Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects

V. B. Deecke
Marine Mammal Research Unit, University of British Columbia, 6248 BioSciences Road, Vancouver B.C. V6T 1Z4, Canada

J. K. B. Ford
Marine Mammal Research Unit, University of British Columbia, 6248 BioSciences Road, Vancouver B.C. V6T 1Z4, Canada and Vancouver Aquarium Marine Science Centre, P.O. Box 3232, Vancouver B.C. V6B 3X8, Canada

P. Spong
OrcaLab, Hanson Island, P.O. Box 258, Alert Bay B.C. V0N 1A0, Canada

(Received 28 April 1998; revised 23 October 1998; accepted 5 January 1999)

A quantitative measure of acoustic similarity is crucial to any study comparing vocalizations of different species, social groups, or individuals. The goal of this study was to develop a method of extracting frequency contours from recordings of pulsed vocalizations and to test a nonlinear index of acoustic similarity based on the error of an artificial neural network at classifying them. Since the performance of neural networks depends on the amount of consistent variation in the training data, this technique can be used to assess such variation from samples of acoustic signals. The frequency contour extraction and the neural network index were tested on samples of one call type shared by nine social groups of killer whales. For comparison, call similarity was judged by three human subjects in pairwise classification tasks. The results showed a significant correlation between the neural network index and the similarity ratings by the subjects. Both measures of acoustic similarity were significantly correlated with the groups' association patterns, indicating that both methods of quantifying acoustic similarity are biologically meaningful. An index based on neural network analysis therefore represents an objective and repeatable means of measuring acoustic similarity, and allows comparison of results across studies, species, and time. © *1999 Acoustical Society of America*. [S0001-4966(99)01004-8]

PACS numbers: 43.80.Ka, 43.80.Lb, 43.80.Jz [FD]

INTRODUCTION

A widespread problem in the study of animal and human vocal communication lies in describing and quantifying the similarity of acoustic signals. A quantitative measure of acoustic similarity is crucial to any study comparing the vocalizations of different species, social groups, or individuals. Current approaches to this problem fall into two categories. *Statistical measures* of acoustic similarity use univariate or multivariate statistics on measures extracted from acoustic signals (e.g., Bailey, 1978; Symmes *et al*., 1979; Clark *et al*., 1987; Buck and Tyack, 1993; for overviews see Martindale, 1980, and Williams and Slater, 1991). *Perceptual measures* quantify acoustic similarity through ratings by human subjects (e.g., Tyack, 1986; Sayigh *et al*., 1990), or by the ability of human or animal subjects to discriminate between classes of signals (e.g., Miller and Nicely, 1955; Loesche *et al*., 1992).

Statistical measures of acoustic similarity have the advantage of being objective and repeatable (Martindale, 1980; Clark *et al*., 1987), making it possible to compare the results from different studies. However, they may not always be the most meaningful, since they only assess the physical properties of the signals and give no information on how they are perceived (see Horn and Falls, 1996). Perceptual measures, although often biologically meaningful, have the problem of observer bias. Whereas ratings of similarity by the same subject are probably comparable, ratings made by different subjects are generally not. In addition, obtaining ratings by human subjects or trained animals becomes a logistic challenge in experiments where the acoustic similarity of multiple samples needs to be assessed in pairwise comparisons, or where sample sizes are large.

In this paper, we introduce the use of an artificial neural network to measure the similarity of discrete calls of killer whales (*Orcinus orca*). Artificial neural networks were developed by modeling biological systems of information processing (for overviews, see Dasgupta, 1991; Hinton, 1992). Due to their ability to classify unknown data based on information obtained from a known training set, neural networks have successfully been used in the automated classification of acoustic signals (e.g., Neumann *et al*., 1992; Ramani *et al*., 1993), including killer whale calls (Spong *et al*., 1993). Since the performance of a neural network depends on the amount of consistent variation between the signal patterns in the training set, we demonstrate that the discrimination error of a neural network can be used to quantify the similarity of signals. Its bio-mimetic nature makes neural network analysis a promising candidate for a measure of similarity which assesses acoustic variation in a biological

meaningful while being objective and repeatable.

Ford (1984, 1989, 1991) showed that different killer whale communities use distinctively different vocal signals. Within the Northern Resident Community of British Columbia, stable kin groups, called pods, have unique vocal repertoires of 7–17 discrete call types. Related pods often use structurally distinct versions of the same call types. Within pods, matrilineal groups, called subpods, again have their own versions of shared call types. Finally, individuals likely have unique ''voices'' due to variation in their sound-producing structures. The vocal communication of killer whales exhibits variation on a variety of levels and provides a challenging field in which to test methods of measuring acoustic similarity.

Many studies have used frequency contours to describe vocalizations (e.g., Bailey, 1978; Sayigh *et al*., 1990; Buck and Tyack, 1993; McCowan, 1995). For tonal signals, a frequency contour gives changes in the fundamental frequency of a vocalization over time. For pulsed signals, such as the discrete calls of killer whales (Schevill and Watkins, 1966; Ford, 1989), the contour describes changes in the pulse repetition rate (pulse-rate contour). The similarity of samples of frequency contours can be assessed using statistical (Bailey, 1978; Buck and Tyack, 1993; McCowan, 1995) or perceptual (Sayigh *et al*., 1990) measures. Using frequency contours to describe vocalizations has the advantage that the signal is analyzed as a unit rather than broken down into disjunct measurements. In addition, irrelevant information, such as background noise or artifacts introduced by the recording apparatus, is eliminated from subsequent analyses. This is especially beneficial in the present study, which compares calls from recordings made in the field with a variety of recording systems.

So far, most automated procedures for extracting frequency contours from spectrograms have been developed for tonal signals (such as bird vocalizations or dolphin whistles) and for recordings obtained under controlled circumstances from captive or temporarily isolated animals (e.g., Buck and Tyack, 1993). In this paper we describe a method to determine the pulse repetition rate from spectrograms of pulsed calls. This method of extracting pulse-rate contours is robust to levels of background noise typical of field recordings. We introduce an index of acoustic similarity based on the performance of a neural network at classifying unknown contours using information obtained from a known training set. We test the contour extraction algorithm and the neural network index on calls of nine matrilineal groups of killer whales. For comparison, we measure the similarity of the same calls using the classification error of three human subjects. To investigate whether both measures of acoustic similarity are biologically meaningful, we compare them to the association patterns of the nine groups.

