Characterizing noise in nonhuman vocalizations: Acoustic analysis and human perception of barks by coyotes and dogs

Tobias Riede[a)]
Institute for Theoretical Biology, Humboldt University of Berlin, D-10115 Berlin, Germany

Brian R. Mitchell[b)]
Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, California 94720

Isao Tokuda[c)]
Institute for Theoretical Biology, Humboldt University of Berlin, D-10115 Berlin, Germany

Michael J. Owren[d)]
Department of Psychology, Cornell University, Ithaca, New York 14853

(Received 16 December 2004; revised 5 April 2005; accepted 18 April 2005)

Measuring noise as a component of mammalian vocalizations is of interest because of its potential relevance to the communicative function. However, methods for characterizing and quantifying noise are less well established than methods applicable to harmonically structured aspects of signals. Using barks of coyotes and domestic dogs, we compared six acoustic measures and studied how they are related to human perception of noisiness. Measures of harmonic-to-noise-ratio (HNR), percent voicing, and shimmer were found to be the best predictors of perceptual rating by human listeners. Both acoustics and perception indicated that noisiness was similar across coyote and dog barks, but within each species there was significant variation among the individual vocalizers. The advantages and disadvantages of the various measures are discussed. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1928748]

PACS number(s): 43.80.Ka [JAS] Pages: 514–522

I. INTRODUCTION

Effective measurement and quantification of characteristics of nonhuman vocalizations is critical to understanding the functional significance of these signals, and great strides have been made in this area as bioacoustic analysis techniques have become more sophisticated (e.g., Hopp *et al.*, 1998). However, not all aspects of vocal signals are equally amenable to analysis. Harmonic components are more easily measured than nonharmonic aspects. In this paper, the latter will be referred to as "noise" or "noisiness," limiting that term specifically to irregular energy that is part of the vocalization itself rather than occurring in the background or as an artifact of the recording process. Our purpose is to compare the performance of a number of measures that can be used to characterize noisiness when applied to a representative mammalian vocalization, namely barks produced by coyotes and dogs. Outcomes from these measures are then examined in light of human ratings of noisiness in the sounds.

In mammalian vocal production, noise is most commonly produced either by irregularity in the oscillating tissue of the sound source (usually the vocal folds) or by turbulence created by constrictions or obstacles that impede airflow in the vocal tract (also known as "additive noise"). These noise components are potentially distinguishable both quantitatively (Tokuda *et al.*, 2002) and perceptually (Schneider, 2004), but they nonetheless present significant challenges in analysis. The greatest difficulty lies in characterizing the complexity of the time-series waveform of a signal and frequency-spectral properties created by noisiness, particularly from irregular vocal-fold vibration (illustrated in Fig. 1). Here, spectrograms of three coyote and three dog barks provide examples of vocalizations that are almost purely harmonic, almost purely noisy, and some combination of the two. This kind of variety and complexity has recently been explained as the expected workings of a nonlinear dynamic system, in this case with mammalian vocal folds acting as coupled oscillators whose behavior can best be modeled using the principles of chaos theory (e.g., Strogatz, 2001).

Research on vocal-fold biomechanics in several mammalian species has demonstrated that harmonically structured utterances reflect quasiperiodic, synchronized vibration of the vocal folds (domestic dogs: Solomon *et al.*, 1994; monkeys: West and Larson, 1999; Jürgens, 1994; Brown *et al.*, 2003). However, the many different kinds of irregularities that also occur reveal the classic workings of nonlin-

[a)]Present address and address correspondence to Tobias Riede, Indiana University, Medical Science Program, 315 Jordan Hall, Bloomington, Indiana 47405. Fax: +1 812 855 4436; electronic mail: tobiasriede@web.de
[b)]Present address: University of Vermont, Rubenstein School of Environment and Natural Resources, 81 Carrigan Drive, Burlington, Vermont 05405-0088. Electronic mail: Brian.Mitchell@uvm.edu
[c)]Present address: Department of Computer Science & Systems Engineering, Muroran Institute of Technology, Muroran, Hokkaido 050-8585, Japan. Electronic mail: i.tokuda@biologie.hu-berlin.de
[d)]Present address: Department of Psychology, Georgia State University, P.O. Box 5010, Atlanta, Georgia 30302-5010. Electronic mail: michaeljowren@gmail.com

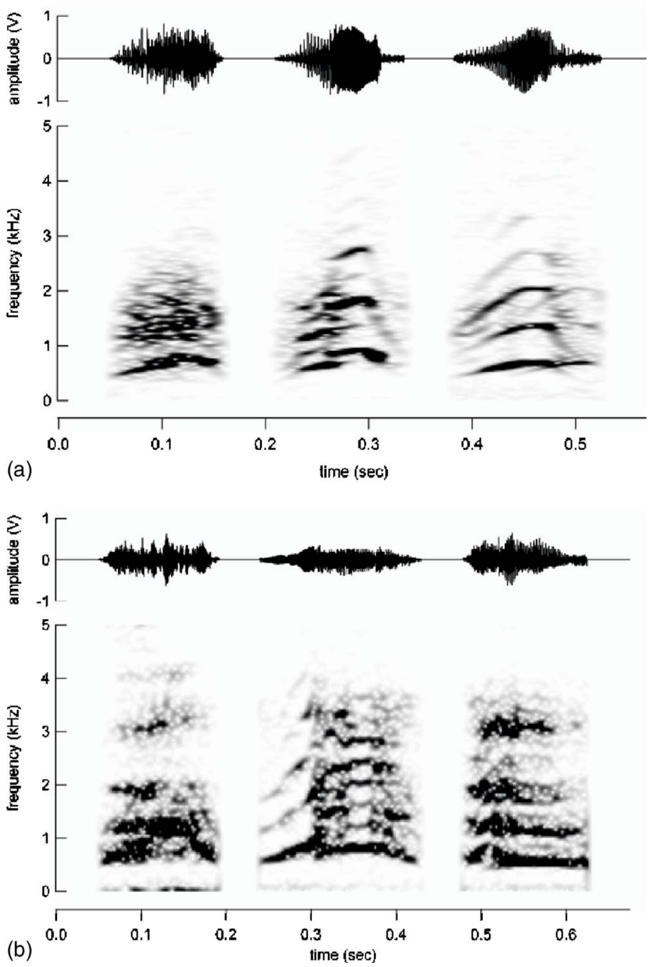

FIG. 1. Time series and spectrograms of (a) three coyote barks and (b) three dog barks. All vocalizations came from different individuals. Sounds were sampled at 22.05 kHz and spectrograms were produced using a 512-point FFT analysis window.

ear dynamic systems (Wilden *et al.*, 1998; Riede *et al.*, 2000, 2004; Fitch *et al.*, 2002; Brown, *et al.* 2003). These irregularities are now often referred to as "nonlinear phenomena," although all vocal-fold vibratory behaviors are technically nonlinear, including those that give rise to harmonically structured sounds.

