of neural networks is that collections of identical, interconnected processing units can achieve functions that the individual units cannot perform alone.

**2.2 Representing Inputs: Transforming Real-World Events into Vectors**

Neural networks are essentially programs for recognizing patterns. Thus, a key step in developing a neural network is deciding the form of the patterns to be recognized. In the case of sound classification, choosing how patterns are represented requires converting selected properties of an acoustic event (e.g., measures of pressure variations over time) into a series of numbers that can be processed by a computer. Neural networks place no constraints on how a sound is digitally represented. This step can have important implications, however, for how a neural network classifies different sounds. The inputs to a neural network are conventionally represented as a set of variables *x1*, *x2*, …, *xn*, which collectively can be denoted as an *N*-by-1 vector **x**. Similarly, the set of weights that determine how a computational unit within a neural network responds to a particular input **x** can be represented with a set of variables *w1*, *w2*, …, *wn*, or an *N*-by-1 *weight vector* **w**. Typically, units within a neural network sum the weighted inputs and the bias, and then compute an output value based on a pre-specified *activation function* (for a graphical depiction of this computation, see Figure 1).

Neural networks are used to differentiate input patterns, either by detecting the presence of particular patterns or by separating patterns into different classes. The set of all input vectors can be denoted as an *N*-by-*M* matrix **X**, where *M* is the number of inputs to be classified. Each input **x** could correspond to a segment of a waveform, a checklist of detected features within a sound, measures of spectral energy, the intensities of pixels within a spectrogram, or all of the above*.* Any mix of measures is permissible as an input pattern to virtually any kind of neural network, but not all possible input representations yield equally desirable outcomes. Furthermore, the use of simpler inputs (e.g., lower dimensional vectors constrained to binary-valued elements) can greatly reduce the computational power required to implement a neural network. Consequently, a common first step in developing a neural network is to minimize the dimensionality of input vectors by converting observed measures into a smaller number of derived measures. This conversion can be done using techniques such as principal-components analysis or autoregressive models, or by simply culling or combining measures that are less likely to differentiate input patterns.

**2.3 Single-Layer Neural Networks**

A neural network’s set of connections or weight vectors (denoted by the weight matrix **W**) determines what its outputs will be for a given input. Most neural networks are deterministic, such that any input produces a corresponding, mathematically defined output. The researcher inputs a set of numbers, and another set comes out. Pragmatically, neural networks usually are implemented as lines of code that perform transformations on matrices. Many neural networks have multiple *layers*, corresponding to successive stages in which input and output vectors are processed – each layer generally corresponds to a separate weight matrix. However, even a *single-layer network* (graphically depicted in Figure 2a), based on a single weight vector, can provide a powerful way to objectively classify different inputs.The number of units in each layer, the possible values a vector element can take, the number of connections between units, and the computations each unit performs constitute the *architecture* of a neural network. Neural-network architectures can be quite complex. For example, the outputs of some units might feed back to the inputs of other units, in turn providing inputs to the units that generated those outputs. Such architectures are called *recurrent* neural networks. In some recurrent networks, the computations performed by units can depend on their own outputs. Neural networks without any recurrent connections are often called *feed-forward networks*. The architecture of a neural network determines the kinds of pattern recognition functions it can perform.

**2.4 Multi-Layer Neural Networks**

Multi-layer networks generally can perform more complex functions than single-layer networks. Numerous multi-layer architectures have been developed. The pros and cons of particular architectures are have been extensively reviewed and so are not described here (e.g., [Haykin, 1994](#_ENREF_35); [Kohonen, 1995](#_ENREF_44)). Each individual layer within a multi-layer network is usually quite similar in form (see Figure 2b) and computational capability to a single-layer network ([Dawson, 2004](#_ENREF_14), [2005](#_ENREF_15)). Processing by a multi-layer, feed-forward network is thus comparable to processing by a cascade of multiple single-layer networks. A multi-layer, recurrent network can also be viewed as a series of single-layer networks in which the outputs of some layers feed back to earlier stages of the series. The units within multi-layer networks that receive inputs from earlier layers and that provide outputs to later layers are often called *hidden units*. Hidden units can increase the functionality of a neural network relative to a single-layer network when their activation functions are nonlinear. Analyses of the weight vectors of hidden units can reveal which features of input vectors a neural network is using to classify patterns.

An appropriately chosen set of weights and activation functions is key to the usefulness of any neural network, because these properties make it possible to map almost any set of acoustic features onto arbitrary categories of sound types. In this respect, neural networks are the ultimate pattern classifiers. Activation functions are architecturally fixed. Weight vectors, however, must be independently developed for each neural network to enable it to classify the patterns of interest. Luckily, one of the advantages of modern neural networks is that one need not know the most appropriate weights or much about the inputs to create a system to perform a desired classification function, because learning algorithms have been developed to automatically adjust the weight vectors of a neural network as needed.

**2.5 Training Neural Networks**

A *learning algorithm* is a mathematically defined rule for changing weights within a neural network. A wide variety of learning algorithms are available ([Haykin, 1994](#_ENREF_35)). Some change connection weights based either on pre-specified targets (called *supervised learning*), or without pre-specified targets (called *unsupervised* learning or *self-organization*). In supervised learning, the learning algorithm is used to *train* the neural network such that when a particular input **x** is presented, the network generates a specific set of output values (often denoted as the output vector **y**). Typically, this type of learning algorithm works by repeatedly calculating the difference between the desired outputs and the actual outputs, and then modifying weights in various ways that ultimately reduce these differences. In unsupervised learning, the weights are usually modified based on the degree to which output units are active. In this case, there is typically no “correct” output state, in which case the neural network cannot make errors.

One of the first steps in developing a neural-network classifier is to choose a specific architecture and to find a learning algorithm that can be used to develop appropriate weights for that architecture. Many common neural-network architectures are associated with a specific set of activation functions as well as a specific learning algorithm, and thus their names reflect both architectural features and the learning algorithms used to find appropriate weights. For example, a prevalent supervised neural networks is the multi-layer perceptron (MLP), which has units arranged in several layers (see Figure 2b), uses a sigmoidal activation function (for details, see [Dawson, 2004](#_ENREF_14), [2005](#_ENREF_15)), and is often trained with a learning algorithm known as *backpropagation* ([Rumelhart, Hinton, & Williams, 1986](#_ENREF_97)). A widely used neural network that involves unsupervised learning is called a *self-organizing map* (SOM). In this architecture, spatially adjacent units are structured such that they respond to similar inputs (see Figure 2c), and a learning algorithm determines changes in weights based on the outcome of competition between units ([Kohonen, 1995](#_ENREF_44)). Although the details of different neural-network architectures and learning algorithms can strongly impact network performance, available software makes it possible to try several different kinds of neural networks in parallel. Thus, armed with only a set of input vectors, and possibly a set of desired output vectors, it is now relatively simple for a bioacoustician to sample various neural-network architectures to empirically determine which works best for a particular application. Descriptions of specific software packages that enable such comparisons are provided below.

**2.6 Advantages and Disadvantages of Neural Networks**

One difference between neural networks and most traditional statistical classification techniques is that some (but not all) neural networks allow nonlinear classification of data. This capability greatly expands the possible range of input to output mappings. It also enables neural networks to learn to recognize systematic variations in input patterns that are not easily detectable by human observers, and might not be revealed through statistical approaches such as clustering algorithms or discriminant analyses.

Another major advantage of neural networks is that, although they do not replicate brain circuits, they are constrained in ways that mimic neural circuits. Therefore, neural networks are more likely to reveal constraints faced by the auditory processing systems of animals than are more traditional techniques. For example, neurons in primary auditory cortex are organized tonotopically such that cells that are physically close together respond to similar frequencies. This spatial organization appears to constrain an animal’s ability to distinguish different pitches and to generalize learned distinctions to novel sounds ([Orduña et al., 2005](#_ENREF_75); [Thompson, 1965](#_ENREF_107)). Neural networks that are spatially organized in ways that mimic the organization of auditory cortex (like some SOMs) can better characterize sounds in ways that predict an animal’s ability to distinguish sounds, as well as how animals are likely to generalize learned responses to sounds.

The variety and flexibility in possible neural network architectures makes them highly customizable and powerful. However, this flexibility also complicates neural network analyses because researchers must decide which of many possible architectures and parameters to apply to a particular classification taskwithout any clear guidelines for selecting an appropriate network. Another limitation is that, for more complex neural networks, it is often difficult to determine how the network classifies sounds after training. Finally, because the set of weights developed by a neural network depends on both the initial weight settings and the specific inputs used to train the network, classifications based on neural networks may be more difficult to replicate than are traditional statistical analyses. Even with a single data set and using a fixed architecture, ten different implementations of a neural network may produce ten different sets of connection weights, with each neural network performing slightly—or even very—differently.

**3. Using Neural Networks to Sort Vocalizations**

In the following sections, we illustrate how neural-network classifications can provide new insights into the structure and function of sounds produced by chickadees (*Poecile atricapillus*), false killer whales (*Pseudoorca crassidens*), and humpback whales (*Megaptera novaeangliae*), as well as how neural networks can be used to supplement subjective sorting of sounds by human observers. Although we focus on sounds produced by birds and cetaceans (whales and dolphins), the classification techniques described are equally applicable to other animals.

**3.1 Categorizing Chickadee Call Sounds with a Multi-Layer Neural Network**

As noted above, chickadees were named for the characteristic call they produce ([Ficken, Ficken, & Witkin, 1978](#_ENREF_25)). This call consists of four distinctive sounds (called *notes*), which researchers have labeled A, B, C, and D (Figure 3), respectively. Chickadees produce these four sound types in a predictable order, but vary the number of times they produce each type, and sometime exclude one or more. Consequently, they can generate a wide variety of note sequences in their calls ([Hailman, Ficken, & Ficken, 1985](#_ENREF_33)). Behavioral evidence suggests that each of the four sound types conveys different information ([Freeberg & Lucas, 2002](#_ENREF_28); [Nowicki, 1983](#_ENREF_73)), and that chickadees recognize when these notes are produced in atypical orders ([Charrier & Sturdy, 2005](#_ENREF_9)). These various notes are subjectively classifiable by human listeners, raising the question of whether they might also constitute natural auditory categories for chickadees.