I. METHODS

A. Extraction of pulse-rate contours: The sidewinder algorithm

The discrete calls of killer whales are pulsed signals in which a tone (of a certain *tonal frequency*) is not emitted continuously but in pulses (given by the *pulse-repetition rate*; Schevill and Watkins, 1966; Watkins, 1967). Unlike in the tonal signals of many birds or other delphinids, the highest amount of energy is therefore not always contained in the first, second, or third harmonic (Watkins, 1967). The pulsed nature of these calls and the fact that the recordings used in this study were made in the field and often contained high levels of background noise meant that extraction algorithms from the literature (such as used by Buck and Tyack, 1993) proved not to be satisfactory.

For the extraction of pulse-rate contours, suitable calls were digitized at a sampling rate of 22 050 Hz from cassette tapes, including at least 100 ms of background noise before the onset of the call. Spectrograms were generated by fast Fourier transform (FFT) using the Canary 1.2.1 sound analysis software (Cornell Laboratory of Ornithology) with a filter bandwidth of 88 Hz, and an FFT size and frame length of 1024 points. Overlap between frames was 87.5%, and a Hamming window function was used for normalization. These parameters give a frequency resolution of 21.53 Hz, and a temporal resolution of 5.81 ms. Contours were extracted using MATLAB 4.2 (The MathWorks, Inc.) for Macintosh with the signal processing toolbox.

The algorithm used in this study assumes that the beginning and the end of the call can be determined visually from the spectrogram. In order to reduce background noise levels, an average noise spectrum was computed from the part of the spectrogram before the onset of the call, and subtracted from all time bins. In a spectrogram of a pulsed vocalization, the pulse repetition rate is given by the spacing between frequency bands (Watkins, 1967). To find the pulse-repetition rate at each point in time, the autocovariance sequence (mean-removed autocorrelation sequence) R was first computed for each individual power spectrum y of the spectrogram using the formula:

$$R_y(n) = \sum_{m=0}^{m=N} [y(n+m)-\bar{y}][y(n)-\bar{y}], \tag{1}$$

where n is the frequency bin number, m is the offset of the spectrum in frequency bins, $\bar{y}$ is the average sound pressure of the spectrum, and N is the number of frequency bins in the spectrum. To save computing time, the sequence was only calculated from $m=0$ to $m=N$, since the segment from $m=-N$ to $m=0$ is an exact mirror image and yields no additional information. The frequencies of any sidebands in the acoustic signal are given by a simple linear relationship, and therefore the autocovariance sequence will show a peak every time m equals a multiple of the frequency spacing of the bands (i.e., of the pulse-repetition rate; Watkins, 1967) and adjacent bands overlap. Because the power spectrum of the background noise tends to decrease with increasing frequency, and adjacent frequency bands generally have similar energy content, the second highest maximum in the autocovariance sequence (after $m=0$) usually corresponds to the frequency bin containing the pulse-repetition rate. Sometimes this maximum represents the second, and in some rare cases the third, harmonic. A simple heuristic algorithm described by Buck and Tyack (1993), which checked for local

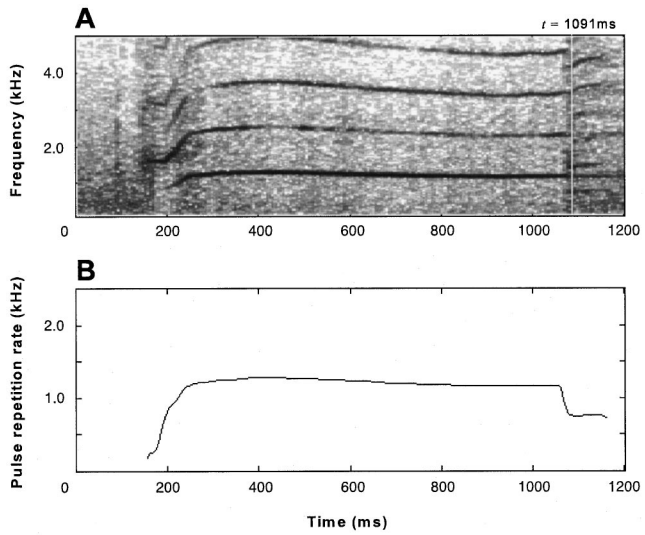

FIG. 1. (A) Spectrogram of an N4 call with a filter bandwidth of 88 Hz. The white line at $t = 1091$ ms shows the position of the power spectrum in Fig. 2. (B) Pulse-rate contour extracted from the spectrogram.

maxima at 1/2 and 1/3 of the offset of the second highest maximum, could account for this.

Figure 1 shows a spectrogram of an N4 call and a pulse-rate contour extracted from it. Figure 2 gives the power spectrum at $t = 1091$ ms (A) and its covariance sequence (B) for the same call. For subsequent analysis, the pulse-repetition rate was determined at 100 equally spaced points throughout the call and presented to the neural network as a vector of 100 numbers. Thus calls were essentially standardized for time; however, call length was entered as a 101 st number

