An Approach to the Analysis of Graded
Vocalizations of Birds

EDWARD H. MILLER

Biology Department, York University, Downsview, Ontario M3J 1P3, Canada

The purpose of this paper is to explore some techniques of representing and
analyzing sound spectrograms, with reference to detecting and characterizing
graded vocalizations. Description and classification of sounds based on their
structure can be separate from research into their functions, and this paper
considers affinities among calls as determined by their appearance as sound
spectrograms only. Calls of two bird species [Least Sandpiper, Calidris minutilla
(Scolopacidae), and Patagonian Black Oystercatcher, Haematopus ater
(Haematopodidae)] are described by a simple method of hand digitization. A set of
similarity measures among calls is derived for each sample of calls, based on the
digitized representations, and multivariate analyses are performed on each of these
sets. Structurally graded series seem to be represented clearly and accurately in
graphical solutions from an ordination procedure (multidimensional scaling),
coupled with results from hierarchical single-linkage clustering. The sandpiper
calls form a simple system of adjacent grading; the oystercatcher calls show only a
tendency toward grading of adjacent elements.

Bird sounds cannot be described and classified accurately without
equipment like the sound spectrograph. This portrays sounds as visual
patterns in the frequency-time plane, whose complexity can be examined
and measured easily. It is common practice to classify bird sounds after
inspecting their sound spectrograms (sonagrams), then to characterize the
classes quantitatively after measuring representative sonagrams for each.
This approach of subjective classification and subsequent measurement is
widespread because it is easy: The great variation among sounds and
among individuals renders numerical classification tedious and usually
demands large sample sizes. Not surprisingly, workers have tended to
eucidate fine structure of representative calls (e.g., Brackenbury, 1978;
Greenewalt, 1968; Latimer, 1977), or have employed numerical tech-

1 Field research and data analysis were supported by the Natural Science and En-
gineering Research Council of Canada (operating grants to I. A. McLaren, Dalhousie
University, and E. H. Miller, York University). I thank Dr. McLaren for encouragement
and financial support during early parts of this work, and Drs. A. R. Gibson and W. M.
Schleidt for their detailed reviews of manuscript drafts.
niques using chosen features of samples of calls (e.g., Bourgeois, 1977; Bourgeois & Couture, 1977; Goldstein, 1978; Hafner & Hafner, 1979; Payne, 1978; Sparling & Williams, 1978; Zann, 1976). These approaches can be combined through analog-to-digital conversion of a taped signal and computer analysis of the digitized waveform. This is common in the study of evoked potentials (e.g., Bennett, MacDonald, Drance, & Ueno, 1971; Donchin, 1966; Schwartz, Ramos, & John, 1976), and will undoubtedly become widely used in bioacoustics too. Zero-crossing analysis gives rapid detailed analysis of sounds of birds (especially whistled sounds) (Staddon, McGeorge, Bruce, & Klein, 1978), and can be used in classifying sounds automatically by computer (Rück, Meier, & Steppuhn, 1977). At present, however, the sound spectrograph is the main tool for depicting bird sounds, and it will continue to be useful for visual portrayal of sounds and for study of sounds on noisy tapes, which are usual in naturalistic studies.

One danger in classifying sonograms subjectively is that we may subdivide a sample of calls whose acoustic morphology varies continuously. A further complication is that most communication systems probably lie between those with discrete categories of signals and those with continuous variation only. Sounds of the Chimpanzee (Pan troglodytes) illustrate this: Cough and Laughter are always “typical” in form, Waa Bark occurs as “intermediate” forms 52% of the time, and 10 other classes of sounds show levels of grading between these (Marler, 1976, Table 7; Marler & Tenaza, 1977, Table 16). It is clear that explicit methods of quantitative description and classification of such sounds are needed. In addition, variation in different characteristics of single classes of sounds needs to be explored. For example, the small monkey species Cebus capucinus has a discrete system of vocal communication which includes a trill-like greeting signal. This shows stronger grading in the duration of interpulse intervals than of pulses (Mal'tsev, 1971). Applications of statistics reviewed by Van Valen (1978) may be useful here.