Early efforts to address this question focused on classifying notes through visual inspection of spectrograms ([Bloomfield, Charrier, & Sturdy, 2004](#_ENREF_4); [Ficken et al., 1978](#_ENREF_25); [Nowicki & Nelson, 1990](#_ENREF_74)). Salient features of spectrographic images that distinguish the four types have often been used to subjectively categorize the notes. The ability of chickadees to distinguish the notes has been verified through behavioral experiments ([Charrier, Lee, Bloomfield, & Sturdy, 2005](#_ENREF_8); [Guillette et al., 2010](#_ENREF_32); [Sturdy et al., 2000](#_ENREF_103)). More analytical approaches to classifying chickadee sounds have emulated this approach of visually sorting spectrograms. For example, spectrograms have been statistically analyzed to reveal similarities and differences in time-varying features ([Nowicki & Nelson, 1990](#_ENREF_74)). Such analyses may have inherited some of the subjective biases of human listeners, because the representativeness of the sounds within the sample being analyzed have typically been balanced to include approximately equal numbers of each note type (i.e., the sounds were subjectively pre-sorted, which can increase the prevalence of prototypical exemplars in the sample).

Another approach to classifying chickadee notes involves directly measuring properties of spectrographic images, and then using those measures to summarize the properties of individual sounds. Such quantitative summaries can be statistically analyzed to reveal acoustic dimensions that differentiate subjectively identified sound types. In the case of chickadee notes, these dimensions include duration, starting frequency, ending frequency, and so on. Standard statistical analysis techniques, such as linear discriminant analyses, multidimensional scaling, principal-components analysis, and cluster analysis can be used to sort these summary representations. The outcomes of such analyses can then be directly compared to the results achieved by humans visually sorting spectrograms to determine which approach agrees most closely with auditory categorization of by expert human listeners. This technique can potentially provide insights into the features that humans use to classify bird sounds, but may not be capturing the features or distinctions that are relevant to chickadees.

To assess whether statistical analyses or neural networks might better characterize how these birds sort the notes, these two approaches were compared to the actual performance of chickadees in auditory discrimination tasks ([Guillette et al., 2010](#_ENREF_32); [Nickerson, Bloomfield, Dawson, & Sturdy, 2006](#_ENREF_72)). Calls were recorded from chickadees housed in individual cages. Spectrograms of notes were visually classified, and nine acoustic features characterizing spectral, temporal, and spectrotemporal features of notes were measured from each spectrogram ([see Dawson, Charrier, & Sturdy, 2006, for details](#_ENREF_16)). A multilayer perceptron with 9 input units, 2 hidden units, and 3 output units was then trained to classify notes as being of type A, B, or C~~;~~ (D notes were not included in the study). This neural network was trained with summary measures from 370 notes in all. Recall that training of a multilayer perceptron involves supervised learning. In other words, the goal of training is for the network to assign every note to a predetermined type.

After training, the multilayer perceptron correctly classified all but 5 of the sounds (98.6% correct). Examination of the two hidden units in this neural network revealed that one unit became tuned to features of A notes, whereas the other was most responsive to notes other than C. Essentially, the network learned to focus on features of frequency contours that were unique to A notes, and on these and other acoustic features that were absent from C notes. Using these two “feature detectors” in concert enabled the network to accurately distinguish the three note types. The fact that both hidden units contributed to the final classification of notes illustrates the distributed nature of the representational schemes learned by neural networks. Discriminant analyses of these same summary measures from spectrograms produced a similar accuracy level (95%). Thus, either approach can potentially replicate the note-discrimination performance of a chickadee.

Whether auditory categorization by a chickadee is more similar to that of a nonlinear, neural-network classification process or of a linear, statistical process remains to be seen. Current behavioral data suggest that chickadees do attend to many of the same spectrotemporal acoustic features that the neural network learned to focus on when classifying notes ([Guillette et al., 2010](#_ENREF_32); [Nickerson et al., 2006](#_ENREF_72)). For instance in recent experiments, chickadees appeared to treat A, B, and C notes as falling along a perceptual continuum. A notes were only confused with B notes, whereas B notes were confused with both A and C notes ([Charrier et al., 2005](#_ENREF_8)).

**3.2 Sorting Gradations in False Killer Whale Sounds with a Self-Organizing Map**

Traditional approaches to characterizing and classifying the sounds of cetaceans are similar to those described above for chickadee call notes. Many cetaceans, however, do not produce calls that can be reliably categorized based on subjective judgments. Instead, they often produce streams of sounds that vary along multiple acoustic dimensions in idiosyncratic ways. Their vocalizations typically have been broadly classified in terms of their aural properties. Short duration broadband sounds are traditionally called *clicks*, whereas longer duration tonal sounds are called *whistles*. Cetaceans may use their sounds for communication or echolocation, and often it is claimed that different functions map directly onto subjectively identified auditory categories. In particular, whistles are assumed to be for communication and clicks are classified as echolocation signals. For a few cetacean species, such as orcas (*Orcinus orca*), extensive efforts have been made to identify and label stereotypical “notes” or call types ([Deecke, Ford, & Spong, 1999](#_ENREF_17)). For most cetaceans, however, no such dictionary of sound types is available.

False killer whales fall into this latter category. Sounds produced by these animals have been described and quantified ([Kamminga & van Velden, 1987](#_ENREF_43); [Murray, Mercado, & Roitblat, 1998a](#_ENREF_67)), but standardized nomenclature has not yet been developed. Relatively little is known about how false killer whales use the sounds they produce. It is possible to aurally and visually sort false killer whale vocalizations into the traditional categories of whistles and clicks. Undoubtedly, statistical correlates of these subjective distinctions could also be identified. However, analyses of the acoustic structure of false whale killer vocalizations provide little justification for dividing their sounds into discrete types ([Murray et al., 1998a](#_ENREF_67)). The animals transition readily and instantaneously across the boundaries of aurally-based categories while maintaining spectral properties and without obvious changes in behavior or state. For example, a train of short duration broadband pulses can gradually transform into a sinusoidal waveform ([Murray et al., 1998a](#_ENREF_67)). Graded transformations of this sort are more comparable to transitions within streams of bat sonar signals than to the switches between stereotyped notes within chickadee calls. Such flexible modulation of acoustic features by false killer whales raises the question of whether human impressions of these sounds bear any relationship to the distinctions a false killer whale perceives.

Using neural networks provides one way of circumventing observer biases when classifying false killer whale vocalizations. For example, individual sounds produced by false killer whales can be sorted using an unsupervised network such as an SOM. Murray and colleagues ([1998b](#_ENREF_68)) used this approach to classify sounds produced by captive false killer whales recorded at close range. Individual sounds were transformed into vectors corresponding to acoustical measurements from short-duration samples (or *frames*) extracted from each vocalization. The SOM was then trained using vectors representing each individual sound. The advantage of this approach was that the neural network classified vocalizations based on the prevalence of certain acoustic features within the sounds, and thus was not constrained by the perceptual impressions of observers. Also, because the SOMs are trained with an unsupervised learning algorithm, they are not limited to a predetermined number or variety of sound categories.

After training, each input was automatically assigned to the one unit in the SOM with a weight vector that was most similar to it. The extent to which the SOM effectively identified sounds within the repertoire was then evaluated by measuring how different the features of individual sounds were from the weights of the unit representing that sound. A set of 500 vocalizations was analyzed. Half the sounds was used to train the SOM, with the rest used to test its generality. The SOM itself consisted of a 5×5, two-dimensional map, initialized to random weights and trained with Kohonen’s (1995) competitive-learning algorithm.

The topography of the resulting map represented the set of vocalizations as in terms of temporal sequences of spectral and waveform features. Figure 4 shows an example of the spatial distribution of prototypical sound features learned by an SOM. The distribution of vectors across the map summarizes the distribution of features within the set of vocalizations analyzed. Units on one side of the map became sensitive to tonal sounds with increasing frequency modulation (i.e., ascending whistles). Six of the twenty-five units showed this sensitivity, attesting to the prevalence of sounds with these features. On the opposite side, six units were responsive to constant frequency pulse trains (i.e., high-frequency clicks occurring in series). These modal categories corresponded closely to subjective impressions of false killer whale sounds. The remaining units, however, represented subjectively less salient classes of sounds, ones that traditionally have been described as *mixed*. The map thus makes clear that such sounds are more prevalent than might be expected based on subjective impressions, and that many sounds have overlapping acoustic properties, consistent with graded modulation of acoustic structure. The SOM classified the 250 novel sounds in similar fashion, showing that the categories developed by the map were representative of the entire sample of sounds produced by false killer whales.

SOM-based classifications of false killer whale vocalizations suggest that subjective human impressions of these sounds are anchored by the two most prevalent types, which consequently influence interpretations of intermediate and non-modal versions. For example, a researcher might describe a vocalization containing features of both modal sound types as, “An initial whistle, followed by clicks,” despite the fact that the sound was acoustically continuous. The fact that not all units within the map corresponded to subjectively salient sound “types” with distinctive perceptual features could be viewed as a methodological weakness. Alternatively, the transparency of the arbitrariness with which sounds were divided into “types” by SOMs could be viewed as a strength, because it highlights the fact that human subjective categories represent only one possible partitioning of an acoustic feature space among many possibilities ([see also Lachlan, Peters, Verhagen, & ten Cate, 2010](#_ENREF_46)). Although the number of “categories” within an SOM and even the topography may vary across simulations, the distribution of sounds across units remains systematic and reflects the structure of the input set. In contrast, human auditory categories can be biased in culturally-determined ways by past subjective experiences. Use of SOMs does not entirely protect against such biases, because the selection of features used as inputs to the SOM may incorporate the subjective biases of the user regarding which elements are most relevant. Selective sampling and measurement of features that are salient to human observers (either aurally or visually in spectrograms) can result in SOMs that are organized in ways that simply recapitulate the subjective impressions of the observers.

