The acoustic features of vervet monkey alarm calls

Michael J. Owren[a] and Robert H. Bernacki[b]
Department of Psychology, Indiana University, Bloomington, Indiana 47405

(Received 29 July 1987; accepted for publication 1 February 1988)

Vervet monkeys routinely produce semantic alarm calls upon detection of various predators encountered in their natural environment. Two of these calls, snake and eagle alarms, were analyzed using digital signal processing techniques in order to identify potentially distinctive acoustic cues. Distinctive cues were sought in the periodicity of the source waveform associated with each call type, the probable vocal tract filtering functions, and in temporal patterning. Results were equivocal with respect to source periodicity, but a variety of distinguishing features were found in both supralaryngeal filtering and timing. These data provide a basis for psychoacoustic perceptual testing with vervets as subjects.

PACS numbers: 43.80.Jz, 43.80.Lb, 43.80.Nd

INTRODUCTION

Recent work has suggested that the natural calling behavior of a variety of nonhuman primates bears fundamental resemblances to certain aspects of human speech communication. For example, humanlike imposition of perceptual categories on continuously varying vocal communication signals has been shown in at least two primate species (Hopp, 1985; Petersen, 1982; Petersen *et al.*, 1984; Snowdon and Pola, 1978). Several studies have confirmed the importance of lateralized processing mechanisms for decoding a phoneticlike contrast used by one of these animals, Japanese macaques (Heffner and Heffner, 1984; Petersen *et al.*, 1978, 1984). Unsuspected complexities in call usage have also been demonstrated. Green (1975), for instance, has shown that subtle variations in tonal sounds used by Japanese macaques are correlated with specific behavioral contexts. In addition, several studies have demonstrated semanticlike referential signaling in various species (Cheney and Seyfarth, 1982; Dittus, 1984; Gouzoules *et al.*, 1984; Seyfarth *et al.*, 1980a,b).

These studies demonstrate the importance of investigations of primate vocal behavior both for gaining insight into the behavior of the animals themselves, finding animal models of human speech processes, and in understanding the evolution of speech. They also illustrate the value of combining field studies of natural calling behavior with laboratory investigations of perception in the same species. Thus far, however, such studies have been limited both by a lack of detailed data about the acoustic features of primate vocalizations and how these features contribute to the communication process.

Given appropriate field documentation, both acoustic analysis and perceptual experimentation are necessary to determine which features are used to distinguish or identify particular sounds (Seyfarth and Cheney, 1984b). In this article, the results of an acoustic analysis of two vervet monkey (*Cercopithecus aethiops*) alarm calls are presented. The potentially important features we uncovered were then tested in a series of perceptual experiments in which vervets and humans classified both natural and synthetic alarm calls. The results of those studies are presented elsewhere (Owren, 1988).

Vervet alarm calls were selected for study because they represent what is arguably the clearest available example of humanlike semantic vocal communication in a nonhuman primate. These animals, who live in groups of 8 to 20 individuals (Seyfarth and Cheney, 1984a), have been shown to produce at least three acoustically distinctive and predator-specific calls (Seyfarth *et al.*, 1980a,b; Struhsaker, 1967). Two of these sounds, snake and eagle calls, are harsh, noisy signals based on the repetition of short iterative elements. They are superficially similar spectrographically, but distinct in sound and function. Hearing snake alarm calls, vervets typically stand on their hind legs, locate the predator through visual scanning, and then mob it. Eagle calls are given only to a few raptor species and elicit aerial scanning or escape into dense brush. The semantic usage of these sounds by vervets has been demonstrated through playing back previously tape-recorded alarm calls in the absence of any predator. Such playbacks elicit the same predator-specific escape responses from listening individuals (Seyfarth *et al.*, 1980a,b).

Vervet calls also represent one of the few instances in which the digital signal processing techniques common in speech research have been applied to nonhuman primate sounds. Seyfarth and Cheney (1984b) used these techniques to successfully characterize subtle acoustic variation in vervet "grunt" calls that were not amenable to spectrographic examination. These calls are acoustically similar to snake and eagle calls.

I. METHOD

A. The alarm call sample

The calls used were tape-recorded from free-ranging vervets in Amboseli National Park, Kenya (see Seyfarth *et al.*, 1980a,b, for details of the recording procedures). Snake and eagle calls consist of one or more phrases (Seyfarth *et al.*, 1980a,b; Struhsaker, 1967) that are, in turn, composed of brief, repeated segments, or iterations. Spectrograms of

[a] Present address: California Primate Research Center, University of California, Davis, CA 95616.
[b] Present address: Exercise Science Laboratories, Rm. 112, School of HPER, Indiana University, Bloomington, IN 47405.

 0001-4966/88/051927-09$00.80

snake and eagle alarm call phrases are shown in Fig. 1.

In all, 89 snake and eagle alarm calls were available for analysis, consisting of calls produced by 21 different animals representing all age/sex classes. Both Struhsaker (1967) and Seyfarth *et al.* (1980a,b) report that, although individual differences are apparent in the acoustic structure of these alarm calls, type-specific cues are always preserved. To maximize the probability of finding these cues in the face of possible individual-, gender-, and age-related variation, the analysis was restricted to calls produced by adult females, who were also best represented in the sample.

Potentially distinctive acoustic cues were sought in the characteristics of the underlying source waveform, in the vocal tract filtering applied to the source, and in temporal aspects of the calls. Although consistent amplitude differences have been reported between these call types (Seyfarth *et al.*, 1980a,b; Struhsaker, 1967), playback studies have shown that this variation is not important in identification (Seyfarth and Cheney, 1982; Seyfarth *et al.*, 1980a,b). Amplitude was therefore not considered to be a potentially distinctive cue.

