Bioacoustics
The International Journal of Animal Sound and its Recording, 2000, Vol. 11, pp. 89–127
0952-4622/00 $10

THE COMPARISON OF HARMONICALLY RICH SOUNDS USING SPECTROGRAPHIC CROSS-CORRELATION AND PRINCIPAL COORDINATES ANALYSIS

KATHRYN A. CORTOPASSI* AND JACK W. BRADBURY†

Cornell Laboratory of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850-1999, USA

ABSTRACT

We explore the effectiveness of spectrographic cross-correlation (SPCC) combined with principal coordinates (PCO) analysis as a method for sound comparison. We do this using synthetic sounds modeled after the individually-distinctive, harmonically-rich contact calls of wild orange-fronted conures *Aratinga canicularis*. Calls with acoustic properties similar to *Aratinga* contact calls are common in other taxa including non-oscine birds, primates and cetaceans. We generated signals with known variations in time-frequency pattern, duration, noise level, harmonic content and harmonic weighting, and applied SPCC-PCO analysis to obtain an ordering of sounds in n-dimensional space. We find that shared time-frequency patterns dominate the positioning of sounds in PCO space. This was true despite high variability in signal-to-noise ratio (from –60 to +40 dB) and duration (150–275 ms). Furthermore, inclusion of naturally-weighted harmonics (versus fundamentals only) enhances, rather than obscures, the separation of call types. We conclude that SPCC-PCO is an effective method for sorting sounds based on overall time-frequency pattern. In addition, the resulting PCO measures can be used in statistical tests of association with extrinsic variables. The method is thus an effective starting point for examining most bioacoustic hypotheses.

Key words: spectrographic cross-correlation, principal coordinates analysis, sound comparison, parrot vocalizations, sound synthesis

INTRODUCTION

Comparison and classification of animal sounds is a common task in bioacoustic research. Typically researchers perform detailed comparison of sound spectrograms and look for correlations between sound time-frequency structure and a variety of extrinsic contextual variables. Researchers might be interested in identifying a relation-

*E-mail: kac53@cornell.edu
†E-mail: jwb25@cornell.edu

ship between sound structure and the habitat or social context in which sounds are used. Or, they may ask whether calls recorded from the same individual, social group, or geographic population are structurally more similar to each other than they are to calls from different individuals, groups, or populations. Still other researchers might be interested in isolating the structural properties used by animals in their behavioral and/or perceptual classification of received sounds, or in relating signal structures to specific sound production and broadcasting mechanisms.

How can one effectively compare and structurally categorize a set of sounds? Commonly used methods include: 1) comparison of sounds based on a set of measured signal time and frequency parameters; 2) comparison of sound spectrograms via spectrographic cross-correlation analysis; and 3) visual comparison of sound spectrograms by a trained human observer. In the first method, a large set of relatively standard time and frequency measures is collected on the sounds being classified. The presumption here is that by exhaustively measuring the sound data, important structural parameters are unlikely to be missed. The resulting data may be compacted via principal components analysis (PCA). PCA extracts dominant and uncorrelated compound measures from a larger set of potentially correlated measure data. One can then look for statistical associations between the PCA measures and the contextual variables of interest using analysis of variance (ANOVA), multivariate analysis of variance (MANOVA), or linear discriminant analysis (LDA).

While a generally useful method for sound classification, the standard measure approach has its limitations. A finite set of measures, no matter how large, could still miss the structural features of a sound most closely associated with a particular context. Furthermore, the selection of measures is undoubtedly susceptible to researcher bias. Criteria for measure selection often include ease and speed of measurement, significance in prior studies, or likely relevance to specific conjectures. None of these criteria guarantee that a critical measure will be included. To circumvent this limitation, Clark et al. (1987) developed the method of spectrographic cross-correlation (SPCC) for sound comparison. Here, the time-frequency spectrograms of two sounds are cross-correlated and the peak correlation value is taken as a measure of sound similarity. The intent of SPCC analysis is to include all discernible structural features of the two sound spectrograms instead of limiting comparison to a predetermined and possibly incomplete set of measures. SPCC typically is used to generate a matrix of similarity values for each pair of signals in an ensemble. Multidimensional scaling (MDS) or cluster analysis techniques are usually applied to the similarity matrix to identify sound groupings. Associations between sound groupings and contextual variables can then be examined. Alternatively, one can compare the

similarity matrix to a contextual hypothesis matrix of the same dimensions using a Mantel test (Mantel 1967, Schnell et al. 1985).

Finally, visual inspection is a widely used method for classifying sound spectrograms into groups. While the visual criteria observers use in classifying a particular set of sounds are rarely specified, the fact that multiple observers converge on similar classification schemes suggests that major structural similarities (at least in the sense of human perception) are being identified. It remains unlikely however, that a visual observer weighs all spectrographic features equally in making classification decisions. Thus there is an inherent subjectivity in this method not present in SPCC. And, there is no guarantee that those features that the observer weighs most heavily in assigning categories are also the ones most correlated with the contextual variable of interest. The human eye is a superb integrator of information, however. Visual inspection may incorporate a larger fraction of the available signal data than does the standard measure approach.

There is continued debate in the literature about the appropriate use of these and other alternative methods of sound comparison. Nowicki and Nelson (1990) compared three techniques for identifying natural categories among notes in the call of the black-capped chickadee: 1) multidimensional scaling applied to similarity values obtained from spectrographic cross-correlation; 2) visual sorting by human observers; and 3) k-means cluster analysis applied to principal component scores obtained from 14 acoustic measures. They found that all three methods produced generally consistent results. The greatest agreement was between SPCC-MDS and visual classification. Janik (1998) compared four techniques for classifying the fundamental contours extracted from the harmonically-rich signature whistles of bottlenose dolphins: 1) visual sorting by a human observer; 2) a method developed by McGowan (1995) which normalizes contours in duration and applies PCA and k-means cluster analysis to successive frequency measures; 3) hierarchical cluster analysis applied to SPCC similarity values; and 4) hierarchical cluster analysis applied to measures of average difference in absolute frequency. Janik found that visual inspection performed best at producing a classification scheme consistent with the individual identity of isolated dolphins. The alternative methods gave classifications less correlated with this contextual variable. He suggested that the other methods failed, in part, because they gave equal weight to each section of the contour in their similarity assessment. As a result, their classifications either reflected spectrographic details that were unimportant to the overall pattern, or did not reflect brief but visually important pattern differences. Finally, Khanna et al. (1997) examined the performance of spectrographic cross-correlation on a set of simple synthetic sounds. They found that the correlation values produced by this method were

sensitive to the size of the analysis window, that is, the fast Fourier transform (FFT) length, used for spectrogram generation. Depending on the FFT length used, the peak correlation values produced by SPCC systematically strayed from the theoretically expected values when comparing signals of either different duration or different overtone content. Furthermore, they found that the correlation value between two identical signals decreased as the background noise level of one of the signals increased. Ultimately, they concluded that SPCC analysis should be used only for simple sounds with no ambient noise and no significant overtone content. Furthermore, they implied that the results of SPCC analysis could be unreliable when comparing sounds with different duration. Their recommendations have been widely cited and many researchers now avoid the technique completely. Those who continue to use SPCC analysis in spite of this face reviewers who are apprehensive and mistrustful of the method.

Admittedly, the debate is open as to which sound comparison technique is best suited to a particular research question. The Khanna et al. (1997) paper however, has had the effect of invalidating SPCC altogether as a suitable technique for sound comparison. While we do not question their results, we do question the interpretation by Khanna et al. of those results and their proscription against the use of SPCC. A key point, and one that we feel was never made explicitly clear in their paper, is that the interaction between cross-correlation value and FFT length that they observed is simply the expected effect of their particular choice of analysis window size for spectrogram generation. It is not a failing of SPCC analysis *per se*. Their paper reconfirmed the well-known time-frequency trade-off inherent in Fourier series representation of signals (see for example Bracewell 1986, Beecher 1988, Bradbury and Vehrencamp 1998). And it rightly demonstrated, and cautioned, that SPCC analysis is not immune to the effects of this trade-off. Their paper, however, demonstrates no innate flaw with SPCC, and it never really examines the utility of SPCC analysis *per se* as a method for sound classification. Despite the obvious fact that SPCC cannot escape the time-frequency inaccuracies inherent in Fourier analysis, as is true with any method of sound comparison based on spectrograms (including visual classification), we contend that it can still be used effectively for the objective comparison of sounds.

In this paper, we explore the utility of spectrographic cross-correlation for sound classification as part of a novel technique based on SPCC combined with the method of principal coordinates analysis (PCO). Beginning with an SPCC similarity matrix, PCO analysis lets one obtain latent composite measures specifying the position of sound objects in n-dimensional space. We explicitly examine the effects of varying duration, background noise level, and harmonic content on the signal classification schemes resulting from the SPCC-PCO method.

We do this using a set of synthetic test signals modeled after the naturally occurring contact calls of wild orange-fronted conures *Aratinga canicularis*; calls which are harmonically rich and individually distinctive in their time-frequency patterns. Though our sounds are modeled after parrot calls specifically, they provide a good model for the harmonically-rich, individually variable signals seen so widely in primates, cetaceans and non-oscine birds (Sebeok 1977, Kroodsma and Miller 1996, Snowdon and Hausberger 1997). Such signals are now the focus of extensive research.

Though rarely seen in the bioacoustic literature, PCO is a data reduction and embedding (ordination) technique used widely in ecology (Gower 1987, Legendre and Legendre 1998). PCO has many parallels with the better-known method of principal components analysis. Both techniques involve the eigen-decomposition of a square symmetric matrix to obtain a new set of orthogonal measures for an ensemble of sound objects. In PCO, as in PCA, the first few measure axes provide the best approximation to the original matrix. The difference is that PCA works on the variance-covariance matrix for a set of raw sound measures, while PCO works on the overall similarity matrix of the sound objects themselves. Thus, PCO decomposes an object-by-object similarity matrix (such as that generated by SPCC) into a set of latent orthogonal object measures. SPCC combined with PCO analysis provides not only visual groupings of sounds (like cluster analysis), but also a set of independent measures against which the association of extrinsic contextual variables can be measured (unlike cluster analysis). The combination of SPCC and PCO may thus provide a solution to Janik's (1998) concerns about SPCC-cluster analysis methods.

We explore the efficacy of the SPCC-PCO method on a variety of test sounds designed to vary systematically from one another with respect to time-frequency pattern, duration, ambient noise level, or harmonic content. The results demonstrate that the combination of SPCC and PCO is surprisingly robust in its ability to separate the effects of shared time-frequency pattern from those of noise, duration, and harmonic variation. Where appropriate, we use multivariate statistics to show the significance of the classifications provided by this method.

METHODS AND MATHEMATICAL BACKGROUND

Synthesis methods and test signal sets

For our SPCC-PCO analyses, we constructed synthetic facsimiles of the harmonically rich contact calls of wild orange-fronted conures *Aratinga canicularis* using sound synthesis tools available in the

program Signal v.3.0 (Engineering Design). The program was run on a 90 MHz Pentium computer (Gateway 2000, P5-90). Wild-caught birds, held captive in Costa Rica during June-August of 1997 and 1998, were recorded using directional microphones (Sennheiser, model MKH 816 P84) and Hi-8 mm video camcorders (Canon, models ES2000 and ES2500). The recorded contact calls were acquired at a sampling rate of 40 kHz using the real-time spectrogram program RTS v.2.0 (Engineering Design), after filtering the signals from 90 Hz–11 kHz through a bandpass filter (Krohn-Hite, model 3550). The spectrograms for the original contact calls and synthetic copies used in this study were generated in Signal using a 256-point fast Fourier transform length, with a Hanning window function and a fixed step size between windows of 0.64 ms (corresponding to a 90 % window overlap). The *Aratinga* contact call is a single continuous note of roughly 200 ms duration whose signal energy lies primarily in the frequency range from 500 Hz to 8 kHz. The call has a complex frequency structure with at least 10 apparent overtones in its dominant frequency range. Since we have yet to find simultaneous components in *Aratinga* contact calls that are not integer multiples of each other, we shall refer to them below as harmonics. The contact call can be divided into three contiguous subunits based on its time-frequency spectrogram: 1) an initial region in which the fundamental ascends from 500 Hz to 1.5 kHz; 2) a middle region in which the fundamental undergoes rapid, stepwise frequency modulations between 3 and 6 kHz; and 3) a final region in which the fundamental descends from 1 kHz to 500 Hz.

