

## VOCAL TRACT LENGTH AND ACOUSTICS OF VOCALIZATION IN THE DOMESTIC DOG (*CANIS FAMILIARIS*)

TOBIAS RIEDE<sup>1,\*</sup> AND TECUMSEH FITCH<sup>2</sup>

<sup>1</sup>*Institut für Biologie, c/o Professor G. Tembrock, Humboldt-Universität zu Berlin, Invalidenstrasse 43, 10115 Berlin, Germany* and <sup>2</sup>*Harvard/MIT Speech and Hearing Sciences, Room 1036, Wm James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA*

\*e-mail: tobias=riede@rz.hu-berlin.de

Accepted 2 August; published on WWW 30 September 1999

### Summary

The physical nature of the vocal tract results in the production of formants during vocalisation. In some animals (including humans), receivers can derive information (such as body size) about sender characteristics on the basis of formant characteristics. Domestication and selective breeding have resulted in a high variability in head size and shape in the dog (*Canis familiaris*), suggesting that there might be large differences in the vocal tract length, which could cause formant behaviour to affect interbreed communication.

Lateral radiographs were made of dogs from several breeds ranging in size from a Yorkshire terrier (2.5 kg) to

a German shepherd (50 kg) and were used to measure vocal tract length. In addition, we recorded an acoustic signal (growling) from some dogs. Significant correlations were found between vocal tract length, body mass and formant dispersion, suggesting that formant dispersion can deliver information about the body size of the vocalizer. Because of the low correlation between vocal tract length and the first formant, we predict a non-uniform vocal tract shape.

Key words: Canidae, dog, growling, formant, source-tract theory, bioacoustics.

### Introduction

The sound production systems of all mammals exhibit a number of fundamental anatomical and acoustic similarities. The primary acoustic signal is generated at a source, typically the vocal folds of the larynx (the glottal source), driven into rapid mechanical oscillations by an expiratory airflow from the lungs. Opening and closing the vocal folds modulates the airflow through the glottal opening, producing a time-varying acoustic signal, the glottal source signal. Many mammals produce a nearly periodic signal in the larynx that can be represented as a Fourier series, with a fundamental (lowest) frequency and integer multiple harmonics. A narrowed but non-oscillating larynx can also generate turbulent noise.

All mammals also have a supralaryngeal vocal tract (hereafter simply referred to as 'vocal tract'), through which the sound generated at the glottal source must pass. The column of air contained in the vocal tract, like any tube of air, has resonant modes that selectively allow certain frequencies in the glottal source to pass and radiate out through the mouth or nostrils into the environment. In other words, the vocal tract acts as a bank of bandpass filters, each of which allows a narrow range of frequencies to pass. These resonances of the vocal tract, along with the spectral peaks they produce in the vocal signal, are given the special name 'formants' (from the Latin *formare*, to shape; after Hermann, 1890). This term has a long history of use in both speech science (Fant, 1960; Titze,

1994) and bioacoustics (Lieberman et al., 1969; Nowicki, 1987; McComb, 1988; Suthers, 1994; Fitch and Hauser, 1995).

In an anatomical investigation of the non-human vocal tract, Lieberman et al. (1969) demonstrated important differences between non-human and human vocal tract shapes and concluded that the production of the full range of vowels found in human speech is impossible without a modern human vocal tract. For a long period after this, there was little research on formants in animal communication (but see Andrew, 1976). Recently, however, there has been a resurgence of interest in animal formant production and perception.

In terms of perception, it has long been known that various animals can discriminate human vowels (baboons, Hienz and Brady, 1988; dogs, Baru, 1975; cats, Dewson, 1964; blackbirds and pigeons, Hienz et al., 1981; macaques, Sommers et al., 1992). Because vowels are perceived and classified primarily on the basis of the two lowest formant frequencies (Peterson and Barney, 1952; Bogert and Peterson, 1957; Kent, 1978, 1979), this work suggests that animals can perceive formants. More directly, Sommers et al. (1992) documented the ability of Japanese macaques to discriminate formant frequencies with an accuracy rivalling that of humans. Finally, Owren and Bernacki (1988) and Owren (1990) used operant techniques to demonstrate that spectral features potentially related to formants are discriminated by vervet monkeys in their own species-specific calls. A number

of studies have demonstrated the potential communicative relevance of formants in several mammal species (cats, Shipley et al., 1991; rhesus macaques, Hauser et al., 1993; baboons, Owren et al., 1997). More detailed discussions of the importance of formants for acoustic communication in animals are given by Fitch (1997) and Owren and Rendall (1997).