Noisy mammalian sounds should therefore not be analyzed based solely on traditional methods, which include measuring the basic rate of vocal-fold vibration (known as fundamental frequency, or F_0), and the filtering effects of resonances of the vocal tract above the larynx (the formants). Here, F_0 is interesting only as a means of testing for the presence of deviations from periodicity, and formant analyses may play little, if any, role when measuring noisiness. If there are identifiable F_0 or harmonic components, one can apply algorithms designed to quantify the relative contribution of periodicity and noise, and to characterize cycle-to-cycle variation or "perturbation" in the waveform. However, some noisy vocalizations may not be truly periodic at all, instead reflecting deterministic chaos (Wilden *et al.*, 1998). An additional approach would involve measuring the degree of nonlinearity in the underlying vibrations. Each of these three approaches, namely periodicity versus noise, perturbation, and relative nonlinearity, are compared in this paper.

Periodicity and noise are typically compared using "harmonic-to-noise ratio" (HNR) algorithms, also known as "signal-to-noise ratio," "harmonicity," and the like. A number of such algorithms have been developed in speech science (see Baken and Orlikoff, 1999), where the typical approach is based on first identifying F_0-related cyclicity present in a waveform (a process referred to as "pitch extraction;" although see Klingholz, 1987). The amount of energy attributable to this periodic component is then used as the basis for inferring how much noise is present, essentially by subtracting periodic energy from total energy. The Praat acoustic analysis system (Boersma and Weeninck 1996, available as freeware from www.praat.org), includes a particularly robust version of this kind of algorithm, which we will refer to as Praat-HNR (Boersma, 1993).

HNR measurement has not been used much in the study of nonhuman vocalizations, although a number of researchers have suggested doing so (e.g., Owren and Linker, 1995; Schrader and Hammerschmidt, 1997; Riede *et al.*, 2001; Darden *et al.*, 2002). One practical impediment has been that pitch-extraction algorithms do not necessarily work well for nonhuman vocalizations, especially when there is only a weak periodic component. Riede *et al.* (2001) therefore developed an alternative that we will refer to as Spectral-HNR. This is a simpler method that operates solely in the frequency domain and is therefore not dependent on pitch extraction. Instead, the approach is to compute a Fourier spectrum of the sound segment of interest, smooth away the harmonic peaks in the spectrum, and estimate HNR as the difference between harmonic peaks in the original spectrum and the average amplitude of the smoothed spectrum. As a result, there are parameter settings for the spectrum computation and the amount of smoothing. The particular settings chosen can have a significant impact on the final outcomes (e.g., Beecher, 1988).

Measuring vocal perturbation, a second technique for quantifying noisiness, has also been more widely used in speech science than in bioacoustics (although see Owren and Linker, 1995). In speech, much of the interest has derived from the fact that irregularity in otherwise synchronized vocal-fold vibration can be a clinical marker of dysfunction (e.g., Kent and Ball, 1999). As a result, measuring irregularity can be an important diagnostic tool in identifying and treating vocal-fold pathologies affecting speech production. "Jitter" is the most widely used perturbation measure, and is defined as cycle-to-cycle F_0 variability. The irregularity in question is variability in the period (or frequency) of successive opening and closings of the vocal folds (referred to as glottal pulses). "Shimmer" is similar, referring to cycle-to-cycle variability in the peak waveform amplitude. This measure is also well known and routinely implemented in quantifying perturbation (e.g., Jiang *et al.*, 2000; Dejonckere *et al.*, 2001).

Nonlinearity has not routinely been measured in either human or nonhuman vocal work, although it is becoming increasingly clear that vocal-fold vibration is best modeled as a nonlinear dynamic process (Sciamarella and d'Alessandro, 2004). Tokuda, *et al.* (2002) have therefore recently developed a technique referred to as the "nonlinear measure" (NLM), which is designed to provide an overall estimate of the strength of nonlinearity in a signal. It does so by comparing the residual errors produced when modeling the signal waveform using deterministic nonlinear versus stochastic linear methods. The rationale is that if signal noise reflects low-dimensional chaos as produced through nonlinear processes, deterministic-nonlinear modeling will produce a relatively small error component. A stochastic linear model, on the other hand, will produce a relatively large error component when applied to such signals. The converse is expected if the noisiness primarily reflects a stochastic process. While thus technically not a measure of noisiness *per se*, we included NLM because high-energy noise in mammalian vocalizations is likely to reflect chaotic vocal-fold vibration. Using the NLM does not require the presence of periodicity, but it can nonetheless be applied to such sounds. As a result, it potentially has a broad applicability to nonhuman vocalizations.

In testing these measures, we first focused on coyote barks as an arguably representative example of a noisy, natural vocalization produced by a wild mammal (illustrated in Fig. 1). We then secondarily selected dachshund dog barks as sounds that are similar to the coyote vocalizations and are familiar to human listeners. The latter consideration was potentially important to comparing acoustic measurement outcomes to human perceptual ratings of noisiness. Investigations of human responses to domestic animal vocalizations has suggested that listeners reliably perceive relatively small differences in vocal noisiness (Feddersen-Petersen, 2000; Nicastro and Owren, 2003; Yin, 2002; Yin and McCowan, 2004).