During the captive study period, each bird produced its own unique signature contact call vocalization. For each harmonic signal that we synthesized in the current study, the spectrogram of an actual *Aratinga* contact call vocalization was used as a template. We used a combination of the Signal "peak ft" and "draw" utilities to trace out the fundamental frequency contour, or first harmonic $f_1(t)$, to be used in constructing the synthetic signal. We constructed the amplitude weighting functions for each of the desired harmonic components of the synthetic signal from the spectrogram of the actual animal vocalization using the "amp ft" utility. The synthetic contact-call-like signal $x(t)$ was then assembled from n–m+1 time components $x_k(t)$, each generated from the time-frequency contour $f_k(t) = kf_1(t)$ of a desired harmonic component k and its corresponding amplitude weighting function $a_k(t)$ (where $k = m, \ldots, n$, m = the first harmonic to be used, and n = the last):

$$x(t) = \sum_{k=m}^{n} x_k(t)$$

where $x_k(t) = a_k(t)\sin(2\pi t f_k(t))$.

For all of our synthetic signals, the first harmonic $m = 1$ and the last harmonic $n = 15$. Because extremely high buffer sampling rates had to be used to accommodate some of the frequencies in the highest harmonic components, once a synthetic signal was assembled, it was digitally lowpass filtered in Signal using a 16 kHz cutoff and the sampling rate was decimated to 40 kHz. K.A.C wrote numerous custom programs in the Signal command language to automate assembly and manipulation of the synthetic signals and to automate spectrographic analysis and cross-correlation.

In order to characterize the performance of our analysis method, we constructed a variety of sample sound sets with signals based on naturally occurring *Aratinga* contact calls. The signals in each set varied systematically with respect to a particular signal parameter and thus had quantifiable similarities and differences. Five sets of model signals were generated. The first sample set contained signals that either possessed the same time-frequency pattern (i.e., fundamental contour, number of harmonics, and harmonic weighting functions) and varied in duration, or possessed the same duration and varied in time-frequency pattern. The second set contained signals that either possessed the same time-frequency pattern and varied in signal-to-noise ratio, or possessed the same signal-to-noise ratio and varied in time-frequency pattern. The third set contained signals that either possessed the same time-frequency pattern and varied in duration and/or signal-to-noise ratio, or possessed the same duration and/or signal-to-noise ratio and varied in time-frequency pattern. The fourth sample set consisted of three subsets whose signals possessed the same time-frequency patterns, durations, and signal-to-noise ratios, but varied in harmonic content and harmonic weighting between subsets. The first subset contained signals composed of the fundamental contour (first harmonic) only, with flat amplitude weighting. The second subset contained signals composed of harmonics 1–15 with flat amplitude weightings applied to all harmonics. The third subset contained signals composed of harmonics 1–15 with natural amplitude weightings applied. The final sample set consisted of signals that varied ("morphed") uniformly in overall time-frequency structure from one signal type to a second different signal type. Specific parameters used in generating each sample sound set are detailed in the Results section when that set is discussed.

Spectrographic cross-correlation

We used Signal to calculate peak spectrographic cross-correlation coefficients and generate a matrix of similarity values between all signals in a sample set. Signal uses the formula

$$R(T) = \frac{\int\int_{\Delta T\,\Delta F} \left(\frac{x(f,t)-\bar{x}}{\sigma_x}\right)\left(\frac{y(f,t+T)-\bar{y}}{\sigma_y}\right) df\,dt}{\Delta T \Delta F}$$

to calculate the cross-correlation function R(T). In all cases, we used matrix-based normalization where $\bar{x}, \bar{y}, \sigma_x$, and σ_y are the means and standard deviations for the entire spectrogram matrices. The maximum time offset for correlation calculation was set to 100% of the signal durations; the frequency range for correlation calculation was 200 Hz–10 kHz. A custom command program was written to generate the matrix of peak cross-correlation values for any arbitrary-sized set of sound signals. Since the matrix of peak cross-correlation values is square symmetric, we reduced computation time by calculating only the lower-triangular matrix plus the diagonal. For a data set of n sounds, this amounts to the calculation $n(n+1)/2$ rather than n^2 values. The custom program checks that all spectrograms being correlated have the same sampling rate, duration, and time-frequency resolution. To obtain the same duration for all spectrograms being cross-correlated, the signals within a data set were compared and zero-padded to the length of the longest signal before their spectrograms were generated.

As stated before, all spectrograms used in our study were generated using a 256-point fast Fourier transform length. For our sounds (sampled at 40 kHz), this was the FFT length that provided a suitable balance between time and frequency domain representation as judged through visual inspection. That is, for all our sounds we attained good separation of signal harmonic components while still seeing signal frequency modulation patterns represented in the time-domain. For our work, a 256-point FFT length brought out the level of spectrogram structure most suitable for call comparison. In fact, the spectrographic structure of our calls was similar for a range of FFT lengths around 256. Harmonic contours became fatter or thinner with shorter or longer FFT lengths, but we still attained the same basic spectrogram patterns—patterns that were a good balance between time and frequency domain representation of the sounds.

Principal coordinates analysis

The technique of principal coordinates analysis, sometimes referred to as classical multidimensional scaling, allows one to calculate the geometric coordinates of data objects in an n-dimensional space given a matrix of similarities or distances between those objects (Gower

1966, 1987, Neff and Marcus 1980, Everitt and Dunn 1991, Wong and Bergeron 1997, Legendre and Legendre 1998). The analysis calculates the coordinates of the objects in such a way that the first principal coordinate axis provides the best first-order approximation to the original inter-object distances. Addition of the second axis provides the best second-order approximation, and so on. Thus, in addition to providing a classification scheme based on the geometric arrangement of objects in PCO space, this method provides a set of independent measures for each sound object, the PCO values. These PCO values can be used like any other structural measure to test hypotheses about sounds and extrinsic variables. For the interested reader, the Appendix at the end of this paper provides a discussion of the mathematical rationale behind the PCO method.

The new PCO axes are computed by extracting the eigenvalues and eigenvectors of a matrix whose elements are equal to a transformation of the original inter-object distance values. If the distances measured between the objects are truly Euclidean, all of the eigenvalues will be positive. If they are not Euclidean however, some of the eigenvalues will be negative resulting in imaginary coordinates. In practice, this is not a problem provided that the negative eigenvalues are small in magnitude (Neff and Marcus 1980, Gower 1987, Everitt and Dunn 1991). If there are numerous large negative eigenvalues however, the use of principal coordinates analysis is inappropriate and the method of metric or non-metric multidimensional scaling (MDS) should be used instead (Kruskal and Wish 1978, Everitt and Dunn 1991). When appropriate, the use of principal coordinates analysis for embedding objects in geometric space is preferred over multidimensional scaling. Since object coordinates are calculated by direct matrix algebra in PCO, rather than an iterative process as in MDS, most statistical packages have no matrix size limit for PCO whereas MDS is usually limited to small matrices. In our experience, the two methods give similar ordinations when applied to the same data set.

The number of dimensions d in PCO space that are needed to provide an accurate representation of the original inter-object distances can be determined using a number of different relationships provided in the literature (for example, see Gower 1987, Everitt and Dunn 1991, Wong and Bergeron 1997). We use the relationship

$$\frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{p} \lambda_i}$$

to quantify the degree of accuracy of the representation in PCO space,

where λ_1, ..., λ_d are the first d positive eigenvalues used in the approximation and λ_1, ..., λ_p are all the positive eigenvalues (Wong and Bergeron 1997).

Thus, starting from a similarity or distance matrix only, one can generate coordinates in n-space for a given set of sound objects. Furthermore, these coordinates (the PCO values) provide a set of intrinsic measure variables for the objects that would otherwise not have been available. This is a major advantage of PCO over cluster analysis. By viewing the sound objects plotted in the first two or three principal coordinate axes, provided these dimensions account for enough of the overall inter-object distance, we can intuitively assess overall object relatedness from geometric proximities. We used the R-Package routine "PCoord" (developed for the Macintosh by Philip Casgrain at the University of Montreal, http://alize.ere.umontreal.ca/~casgrain/R) to calculate the principal coordinate values for each set of sound data from the matrix of peak SPCC values.

Statistical tests of association

Given that they have approximately normal distributions and similar variances and covariances, we can use any subset of the PCO measures generated for a set of sounds in multivariate statistical tests of association. Associations of PCO values with categorical extrinsic variables can be examined with multivariate analysis of variance (MANOVA) or linear discriminant analysis (LDA). MANOVA provides standard variance partitioning tables, whereas LDA allows one to use bootstrap methods such as cross-validation to test for robustness of the classification. MANOVA and LDA are robust to skew in the data, but are sensitive to outliers and kurtosis (Tabachnick and Fidell 1996). We have found that the PCO distributions from field samples of conure calls rarely have outliers but do tend to be platykurtic, that is, to have flattened distributions compared to the normal distribution. As long as the sample sizes per factor level are similar, platykurtosis makes it harder to obtain a significant result using MANOVA or LDA (Sharma 1996, Tabachnick and Fidell 1996). These tests are thus conservative. Both methods compute linear combinations of the continuous variables, called canonical variates, which maximally segregate calls according to the alternative factor levels. The canonical variates are essentially linear transformations of the PCO coordinates, and thus constitute a rotation of the PCO axes. From the new rotational perspective, calls in the same factor level appear maximally clustered with each other. The relative positions of the calls in PCO space are unchanged by the rotation. When the assumptions underlying canonical variate analysis cannot be accepted, logistic regression can

be used instead. Where extrinsic variables were continuous, we also used analysis of covariance (ANCOVA) or multiple linear regression as appropriate. Statistics were run on either the JMP (version 3.2, SAS Institute) or SPSS (version 8.0, SPSS Inc.) statistical packages.

RESULTS

Examination of FFT length issues

Before presenting our results using the SPCC-PCO method for sound classification, we feel it is important to explicitly characterize the effects of FFT length on spectrogram structure. We do this using the synthetic test sounds of Khanna et al. (1997). These sounds were replicated using Signal as described in their paper. Our intention is not to rebut or invalidate the work of Khanna et al. We wish merely to show that SPCC values are simply, and above all, a reflection of the structural information available in the spectrograms used. And we wish to fully elucidate why we continue to accept SPCC as a valid method of sound comparison.

First, Khanna et al. (1997) examined the SPCC values generated when pure-tone signals of the same frequency but different duration were correlated. In this test, and all others we discuss in this section, sounds had a 20 kHz sampling rate, and the frequency range used for SPCC analysis was 1–8 kHz. As expected, Khanna et al. found that the ratio of signal durations, measured as note length ratio (NLR), and not absolute difference in duration, determined sound similarity. They also found, however, that longer FFT frame lengths resulted in slightly, but systematically higher, SPCC values than theoretically expected. They reported this as a sensitivity of SPCC to choice of analysis window (i.e., FFT) length when signal durations varied. We replicated their analyses and examined each spectrogram before generating peak SPCC values. Using a threshold of –60 dB from the peak amplitude value to estimate the beginning and end of sound spectrograms, we found that FFT frame size directly alters effective note length ratio. The consequent SPCC values faithfully reflect these changes in spectrogram NLR but were otherwise independent of frame size.

This change in effective note length ratio is directly predictable from Fourier transform theory. Longer FFT length translates directly into a longer Δt, that is, a wider time bin, for spectrogram generation. Thus, the longer the FFT length the more the signal energy is smeared out in time. Precise resolution of signal temporal features will be impossible. For example, Khanna et al. compared two 4-kHz notes, one with a duration of 1000 ms and the other with a duration of 100 ms. In their study, and our replicate, spectrograms of the two notes were

generated using an 8192-point FFT and a 32-point FFT. At an FFT size of 8192 points, the spectrogram of the 1000 ms tone shows a duration of 1309 ms. That of the 100 ms tone shows a duration of 431 ms. At an FFT size of 32 points, the spectrogram of the 1000 ms tone shows a duration of 986 ms. That of the 100 ms tone shows a duration of 92 ms. Thus, longer FFT lengths for spectrogram generation make these two signals appear more similar in duration than they really are. For this pair of sounds, spectrographic NLR is equal to 0.09 for the 32-point FFT and 0.33 for the 8192-point FFT, compared to the theoretical NLR of 0.1 based on actual time waveform durations. The computed SPCC values reflect this change in NLR and thus increase. Similar results were found for all other comparisons in the duration tests. Figure 1 shows the effective NLR ratios (1b) for note combinations considered by Khanna et al. and the corresponding SPCC values (1a) taken from Table 1 of their paper. The average correlation coefficient between effective NLR and corresponding SPCC value across all 10 contrasts in their Table 1 was r = 0.96 (range 0.89–0.99). We conclude that the differences in SPCC value with FFT frame length are entirely a by-product of changes in the spectrogram NLR values; the SPCC process itself adds no additional errors.