In terms of production, formant frequencies are strongly influenced by vocal tract length and shape (Fant, 1960; Titze, 1994). Carterette et al. (1979) found a good correspondence between the formant frequencies of isolation calls by domestic kittens and the formant frequencies predicted theoretically for a uniform tube (closed at one end) of the same length as the kitten vocal tract. The effect of vocal tract length was studied further by Fitch (1997), who measured vocal tract length, body size and formant frequencies in rhesus macaques (*Macaca mulatta*) and found that formant frequency spacing is a reliable correlate of body size in that species. Further investigations on the relationship between vocal tract length, body size and formant characteristic suggested that the prediction of body size from the acoustic signal (and *vice versa*) may also be applicable to humans (T. Fitch and J. Giedd, in preparation). Given the study of Baru (1975), which indicated that formants in human speech can be perceived by dogs, it is reasonable to ask whether formant frequencies could potentially be used in intraspecific size assessment in this species. In canid communication, the advertisement of body size and strength plays an important role (mentioned as early as Darwin, 1872). In dominant dogs, large size is exaggerated by the stiff upright threat posture, while the crouched posture of submission decreases apparent size. Low-frequency, broad-band barking or growling often accompanies threats, vocalizations ideally suited for accurately outlining the vocal tract transfer function. If vocal tract length is related to body size in canids, growls could possibly convey an accurate impression of size to listeners. In contrast, the high-frequency whining that accompanies the crouched posture of submission provides little information on body size because such a high-frequency source contains energy at only a few widely spaced frequencies (Ryalls and Lieberman, 1982).

At least 12 000 years of domestication (Clutton-Brock, 1995; Vila et al., 1997) have resulted in a remarkable diversity of dog breeds, differing in behavioural traits and overall body size variables. Dogs range in mass from chihuahuas to Saint Bernards, a 100-fold difference. However, this difference in body size does not necessarily imply an equivalent difference in proportions (Clutton-Brock, 1995). Most dog skulls are actually quite similar in proportion, except for the extremely long-faced breeds such as the borzoi or short-faced breeds such as the boxer. Morey (1992) gives evidence that skull proportions have been relatively constant since prehistoric times. Because of these significant size differences among breeds, it is not obvious that the correlation between body size, vocal tract length and formant frequencies documented for macaques (Fitch, 1997) will hold for domestic dogs of different breeds. The goals of this study were therefore to analyze the correlation between vocal tract length (VTL), body mass and formant frequencies in the domestic dog and to determine

whether the huge intraspecific size differences generated by selective breeding have affected the relationship between body size and the acoustic cues related to vocal tract length.

## Materials and methods

### *Formant frequencies and vocal tract length*

Acoustic theory predicts that formant centre frequencies will relate to simple anatomical measures of the vocal tract, mainly its length and shape (particularly variations in cross-sectional area along the length of the tract). The length is the single most important anatomical variable that influences the frequencies of the formants. As a first approximation, we can consider the resonant frequencies (natural modes) of an air-filled tube that does not vary in cross-sectional area, a 'uniform tube'. For a uniform tube with one end closed, resonant frequencies are described by:

$$F_i = \frac{(2i+1)c}{4VTL}, \quad (1)$$

and with both ends closed by:

$$F_i = \frac{ic}{2VTL}, \quad (2)$$

where  $i$  is the formant number,  $c$  is the speed of sound ( $350 \text{ m s}^{-1}$ ),  $VTL$  is vocal tract length (in m) and  $F_i$  is the frequency (in Hz) of  $i$ th formant. Note that, for a uniform tube regardless of end conditions, the frequency difference between the successive resonances is constant and given by:

$$F_i - F_{i-1} = c/2VTL. \quad (3)$$

Thus, if formants from a uniform tube can be measured (which requires an excitatory signal with appreciable energy at the formant frequency), their frequencies provide direct information about the length of the tube. Specifically, the difference between successive resonances should theoretically provide an accurate estimate of tube length. The only effect of the boundary conditions (whether the tube is open or closed at the ends) is to shift the entire pattern up or down in frequency, offsetting the absolute frequencies of the formants (e.g. of the lowest formant  $F_1$ ), but not changing their spacing. A measure of formant spacing therefore overcomes the need to make assumptions about these vocal tract boundary conditions, which are known to vary in humans but which are not known for non-humans. These factors led Fitch (1997) to introduce a measure he termed 'formant dispersion', which is the averaged difference between successive formants:

$$D_f = \frac{\sum_{i=1}^{N-1} (F_{i+1} - F_i)}{N-1}, \quad (4)$$

where  $D_f$  is the formant dispersion (in Hz) and  $N$  is the total number of formants measured. Formant dispersion is the mean of the formant spacings. This is a reasonable choice, since the

mean is theoretically the best predictor (in a least-squared error sense). Averaging the differences should make the measure less sensitive to deviations from the first approximation (caused, for example, by non-uniform shape) and thus provides an overall estimate of spectral dispersion. Higher-order statistics, e.g. the standard deviation of the formant intervals  $F_{i+1}-F_i$ , can be used to evaluate the extent to which the uniform tube approximation holds. For example, four formants of 500, 1500, 2500 and 3500 Hz (as in a 17.5 cm long uniform tube) would yield a  $D_f$  of 1000 Hz. However, a set of formants of 700, 1200, 2200 and 3700 Hz (as in the human spoken vowel) also yields a  $D_f$  of 1000 Hz, but does not approximate those of a uniform tube, which has evenly spaced formants. The uniform tube approximation is a good one to the extent that the intervals between  $F_1$  and  $F_2$ ,  $F_2$  and  $F_3$ ,  $F_3$  and  $F_4$ , etc. are nearly equal, so that the standard deviation of the formant intervals will be small.