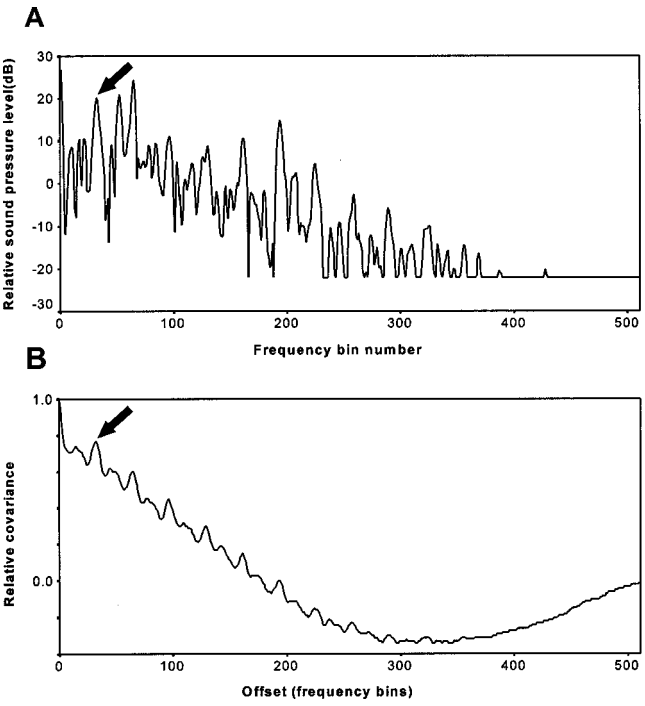

FIG. 2. (A) Power spectrum of the terminal component of an N4 call at $t = 1091$ ms (see Fig. 1). Filter bandwidth is 88 Hz and frequency resolution is 21.53 Hz per frequency bin. (B) Autocovariance sequence of the power spectrum. The arrows indicate the frequency bin containing the pulse-repetition rate.

into the analysis to allow discrimination of calls which differed consistently in length, but not in structure.

B. Analysis of acoustic variation in the N4 call

To test the performance of the neural network index on biological data, we used recordings of nine matrilineal groups, or subpods, of killer whales. Matrilineal groups consisted of between two and seven individuals, and belong to A-subclan of the Northern Resident Community (Ford, 1991) which inhabits the waters of British Columbia, Canada. The recordings analyzed in this study were made in the fjords and straits of the southern coast of British Columbia in weakly stratified or unstratified waters of depths of up to 400 m and often contained low to moderate levels of shipping noise. Recordings were contributed by a number of researchers using a variety of recording systems. All systems had a flat frequency response from 0.1 to 7 kHz, although for some systems the range of the flat response extended up to 20 kHz.

All members of the Northern Resident Community can be identified consistently from natural markings (Bigg *et al.*, 1990; Ford *et al.*, 1994). The analysis was restricted to recordings which could be attributed to a certain matrilineal group because was the only group within recording range and its identity was confirmed visually or photographically. We chose the N4 call (Ford, 1989, 1991; see Fig. 3) for this study because it is shared by all nine groups and it is one of the most frequently used call types in their repertoire. Structurally, the N4 calls of A08, A09, A23, A25, and A36 all have relatively low peak pulse repetition rates, and a pronounced terminal component at the end of the call (see Fig. 3). The versions of N4 made by A12 and A30 subpods usually lack the terminal component and have a relatively higher peak pulse repetition rate. Finally the N4 calls of A11 and A24 subpods (A4 pod of Ford, 1991) tend to be longer than those of any other matrilineal group and generally end in an upsweep.

N4 calls with adequate signal-to-noise ratios were identified acoustically and visually from recordings, and were digitized using the Canary 1.2.1 sound analysis software. Spectrograms were computed and pulse-rate contours extracted with the sidewinder algorithm. Since the performance of a neural network is highly dependent on the number of examples for each signal pattern in the training set, sample size for all matrilineal groups was standardized to 24, the size of the smallest sample. For each group we included calls from as many independent recording sessions as possible, to present the neural network and the human subjects with calls from a wide range of behavioral contexts, which are known to affect call structure (Ford, 1989). No less than three independent recording sessions were used for any one matrilineal group.

Association patterns of the different matrilineal groups were analyzed by generating an association matrix giving the *half-weight index of association* (Ginsberg and Young, 1992) for each pair of matrilineal groups. This index gives the number of observations of two groups traveling together as a proportion of half the total number of observations for the two groups. The association data came from a sightings da-