Khanna et al. (1997) also reported that when the spectrograms of two equal-duration signals, each possessing a different-valued set of component frequencies, were cross-correlated, shorter FFT frame sizes resulted in systematically higher peak correlation values. Furthermore, the expected correlation values were achieved at frame sizes twice that necessary, in theory, to resolve the harmonic components. The first effect, again, is predictable from Fourier transform theory. Shorter FFT length translates directly into a larger Δf, that is, a wider frequency band for spectrogram generation. Thus, the shorter the FFT length the more the signal energy will be smeared out in frequency. Precise resolution of signal spectral features will be impossible. Figure 2 shows the spectrograms of two of the signals used in Khanna et al.'s test of cross-correlation frequency sensitivity. As in their paper, spectrograms of the two notes, one containing frequencies of 2200, 2400, and 2600 Hz and the other containing frequencies of 2000, 2500, and 3000 Hz, were generated using a 256-point FFT (Figure 2a,b), and a 2048-point FFT (Figure 2c,d). A frame size of 256 points should theoretically provide just enough frequency resolution (Δf = 78 Hz at a 20 kHz sampling rate) to separate the different component frequencies in the two signal types. However, it is apparent from Figure 2a,b that there is considerable spectral smearing and overlap of the frequency content of the two spectrograms at this FFT length. Although not plotted here, spectrograms show much better separation of frequency content at the next higher frame size of 512 points. The 2048-point FFT length used for the spectrograms in Figure 2c,d easily resolves the frequency components and the SPCC values Khanna et al.

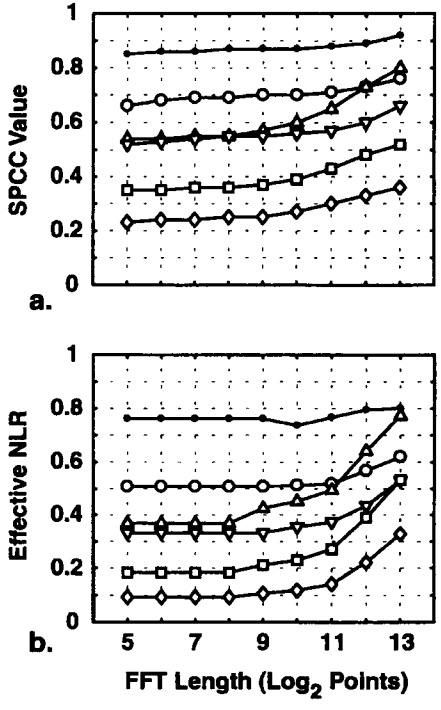

Figure 1. SPCC correlation values from Khanna et al. (1997) (1a) and effective note length ratios measured from sound spectrograms (1b) plotted against analysis FFT length for various actual note length ratios. Actual NLR equals: 0.75 (dots), 0.5 (circles), 0.4 (up triangles), 0.33 (inverted triangles), 0.2 (squares), and 0.1 (diamonds).

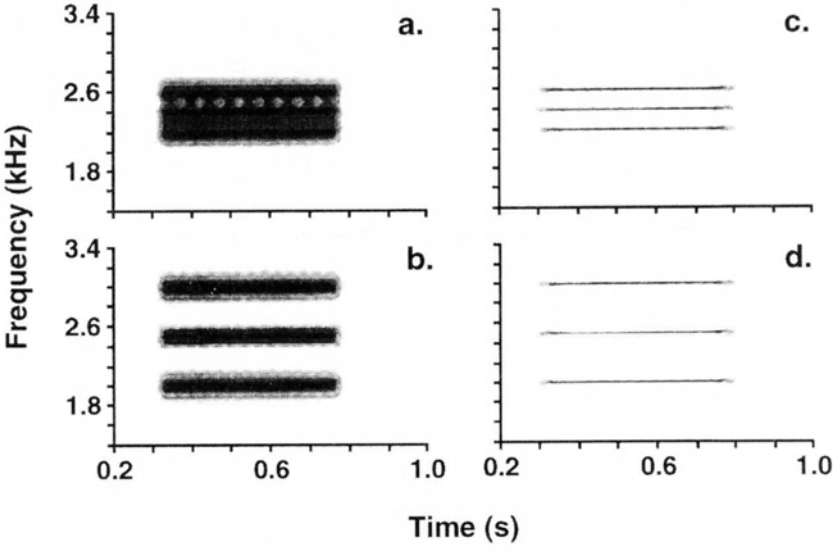

Figure 2. Spectrograms of two synthetic test sounds with different overtone content generated at different FFT analysis lengths (per the methods of Khanna et al. 1997). Signal component frequencies are 2.2, 2.4, and 2.6 kHz (2a,c) and 2.0, 2.5, and 3.0 kHz (2d,b). Spectrogram FFT lengths are 256 points (2a,b) and 2048 points (2c,d).

obtained for these spectrograms are exactly as expected. Given the increased spectral smearing that occurs at shorter FFT lengths, it makes sense that shorter FFT lengths result in systematically higher peak correlation values. Thus, their only remaining puzzle with overtone-rich sounds was the need to use frame sizes at least twice the theoretical size to obtain expected SPCC values. Again, this deviation is not due to SPCC. Instead it is due to the finite duration windowing inherent in discrete time Fourier transformation. Whether rectangular or cosine-based (i.e., Hanning or Hamming) such windowing is known to shift effective analysis bandwidths to higher values than the theoretical value of $\Delta f = 1/\Delta t$. The loss of effective frequency resolution from windowing is due to the smearing that occurs when the spectrum of the signal is convolved with the spectrum of the analysis window. The interested reader is referred to Oppenheim and Schafer (1989) section 7.4. Consequently, one must use FFT lengths longer than the theoretical minimum before the desired frequency resolution is obtained in spectrograms. Again, the effects reported by Khanna et al. do not arise from SPCC analysis in itself, but from a failure to use appropriate analysis bandwidths so that spectrograms resolve the temporal or spectral features of interest.

Khanna et al.'s (1997) tests of SPCC analysis on frequency modulated (FM) signals showed no apparent anomalies. As discussed by Beecher (1988), optimal representation of FM in spectrograms occurs at intermediate bandwidths where the effective Δf is small enough to identify the modulated frequencies accurately, but not so small that the time pattern is lost and the modulation is broken fully into its frequency domain components. Their FM test of SPCC basically confirmed Beecher's conclusions. Making allowances for windowing effects, the peak SPCC values obtained by Khanna et al. are as one would expect.

Thus, we were unable to find any anomalous effects of varying FFT length on SPCC analysis. The effects described by Khanna et al. (1997) as SPCC sensitivity to analysis bandwidth are in fact solely consequences of spectrogram resolution, that is, the time-frequency inaccuracy introduced by the discrete-time Fourier analysis of signals. If the spectrograms show no resolution of a signal temporal or spectral feature, SPCC analysis does neither better nor worse.

Synthetic parrot sounds

I. Duration series

While SPCC analysis itself does not introduce artifacts when comparing signals of varying duration, it is unclear whether differences in signal duration might obscure overall similarities or differences in spectrographic time-frequency pattern. The sensitivity of

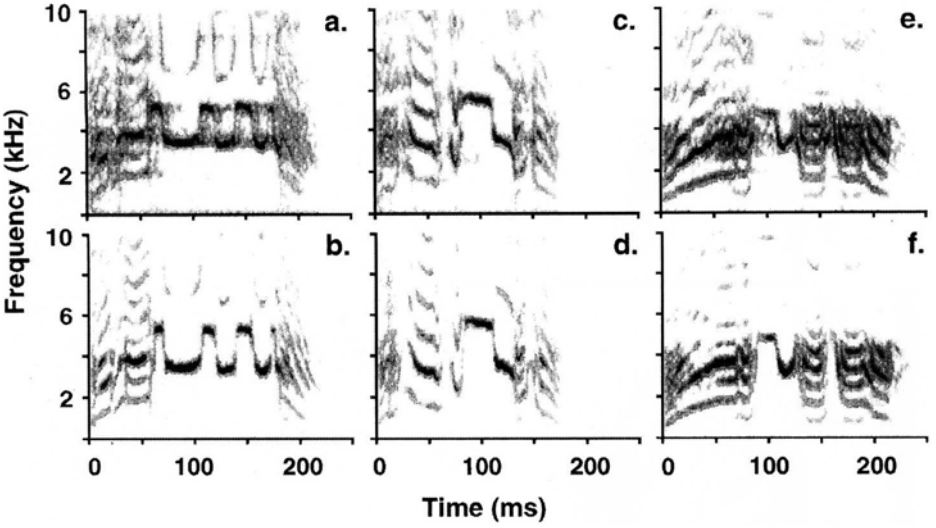

Figure 3. Spectrograms of actual Aratinga contact calls (3a,c,e) and their synthetic replicas (3b,d,f). The individuals and call types are Thalia (3a,b), Randy (3c,d), and Thoreau (3e,f).

comparison techniques to the time expansion and compression of signals is of concern to many researchers, especially those working with cetacean sounds (Reiss and McCowan 1993, McCowan 1995). Methods have been suggested that employ either linear or nonlinear time normalization (time-warping) before signal comparison (Buck and Tyack 1993, McCowan 1995). As a first step in exploring the efficacy of SPCC-PCO, we wished to determine its sensitivity to the expansion and contraction of signal duration. Specifically, we wished to determine if the method sorted signals according to shared time-frequency pattern or shared duration. We examined this issue using the contact calls of three *Aratinga* individuals. The contact call spectrograms of these three individuals, Randy, Thalia, and Thoreau (Figure 3a,c,e) are clearly distinct from one another and possess individually unique time-frequency signatures. The synthesized replicas of these contact calls are shown in Figure 3b,d,f. A custom command program generated the appropriate harmonics from the fundamental contour, extracted the harmonic amplitude weighting functions from a spectrogram of the original call, expanded or contracted the durations of the harmonics and amplitude weighting functions, and then weighted and assembled all components of the synthetic sound. In this way, we constructed a set of 33 sound signals that were either identical in their global time-frequency pattern but different in duration, or different in their global time-frequency pattern but identical in duration. Signal durations ranged from 150 ms to 275 ms in increments of 12.5 ms. We also included the three signals of unaltered duration equalling 178 ms for the Randy call type,

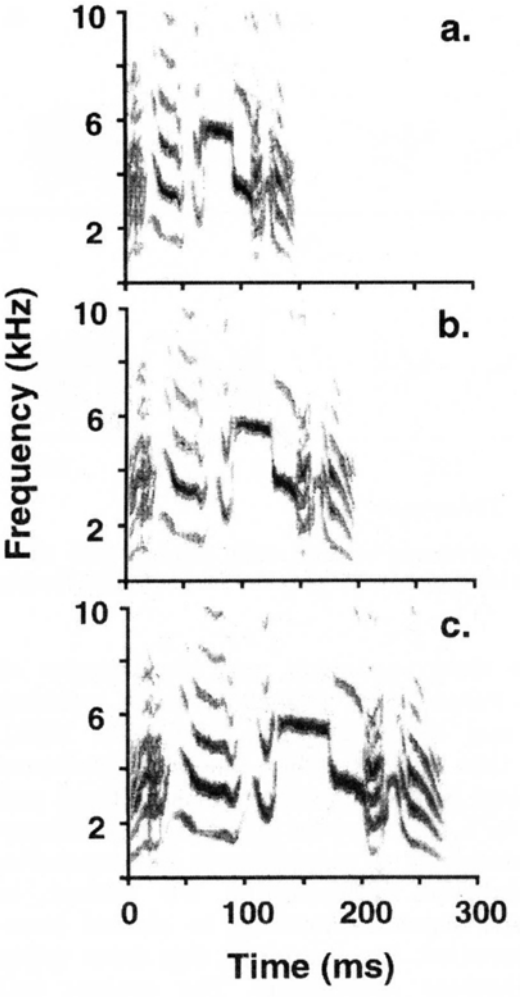

Figure 4. Examples from the duration series for the Randy call type. Durations equal 150 ms (4a), 200 ms (4b), and 275 ms (4c).

220 ms for the Thalia call type, and 243 ms for the Thoreau call type. This made a total of 36 sound signals in the duration series. Figure 4 shows examples from the time duration series for the Randy call type.