B. Apparatus

All acoustic analyses were done using the facilities of the Indiana University Speech Research Laboratory, which have been described in detail by Forshee (1975; Forshee and Nusbaum, 1984). Digitally based analysis was performed either on a DEC PDP-11/34, or a VAX 11/750 computer system. Both were equipped with digitizing, display, printout, and hard copy capabilities. Spectrographic analysis was performed on a Voiceprint Identification Inc. 700 series instrument.

Programs implemented on the PDP-11/34 were WAVES, WAVMOD, and SPECTRUM. WAVES, a waveform editor (Luce and Carrell, 1981), was initially used to digitize signals at a 10-kHz sampling rate with automatic low-pass filtering at 4.8 kHz. In addition, temporal parameters of the sampled calls were measured on a point-by-point basis using this program. WAVMOD, a digital waveform modification program (Bernacki, 1981), was used to measure waveform amplitudes. SPECTRUM (Kewley-Port, 1979) was used for analysis of various spectral properties of the signals using the FFT and LPC methods. Cepstral fundamental frequency extraction was performed with the ILS digital signal processing package (Interactive Laboratory System, version 4.2; Signal Technology Inc.) implemented on the VAX-11/750. The BMDP software package (Dixon, 1983) was used for most statistical analyses.

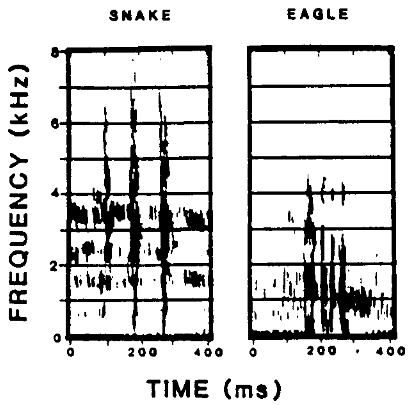

FIG. 1. Wideband spectrograms of single phrases taken from snake and eagle alarm calls.

C. Procedure

1. Preliminary screening

Due to the relatively small number of calls available for analysis, statistical treatment of sets of calls produced by single individuals was not possible. As analysis was intended to provide hypotheses for future testing rather than to identify critical cues indisputably, some liberty was taken in applying statistical tests. Specifically, every call was treated as an independent event even though the number of calls analyzed far exceeded the number of callers.

Noise reduction manipulations were not attempted on the sample. Instead, signal-to-noise ratios were calculated for the adult female snake and eagle calls, and a small set of the highest quality phrases was selected. To do so, the beginning and end of each iteration was marked at a zero crossing in the WAVES display (see Fig. 2). It was apparent from calls with low levels of background noise that interiteration intervals within phrases contained no signal energy. Measurements of these intervals were therefore combined with measurements taken over approximately 100-ms segments immediately preceding and following each phrase to obtain the mean background noise level. The ratio of mean iteration amplitude to mean background amplitude was calculated as

$$\mathrm{dB}_{sn} = S - N, \tag{1}$$

where S represents the signal (vocalization) energy and N is the background energy. The best single phrase from each call was identified, and was included in analysis only if this ratio exceeded 10.0 dB.

Preliminary screening produced a final set of 27 analysis phrases. Twelve of these were snake alarms, representing five different individuals. Fifteen were eagle alarms, produced by eight different callers. Three animals were represented by at least one phrase of each type. Signal-to-noise ratios of the selected snake and eagle phrases were not statistically different, $t(25) = -1.28, p > 0.05$, two tailed.

2. Source function analysis

Due to present ignorance about the physical mechanisms of vervet call production, only limited examination of possible source function parameters was possible. Source functions that are a mixture of voicing and noise are not well modeled using current digital signal processing techniques, and are also difficult to isolate in spectrographic analysis. However, Seyfarth and Cheney (1984b) have reported successful extraction of irregular periodicity in vervet grunt calls using cepstral fundamental frequency extraction. A combination of this technique with inspection of spectrographic representations was therefore employed.

Evidence of vertical striations were sought in wideband (450 Hz) spectrograms produced both at normal and one-half tape speed (lowering the speed increases temporal resolution), and with regular and lowered (-15 dB) gain set-