Extensive grading in communication systems is probably an adaptive feature, and offers advantages lacking in one comprised of discrete categories (see Green, 1975; Marler, 1975, 1976; Marler & Tenaza, 1977; Morton, 1977; Morton & Shalter, 1977). Hence, species with such a system should “process their graded vocalizations in noncategorical fashion for this advantage to be realized” (Marler, 1975, p. 31; but see Snowdon & Pola, 1978). Knowledge of how sound signals are perceived can indicate how accurately a physical description corresponds to the perceptual capacity of a species, what transformations must be applied to the physical description to improve its correspondence with perceptual capacities, etc. However, these considerations do not need to influence initial description and classification of calls. We should not bias our choice of characters and scales of measurement toward those of sus-
pected or possible functional (i.e., perceptual) importance, for several reasons. First, we do not have adequate knowledge of hearing abilities in birds to apply perceptually "correct" transformations of scale, or to measure only perceptually important variables. In addition, there are major species differences in hearing abilities. Finally, it may be desirable to have information on characteristics which are evolutionarily conservative (e.g., in taxonomic studies), and these are probably to be found in features of sounds which are perceptually unimportant or undetectable. In summary, we should strive for relatively complete, unbiassed quantitative descriptions of sounds, unless the descriptions are being made for a restricted purpose. Marshall (1977) suggests that a logarithmic scale of frequency be used when preparing sonograms, because it represents how birds perceive sounds better than does a linear scale. This rationale exemplifies the kind of bias I refer to. It is legitimate to choose a logarithmic or any other kind of scale, but the choice should not be determined just by considerations of signal function.

In practice, such biases are hard to avoid entirely. The sound spectrograph was developed to aid in the study of functionally important speech patterns in humans, and it displays a sound's frequency over time. Amplitude was represented crudely in early sound spectrographs. More complete displays of amplitude, frequency, and time can be obtained with modern equipment, but these are still portrayed infrequently (e.g., Ford & Fisher, 1978; Singleton & Poulter, 1967). I present some preliminary analyses of sonograms in this paper, acknowledging that they are somewhat limited and biased representations of sounds.

I am concerned here with the objective classification of graded sounds based upon their visual characteristics as sonograms, and of their structural affinities to one another. My approach is to retain as much information as possible in descriptions of the calls chosen for analysis. The call types examined here are brief, so I employ a method of hand digitization with small cell sizes rather than measuring each call on a set of rectilinear variables (e.g., duration, frequency range, etc.), which would be few and prone to large measurement error. Furthermore, I assume that the appearance of sonograms is related systematically to sound characteristics. This includes some artifacts and inaccuracies but these form part of the digitized "description" anyway, for they presumably occur across calls nonrandomly and hence contribute meaningfully to call descriptions. It is unimportant for descriptive purposes that such "noise" may be meaningless functionally.

Two different call samples were analyzed: (1) Two brief series of calls from within a complex kind of call ("Song") of a sandpiper species (Fig. 2). These series were chosen because they suggest a clearly graded sequence; and (2) a collection of calls which vary much and seem to
intergrade, but for which I could not derive a satisfactory classification by just inspecting sonagrams (oystercatcher: Fig. 4).

METHODS

Samples of calls were chosen from two species of shorebirds I am studying currently. Songs of the Least Sandpiper, *Calidris minuilla* (Scolopacidae), were recorded on Scotch tape 176, with a Uher 4200 Report Stereo 1C tape recorder and a Uher M517 dynamic microphone mounted in a Dan Gibson parabolic reflector. They were taped on Sable Island, Nova Scotia, in spring, 1976. Calls of the Patagonian Black Oystercatcher, *Haematopus ater* (Haematopodidae), were recorded on Punta Tombo, Argentina, in November, 1977. I used a Nagra IS tape recorder and Sennheiser MKH 816 “shotgun” microphone and Scotch tape 208, matched to the Nagra. Recording speed for both species was 19 cm sec⁻¹. Sonagrams were prepared on a Kay Sonograph 7029A, using the FL-1 setting and wide band (300 Hz) filter. The frequency scale in Fig. 2 is based on narrow band (45 Hz) calibration (see Greenewalt, 1968, p. 8).

One song from each of two male sandpipers and 24 calls from a probable female oystercatcher were analyzed. The oystercatcher calls constituted part of a long series given by the bird near its nest and eggs, in my presence. They were emitted irregularly at intervals from about 0.1 to 1.6 sec. For behavioral observations on vocalizations of the two species, see Miller (1977, 1979) and Miller & Baker (1979).