**3.3 Describing the Temporal Dynamics of Humpback Whale Songs with Self-Organizing Maps**

Characterizing the structural organization of vocal sequences traditionally has been considered to be less challenging than classifying individual sounds of animals. A common approach is to describe sequences using a series of symbols, each of which denotes the category for an individual sound in the sequence (Kershenbaum et al., 2016). For example, the seven-note chickadee call shown in Figure 3 can be symbolically described as AABCDDD. It is not straightforward, however, to establish which structural features of vocal sequences are perceived by an animal. Just as bioacousticians cannot be certain that their impressions of differences in sound types are comparable to those of other animals, they also cannot be certain that their descriptions of structural properties of sequences reflect how animals organize the sequences. If chickadees only encode changes in notes, for instance, then a more biologically-relevant representation of the note sequence shown in Figure 3 might be 111, where each one indicates the detection of a note change. If chickadees instead live only in the present ([Roberts, 2002](#_ENREF_93)), retaining no memory of a note once it is over, then they would lack the ability to perceive sequences. In that case, researchers’ descriptions of chickadee call structure would differ radically from how chickadees represent calls. Bioacousticians can use neural networks to more objectively characterize how animals vary individual sounds over time and to detect non-obvious regularities within sound sequences. In the following, we illustrate how neural networks can be used to describe the temporal dynamics of sound sequences, focusing specifically on analyses of humpback whale songs.

**3.3.1 Humpback whale song content and structure**

Humpback whales may sing continuously for hours ([R. S. Payne & McVay, 1971](#_ENREF_82); [Winn & Winn, 1978](#_ENREF_111)). When singing, they rhythmically produce hundreds of sounds in repeated patterns. Neither the sound repertoire used to construct songs nor the temporal or spectral features of sounds within songs appear to be fixed ([K. Payne & Payne, 1985](#_ENREF_80)). Both can change dramatically over the span of a few years. No other mammals are known to modulate the structure and content of their vocalizations throughout their lives to such a great extent.

Developing a scheme for classifying such highly variable sequences is quite difficult. Researchers traditionally have analyzed structural patterns within humpback whale songs by first converting recordings into sequences of letters or labels, and then by measuring features of these symbol sequences ([Miksis-Olds, Buck, Noad, Cato, & Stokes, 2008](#_ENREF_64); [R. S. Payne & McVay, 1971](#_ENREF_82); [Suzuki, Buck, & Tyack, 2006](#_ENREF_104); [Winn & Winn, 1978](#_ENREF_111)). Occasionally, temporal regularities within songs also have been taken into account ([Handel, Todd, & Zoidis, 2009](#_ENREF_34); [Mercado, Herman, & Pack, 2003](#_ENREF_60)). Past analyses have revealed that the sound sequences produced by singing humpback whales are periodic at multiple time scales. Cyclical features are evident seasonally, daily, and within multi-hour sessions. Repeated patterns are also seen within songs across durations ranging from minutes to milliseconds (for details, see [Cholewiak, Sousa-Lima, & Cerchio, 2013](#_ENREF_10); [K. Payne, Tyack, & Payne, 1983](#_ENREF_81)).

**3.3.2 Classifying sounds within humpback whale songs**

Most past approaches to describing humpback whale songs rely heavily on the subjective impressions of observers. When quantitative approaches have been used, the emphasis has remained on classifying sounds into discrete categories as a first step in creating symbol sequences to describe repeating patterns within songs ([Helweg, Cato, Jenkins, Garrigue, & McCauley, 1998](#_ENREF_38); [Suzuki et al., 2006](#_ENREF_104); [Walker et al., 1996](#_ENREF_109)). This approach treats all sounds of a designated type as being equivalent and discards most information about acoustic differences between sound types. For instance, if a low intensity sound lasting 100 ms is symbolically designated as *a*, and a subsequent high intensity sound lasting 10 s is designated as *b*, then the symbolic sequence *ababab* obscures the fact that the whale spent much more time belting out *b*s. Furthermore, recent analyses suggest that the repertoire of sounds used by singing humpback whales is graded, including many sounds that vary along multiple dimensions ([Mercado, Herman, & Pack, 2005](#_ENREF_59); [Mercado & Kuh, 1998](#_ENREF_61); [Mercado, Schneider, Pack, & Herman, 2010](#_ENREF_63); [Winn & Winn, 1978](#_ENREF_111)).

As with the sounds of false killer whales, the structure of sounds within humpback whale songs suggests that sorting individual sounds into perceptually-based categories is not warranted and may actually obscure the temporal structure of songs ([Mercado & Handel, 2012](#_ENREF_57)). Most traditional sound-classification approaches are ill-suited for describing sounds that grade continuously from one type into another. SOMs have no problem classifying graded repertoires, however, because they construct spatial representations of overlapping sound “types” such that units that are physically close within the map respond to similar inputs. Graded similarities and differences between sounds can then be quantitatively described in terms of the physical distances between SOM units, or by the numerical differences between their weight vectors.

In the SOM analyses of false killer whale vocalizations described above, each vocalization was represented as a vector of measures from sequentially sampled frames. Similar approaches can be used to classify individual sounds produced by singing humpback whales ([Mercado & Kuh, 1998](#_ENREF_61); [Pace, Benard, Glovin, Adam, & White, 2010](#_ENREF_76)). Alternatively, an SOM can be trained to classify individual frames from sounds rather than individual sounds. Figure 5 shows a 10×10 SOM trained with 5,516 frames (each 50-ms long) from 248 individual sounds automatically extracted from a single, 10-min long humpback whale song. Linear predictive coding (LPC) was used to estimate the frequency content of each frame. LPC spectra were then transformed using a computational model of transduction by a humpback whale cochlea to emphasize spectral information that is likely to be available to humpback whales ([Branstetter & Mercado, 2006](#_ENREF_5); [Mercado, Green, & Schneider, 2008](#_ENREF_56)). These transformed spectra were used as inputs to train the SOM.

Using this approach, any individual sound within the humpback whale song can be characterized in terms of the sequence of SOM units it activates. Note that no unit in this SOM corresponds to a named sound type. In principle, the set of frames from an individual sound could activate every unit in the map or might only activate a single unit. Given that humpback whales must process sounds in real-time as they are hearing the song, this type of frame-based decomposition may more closely approximate the auditory computations that listening humpback whales employ. This approach also avoids distortions that can be introduced when symbols are substituted for sounds as a first step in analyses of sequential structure (Kershenbaum et al., 2016). Comparisons within and across individuals, populations, seasons, and contexts can be made by training SOMs with frames collected from sounds recorded from different whales at different times.

**3.3.3 Classifying sequences within humpback whale songs**

Activation trajectories across the frame-based SOM shown in Figure 5 can be described in the same way that properties of frequency contours for chickadee notes were measured from spectrograms in the simulations described above. Alternatively, one can introduce an additional SOM that analyzes patterns of activity in the first. For instance, sequences of unit activation can be represented as the number of times each unit within the larger, first map was activated, as well as the transitions that occurred across different quadrants of the map (see Figure 6). The number of unit activations characterizes the duration and spectral content of a sound, whereas unit transitions describe vocal trajectories over time. In this approach, the second, 5×5 SOM sorts individual sounds based on the distribution of features detected when frames from a single sound are sequentially input into the 10×10 SOM. Figure 7 shows the result of simulations using this dual SOM-approach. The topographic structure of the 5×5 SOM revealed modal sound types corresponding to several that had been proposed based on subjective impressions, as well as other sounds that were less easy to subjectively classify. The most identifiable prototypes invariably corresponded to sounds that occurred frequently, indicating that human subjective categories reflect prevalent sound features.

An unexpected finding from the dual-SOM simulation was that the topographic structure of the 5×5 map reflected the order of individual sounds within the single song being analyzed (Figure 7), despite the fact that no information about the order of sounds (or frames) had been provided to either SOM. Certain sound types were evident throughout the song, whereas others occurred only at certain stages of the song. Apparently, in learning to distinguish sounds within the song, the 5×5 SOM picked up on a gradual modulation in some acoustic features of a subset of the sounds within this song that were changing systematically throughout the 10-min recording ([see also Mercado & Handel, 2012](#_ENREF_57)). This finding illustrates how neural networks can reveal structure within sequences that may not be subjectively obvious, and that would be impossible to detect through analyses of symbol sequences.

The graded changes in the sounds within humpback whale songs raise important questions about how songs might best be characterized. For example, should sound categories be defined based on the absolute acoustic features of these sounds (e.g., high versus low frequency content, short or long duration). Or, might these features be less relevant than the pattern of acoustic changes within sounds, analogous to words sung at different pitches and rates? There has been heated debate among cetacean researchers about how much a vocalization should be pre-processed or normalized before attempting to match it to categorical prototypes (e.g., [Buck & Tyack, 1993](#_ENREF_6); [Janik, 1999](#_ENREF_41); [McCowan, 1995a](#_ENREF_51), [1995b](#_ENREF_52)). The basic issue concerns the degree and kinds of similarity needed for two sounds to be considered instances of the same category. This issue also arises when analyzing sound patterns. The song “Row, row, row your boat” is recognizable if sung by a 3-year-old in English, an 80-year-old in Chinese, or a parrot--despite the fact that the individual sounds would have quite different acoustic properties. In this case, one recognizes and classifies the patterns of changes in pitch and timing across sounds, not the absolute acoustic features of the sounds. Humpback whales may do something similar when producing and perceiving species-typical sound sequences.