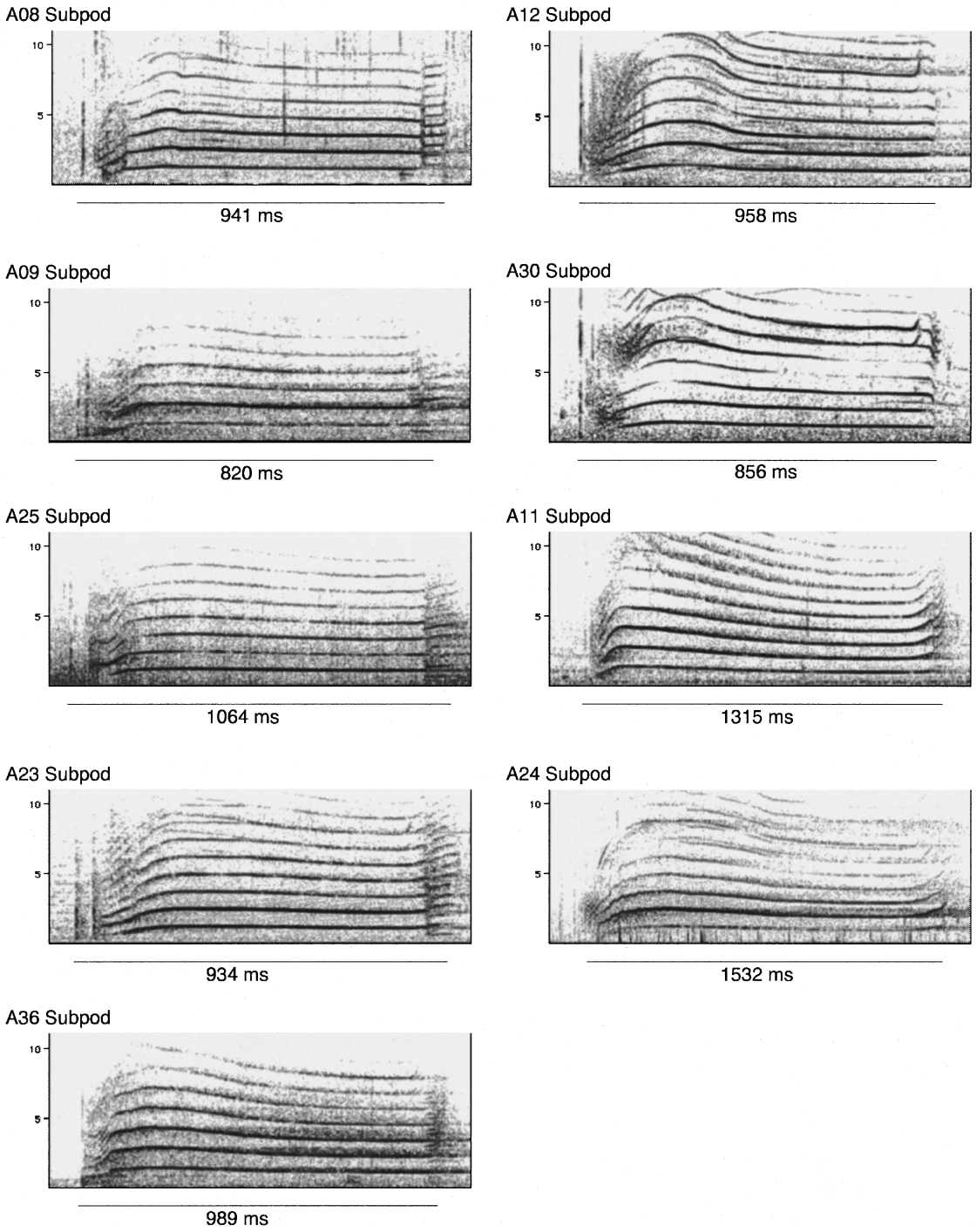

FIG. 3. Examples of spectrograms of N4 calls from the nine matrilineal groups of A-subclan.

tabase for the Northern Resident Community spanning the years 1990–1995. The total number of sightings of one or more A-subclan matrilineal group was 757, while numbers of sightings of any one matrilineal group ranged between 147 (A25 subpod) and 415 (A30 subpod).

C. The neural network index of acoustic similarity

Neural network analysis was done with the neural network toolbox of MATLAB 4.2 for the Macintosh (The MathWorks, Inc.). We used a standard back-propagation network (e.g., Rumelhart *et al*., 1986) with momentum and an adaptive learning rate (Vogl *et al*., 1988). Back-propagation networks can be trained to classify unknown patterns by ''learning'' to associate certain known input patterns with certain outputs. In our case inputs consisted of pulse-rate contours plus call length from two social groups, and the expected outputs were the matrices [0 1] and [1 0], depending on which group the contour came from. After training, the performance of a neural network can be tested by presenting it with data not used in training, and determining how closely the observed output matches the expected one.

To determine the network's performance during the training process, the training algorithm computes the sum-

square error of observed against expected output. We used a modified version of this parameter, the *discrimination error*, to determine network performance when testing a network with unknown contours. The discrimination error is calculated by subtracting the observed output from the expected output ([0 1] or [1 0]) and taking the mean of the absolute differences. The average of the discrimination error of all networks trained on one comparison was chosen over the proportion of misclassifications because the discrimination error not only yields information on whether a classification is correct, but also gives a relative measure of the quality of discrimination. For example, even an untrained network might easily generate [0.49 0.51] for an expected output of [0 1]. Since in both cases the second output exceeds the first one, the classification is considered correct and the neural network has a classification error of 0, although the classification was hardly better than random. The discrimination error of 0.49 much better reflects the poor quality of this classification.

The optimal neural network architecture for the discrimination tasks was determined in a network design experiment which consisted of training neural networks on a range of comparisons and varying the number of neurons in the hidden layer, as well as the length of training. For all comparisons, discrimination did not increase detectably when using more than 20 neurons in the hidden layer and when training for more than 5000 iterations, so that these parameters were used in all subsequent analyses. Neural networks were initialized with random weights, and a small number of neural networks did not improve in performance from the initialized state. Since this failure to train results from the configuration of weights at initialization rather than from a lack of consistent variation in the training set, such networks were eliminated from the analysis by setting a criterion of a 20% decrease of the sum-square error during the first 150 iterations.

To arrive at an acoustic similarity matrix for the N4 calls of the nine groups, we trained and tested neural networks on all 36 possible pairwise comparisons. We intended to train as many independent neural networks as possible on each comparison to eliminate the stochastic component of neural network analysis. To do this, one contour was excluded from the training set, a neural network was trained on the remaining 47 contours, the neural network was tested using the excluded contour and the discrimination error was determined. The test contour was then added back to the training set, another one was removed, and this procedure was repeated until each contour had served as the test contour. We tested each network with only a single contour in order to have as many contours for training as possible. Networks trained with fewer contours and tested with more gave consistently higher discrimination errors, probably due to overtraining (Rumelhart *et al*., 1986). The *neural network index of acoustic similarity* for each pairwise comparison is the average of the discrimination errors of all 48 neural networks trained this way.

D. Acoustic similarity ratings by human subjects

The same nine samples of 24 calls each were used to determine the classification errors of human subjects in pairwise computer-based discrimination tasks. Three female subjects, none of whom had any previous knowledge of killer whale vocal communication, were presented with the discrimination tasks. Subject A was 20 years old and had no musical background. Subject B was 22 years old and had played the flute for 2 years, and subject C was 20 years old and had played the piano for 13 years.

Since human subjects cannot be trained more than once on the same problem without seeing an increase in performance, we used a somewhat modified training and testing protocol for this part of the analysis. In each discrimination task, the subject was first presented with a training set of 16 calls belonging to two categories (A or B) according to the group they came from. The subject could listen to the calls and view their spectrograms, and was then asked to assign a test set of 32 unfamiliar calls to the appropriate category. The *rating of acoustic similarity* gives the proportion of misclassifications among these 32 calls. During the testing, the subject was allowed to return to the training set, but in order to cause her to generalize, was asked not to do so more than three times for each discrimination task. Following the experiments, the subjects completed a questionnaire asking whether they classified the calls primarily using acoustic or visual cues.

For visual comparison, average linkage dendrograms were generated from the four acoustic similarity matrices (one neural network index and three human subject ratings) as well as from the association matrix. Average linkage is a hierarchical tree-building algorithm and will group subpods with high indices of acoustic similarity or association into common clusters in a dendrogram (see Johnson, 1967). The acoustic similarity matrices and the association matrix were compared statistically by generating the matrix correlation coefficient for all possible comparisons. A Mantel test was used to test for significance.

