Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (Cervus elaphus) stags

Article (Unspecified)

Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (*Cervus elaphus*) stags

David Reby[a)]
Department of Psychology, University of Sussex, Brighton BN1 9QH, United Kingdom

Régine André-Obrecht, Arnaud Galinier, and Jerome Farinas
Institut de Recherche en Informatique de Toulouse, UPS, 31062 Toulouse, France

Bruno Cargnelutti
UR Comportement et Ecologie de la Faune Sauvage, INRA, 31326 Toulouse, France

(Received 28 December 2005; revised 27 July 2006; accepted 5 September 2006)

Bouts of vocalizations given by seven red deer stags were recorded over the rutting period, and homomorphic analysis and hidden Markov models (two techniques typically used for the automatic recognition of human speech utterances) were used to investigate whether the spectral envelope of the calls was individually distinctive. Bouts of common roars (the most common call type) were highly individually distinctive, with an average recognition percentage of 93.5%. A "temporal" split-sample approach indicated that although in most individuals these identity cues held over the rutting period, the ability of the models trained with the bouts of roars recorded early in the rut to correctly classify later vocalizations decreased as the recording date increased. When Markov models trained using the bouts of common roars were used to classify other call types according to their individual membership, the classification results indicated that the cues to identity contained in the common roars were also present in the other call types. This is the first demonstration in mammals other than primates that individuals have vocal cues to identity that are common to the different call types that compose their vocal repertoire. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2358006]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.80.Ev [DOS] Pages: 4080–4089

I. INTRODUCTION

Individual differences in the acoustic structure of vocalizations have been described in several mammal species (e.g., spider monkeys, *Ateles geoffroyi:* Champman and Weary, 1990; mouse lemurs, *Microcebus murinus*: Zimmerman and Lerch, 1993; timber wolves, *Canis lupus*: Tooze *et al.*, 1990; arctic foxes, *Alopex lagopus*: Frommolt *et al.*, 1997; swift foxes, *Vulpes velox*: Darden *et al.*, 2003; spotted hyenas, *Crocuta crocuta*: East and Hofer, 1991; harbour seals, *Phoca vitulina*: Hanggi and Schusterman, 1994; sea otters, *Enhydra lutris*: McShane *et al.*, 1995; elephants, *Loxodonta africana*: McComb *et al.*, 2003; Clemins *et al.*, 2005; bottlenose dolphins, *Tursiops truncatus*: Tyack, 1986; Sayigh *et al.*, 1990; Janik *et al.*, 2006). In deer, studies of individual recognition based on acoustic cues have focused on the vocalizations emitted during early mother/young interactions, and have described how information on individual identity present in vocalizations facilitated either mutual (reindeer, *Rangifer tarandus*: Espmark, 1971, 1975) or partial (red deer: Vankova and Malek, 1997; Vankova *et al.*, 1997) recognition. Individual vocal cues have also been found in the barks given by roe deer (*Capreolus capreolus*) bucks during inter- and intraspecific interactions (Reby *et al.*, 1999) and in the groans of fallow deer (*Dama dama*) bucks during the rutting period (Reby *et al.*, 1998). Although roaring in red deer stags has been extensively studied (Clutton-Brock and Albon, 1979; McComb, 1987, 1988, 1991; Reby *et al.*, 2001; Reby and McComb, 2003a, 2003b; Reby *et al.*, 2005), the potential for red deer rutting calls to convey information on the identity of the caller has not been systematically investigated. Red deer stags give loud and repeated calls during the period of reproduction. Although the roar has received most attention red deer stags actually give four different call types: common roars, harsh roars, chase barks, and barks, each differing in their temporal and spectral acoustic structure, and each being associated with specific postures, social contexts and motivational levels (Reby and McComb, 2003b).

The aim of this study is to evaluate the interindividual variability of the most frequent call type (the *common roar*) and to assess the temporal variation in this identity information over the rutting period. We also assess whether the identity information we detect in the common roars is also present in the other three call types (harsh roars, chase barks, and barks). As three of the studied call types (common roars, harsh roars, and chase barks) are typically composed of more than one vocalization, we use signal detection and classification tools that are compatible with the analysis of series of nonstereotypical signals (rather than focusing our analyses on the first vocalization in the series or treating each vocalization as independent). For this, we use digital signal processing techniques initially developed for the automatic clas-

[a)]Electronic mail: reby@sussex.ac.uk

sification of human speech utterances, and based on the source-filter theory of voice production.

Despite the fact that the source-filter theory was initially designed for the study of human speech production, several recent studies have shown that it can be successfully generalized to most vocalizations emitted by terrestrial mammalian species (Fitch and Hauser, 1995; Fitch, 1997; Rendall *et al.*, 1998; Fitch and Reby, 2001; Reby and McComb, 2003a, 2003b; McComb *et al.*, 2003; Reby *et al.*, 2005). According to this theory, the spectral structure of mammalian voiced vocalizations results from two successive and independent mechanisms. The glottal wave is generated by the vibration of the vocal folds caused by the passage of air through the closed glottis. It is characterized by its fundamental frequency (F0) and its series of harmonic overtones, which are determined by variation in the subglottal pressure and tension of vocal folds (Titze, 1994) and affect the pitch of the vocalization. The relative amplitude of these frequency components is then modulated due to resonances occurring in the supralaryngeal vocal tract. This supralaryngeal filtering generates broadband frequency components in the sound spectrum, which are called vocal tract resonances or formants. Variation of the relative positions and movement of articulators (the larynx, mandibles, tongue, and lips) throughout the call and among different call types will affect the shape of the vocal tract and therefore the formant characteristics (Liebermann, 1968, 1969; Fitch and Hauser, 1995; McComb, 1988; Owren and Rendall, 1997). Both the individual morphology of the animal's vocal tract and the individual variation in its operation are likely to yield individual differences in the central frequencies and bandwidth of formant frequencies, affecting the "timbre" of the vocal signal.