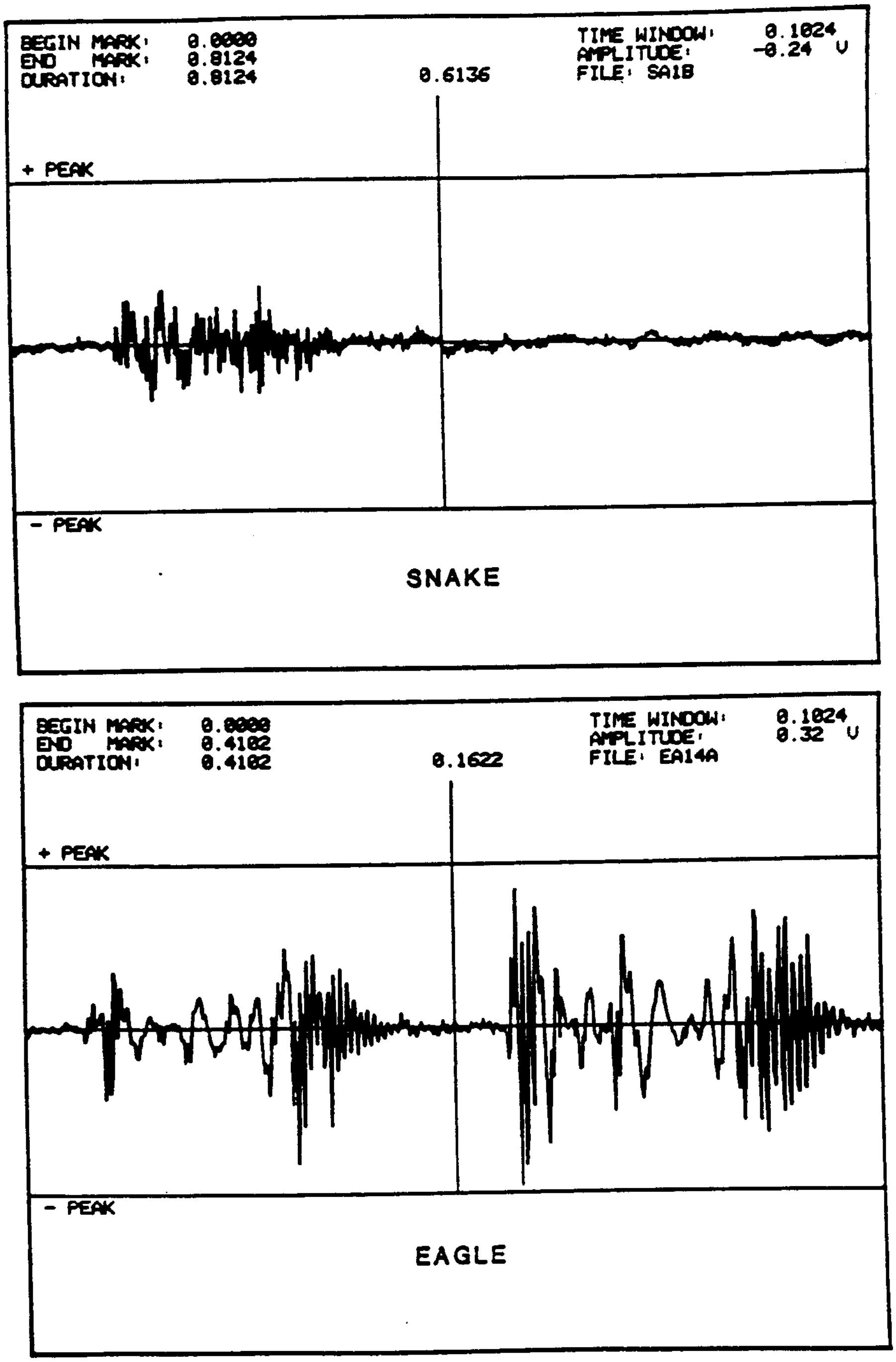

FIG. 2. Waveform displays of snake and eagle iterations, produced by the WAVES program.

tings. Narrow-band (45 Hz) spectrograms were examined for evidence of regularly spaced harmonics. Cepstral pitch analysis was then performed on each phrase using the ILS command API. Analysis parameters were determined through trial and error exploration so that as many voiced frames as possible were found (API command parameters: Periodicity buffer = 12 ms; rectangular cepstral window, center frequency = 300 Hz; minimum frequency = 200 Hz; maximum frequency = 500 Hz; slope factor = − 40; voicing decision criterion = − 500).

3. Spectral filtering analysis

Most spectral filtering analysis was done with the SPECTRUM program, including FFTs, LPC modeling, and mean amplitude analysis. Formant frequencies, bandwidths, and peak amplitudes were computed from LPC coefficients using the POLYRT subroutine (IBM Scientific Subroutine Package).

All phrases were examined at the iteration level. Prior to analysis, each iteration was placed in a separate file with the context (frame size) set to one-fourth of the overall signal duration. Iterations were examined as a whole (or up to 51.2 ms from the beginning of the signal), and as five overlapping segments with the analysis window set to twice the frame size. Hamming windows were used throughout.

In FFT analysis, the resolution of the transform was matched as well as possible to the length of the analysis window. In LPC modeling, parameter values for a number of reflection coefficients and preemphasis were selected after both systematic examination of the effects of various settings on the results from two phrases of each type, and more cursory exploration of a larger number of calls. Six reflection coefficients were found to yield maximum consistency in peak frequencies and bandwidths, and also resulted in the closest subjectively estimated "fit" between the LPC spectral envelope and the FFT. Preemphasis was set to zero as a conservative measure in light of the unknown glottal and radiation characteristics of vocalizing vervets. In addition, however, elimination of preemphasis allowed the use of the first LPC reflection coefficient as a measure of overall spectral flatness or tilt (cf. Markel and Gray, 1976).

Measurement variables used were: *PKx* (the center frequency of peak *x* in the envelope, in Hz); *BWx* (the bandwidth of peak *x*, in Hz); *DBx* (the amplitude of peak *x* relative to the average amplitude of the total spectrum, in dB); *TILT* [$RC(1)\times 100$; the overall tilt of the envelope, i.e., rising $(0 < TILT <= +100)$]; *flat* $(TILT = 0)$; *falling* $(-100 <= TILT < 0)$; *Qx* (the peak to bandwidth ratio for peak *x*); *PKy*/*PKx* (the ratio between each higher frequency peak *y*, and lower frequency peak *x*); and *BWy*/*BWx* (the ratio between bandwidths of higher and lower peaks).

4. *Timing analysis*

The variables of interest in segment duration analysis were number of iterations in each phrase, iteration duration, the total duration of each phrase, cumulative durations of all iterations in a phrase, the percentage of signal energy in each phrase, and the interiteration interval durations in each phrase. Each phrase received one score for each of these variables. Iteration-related scores were mean values for all iterations in the phrase.