II. RESULTS

The sidewinder algorithm proved effective at extracting pulse-rate contours from recordings obtained under a variety of recording conditions. Contours could be obtained even from recordings with high levels of ambient noise, if the call was clear and the energy in two or more frequency bands exceeded the background noise level. Only recordings containing boat noise with harmonic content, and recordings with a great amount of acoustic reverberation or strong echoes, caused problems in the contour extraction.

The values for the neural network index of acoustic similarity for the pairwise comparisons of N4 calls are given in Table I. The neural network could best discriminate between the N4 calls of A23 and A24 subpods (neural network index: 0.01). A09 subpod and A25 subpods gave the poorest discrimination (neural network index: 0.48). The average value for the neural network index for all discrimination tasks was 0.15. The neural network index grouped the nine matrilineal groups into three major clusters according to the

TABLE I. Acoustic similarity matrix for the N4 call of the nine matrilineal groups based on the neural network index of acoustic similarity. The values give the neural network performance (average discrimination error) for each pairwise comparison.

A09	0.34							
A11	0.09	0.04						
A12	0.05	0.03	0.08					
A23	0.18	0.19	0.04	0.04				
A24	0.10	0.03	0.29	0.08	0.01			
A25	0.43	0.48	0.04	0.05	0.23	0.03		
A30	0.14	0.06	0.07	0.37	0.09	0.10	0.06	
A36	0.27	0.40	0.07	0.06	0.19	0.06	0.37	0.10
	A08	A09	A11	A12	A23	A24	A25	A30

similarity of their N4 calls. These are A08–A09–A23–A25–A36, A12–A30, and A11–A24 (Fig. 4). These clusters are consistent with structural differences in the calls shown in Fig. 3.

Table II gives the ratings of acoustic similarity (proportion of misclassifications) for the 3 subjects and the 36 classification tasks. The table shows that subjects B and C classified all calls correctly in at least one comparison. The highest proportion of misclassification was higher than random (0.63, A08 vs A09 by subject C). The average proportion of misclassification for all discrimination tasks was 0.25, 0.18, and 0.15 for subjects A, B, and C, respectively, and a

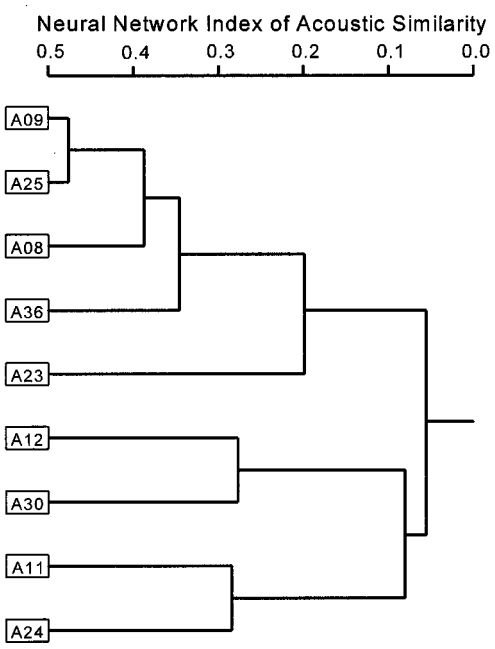

FIG. 4. Average linkage dendrogram giving acoustic similarity of the N4 call of the nine matrilineal groups based on the neural network index (generated from Table I). The position of the vertical lines linking groups or clusters of groups with respect to the scale bar above indicates the similarity of their N4 call based on the neural network index. Comparisons of N4 calls from groups which are linked on the left-hand side of the graph gave higher average discrimination errors (suggesting higher similarity) than those linked on the right-hand side.

TABLE II. Acoustic similarity matrix for the N4 call of the nine matrilineal groups generated by three human subjects. The values give the subjects' performance (proportion of misclassifications) for each pairwise comparison.

	Subject								
A09	A	0.44							
	B	0.25							
	C	0.63							
A11	A	0.13	0.22						
	B	0.25	0.06						
	C	0.16	0.03						
A12	A	0.16	0.06	0.06					
	B	0.09	0.09	0.09					
	C	0.06	0.09	0.03					
A23	A	0.50	0.25	0.03	0.25				
	B	0.19	0.47	0.03	0.09				
	C	0.19	0.53	0.00	0.03				
A24	A	0.25	0.28	0.44	0.28	0.03			
	B	0.16	0.06	0.38	0.22	0.03			
	C	0.19	0.06	0.28	0.00	0.03			
A25	A	0.44	0.50	0.03	0.22	0.41	0.16		
	B	0.44	0.50	0.13	0.09	0.38	0.03		
	C	0.25	0.34	0.13	0.00	0.38	0.00		
A30	A	0.13	0.13	0.22	0.31	0.09	0.38	0.06	
	B	0.09	0.13	0.09	0.38	0.00	0.03	0.06	
	C	0.06	0.13	0.00	0.41	0.00	0.00	0.00	
A36	A	0.34	0.25	0.22	0.34	0.44	0.22	0.50	0.38
	B	0.09	0.28	0.16	0.16	0.16	0.13	0.31	0.31
	C	0.22	0.19	0.13	0.13	0.28	0.06	0.25	0.28
		A08	A09	A11	A12	A23	A24	A25	A30

sign test showed that subject A made significantly more misclassification than the other two subjects ($p<0.001$). The proportions of misclassification of the three subjects for any one comparison differed on average by 0.11, and these differences ranged from 0 to 0.31. Subjects A and C said that they used mainly acoustic and some visual cues to do the discrimination, subject B said she relied mainly on the spectrogram, with some acoustic cues. Figure 5 shows that all three subjects grouped the calls of the nine matrilineal groups into three major clusters which correspond to the clusters generated by the neural network index (Fig. 4). However, the results from individual subjects differ in the relationship of matrilineal groups within the three clusters, as well as in the positions of the clusters with respect to each other.