Analyses of the fundamental frequency in red deer roars have suggested that the fundamental frequency varies with motivational state (although the average F0 in adults is 107 Hz, it can drop as low as 20 Hz in "lazy roars") (Reby and McComb, 2003a, 2003b, and unpublished data). Moreover, three of the four call types studied here are either largely (harsh roars) or totally (chase barks and single barks) aperiodic, and therefore do not contain measurable fundamental frequency and harmonics. On this basis, we decided to focus instead on interindividual variation in the filter-related formant frequencies (as in Rendall *et al.*, 1998). In order to separate the characteristics of the formant frequencies (filter) from the fundamental frequency contour (source), we use "homomorphic analysis," a method based on the source-filter paradigm of voice production (Oppenheim and Schafer, 1968). We then run a series of classification experiments using hidden Markov models, in which the bouts of roars are modeled as a succession of silences and roars, and each roar is modeled as a succession of states of the filter-related frequency components. First, we train a model of each individual's bout of roars using the most commonly uttered vocalization in the repertoire, the bout of common roars. Different identification tests are then performed to evaluate the model's ability to recognize and predict the individual membership of these bouts of vocalizations. Second, we test the stability of the information on individual identity conveyed by the formants throughout the rut. For this, we train a model with the bouts uttered in the first days of vocal activity, and we test the remaining bouts as additional cases. Finally, we test whether this individuality holds across the different vocalizations that compose the vocal repertoire of the stags during the rut, i.e., whether red deer stags have individual voice characteristics. For this, we classify the other call types as additional cases, using a model exclusively trained with bouts of common roars.

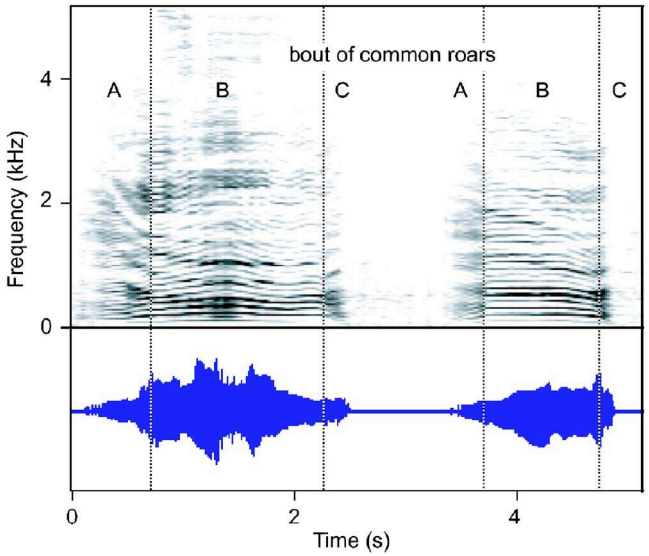

FIG. 1. Narrow band spectrogram of a bout of common roars. The common roar typically includes three phases, A, B, and C. In phase A, the formants fall while the fundamental frequency increases. During phase B the formants are more stationary. Phase C is shorter, with rising formants and a decreasing fundamental frequency.

II. DATA

A. Study animals

We recorded the vocalizations of three adult red deer stags (aged 5, 9, and 12) at the Picarel red deer farm (Southwest France) between September 25 and October 18, 1995, and from four additional adult stags (aged 5, 6, 6, and 8 years, and, respectively, weighing 210, 210, 215, and 230 kg) at the INRA experimental station of Redon (Puy de Dôme) between September 13 and October 4, 1996.

B. Sound recording

Vocalizations were recorded with a Telinga pro-III-S /DAT Mike microphone and a DAT Sony TCD7 recorder, (amplitude resolution: 16 bits, sampling rate: 48 kHz). Digital signals were directly transferred on to a Quadra 950 Macintosh computer using an Audiomedia II sound card and Sound Designer software. Each sound file consisted of a series (bout) of 1–10 consecutive vocalizations (roars) uttered by a stag during a single exhalation. Canary 1.2 (Charif *et al.*, 1995) was used to edit spectrograms of vocalizations. We considered 696 bouts of vocalizations from the seven males. Bouts were classified into four different categories on the basis of their acoustic structure and the postural and social context in which they were given.

We recorded 625 bouts of *common roars* (Fig. 1) from

TABLE I. Distribution of stags' recordings across the period of vocal activity. Each cell represents the number or recorded bouts of common roars. Day 1 is the first day when the stag is heard to vocalize. Bold figures indicate the vocalizations used in the training set of the "temporal" classification test.

	Days of recording																			
Stag	1	2	3	4	5	6	8	10	11	12	13	14	15	16	18	19	22	23	24	25
1	**14**		35					24								3				
2	**20**	6	11	24	22										54		28			
3	**22**	24			7	13	17					7					7			
4	**5**	**10**	11						26							43				
5	**7**								**3**	**8**			4	13	9		16	5	3	
6	**5**	**2**	**2**				31				4									
7	**18**								48	3			9	5	14		1			11

the seven stags, regularly distributed across the periods of vocal activity (Table I). Bouts of common roars contain between 1 and 11 roars, and each roar within the bout is typically composed of three distinct phases that reflect changes in vocal fold vibration and vocal tract shape that occur during the production of the roar [described in detail in Fitch and Reby (2001) and in Reby and McComb (2003a, 2003b)]. In the first phase the stag lowers its larynx and extends its neck to lengthen its vocal tract, inducing the decrease of the formants frequencies and spacing. During the second phase, the vocal tract remains extended and formant spacing remains minimal. Finally, the stag relaxes its vocal tract in the last (and usually shorter) phase, causing formants to rise. We recorded 40 bouts of *harsh roars* from six different individuals (Fig. 2). These bouts are less frequent, and usually characteristic of high motivational states following a contest or a period of intensive herding. Typically the bout starts with a series of short roars (also called grunts) followed by a couple of longer roars with comparable formants. The harsh roar is louder and less periodic than the common roar, and often contains no noticeable harmonics. It is also characterized by little or no formant modulation, reflecting the static body posture adopted by the animal while producing a bout of harsh roars (the larynx is fully lowered and the neck fully extended before the onset of the call and both remain almost static throughout the production of the bout). We also recorded 13 series of *chase barks* (Fig. 3) from three different stags. These calls are short series of short, loud, and explosive barks typically emitted by stags while they chase a hind or a young stag (Clutton-Brock *et al.*, 1982). Finally, we recorded 18 *single barks* (Fig. 4) from five different stags. These louder and longer calls are typically given by stags immediately before a bout of roaring or sometimes singly, and appear to be directed at females (Reby and McComb, 2003b).

III. METHODS

A. Signal processing and analyses

Sound files were low-pass filtered, converted to 8 bits, 8 kHz, SunAU files format, and transferred to a Sun SPARC station. In order to detect the time labels indicating the beginning and ending of each roar in the recorded bouts, we used a preprocessing automated segmentation technique followed by a relative threshold voice detection technique.