To test the possibility that the call types might exhibit distinctive amplitude patterns within phrases, each phrase was characterized by three scores normalized to the overall mean amplitude of all the component iterations. The first and third scores were based on the initial and final iterations, respectively. If three or five iterations were present, the second score reflected the amplitude of the middle iteration. If four iterations were present, the second and third were averaged, and if six were present, iterations three and four were combined.

II. RESULTS

A. Source function

To the human ear, there is subtle, but consistent, source periodicity in eagle alarms relative to snake alarms. Both spectrographic and cepstral analyses tended to support this conclusion, but the results were not clear-cut.

In spectrographic analysis, snake phrases revealed no appreciable energy below 1 kHz, the expected location of any fundamental frequency excitation. Broadband analysis did not result in the appearance of indisputable vertical striations, and no harmonic structure was found in narrow-band representations. Broadband analysis of eagle calls consistently showed energy below 1 kHz but other evidence was equivocal. Only with the tape slowed and the gain lowered did a majority of eagle phrases appear to exhibit regular vertical striations. In narrow-band analysis, however, no harmonic structure was apparent in these calls either.

Cepstral analysis revealed regularity in 52 of a total of 778 20-ms eagle frames, and 10 of the 839 snake frames. This difference was statistically significant, $\lambda^2(1,N = 1617) = 31.6, p < 0.001$. Of 56 eagle iterations examined, 10 contained periodicity. Only 3 of 43 snake iterations were so characterized but this difference was not significant, $\lambda^2(1,N = 99) = 1.66, p > 0.10$. Seven of 15 eagle phrases, and 3 of 12 snake phrases contained some periodicity. These proportions could easily have occurred by chance: Fisher Exact Probability test, $p = 0.33$. The fundamental frequencies returned by the algorithm for the seven eagle phrases were quite consistent (303.3, 313.0, 323.6, 323.0, 347.8, 351.0, and 368.1). For snakes, the values returned bracketed the eagle scores (224.5, 370.7, 385.0). A Wald–Wolfowitz runs test failed to show that these scores were significantly different, however, $r(7,3) = 3, p > 0.05$.

B. Spectral filtering

The number of peaks found in individual iterations varied both within and between phrases. In most cases, two peaks were present but a number of three-peak iterations were also found. Overall, three peaks occurred more often in eagle phrases, as shown in Table I. Further analysis was conducted on two- and three-peak iterations separately. For all phrases, mean scores were determined on each variable based on the two- or three-peak iterations. As a result, 11 two-peak scores for both snakes and eagles, and 6 snake and 10 eagle three-peak scores were obtained. The difference in number of snake and eagle phrases containing two- and three-peak iterations was not statistically significant, $\lambda^2(1) = 2.41, 0.10 < p < 0.20$, two tailed.

1. *Two-peak iterations*

Differences between snake and eagle phrases for two-peak data were first compared via analysis of variance. The results are shown in Table II. In summary, snake phrases had a significantly higher frequency, broader bandwidth, and lower amplitude first peak than eagle phrases. The second peak was similar in frequency but had a narrower bandwidth and higher amplitude in snakes. Snake spectra were also significantly flatter than those of eagles, which had a clear downward tilt with increasing frequency. Prototypical spectral shapes drawn from mean data for two-peak iterations are shown in Fig. 3. It is apparent from the results that any of these variables could potentially be used to differen-

TABLE I. Distribution of spectral peaks in the analysis phrases.

		Call type	
Acoustic unit	No. of peaks	Snake	Eagle
Iteration	1	1	0
	2	33	32
	3	9	20
Phrase	1	0	0
	2	6	5
	3	1	4
	2,3	5	6

TABLE II. Analysis of variance results for two-peak phrases. (Note: Scores shown are means; see text for descriptions of each variable.)

Variable	Call type		F	p
	Snake	Eagle		
$PK1$	1600	1148	21.01	0.000[a]
$BW1$	1152	362	26.67	0.000[a]
$DB1$	2.1	9.7	42.86	0.000[a]
$PK2$	3059	2966	0.22	0.642
$BW2$	419	858	10.76	0.004[a]
$DB2$	4.5	0.8	24.62	0.000[a]
$TILT$	−0.6	−90.0	137.36	0.000[a]
$Q1$	1.6	4.2	7.32	0.014[b]
$Q2$	8.7	4.1	5.87	0.025
$PK2/PK1$	1.9	2.6	23.83	0.000[a]
$BW2/BW1$	0.4	2.6	27.47	0.000[a]

[a] $p<0.01$.
[b] $p<0.05$.

tiate two-peak iterations. Every comparison was statistically significant, with the exception of $PK2$ and $Q2$.

A discriminant analysis was performed to determine the efficacy of statistical classification that is possible on the basis of these measurements. $TILT$ alone allowed perfect classification of all 27 phrases with an F-to-enter value of 135.36. This variable greatly overshadowed all others, and the analysis was performed again without it. Variables $DB1$, $Q2$, $PK2$, and $PK1$ again allowed perfect classification.

As might be expected, the two-peak variables were strongly intercorrelated. Thirty-nine correlation coefficients computed in factor analysis of the two-peak data were statistically significant, out of a possible total of 55 binary permutations of these 11 variables. Three factors were identified, and could account for 86% of the variance. $TILT$ was significantly correlated with all other variables except $PK2$, and figured in all three factors. Factor 1 was most strongly associated with variables $Q1$, $DB1$, $BW2/BW1$, $BW1$, and $PK1$. It seemed to mainly reflect the relative definition and prominence of the first peak. Factor 2 similarly characterized the second peak, being associated mostly with $DB2$, $PK2/PK1$, and $BW2$. Factor 3 did not clearly differ from factor 2, with $Q2$ and $PK2$ receiving the heaviest loadings. Factor 1 accounted for 44% of the total variance explained by all three, whereas factors 2 and 3 accounted for 37% and 19%, respectively.