II. METHODS

A. Study animals

The vocalization sample included barks from five coyotes (*Canis latrans*) and five dachshunds (*Canis familiaris*), with each species represented by three males and two females. Coyotes were recorded at the United States Department of Agriculture's National Wildlife Research Center (NWRC) field station in Logan, Utah, between July 8 and July 27, 1998. These animals were housed in outdoor 0.1-ha pens, and were between 2 and 5 years old. Their weights ranged from 8.4 to 15.0 kg. A Tascam DA-P1 digital tape recorder (DAT) was used with a Sennheiser MKH-70 shotgun microphone at a recording height of 1.0 m. The recording distance varied between 5 and 20 m. The dachshunds were dogs that had been brought to the Small Animal Veterinary Clinic of the Free University of Berlin, Germany for routine examination and were subsequently recorded at their owners' homes. All had been found to be free of disease and any clinical peculiarities, and ranged in age from 9 months to 11 years and in weight from 6.8 to 10.0 kg. These recordings were made with a Marantz PMD 222 tape recorder and a Sennheiser microphone (ME80 head with K3U power module) on BASF ChromeSuper II tapes. The distance between dog and microphone varied between 0.5 and 1.5 m.

B. Acoustic analysis

We analyzed ten barks from each of the ten animals. Coyote barks were selected as those having the least extraneous background noise among all the recordings available for a particular individual. For the dachshunds, we used the first ten barks recorded in that individual's session, while avoiding those contaminated by extraneous sounds or distorted through clipping during audio recording. Coyote vocalizations were originally recorded digitally at 48.0 kHz and downsampled to 25.0 kHz, while dog barks were recorded on analog tape and then digitized at 44.1 kHz. All recordings were downsampled to 22.05 kHz for use in the present investigation.

All barks were analyzed using two HNR methods (Spectral-HNR and Praat-HNR), the percentage of voiced frames identified through pitch extraction (%VF), Koike's (1973) "relative average perturbation" method of computing jitter (Jitter), shimmer computed from adjacent cycles (Shimmer), and the NLM measure. Measures were computed using Praat software (www.praat.org), with the exception of Spectral-HNR and NLM. Spectral-HNR was computed as described by Riede *et al.* (2001), using a routine written and implemented in Signal 4.0 (Engineering Design, www.engdes.com). Briefly, we computed an FFT transform from the middle of each bark and estimated noise level by passing a moving-average filter across the spectrum. The Spectral-HNR value was then set as the maximum amplitude difference between the original spectrum and the moving average spectrum, in dB. The computation involved 3 adjustable parameters: duration of the vocalization segment (segment duration, extracted from the middle of the vocalization), width of the Fourier window (FFT size), and number of points in the moving average (smoothing factor). We evaluated 27 parameter combinations by systematically varying segment duration (25 ms, 50 ms, or 75 ms), FFT size (256, 512, or 1024 points), and smoothing factor (5, 10, or 15 points). However, because a 1024-point FFT is longer than 25 ms for 22.05 kHz sounds, the 3 smoothing widths were not evaluated for the 25 ms segment and 1024-point FFT. NLM was computed using the method described by Tokuda *et al.* (2002), using a custom-written program, which is available at http://itb.biologie.hu-berlin.de/~tokuda/NLM/, implemented on a Linux workstation (Red Hat Linux 7.3, Kernel 2.4.18-3 on an i686, gcc version 2.96).

C. Perceptual rating

Perceptual testing was conducted in a room with five booths equipped with Beyerdynamic DT 831 headphones and Tucker-Davis Technologies (TDT) response boxes. The booths were operated from an adjacent room using TDT modules, a computer, and custom software (B. Tice & T. Carrell, http://hush.unl.edu/LabResources.html). Stimuli were prepared with Praat.

Participants were 21 female and 21 male Cornell University undergraduates, who rated either all the coyote or all

TABLE I. Variables and perceptual ratings for coyotes and dogs (mean of 10 barks per individual ± SD). ANOVA comparisons were made within species based on 10 bark vocalizations from each of 5 individuals, with two-sample t tests used for between-species comparisons based on mean values computed separately by an individual. Outcomes for two settings of the Spectral-HNR (segment duration—FFT size—smoothing factor) are given. Significant values are marked with an asterisk ($p<0.01$).

	Sex/age	Mean perceptual rating	Spectral-HNR 75-256-10	Spectral-HNR 75-512-10	Praat-HNR	%VF	Jitter	Shimmer	NLM
Coyote 1	M/adult	2.53±0.33	8.7±2.1	9.4±1.5	17±10.7	29±21	0.07±0.2	0.32±0.08	0.11±0.07
Coyote 2	M/adult	3.47±0.33	16.6±3.3	13.7±2.9	26±6.9	75±16	0.1±0.2	0.19±0.1	0.11±0.08
Coyote 3	M/adult	4.07±0.33	15.1±3.4	13.5±2.9	23±6.9	73±17	0.03±0.1	0.15±0.09	0.28±0.26
Coyote 4	F/adult	4.33±0.33	18.5±5.2	17.1±4.9	21±6.7	82±16	0.04±0.003	0.13±0.08	0.33±0.2
Coyote 5	F/adult	4.21±0.33	19.9±2.1	15.5±3.5	24±7.5	85±12	0.07±0.1	0.15±0.05	0.27±0.11
All coyotes		3.6±0.9	15.7±5.1	13.8±4.1	7.0±3.9	69.1±25	0.05±0.1	0.18±0.1	0.22±0.18
F values		12.5*	16.3*	7.4*	20.9*	18.5*	0.6	7.1*	3.7
Dog 1	M/9 mo	3.31±0.4	12.4±2.4	10.1±1.0	26±9.9	53±19	0.1±0.2	0.29±0.06	0.06±0.04
Dog 2	M/11 yr	3.6±0.4	15.3±1.3	13.6±2.6	26±6.6	63±16	0.06±0.1	0.17±0.07	0.05±0.02
Dog 3	M/3 yr	3.86±0.4	17.0±3.8	20.0±4.3	27±6.9	88±10	0.15±0.1	0.2±0.07	0.29±0.38
Dog 4	F/5 yr	3.97±0.4	15.8±2.8	16.6±4.7	27±6.9	84±11	0.008±0.003	0.21±0.06	0.19±0.17
Dog 5	F/9 yr	3.08±0.4	16.9±3.5	15.4±4.4	22±9.8	95±10	0.001±0.0004	0.22±0.04	0.05±0.05
All dogs		3.7±0.5	15.4±3.3	15.1±4.8	5.8±1.9	77±21	0.07±0.1	0.22±0.07	0.13±0.2
F values		5.3*	4.1*	9.9*	4.8*	16.4*	2.6	4.2*	3.2
Btwn spp t values		0.43	0.13	−0.62	−1.9	−0.59	−0.06	−0.75	1.4