The segmentation was performed with the *a priori*

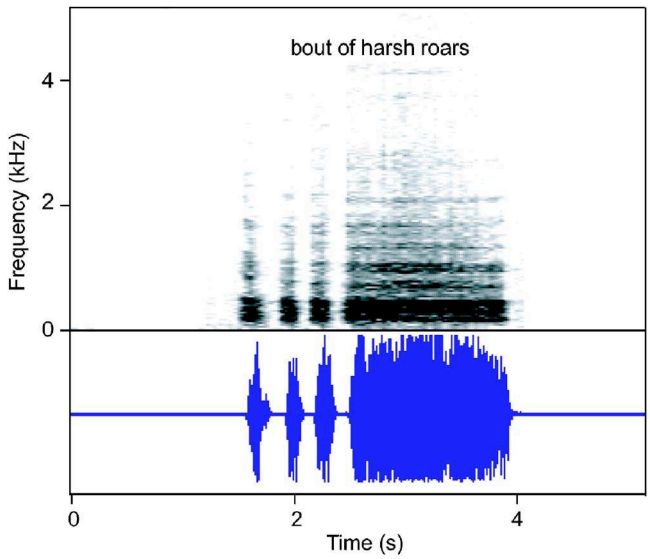

FIG. 2. Narrow band spectrogram of a bout of harsh roars. Compared to common roars, harsh roars are louder, atonal, and characterized by little frequency or energy modulation.

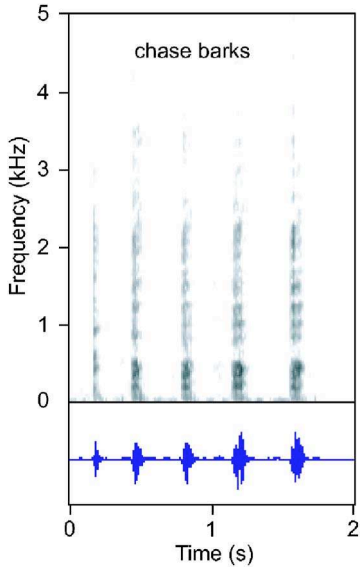

FIG. 3. Narrow band spectrogram of a chase bark series. Chase barks are short vocalizations that are emitted in series.

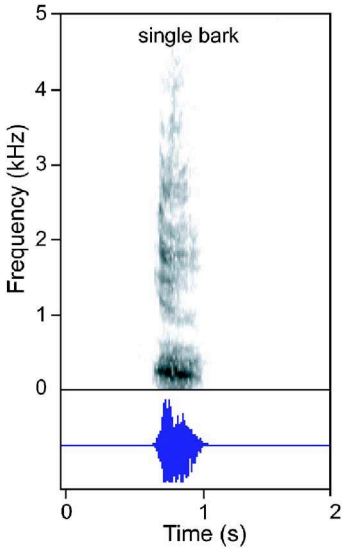

FIG. 4. Narrow band spectrogram of a single bark. Single barks are typically longer than chase barks.

"forward-backward divergence" algorithm (André-Obrecht, 1988). By detecting changes in the parameters of an autoregressive model, this method fragments the signal into stationary segments of variable size, on which statistical parameters can be computed.

Then, in order to define the vocalization boundaries in the sound file by separating intervals of "silence" from intervals of "vocalization," we used the relative energy of each segment. For this, we (1) identified the least energetic segment in the bout, presumably consisting of background noise; (2) calculated the difference between the energy of each segment in the bout, and the energy of the least energetic one; (3) calculated the ratio of each segment's differences to the highest difference. If *Ei* is the energy of a segment *i*, and *n* the number of segments in the bout, the ratio *k* for the considered segment was calculated as follows:

$$k_i = \frac{E_{i-\mathrm{MIN}_{i=1}^{n}}(E_i)}{\mathrm{MAX}\left[E_{i-\mathrm{MIN}_{i=1}^{n}}(E_i)\right]}.$$

Each segment was considered as vocalization if this ratio was greater than 0.75, and silence (or background noise) if it was less than 0.75. This threshold value was determined experimentally with the aim of minimizing the number of misclassified segments. Examination of spectrograms showed that this technique was highly successful at identifying voiced segments; almost all the misclassified segments were very short segments located at the end of the vocalizations. Consecutive segments of silence were then merged into silence phases, and consecutive vocalization segments were merged into vocalization phases. An example of this automated segmentation and energy threshold computation is presented in Fig. 5. The resulting time labels were used to indicate the location of common roars and silences in the bout file for the training phase of the hidden Markov model classifications.

As mentioned previously, we used homomorphic analysis (Oppenheim and Schafer, 1968; Deller, 1999; Quatieri, 2002) to separate the contributions of the excitation source and the vocal tract filter to the sound wave. According to the source-filter theory, the sound wave is produced by filtering the output of the excitation source through the vocal tract filter. In the wave form domain this process can be thought of as the convolution of the excitation wave form with the

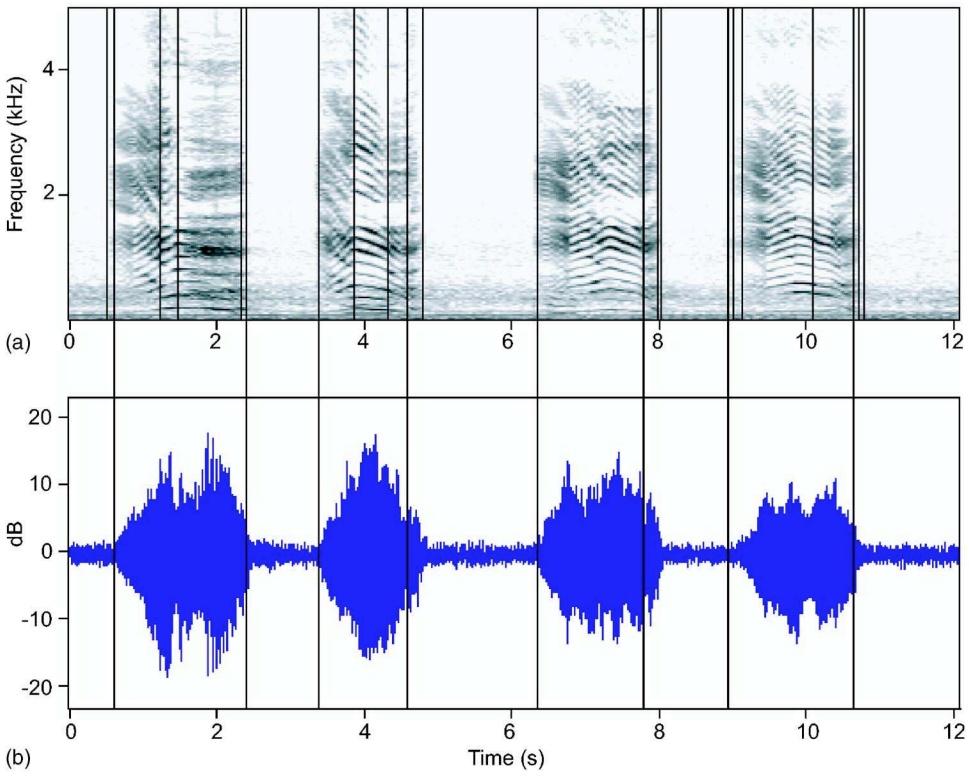

FIG. 5. Automatic detection of vocalization and silence phases in a bout of common roars. (a) Segmentation: the "forward-backward divergence" algorithm fragments the signal into stationary segments of variable size. (b) Energy thresholding: segments are classified as silence or vocalization using the relative energy of each segment. Consecutive silence segments are merged into silence phases and consecutive vocalization segments are merged into vocalization phases.