To assess this possibility, SOMs comparable to those described above were used to analyze input vectors corresponding to measures of the absolute features of individual song sounds, relative features across consecutive sounds (e.g., the change in frequency content between sounds), or both ([Green, Mercado, Pack, & Herman, 2007](#_ENREF_31)). Statistical analyses of transitional probabilities in SOM unit activation were then performed to assess the number of repeated patterns revealed by each of these maps. Surprisingly, the SOM trained only with relative features of sounds (and no absolute features) identified the most repeated patterns within songs, suggesting that relative changes were stable even when absolute features changed over time. This pattern of results held not only for songs produced within a single season, but also when SOMs were trained on songs from multiple years in which sound patterns change progressively over time ([Green, Mercado, Pack, & Herman, 2011](#_ENREF_30)). The observed patterns of relative change occurred in parallel along multiple acoustic dimensions, and were therefore more complex than the shifts in rate or key that are typical of musical sequences. Given that the absolute acoustic features of humpback whale song sounds are highly salient to human listeners, it is unlikely that these stable patterns of relative change would have been discovered without the use of neural networks.

**4. Using Neural Networks to Understand How Animals Classify Sounds**

Many past neural-network analyses of animal vocalizations have focused on sorting sounds into different types corresponding to subjective labels, individuals, or behavioral situations. Neural networks can also be used, however, to emulate how animals themselves classify their sounds. Neural networks are especially useful for such simulations when an animal’s capacities are changing over time, such as during development or when the animal is learning about sounds. In the following sections, we discuss some recent examples of how neural networks can be used to quantitatively instantiate and test hypotheses about how nonhuman animals classify sounds.

**4.1 Simulating Distance Estimation by Whales with a Single-Layer Perceptron**

Many applications of neural networks in bioacoustics, including all of the examples discussed above, involve using networks to identify different classes of vocalizations. Another approach, however, is to classify features of animal sounds that are not produced by the vocalizer ([Datum et al., 1996](#_ENREF_13); [Neti et al., 1992](#_ENREF_71); [Wotton & Jenison, 1997](#_ENREF_113)). For example, neural networks have been used to classify echoes from the clicks generated by dolphins ([Roitblat et al., 1995](#_ENREF_94); [Roitblat et al., 1993](#_ENREF_95); [Wisniewski, DeLong, Heberle, & Mercado, 2013](#_ENREF_112)). In those simulations, the relevant information was not the properties or types of clicks produced, but the properties of the reflected echoes that provided information about the object to the dolphins. Essentially, the neural networks learned to recognize features of echoes that were a consequence of differences in the material, shape, and size of the objects being echolocated. Often, the acoustic features that are correlates of such object differences are subtle and may not be salient to human listeners ([however, see DeLong, Au, Harley, Roitblat, & Pytka, 2007](#_ENREF_19)).

This neural-network approach is not limited to processing echoes from sonar signals. It can also be used for tasks such as assessing changes in the agitation of individuals ([Manteuffel et al., 2004](#_ENREF_48); [Schon et al., 2001](#_ENREF_99)), separating sounds from noise ([Erbe, 2000](#_ENREF_23); [Mellinger & Clark, 2000](#_ENREF_54)), and to extract information from sounds that reveals the size or location of the vocalizing animal. One piece of information of particular relevance to many animals is the distance of the vocalizing animal from the receiver. For birds, this information is pertinent for maintaining the integrity of territories and for locating mates ([Naguib & Wiley, 2001](#_ENREF_69); [Nelson & Stoddard, 1998](#_ENREF_70); [Phillmore, Sturdy, Ramsay, & Weisman, 1998](#_ENREF_85)). For cetaceans, it is critical for assessing the spatial distribution of conspecifics underwater, often involving relatively long distances and little or no available visual information.

For example, humpback whales do not maintain territories, and travel thousands of miles during yearly migrations. Individual whales are often separated by several kilometers. During the winter, males appear to actively compete for females ([Clapham, 1996](#_ENREF_11)). In this context, it is important that males be able to localize likely competitors, and that both males and females be able to determine the locations of possible mates from more than a kilometer away. Because whales cannot see each other underwater at this distance, they must gain spatial information through hearing. Little is known about how whales do this, but simulations and field experiments suggest that propagation-related distortion of the spectral content of vocalizations plays an important role ([Mercado & Frazer, 1999](#_ENREF_55); [Mercado et al., 2008](#_ENREF_56); [Mercado et al., 2007](#_ENREF_62)).

To assess whether such distortions might be perceptible to humpback whales, sounds were recorded in humpback whale habitats after having traveled various distances ([Mercado et al., 2007](#_ENREF_62)). For this data set, the distance each sound traveled was known, and the question addressed was whether spectral distortions provide sufficient cues for a neural network to learn to determine how far a sound had traveled ([Mercado et al., 2008](#_ENREF_56)). Two types of neural networks, a multi-layer perceptron (containing 8-12 units in the first layer, 3 hidden units, and 1 output unit) and a single-layer perceptron (containing 8-12 units), were trained to classify recorded sounds based on distance traveled. Inputs to the network represented the frequency content of the recorded sounds. Input vectors with 6688 elements representing spectral energy were reduced to a smaller set (8-12) of values representing the coefficients of the principal components required to account for 90% of the variance within all the input vectors.

Training and testing of the networks was similar to the procedures described above for classifying chickadee call notes. In this case, however, the output of the network was not a subjectively chosen label, but a value corresponding to the distance between source and receiver. Both neural-network architectures were successful at sorting received sounds, showing that transmission-related distortion of spectral cues were sufficient for estimating distance to the sound source. These simulations also showed that the networks’ accuracy depended on the acoustic properties of the sounds used as inputs. In other words, it was easier for the network to estimate source distance for a subset of sound types. This finding suggested that singing humpback whales might be able to acoustically modulate the acoustic features of their sounds to control how easily a listening whale could judge its distance from the singer. In addition, the simulations suggested that different sounds varied in their localizability for specific distances, meaning that a singing whale might be able to produce sounds that were maximally localizable by listeners at particular distances. Thus, the simulations not only established that distance cues were present within received whale sounds, but also generated several new hypotheses about how a vocalizing whale might modulate properties of its sounds to facilitate or hamper the abilities of listening whales to judge its location. Such subtle variations in the sounds recorded from humpback whales are not likely to be salient to humans, but may be highly salient and functionally relevant for a humpback-whale listener.

**4.2 Simulating Note Discrimination Learning by Chickadees with a Single-Layer Perceptron**

A neural network trained to classify some animal sounds can also be used to classify other similar sounds. By comparing classifications of novel sounds by neural networks to classifications by animals, one can assess whether the networks and animals are using similar cues ([Nickerson et al., 2006](#_ENREF_72); [Wisniewski et al., 2013](#_ENREF_112)). In analyses of chickadee notes, simulations have been conducted in which networks were trained to make the same note distinctions that chickadees learned to perform in the laboratory ([Nickerson et al., 2006](#_ENREF_72)). Afterward, generalization to new notes was tested for both chickadees and networks.

In these behavioral experiments, chickadees were trained to discriminate two note types by responding after hearing one but not the other. For example, a bird might be trained to respond to B but not A, or to A but not B. To simulate this training task, single-layer perceptrons were trained to distinguish 9-element vectors summarizing features of spectrograms (Nickerson et al. 2006). Each perceptron had 9 input units and 1 output unit. Networks were trained with 40 patterns: 20 exemplars of one of the two note types and 20 exemplars of the second type. Like the birds, the neural networks were trained to respond to only one of the two note types, and had little difficulty learning this discrimination task.

After training, both birds and the networks were tested with call notes that had been frequency shifted up or down by 0.5-2.5 standard deviations. Birds responded to these shifts by classifying the notes differently. For example, when the A note was shifted downward, chickadees were more likely to classify it as a B note. Conversely, if a B note was shifted up in frequency, birds were more likely to respond to it as an A note. Figure 8 shows that single-layer perceptrons displayed the same basic pattern of generalization. The overall correlation between generalization in the chickadees and the neural networks was 0.69. In both cases, it appeared that absolute pitch was a strong determinant of how a note would be classified.

Another interesting similarity between chickadee and neural-network performance was that the time required for a network to learn a particular discrimination paralleled the time the birds needed for the same discrimination ([Guillette et al., 2010](#_ENREF_32)). For example A vs. B discriminations required many more training trials for both birds and neural networks than did B vs. C distinctions*.* These differences in difficulty likely reflect relative similarities among note types. More similar note types were harder to distinguish, increasing the time required to learn their distinctive features. This finding suggests that the properties of call notes that made discrimination easy or difficult for chickadees were captured by the spectrographic measures used. Differences in learning rate also provide support for the idea that adjacent notes fall along a perceptual continuum. Note types that occurred farther apart in the calls (e.g., A-C) were easier to distinguish than did those occurring close together (e.g., A-B, or B-C).

A final test involved notes that had not been included in the training set, to determine whether generalization patterns reflected evident acoustic similarities among calls. For novel sounds, both birds and networks generalized differently, depending on the sound pair they had been trained with. Birds trained to respond to B but not to A responded more strongly to C (a novel sound) than to B (the familiar sound). Birds trained to respond to B but not C~~,~~ showed strong responses to A when it is was first presented. Single layer perceptrons showed the same pattern of generalization when tested with vectors corresponding to untrained note types.