In summary, the spacing between the resonant frequencies of a uniform tube decreases as its length increases, implying that individuals with longer vocal tracts should exhibit smaller formant dispersions. In this study, we tested this prediction by measuring *VTL* from radiographs (X-ray photographs) of domestic dogs. We then used linear prediction to measure the formant frequencies from the growls of the same dogs, calculating the formant dispersion using equation 4 and comparing it with the predicted formant dispersion from equation 3. Finally, we tested the correlation between these variables and body mass to see how closely these acoustic variables predict body size.

#### Subjects

The subjects were 47 domestic dogs of various ages representing 21 different breeds. These dogs were patients at a veterinary practice. All dogs were treated for broken legs, an affliction unlikely to have a substantial impact on the vocalizations of the animal. The animals were between 0.5 and 15 years old and between 2.5 and 50 kg in mass. Dogs were weighed using a 100 kg scale (accuracy 100 g). All animals were within the mass range typical of their breed (according to Clark and Brace, 1996).

#### Anatomical measures

Radiographs were taken of 33 anaesthetised dogs immediately after surgery. Each animal was placed on its side on a radiographic table, and lateral images were made of the head and the neck region. For calibration, a 1 cm lead reference square was positioned at the midsagittal level of the head. Vocal tract length (*VTL*) was determined from tracings of the X-ray images scanned using a Microtek MRS-600Z scanner and measured using NIH Image (version 1.58). Image clarity was sufficient to delineate the outlines of the oral vocal tract, although the glottis itself was not visible in all cases. Therefore, the midpoint of the thyrohyoid bone, which was always visible, was used as the origin of our vocal tract length measurement. The thyrohyoid always appears on such radiographs just on the cranial side of the glottis

(Gaskell, 1974; Kneller, 1994). Thus, our measure was slightly shorter than the actual vocal tract length, but in a manner consistent from dog to dog. A curvilinear line from the midpoint of the thyrohyoid cartilage along the line of the soft and hard palates to the front of the incisors was drawn and measured (Fig. 1). The *VTL* was determined with reference to the calibration square. The skull length was measured as the distance between the front of the incisors and the protuberantia occipitalis externa of the os occipitale; the latter was sometimes off the radiograph, and the skull length measurement was therefore missing in some cases. The repeatability of these measurements was high: in 10 animals, the measurements were repeated 10 times, yielding standard deviations of 0.5 cm (2.8%) for *VTL* and 0.1 cm (<1%) for skull length.

#### Acoustic measurements

Dog vocalisations were recorded using a Sony WMD3 Walkman Professional cassette recorder and Sennheiser microphone (ME80 head with K3U power module) on BASF ChromeSuper II tapes whilst the animals were in their stalls in the clinic. Only growl vocalisations were considered for analysis, since growls are uttered with a nearly closed mouth and therefore use the full length of the vocal tract. Growling is a low-frequency, broadband signal, which is uttered in sequences of variable duration consisting of growls with interspersed pauses. Because of their low frequency and broad bandwidth, growls are well-suited for formant estimation. Growling was induced by staring into the dog's eyes (a mild threat to the animal). If growling could not be provoked in this way, only radiographic data were used for that individual. This opportunistic approach resulted in an overlap of anatomical and acoustic data sets for 12 dogs representing eight breeds. To enlarge the acoustic data set, we also recorded growling for 14 dogs from which we did not take radiographs.

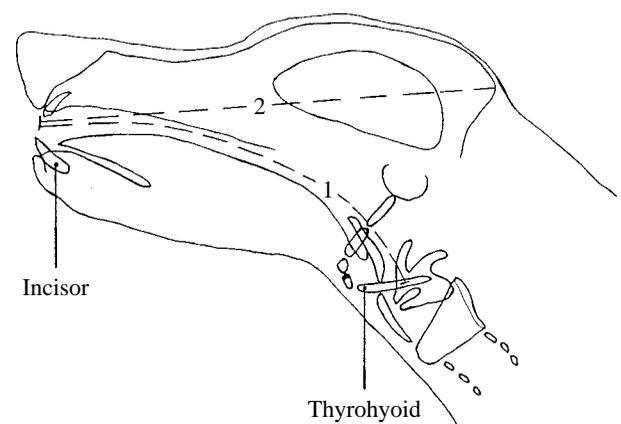


Fig. 1. Schematic drawing of the anatomical features and the morphometric features used in this study as observed by radiography. The lines represent vocal tract length (*VTL*) (1) and skull length (2) and were measured on digitized images of radiographs in NIH image with reference to a 1 cm calibration square.

Recordings were digitised at 16-bit quantization and 24 kHz sampling rate using an Audiomeia II sound card and Sound Designer 2 software (Digidesign, Palo Alto, CA, USA).