Spectrograms were generated for all 36 sounds using the analysis window size (256 points) that we felt provided the best time and frequency domain display of signal structure. The sound spectrograms were cross-correlated and principal coordinates analysis was applied to the resulting matrix of peak correlation values. Figure 5a shows a three-dimensional scatter of the sound objects plotted on the first three PCO axes which together account for 68% of the original inter-object distances. The figure also shows two-dimensional scatter plots of PCO1 and PCO3 versus PCO2 (Figure 5b,c), and scatter plots of the first three PCO measures versus signal duration in milliseconds

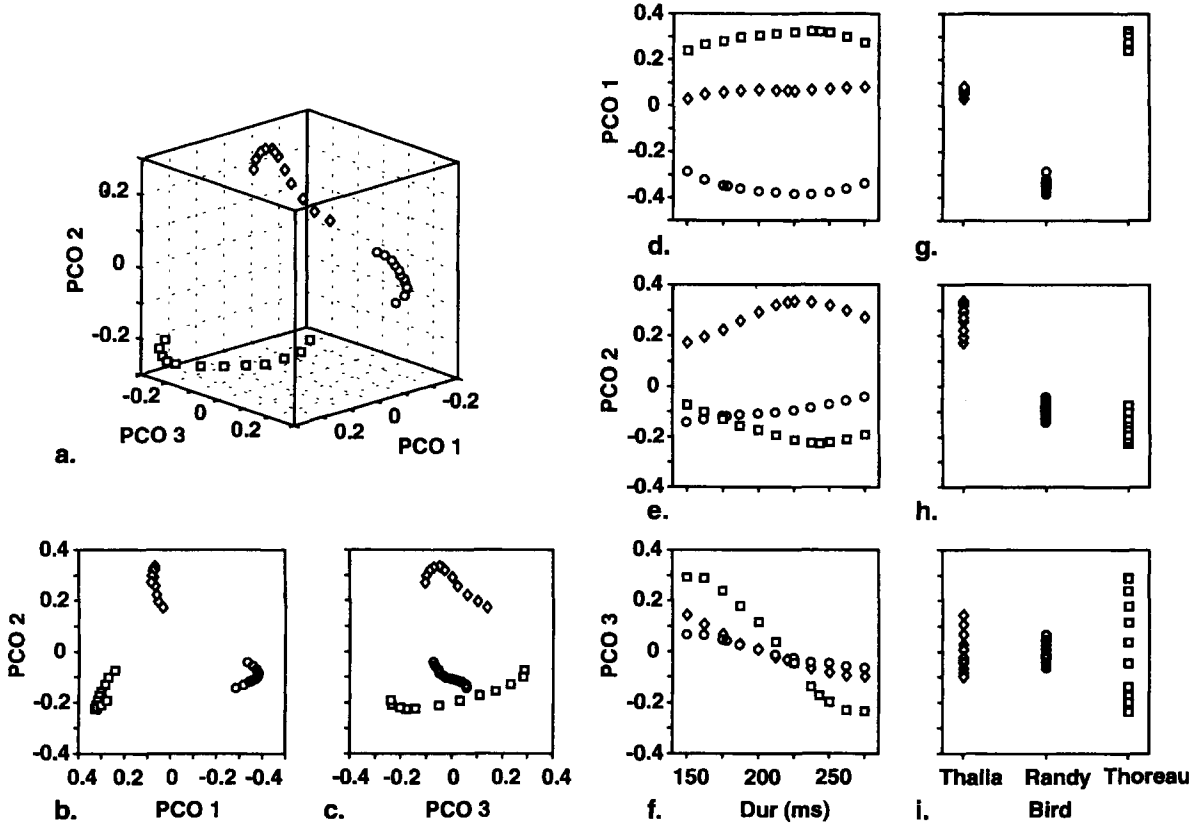

Figure 5. Two and three-dimensional PCO plots for the duration series, and plots of PCO1-3 versus duration in ms and call type. The different calls types are Thalia (diamonds), Randy (circles), and Thoreau (squares).

(Figure 5d-f) and call type (Figure 5g-i). Although there is systematic spread of the sound objects associated with duration change, the three call types clearly cluster into their own distinct regions of three-dimensional PCO space. Furthermore, the nearest neighbor to any sound object in PCO space is that sound of the same call type with the most similar duration. Thus, the geometric embedding of sounds reflects the sequential ordering of duration within each call type. Despite a 125 ms spread in duration, the three call types can still be sorted from one another. Statistically, PCO1 provided perfect separation of signals according to call type, but was unrelated to signal duration (ANCOVA: call type: $F_{2,20} = 1704$, $P < 0.0001$; relative duration: $F_{10,20} = 0.11$, NS). PCO3, in contrast, was significantly correlated with duration (entered as percent of the unaltered signal duration), but not with call type (ANCOVA: call type: $F_{2,20} = 0.39$, NS; relative duration: $F_{10,20} = 7.46$, $P < 0.0001$). These two coordinates separated overall time-frequency pattern from signal duration and provided an independent measure of each.

II. Noise series

A second aspect we wished to examine was how the analysis method performed in the presence of noise. It has been shown that peak SPCC values decrease systematically as signal-to-noise ratio (SNR) decreases when a clean reference signal is correlated with a noisy version of the same signal (Khanna et al. 1997). We wished to see whether SPCC combined with PCO might permit the extraction of basic time-frequency patterns from common background noise levels. We used the three distinct call types of the prior section and added different amounts of uniform random noise generated using the Signal "ran" function. A custom command program was written to normalize the root-mean-square (RMS) amplitude of the sound signal (working directly on the time waveform), and generate buffers of uniform random noise. These noise buffers were normalized and appropriately scaled so that they resulted in specified signal-to-noise ratios when added to the sound signal. A new random noise buffer was generated for each new noisy signal we constructed so that the noise between signals remained truly random. In this way, we constructed a set of 66 sound signals that were either identical in their global time-frequency pattern but different in signal-to-noise ratio, or different in their global time-frequency pattern but identical in signal-to-noise ratio. The sounds had signal-to-noise ratios (measured using V_{rms} values) ranging from –60 to +40 dB increasing in increments of 5 dB. This set also contained the three unaltered sounds which had infinite signal-to-noise ratios. Figure 6 shows examples from the series of noisy signals for the Thalia call type.

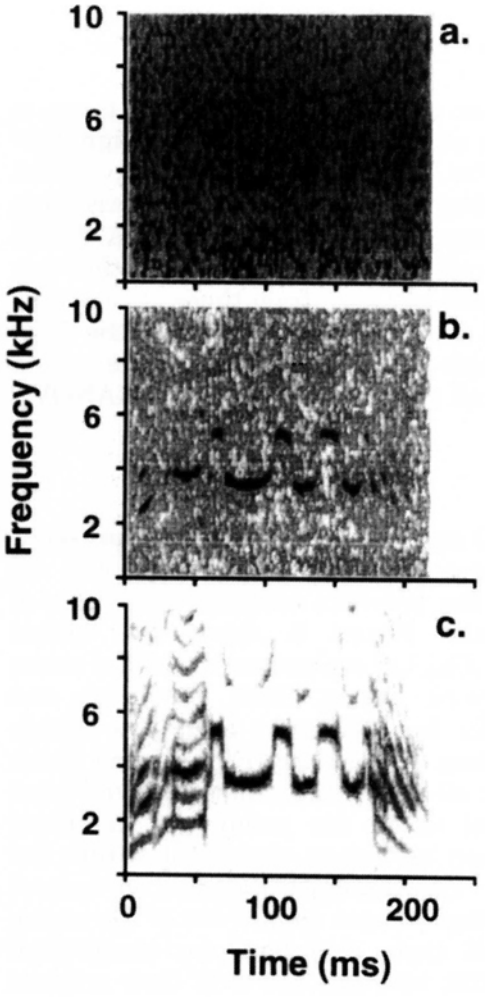

Figure 6. Examples from the noise series for the Thalia call type. Signal-to-noise ratios equal –20 dB (6a), 0 dB (6b), and +20 dB (6c).

All 66 sound spectrograms were cross-correlated and the ordering of the ensemble was examined in PCO space. Figure 7 shows a three-dimensional scatter of the sounds plotted on the first three principal coordinate axes (7a) which together account for 52% of the original inter-object distances. It also shows two-dimensional scatter plots of PCO1 and PCO2 versus PCO3 (7b,c), and scatter plots of the first three PCO measures versus signal-to-noise ratio in dB (7d-f) and call type (7g-i). Again, although there is systematic spread associated with change in signal-to-noise ratio, the three call types clearly occupy distinct regions of PCO space. For this ensemble, PCO1 decreases as signal to noise ratio increases. The pattern of these points in 3-dimensional PCO space is interesting. At the lowest signal-to-noise ratios (< –20 dB), the points for different call types lie close together. However, even at the lowest signal-to-noise ratio, there is no sound

whose nearest neighbor is the sound of another call type. As signal-to-noise ratio increases, points for each call type follow separate trajectories radiating away from the convergent area and from the trajectories of other call types. For all 3 PCO coordinates, there is perfect separation of sound objects according to call type for signal-to-noise ratios of –5 dB or higher. For PCO1, there is, as well, perfect separation of sound objects according to call type for SNR less than –20 dB. Over the full 100-dB range of signal-to-noise ratios, PCO2 completely separates the Thoreau call type from those of the other two birds, and PCO3 separates the Thalia call type from those of the other two birds. Although this data set is not normal by design, the PCO plots suggest that a random sample of calls from this series would separate completely according to call type were we to apply MANOVA or LDA analysis.

III. Duration and noise series

We next examined how SPCC-PCO analysis performed in the presence of both duration and noise fluctuations. For this sample set, we combined the three call types of the previous sections with those of seven other *Aratinga* individuals. Figure 8 shows the actual (8a,c,e,g,i,k,m) and synthetic (8b,d,f,h,j,l,n) contact calls of these seven additional individuals, Diane, Eeyore, Emma, Hera, Jane, Pat, and Whitman. Again, the contact calls from different birds are clearly distinct from one another and from the contact calls of the three individuals shown in Figure 3. For all ten contact call types, we varied duration and signal-to-noise level using the techniques described above. This time, however, duration variations were kept within the ranges typical of real recordings. The standard deviation in contact call duration from individual *Aratinga* ranges from 5–7% of the mean. For each of the ten contact call types therefore, we constructed synthetic calls that were 90, 95, 100, 105, and 110 % of their original durations. Then, for each of these expanded and contracted signals, uniform random noise was added to generate signal-to-noise ratios of –5, 0, +5, and +10 dB. As we saw in the previous section, this is the range of signal-to-noise ratios in which PCO values show the most spread and thus the most overlap between call types (Figure 7d). We also included the signals with no noise added. The resulting sample set included ten call types each with five signal durations (90 to 110% of the original duration) combined with five signal-to-noise ratios (–5 to +10 dB, and infinity). Thus, the 250 sound signals were either identical in their global time-frequency pattern but different in duration and/or signal-to-noise ratio, or different in their global time-frequency pattern but identical in duration and/or signal-to-noise ratio. We also included the ten actual field recordings after which the synthetic calls were modeled. A total of 260 spectrograms were thus generated and cross-correlated with one another.