The fundamental frequency of each call was traced from the sonagrams and the tracing was overlain on graph paper, left-justified to a reference mark, and adjusted vertically to a reference frequency marker. The call was represented by a matrix of 0, 1, and 2 according to its coverage of cells on the grid, as follows: if more than half of a cell was covered by the tracing it was scored 2; if a cell was covered to less than half, it was scored 1; and if a cell was not covered at all by the tracing, it was scored 0 (Fig. 1). The number of cells scored was set by the maximal duration and maximal and minimal frequencies over the entire sample of calls; thus, all calls had many cells with “0” scores. Examples of scoring are shown in Fig. 1.

To assess similarities among calls I compared them pairwise on the basis of their scores across all cells; i.e., each cell was treated as a single character. Cells that were invariant over the sample of calls were disregarded (e.g., those cells with “1” in every call, etc.). Using the remaining cells, similarity was defined as follows: For those cells that took only two states (0 and 1 or 0 and 2, or 1 and 2) in the entire sample of calls, matches counted 1 and mismatches counted 0; for those cells that took three states (0 and 1 and 2) in the call sample, matches counted 1, partial matches counted 0.5, and mismatches counted 0. Between any two calls then, my measure of overall similarity (hereafter S) was simply the sum of these
Fig. 1. Illustration of the scoring method used to digitize calls by hand. In (A) is shown a call overlain on a grid. Grid dimensions were established by the maximal and minimal frequencies recorded in the sample of calls. Cells of the grid are scored as 0, 1, or 2 according to whether they are covered by none, less than half, or more than half of the tracing, respectively. Examples of digitized representations arrived at in this manner are shown in (B), for two calls of *H. ater* from a long series. The matrix labeled MASTER simply indicates which cells of the grid are invariant ("1"), which take two values ("2"), and which take three values ("3") over a large sample of oystercatcher calls. Only cells in MASTER designated by "2" or "3" are used in computing similarity measures.

cell-by-cell scores. I scaled the S values over the range 0 to 1 for the following analyses.

A symmetric proximity ("correlation") matrix of S values was obtained for each data set. The matrices were analyzed through hierarchical single-linkage clustering (program HICLUST of Bell Laboratories; see Johnson, 1967), and nonmetric multidimensional scaling, or MDS (program MDSCAL, version 5MS, of Bell Laboratories; see Kruskal, 1964a, 1964b). MDSCAL was run several times for each data set, using different starting configurations. This helps to avoid suboptimal solutions, but may be slightly less effective than using a starting configuration generated through a principal components analysis (Arabie 1978a, 1978b; Spence and Young, 1978). General accounts of clustering and MDS with ethological applications are given by De Ghett (1978), Morgan, Simpson, Hanby, & Hall-Craggs (1976) and Spence (1978); Blashfield & Aldenderfer (1978) provide a useful overview of the uses and proliferation of cluster-analytic techniques. As used here, MDS is a straightforward ordination procedure
which attempts to position calls that are similar to one another (high S value) close together and dissimilar calls far apart, in a space of few dimensions. Success at achieving this is reflected in “stress” values at each dimensionality: low stress indicates that a good fit between the similarity structure in the S matrix and positioning of the calls in space has been obtained. See Spence (1978) for further details. Stress values given below were computed with stress formula 2.2

The 24 oystercatcher calls defined a 16 by 9 matrix with 109 cell comparisons [i.e., there were 109 with varying states in the call sample; thus \((16 \times 9) - 109 = 35\) cells were invariant across the sample]. For sandpiper male 1, there were 68 cell comparisons in a 10 by 9 matrix; for male 2, there were 77 comparisons in a 12 by 8 matrix.

RESULTS

Sonagrams of the sandpiper songs are shown in Fig. 2. The graded elements in each song are designated by numbers corresponding to the points in the MDS plots (Figs. 2A, B) and clustering solutions (Fig. 3). One dimension represents the data for male 1 well, and his calls fall along this dimension much as they are ordered in the song [Kendall’s rank correlation coefficient = 0.78, \(p\) (two tailed) = 0.001]. Male 2 had a similar number of calls of this type in his song, but two dimensions are needed to reduce stress to a satisfactory level. His calls fall along the second dimension close to their order in the song [Kendall’s rank correlation coefficient = 0.56, \(p\) (two tailed) = 0.016]. Terminal calls in the songs of both males seem to be more similar to one another than are initial calls, to judge from their relative spacing in the MDS plots. This is shown more clearly in the dendrograms in Fig. 3, where terminal calls form the first clusters, and calls from increasingly earlier in the series are added successively.