Linear predictive coding (LPC) is a spectral modelling

technique widely used to estimate formant frequencies in human speech. LPC uses least-squares curve-fitting to estimate the value of a point in a time-domain waveform based on the past  $n$  points, where  $n$  is the order of the LPC analysis. LPC

Table 1. Raw morphological and acoustic data for 47 individual dogs

Breed	Mass (kg)	VTL (cm)	Skull length (cm)	Number of formants	$D_f$ (Hz)	s.d. of $D_f$ (Hz)	CV of $D_f$	$F_1$ (Hz)	$F_2$ (Hz)
Yorkshire terrier	2.5	9.6	8.5						
Yorkshire terrier	3.1	6.9	8.1	7	1703	714	41.9	891	1828
Yorkshire terrier	3.5	9.9	10.4						
Pekingese	3.5	10.4	8.2						
Pekingese	3.5	7.5	8.4						
Yorkshire terrier	3.6	8.8	9.0	7	1910	748	39.1	609	1688
West Highland terrier	6.6	11.8	13.9						
Shih tzu	7.5	10.1	10.9	8	1466	405	27.6	984	1594
Yorkshire terrier	8.0	13.4							
West Highland terrier	8.0	15.2	16.0	7	1381	370	26.8	625	2063
Fox terrier	8.7	17.2	17.4	8	1467	620	42.2	797	1813
West Highland terrier	8.7	15.5							
Tibet terrier	9.5	14.0	15.9	8	1500	820	54.6	1031	2438
West Highland terrier	10.0	14.1	15.7	8	1493	950	63.6	1078	2391
Dachshund	10.0	16.2							
Mongrel	10.0	13.9	15.3	7	1510	470	31.1	469	2281
Dachshund	10.5	15.7	17.0	7	1229	572	46.5	375	844
Dachshund	10.6	13.4							
Mongrel	12.0	15.1	17.6						
Beagle	17.5	19.0	18.1						
Pharaoh hound	19.0	20.6	18.8	7	867	254	29.3	656	1166
Irish setter	23.0	15.9	16.7						
Irish setter	23.0	19.5	22.3	8	733	513	70	586	961
Irish setter	24.2	18.6	20.2	7	1035	742	71.7	492	820
Irish setter	24.5	20.9							
Staffordshire terrier	25.0	18.1	21.4						
Hunting terrier	30.0	21.8	22.1						
Giant schnauzer	31.0	18.9							
German shepherd	33.5	19.3	20.0						
German shepherd	34.0	20.4	23.0						
Dobermann	34.0	21.8							
Rottweiler	40.0	22.4							
German shepherd	50.0	21.6	25.1						
Pekingese	4.5			5	1960	558	28.5	1173	3343
Pekingese	6.7			7	1630	500	30.6	688	1469
Mongrel	7.5			4	1469	1008	68.6	328	1102
Mongrel	8.0			8	1546	525	33.9	563	2578
German shepherd	15.0			5	1102	732	66.4	515	891
Boxer	22.0			3	705	—		450	920
Rottweiler	25.0			7	678	226	33.3	328	680
Rottweiler	28.0			7	679	320	47.1	234	539
Giant schnauzer	31.0			9	568	317	55.8	422	1055
Mongrel	33.5			7	734	327	44.5	281	656
Pit bull terrier	35.0			7	739	116	15.9	688	1375
Dobermann	38.5			7	642	71	11	383	1031
Giant schnauzer	39.0			7	689	244	35.4	258	586
Hovawarth	50.0			7	717	278	38.8	656	1512

VTL, vocal tract length;  $D_f$ , formant dispersion; s.d., standard deviation of the formant dispersion; CV, coefficient of variation of the formant dispersion;  $F_1$ , lowest formant frequency;  $F_2$ , second formant frequency.

Table 2. Descriptive data for acoustic and anatomical variables of the dogs used in the study

	<i>N</i>	Mean	S.D.	S.E.M.	Min	Max
Age (years)	47	5.4	4.2	0.6	0.5	15
Mass (kg)	47	18.9	13.4	1.96	2.5	50.0
log <sub>10</sub> (body mass)	47	1.14	0.37	0.053	0.39	1.69
Skull length (cm)	25	16.0	5.27	1.08	8.1	25.1
<i>VTL</i> (cm)	33	15.7	4.05	0.71	6.9	22.4
<i>D<sub>f</sub></i> (Hz)	26	1101	429	88.6	568	1910
<i>F<sub>1</sub></i> (Hz)	26	585	239	49	234	1173

*N*, sample size; S.D., standard deviation; S.E.M., standard error of the mean; Min, Max, minimum and maximum values, respectively; *VTL*, vocal tract length; *D<sub>f</sub>*, formant dispersion; *F<sub>1</sub>*, lowest formant frequency.