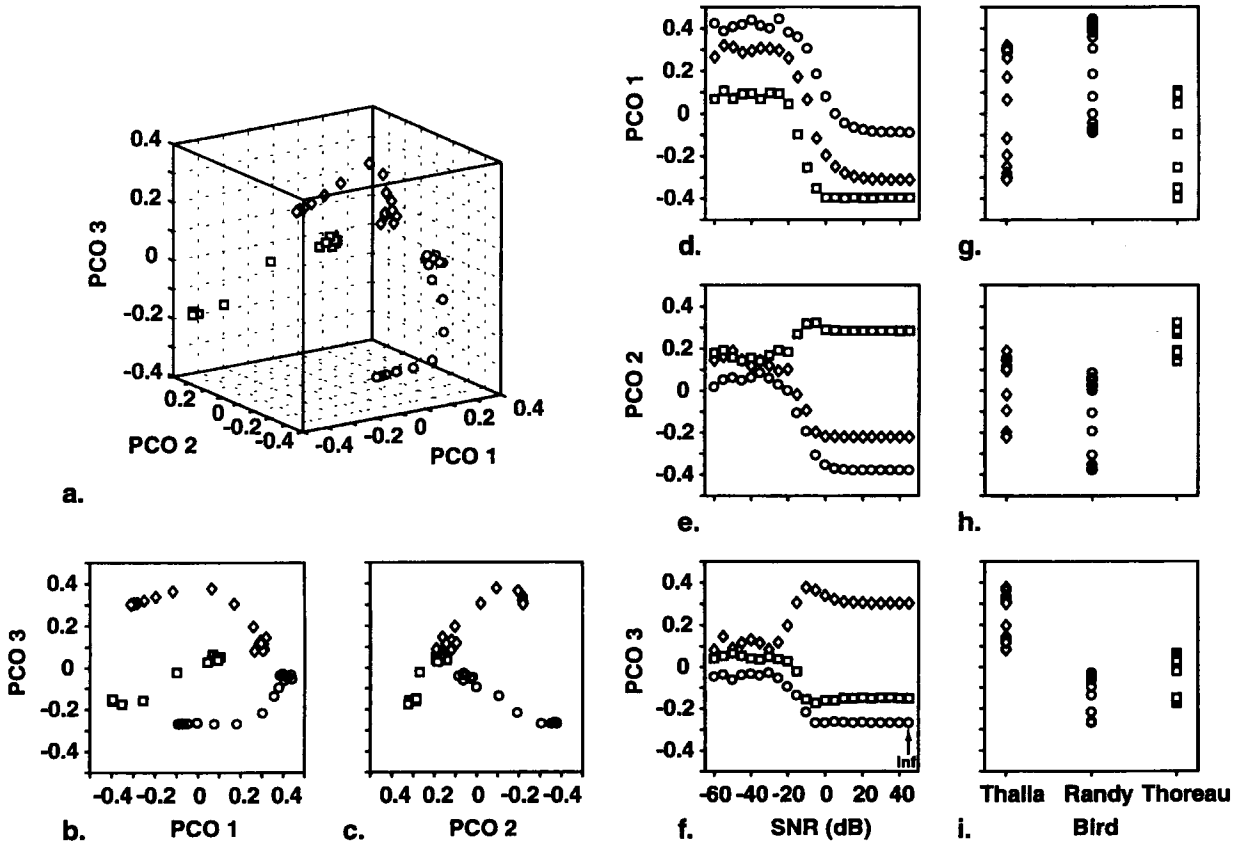

Figure 7. Two and three-dimensional PCO plots for the noise series, and plots of PCO1-3 versus signal-to-noise ratio in dB and call type. The different calls types are Thalia (diamonds), Randy (circles), and Thoreau (squares).

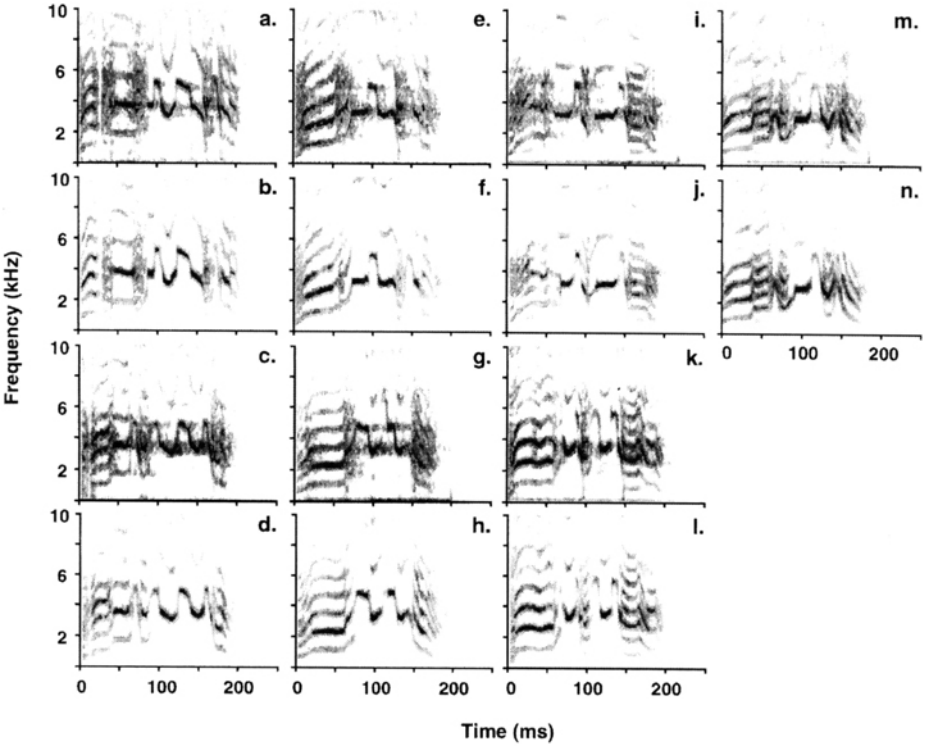

Figure 8. Spectrograms of actual *Aratinga* contact calls (8a,c,e,g,i,k,m) and their synthetic replicas (8b,d,f,h,j,l,n). The individuals and call types are Jane (8a,b), Hera (8c,d), Diane (8e,f), Pat (8g,h), Whitman (8i,j), Eeyore (8k,l), and Emma (8m,n).

Figure 9 shows a three-dimensional scatter of the 260 sounds plotted on the first three principal coordinate axes (9a) which account for 41% of the original inter-object distances, two-dimensional scatter plots of PCO1 and PCO2 versus PCO3 (9b,c), and scatter plots of the first three PCO measures versus call type (9d-f). Again, although there is systematic spread associated with changes in duration and signal-to-noise ratio, the ten call types occupy clearly distinct regions of three-dimensional PCO space. Because of its large size and particular design, the PCO measures for this sample set approached normality, as would the measures for any typical ensemble of natural sounds. We thus performed MANOVA and LDA analyses using the first five PCO measures as dependent variables and call type (based on identity of the source bird) as a factor. The first five PCO measures account for 57% of the original inter-object distances. The MANOVA analysis strongly rejected the null hypothesis of equal means (Wilk's $\Lambda < 0.0001$, $F_{45,1100} = 1732$, $P < 0.0001$). All five canonical variates showed significant separation ($P < 0.001$), and LDA analysis correctly

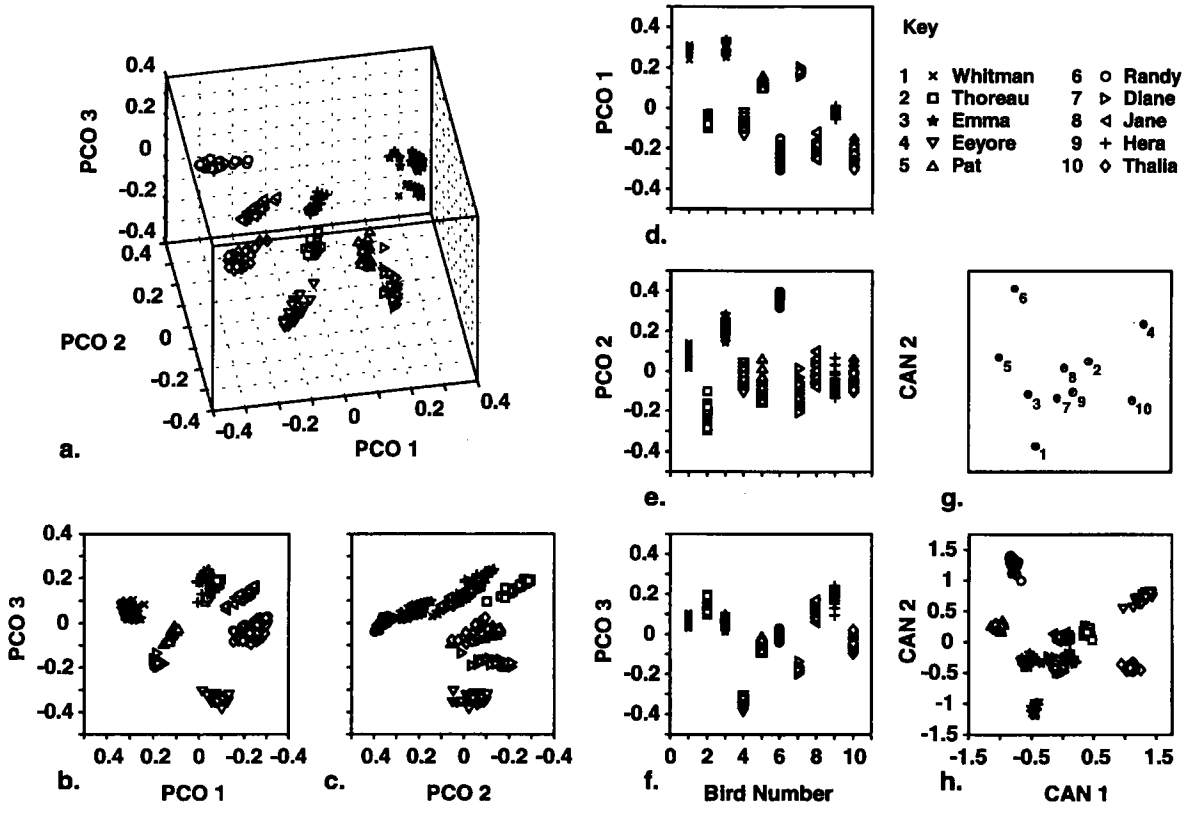

Figure 9. Two and three-dimensional PCO plots for the duration plus noise series, plots of PCO1-3 versus call type, and plots of canonical variate 1 versus 2. The different calls types and their associated number and symbol are shown in the key.

classified 100% of the calls, even with cross-validation bootstrapping. Figure 9h shows a scatter plot of the sound objects on the first two canonical axes, with 95% confidence regions around the call-type centroids shown in Figure 9g. The 95% confidence regions are so small in this data set, that they are comparable to the size of the dots used to indicate the centroids. Thus, even in the face of temporal fluctuations and varying levels of random noise, the canonical separation of each of the ten different call types remains clear.

IV. Harmonics series

We examined next how the presence of harmonics other than the fundamental affects the SPCC-PCO classification of sounds. It is a common practice among cetacean researchers to remove all harmonics except for the fundamental when comparing harmonically rich sounds (Buck and Tyack 1993, McCowan 1995, Janik 1999). In fact, based on their concerns around the analysis of over-tone rich signals, Khanna et al. (1997) suggest that is appropriate to compare only fundamental contours when using SPCC. However, the degree to which harmonic content affects the classification of sounds by SPCC has never been explicitly examined. We used a set of calls modeled after three conures (Emma, Eeyore, and Jane) to examine explicitly the effect of including naturally-weighted harmonics, as opposed to the fundamental contour only, on the ability of SPCC-PCO analysis to sort sounds based on common time-frequency pattern. Rather than using the same three call types as in sections I and II, we decided to randomly pick a new set of calls. This ensured that our results were not contingent on something peculiar to that first set.

Three signal sets with different harmonic content and amplitude weighting were constructed. The first set of signals was made up of the fundamental contour only, with uniform (flat) amplitude weighting along the entire contour. We shall refer to this as the fundamental case. The second set was composed of the fundamental and next 14 harmonics, with uniform amplitude weighting along the entire contour for all harmonics (the uniformly weighted case). The last set was composed of the fundamental and next 14 harmonics, with each contour for each harmonic weighted as in the original call spectrogram (the naturally weighted case). Each of the three signal sets contained synthetic sounds modeled after the three *Aratinga* call types. These signals were varied in duration (90% to 110% of the original) and signal-to-noise level (–5dB to +10 dB and infinity) as in section III. Thus, each of the sample sets contained a total of 75 signals.