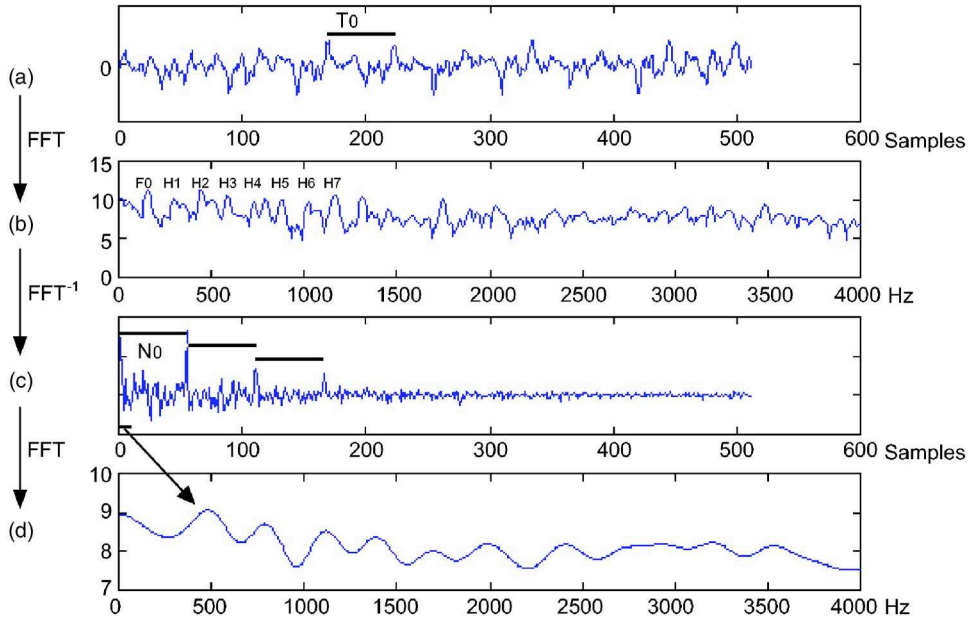

FIG. 6. Homomorphic analysis performed on a 512 samples window of a red deer stag common roar (sampling rate: 8 kHz). Panel A represents the sound wave in the time domain; the signal is periodic with a period T_0. Panel B represents the spectrum (fast Fourier transform) of this sample, with the fundamental frequency (F0) and its harmonic series (the first six harmonics H1–H6 are labeled). Panel C shows the cepstrum Yn. The cepstrum is calculated by taking the inverse Fourier transform of the logarithm of the energy spectrum of the signal. The contribution of the glottal source is represented by impulses spaced by N_0 samples (corresponding to the pitch period), while the contribution of the filter is represented by the lower part of the cepstrum. Finally, panel D shows the frequency spectrum obtained by applying a Fourier transform to the first eight coefficients of the cepstrum, illustrating the smoothing effect of the deconvolution process.

impulse response of the vocal tract. In the spectral domain this same process can be thought of as multiplying the spectrum of the excitation function by the vocal tract's transfer function. Taking the logarithm of the energy spectrum changes this multiplication to an addition, and homomorphic analysis decomposes these additive components of the log spectrum into *cepstral* components, in an exactly analogous way to that in which frequency components are obtained from a complex sound wave. The low "quefrency" cepstral coefficients represent slowly changing aspects of the spectrum—namely the formant frequencies imposed by the vocal tract filter, whereas the high quefrency cepstral coefficients represent rapidly changing aspects of the spectrum—the spectral ripple that is the harmonic structure (the fundamental frequency and its harmonic series). In order to selectively capture the contribution of the vocal tract we used the low quefrency cepstral coefficients. The application of cepstral analysis to a red deer roar is represented in Fig. 6.

When the Mel scale (Stevens *et al.*, 1937), a human logarithmic perceptual scale, is applied to the signal in the frequency domain in order to reduce the dimensionality of the feature vector, these coefficients are called Mel frequency cepstrum coefficients (MFCC). The use of the Mel scale in the classification of red deer vocalizations is supported by the fact that the hearing range of hoofed mammals is comparable to that of humans (Flydal *et al.*, 2001), and that studies of the mammalian auditory system indicate that frequencies are perceived along a roughly logarithmic scale (Fay, 1974; Greenwood, 1990; Clemins, 2005). In our study, we analyzed windows of 25 ms (200 samples at the 8 kHz sampling rate), with a 10 ms overlap. Each window was considered stationary, and the first eight MFCC were retained. For each recorded roar, we obtained a sequence of observation vectors $Y=(Y_1, Y_2, \ldots, Y_T)$ each corresponding to the eight cepstral coefficients of the T subsequent analysis windows.