Body mass was measured in kilograms.

algorithms then construct the best-fitting all-pole model to account for the waveform. ‘All-pole’ means that only vocal tract resonances (‘poles’) are estimated, and not anti-resonances (‘zeroes’). Such a spectral model appears to be a valid first approximation for most human speech (Markel and Gray, 1976) as well as for the dog growls we analysed here. The specifics of the many algorithms used to perform LPC are well-documented elsewhere (see Markel and Gray, 1976, for the mathematical details, or Owren and Bernacki, 1998, for applications in bioacoustics); we used the autocorrelation technique provided by the signal processing toolbox of MATLAB 4.2 (The Mathworks, Natick, MA, USA) in the current analysis. This technique provides as output the coefficients of an *n*th-order all-pole digital filter whose frequency response best approximates (in a least-squares sense) the spectrum of the input signal. Given a broadband source signal and an appropriate model order (approximately 2 plus twice the number of formants), LPC analysis can provide an extremely accurate estimate of formant centre frequencies in both human speech and animal sounds (see Fitch, 1997). Signal analysis was performed using MATLAB with 18–26 coefficients and pre-emphasis settings of 0.8–0.99. All LPC measurements were visually verified by superimposing the LPC-derived frequency response over a 512-point fast Fourier transform (FFT) of the same time slice, allowing the user to select the optimum number of coefficients for each call by trial and error. When possible, we measured five formants in each growl, but the number of formants that could be extracted was highly subject-dependent.

### Results

Raw data are presented in Table 1, and a summary of the data is presented in Table 2. Of the 33 dogs X-rayed, vocal tract length varied between 6.9 cm (a Yorkshire terrier) and 22.4 cm (a Rottweiler). In 12 of these 33 dogs, growls were elicited for acoustic analysis. There were no significant differences between females and males (unpaired *t*-test,  $t=-1.08$ ,  $N_f=19$ ,  $N_m=26$ ,  $P=0.57$ ) and therefore the data for both sexes were combined for further analysis. Normality testing indicated that head length, *VTL* and body mass measurements were not normally distributed, while formant dispersion and age were. All correlation analyses were

therefore performed using both Pearson parametric correlation and Spearman non-parametric correlation. No significant differences in the results of these two analyses were found, so we report only the Pearson values. Correlation analyses are presented in Table 3. High and statistically significant correlations were found between all the variables measured except for skull length and the lowest formant frequency (*F<sub>1</sub>*) and between *VTL* and *F<sub>1</sub>*. In contrast to *F<sub>1</sub>*, correlations between formant dispersion and *VTL* or skull length were highly significant ( $P<0.001$ ). *VTL* increased with body mass, and formant dispersion decreased with increasing *VTL* and with increasing body mass (Fig. 2).

For 12 dogs (for which both anatomical and acoustic data were available), the theoretical value calculated using the *VTL* obtained from the radiograph and equation 3 was compared with that calculated from vocalisation recordings using equation 4. The predicted and measured formant dispersion were not significantly different (paired *t*-test:  $t=0.26$ ;  $N=12$ ;  $P=0.79$ ).

The longest growl that we recorded lasted 3 s. The length of the growl sequences from different dogs varied from a few seconds to several minutes. In four animals with very long growl

Table 3. Pearson correlation coefficients between the various acoustic and anatomical variables measured in this study

	log <sub>10</sub> (body mass)	Skull length	<i>VTL</i>	<i>F<sub>1</sub></i>
Skull length	0.96*** ( <i>N</i> =25)	1.00		
<i>VTL</i>	0.93*** ( <i>N</i> =33)	0.95*** ( <i>N</i> =25)	1.00	
<i>F<sub>1</sub></i>	-0.53** ( <i>N</i> =26)	-0.37NS ( <i>N</i> =12) <i>P</i> =0.117	-0.38NS ( <i>N</i> =12) <i>P</i> =0.107	1.00
<i>D<sub>f</sub></i>	-0.88*** ( <i>N</i> =26)	-0.87*** ( <i>N</i> =12)	-0.87*** ( <i>N</i> =12)	0.62*** ( <i>N</i> =26)

\*\*\* $P<0.001$ ; \*\* $P<0.01$ ; \* $P<0.05$ ; NS, not significant ( $P>0.05$ ).

*F<sub>1</sub>*, lowest formant frequency; *VTL*, vocal tract length; *D<sub>f</sub>*, formant dispersion.