2. Changes within iterations and phrases

Visual examination of call spectra from both frame-by-frame and overall analysis of iterations revealed no consistent patterns of change in peak frequencies, bandwidths, or amplitudes. The possibility of patterned changes within phrases was tested through analysis of variance of peak-related difference scores for adjacent pairs of two-peak iterations. No significant trends were found.

3. Three-peak iterations

Analysis of variance results for three-peak data were essentially similar to those of two-peak iterations, but with proportionately fewer significantly different variables. Prototypical frequency spectra drawn from mean data appear in Fig. 3. To summarize, the first and second peaks were of higher frequency in snake phrases, with a lower amplitude and wider bandwidth in the lowest peak. The third peak was of higher amplitude, the tilt flatter, and the three peaks were more closely spaced than in eagle phrases. $TILT$ was again predominant among the variables.

C. Timing analysis

The results of the measurements on the 12 snake and 15 eagle phrases are shown in Table III. The only variable that clearly did not differ between call types was the mean number of iterations in a phrase. The mean iteration duration did

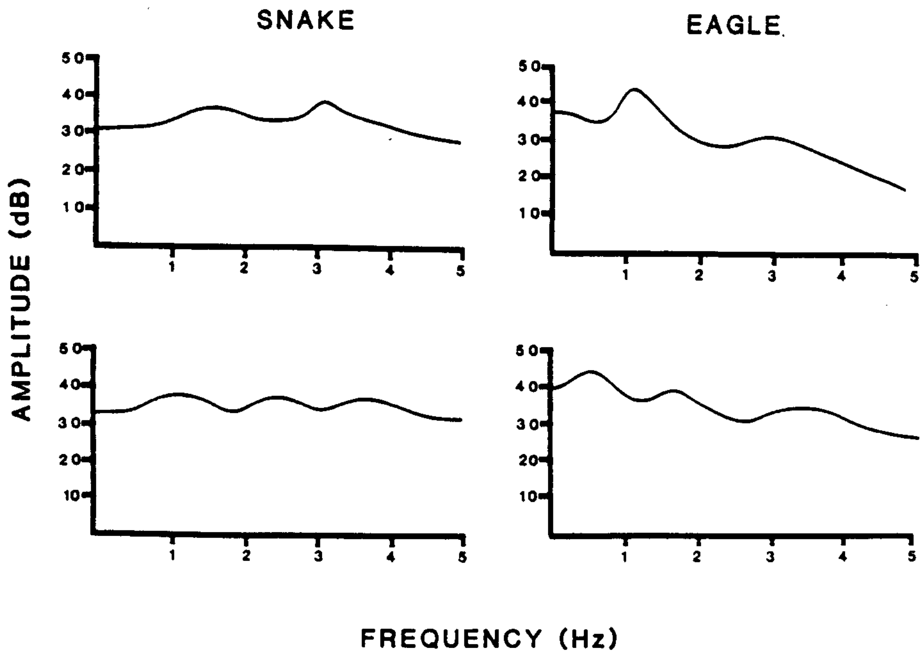

FIG. 3. Prototypical spectral envelopes for two- and three-peak iterations based on mean data.

TABLE III. Analysis of variance results for temporal measurements. (Note: Scores shown are means, with durations in milliseconds.)

Variable	Call type: Snake	Eagle	F	p
Iterations per phrase	3.6	3.5	0.07	0.794
Iteration duration	37.1	29.6	3.87	0.060
Total phrase duration	252.5	138.3	18.79	0.000[a]
Total signal per phrase	129.5	93.0	11.84	0.002[a]
Percent signal per phrase	55.5	67.8	4.90	0.036[b]
Interiteration duration	44.7	19.4	20.97	0.000[a]

[a] $p<0.01$.
[b] $p<0.05$.

not show a formally significant difference but did closely approach the 0.05 level. All other comparisons were statistically significant. At the lowest level of temporal measurement, iteration duration and spacing results showed that snake iterations are longer and more widely spaced than eagle iterations. The difference in total iteration duration confirmed the reliability of the apparent, but statistically insignificant, difference in iteration duration. Eagle calls contained proportionately more signal than snake calls. Further comparisons revealed that iterations in eagle calls were significantly longer than the interiteration intervals, $t(26) = 2.99, p < 0.05$, two tailed, whereas no difference was found between iterations and intervals in snake calls, $t(22) = -1.21, p > 0.05$, two tailed.

Both call types showed substantial variability with respect to each significant variable. Discriminant analysis correctly classified a maximum of 89% of the phrases, and did so on the basis of iteration duration and percent signal in the phrase alone. However, temporal variables were also found to be highly intercorrelated; 9 out of a total of 15 possible correlations between these 6 variables were statistically significant.

The mean amplitude ranges in the phrases were 4.3 dB for snakes and 5.7 for eagles. This difference was not statistically significant, $t(21) = -1.46, p > 0.05$, two tailed. Analysis of variance on relative amplitude measurements showed that snake and eagle calls did not differ in their overall deviation around the phrase mean, $F = 0.44, p > 0.05$. However, each did exhibit a characteristic phrase amplitude pattern when their three normalized scores were compared, $F = 18.75, p < 0.001$. Both types of phrases began at a relatively low amplitude. In snake phrases, the amplitude rose in the middle and either remained at this level or continued to rise. Eagle phrases reached peak amplitude in the middle and fell to the original level by the end. Discriminant analysis correctly classified 87% of the phrases on the basis of the third amplitude score alone.