The ordination and clustering solutions portray the graded nature of the call series within sandpiper songs well; this is probably a fairly typical example of grading of adjacent elements in bird songs (see Marler, 1976; Marler & Tenaza, 1977). I will now apply these techniques to the more complex and seemingly heterogeneous sample of oystercatcher calls.

Oystercatcher calls show suggestions of grading in the first two dimensions of the MDS solution, but grading is most apparent in the second and third dimensions (Fig. 4). Several clusters are apparent in Fig. 4, at a level of \(S \geq 0.70\) (i.e., all calls in each cluster have a similarity level of at least 0.70 with respect to one or more calls in the same cluster). The largest of these clusters includes calls which are strongly bi-peaked (No. 21), weakly bi-peaked (No. 5), and single-peaked with strongly ascending

2 Kruskal, J. B., and Carmone, F. How to Use M-D-SCAL, a Program to Do Multidimensional Scaling and Multidimensional Unfolding. Unpublished ms., Bell Telephone Laboratories, Murray Hill, N. J.
Fig. 2. Upper: Sonagrams of songs from two males of *C. minuta*. The numbers beneath the graded series in each song designate particular calls, and are referred to in the MDSCAL solution below. Lower: (A) and (B) represent one- and two-dimensional MDSCAL solutions for graded calls from the songs of the two males. The sequences of calls along dimension I (for male 1) and II (for male 2) were tested for correlation with their sequence in the songs (for example, male 1’s sequence along dimension I is 1-2-3-4-5-9-10-7-8, which was tested for correlation with the sequence 1-2-3-4-5-6-7-8-9-10; see text).

frequency (No. 10). The paths from calls Nos. 21 to 5, and 5 to 10 are roughly orthogonal and parallel to the second and third dimensions of the MDS solution, respectively. This may indicate that the call complex around No. 5 is a transition point between two systems of grading (21–5 and 5–10). At the very least, the cluster reveals a smooth gradation between bi-peaked and single-peaked calls.

Two other points bear mention. First, a small cluster of single-peaked calls occurs (12–22). These are much briefer than similar calls surrounding call No. 10, and appear to ascend in frequency less strongly. Subjectively, these might have been grouped with the complex around that call.
Fig. 3. Dendrograms representing clustering solutions obtained with HICLUSST (using "connectedness"
"minimum", or single-linkage method; see Blashfield & Aldenderfer, 1978), for two males of *C. minuella*. The
numbers correspond to calls in Fig. 2 (upper).
Second, many calls owe their membership in the large cluster to their high similarity with calls Nos. 5 and 10. These two calls may therefore represent modes around which grading can occur in several ways.

Grading in the sample of oystercatcher calls is suggested strongly by the analytic results. There is continuous grading from bi-peaked to single-peaked calls, at least; the number of call types recognized would depend upon the $S$ level chosen to delimit clusters. These calls are not emitted in a rapid rhythmic sequence, unlike in the sandpiper song. It is therefore of interest to examine similarity measures between adjacent oystercatcher calls, to determine whether an adjacent system of grading exists (see Marler, 1975, 1976; Marler & Tenaza, 1977). This was done by examining a long series of calls from which 89 $S$ values between calls in the same series were chosen randomly. Calls next to one another in the sequence tended to be much more similar to one another than were calls chosen...
randomly $0.01 < p$ (one tailed) $< 0.025$, by median test; Fig. 5). Thus, this system of grading lies between the extremes of adjacent and nonadjacent systems.

DISCUSSION

Digitization and Similarity

There are many ways to assess similarity among objects (e.g., calls). Usually calls are measured on several rectilinear variables which are then used in computing a similarity measure (see Introduction). This is useful because patterns of clustering and ordination among calls may be interpretable in terms of their physical characteristics; this approach is used widely in exploring the structure and causation of behavior (e.g., Aspey & Blankenship, 1976; Balthazart, 1973; Conley, 1976; Mihok, 1976; Miller, 1975; Robertson & Sale, 1974; van Hooff, 1975). If similarity measures between objects are based on undefined characters (e.g., Rohlf & Sokal, 1967; Sokal & Rohlf, 1966; this study), then only Q-approaches are possible (see Sneath & Sokal, 1973). The distinction between Q-technique (the study of similarity relations among objects) and R-technique (the study of similarity relations among variables or characters) is largely artificial. Nevertheless, the distinction is useful to maintain in studies like mine, in which post hoc assessment of affinities in terms of particular characteristics of calls is impossible. The purposes of my study

![Figure 5](image-url)  
**Fig. 5.** Frequency histogram for 89 $S$ values between adjacent calls and between randomly chosen calls, in a long sequence of calling in *H. ater.*
were to determine whether grading could be characterized adequately by a digitized representation, and could be detected through simple multivariate procedures; questions about the nature and function of this grading must be approached separately (see Introduction).