Collectively, these simulations suggest that single-layer perceptrons trained with spectrographic measures of frequency contours can provide a reasonable approximation to classification by chickadees, at least in the sense of showing similar sensitivities to differences among call notes and similar generalization of learned responses. Furthermore, these simulations illustrate that neural networks can not only be useful for objectively sorting animal sounds, but also for modeling how species members themselves might classify and respond to these sounds. When the networks made errors with chickadee notes, those mistakes were similar those the birds made in experimental tasks. Thus, the neural networks used provided a way to predict chickadeeperformance before training and testing ([Guillette et al., 2010](#_ENREF_32)), an approach that is likely applicable across a wide variety of species and sounds.

**5. Conclusions**

**5.1 Summary**

Humans naturally classify sounds into categories, often using descriptive labels that emphasize salient features of the sounds. Bioacousticians have supplemented these subjective sorting strategies with statistical analyses of precisely measured physical features of animal sounds. Using neural networks to classify such sounds extends this process one step further, facilitating quantitative, descriptive analysis of both individual sounds and sequences of sounds. Critical aspects of developing a neural network include choosing a particular architecture, converting measurements to input vectors, and training the network to recognize a particular set of inputs. When the goal is to sort vocalizations based on defined criteria, supervised learning algorithms make it possible for a neural network to do this with high accuracy and speed. When the goal is to sort vocalizations based on similarities between measured properties, unsupervised learning algorithms can be used to create networks that objectively sort sounds or quantify sequential properties of sounds. Neural networks can also be useful for understanding how animals might themselves classify the sounds they hear and for developing specific testable hypotheses about the functions of these classifications.

**5.2 Practical Advice**

Neural networks have often been used as an alternative to sound classification by human observers. Using networks in this way can greatly speed up analysis of large data sets. However, it is a mistake to view this approach as increasing the objectivity of sound classification. If a network is designed to assign sounds to predetermined subjective categories through selective training with prototypical sounds and tweaking of network parameters, then the network is likely to inherit the subjective biases of those researchers that developed the categories. In this context, networks typically serve as automated spectrogram sorters. Choosing signal processing strategies and input representations that parallel those used by human observers can increase the accuracy with which networks match human performance ([e.g., Deecke & Janik, 2006](#_ENREF_18)). If the goal is to develop a neural network system that classifies sounds in biologically meaningful ways, however, then designing the system to mimic human subjective biases may be counterproductive.

When deciding whether to attempt to use a neural network for a sound classification problem, a simple approach is to search the literature for studies that have solved similar problems (see Table 1). It is worth looking beyond journals devoted to animal research. Papers describing the most advanced applications of networks have been published in the fields of music, speech, underwater acoustics, signal processing, and neuroscience. Typically, multiple network architectures can achieve comparable performance levels. Consequently, it is worth searching for the simplest approaches to maximize the transparency of the methods. Selecting an appropriate input representation is probably more important than the particular kind of network used. For instance, pre-processing sounds within a sample to remove acoustic features that are not subjectively salient (e.g., extracting frequency contours from sounds) radically constrains the kinds of classifications networks can learn to perform. Experimenting with different input representations can clarify which features are most relevant for a particular classification task (e.g., [Mercado et al., 2008](#_ENREF_56); [Wisniewski et al., 2013](#_ENREF_112)).

If the goal is to create a tool for automated classification, running multiple simulations with different initial weight settings, parameter values, numbers of layers, and numbers of units may be needed to find one with the desired levels of performance. The size of the input set is less critical than whether the inputs are representative of the full range of input features and recording conditions. Neural networks typically make no distinction between vocalizations, background noise, and artifacts, using all available cues to sort different patterns. Increasing the variety of samples can decrease the chances that the network learns to use features other than vocalizations to sort sounds. If features vary systematically across inputs (e.g., certain sound types are only produced when an animal is high in a tree, and others only when near the ground), network may learn to recognize input patterns beyond those that would be obvious to a human observer. Use of such features can cause the network to misclassify some sounds that subjectively seem to be unambiguously members of a particular category.

Ideally, a trained network will be able to classify sounds beyond the specific examples it was trained to classify. In some cases, the ability of a network to generalize across similar sounds can be increased by adding an additional layer of processing with a smaller number of units than an earlier stage of processing. Halting training before the network achieves the highest levels of performance can also sometimes increase its ability to successfully classify novel sounds. Finally, adding random noise to inputs during the training stage can increase generalization performance (for other useful tips on development neural network classifiers, see Montavon, Orr, & Müller, 2012).

**5.3 Future Prospects**

The simulations described above involved relatively simple neural-network architectures (perceptrons and SOMs). Many more sophisticated networks and learning algorithms are available, including radial-basis networks, multi-layer, recurrent networks, and more ([for review, see Ripley, 2008](#_ENREF_92)). As illustrated in Table 1, some of these other techniques are also beginning to be applied to bioacoustic analyses. There has been a recent surge of interest in “deep learning” approaches to neural networks (LeCun, Bengio, & Hinton, 2015), which basically involves using multiple layers of hidden units that automatically extract features from raw input data when learning to classify patterns. Deep learning approaches are particularly useful for analyzing large data sets. For example, they can be used to automatically segment and classify bird songs from audio recordings (Koops, van Balen, & Wiering, 2014). No one neural network approach is ideally suited to all sound-classification tasks ([Haykin, 1994](#_ENREF_35)). Even for a single architecture, investigators must make decisions about the number of units, number of layers, initial settings, and constraints on adaptive changes. This large array of options makes standardization of neural-network classification approaches problematic.

One way to address this issue might be to use neural networks that adaptively “design themselves,” for example by choosing the number of units and layers based on the demands of the classification problem ([Heinke & Hamker, 1998](#_ENREF_37)). If standardized criteria for “growing” or “evolving” neural-network classifiers can be developed, then the various parameters used in the neural-network classification of various animal sounds will become an empirical question rather than a more arbitrary methodological choice. Neural-network developers have made advances along these lines ([do Rego, Araujo, & de Lima Neto, 2010](#_ENREF_22); [Marsland, Shapiro, & Nehmzow, 2002](#_ENREF_49)), but as of yet there are no standardized algorithms or architectures for self-designing networks.

Bioacousticians attempting to classify animal sounds historically have often tended to rely on analysis approaches originally developed for human speech. Speech-analysis techniques generally are more technically sophisticated than analyses applied to animal sounds, which is also true for neural-network architectures typically used in each case ([Juang, Lin, & Tu, 2010](#_ENREF_42); [Wysoski, Benuskova, & Kasabov, 2010](#_ENREF_114)). A major difference between classification of speech versus animal sounds is that the former allows clear performance standards, which is not possible for the latter. Network classification of bioacoustic material can be measured relative to a human’s perception, but there is often no way to objectively assess how reliable and valid that listener’s performance actually is. Assessing network performance might be possible when the goal is to identify individual animals by listening, or to localize a vocalizing animal. If the goal is to sort vocalizations based on their inherent qualities, however, basically all that can be determined is whether multiple human listeners apply consistent criteria and produce similar classification outcomes. A neural-network classifier that sorts animal sounds differently from a human listener might be performing poorly, or instead be classifying the sounds as the study species would. An important direction for future research is to develop techniques to directly compare the sorting strategies and outcomes of non-human auditory categorization to those of computational approaches.

Unfortunately, it is unclear what techniques might make such comparisons possible. One approach is to collect psychophysical measures of auditory perceptual sensitivities from animals ([Orduña et al., 2005](#_ENREF_75); [Sturdy et al., 2000](#_ENREF_103)). The downside of this approach is that it is time-consuming, can only provide glimpses into the sorts of distinctions animals make, and even then just for a small set of sounds in a specific context. Little is revealed about how they naturally sort sounds. A second option is to try to better understand what exactly animals’ brains do when sorting sounds. This approach depends on a combination of neurophysiological experiments and computational models of neural-circuit function ([Fiete, Hahnloser, Fee, & Seung, 2004](#_ENREF_26); [Fiete & Seung, 2009](#_ENREF_27); [Palakal & Wong, 1999](#_ENREF_77); [Reijniers & Peremans, 2010](#_ENREF_91); [Yamashita et al., 2008](#_ENREF_115)). So far, it has been applied to analysis of both sound production and perception in birds and bats. Although the network architectures described above are neurally inspired, they do not begin to approach the sophistication of processing by actual neural circuits. As more is learned about the neural mechanisms of category learning and auditory processing in various species, advanced computational approaches that simulate neural processing will undoubtedly be developed that clarify how animals organize and synthesize the sounds they use and hear.

**6. Software Notes**

Numerous computer programs for implementing neural networks are available commercially and freely via the internet. MATLAB (The MathWorks, Inc., Natick, Massachusetts, USA) provides a set of functions within its Neural Network Toolbox that make it possible to implement many commonly used network architectures and learning algorithms. MATLAB is useful for combining sophisticated signal processing (e.g., wavelet analysis) with neural- network classification, and for developing advanced graphical displays of network performance. Many other neural-network toolboxes have been developed for use with MATLAB, including PDPTool, which was developed in the laboratory of James McClelland. This software is freely available online at http://www.stanford.edu/group/pdplab/resources.html. A free data-mining program called ORANGE provides a range of basic neural-network processing capabilities, as well as a more graphically oriented interface ([Demsar, Zupan, Leban, & Curk, 2004](#_ENREF_20)). For basic neural-network classification needs, Michael Dawson has developed several simple neural network programs with graphical user interfaces, available at no cost at <http://www.bcp.psych.ualberta.ca/~mike/Software/>.

**7. Acknowledgments**

Preparation of this chapter was made possible by a fellowship from the Center for Advanced Studies in the Behavioral Sciences at Stanford to E. Mercado III and NSF grant IOS 0718004, along with a Natural Sciences and Engineering Research Council of Canada

(NSERC) Discovery Grant and Discovery Accelerator Supplement, an Alberta Ingenuity Fund (AIF) New Faculty Grant, a Canada Foundation for Innovation (CFI) New Opportunities Fund (NOF) and Infrastructure Operating Fund (IOF) grants along with start-up funding and CFI partner funding from the University of Alberta (UofA) to C. B. Sturdy.