Separate SPCC similarity matrices were calculated for the fundamental, uniform weighting, and natural weighting cases. We then applied PCO analysis to each matrix and submitted the resulting principal coordinates to MANOVA analysis to assess how well spectrograms could be separated according to call type. All three

sample sets showed a statistically significant association between PCO coordinates and call type ($P < 0.0001$ in all three cases). Furthermore, the significance of the association increased with the number of successive PCO coordinates included in the analysis. For all sample sets, only the first two PCO measures were necessary to achieve significant separation. The three sample sets did differ, however, in the degree of significance of the group separations accomplished by MANOVA. To quantify this, we plotted Wilk's Λ versus the number of PCO measures included in the MANOVA analysis. Wilk's Λ decreases as the statistical significance between group separation accomplished by MANOVA increases. The plot in Figure 10a shows that signals composed of naturally weighted harmonics provide better group separation than do signals composed of only the uniformly weighted fundamental for up to five PCO coordinates. When more coordinates are included in the analysis, the fundamental and naturally weighted cases give similar resolution. Figure 10b compares the results for signals composed of uniformly versus naturally weighted harmonics. It is clear that naturally weighted harmonics provide much stronger associations between PCO coordinates and call type, and thus better classification based on call type, than do uniformly weighted harmonics. The fundamental case is not plotted in Figure 10b because on this scale it would completely overlap the naturally weighted case. Thus, the signal set composed of uniformly weighted harmonics does much worse at separating calls by time-frequency structure than do

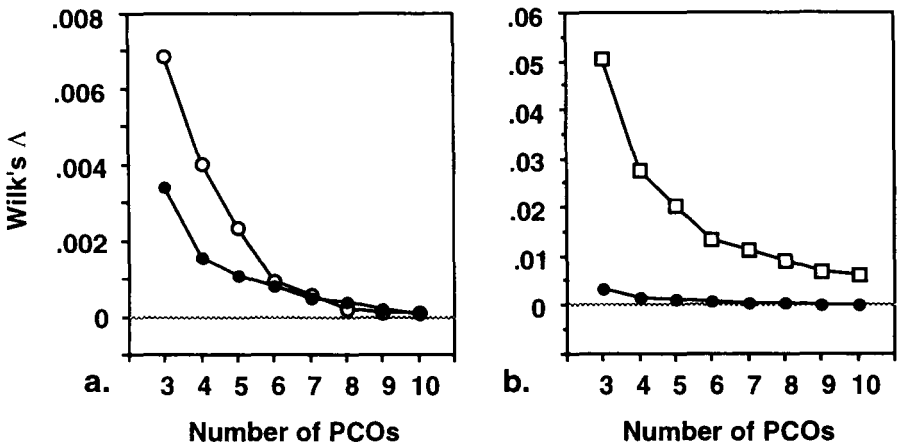

Figure 10. Plots of Wilk's Λ versus the number of principal coordinates included in the MANOVA analysis for three sets of sounds containing different numbers of harmonics and harmonic weightings. Set one was composed of sounds containing the fundamental contour only with uniform weighting (open circles); set two sounds containing the fundamental contour plus the next 14 harmonics with uniform weightings (open squares); set three sounds containing the fundamental contour plus the next 14 harmonics with natural weightings (closed circles).

the signal sets composed of either the fundamental alone or naturally weighted harmonics. And, the signal set composed of naturally weighted harmonics does better than the signal set composed of the fundamental contour alone when classifying signals using only the first few PCO coordinates.

V. Morphing series

Finally, we wished to examine how the method performed with a set of signals that represented a continuous transition from one signal type to a second completely different signal type. Transitions in call types have been noted, for example, during new group formation in captive budgerigars (Farabaugh et al. 1994, Farabaugh and Dooling 1996) and as part of the natural variation in the repertoires of some primates and cetaceans (Snowdon and Hausberger 1997). We knew from analysis of the previous sample sets that the SPCC-PCO method did well at grouping similar signal types for sets containing a few discrete time-frequency patterns. It was unknown, however, how it would array sounds in PCO space that showed continuous changes in time-frequency structure. The morphing series was an attempt to examine this question.

We used synthetic signals modeled after the contact calls of Randy, Thalia, and Thoreau. In this case, we constructed two series of signals that resulted in both the Randy call type and the Thoreau call type metamorphosing into the Thalia call type. Each series consisted of the two endpoint call types and nine intermediate call types that moved between the endpoints in increments of 10%. This made a total of 21 signals in all for the morphing series.

There are many ways in which two contact call types can be morphed from one into the other. To make an intermediate signal that is x% of call type B, starting from call type A, we selected the following procedure. First, we adjusted the duration of the endpoint harmonic contours and corresponding endpoint amplitude weighting functions to the appropriate intermediate duration. If call type A has duration T_A and call type B has duration T_B, this intermediate duration equals:

$$T_A + \frac{x(T_B - T_A)}{100}.$$

Then, we calculated the appropriate weighted averages of the endpoint harmonic contours and the endpoint amplitude weighting functions for that intermediate. For example, the intermediate fundamental contour will equal x% of the type B fundamental plus (1-x)% of the type A fundamental. The amplitude weighting function of the intermediate fundamental contour will equal x% of the type B fundamental amplitude weighting function plus (1-x)% of the type A

fundamental amplitude weighting function, and so on. Finally, we assembled the intermediate signal from the appropriately calculated intermediate contours and weightings. In equation form, the signal which is $x\%$ of call type B starting from call type A equals

$$\sum_{k=1}^{15}\left(\frac{x}{100}a_{Bk}(t)+\left(1-\frac{x}{100}\right)a_{Ak}(t)\right)\left(\sin 2\pi t\left(\frac{x}{100}f_{Bk}(t)+\left(1-\frac{x}{100}\right)f_{Ak}(t)\right)\right)$$

where the durations of all functions equal $T_A+\frac{x(T_B-T_A)}{100}$.

While there are other possible algorithms for morphing between two call types, simply averaging the two endpoint spectrogram matrices is not one of them. The spectrogram resulting from such an operation would not resemble a single, smoothly varying contact call type. Rather, it would resemble the discontinuous pattern of two overlapping contact calls. Figure 11 shows part of our transformation series taking the Randy (11a) and Thoreau (11e) call types to the Thalia (11i) call type.

The 21 signals in the morphing series plus seven additional natural signal types (based on the individuals added in section III) were cross correlated and ordered using principal coordinates analysis. Figure 12 shows a three-dimensional scatter of the sounds plotted in the first three axes of PCO space (12a) which together account for 46% of the original inter-object distance. It also shows two-dimensional scatter plots of PCO1 and PCO3 versus PCO2 (12b,c), and scatter plots of the first three PCO measures versus signal order in the morphing series (12d-f). The signals in the morphing series are ordered from 1 to 11 beginning with either the Randy and Thoreau call types (both having order 1) and ending with the Thalia call type (having order 11). It can be seen from the scatter plots of Figure 12 that starting from the Randy and Thoreau call types, the two series of intermediate calls follow separate trajectories toward the Thalia call type. The sounds in the transformation series are arrayed in three-dimensional PCO space according to their generation order, as would be expected. However, the end points of the transformation series are closer to one another than each is to many of its intermediates. In addition, all of the natural call types plot closer together in PCO space than do many of the intermediates to either of the call types on which they are based.

These results are a consequence of the SPCC values themselves and not a by-product of ordination by PCO analysis. Figure 13 shows selected SPCC values as a function of signal order in each morphing series. In addition to cross-correlation values for the full signals of the morphing series (that is, signals with naturally weighted harmonics), Figure 13 shows correlation values for signals composed of uniformly weighted fundamental contours only. Although it is most clear in the plots for the full signal correlations, even the plots for the fundamental

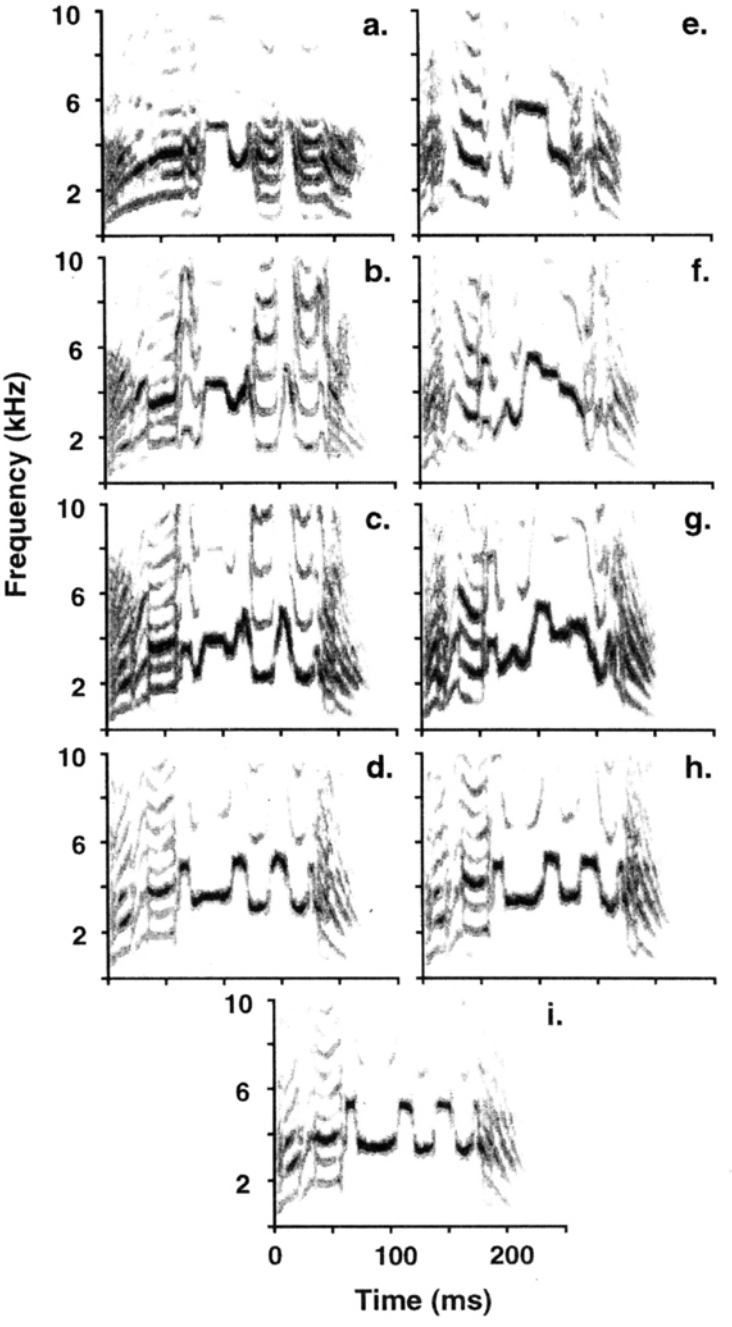

Figure 11. Examples from the morphing series showing the Thoreau (11a), Randy (11e), 70% Thoreau (11b), 70% Randy (11f), 40% Thoreau (11c), 40% Randy (11g), 10% Thoreau (11d), 10% Randy (11h), and Thalia (11i) call types.

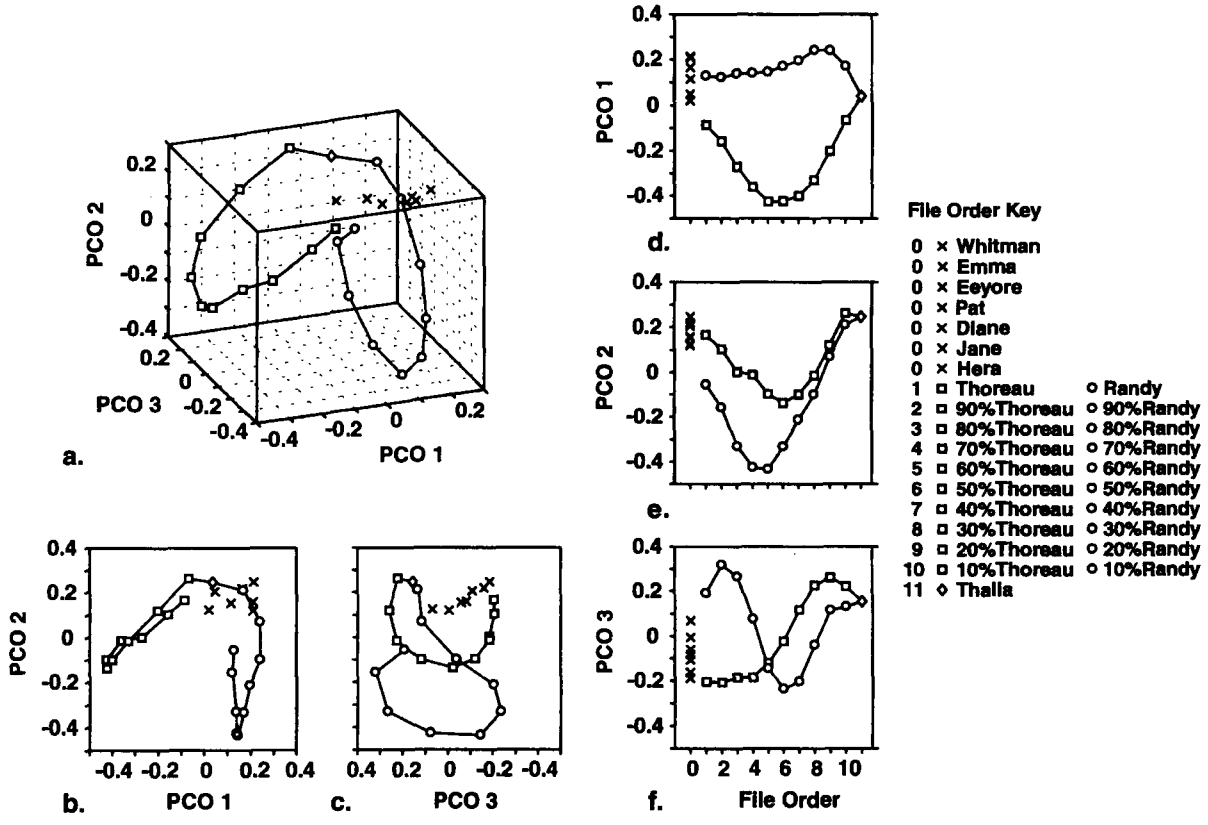

Figure 12. Two and three-dimensional PCO plots for the morphing series, plus plots of PCO1-3 versus signal order in the series. The different calls types and their associated order and symbol are shown in the file order key.

contour comparisons show that the peak correlation values do not fall off monotonically as might be expected given the way the series was mathematically constructed. Rather, the slopes of the SPCC versus signal order curves start out negative, reach zero at the most intermediate call types, and then reverse. For instance, in Figure 13a the Randy call type (number 1 = 100% Randy) has a higher correlation value with the Thalia call type (number 11 = 0% Randy) than it does with the intermediate call type that is 70% Randy and 30% Thalia (number 4 = 70% Randy). If one examines the plots of Figure 13 carefully, it becomes clear that adding naturally weighted harmonics amplifies, but does not distort, the patterns that are already visible in comparisons of fundamental contours.