the dog barks. The coyote barks were rated by 12 female and 9 male students, while the dachshund barks were rated by 9 female and 12 male students. Participants provided informed consent, were rewarded with course credit, and reported themselves to be free of speech or hearing impairments. Their instructions included the information that the stimuli being presented had been recorded from either coyotes or dogs, and that they should rate the relative noisiness versus tonality of each sound using a seven-point scale. That scale was defined through labels on the response box. Buttons were arrayed in a line on the box, with the first, third, fifth, and seventh buttons labeled "very noisy," "noisy," "tonal," or "very tonal," respectively. We counterbalanced for label position by reversing the order of this scale on alternate days, and later rescored the scaled responses using 1 for most noisy and 7 for most tonal.

III. RESULTS

A. Acoustic measures

Species- and individual-level comparisons. None of the acoustic measures revealed significant differences between species (Table I). In contrast, only Jitter and NLM failed to reveal significant differences among individuals within species.

1. Spectral-HNR

For simplicity, Tables I, III, and IV present only two representative outcomes for Spectral-HNR, drawn from the 27 total combinations of parameter settings that were tested. Parameter settings did in fact have a major impact on mean measurement values, as illustrated both in Table II and Fig. 2. The smoothing factor had the largest effect. With a smoothing factor of 5, for example, the overall means were 9.3 and 10.3 dB for coyotes and dogs, respectively. In contrast, a smoothing factor of 15 points produced means of 21.3 and 21.1 dB (differences of 12.0 and 10.8 dB). The segment duration had a smaller effect, the 75 ms segment produced overall Spectral-HNR means of 13.9 and 15.1 dB for coyotes and dogs, while 25 ms duration yielded corresponding means of 18.3 and 18.7 dB (differences of only 4.4 and 3.6 dB). Finally, the FFT size showed the least influence; a setting of 1024 points produced mean Spectral-HNR values

TABLE II. Mean Spectral-HNR values for coyotes and dogs. ANOVA results are shown for comparisons among individuals within species and between species (for each of 9 parameter combinations). In each case, the target setting is paired with medium values for the two other parameters (i.e., segment duration=50 ms, FFT size=512, Smoothing Factor=10). Significant F values are marked with an asterisk ($p<0.05$).

		Spectral-HNR coyotes	F values among coyotes	Spectral-HNR dogs	F values among dogs	F values between spp
Segment duration	25	18.3±5.3	4.6*	18.7±5.3	6.0*	0.1
	50	16.3±6.4	8.3*	16.8±5.1	9.1*	0.2
	75	13.9±4.1	7.4*	15.1±4.8	9.8*	1.7
FET size	256	17.3±5.0	12.2*	16.2±3.3	4.8*	1.7
	512	16.2±5.3	8.3*	16.7±5.1	9.1*	0.2
	1024	15.7±3.4	3.3*	16.8±2.8	5.1*	3.2
Smoothing factor	5	9.3±2.1	5.1*	10.3±1.6	3.3*	8.3*
	10	16.3±5.3	8.3*	16.8±5.1	9.2*	0.2
	15	21.3±5.7	12.5*	21.1±5.6	12.4*	0.7

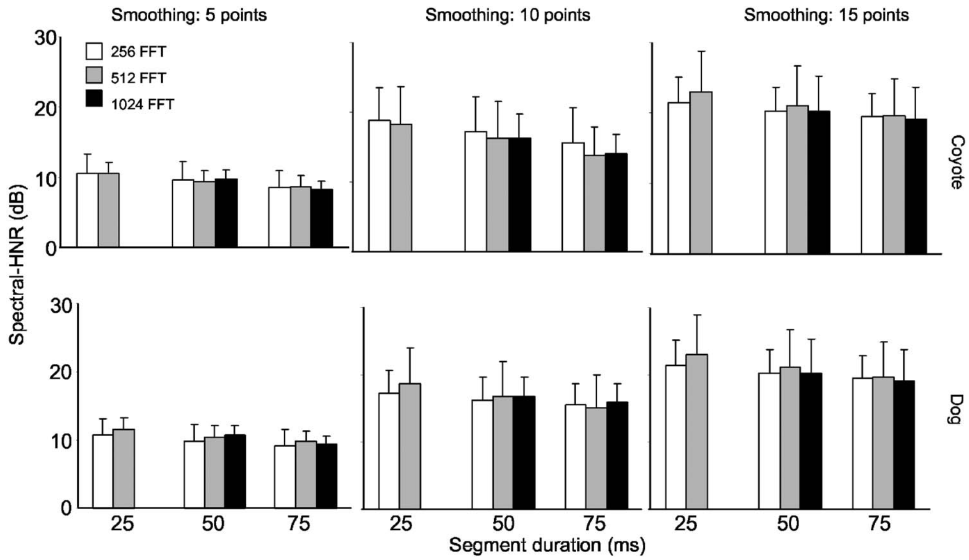

FIG. 2. Means and standard deviations of the Spectral-HNR measurements for different settings of segment duration, FFT size, and smoothing factor.

of 15.7 dB and 16.8 dB for coyotes and dogs, respectively. At 256 points, those values became 17.3 and 16.2 dB (differences of 1.6 and 0.6 dB).