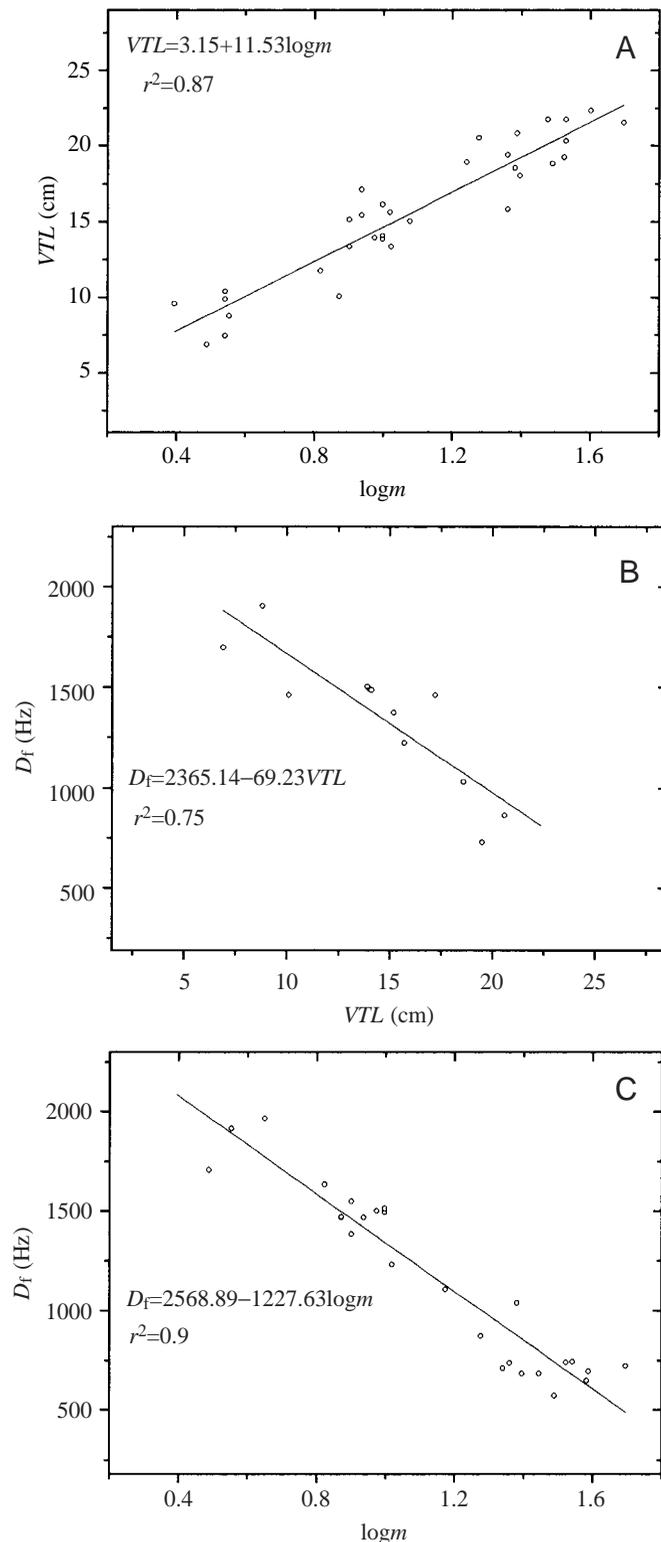


Fig. 2. Relationship between (A) vocal tract length,  $VTL$ , and body mass ( $m$  in kg), (B)  $VTL$  and formant dispersion,  $D_f$ , and (C) formant dispersion and body mass. The lines represent the linear regression lines, with the equations and  $r^2$  values given in each case.

sequences, the stability of formant frequencies was investigated. Within a single growl utterance, the formant pattern was rather stable (Fig. 3 gives a representative example). In the four

animals for which multiple growls were examined, the formant distribution was found to vary at least slightly between growls. In particular, certain formants sometimes 'dropped out' of the signal (Fig. 4), suggesting that the dogs were able to modify their source characteristics or vocal tract configuration (e.g. using lip or tongue movements) somewhat during a growl sequence.

Standard deviations of formant spacing ('formant deviation') were quite variable between dogs, ranging from 11% to 71% of the formant dispersion (mean 42%). Thus, some dogs closely approximated the uniform tube approximation (e.g. dog 1 in Fig. 4), while others deviated considerably from this approximation (e.g. dog 4 in Fig. 4).

### Discussion

The results presented above indicate that formant frequency spacing (as measured by  $D_f$ ) is closely correlated with vocal tract length ( $VTL$ ) (Fig. 2B). Because  $VTL$  is in turn related to body mass (Fig. 2A), the growling signals analysed here provide a reliable indication of the size of the animal that produced them (Fig. 2C). We now consider some of the acoustic, anatomical and methodological issues raised by our data and end with a brief consideration of some of the implications of these data for canine communication.

Formant frequencies are held relatively constant across the duration of an individual growl (Fig. 3), but show some changes across growls produced by the same animal (Fig. 4). These changes may result from a process paralleling that used in human speech: active variation of the shape (mid-sagittal cross-sectional area function) of the vocal tract. Vocal tract shape can be changed voluntarily by a vocaliser by varying the amount of mouth opening (Stevens and House, 1955; Ohala, 1984), as has been shown in both monkeys (Hauser et al., 1993) and birds (Westneat et al., 1993; for a review, see Gaunt and Nowicki, 1998). Mouth opening in humans has a greater effect on the first formant than on other formants (Stevens and House, 1955). Vocal tract shape can also be modified by movements of the body or blade of the tongue, by raising or lowering the larynx or by using the velum to open or close the nasal passages. Although such changes have not been documented in non-humans, comparative anatomy suggests that dogs may have similar capabilities to those of humans in these respects.

The idea that dogs can vary their vocal tract shape receives further support from the data on deviation from even spacing of the formants. For a uniform vocal tract, the formant spacing is expected to be constant. This is not the case in all the dogs' growls we recorded. The standard deviations of formant spacing show considerable variability for an individual dog, with some individuals approaching uniform spacing and others demonstrating quite high deviation from even spacing. For example, in Fig. 4, we can see that dog 1 has approximately equal formant spacing (in growls 1–4), while dog 4 does not. Because our radiographs clearly demonstrate some deviations from uniform cross-sectional areas in the vocal tracts of many of the dogs, we hypothesise that the deviations observed in the acoustic signal result from non-uniform vocal tract shapes.