III. DISCUSSION

A. Source function

Overall, source function evidence was equivocal. Snake calls were found not to exhibit source periodicity, but there was weak evidence of regularity in eagle calls both from the spectrographic and cepstral analyses. However, the overall spectral shape of these calls constitutes additional evidence of differences in the source function periodicity. In humans, source waveform periodicity is associated with a 12-dB per octave downward tilt (Fant, 1973). This shape is reflected in LPC modeling of vowels and is actually used as a "voicing decision" factor in the ILS-API algorithm. It has not been definitively shown that these alarm calls are laryngeally produced. However, in the absence of any indications to the contrary, this is a safe assumption. The spectrum of the eagle call is not as strongly tilted as a typical vowel spectrum, but is clearly different from the flat snake spectrum.

Nonetheless, no unequivocal conclusion can be drawn. The present cepstral analysis results were much weaker than results reported by Seyfarth and Cheney (1984b) for the similarly harsh-sounding vervet grunts. Their examination of four grunts from a single adult female revealed "unstable" periodicity in over 40% of 556 segments analyzed for 32 calls. The comparable figure for the eagle phrases was only 7%. For the purposes of perceptual testing, the possibility of periodicity in eagle calls can be tentatively hypothesized. A semiperiodic vibration mechanism not used in snake alarms may be present in these sounds. The inconsistency of the results obtained may indicate the absence of any real periodicity in eagle calls, but inconsistent results could also be expected if the tests employed were simply inadequate in capturing regularity that does exist.

B. Spectral filtering

The comparison of two- and three-peak results indicates that two-peak iterations are more distinctive. All the differences found in three-peak iterations were present in two-peak data, but not the converse. Several interpretations are possible given this pattern of results. First, three-peak iterations may have resulted primarily from specifying an excessive number of coefficients in LPC modeling. However, comparisons of FFTs and spectral envelopes in three-peak cases do not support this interpretation. Second, three-peak iterations could result from "sloppy" vocal tract placement during call production, and may actually be perceptually less distinctive to vervets. This speculation provides a testable hypothesis for future research. Finally, the spectral cues of both two- and three-peak patterns may be valid and provide adequately distinctive features.

The analysis of a number of aspects of two-peak iterations and phrases indicates that changes in spectral patterns within individual calls do not distinguish the snake and eagle call types. There were no indications of patterned changes in the spectral properties of these sounds either at the level of iterations or phrases. Comparisons of prototypical two- and three-peak spectral patterns are consistent with the evidence derived from two-peak factor analysis.

In accordance with the results of two-peak discriminant analysis, *TILT* appeared to be an independently predictive variable. It was strongly correlated with other variables, but uniformly so. The *TILT* information was rather evenly distributed across the three factors that were derived. Although it can often be difficult to interpret clustering of variables in factor analysis, in the present case, factor 1 appeared to be related to the first peak, whereas factors 2 and 3 were mainly related to the second peak. The distinctiveness of snake and eagle calls may, therefore, derive either from differences in the simple overall tilt of their respective filtering functions and resulting frequency spectra, or from more detailed aspects of these prototypical patterns. In the latter case, the pattern as a whole is more likely to be important than absolute peak locations, bandwidths, or amplitudes given the occurrence of both two- and three-peak versions.

Two formal descriptions of these hypotheses were derived and tested. (1) The distinctive feature is overall spectral tilt. Snake calls have the value *FLAT* on this feature, and eagle calls are *FALLING*. In terms of *TILT* measurements, a score greater than or equal to -50.0 is designated as being *FLAT*, and anything less is *FALLING*. Examination of all iterations, both two- and three-peak, resulted in 96.8% correct classification (90/93) using this rule. One snake iteration was misclassified as an eagle, and two eagle iterations from a single phrase showed snakelike tilt. (2) Simple spectral patterns characterize these calls. If two or three spectral peaks occur but do not differ more than 3 dB in amplitude, or one stronger peak above 1750 Hz occurs, it is a snake call. If two or three spectral peaks occur, one peak below 1750 Hz is at least 3 dB stronger than any other peak, and an identifiable peak occurs above 1750 Hz, it is an eagle call. With these rules 89.2% of the iterations were correctly classified (83/93). One snake iteration was misclassified as an eagle, one eagle iteration was misclassified as a snake, and two snake and seven eagle iterations did not conform to either pattern.

It should be noted that snake calls frequently appear to contain at least some energy above 5 kHz, which was the upper frequency limit in this analysis due to sampling rate limitations in digitization. Seyfarth *et al.* (1980a,b) reported the presence of energy up to 16 kHz in some snake call spectrograms. However, the presence of unanalyzed energy in these calls does not affect the gist of the spectral filtering hypotheses. Inspection of spectrograms of all the snake calls available gave no indication of energy patterns inconsistent with a noisy source and generally flat vocal tract filtering. Peaks above 5 kHz may be present in some calls, but, if so, only serve to further differentiate these sounds from the low-frequency eagle calls.