We believe that these results arise because of the nature of the morphing algorithm we selected and the specific character of *Aratinga* contact calls. The latter typically contain a series of rectangular frequency modulations in the center of the call moving between roughly 3 and 6 kHz. Calls of different individuals have different numbers of these modulations located at different points within the central section of the call. As a consequence, our morphing rule produced intermediates in which the modulations of one endpoint call have largely disappeared and those of the second endpoint call have yet to become apparent. The result is a series of gentler modulations around 4.5 kHz in the middle of the intermediate call. This intermediate call type clearly is so different from either endpoint call type, or in fact any other naturally occurring contact call type that we have recorded, that it results in a low SPCC value when compared to the latter. Again, it appears that the resulting SPCC values faithfully reflected the spectrographic information provided.

DISCUSSION

Our results have a number of important implications. First, SPCC-PCO analysis appears to perform well with the classification and sorting of highly variable signal recordings. Even in the presence of random noise (with SNR ranging from –60 to +40 dB) and substantial fluctuations in signal duration (from 150 to 275 ms), signals with common time-frequency patterns formed discrete clusters in PCO space. Shared time-frequency pattern appears to be more important to the grouping of sound spectrograms using this method than are shared duration or signal-to-noise ratios. Second, SPCC-PCO analysis performs well in the comparison of harmonically rich sounds. Given an analysis bandwidth for spectrogram generation small enough to resolve signal overtones, a typical requirement for any spectrogram-based classification method, SPCC-PCO analysis sorted our harmonically-rich test sounds according to common time-frequency

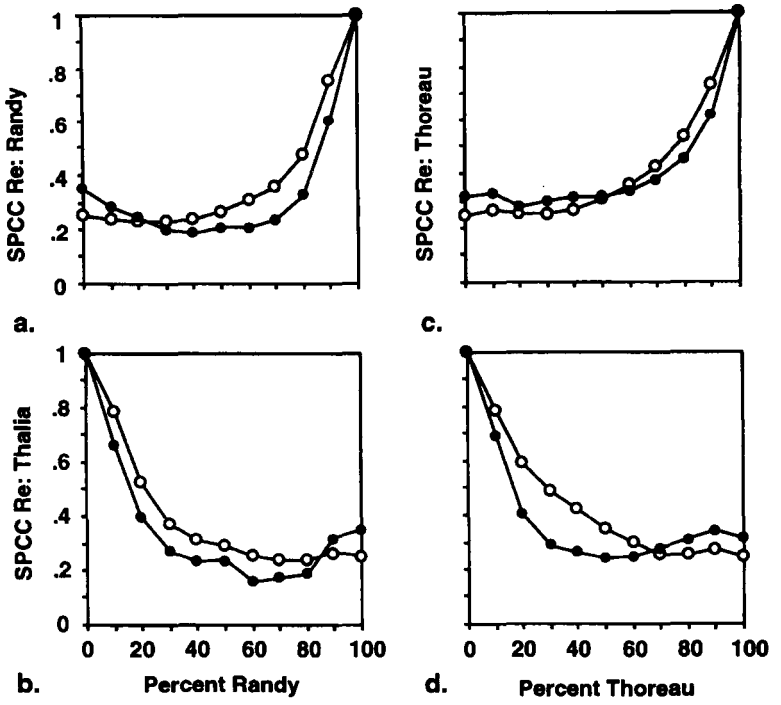

Figure 13. Plots of SPCC value re: the Thalia (13b,d), Randy (13a), or Thoreau (13c) call type versus file order in the morphing series (expressed as percent Randy or Thoreau call type). Plots are shown for SPCC comparisons made with sounds containing the fundamental contour only with uniform weighting (open circles), and the fundamental contour plus the next 14 harmonics with natural weightings (closed circles).

patterns. Again, the method was able to accomplish this despite the presence of considerable variation in signal duration, background noise level, and harmonic weighting pattern. Finally, retention of the naturally-weighted higher harmonics in harmonically-rich signals does not hinder their classification via SPCC-PCO. In our study, the method was better able to generate statistically significant groupings for sample sets in which the sounds contained the naturally-weighted higher harmonics rather than the fundamental contour only.

The implications of our findings are in contrast to the recommendations of Khanna et al. (1997). A detailed examination of their data showed that the reported sensitivity of SPCC values to FFT length when signals vary in duration or overtone content is largely a side effect of the time-frequency trade-off inherent in Fourier analysis. Spectrographic cross-correlation analysis is no more sensitive to selection of analysis window (FFT) size than is any other classification method based on spectrograms. That is, if a spectrogram does not resolve structural details well enough for standard measure or visual

classification methods, then it is unreasonable to expect that it will resolve them well enough for SPCC analysis. Only structural differences detectable within the effective resolution of a spectrogram will be detectable by spectrographic cross-correlation.

The Khanna et al. study correctly reveals the danger of performing SPCC comparisons without examining the relevant spectrograms. Digital sound analysis programs, such as Signal or Canary, have made it easy for researchers to move from sound capture to SPCC similarity matrices directly. Those who use these methods, but never examine the relevant spectrograms, risk facing the pitfalls cited by Khanna et al. (1997). In choosing the analysis window length used in constructing spectrograms, a researcher imposes certain constraints on the resolution of the analysis. Patterns of signal time-frequency structure that are smaller than the specified spectrogram resolution (which is a function of both windowing and FFT length) are undetectable. This has the advantage of letting the researcher decide what level of time-frequency structure is of interest for a particular set of sounds; however, it puts the burden of intelligent choice of analysis window and subsequent interpretation on the researcher. Obviously, only structural features that are visible given the effective resolution of the spectrogram will be detectable by cross-correlation analysis. A poor choice of analysis window size could result in biologically important features being missed. Our recommendation, therefore, is to always examine spectrograms before using SPCC to be sure that time-frequency features are portrayed as expected. Researchers interested in comparing sound sets at multiple structural levels can always use spectrograms generated at the appropriate multiple FFT lengths.

The combination of SPCC and PCO analysis appears to have significant advantages over the current practice of applying hierarchical or k-means cluster analysis to SPCC similarity matrices. First, the eigen-decomposition of similarity matrices by PCO appears to allow the investigator to separate out contributions to SPCC values that are due to generally shared features from those idiosyncratic to single sounds. Because the order of principal coordinate extraction is determined by the amount of the original inter-object distance explained (akin to the amount of variance explained), the first few PCO coordinates turn out to be the ones that reflect the most generally shared structural details. Subsequent coordinates represent less of the overall variation; thus, they are more likely to reflect idiosyncrasies of individual calls. In contrast, cluster analysis is based on the entire similarity matrix: any kind of variation may affect the final classification. Another advantage of PCO is the reduction of the data space to a few important PCO axes. This provides the opportunity to view the ensemble of sounds in two or three dimensions. Such visual inspection, especially of animated (rotating) three-dimensional plots, is

extremely useful for seeing how different calls are related to each other. Similarly, one can calculate the location of and compute the distances between the centroids of different clusters of sounds in PCO space. Often researchers are interested in whether the clustered renditions of a song type of interest (type A) are on average more similar to one song type (type B) or to another (type C). Calculation of centroid distances can provide quantitative answers to such questions. Finally, principal coordinates analysis provides a set of composite measures (the PCO values) that can be used in statistical tests of association between sound structure and various external variables. These options are not available using cluster analysis.

While the SPCC-PCO method appears robust when applied to single continuous sounds, it is likely to break down if applied to long, multi-syllable signals. Songs in which syllable order and structure were identical but the silent spaces between syllables varied would be difficult to align for cross-correlation. Low correlation values could arise simply because of variation in the length of silent sections between syllables. Readers should be aware of this if applying the SPCC-PCO method to long, multi-part sounds. However, if such sounds were highly stereotyped in syllable placement there is no reason why the method might not work well. We have not yet explored this application systematically and would be interested in the results of such efforts.

As noted in the introduction, classification of sounds is only useful if class membership is significantly associated with some extrinsic contextual variable of interest. Janik (1998) refers to this as "external validity". While Janik emphasized the correspondence of sound classification and caller identity, it is important to note that there may be multiple criteria of external validity for a given set of sounds. Appropriate classification schemes might differ depending on whether the contextual variable were, for example: 1) habitat and its associated effects on sound propagation; 2) identity of the caller or its social unit; 3) the mechanism of sound production; 4) the way in which a receiver of that species categorizes the sounds; or 5) the behavioral function of those sounds. There is no reason to expect to find a single classification scheme for a set of sounds that is valid for all of these different contexts. Our method allows for statistical tests of external validity using different combinations of composite sound measures based on principal coordinates. Several coordinates can be considered at one time using MANOVA or LDA. These statistical analyses essentially perform a rotation of the PCO axes until a maximal level of sound separation is achieved based on factor level. If the separation between clusters is sufficiently great, given within-cluster spread, the association will be statistically significant. It may be necessary to rotate the PCO axes in a different way to show significant association with a different external factor. The canonical variates generated by

MANOVA indicate which principal coordinates are most important in effecting this association and the direction of their effect. In the case of the LDA, one can use stepwise LDA to find the subset of PCO coordinates that is best correlated with the external variable of interest.

There is an obvious limitation of SPCC-PCO analysis when compared to standard measure techniques. Even if one finds a significant association between an external variable and a specific subset of the principal coordinates, one does not know which structural details in the spectrograms are the salient ones. In many studies, this is not a problem: one may simply want to know whether or not a significant association between sound structure and an external variable is present. The specific signal details generating this association may not be important to the question at hand. However, when it is necessary to identify the important structural details of the sound, one can subsequently take a set of standard measures and determine which ones are correlated with the significantly associated principal coordinates. Why bother with SPCC-PCO analysis if in the end standard measures must be taken anyway? The answer is that by using SPCC, as opposed to standard measures, one is less likely to miss the critical structural detail. As stated before, a finite set of measures, no matter how large, could still miss the structural features of a sound most closely associated with a particular external variable. All features of the sound spectrograms are taken into account, however, when making SPCC comparisons. Once a significant association between an external variable and a principal coordinate is found, a search criterion is established. One can then begin to look for those spectrographic details that are both correlated with the principal coordinate and with the external variable of interest. Thus, SPCC-PCO analysis could be used effectively for the initial screening of data sets.

Finally, while it is widely considered appropriate to compare only the fundamental contours of harmonically rich signals (Buck and Tyack 1993, McCowan 1995, Khanna et al. 1997, Janik 1999), our results suggest that this practice should be reconsidered. First, elimination of higher harmonics does not appear necessary. We find no evidence that harmonic structure undermines the validity or robustness of the SPCC-PCO method. In fact, our analyses demonstrate that comparison of spectrograms with higher harmonics and their natural weightings tends to enhance the group separation patterns seen in comparison of their fundamental contours only. Second, although the frequency information contained in the harmonics is redundant given the fundamental, the pattern information contained in the relative amplitude weightings of those harmonics is not. These amplitude weighting patterns could be correlated with extrinsic variables related to sound propagation, sound generation, caller identity, or sound function. Unless one knows *a priori* that the harmonic

amplitude weightings are irrelevant, it seems ill advised to exclude them from consideration. The pattern of the harmonic amplitude weighting functions provides an additional dimension for sound comparison not available with the fundamental contours alone. If the harmonics are equally weighted for all sounds in a set of sounds, however, retention of higher harmonics would provide no additional pattern information. Given the results of section IV it may be prudent to compare fundamental contours only in these cases. In general however, it seems that SPCC values generated by the comparisons of sounds with their full harmonic content will be most representative of the total information available to sound receivers.