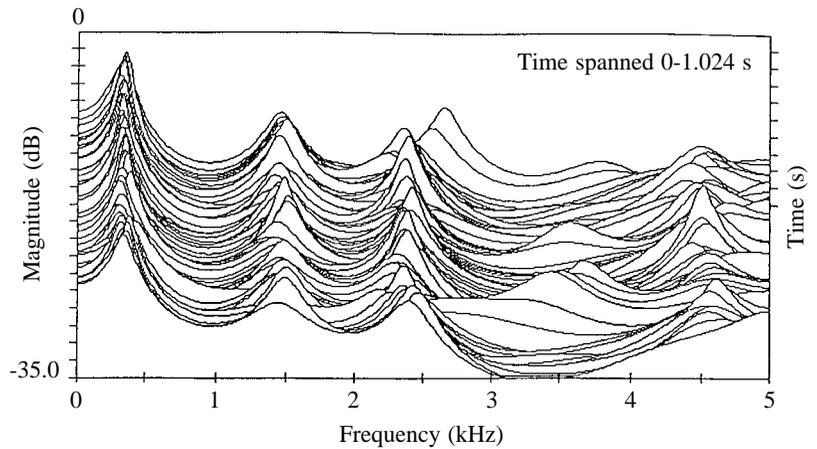


Fig. 3. Waterfall representation of the linear predictive coding (LPC) curves of a growl showing the formant distribution over approximately 1 s. This three-dimensional spectrogram display shows the Fourier transform spectra of several time slices. The actual time slices overlap by 75 %.

Unfortunately, our data provide only static images of the vocal tract of anaesthetised dogs, while the relevant vocal tract shape is obviously that adopted during vocalisation. We are therefore unable to determine whether different degrees of formant deviation result from differences across dogs in the vocal posture adopted during growling, although this seems a plausible hypothesis. Testing this will require anatomical images of the vocal tract of dogs taken during vocalisation; cineradiography (moving picture radiography) would be ideally suited for this task. Another useful extension of our results would involve three-dimensional reconstruction of the vocal tract using non-invasive imaging techniques such as magnetic resonance imaging. Such techniques have proved exceedingly valuable in understanding human speech acoustics (Baer et al., 1991; Story et al., 1996) and could be used with

other species as veterinary use of non-invasive imaging techniques increases.

Because of deviations from even spacing, the amount of information contained in any one formant is less than that contained in the entire ensemble of formants. A good illustration of this is that the lowest formant,  $F_1$ , provided a much less reliable indication of body size than did formant dispersion,  $D_f$  (Table 3). The correlation coefficient between  $\log_{10}m$  (where  $m$  is body mass) and  $F_1$  was only  $-0.53$  ( $N=12$ ) compared with that between  $D_f$  and  $\log_{10}m$  of  $-0.88$ . In addition, there tended to be a negative correlation between  $F_1$  and  $VTL$ , although it was not statistically significant ( $r=-0.38$ ,  $P=0.1$ ), while the correlation between  $VTL$  and  $D_f$  was highly significant ( $r=-0.87$ ,  $P<0.001$ ). However, it should be noted that the sample size for this analysis is too small ( $N=12$ ) to reject