A more general question is whether the LPC-based quantification of spectral filtering that was employed is appropriate for investigations of nonhuman vocalizations. The LPC was adapted to speech analysis and synthesis based on a vocal tract model that takes only resonances into account (Wakita, 1976). At present, too little is known about the filtering properties of nonhuman vocal tracts to allow uncritical application of techniques like LPC modeling to primate vocalizations. In one instance, in fact, antiresonances have been shown to occur in the calls of a nonhuman primate (Haimoff, 1983). Even for speech, however, LPC is simply a best, first approximation to actual vocal tract filtering. Although vowels and vowellike phones are well suited to LPC analysis, segments like nasal consonants, unvoiced stops, fricatives, affricates, and unvoiced portions of otherwise voiced sounds are known to involve vocal tract zeros (Fant, 1973). Atal and Schroeder (1975) have suggested that the perceptible differences between actual speech and LPC-based synthetic speech are due, at least in part, to the influence of vocal tract antiresonances. Nonetheless, LPC synthesis of speech is quite successful and demonstrates "amazing fidelity" to the natural waveform (Wakita, 1976, p. 5). The expectation of some robustness to the LPC assumption of all-pole filtering seems warranted. Future investigations would, however, clearly benefit if filtering models taking both poles and zeros into account were implemented in processing packages (see Markel and Gray, 1976, for descriptions of such models).

Two further observations suggest that vocal tract zeros do not, in fact, play an important role in the particular calls under study here. First, both snake and eagle alarms are produced with the mouth open (Struhsaker, 1967, personal observation). Resonance characteristics of the vervet oral cavity alone are therefore likely to be of overriding importance in the spectral filtering that occurs. Second, the spectra derived for the analysis phrases showed no sign of antiresonances. The presence of zeros in spectral filtering often results in visible "dips" in the FFT (cf. Markel and Gray, 1976). Snake and eagle spectra, on the other hand, appear to be completely determined by frequency peaks.

Finally, the appropriateness of LPC analysis for these sounds has been confirmed in perceptual testing. Vervets and humans classifying snake and eagle alarm calls perform equally well with natural calls and LPC-based synthetic replicas (Owren, 1988).

C. Timing

The measurements made on various temporal aspects of the iterations and intervals in each phrase show that snake and eagle calls are distinguishable on the basis of energy distribution through time. Classification was more difficult from these variables than from spectral cues, but was quite successful nonetheless. It was not apparent, however, how to formally state a classification rule that could be used in differentiating either single iterations, or overall phrases.

The amplitude data were not problematical in this regard, but also did not permit completely accurate classification of all phrases. The comparison of interest was the relative amplitude of the last segment, which was scored as HIGH (greater than the mean amplitude of the phrase) or LOW (less than the mean amplitude of the phrase). This simple rule correctly sorted snake alarms as HIGH and eagle alarms as LOW for 87% of the analysis phrases.

D. Conclusions

In spite of the equivocal results of cepstral analysis, the present study reinforces Seyfarth and Cheney's (1984b) demonstration of the power of digital processing techniques in characterizing nonhuman primate sounds. Both vervet grunt calls and snake and eagle alarms are relatively inaccessible to spectrographic examination due to their noisy, broadband structure. In addition to greatly improving measurement accuracy both in the frequency and time domains, digital treatment in both cases revealed potentially distinctive cues that cannot be discerned in spectrograms of these calls. Nonetheless, our results also point up the need for studies in which source waveforms and vocal tract filtering are examined directly and in isolation.

The current results extend findings from previous studies that have attempted to establish the "phonetic contrasts" of which various primates are capable. Analyzing spectrograms, Lieberman (1975) concluded that primates can produce humanlike contrasts based, for instance, on glottal vibration, increases and decreases in the rate of glottal vibration, and simple coupling of nonoral cavities to the resonating vocal tract. Conspicuously absent from the list, though, was the ability to flexibly alter vocal tract shape to produce distinctive resonance patterns. This ability is important in speech, where formant patterning is a critical component of vowel perception (cf. Joos, 1948).

Evidence of variable spectral patterns in the calls of baboons (*Papio hamadryas*) and gelada monkeys (*Theropithecus gelada*) has been found in spectrographic analyses performed by Andrew (1976) and Richman (1976), respectively. Our findings, together with Seyfarth and Cheney's (1984b) data, raise the possibility that vervets may also routinely manipulate vocal tract resonance characteristics during call production. Follow-up perceptual testing in which spectral patterning was manipulated in synthetic snake and eagle alarm calls has provided evidence that such variation is also communicatively significant to vervets (Owren, 1988).

ACKNOWLEDGMENTS

This article is based on a portion of a dissertation submitted by Michael Owren to Indiana University in partial fulfillment of the requirements for the Ph.D. degree. We thank Robert Seyfarth and Dorothy Cheney for providing the alarm calls used in this study, and David Pisoni for making available the facilities of the Indiana University Speech Research Laboratory. The impetus for the work came from ongoing investigations in the Primate Communication Laboratory at Indiana University, under the direction of Michael Petersen. Robert Seyfarth and two anonymous reviewers provided helpful comments on an earlier draft of this article. The research was supported by National Science Foundation Grant BNS 79-24477, a Harry Frank Guggenheim Foundation Grant, and a Sloan Foundation Research Fellowship to Michael Petersen, as well as an Indiana University Doctoral Grant-in-Aid, and crucial awards from the Marie Znaiden Sinnott Memorial Fund to Michael Owren, who was a NIH Postdoctoral Trainee during the preparation of this article. Support of the instrumentation in the Speech Research Laboratory was provided, in part, by NIH Research Grant NS-12179 and, in part, by a contract with the United States Air Force Office of Scientific Research.

Andrew, R. J. (1976). "Use of formants in the grunts of baboons and other nonhuman primates," Ann. N.Y. Acad. Sci. 280, 673–693.