B. Models

Hidden Markov models are doubly stochastic processes characterized by an underlying stochastic process that is not observable (it is hidden), but can be assessed through another stochastic process that produces the sequence of observed symbols or vectors. Hidden Markov models (HMM) (Rabiner and Juang, 1986) are typically used to model the processes underlying a sequential behavior whose inner workings cannot be directly observed. Here, we make the hypothesis that interindividual differences in the way vocalizations are produced will result in observable interindividual differences in the acoustic structure of the vocalizations. Although we cannot directly observe the individual vocal gestures that are at the origin of the observed individual differences in the acoustic structure of the calls, we can use a HMM to model these underlying mechanisms, and then use these models to predict the individual membership of additional vocalizations. The analyses were run using HTK version 2.2 (Cambridge University Engineering Department). Our Markov model analysis can be formally described as follows: our purpose was to identify one deer among N through the analysis of its bout of vocalization. As a bout consists of a series including up to 11 vocalizations, the bout model M^k_{bou} of the deer D^k is sequence of alternating silence models M_{sil} and vocalization models M^k_{voc}, where the number of vocalizations is variable [Fig. 7(a)]. Each elementary model (M_{sil}, M^k_{voc}, $k=1, \ldots, N$) is a HMM with a Bakis to-

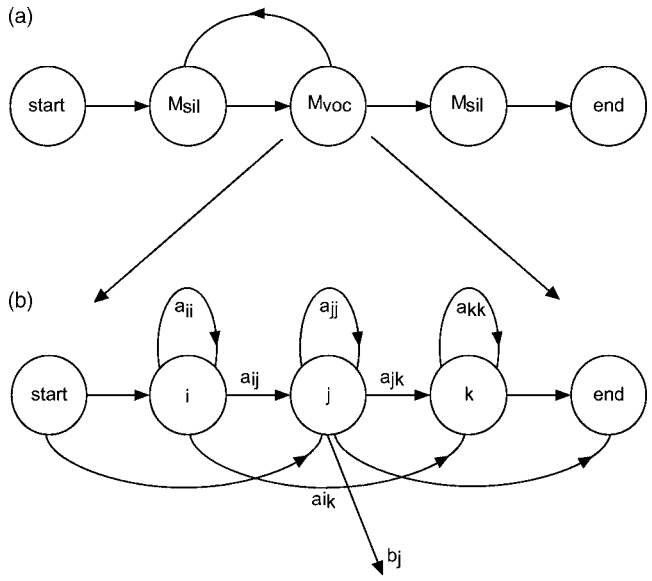

FIG. 7. (a) The model of the roar bout is a succession of silences (M_{sil}) and vocalizations (M_{voc}). The silence model is independent of the considered individuals. (b) In contrast, each individual has its own roar model, a hidden Markov model of three states, where each state emits a vector of eight cepstral coefficients according to a Gaussian mixture probability distribution. Each state is assumed to correspond to one of the three phases that characterize the roar (see Fig. 1).

pology [Fig. 7(b)]. In a Bakis topology, each state can be repeated or omitted. This topology is used in speech recognition in order to take into account the rhythm differences that typically occur in speech sequences. In the case of red deer roaring, this topology enables the HMM automates to model the variability that characterizes deer vocalizations.

The silence model M_{sil} is independent of the considered deer D_k, so that:

$$M^k_{bou} = (M_{sil}, M^k_{voc}).$$

In our study, the hidden process is a finite state, first order Markov chain, meaning that each transition only depends on the very preceding state (and not on the way that state was reached). At each time step, a new state is entered based upon a transition probability distribution ($a_{i,j}$) which depends on the previous state (the Markov property), and an observation output symbol (or vector Y) is produced according to a probability distribution which depends on the current state (b_j). In our case, the distributions b_j are Gaussian mixture models (order 5). During the training phase, we use a subset of records of each deer D_k to adjust the parameters of the corresponding model $M^k_{voc}(a^k_{i,j}, b^k_j)$, using the Baum Welch algorithm (Rabiner and Juang, 1986). The silence model M_{sil} is estimated using all the silence segments available within the training set. During the test phase [performed using the Viterbi algorithm (Forney, 1973)], for each unknown bout characterized by an observation vector sequence $Y=(Y_1, Y_2, \ldots, Y_T)$, and for each individual bout of roar model M^k_{bou}, the likelihood $P(Y|M^k_{bou})$ is calculated. The predicted membership is determined by the best likelihood.

TABLE II. Confusion matrix from the hidden Markov model validation classification computed on the cepstral coefficients from 654 roaring bouts from seven red deer stags. 93.4% of tested bouts are correctly classified.

	Predicted group membership								
Stag	1	2	3	4	5	6	7	% correct	N
1	**73**	1	0	0	1	0	1	96.0	76
2	1	**149**	9	1	1	0	4	90.3	165
3	0	10	**87**	0	0	0	0	89.7	97
4	0	1	0	**94**	0	0	0	98.9	95
5	0	0	0	0	**63**	4	1	92.6	68
6	0	1	0	0	3	**38**	2	86.4	44
7	1	0	0	0	0	1	**107**	98.1	109

C. Classification experiments

Several data sets were constituted in relation to the individual and call type memberships of the vocalization bouts. The first stages (training stage and validation stage) consisted of training the HMM to establish a vocalization model for each individual, with all the 654 bouts of roars. All these bouts were then reclassified using this model in order to test its ability to memorize the dataset (reclassification performance). In the second stage, we tested the model's ability to generalize by performing a random cross-validation test. This evaluated the model's ability to classify additional vocalizations (prediction performance). For this purpose, we trained a HMM with a sample which constituted two thirds of each individual's vocalizations (N=436). This model was then tested with its validation set of remaining vocalizations (N=218). In order to assess the possible degradation of acoustic cues to identity in the course of the rutting period, we conducted a temporal cross validation. To achieve this, we constituted individual training sets including only the vocalizations recorded in the early day(s) of vocal activity (N =165, Table I). We performed a logistic regression on the classification results of the vocalizations recorded later in the rut (N=489) in order to assess the time-related change of the prediction performances of each individual's model. To test if the individuality modeled in common roars holds in the three other vocalization types, we used the subset of common roar bouts as the training set (n=625) and all the other vocalizations (n=71) as test sets. Because stags relatively rarely produce harsh roars, chase barks, and barks, our vocalizations sets are unbalanced among call types, with samples too small to conduct a split-sample approach (Rendall *et al.*, 1998). However, in our case, from the biological point of view, our approach is consistent as recipients are more likely to learn individuality from the most currently uttered call type. Therefore, we do not compare individuality among call types, but we instead test if individuality in the most currently emitted one carries over into the others.

IV. RESULTS

A. Classification of common roars

In the validation stage, 93.4% of the roars were correctly attributed (Table II), with individual scores ranging from

TABLE III. Confusion matrix from the hidden Markov model classification computed on the cepstral coefficients from 654 roaring bouts from seven red deer stags. The model is trained with two-thirds of the available bouts randomly selected within each individual, and the remaining third ($N=218$) are tested as additional cases. 84.9% of tested bouts are correctly classified.