Despite the effects of parameter settings on the absolute value of the Spectral-HNR measurements, ANOVA comparisons confirmed that differences among individuals in both species were significant across a subset of nine combinations of Spectral-HNR settings (Table II). In contrast, differences between species were significant for only one of the nine combinations. Spectral-HNR values obtained using different parameter settings were routinely significantly correlated (Spearman *r* values, $p < 0.01$), although four combinations of parameter settings often failed to correlate with other settings.

In both species, these included combinations of 50-1024-5 (Segment duration-FFT size-smoothing factor) and 75-1024-5. In addition, combination 75-1024-10 produced nonsignificant correlations among coyote barks, while 75-512-5 yielded nonsignificant correlations among dog barks. For comparison to other measures, we chose to specifically focus on Spectral-HNR values obtained from two parameter combinations: 75-256-10 and 75-512-10.

2. Praat-HNR

Praat-HNR also requires parameter selection related to pitch extraction, specifically setting an analysis window length and an expected range of F_0 values. These settings are chosen based on the particular kind of signal being analyzed, and it is straightforward to make the necessary adjustments. It is therefore less interesting to examine Praat-HNR across a range of its potential settings. We set the values at levels that produced the most reliable pitch extraction, with resulting HNR means and standard deviations shown in Table I. The values were significantly correlated with outcomes derived using Spectral-HNR (Table III).

3. %VF, jitter, shimmer, and NLM

Mean and standard deviations for percent-voiced frames, vocal perturbation, and nonlinearity are also summarized in Table I. Of these four measures, jitter clearly stood out as being uncorrelated with other acoustic measures (Table III). In contrast, although %VF results varied dramatically among the barks of individual vocalizers, its overall values were significantly correlated with those of other measures. Vocal perturbation measured as shimmer was less variable, and was also strongly correlated with other outcomes. Finally, NLM showed intermediate variability and strong correlations.

B. Perceptual ratings of noisiness

1. Species- and individual-level comparisons

The overall noisiness of barks was rated as 3.6 ± 0.7 by the 42 participants tested. This outcome was modestly, but

TABLE III. Matrix of Pearson correlations among acoustic outcomes. Values for coyote barks ($n=50$) are on the left, while values for dog barks ($n=50$) are on the right. Significant correlations ($p < 0.05$) are marked with an asterisk.

	Spectral-HNR (75-256-10)	Spectral-HNR (75-512-10)	Praat-HNR	%VF	Jitter	Shimmer
Spectral-HNR (75-512-10)	0.79*/0.74*					
Praat-HNR	0.85*/0.49*	0.71*/0.70*				
%VF	0.65*/0.41*	0.42*/0.50*	0.80*/0.43*			
Jitter	−0.12/0.03	0.06/−0.13	−0.14/0.02	−0.18/−0.13		
Shimmer	−0.64*/−0.43*	−0.65*/−0.60*	−0.69*/−0.55*	−0.48*/−0.07	0.05/0.14	
NLM	0.35*/0.44*	0.47*/0.60*	0.52*/0.54*	0.32*/0.22	−0.19/−0.08	−0.54*/−0.33*

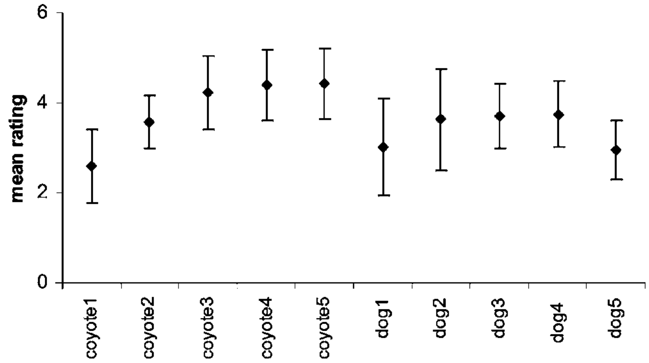

FIG. 3. Mean perceptual ratings (and standard deviations) for ten barks from each of five dogs and five coyotes. Two groups of 21 listeners rated the barks from each species using a 7-point scale that ranged from "very noisy" to "very tonal."

significantly, different from neutral, which in this scoring system would have been 4.0 ($t=-4.7, p<0.05$). Mean ratings of coyote and dog barks were 3.6±0.9 and 3.7±0.5, respectively, which was not different across species (see Table I and Fig. 3). However, ratings did differ significantly among individuals within each species, as was the case for most of the acoustic measures (Table I and Fig. 4). While the coyote sample yielded a broader range of mean ratings, mean coefficients of variation within individuals were higher in dogs (9.46% and 11.37% in coyotes and dogs, respectively).

2. HNR measures

Mean perceptual ratings were significantly correlated with Spectral-HNR (Table IV). For coyotes, the highest correlation ($r_c=0.69$) occurred using the smallest FFT size, the longest segment duration, and the intermediate smoothing factor (Fig. 4). For dogs, the highest correlation ($r_d=0.63$) occurred with intermediate FFT size, the longest segment duration, and intermediate smoothing factor (Fig. 4). However, as shown in the figure, the strength of the correlation was quite variable. Praat-HNR was also significantly correlated with perceptual ratings in both species, at levels approaching the best outcomes for Spectral-HNR ($r_c=0.63$, $r_d=0.52$).

3. %VF, jitter, shimmer, and NLM

Outcomes for the other noisiness measures generally paralleled those reported above. Shimmer and NLM showed significant correlations with human perceptual responses, and %VF did so for coyote barks, but not for dogs (Table IV). Jitter was the exception, failing to show a correlation for either species.