I tried to retain the approximate shape of sonagrams by digitizing them and comparing the digitized representations with one another (see Bertram, 1970; Field, 1976; Schleidt, 1974). Koepp, Hoffman, & Nadler (1978) use a similar technique. There are many other methods for characterizing shape quantitatively (e.g., Cheetham & Lorenz, 1976; Gates, 1977, 1978a, 1978b; Ramaekers, 1975; Waters, 1977; Younker & Ehrlich, 1977), but these take more time and rely upon a level of accuracy and reproducibility in measurements which sonagrams of brief calls cannot assume comfortably.

There are two simple changes to my approach that may be desirable: (1) shift the calls relative to one another along the time axis until they show maximal overlap or similarity. This would allow for the possibility that a call is compound and includes another call type entirely [e.g., compare call No. 5 and its relatives (Fig. 4) with call No. 60 (Fig. 1)]. Comparable techniques are used in geology and biochemical genetics, in which one sequence is slid past another in steps, and matches and mismatches at each step are examined (e.g., Gibbs & McIntyre, 1970; Sackin, 1971; Sackin & Merriman, 1969). It would be necessary to modify or replace my similarity measure in such a case. (2) Center calls in time, or frequency, or both, before comparing them. This alteration would imply that frequency and time components of the digitized representation are arbitrary to some extent. This seems unrealistic for frequency components, though for time components the method of left justification may weight the similarity measure too much by how calls start. It may therefore be desirable to center calls horizontally before comparing them.

The similarity measure used could be changed. As defined it considers only matches and mismatches at varying positions, and these positions depend upon variation in the sample of calls. Similarity between two calls increases if one adds an aberrant call (e.g., long or of high frequency) to the sample, and declines if one deletes such a call from the sample. Similarity measures that are not influenced by the range of variation in the sample may be preferred.

General Comments

The method of hand digitization is useful for small to moderate samples of narrow band calls, and is thus applicable mainly to such sounds as bird calls, bat ultrasound, and so on. The digitization procedure stands apart from the procedures used to compute similarity, and from the clustering and ordination procedures used. If one is faced with problems in assessing grading among calls or in classifying them on the basis of sonagrams, then
it is desirable to retain as much information as possible in call descriptions. Certainly more (but different) information is retained if calls are digitized using small cell sizes than if they are measured on a set of rectilinear variables. It is more desirable that calls be treated as sounds than as visual patterns of course. Nevertheless, even with the advent of minicomputer applications in analysis of animal sounds, and good filtering techniques, it is likely that many modest problems with sounds recorded in field conditions will continue to be tackled as if they were visual patterns (i.e., as sonagrams).

REFERENCES

ANALYSIS OF GRADED CALLS OF BIRDS


Protistologica, 14, 125–132.


use with amino acid and nucleotide sequences. European Journal of Biochemistry, 16, 
1–11.

Goldstein, R. B. (1978). Geographic variation in the “hoy” call of the bobwhite. Auk, 95, 
85–94.

Green, S. (1975). Variation of vocal pattern with social situation in the Japanese monkey 
Academic Press.

Institution.


Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric 
hypothesis. Psychometrika, 29, 1–27.

ometrika, 29, 28–42.


Mal’tsev, V. P. (1971). Some types of acoustical signals of capuchin monkeys (Cebus 

Cutting (Eds.), The Role of Speech in Language, pp. 11–37. Cambridge, Massachu 
setts: M.I.T. Press.

Marler, P. (1976). Social organisation, communication and graded signals: The chimpanzee 
and the gorilla. In P. P. G. Bateson and R. A. Hinde (Eds.), Growing Points in Ethology, 

Bloomington: Indiana Univ. Press.


Mihok, S. (1976). Behaviour of subarctic red-backed voles (Clethrionomyys gapperi athabas 

Miller, E. H. (1975). Walrus ethology. I. The social role of tusks and applications of 


Miller, E. H. (1979). Functions of display flights by males of the least sandpiper, Calidris 
minutilla (Vieill.) on Sable Island, Nova Scotia. Canadian Journal of Zoology, 57, 
176–193.

ms. (available from E. H. M. upon request).