	Predicted group membership								
Stag	1	2	3	4	5	6	7	% correct	N
1	**24**	0	0	1	0	0	0	96.0	25
2	2	**48**	4	1	0	0	0	87.3	55
3	0	5	**27**	0	0	0	0	84.4	32
4	0	1	0	**30**	0	0	1	93.8	32
5	0	1	0	0	**16**	1	5	69.6	23
6	0	0	1	0	4	**9**	1	60.0	15
7	0	0	0	0	4	1	**31**	86.1	36

TABLE IV. Confusion matrix from the hidden Markov model classification computed on the cepstral coefficients from 654 roaring bouts from seven red deer stags. The model is trained with the bouts uttered on the first days of vocal activity ($N=165$), and the bouts uttered during the rest of the period of vocal activity ($N=489$) are tested as additional cases. 58.1% of tested bouts are correctly classified.

	Predicted group membership								
Stag	1	2	3	4	5	6	7	% correct	N
1	**49**	8	3	1	0	0	1	79.0	62
2	0	**95**	35	4	1	0	10	65.5	145
3	1	15	**56**	0	0	0	3	74.7	75
4	6	1	17	**35**	0	0	21	43.8	80
5	0	3	0	0	**12**	4	21	24.0	50
6	0	1	0	0	3	**1**	30	2.9	35
7	2	0	1	0	1	2	**36**	85.7	42

86.4% to 98.9%. In the one-third holdout cross validation, 84.9% of the 218 randomly selected and tested bouts of common roars are correctly classified (Table III). Individual percentages range between 60.0% and 96.0%.

B. Degradation of individuality in common roars with time

In the temporal cross validation, 58.1% of the roars were correctly classified with models constituted with the roars uttered on the first days of vocal activity (Table IV). Percentages were highly variable between individuals, ranging from 2.9% for stag 6 to 85.7% for stag 7. A logistic regression performed on the classification scores of each individual shows that, for three of the seven stags (stag 1: $R=-0.361$, $p<0.005$; stag 2: $R=-0.205$, $p<0.005$, and stag 4: $R=-0.114$, $p=0.06$) the percentage of correctly classified bouts decreases significantly across the period of vocal activity.

C. Across call recognition

In the cross validation performed with the model trained on common roars, 63.4% of the chase barks, harsh roars, and barks are correctly classified (Table V). Last, when chase barks, harsh roars, and barks were included with the common roars in the training set, in the validation phase, the classification score of the common roars was not affected (93.3%) and 91.5% of the calls from the three other types were correctly recognized.

V. CONCLUSIONS

A. Automatic analysis of vocalization sequences

In this paper, we use entirely automated analysis techniques that are particularly appropriate for the processing of large amounts of acoustic data of variable format. The automatic segmentation is particularly well adapted for the detection of calls given in series, and it could be generalized for the automatic detection and identification of animal signals in the context of wildlife population monitoring for conservation or management purposes. The homomorphic analysis is particularly appropriate for disentangling the formants from the fundamental frequency contour in harmonically rich vocalizations, and it has the advantage of characterizing the filter function with a set of largely uncorrelated coefficients suitable for multivariate classifications (Clemins *et al.*, 2005).

In red deer roars, the movement of the larynx causes variation in the filter components. The use of Markov models

TABLE V. Classification of chase barks (cb), barks (ba), and harsh roars (hr) from six stags, using Hidden Markov Models trained with the cepstral coefficients from 625 common roars from seven red deer stags. 63.4% correctly classified. Chase barks: 84.6%, $N=13$; barks: 55.5%, $N=18$; harsh roars: 60%, $N=40$.

	Predicted group membership																						
	1			2			3			4			5			6			7				
Stag	cb	ba	hr	cb	ba	hr	cb	ba	hr	cb	ba	hr	cb	ba	hr	cb	ba	hr	cb	ba	hr	N correct	N total
1	-	-	**1**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1
2	-	-	-	**10**	-	**8**	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	18	20
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	1	7	-	-	1	-	-	**6**	-	-	-	-	-	-	-	1	-	6	16
5	-	-	-	-	2	-	-	-	-	-	-	-	**1**	**9**	**4**	-	1	-	-	-	1	14	18
6	-	-	-	-	-	-	-	-	-	-	-	-	2	1	1	-	-	-	-	-	-	0	4
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-	1	-	-	**1**	**5**	6	12

enables us to take into account these different states as well as the transition probabilities between these states. However, it is important to note that the process being *hidden*, we cannot verify whether the states used in the model actually correspond to those anticipated on the basis of our knowledge of formants production in red deer roaring. It would be interesting in further investigations to assess the effect of varying the number of states and the possible transitions on the predictive performance of the different models. This method has recently been applied successfully to the automatic recognition of call types and individuals from elephant vocalizations (Clemins *et al.*, 2005). The use of Markov models also enables us to include bouts of vocalizations. Such techniques are particularly suited for the study of the acoustic variability of vocalizations emitted in bouts or series, which is the case in many animal acoustic signals.

B. Individual differences in common roars

The results of the validation phase and 1/3 random sample test classifications show that common roar bouts uttered during the rutting period by red deer stags are highly individually structured. Individuality is relatively stable across the period of vocal activity, as a model trained with the vocalizations uttered over a few days at the onset of vocal activity was sufficient to predict the group membership of a majority of the vocalizations uttered later in the rut. In three of the stags studied, we observe a significant decrease in membership prediction, probably resulting from a progressive alteration of the formant characteristics. The very low score obtained for stag 6 may indicate that a drastic change had occurred between the roars given in the first days and those from the rest of the rutting period. It may also be a consequence of the small number of bouts available in the training set ($n=9$) for this individual.

These results suggest that cues to caller's identity exist in the filter-related components of red deer stags' common roars. This variability is likely to result from interindividual differences in the shape of the vocal tract. These differences may have three origins: (1) differences in body size affecting vocal tract length, (Reby and McComb, 2003a), (2) interindividual differences in vocal tract shape independent from body size, and (3) interindividual differences in vocal gesture control of vocal tract length and shape involving larynx, mandible, tongue, and lip positions.

It is notable that classification percentages indicate that cues to individual identity also appear to vary over time. This suggests that the temporal approach described in this paper should be used more often when designing training and testing sets in studies of individual differences based on modeling and classification experiments. Indeed, pooling recordings from different dates, and using classifications percentages from validation phases, *leave-one-out* validations or any cross validations where recordings made on the same date as the tested case(s) are included in the training sample is very likely to result in serious over-estimations of the actual predictive potential of the models.