References

Au, W. W. (1994). Comparison of sonar discrimination: Dolphin and an artificial neural network. *Journal of the Acoustical Society of America, 95*, 2728-2735.

Au, W. W., Andersen, L. N., Rasmussen, A. R., Roitblat, H. L., & Nachtigall, P. E. (1995). Neural network modeling of a dolphin's sonar discrimination capabilities. *Journal of the Acoustical Society of America, 98*, 43-50.

Aubin, T., Jouventin, P., & Hildebrand, C. (2000). Penguins use the two-voice system to recognize each other. *Proceedings of the Royal Society of London Series B - Biological Sciences, 267*, 1081-1087.

Bloomfield, L. L., Charrier, I., & Sturdy, C. B. (2004). Note types and coding in Parid vocalizations. II. The chick-a-dee call of the mountain chickadee (*Poecile gambeli*). *Canadian Journal of Zoology, 82*, 780-793.

Branstetter, B. K., & Mercado, E., III. (2006). Sound localization by cetaceans. *International Journal of Comparative Psychology, 19*, 26-61.

Buck, J. R., & Tyack, P. L. (1993). A quantitative measure of similarity for *Tursiops truncatus* signature whistles. *Journal of the Acoustical Society of America, 94*, 2497-2506.

Burnett, S., & Masters, W. (1999). The use of neural networks to classify echolocation calls of bats. *Journal of the Acoustical Society of America, 106*, 2189.

Charrier, I., Lee, T. T.-Y., Bloomfield, L. L., & Sturdy, C. B. (2005). Acoustic mechanisms of note-type perception in black-capped chickadee calls. *Journal of Comparative Psychology, 119*, 371-380.

Charrier, I., & Sturdy, C. B. (2005). Call-based species recognition in black-capped chickadees. *Behavioral Processes, 70*, 271-281.

Cholewiak, D., Sousa-Lima, R., & Cerchio, S. (2013). Humpback whale song hierarchical structure: Historical context and discussion of current classification issues. *Marine Mammal Science, 29*, E312-E332.

Clapham, P. J. (1996). The social and reproductive biology of humpback whales: An ecological perspective. *Mammal Review, 26*, 27-49.

Clemins, P. J., Johnson, M. T., Leong, K. M., & Savage, A. (2005). Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations. *Journal of the Acoustical Society of America, 117*, 956-963.

Datum, M. S., Palmieri, F., & Moiseff, A. (1996). An artificial neural network for sound localization using binaural cues. *Journal of the Acoustical Society of America, 100*, 372-383.

Dawson, M. R. W. (2004). *Minds and machines: Connectionism and psychological modeling*. Oxford: Blackwell.

Dawson, M. R. W. (2005). *Connectionism: A hands-on approach*. Oxford: Blackwell.

Dawson, M. R. W., Charrier, I., & Sturdy, C. B. (2006). Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types. *Journal of the Acoustical Society of America, 119*, 3161-3172.

Deecke, V. B., Ford, J. K., & Spong, P. (1999). Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects. *Journal of the Acoustical Society of America, 105*, 2499-2507.

Deecke, V. B., & Janik, V. M. (2006). Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls. *Journal of the Acoustical Society of America, 119*, 645-653.

DeLong, C. M., Au, W. W. L., Harley, H. E., Roitblat, H. L., & Pytka, L. (2007). Human listeners provide insights into echo features used by dolphins (*Tursiops truncatus*) to discriminate among objects. *Journal of Comparative Psychology, 121*, 306-319.

Demsar, J., Zupan, B., Leban, G., & Curk, T. (2004). Orange: From experimental machine learning to interactive data mining *Lecture notes in computer science*: Faculty of Computer and Information Science, University of Ljublijana, Slovenia.

Deregnaucourt, S., Guyomarc'h, J., & Richard, V. (2001). Classification of hybrid crows in quail using artificial neural networks. *Behavioural Processes, 56*, 103-112.

do Rego, R. L., Araujo, A. F., & de Lima Neto, F. B. (2010). Growing self-reconstruction maps. *IEEE Transactions on Neural Networks, 21*, 211-223.

Erbe, C. (2000). Detection of whale calls in noise: Performance comparison between a beluga whale, human listeners, and a neural network. *Journal of the Acoustical Society of America, 108*, 297-303.

Ereshefsky, M. (2000). *The poverty of the Linnean Hierarchy: A philosophical study of biological taxonomy*. Cambridge, UK: Cambridge University.

Ficken, M. S., Ficken, R. W., & Witkin, S. R. (1978). Vocal repertoire of the black-capped chickadee. *Auk, 95*, 34-48.

Fiete, I., Hahnloser, R., Fee, M., & Seung, H. (2004). Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *Journal of Neurophysiology, 92*, 2274.

Fiete, I., & Seung, H. (2009). Birdsong learning. In L. Squire (Ed.), *Encyclopedia of Neuroscience* (pp. 227-239). Amsterdam: Elsevier.

Fernández-Juricic , E., del Nevo, A. J., Poston, R. (2009). Identification of individual and population-level variation in vocalizations of the endangered southwestern willow flycatcher (*Empidonax traillii extimus*). *Auk, 126*, 89-99.

Freeberg, T. M., & Lucas, J. R. (2002). Recievers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinus*. *Animal Behaviour, 63*, 837-845.

Ghosh, J., Deuser, L., & Beck, S. (1992). A neural network based hybrid system for detection, characterization, and classification of short-duration oceanic signals. *IEEE Journal of Oceanic Engineering, 17*, 351-363.

Green, S. R., Mercado, E., III, Pack, A., & Herman, L. (2011). Recurring patterns in the songs of humpback whales (*Megaptera novaeangliae*). *Behavioural Processes, 86*, 284-294.

Green, S. R., Mercado, E., III, Pack, A. A., & Herman, L. M. (2007). Characterizing patterns within humpback whale (*Megaptera novaeangliae*) songs. *Aquatic Mammals, 33*, 202-213.

Guillette, L. M., Farrell, T. M., Hoeschele, M., Nickerson, C. M., Dawson, M. R. W., & Sturdy, C. B. (2010). Mechanisms of call note-type perception in black-capped chickadees (*Poecile atricapillus*): Peak shift in a note-type continuum. *Journal of Comparative Psychology, 124*, 109.

Hailman, J. P., Ficken, M. S., & Ficken, R. W. (1985). The 'chick-a-dee" calls of *Parus atricapillus*: A recombinant system of animal communication compared with written English. *Semiotica, 56*, 191-224.

Handel, S., Todd, S. K., & Zoidis, A. M. (2009). Rhythmic structure in humpback whale (*Megaptera novaeangliae*) songs: Preliminary implications for song production and perception. *Journal of the Acoustical Society of America, 125*, EL225-230.

Haykin, S. (1994). *Neural networks: A comprehensive foundation*. Upper Saddle River, NJ: Prentice Hall.

Heimbauer, L. A., Beran, M. J., & Owren, M. J. (2011). A chimpanzee recognizes synthetic speech with significantly reduced acoustic cues to phonetic content. *Current Biology, 21*, 1210-1214.

Heinke, D., & Hamker, F. H. (1998). Comparing neural networks: A benchmark on growing neural gas, growing cell structures, and fuzzy ARTMAP. *IEEE Transactions on Neural Networks, 9*, 1279-1291.

Helweg, D. A., Cato, D. H., Jenkins, P. F., Garrigue, C., & McCauley, R. D. (1998). Geographic variation in South Pacific humpback whale songs. *Behaviour, 135*, 1-27.

Houser, D. S., Helweg, D. A., & Moore, P. W. (1999). Classification of dolphin echolocation clicks by energy and frequency distributions. *Journal of the Acoustical Society of America, 106*, 1579-1585.

Janata, P. (2001). Quantitative assessment of vocal development in the zebra finch using self-organizing neural networks. *Journal of the Acoustical Society of America, 110*, 2593-2603.

Janik, V. M. (1999). Pitfalls in the categorization of behaviour: A comparison of dolphin whistle classification methods. *Animal Behavior, 57*, 133-143.

Juang, C. F., Lin, Y. Y., & Tu, C. C. (2010). A recurrent self-evolving fuzzy neural network with local feedbacks and its application to dynamic system processing. *Fuzzy Sets and Systems, 161*, 2552-2568.

Kamminga, C., & van Velden, J. G. (1987). Sonar signals of *Pseudorca crassidens* in comparison with *Tursiops truncatus*. *Aquatic Mammals, 13*, 43-49.

Kershenbaum, A., Blumstein, D., Roch, M., Ackay, C., Backus, G., Bee, M., … Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews*, *91*, 13-52

Kohonen, T. (1995). *Self organizing maps*. Berlin: Springer Verlag.

Koops, H. V., van Balen, J. Wiering, F. Automatic segmentation and deep learning of bird sounds. *Experimental IR Meets Multilingulity, Multimodality, and Interaction, 9283*, 261-267.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*, 69-72.

Lachlan, R. F., Peters, S., Verhagen, L., & ten Cate, C. (2010). Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taenopygia guttata*), and swamp sparrows (*Melospiza georgiana*). *Journal of Comparative Psychology, 124*, 92-108.

LeCun, Y., & Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436-444.

Lengagne, T., Aubin, T., Jouventin, P., & Lauga, J. (2000). Perceptual salience of individually distinctive features in the calls of adult king penguins. *Journal of the Acoustical Society of America, 107*, 508-516.