ACKNOWLEDGEMENTS

This work was funded by NSF grant IBN-94-06217. We thank Sandra Vehrencamp and Tim Wright for frequent feedback and help in the development of the method and Michael Beecher, Sandra Gaunt, Alejandro Purgue, Kurt Fristrup and an anonymous reviewer for thoughtful critique of the manuscript.

REFERENCES

Beecher, M. D. (1988). Spectrographic analysis of animal vocalizations: implications of the "uncertainty principle." *Bioacoustics*, **1**, 187-208.

Bracewell, R. N. (1986). *The Fourier Transform and Its Applications, Second Edition, Revised.* WCB/McGraw-Hill; Boston, Massachusetts.

Bradbury, J. W. & Vehrencamp, S. L. (1998). *Principles of Animal Communication.* Sinauer Associates; Sunderland, Massachusetts.

Buck, J. R. & 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. & Beeman, K. (1987). Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology*, **76**, 101-115.

Everitt, B. S. & Dunn, G. (1991). *Applied Multivariate Data Analysis.* Arnold; London. John Wiley & Sons, Inc.; New York.

Farabaugh, S. M., Linzenbold, A., & Dooling, R.J. (1994). Vocal plasticity in budgerigars (*Melopsittacus undulatus*): evidence for social factors in the learning of contact calls. *J. Comp. Psych.*, **108**, 81-92.

Farabaugh, S. M. & Dooling R. J. (1996). Acoustic communication in parrots: laboratory and field studies of budgerigars, *Melopsittacus undulatus*. In *Ecology and Evolution of Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, eds). Comstock Publishing Associates, Cornell University Press; Ithaca, pp. 97-117.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325-338.

Gower, J. C. (1987) Introduction to ordination techniques. In *Developments in Numerical Ecology, NATO ASI Series, Vol. G14* (P. Legendre and L. Legendre, eds). Springer-Verlag; Berlin, pp. 5-64.

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

Khanna, H., Gaunt, S. L. L., & McCallum, D. A. (1997). Digital spectrographic cross-correlation: tests of sensitivity. *Bioacoustics*, **7**, 209-234.

Kroodsma, D. E. & Miller, E. H. (1996). *Ecology and Evolution of Acoustic Communication in Birds*. Comstock Publishing Associates, Cornell University Press; Ithaca.

Kruskal, J. B. & Wish, M. (1978). *Multidimensional Scaling*. Sage Publications; London.

Legendre, P. & Legendre, L. (1998). *Numerical Ecology, Second English Addition: Developments in Environmental Modelling, 20*. Elsevier Science B.V.; Amsterdam.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209-220.

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.

Neff, N. A. & Marcus, L. F. (1980). *A Survey of Multivariate Methods for Systematics*. Privately published, American Museum of Natural History; New York.

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

Oppenheim, A. V. & Schafer, R. W. (1989). *Discrete-time Signal Processing*. Prentice Hall, Inc.; New Jersey.

Reiss, D. & McCowan, B. (1993). Spontaneous vocal mimicry and production by Bottlenose dolphins (Tursiops truncatus): Evidence for vocal learning. *J. Comp. Psych.*, **107**, 301-312.

Schnell, G. D., Watt, D. J., & Douglas, M. E. (1985). Statistical comparison of proximity matrices: applications in animal behaviour. *Animal Behav.*, **33**, 239-253.

Sebeok, T.A. (1977). *How Animals Communicate*. Indiana University Press; Bloomington.

Sharma, S. (1996). *Applied Multivariate Techniques*. John Wiley & Sons, Inc.; New York.

Snowdon, C. T. & Hausberger M. (1997). *Social influences on vocal development*. Cambridge University Press; Cambridge.

Tabachnick, B. G. & Fidell, L. S. (1996). *Using Multivariate Statistics, Third Edition*. Harper Collins College Publishers; New York.

Wong, P. C. & Bergeron, R. D. (1997). Multivariate visualization using metric scaling. In *Visualization '97: Proceedings of the 8th IEEE Conference on Visualization* (R. Yagel and H. Hagen, eds). ACM Press; New York, pp. 111-118.

APPENDIX

Principal coordinates analysis: obtaining a multivariate representation of data from a matrix of between-object similarity or distance values

What follows is a brief overview of the mathematical rationale behind the PCO method. The information provided here should enable interested researchers to implement the algorithm for performing the PCO analysis of SPCC or other similarity or distance matrices. A matrix-based programming environment such as Matlab (The Math Works, Inc.) would make such implementation especially straightforward. Readers interested in a more complete treatment of the PCO method are referred to Gower (1966, 1987), Neff and Marcus (1980), and Legendre and Legendre (1998).

Suppose we are given a matrix **D** of distance measures between the n objects of a set

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{pmatrix}.$$

Alternatively, we can begin with a matrix of similarity measures s_{ij} that we convert to distance measures d_{ij} by some standard transform (for example, $d_{ij} = 1 - s_{ij}$, or $d_{ij} = \sqrt{1 - s_{ij}}$). The entries d_{ij} can be transformed into new variables b_{ij} such that

$$b_{ij} = -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 - \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 \right).$$

Why make this transformation? Suppose that we had access to the theoretical raw data matrix $\mathbf{X}$ consisting of the values $x_{ij} \in \Re$ of some set of p variables measured for each of the n objects

$$\mathbf{X} = \begin{matrix} \\ p \text{ vars} \\ \downarrow \end{matrix} \begin{matrix} n \text{ objs} \rightarrow \\ \end{matrix} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix} = \begin{pmatrix} \mathbf{y_1} & \mathbf{y_2} & \cdots & \mathbf{y_n} \end{pmatrix}.$$

The coordinate vectors of any two given objects i and j relative to a basis composed of the p variables are

$$\mathbf{y_i} = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{pi} \end{pmatrix} \text{ and } \mathbf{y_j} = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{pmatrix}.$$

So, the squared distance between the objects i and j is

$$d_{ij}^2 = \left\| \mathbf{y_i} - \mathbf{y_j} \right\|^2 = \left\langle \mathbf{y_i} - \mathbf{y_j}, \mathbf{y_i} - \mathbf{y_j} \right\rangle$$

$$= \left(\mathbf{y_i} - \mathbf{y_j} \right)^{\mathbf{t}} \left(\mathbf{y_i} - \mathbf{y_j} \right) = \mathbf{y_i^t y_i} + \mathbf{y_j^t y_j} - 2\mathbf{y_i^t y_j}$$

where $\langle\ \rangle$ denotes the standard inner product between two vectors. Assuming that our data have been standardized (so that there is zero mean and unit variance), then

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{y_i^t y_j} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n} x_{ki}x_{kj} = \sum_{k=1}^{n} x_{kj}\frac{1}{n}\sum_{i=1}^{n} x_{ki} = 0.$$

So,

$$\frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 = \mathbf{y_j^t y_j} + \frac{1}{n}\sum_{i=1}^{n} \mathbf{y_i^t y_i},$$

and similarly,

$$\frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 = \mathbf{y_i^t y_i} + \frac{1}{n}\sum_{j=1}^{n} \mathbf{y_j^t y_j},$$

therefore,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{y_i^t y_i} + \frac{1}{n}\sum_{j=1}^{n} \mathbf{y_j^t y_j}.$$

Now, define an inner product matrix **B** with elements $b_{ij} = \langle \mathbf{y_j}, \mathbf{y_i} \rangle = \mathbf{y_i^t y_j} = \left(\mathbf{y_j^t y_i} = b_{ji}\right)$. We see that,

$$b_{ij} = -\frac{1}{2}\left(d_{ij}^2 - \mathbf{y_i^t y_i} - \mathbf{y_j^t y_j}\right) = -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2\right)$$

the transform from above. So, by making this transformation, we can convert our matrix of absolute distances between objects to a matrix of inner products between object vectors whose common origin is the centroid of the standardized data. The absolute distances between the objects are unchanged; we have simply expressed their positional relationship in a different way by using vector inner products. Now, since

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} \mathbf{y_1^t y_1} & \mathbf{y_1^t y_2} & \cdots & \mathbf{y_1^t y_n} \\ \mathbf{y_2^t y_1} & \mathbf{y_2^t y_2} & \cdots & \mathbf{y_2^t y_n} \\ \vdots & \vdots & & \vdots \\ \mathbf{y_n^t y_1} & \mathbf{y_n^t y_2} & \cdots & \mathbf{y_n^t y_n} \end{pmatrix} = \begin{pmatrix} \mathbf{y_1^t} \\ \mathbf{y_2^t} \\ \vdots \\ \mathbf{y_n^t} \end{pmatrix} \begin{pmatrix} \mathbf{y_1} & \mathbf{y_2} & \cdots & \mathbf{y_n} \end{pmatrix}$$

$$\mathbf{B} = \mathbf{X^t X}.$$

And since **B** is square symmetric, **B** is orthogonally diagonalizable, so

$$\mathbf{B} = \mathbf{Q}\Lambda\mathbf{Q^t} = (\mathbf{u_1}\ \mathbf{u_2}\ \cdots\ \mathbf{u_n}) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \mathbf{u_1^t} \\ \mathbf{u_2^t} \\ \vdots \\ \mathbf{u_n^t} \end{pmatrix}$$

$$= (\mathbf{u_1}\ \mathbf{u_2}\ \cdots\ \mathbf{u_n}) \begin{pmatrix} \lambda_1 \mathbf{u_1^t} \\ \lambda_2 \mathbf{u_2^t} \\ \vdots \\ \lambda_n \mathbf{u_n^t} \end{pmatrix} = \sum_{i=1}^{n} \lambda_i \mathbf{u_i u_i^t}$$

where the λ_i are the eigenvalues and $\mathbf{u_i}$ the corresponding orthonormal eigenvectors of **B**. From the summation, we see that the best approximation to the matrix **B** is provided by its largest eigenvalue and corresponding eigenvector. If we order the eigenvalues along the diagonal of Λ from largest to smallest, then

$$\mathbf{B} \cong \lambda_1\ \mathbf{u_1 u_1^t} = \lambda_1 \begin{pmatrix} u_{11} \\ u_{21} \\ \vdots \\ u_{n1} \end{pmatrix} (u_{11}\ u_{21}\ \cdots\ u_{n1}) = \begin{pmatrix} \hat{\mathbf{y}}_1^t \\ \hat{\mathbf{y}}_2^t \\ \vdots \\ \hat{\mathbf{y}}_n^t \end{pmatrix} (\hat{\mathbf{y}}_1\ \hat{\mathbf{y}}_2\ \cdots\ \hat{\mathbf{y}}_n).$$

Thus, $\hat{\mathbf{y}}_1=(\lambda_1)^{\frac{1}{2}}u_{11}, \hat{\mathbf{y}}_2=(\lambda_1)^{\frac{1}{2}}u_{21}, \cdots, \hat{\mathbf{y}}_n=(\lambda_1)^{\frac{1}{2}}u_{n1}$ are the coordinate vectors for the objects that best approximate their original inner products $\langle \mathbf{y_j}, \mathbf{y_i} \rangle = \mathbf{y_i^t y_j}$. These coordinate vectors are actually just scalars $\hat{y}_i$ and they give the coordinates of the objects on the first axis (PCO1) that does the best job of approximating their original inner products. We see that these coordinates are simply the coordinates of the first eigenvector times the square root of its corresponding eigenvalue

$$\hat{\mathbf{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = (\lambda_1)^{\frac{1}{2}}\mathbf{u_1}.$$

The coordinates of the objects on the second best axis (PCO2) are the coordinates of the second eigenvector of **B** times the square root of its corresponding eigenvalue, and so on.

Various measures can be used to assess the goodness of fit of the d PCO dimensions used. Some common relationships to quantify the degree of accuracy of the representation include $\frac{\sum_{i=1}^{d}\lambda_i^2}{\sum_{i=1}^{n}\lambda_i^2}$ (Gower 1987), $\frac{\sum_{i=1}^{d}\lambda_i}{\sum_{i=1}^{n}|\lambda_i|}$ (Everitt and Dunn 1991), and $\frac{\sum_{i=1}^{d}\lambda_i}{\sum_{i=1}^{p}\lambda_i}$ (Wong and Bergeron 1997), where $\lambda_1, \ldots, \lambda_d$ are the first d positive eigenvalues used in the approximation, $\lambda_1, \ldots, \lambda_p$ are all the positive eigenvalues, and $\lambda_1, \ldots, \lambda_n$ are all the eigenvalues.

Received 2 July 1999, revised 30 November 1999 and accepted 19 January 2000