C. Across calls recognition

When we tested the membership of the three other components of the males' rutting vocal repertoire (harsh roars, chase barks, and barks) using models trained on the cepstral coefficients of the 625 common roars, we obtained percentages of correct classification higher than expected if the membership had been determined randomly. Our sample is too small and our data set is too unbalanced among individuals and call types to allow a comparison of the percentage of recognition between the three types of vocalizations. Nevertheless, our results suggest that although the four vocalizations are produced in different body postures, likely to affect the length and shape of the vocal tract, their formant frequencies share cues to identity. This result indicates that red deer stag have individual voice characteristics, as seen in humans (Doddington, 1985; Furui, 1997) and rhesus monkeys (Rendall *et al.*, 1998). The percentages of correct classification obtained in the validation phase using models trained with tokens from all four vocalization types are higher, showing that the individuality of the voice may consist of individual features shared by all call types as well as individual features specific to each call type (the later being partially lost when a particular call type is not used for the training of the model).

This is the first demonstration of across call individuality in a nonprimate mammal (for primates, see Cheney and Seyfarth, 1988; Rendall *et al.*, 1998). Indeed, to our knowledge, all previous studies on individual cues in acoustic communication in nonprimate mammals have been conducted on the individual differences occurring within each type of call, never across several types of calls (Lambrecht and Dhondt, 1995). Rendall *et al.* (1998) found more mixed evidence for individual voice characteristics across the vocal repertoire of rhesus monkey *Macaca mulatta* (harmonically rich coos were more individually distinctive than either grunts or noisy screams), raising the interesting possibility of interspecific differences in the "individual voice" phenomenon. The ability of red deer receivers to discriminate the identity information discussed above and to transfer it from one call type to another could be assessed by means of playback experiments using the habituation/discrimination paradigm (Rendall *et al.*, 1996; Reby *et al.*, 2001).

D. Potential biological significance of cues to identity in red deer roaring

Studying the acoustic structure of the first roar emitted in a bout, Reby and McComb (2003a) have found that in red deer, formant frequencies and their spacing decreased with increasing age and/or body weight, and that stags attended to these cues during agonistic interactions (Reby *et al.*, 2005). Formant spacing is correlated with the length of the vocal tract and therefore indirectly related to overall body size and body weight. In the present study the recorded males are all adult farmed animals, which are likely to have reached their maximum body weight. The body weight of the four stags for which we had access to biometrical data ranged between 210 and 230 kg, and their roars were characterized by very similar formant frequency spacing corresponding to esti-

mated vocal tract lengths of 81.0, 81.5, 81.5, and 81.8 cm (Reby and McComb, 2003a). Therefore, the individual differences modeled here are more likely to rest in the relative positioning and bandwidth of individual formants rather than in the size-related overall spacing of the formants in the frequency domain. Playback experiments have suggested that females may be preferentially attracted to males with high roaring rates, but indifferent to differences in roar pitch (McComb, 1991). As females often leave or enter male harems, McComb (1991) suggested that females choose which harem to join on the basis of male roaring rate, a potentially reliable cue of the stag's fitness. More recently, Reby *et al.* (2001) have shown that red deer hinds could discriminate between the common roars of their current harem holder and the roars of neighboring males (Reby *et al.*, 2001), and suggested that estrus hinds may choose to mate with stags that they are most familiar with (familiarity being an indicator of the stag's ability to hold mating stands for significant periods), a choice that may partially rely on acoustic individual recognition. The results presented here suggest that hinds may use characteristics of formant frequencies to achieve this individual discrimination, and that these characteristics are available both within and across call types, constituting the equivalent of an individual voice.

ACKNOWLEDGMENTS

The authors thank Marcel Verdier and Alain Brelurut for letting us record the stags at INRA and Catherine Souef at Picarel le Haut, Dominique Pépin, George Gonzales and Mark Hewison for help with the data collection, and Chris Darwin, Karen McComb, Ben Charlton, Loic Hardouin, Drew Rendall and three anonymous referees for their very helpful comments on earlier versions of the manuscript.

Andre-Obrecht, R. (**1988**). "A new statistical approach for the automatic segmentation of continuous speech signals," IEEE Trans. Acoust., Speech, Signal Process. **36**(1), 29–40.

Chapman, C. A., and Weary, D. M. (**1990**). "Variability in spider monkeys' vocalisations may provide basis for individual recognition," Am. J. Primatol **22**, 279–284.

Charif, R. A., Mitchell, S., and Clark, C. W. (**1995**). *Canary 1.2 User's Manual* (Cornell Laboratory of Ornithology, Ithaca, NY).

Cheney, D. L., and Seyfarth, R. M. (**1988**). "Assessment of meaning and the detection of unreliable signals in by vervet monkeys," Anim. Behav. **36**, 477–486.

Clemins, P. J. (**2005**). "Automatic classification of animal vocalizations," Ph.D. thesis, Marquette University.

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. (**2005**). "Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations," J. Acoust. Soc. Am. **117**, 956–963.

Clutton-Brock, T. H., and Albon, S. D. (**1979**). "The roaring of red deer and the evolution of honest advertisement," Behaviour **69**, 124–134.

Clutton-Brock, T. H., Guinness, F. E., and Albon, S. D. (**1982**). *Red Deer: Behavior and Ecology of Two Sexes* (The University of Chicago Press, Chicago, IL).

Darden, S. K., Dabelsteen, T., and Pedersen, S. B. (**2003**). "A potential tool for swift fox (Vulpes velox) conservation: Individuality of long-range barking sequences," J. Mammal. **84**, 1417–1427.

Deller, J. R., Hansen, J. H. L., and Proakis, J. G. (**1999**). *Discrete-Time Processing of Speech Signals* (Wiley-IEEE, New York).

Doddington, G. (**1985**). "Speaker Recognition—Identifying people by their voices," Proc. IEEE **73**, 1651–1662.

East, L. E., and Hofer, H. (**1991**). "Loud calling in a female dominated society: I. Structure and composition of whooping bouts of spotted hyenas, *Crocuta crocuta*," Anim. Behav. **42**, 637–649.

Espmark, Y. (**1971**). "Individual recognition by voice in reindeer mother-young relationship," Behaviour **40**, 295–301.

Espmark, Y. (**1975**). "Individual characteristics in the calls of reindeer calves," Behaviour **54**, 50–59.

Fay, R. R. (**1974**). "Auditory frequency discrimination in vertebrates," J. Acoust. Soc. Am. **56**, 206–209.

Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **102**, 1213–1222.

Fitch, W. T., and Hauser, M. D. (**1995**). "Vocal production in nonhuman primates: acoustics, physiology and functional constraints on honest advertising," Am. J. Primatol. **37**, 191–219.

Fitch, W. T., and Reby, D. (**2001**). "The descended larynx is not uniquely human," Proc. R. Soc. London **268**, 1669–1675.