The association matrix for the nine matrilineal groups is given in Table III. Association indices range from 0.14 for A09 and A11 subpod to 0.95 for A11 and A24 subpod. The average linkage dendrogram (Fig. 6) shows that their association patterns group the nine matrilineal groups into the same three clusters as the acoustic analyses, with the difference that the A36 subpod associates more often with A12 while being acoustically more similar to A08–A09–A23–A25.

Table IV gives the correlation matrix of the ratings of

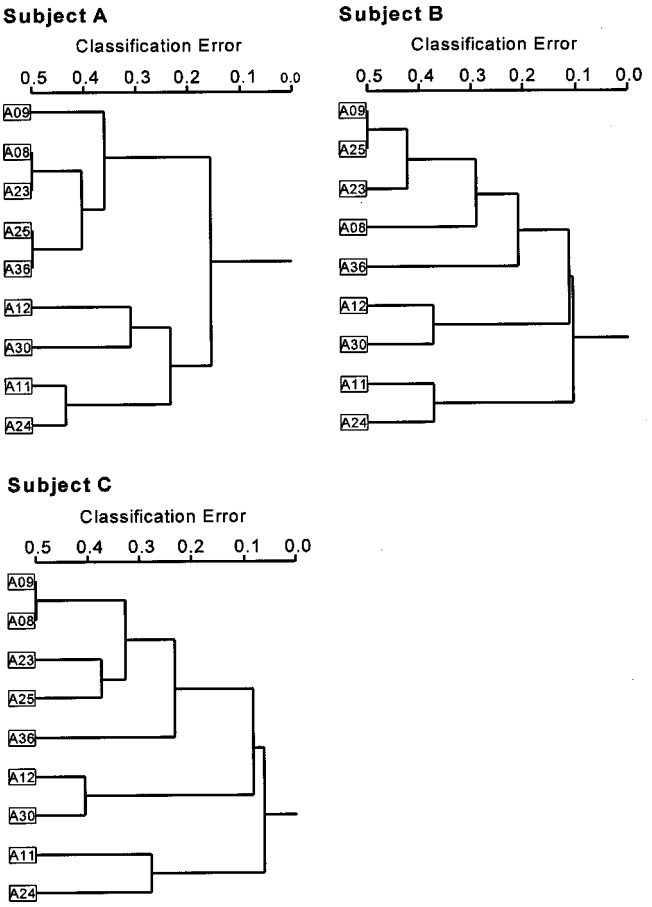

FIG. 5. Average linkage dendrogram giving acoustic similarity of the N4 call of the nine matrilineal groups based on the classification errors of the three subjects (generated from Table II). Comparisons of N4 calls from groups which are linked on the left-hand side of the graph gave higher classification errors (suggesting higher similarity) than those linked on the right-hand side

acoustic similarity by the three subjects, by the neural network, as well as of the association indices. All correlations are significant with $p<0.001$ (Mantel's test). Two correlation coefficients comparing ratings of different subjects (subjects A and B; subjects A and C) are lower than the correlation coefficients comparing human subject ratings and the neural network indices. All measures of acoustic similarity gave significant correlations with the groups' association indices.

TABLE III. Association matrix for the nine matrilineal groups. The values give the half-weight index of association.

A09	0.73							
A11	0.42	0.24						
A12	0.51	0.40	0.41					
A23	0.77	0.69	0.33	0.47				
A24	0.42	0.25	0.97	0.41	0.32			
A25	0.69	0.61	0.27	0.37	0.78	0.26		
A30	0.33	0.25	0.33	0.54	0.35	0.33	0.27	
A36	0.40	0.38	0.34	0.52	0.39	0.33	0.27	0.37
	A08	A09	A11	A12	A23	A24	A25	A30

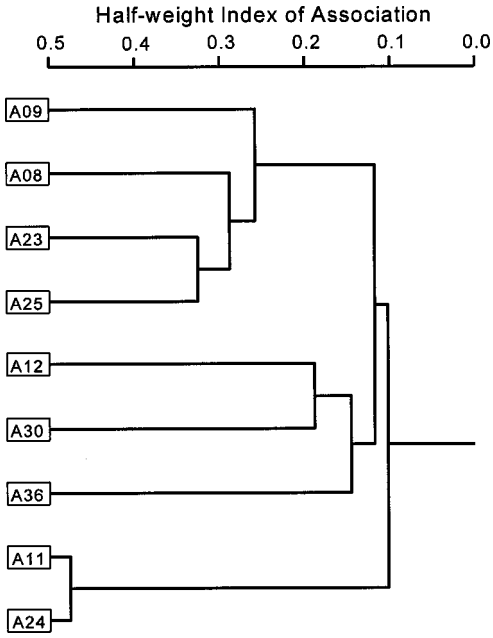

FIG. 6. Average linkage dendrogram giving association patterns of the nine matrilineal groups based on the half-weight index of association (generated from Table III). Groups which are linked on the left-hand side of the graphs spend more time traveling together than those linked on the right-hand side.

III. DISCUSSION

The contour extraction algorithm based on autocovariance in the frequency domain proved good at extracting pulse-rate contours even from recordings with poor signal-to-noise ratios. Unless the noise itself had harmonic content, it was canceled out in the autocovariance sequence, whereas the harmonic signals were amplified. We suggest that pulse-rate contours are an effective way to describe pulsed vocalizations and believe that this algorithm would be useful for extracting contours from noisy recordings of the pulsed calls of a wide variety of species.

The shortcomings of this algorithm are that it cannot be applied to broadband or pure-tone signals, and that compared to alternative algorithms, it is computationally expensive. Mixed signals, however, can still be analyzed by switching to another algorithm (e.g., that of Buck and Tyack, 1993) if the autocovariance sequence fails to detect harmonic content. Recent developments in computer hardware are likely to further reduce computing time, making real-time extraction of pulse-rate contours a possibility.

The advantage of analyses of acoustic similarity based on frequency contours over those based on isolated measurements of the spectrogram lies in the fact that analysis of frequency contours requires no, or very little, prior knowledge of where to expect the differences in the signals (Bailey, 1978). Subtle and very localized differences between two signal patterns are easily missed in conventional analyses by taking measurements of a limited number of structural variables. Unlike discrimination and classification analyses of bioacoustic signals where the input is the waveform (e.g., Neumann *et al.*, in press), or the spectrogram (e.g.