IV. DISCUSSION

A. Acoustic measures

1. Spectral-HNR

Spectral-HNR measures the difference between the amplitude of the harmonic peaks and the noise floor in the spectrum. We found that it performed best using a relatively long segment duration (75 ms), an intermediate smoothing factor (10 points), and a medium-to-short FFT size (coyotes: 256 points; dogs: 512 points). This outcome is reasonable based on understanding barks as a combination of harmonics and noise, where the harmonic frequencies in particular are likely to be nonstationary (meaning that they change through the course of the call). As a result, using longer segment durations for these sounds tends to produce a spectrum whose

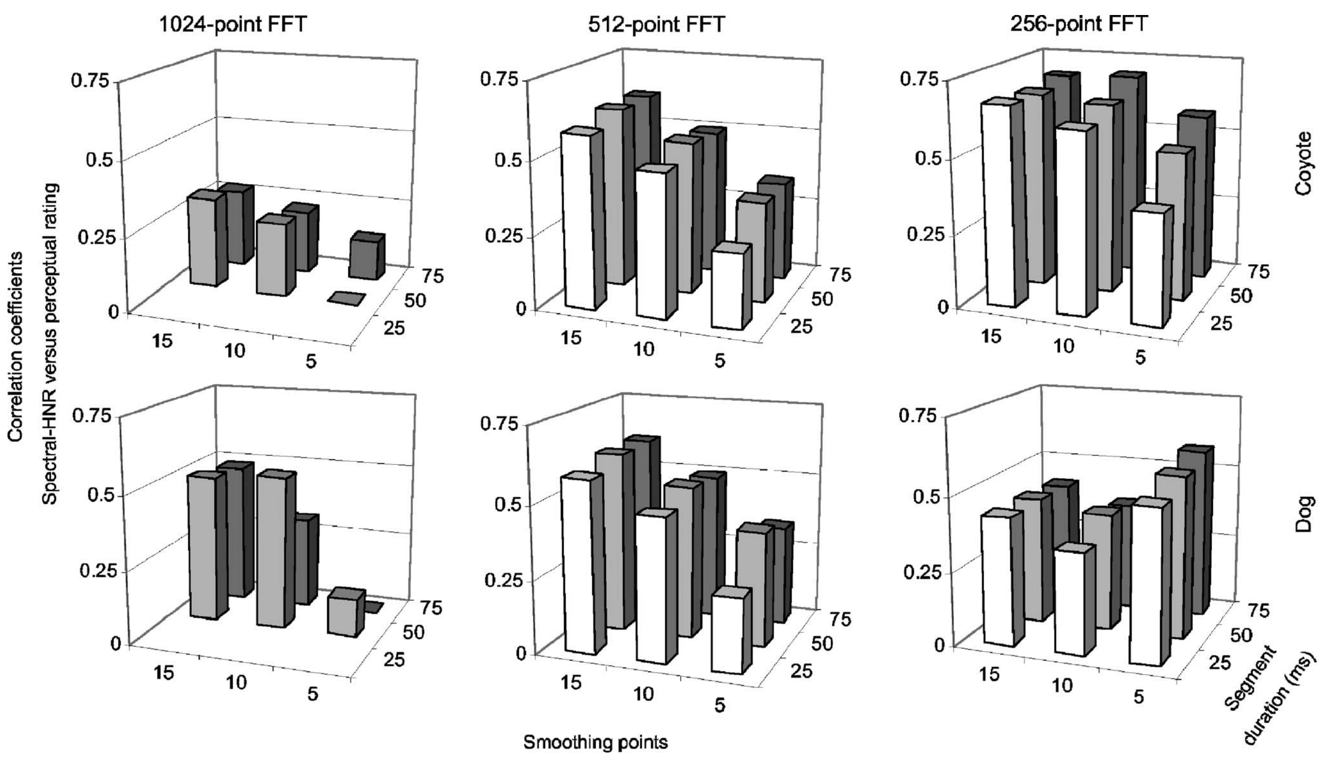

FIG. 4. Correlations between Spectral-HNR and human perceptual ratings, broken down by species and parameter settings.

TABLE IV. Regression analyses of perceptual ratings (dependent variable) against the various acoustic measures (independent variables). Significant outcomes are marked with an asterisk ($p < 0.01$).

Independent variable	r^2		Slope		F value	
	Coyotes	Dogs	Coyotes	Dogs	Coyotes	Dogs
Spectral-HNR (75-256-10)	0.47	0.14	0.12	0.06	42.8*	7.9*
Spectral-HNR (75-512-10)	0.25	0.39	0.11	0.06	16.2*	31.4*
Praat-HNR	0.63	0.52	0.18	0.18	82.8*	53.5*
%VF	0.44	0.006	0.02	0.002	37.8*	0.27
Jitter	0.03	0.02	−1.13	−0.45	1.6	1.0
Shimmer	0.39	0.34	−5.39	−3.93	29.8*	25.6*
NLM	0.41	0.37	3.17	1.46	33.8*	28.4*

harmonic frequencies have been smeared through averaging. The net effect is to broaden the bandwidth of each harmonic, thereby also lowering peak amplitudes. Using longer segment durations thus tended to lower Spectral-HNR values, as illustrated in Table II. In contrast, the smoothing factor had its largest effect on the estimated noise level. In this case, higher values produced smoother and hence lower-amplitude spectra, which increased the differentiation from harmonic peaks. Data in Table II show that higher smoothing factors did, in fact, produce higher Spectral-HNR values.

We suggest that when using Spectral-HNR, the best approach is to choose intermediate values for segment duration and the smoothing factor, and to mainly focus on the FFT size. This parameter has the most complex effects, with optimal FFT size being heavily dependent on signal characteristics. For example, because lower F_0 values are associated with more closely spaced harmonics, FFTs should probably be longer in order to achieve higher-frequency resolution (Beecher, 1988). On the other hand, if the harmonics are nonstationary, the positive effects of increasing frequency resolution can be offset by inaccuracy due to averaging effects. The upshot is therefore that some preliminary work be done in order to optimize analysis parameters. We suggest selecting fixed values for both the segment duration and the smoothing factor, but conducting the analyses with two or more different values for FFT size. Resulting outcomes may be somewhat different, and potentially more revealing at one setting than another.