Manteuffel, G., Puppe, B., & Schon, P. C. (2004). Vocalization of farm animals as a measure of welfare. *Applied Animal Behaviour Science, 88*, 163-182.

Marsland, S., Shapiro, J., & Nehmzow, U. (2002). A self-organising network that grows when required. *Neural Networks, 15*, 1041-1058.

May, L. (1998). Individually distinctive corncrake *Crex crex c*alls: A further study. *Bioacoustics, 9*, 135-148.

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

McCowan, B. (1995b). Quantitative comparison of whistle repertoires from captive adult bottle-nosed dolphins (Delphinidae, *Tursiops truncatus*) - a reevaluation of the signature whistle hypothesis. *Ethology, 100*, 194-209.

McIlraith, A., & Card, H. (1997). Birdsong recognition using backpropagation and multivariate statistics. *IEEE Transactions on Signal Processing, 45*, 2740-2748.

Mellinger, D. K., & Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *Journal of the Acoustical Society of America, 107*, 3518-3529.

Mercado, E., III, & Frazer, L. N. (1999). Environmental constraints on sound transmission by humpback whales. *Journal of the Acoustical Society of America, 106*, 3004-3016.

Mercado, E., III, Green, S. R., & Schneider, J. N. (2008). Understanding auditory distance estimation by humpback whales: A computational approach. *Behavioural Processes, 77*, 231-242.

Mercado, E., III, & Handel, S. (2012). Understanding the structure of humpback whale songs (L). *Journal of the Acoustical Society of America, 132*, 2947-2950.

Mercado, E., III, & Henderson, C. M. (2012). Neurally inspired models of psychological processes. In I. B. Weiner, R. J. Nelson & S. Mizumori (Eds.), *Handbook of psychology: Behavioral neuroscience* (pp. 620-642). New York: John Wiley and Sons.

Mercado, E., III, Herman, L., & Pack, A. (2005). Song copying by humpback whales: Themes and variations. *Animal Cognition, 8*, 93-102.

Mercado, E., III, Herman, L. M., & Pack, A. A. (2003). Stereotypical sound patterns in humpback whale songs: Usage and function. *Aq. Mamm., 29*, 37-52.

Mercado, E., III, & Kuh, A. (1998). Classification of humpback whale vocalizations using a self-organizing neural network. *Proceedings of IJCNN '98*, 1584-1589.

Mercado, E., III, Schneider, J. N., Green, S. R., Wang, C., Rubin, R. D., & Banks, P. N. (2007). Acoustic cues available for ranging by humpback whales. *Journal of the Acoustical Society of America, 121*, 2499-2502.

Mercado, E., III, Schneider, J. N., Pack, A., & Herman, L. (2010). Sound production by singing humpback whales. *Journal of the Acoustical Society of America, 127*, 2678-2691.

Miksis-Olds, J. L., Buck, J. R., Noad, M. J., Cato, D. H., & Stokes, M. D. (2008). Information theory analysis of Australian humpback whale song. *Journal of the Acoustical Society of America, 124*, 2385-2393.

Montavon, G., Orr, G., & Müller, K. R. (2012). *Neural Networks: Tricks of the Trade*. Berlin: Springer.

Moore, P., Roitblat, H., Penner, R., & Nachtigall, P. (1991). Recognizing successive dolphin echoes with an integrator gateway network. *Neural Networks, 4*, 701-709.

Moss, C. F., Chiu, C., & Surlykke, A. (2011). Adaptive vocal behavior drives perception by echolocation in bats. *Current Opinion in Neurobiology, 21*, 645-652.

Murray, S. O., Mercado, E., & Roitblat, H. L. (1998a). Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations. *Journal of the Acoustical Society of America, 104*, 1679-1688.

Murray, S. O., Mercado, E., & Roitblat, H. L. (1998b). The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations. *Journal of the Acoustical Society of America, 104*, 3626-3633.

Naguib, M., & Wiley, R. H. (2001). Estimating the distance to a source of sound: Mechanisms and adaptations for long-range communication. *Animal Behaviour, 62*, 825-837.

Nelson, B. S., & Stoddard, P. K. (1998). Accuracy of auditory distance and azimuth perception by a passerine bird in a natural habitat. *Animal Behaviour, 56*, 467-477.

Neti, C., Young, E. D., & Schneider, M. H. (1992). Neural network models of sound localization based on directional filtering by the pinna. *Journal of the Acoustical Society of America, 92*, 3140-3156.

Nickerson, C. M., Bloomfield, L. L., Dawson, M. R. W., & Sturdy, C. B. (2006). Artificial neural network discrimination of black-capped chickadee (*Poecile atricapillus*) call notes. *Journal of the Acoustical Society of America, 120*, 1111-1117.

Nowicki, S. (1983). Flock-specific recognition of chickadee calls. *Behavioral Ecology and Sociobiology, 12*, 317-320.

Nowicki, S., & Nelson, D. A. (1990). Defining natural categories in acoustic-signals - Comparison of 3 methods applied to chick-a-dee call notes. *Ethology, 86*, 89-101.

Orduña, I., Mercado, E., III, Gluck, M. A., & Merzenich, M. M. (2005). Cortical responses in rats predict perceptual sensitivities to complex sounds. *Behavioral Neuroscience, 119*, 256-264.

Pace, F., Benard, F., Glovin, H., Adam, O., & White, P. (2010). Subunit definition and analysis for humpback whale call classification. *Applied Acoustics, 71*, 1107-1112.

Palakal, M. J., & Wong, D. (1999). Cortical representation of spatiotemporal pattern of firing evoked by echolocation signals: Population encoding of target features in real time. *Journal of the Acoustical Society of America, 106*, 479-490.

Parsons, S. (2001). Identification of New Zealand bats (*Chalinolobus tuberculatus* and *Mystacina tuberculata*) in flight from analysis of echolocation calls by artificial neural networks. *Journal of Zoology, 253*, 447-456.

Parsons, S., & Jones, G. (2000). Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *Journal of Experimental Biology, 203*, 2641-2656.

Payne, K., & Payne, R. S. (1985). Large scale changes over 19 years in songs of humpback whales in Bermuda. *Zeitschrift fur Tierpsychologie, 68*, 89-114.

Payne, K., Tyack, P., & Payne, R. S. (1983). Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): A detailed analysis of two seasons in Hawaii. In R. Payne (Ed.), *Communication and behavior of whales* (pp. 9-57). Boulder, CO: Westview Press.

Payne, R. S., & McVay, S. (1971). Songs of humpback whales. *Science, 173*, 585-597.

Pepperberg, I. M. (2005). Insights into vocal imitation in African grey parrots (*Psittacus erithacus*). In S. Hurley & N. Chater (Eds.), *Perspectives on imitation, vol 1: Mechanisms of imitation and imitation in animals* (pp. 243-262). Cambridge, MA: MIT Press.

Phelps, S. M., & Ryan, M. J. (1998). Neural networks predict response biases of female tungara frogs. *Proceedings of the Royal Society of London Series B - Biological Sciences, 265*, 279-285.

Phillmore, L. S., Sturdy, C. B., Ramsay, S. M., & Weisman, R. G. (1998). Discrimination of distance cues in black-capped chickadees (*Poecile atricapillus*) and zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology, 112*, 282-291.

Placer, J., & Slobodchikoff, C. N. (2000). A fuzzy-neural system for identification of species-specific alarm calls of Gunnison's prairie dogs. *Behavioural Processes, 52*, 1-9.

Potter, J. R., Mellinger, D. K., & Clark, C. W. (1994). Marine mammal call discrimination using artificial neural networks. *Journal of the Acoustical Society of America, 96*, 1255-1262.

Pozzi, L., Gamba, M., & Giacoma, C. (2010). The use of artificial neural networks to classify primate vocalizations: A pilot study on black lemurs. *American Journal of Primatology, 72*, 337-348.

Reby, D., Lek, S., Dimopoulos, I., Joachim, J., Lauga, J., & Aulagnier, S. (1997). Artificial neural networks as a classification method in the behavioural sciences. *Behavioural Processes, 40*, 35-43.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience, 13*, 87-103.

Reijniers, J., & Peremans, H. (2010). On population encoding and decoding of auditory information for bat echolocation. *Biological Cybernetics, 102*, 311-326.

Ripley, B. D. (2008). *Pattern recognition and neural networks*. Cambridge: Cambridge University.

Roberts, W. A. (2002). Are animals stuck in time? *Psychological Bulletin, 128*, 473-489.

Roitblat, H., Au, W., Nachtigall, P., Shizumura, R., & Moons, G. (1995). Sonar recognition of targets embedded in sediment. *Neural Networks, 8*, 1263-1273.

Roitblat, H., Moore, P., Helweg, D., & Nachtigall, P. (1993). Representation and processing of acoustic information in a biomimetic neural network. In J.-A. Meyer, H. L. Roitblat & S. W. Wilson (Eds.), *From animals to animats 2: Proceedings of the second International Conference on Simulation of Adaptive Behaviour* (pp. 90-99). Cambridge, MA: MIT.

Roitblat, H., Moore, P., Nachtigall, P., Penner, R., & Au, W. (1989). Natural echolocation with an artificial neural network. *Neural Networks, 1*, 707-716.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by backpropagating errors. *Nature, 323*, 533-536.

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructures of cognition. Volume 1: Foundations*. Cambridge, MA: MIT.

Schon, P. C., Puppe, B., & Manteuffel, G. (2001). Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*). *Journal of the Acoustical Society of America, 110*, 1425-1431.

Schwenker, F., Dietrich, C., Kestler, H., Riede, K., & Palm, G. (2003). Radial basis function neural networks and temporal fusion for the classification of bioacoustic time series. *Neurocomputing, 51*, 265-275.

Simmons, J. A., & Stein, R. A. (1980). Acoustic imaging in bat sonar: Echolocation signals and the evolution of echolocation. *Journal of Comparative Physiology, 135*, 61-84.