Atal, B. S., and Schroeder, M. R. (1975). "Recent advances in predictive coding-application to speech synthesis," in *Speech Communication (Vol. 1): Speech Wave Processing and Transmission*, edited by G. Fant (Wiley, New York), pp. 27–31.

Bernacki, B. (1981). "WAVMOD: A program to modify digital waveforms," in *Research in Speech Perception Progress Report No. 7* (Department of Psychology, Indiana Univ., Bloomington, IN), pp. 275–286.

Cheney, D. L., and Seyfarth, R. M. (1982). "How vervet monkeys perceive their grunts: Field playback experiments," Anim. Behav. 30, 739–751.

Dittus, W. P. J. (1984). "Toque macaque food calls: Semantic communication concerning food distribution in the environment," Anim. Behav. 32, 470–477.

Dixon, W. J. (Ed.) (1983). *BMDP Statistical Software* (University of California, Berkeley).

Fant, G. (1973). *Speech Sounds and Features* (MIT, Cambridge, MA).

Forshee, J. D. (1975). "Speech perception research laboratory: The state of the computer system," in *Research in Speech Perception Progress Report No. 2* (Department of Psychology, Indiana Univ., Bloomington, IN), pp. 202–220.

Forshee, J. D., and Nusbaum, H. C. (1984). "An update on computer facilities in the speech research laboratory," in *Research in Speech Perception Progress Report No. 10* (Department of Psychology, Indiana Univ., Bloomington, IN), pp. 453–462.

Gouzoules, S., Gouzoules, H., and Marler, P. (1984). "Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid," Anim. Behav. 323, 182–193.

Green, S. (1975). "Variations of vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): A field study," in *Primate Behavior: Developments in Laboratory and Field Research Vol. 4*, edited by L. A. Rosenblum (Academic, New York), pp. 1–102.

Haimoff, E. H. (1983). "Occurrence of anti-resonance in the song of the siamang (*Hylobates syndactylus*)," Am. J. Prim. 5, 249–256.

Heffner, H. E., and Heffner, R. S. (1984). "Temporal lobe lesions and perception of species-specific vocalizations by macaques," Science 226, 75–76.

Hopp, S. L. (1985). "Differential Sensitivity of Japanese monkeys (*Macaca fuscata*) to variations along a synthetic vocal continuum," unpublished doctoral dissertation, Indiana Univ., Bloomington, IN.

Joos, A. M. (1948). "Acoustic phonetics," Language Suppl. 24, 1–136.

Kewley-Port, D. K. (1979). "SPECTRUM: A program for analyzing the properties of speech," in *Research on Speech Perception Progress Report No. 5* (Department of Psychology, Indiana Univ., Bloomington, IN), pp. 475–492.

Kuhl, P. K. (1979). "Models and mechanisms in speech perception: Species comparisons provide further contributions," Brain Behav. Evol. 16, 374–408.

Lieberman, P. (1975). *On the Origins of Language* (Macmillan, New York).

Luce, P., and Carrell, T. D. (1981). "Creating and editing waveforms using WAVES," *Research on Speech Perception Progress Report No. 7* (Department of Psychology, Indiana Univ., Bloomington, IN), pp. 287–297.

Markel, J. D., and Gray, A. D., Jr. (1976). *Linear Prediction of Speech* (Springer, Berlin).

Owren, M. J. (1988). "Perception of semantic alarm calls by vervets (*Cercopithecus aethiops*) and humans," in preparation.

Petersen, M. R. (1982). "The perception of species-specific vocalizations by primates: A conceptual framework," in *Primate Communication*, edited by C. T. Snowdon, C. H. Brown, and M. R. Petersen (Cambridge U.P., New York), pp. 171–211.

Petersen, M. R., Beecher, M. D., Zoloth, S. R., Moody, D. B., and Stebbins, W. C. (1978). "Neural lateralization of species-specific vocalizations by Japanese macaques (*Macaca fuscata*)," Science 202, 324–327.

Petersen, M. R., Zoloth, S. R., Beecher, M. D., Green, S., Marler, P. R., Moody, D. B., and Stebbins, W. C. (1984). "Neural lateralization of vocalizations by Japanese macaques: Communicative significance is more important than acoustic structure," Behav. Neurosci. 98, 779–790.

Richman, B. (1976). "Some vocal distinctive features used by gelada monkeys," J. Acoust. Soc. Am. 60, 718–724.

Seyfarth, R. M., and Cheney, D. L. (1982). "How monkeys see the world: A review of recent research on East African vervet monkeys," in *Primate Communiation*, edited by C. T. Snowden, C. H. Brown, and M. R. Petersen (Cambridge U.P., New York), pp. 239–252.

Seyfarth, R. M., and Cheney, D. L. (1984a). "The natural vocalizations of non-human primates," Trends Neurosci. 7, 66–73.

Seyfarth, R. M., and Cheney, D. L. (1984b). "The acoustic features of vervet monkey grunts," J. Acoust. Soc. Am. 75, 1623–1628.

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980a). "Vervet monkey alarm calls: Semantic communication in a free-ranging primate," Anim. Behav. 28, 1070–1094.

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980b). "Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication," Science 210, 801–803.

Snowdon, C. T., and Pola, Y. V. (1978). "Interspecific and intraspecific responses to synthesized pygmy marmoset vocalizations," Anim. Behav. 26, 192–206.

Struhsaker, T. T. (1967). "Auditory communication among vervet monkeys (*Cercopithecus aethiops*)," in *Social Communication among Primates*, edited by S. A. Altman (University of Chicago, Chicago, IL), pp. 281–324.

Wakita, H. (1976). "Instrumentation for the study of speech acoustics," in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass (Academic, New York), pp. 3–40.