2. Comparing the HNR measures

The HNR measure implemented in Praat works by extracting pitch periods from short segments of the waveform through cross-correlation, then using the peak amplitude of the resulting cross-correlation function to estimate harmonicity (Boersma, 1993). When pitch extraction is reliable, Praat-HNR appears to provide a robust and straightforward measure of noisiness. Our evidence suggests that it performed well even when applied to these noisy and variable barks, producing values that correlated strongly with human noisiness ratings. These outcomes were better than those associated with many of the results for Spectral-HNR, but were not quite equal to the best performance of that algorithm. Praat-HNR is likely to perform well as a measure of noisiness for a variety of signals, but will work best for sounds with a clear harmonic structure. Spectral-HNR can be applied to any signal, regardless of the relative prominence of its harmonic peaks, but it requires more careful parameter tuning for best performance.

3. %VF, jitter, shimmer, and NLM

Quantifying the percentage of voiced frames through pitch extraction is a simple way of gauging signal noisiness, and one we found to be highly successful. This confirms findings from earlier studies on the vervet monkey (*Cercopithecus aethiops*) vocalization (Seyfarth and Cheney, 1984; Owren and Bernacki, 1988). %VF was positively correlated with perceptual rating and results from both HNR methods, and negatively correlated with shimmer. Each of these outcomes was appropriate for a measure of noisiness. In contrast, we did not find jitter to be useful. This is not necessarily surprising, as jitter is not meant to measure noisiness *per se*, but rather to characterize perturbation in largely periodic signals (Lieberman, 1963). Shimmer is also designed to quantify vocal perturbation, but may also have value as a measure of noisiness. This metric has, for instance, been applied to diagnosing pathological roughness in human voices (Baken and Orlikoff, 1999). We found shimmer to show strong negative correlations with perceptual rating values, both HNR measures, and %VF.

NLM results were positively correlated with perceptual ratings and HNR measurements, and negatively correlated with shimmer values—in other words, the opposite of what would be expected for a noisiness measure. However, such outcomes are understandable in that the NLM responds to all nonlinear phenomena, including not only the broadband noise of deterministic chaos, but also harmonic structures that reflect limit-cycle attractors (Tokuda *et al.*, 2002). In the current work, the pattern of correlations thus strongly suggests an influence of harmonic sound elements. The NLM is also not responsive to either high-dimensional or stochastic noise, with the latter here meaning any noisiness resulting from airflow turbulence. As a result, we suggest that NLM is likely to be most useful when specifically testing for the presence of low-dimensional chaos in noisy sounds, but less so when quantifying the overall degree of signal noisiness.

B. Perceptual ratings and species versus individual effects

Rating tests revealed that Spectral-HNR, Praat-HNR, %VF, shimmer, and NLM were all significantly correlated with the human perception of noisiness in coyote and dog barks, which provides general confirmation that humans are sensitive to such noise. In addition, both acoustic and perceptual findings suggest comparable noise levels in coyote and dog barks, although we may have biased the outcome by selecting a breed roughly matched to coyotes in overall size. There is nonetheless little reason to believe that size plays a critical role in vocal noise, or that dachshunds are more like coyotes in other physical features than are other dog breeds. Listeners were also found to be sensitive to differences among individual vocalizers within each species, and produced a larger range of differences for the coyotes than for the more inbred domestic dogs. We therefore suggest that these data provide preliminary evidence of general similarities among canid bark vocalizations, at least as far as noisiness is concerned.

The results are also indicative of individual variation in the noisiness of barks, with the caveat that the sample included both males and females within each species. Both objective acoustic measurement and subjective perceptual evaluation showed significant, vocalizer-specific variation in noisiness levels. As there was not a perfect correspondence in measurement outcomes and perceptual ratings when tabulated on a vocalizer-by-vocalizer basis, we cannot conclude that the particular acoustic measures tested here are optimal for capturing all perceptually salient nuances of noisiness, either in these barks or in other vocalizations. However, the overall similarity in outcomes is of interest nonetheless, particularly in light of experiments like those of Yin and McCowan (2004). They reported that human participants were able to discriminate among barks produced by dogs that were experiencing differentiated social/behavioral contexts. Our results suggest that noisiness in particular is likely to be playing a role in dog–human communication.

C. Conclusions

We believe that jitter is not suited to measuring noisiness in vocalizations, while remaining a useful measure of vocal perturbation. NLM outcomes were also of limited value, at least as far as characterizing the amount of noise is concerned. This outcome is understandable, as NLM is by design specifically responsive to all nonlinear phenomena that can occur in a vocalization, whether or not those events create perceptual noisiness. At this point, the best use of NLM appears to be in distinguishing chaos from turbulence as a contributing factor in noise generation.

Measures of HNR, %VF, and shimmer can all be valuable metrics for estimating noisiness in bark vocalizations. The two HNR approaches tested were the most successful, and have the advantage of providing a graded value for any given point in a signal. HNR can therefore be used not only for an overall characterization of the sound, but also to map noisiness throughout the course of the vocalization. %VF can also be useful, and is certainly simple in requiring only that pitch extraction can be performed. On the other hand, using successful pitch extraction as the criterion of noisiness means that each segment of the sound can only be scored in a one–zero fashion, resulting in a rating that applies only to the sound as a whole rather than any particular point within it. In contrast, both HNR measures can be used to return detailed information about any given point in a sound, and can therefore be used to map relative noisiness along its entire length. Shimmer also proved to be of value, in this case beyond its origins as a vocal perturbation measure. Like HNR, it quantifies noisiness in a continuous fashion over short segments of sound, and can be used to trace changes across a signal.

Significant correlations among these measures indicate that they were capturing overlapping attributes of the noise in the barks we tested. However, these correlations ranged from as low as 0.35 to as high as 0.85, which we take to be indicative of the additive value in accounting for noise-related variance. Our best advice at this point is therefore to make use of each technique, screening out any measurements that prove to be redundant after the fact. If one were to select only one approach, HNR appears to be the best single measure, taking into account the tradeoffs between a pitch-extraction-based and a spectrally based method. However, if pitch extraction can be performed, there appears to be little reason not to include %VF and shimmer in the characterization.