TABLE IV. Correlation matrix giving matrix correlation coefficients for the ratings of acoustic similarity by the human subjects, the neural network index, and the association indices of the nine matrilineal groups. All correlations are significant with $p<0.001$ (Mantel test).

Similarity ratings					
Subject A	1				
Subject B	0.63	1			
Subject C	0.60	0.79	1		
Neural network index	0.69	0.78	0.71	1	
Index of association	0.57	0.67	0.66	0.54	1
	Subject A	Subject B	Subject C	Neural network index	Index of association
		Similarity ratings			

Spong *et al*., 1993; Erbe *et al*., in press), amplitude information is excluded from the analysis of frequency contours. Although this may be a disadvantage in some studies analyzing recordings obtained in controlled environments, it will prove beneficial in others where differences in recording equipment and in the composition of background noise introduces spurious variability into the data. In the study of Spong *et al*. (1993), for example, it cannot be ruled out that the neural network discriminated along differences in background noise composition rather than individual-specific vocal differences.

Although the ratings of similarity by human subjects agreed on a fundamental level, this study suggests that individual human subjects perceive similarity of killer whale calls differently. The ratings of similarity disagree between subjects in the acoustic relationships of matrilineal groups within the three clusters, as well as in the position of these clusters with respect to each other. The subject who had never played a musical instrument had significantly higher classification errors than the other two subjects, which may suggest that the amount of musical exposure contributes to observer bias (see also Halpern *et al*., 1995; Baribeau *et al*., 1996; Halpern *et al*., 1996).

Comparing the ratings of acoustic similarity by the neural network with those of the human subjects shows that both ways of quantifying acoustic variation gave similar results. The matrix correlation coefficients (Table IV) suggest that the differences between ratings from individual subjects are greater than are differences between subject ratings and the neural network index. Since multiple independent neural networks are trained on the same problem in each comparison, the neural network index will give essentially identical results given the same input data. The neural network therefore represents an objective and repeatable means of measuring acoustic similarity, and allows the comparison of results across studies, species, and time.

Like discriminant function analysis (e.g., Job *et al*., 1995), or analysis of confusion frequencies (e.g., Miller and Nicely, 1955; Loesche *et al*., 1992), the neural network index of acoustic similarity is based on the premise that similarity and discrimination are inversely related. All three methods rate patterns as similar if the analysis is unable to tell them apart, and conversely consider patterns distinct if the analysis can consistently discriminate between them. This concept of similarity differs from that underlying other methods which use the geometric distance between two patterns as a measure of their similarity. Examples for the latter are cross correlation (e.g., Clark *et al*., 1987), and cluster analysis (e.g., McCowan, 1995). Arguably the first concept of similarity is more applicable to the study of communication, since the information value of a signal is largely determined by how well a receiving animal can distinguish it from other signals (Beecher, 1989).

The training procedure, which involves error back-propagation to discriminate between different patterns, is essentially a self-organizing process and does not depend on strictly linear relationships in the input data. For this reason a neural network index will be able to detect and integrate differences between the input patterns that would be missed by most conventional statistical analyses. Research into biological neural systems suggests that these also operate in a nonlinear and self-organizing way (Kelso, 1995), which may explain why a neural network based approach is often the best way to model biological signal processing tasks (Hunt, 1993; Erbe *et al*., in press). The fact that the neural network index of acoustic similarity shows a significant correlation with the association patterns of the different matrilineal groups suggests that the index rates acoustic similarity in a biologically meaningful way.

An index of acoustic variation based on neural network analysis can be viewed as a hybrid between statistical and perceptive approaches of measuring acoustic similarity. It combines the objectivity and repeatability of a strictly statistical approach with the self-organizing nonlinear nature of acoustic perception and biological signal processing, and therefore holds great potential in the study of human and animal communication.

IV. CONCLUSIONS

This study demonstrates that autocovariance in the frequency domain is a useful way to extract contours of the pulse-repetition rate from noisy recordings of pulsed signals. This study also shows that discrimination of frequency contours using a back-propagation neural network is an effective and repeatable way to measure the similarity of animal sounds. The significant correlation between the neural network based acoustic similarity index and a biological param-

eter, the groups' association patterns, suggests that the index assesses acoustic similarity in a biologically meaningful way.

ACKNOWLEDGMENTS

This study would not have been possible without the contribution of recordings by D.E. Bain, J. Borrowman, D. Briggs, G. Ellis, B. and D. Mackay, A. Morton, F. Thomsen, and S. Wischniowski, and we are very grateful for their generosity. We would like to thank L. Farroway, S. Midwinter, and C.A. Shankel for acting as research subjects for the discrimination tasks, as well as D.E. Bain, J.M. Gosline, J.M.N. Smith, and H. Symonds for their suggestions and comments throughout the study. Earlier drafts of this article benefited greatly from comments by L. Barrett-Lennard, L. Deecke, V. M. Janik, D. Kapan, P.J.B. Slater, H. Symonds, and R.M. Williams, and two anonymous reviewers. We would like to thank The MathWorks, Inc. for generously donating the software and DKMC for logistic support.

Bailey, K. (**1978**). ''The structure and variation of the separation call of the bobwhite quail (*Colinus virginianus*, Odontophorinae),'' Anim. Behav. **26**, 296–303.

Baribeau, J., Berman, B., Atkin, A., and Roth, R. M. (**1996**). ''Musical experience and auditory P300 in a divided attention task,'' Brain Cogn. **30**, 378–380.

Beecher, M. D. (**1989**). ''Signalling systems for individual recognition: an information theory approach,'' Anim. Behav. **38**, 248–261.

Bigg, M. A., Olesiuk, P. F., and Ellis, G. M. (**1990**). ''Social organization and genealogy of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State,'' Rep. Int. Whaling Comm. Spec. Issue No. **12**, 383–405.

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

Clark, C. W., Marler, P., and Beeman, B. (**1987**). ''Qualitative analysis of animal vocal phonology and application to swamp sparrow song,'' Ethology **76**, 101–115.