Sturdy, C. B., Phillmore, L. S., Price, J. L., & Weisman, R. G. (1999). Song-note discriminations in zebra finches (*Taeniopygia guttata*): Categories and pseudocategories. *Journal of Comparative Psychology, 113*, 204-212.

Sturdy, C. B., Phillmore, L. S., & Weisman, R. G. (2000). Call-note discriminations in black-capped chickadees (*Poecile atrica pillus*). *Journal of Comparative Psychology, 114*, 357-364.

Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. *Journal of the Acoustical Society of America, 119*, 1849-1866.

Tanttu, J., Turunen, J., Selin, A., & Ojanen, M. (2006). Automatic feature extraction and classification of crossbill (*Loxia spp*.) flight calls. *Bioacoustics, 15*, 251.

Terry, A., & McGregor, P. (2002). Census and monitoring based on individually identifiable vocalizations: The role of neural networks. *Animal Conservation, 5*, 103-111.

Thompson, R. F. (1965). The neural basis of stimulus generalization. In D. I. Mostofsky (Ed.), *Stimulus generalization* (pp. 154-178). Stanford, CA: Stanford University.

Van Der Schaar, M., Delory, E., Catala, A., & Andre, M. (2007). Neural network-based sperm whale click classification. *Journal of the Marine Biological Association of the UK, 87*, 35-38.

Walker, A., Fisher, R., & Mitsakakis, N. (1996). Singing maps: Classification of whalesong units using a self-organizing feature mapping algorithm *DAI Research Paper no. 833*: Department of AI, University of Edinburgh, UK.

Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience, 5*, 279-290.

Winn, H. E., & Winn, L. K. (1978). The song of the humpback whale *Megaptera novaeangliae* in the West Indies. *Marine Biology, 47*, 97-114.

Wisniewski, M. G., DeLong, C. M., Heberle, A. L., & Mercado, E., III. (2013). Exploring the capacity of neural networks to recognize objects from dolphin echoes across multiple orientations. *Proceedings of Meetings on Acoustics, 19*, 010035.

Wotton, J. M., & Jenison, R. L. (1997). A backpropagation network model of the monaural localization information available in the bat echolocation system. *Journal of the Acoustical Society of America, 101*, 2964-2972.

Wysoski, S. G., Benuskova, L., & Kasabov, N. (2010). Evolving spiking neural networs for audiovisual information processing. *Neural Networks, 23*, 819-835.

Yamashita, Y., Takahasi, M., Okumura, T., Ikebuchi, M., Yamada, H., Suzuki, M., . . . Tani, J. (2008). Developmental learning of complex syntactical song in the Bengalese finch: A neural network model. *Neural Networks, 21*, 1224-1231.

Zimmerman, A. (1995). Artificial neural networks for analysis and recognition of primate vocal communication. In E. Zimmerman, J. D. Newman & U. Jürgens (Eds.), *Current topics in primate vocal communication* (pp. 29-46). New York: Plenum.

**Figure Captions**

**Figure 1:** The input vector **x** consists of an array of values (in this case four, *x*1-4). The computation performed by a single unit within a neural network consists of multiplying each of these values times a weight (*w*1-4), summing the products (plus a constant

value called the *bias*), and transforming that sum into another value *y* using the function *f*. The function *f*, called the activation function, can take many forms. One simple variant is the step function. It assigns any value greater than a fixed threshold a value of one, with a lower values set to zero. In the example above, if *f* is a step function having a threshold of two, then if x = (1 1 1 0), y = 1. If x = (0 0 0 1), then y = 0.

**Figure 2:** (A) A single-layer neural network has one set of connections and associated weights. (B) A multi-layer network can be viewed as a series of single-layer networks. Layers that are between inputs and outputs are called hidden layers. In the example above, the hidden layer consists of three units. (C) Self-organizing maps typically consist of single layer with many output units that compete to become active. Each output unit is associated with a weight vector. When an output unit wins a competition, it becomes active (illustrated here by the white unit). Nearby units are also activated, but to a lesser degree.

**Figure 3:** A sound spectrogram of black-capped chickadee call notes (1024-point Hamming window, 21-Hz frequency resolution). Adapted from Dawson et al. (2006).

**Figure 4:** The topography of an SOM trained to classify false killer whale vocalizations. Element number shows the value~~s~~ of each unit’s weight vector. The first 30 elements in each figure represent measures of waveform structure, while the last 30 elements represent peak frequencies. Adapted from Murray et al. (1998).

**Figure 5:** The feature space of an SOM trained to classify short-duration frames from humpback whale sounds. Each plot corresponds to the weight vector of one unit, representing the prototypical, perceptually warped spectrum for one frame. Note that many units show only subtle differences in sensitivity.

**Figure 6:** A scheme for characterizing SOM activity. The 10 × 10 SOM shown in Figure 5 can be divided into discrete regions to describe its activity patterns. Because sounds of differing durations contain different numbers of frames, a variable number of neurons will be sequentially activated in response to each sound (SOMs, however, require input vectors to be of equal length). To create suitable inputs for the second-level SOM, activity can be described using two frequency distributions. One distribution measures how often units become active in each of the 25 activity regions shown. A second distribution measures how often activity crosses between regions of the map (with 16 possible transitions). These two distributions can be combined to create a 41-element vector summarizing the positions and dynamics of unit activation for individual sounds.

**Figure 7:** A “hit” distribution for a second-level SOM constructed using the frame-based SOM shown in Figure 5 and the coding scheme described in Figure 6. Each hexagon corresponds to a unit in a 5 × 5 SOMtrained with 248 individual sounds from a single humpback whale song., The size of the hexagon reflects the number of times that unit became active after all individual sounds were presented to the SOM. Subjectively recognizable categories are labeled. Dotted arrows indicate the progression of unitactivity across the map when sounds were input in the order they occurred within the song.

**Figure 8:** The mean response of a set of networks to shifted notes after the networks were trained to discriminate A notes from B notes. Solid lines show results from trials in which networks were trained to respond to a note, while dashed lines show results from trials in which the networks were trained to not respond to a note. Adapted from Nickerson et al. (2006).

Table 1

Representative applications of neural networks to the classification, simulation, and assessment of animal vocalizations.

|  |  |  |  |
| --- | --- | --- | --- |
| Focus | Species | Study | Network Type |
| Classifying calls | Chickadee | Dawson et al. (2006) | MLP |
|  | Chickadee | Guillette et al. (2010) | SLP |
|  | Crossbill | Tanttu et al. (2006) | SOM |
|  | Quail | Deregnaucort et al. (2001) | MLP |
|  | Lemur | Zimmerman (1995) | SOM/SLP |
|  | Lemur | Pozzi et al. (2010) | MLP |
|  | Prairie dog | Placer & Slobodchikoff (2000) | MLP |
|  | Prairie dog | Placer et al. (2006) | SOM |
|  | Bowhead whale | Potter et al. (1994) | MLP |
|  | Blue whale | Bahoura & Sinourd (2009) | MLP |
|  | Dolphin, Orca | Deecke & Janik (2006) | ART |
|  | Pseudoorca | Murray et al. (1998) | SOM |
|  | Sperm whale | Van der Schaar et al. (2007) | RBF |
|  | Humpback whale | Walker et al. (1996) | SOM |
| Classifying songs | Cricket (various) | Schwenker et al. (2003) | RBF |
|  | Bird (various) | Selouani et al. (2006) | TDNN |
|  | Bird (various) | Juang & Chen (2007) | SRNFN |
|  | White-crowned sparrow | Ranjard & Ross (2008) | Evolving tree (SOM) |
|  | Humpback whale | Helweg et al. (1998) | ART 2-A |
| Classifying sonar | Bats | Burnett & Masters (1999) | SOM |
|  | Dolphin | Houser et al. (1999) | Counterpropagation |
| Simulating spatial processing | Bat | Wotton & Jenison (1997) | MLP |
|  | Cat | Neti et al. (1992) | MLP |
|  | Dolphin | Roitblat et al. (1989) | Counterpropagation |
|  | Dolphin | Moore et al. (1991) | Integrator gateway |
|  | Dolphin | Roitblat et al. (1992) | BNN |
|  | Dolphin | Au (1994); Au et al. (1995) | Counterpropatation, MLP |
|  | Dolphin | Helweg et al. (2006) | PNN, MLP, GLN |
|  | Humpback whale | Mercado et al. (2008) | MLP |
|  | Any with two ears | Datum et al., 1996 | MEKA |
| Assessing status | Zebra finch | Janata (2001) | SOM |
|  | Pig | Schon et al. (2001) | SOM |
|  | Beluga | Erbe (2000) | MLP |
|  | Marine animals | Ghosh et al. (1992) | RBF, AKC, PSN |
| Assessing identity (individual or species) | Frog | Phelps & Ryan (1998) | MLP |
|  | Bird (various) | McIlraith & Card (1997) | MLP |
|  | Corncrake | Terry & McGregor (2001) | MLP, PNN, SOM |
|  | Corncrake | May (1998) | MLP |
|  | Flycatcher | Fernández-Juricic et al. (2009) | PNN |
|  | Deer | Reby et al. (1997) | MLP |
|  | Bat (various) | Parsons & Jones (2000); Parsons (2001) | MLP |

*Note.* MLP = multi-layer perceptron; SLP = single-layer perceptron; SOM = self-organizing map; ART = Adaptive resonance theory; RBF = radial basis function; TDNN = time delay neural network; SRNFN = singleton-type recurrent neural fuzzy networks; BNN = biomimetic neural network; PNN = probabilistic neural network, GLN = genetic reinforcement learning; MEKA = multiple extended Kalman algorithm; PSN = Pi-Sigma network; AKC = Adaptive kernel classifier.