Flydal, K., Hermansen, A., Enger, P. S., and Reimers, E. (**2001**). "Hearing in red deer," J. Comp. Physiol.. **187**, 265–269.

Forney, G. D. (**1973**). "The Viterbi algorithm," Proc. IEEE **61**, 268–278.

Frommolt, K. H., Kruchenkova, E. P., and Russig, H. (**1997**). "Individuality of territorial barking in arctic foxes, *Alopex lagopus*," in *Proceedings of the First International Symposium on Physiology and Ethology of Wild and Zoo Animals*, edited by F. Klima and R. R. Hofman (Gustav Fisher), Jena, pp. 66–70.

Furui, S. (**1997**). "Recent advances in speaker recognition," in *Audio and Video-Based Biometric Person Authentification*, edited by J. Bigun, G. Chollet, and G. Borgefors (Springer-Verlag), Berlin, pp. 237–252.

Greenwood, D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2650.

Hanggi, E. B., and Schusterman, R. J. (**1994**). "Underwater acoustic displays and individual variation in male harbour seals, *Phoca vitulina*," Anim. Behav. **48**, 1275–1283.

Janik, V. M., Sayigh, L. S., and Wells, R. S. (**2006**). "Signature whistle shape conveys identity information to bottlenose dolphins," Proc. Natl. Acad. Sci. U.S.A. **103**, 8293–8297.

Lambrecht, M. M., and Dhondt, A. A. (**1995**). "Individual voice discrimination in birds," in *Current Ornithology*, edited by D. M. Power (Plenum, New York) Vol. 12, pp. 115–139.

Lieberman, P. (**1968**). "Primate vocalization and human linguistic ability," J. Acoust. Soc. Am. **44**, 1574–1584.

Lieberman, P., Klatt, D. H., and Wilson, W. H. (**1969**). "Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates," Science **164**, 1185–1187.

McComb, K. (**1987**). "Roaring by red deer stags advances the date of oestrus in hinds," Nature (London) **330**, 648–649.

McComb, K. (**1988**). "Roaring and reproduction in red deer, *Cervus elaphus*," Ph.D. thesis, University of Cambridge.

McComb, K. (**1991**). "Female choice for high roaring rate in red deer, *Cervus elaphus*," Anim. Behav. **41**, 79–88.

McComb, K., Reby, D., Baker, L., Moss, C., and Sayialel, S. (**2003**). "Long-distance communication of social identity in African elephants," Anim. Behav. **65**, 317–329.

McShane, L. J., Estes, J. A., Riedman, M. L., and Staedler, M. M. (**1995**). "Repertoire, structure, and individual variation of vocalisations in the sea otter," J. Mammal. **76**, 414–427.

Oppenheim, A. V., and Schafer, R. W. (**1968**). "Homomorphic analysis of speech," IEEE Trans. Audio Electroacoust. **16**(2), 221–226.

Owren, M. J., and Rendall, D. (**1997**). "An affect-conditioning model of nonhuman primate vocal signaling," in *Perspectives in Ethology: Vol. 12. Communication*, edited by D. H. Owings, M. D. Beecher, and N. S. Thompson (Plenum, New York), pp. 299–346.

Quatieri, T. F. (**2002**). *Discrete-Time Speech Signal Processing, Principles and Practice* (Prentice-Hall, Upper Saddle River, NJ).

Rabiner, L. R., and Juang, B. H. (**1986**). "An introduction to hidden Markov models," IEEE ASSP Mag. **3**(1), 4–16.

Reby, D., Joachim, J., Lauga, J., Lek, S., and Aulagnier, S. (**1998**). "Individuality in the groans of fallow deer (*Dama dama*) bucks," J. Zool. **245**, 79–84.

Reby, D., Cargnelutti, B., Joachim, J., and Aulagnier, S. (**1999**). "Spectral acoustic structure of barking in roe deer (*Capreolus capreolus*). Sex-, age- and individual-related variations," C. R. Acad. Sci. III **322**, 271–279.

Reby, D., Izquierdo, M., Hewison, A. J. M., and Pepin, D. (**2001**). "Red deer (*Cervus elaphus*) hinds discriminate between the roars of their current harem holder stag and those of neighbouring stags," Ethology **107**, 951–959.

Reby, D., and McComb, K. (**2003a**). "Anatomical constraints generate hon-

esty: Acoustic cues to age and weight in the roars of red deer stags," Anim. Behav. **65**, 519–530.
Reby, D., and McComb, K. (**2003b**). "Vocal communication and reproduction in deer," Adv. Stud. Behav. **33**, 231–264.
Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., and Clutton-Brock, T. H. (**2005**). "Red deer stags use formants as assessment cues during intrasexual agonistic interactions," Proc. R. Soc. London **272**, 941–947.
Rendall, D., Rodman, P. S., and Edmond, R. E. (**1996**). "Vocal recognition of individuals and kin in free-ranging monkeys," Anim. Behav. **51**, 1007–1015.
Rendall, D., Owren, M. J., and Rodman, P. S. (**1998**). "The role of vocal tract filtering in identity cueing in rhesus monkey *(Macaca mulatta)* vocalizations," J. Acoust. Soc. Am. **103**, 602–614.
Sayigh, L. S., Tyack, P. L., Wells, R. S., and Scott, M. D. (**1990**). "Signature whistles of free-ranging bottlenose dolphins (*Tursiops truncatus*): Stability and mother-offspring comparisons," Behav. Ecol. Sociobiol. **26**, 247–260.
Stevens, S. S., Volkmann, J., and Newman, E. B. (**1937**). "A scale for the measurement of the psychological magnitude pitch," J. Acoust. Soc. Am. **8**, 185–190.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
Tooze, Z. J., Harrington, F. H., and Fentress, J. C. (**1990**). "Individually distinct vocalisations in timber wolves, *Canis lupus*," Anim. Behav. **40**, 723–730.
Tyack, P. L. (**1986**). "Whistle repertoires of two bottlenose dolphins, *Tursiops truncatus*: Mimicry of signature whistles?," Behav. Ecol. Sociobiol. **18**, 251–257.
Vankova, D., and Malek, J. (**1997**). "Characteristics of the vocalisations of red deer (*Cervus elaphus*) hinds and calves," Bioacoustics 7, 281–289.
Vankova, D., Bartos, L., and Malek, J. (**1997**). "The role of vocalizations in the communication between red deer hinds and calves," Ethology **103**, 795–808.
Zimmerman, E., and Lerch, C. (**1993**). "The complex acoustic design of an advertisement call in male Mouse lemurs (*Microcebus murinus*) and sources of its variation," Ethology **93**, 211–224.