ACKNOWLEDGMENTS

BRM was supported by the United States Department of Agriculture's National Wildlife Research Center through a cooperative agreement with the University of California, Berkeley (12-03-7405-0235 CA). IT was supported by a research fellowship from the Alexander-von-Humboldt Foundation. MJO received support from NIMH award 1 R01 MH65317-01A2 Subcontract 8402-15235-X. TR thanks Hanspeter Herzel (Berlin) for support.

Baken, R. J., and Orlikoff, R. F. (**1999**). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular Publishing, San Diego).

Beecher, M. D. (**1988**). "Spectrographic analysis of animal vocalization: Implications of the uncertainty principle," Bioacoustics 1, 187–208.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proc. Inst. Phonetic Sci. of the University of Amsterdam **17**, 97–110.

Boersma, P., and Weenink, D. (**1996**). "Praat: A system for doing phonetics by computer," Technical Report 132, Institute of Phonetic Sciences of the University of Amsterdam.

Brown, C., Alipour, F., Berry, D. A., and Montequin, D. (**2003**). "Laryngeal biomechanics and vocal communication in the squirrel monkey (*Saimiri boliviensis*)," J. Acoust. Soc. Am. **113**, 2114–2126.

Darden, S. K., Pedersen, S. B., and Dabelsteen, T. (**2002**). "Methods of frequency analysis of a complex mammalian vocalization," Bioacoustics 13, 247–263.

Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., Van De Heyning, P., Remacle, M., and Woisard, V. (**2001**). "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques: Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS)," Eur. Arch. Otorhinolaryngol. **258**, 77–82.

Feddersen-Petersen, D. (**2000**). "Vocalization of European wolves (*Canis lupus lupus L.*) and various dog breeds (*Canis lupus f. fam.*)," Arch. Anim. Breed **43**, 387–397.

Fitch, W. T., Neubauer, J., and Herzel, H. (**2002**). "Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal communication," Anim. Behav. **63**, 407–418.

Hopp, S. L., Owren, M. J., and Evans, C. (**1998**). *Animal Acoustic Research: Sound Analysis and Research Methods* (Springer-Verlag, Heidelberg).

Jiang, J., Lin, E., Wu, J., Gener, C., and Hanson, D. G. (**2000**). "Effects of simulated source of tremor on acoustic and airflow voice measures," J. Voice **14**, 47–57.

Jürgens, U. (**1994**). "The role of the periaqueductal grey in vocal behavior," Behav. Brain Res. **62**, 107–117.

Kent, R. D., and Ball, M. J. (**1999**). *Voice Quality Measurement* (Singular Publishing, San Diego).

Klingholz, F. (**1987**). "The measurement of signal-to-noise ratio (SNR) in continuous speech," Speech Commun. **6**, 15–26.

Koike, Y. (**1973**). "Application of some acoustic measures for evaluation of laryngeal dysfunction," J. Acoust. Soc. Am. **45**, 839–844.

Lieberman, P. (**1963**). "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges," J. Acoust. Soc. Am. **35**, 344–353.

Nicastro, N., and Owren, M. J. (**2003**). "Classification of domestic cat (*Felis catus*) vocalizations by naive and experienced human listeners," J. Comp. Psychol. **117**, 44–52.

Owren, M. J., and Bernacki, R. H. (**1988**). "The acoustic features of vervet monkey alarm calls," J. Acoust. Soc. Am. **83**, 1927–1935.

Owren, M. J., and Linker, C. (**1995**). "Some analysis methods that may be useful to acoustic primatologists," in *Current Topics in Primate Vocal Communication*, edited by E. Zimmermann, J. D. Newman, and U. Jürgens (Plenum Press, New York), pp. 1–27.

Riede, T., Arcadi, A. C., and Owren, M. J. (**2004**). "Nonlinear acoustics in the pant hoot vocalizations of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos," Am. J. Primatol **64**, 277–291.

Riede, T., Herzel, H., Hammerschmidt, K., Brunnberg, L., and Tembrock, G. (**2001**). "The harmonic-to-noise-ratio applied to dog barks," J. Acoust. Soc. Am. **110**, 2191–2197.

Riede, T., Herzel, H., Mehwald, D., Seidner, W., Trumler, E., Böhme, G., and Tembrock, G. (**2000**). "Nonlinear phenomena in the natural howling of a dog-wolf mix," J. Acoust. Soc. Am. **108**, 1435–1442.

Schneider, G. T. (2004). "Magnitude estimation of chaotic and white noises: Listener evaluations of loudness and annoyingness," unpublished Baccalaureate thesis, Cornell University.

Schrader, L., and Hammerschmidt, and K. (**1997**). "Computer-aided analysis of acoustic parameters in animal vocalizations: a multiparametric approach," Bioacoustics 7, 247–265.

Sciamarella, D., and d'Alessandro, C. (**2004**). "On the acoustic sensitivity of a symmetrical two-mass model of the vocal folds to the variation of control parameters," Acta. Acust. Acust. **90**, 746–761.

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

Solomon, N. P., Luschei, E., and Kang, L. (**1994**). "Fundamental frequency and tracheal pressure during three types of vocalizations elicited from anaesthetized dogs," J. Voice **9**, 403–412.

Strogatz, S. H. (**2001**). *Nonlinear Dynamics and Chaos* (Perseus Books, Reading, MA).

Tokuda, I., Riede, T., Neubauer, J., Owren, M. J., and Herzel, H. (**2002**). "Nonlinear analysis of irregular animal vocalizations," J. Acoust. Soc. Am. **111**, 2908–2919.

West, R., and Larson, C. R. (**1999**). "Laryngeal and respiratory activity during vocalization in macaque monkeys," J. Voice **7**, 54–68.

Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (**1998**). "Subharmonics, biphonation, and deterministic chaos in mammal vocalization," Bioacoustics 9, 171–196.

Yin, S. (**2002**). "A new perspective on barking in dogs (*Canis familiaris*)," J. Comp. Psychol. **116**, 189–193.

Yin, S., and McCowan, B. (**2004**). "Barking in domestic dogs: Context specificity and individual identification," Anim. Behav. **68**, 343–355.