Dasgupta, C. (**1991**). ''Neural networks: An overview,'' J. Indian Inst. Sci. **71**, 491–502.

Deng, L. (**1992**). ''Processing of acoustic signals in a cochlear model incorporating laterally coupled suppressive elements,'' Neural Networks **5**, 19–34.

Erbe, C., King, A. R., Yedlin, M., and Farmer, D. M. (**in press**). ''Computer models for masked hearing experiments with beluga whales (*Delphinapterus leucas*),'' J. Acoust. Soc. Am.

Ford, J. K. B. (**1984**). ''Call traditions and vocal dialects of killer whales (*Orcinus orca*) in British Columbia,'' Ph.D. Dissertation (University of British Columbia, Vancouver).

Ford, J. K. B. (**1989**). ''Acoustic behaviour of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia,'' Can. J. Zool. **67**, 727–745.

Ford, J. K. B. (**1991**). ''Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia,'' Can. J. Zool. **69**, 1454–1483.

Ford, J. K. B., Ellis, G. M., and Balcomb, K. C. (**1994**). *Killer Whales—the Natural History and Genealogy of Orcinus orca in British Columbia and Washington State* (UBC Press, Vancouver).

Ginsberg, J. R., and Young, T. P. (**1992**). ''Measuring associations between individuals or groups in behavioural studies,'' Anim. Behav. **44**, 377–379.

Halpern, A. R., Bartlett, J. C., and Dowling, W. J. (**1995**). ''Aging and experience in the recognition of musical transpositions,'' Psychol. Aging **10**, 325–342.

Halpern, A. R., Kwak, S., Bartlett, J. C., and Dowling, W. J. (**1996**). ''Effects of aging and musical experience on the representation of tonal hierarchies,'' Psychol. Aging. **11**, 235–246.

Hinton, G. E. (**1992**). ''How neural networks learn from experience,'' Sci. Am. **268**, 145–151.

Hunt, E. (**1993**). ''A proposal for computer modelling of animal linguistic comprehension,'' in *Language and Communication: Comparative Perspectives*, edited by H. L. Roitblat, L. M. Herman, and P. E. Nachtigall (Lawrence Erlbaum, Hillsdale, NJ), pp. 85–94.

Horn, A. G., and Falls, J. B. (**1996**). ''Categorization and the design of signals: the case of song repertoires,'' in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Comstock Publishing, Ithaca, NY), pp. 121–135.

Job, D. A., Boness, D. J., and Francis, J. M. (**1995**). ''Individual variation in nursing vocalizations of Hawaiian monk seal pups, *Monachus schauinslandi* (Phocidae, Pinnipedia), and lack of maternal recognition,'' Can. J. Zool. **73**, 975–983.

Johnson, S. C. (**1967**). ''Hierarchical clustering schemes,'' Psychometrika **32**, 241–54.

Kelso, J. A. S. (**1995**). *Dynamic Patterns: The Self-Organization of Brain and Behavior* (MIT Press, Cambridge, MA).

Loesche, P., Beecher, M. D., and Stoddard, P. K. (**1992**). ''Perception of cliff swallow calls by birds *Hirundo pyrrhonota* and *Sturnus vulgaris* and humans *Homo sapiens*,'' J. Comp. Psych. **106**, 239–247.

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

Martindale, S. (**1980**). ''On the multivariate analysis of avian vocalizations,'' J. Theor. Biol. **83**, 107–110.

Miller, E. H. (**1979**). ''An approach to the analysis of graded calls of birds,'' Behav. Neural Biol. **27**, 25–38.

Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.

Neumann, E. K., Wheeler, D. A., Bernstein, A. S., Burnside, J. W., and Hall, J. C. (**1992**). ''Artificial neural network classification of *Drosophila* song mutants,'' Biol. Cybern. **66**, 485–469.

Ramani, N., Hanson, W. G., Patrick, P. H., and Sheehan, R. W. (**1993**). ''An automated environmental monitoring system—phase 1. amphibian species identification from calls,'' Proceedings of the World Conference on Neural Networks **1**, 304–307.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (**1986**). ''Learning internal representation by error propagation,'' in *Parallel Data Processing, Vol. 1*, edited by D. E. Rumelhart and J. McLelland (MIT Press, Cambridge, MA), pp. 318–362.

Sayigh, L. S., Tyack, P. L., Wells, R. S., and Scott, M. D. (**1990**). ''Signature whistles of free-ranging bottlenose dolphins *Tursiops truncatus*: Stability and mother-offspring comparisons,'' Behav. Ecol. Sociobiol. **26**, 247–260.

Schevill, W. E., and Watkins, W. A. (**1966**). ''Sound structure and directionality in *Orcinus* (killer whale),'' Zoologica **51**, 70–76.

Spong, P., Symonds, H., Gaetz, W., Jantzen, K., and Weinberg, H. (**1993**). ''A neural network method for recognition of individual *Orcinus orca* based on their acoustic behavior: Phase 1,'' Paper presented at Oceans 1993, Conference of the IEEE, Victoria, B.C.

Symmes, D., Newman, J. D., Talmage-Riggs, G., and Katz Lieblich, A. (**1979**). ''Individuality and stability of isolation peeps in squirrel monkeys,'' Anim. Behav. **27**, 1142–1152.

Tyack, P. L. (**1986**). ''Whistle repertoires of two bottlenose dolphins (*Tursiops truncatus*): Mimicry of signature whistles?'' Behav. Ecol. Sociobiol. **18**, 251–257.

Vogl, T. P., Mangis, J. K., Ziegler, A. K., and Alkon, D. L. (**1988**). ''Accelerating the convergence of the backpropagation method,'' Biol. Cybern. **59**, 257–263.

Watkins, W. A. (**1967**). ''The harmonic interval: fact or artifact in spectral analysis of pulse trains,'' in *Marine Bioacoustics, Vol. 2*, edited by W. N. Tavolga (Pergamon New York), pp. 15–43.

Williams, J. M., and Slater, P. J. B. (**1991**). ''Computer analysis of bird sounds: a guide to current methods,'' Bioacoustics **3**, 121–128.