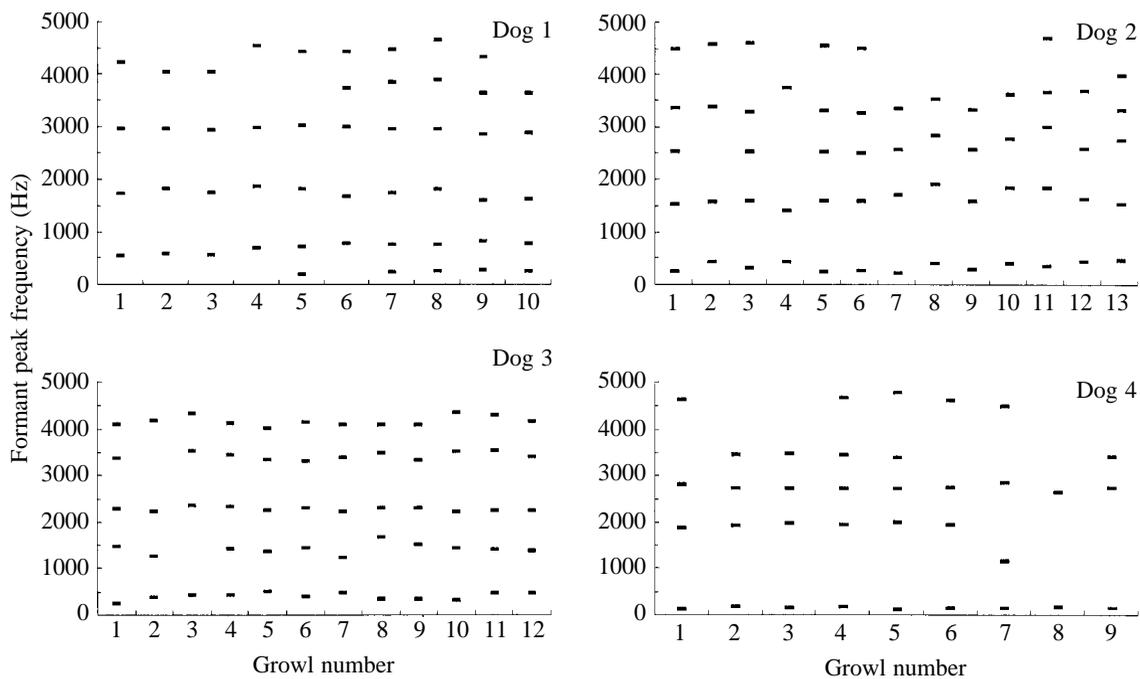


Fig. 4. Consistency of formant frequencies measured across different growls for four individual dogs. Each growl is represented by a set of formant frequencies. For instance, the first growl of dog 1 has four formant frequencies at approximately 500 Hz, 1800 Hz, 3000 Hz and 4200 Hz. Dog 1, dachshund; dog 2, Rottweiler; dog 3, Irish setter; dog 4, mongrel (mass 10 kg).

reliably an inverse correlation between  $F_1$  and  $VTL$ , but the difference between the findings for  $F_1$  and for  $D_f$  (which had the same sample size) invites some comments. One possibility for the higher variance in  $F_1$  is measurement error: LPC estimates of lower formants can be affected by low frequency or direct current poles due to radiation, room reverberation characteristics or other phenomena. Also, variability in the amount of low-frequency energy in the source signal will have the effect of decreasing the accuracy with which  $F_1$  can be measured (although this decrement would also be suffered by perceivers and is therefore not restricted to our algorithm).

Physiological explanations for the low reliability of  $F_1$  include the deviations from uniform vocal tract shape mentioned above (changes in mouth opening have a disproportionate effect on  $F_1$ , and potential nasal resonances might also play a role in obscuring  $F_1$ ). Another important factor may be the effect of tissue compliance. At low frequencies, the tissues in the walls of the vocal tract absorb energy, and tissue compliance therefore places a bound on  $F_1$  below which it cannot fall (approximately 180–200 Hz for humans; Fujimura and Lidqvist, 1970). Finally, variation in the end conditions (e.g. from a half-open to a fully open tube) will have a large effect on individual formant frequencies but no effect on formant spacing (see Introduction). In summary, an algorithm for estimating body size from formant locations (whether on a computer or by a perceiver) will be most reliable if it includes information from all formants and does not focus on a single formant (especially not on  $F_1$ , which for the reasons detailed above is likely to be the least reliable formant).

Although dog breeds exhibit high variability in skull shape (Miller, 1964), our results indicate that this variability does not obscure the strong negative relationship between formant dispersion and  $VTL$  and/or body mass. Formants therefore provide an indication of the body size of a growling dog. These results parallel those reported for rhesus macaques by Fitch (1997), who discusses further the implications of these relationships for acoustic communication. The artificial selection for brachycephalic or dolichocephalic skull shapes over the course of domestic breeding seems not to have had a pronounced effect on  $VTL$ /body size correlations. However, there are indications from other work that the shortening of the skull during domestication in some breeds has had an impact on vocal tract shape, particularly in the pharyngeal area. In extreme cases of head shortening (brachycephalic breeds such as bulldogs), a pharyngeal constriction is often observed and is known to cause respiratory problems (Wykes, 1991; Hendricks, 1992). More research is required to determine what acoustic effects might result from this skull and vocal tract shape diversity in domestic dogs.

Although the current work has focused only on formant production in canines, earlier work by Baru (1975) clearly demonstrated that dogs are also capable of perceiving formants with high accuracy. It would be quite surprising if dogs could perceive formants in synthetic human speech but not in their own species-specific vocalisations. Thus, we hypothesise that formants could play a role in canine communication. In particular, the information that formants convey in growls about

body size should be available to other listening dogs and may well play a role in social behaviour. Perceptual experimentation will be necessary to determine whether this is the case. For example, a dog could be given the choice of entering two darkened enclosures. From one of these, a synthetic growl stimulus with low, narrowly spaced formants would emanate, while a growl with widely spaced formants would come from the other. All other acoustic aspects of the stimuli would be identical. If dogs perceive formants as cues to body size, we would predict that the dogs would be more willing to enter the chamber with the 'small' dog (widely spaced formants).

In conclusion, we have found clear evidence that vocal tract length is correlated with body size in domestic dogs, despite the apparent variation in skull and vocal tract shape induced by selective breeding. As predicted by acoustic theory, vocal tract length was inversely correlated with the spacing between formant frequencies, which means that formant spacing provides a reliable cue to body size [ $\log(\text{body mass})$ ] in the sample of dogs studied here. Single formants were less reliably related to body size or vocal tract length than was the ensemble of formants, suggesting that acoustic estimates of body size will be most reliable when integrating information from multiple formants rather than just one. We found some deviations from the first-order approximation of a uniform tube, which probably result from the deviations from uniformity in vocal tract shape visible in the radiographs. However, further data are necessary to test this since our radiographs came from anaesthetised (not vocalising) dogs. Because dogs have been shown to perceive formants in synthetic human speech sounds, it seems plausible that they could perceive formant data in their own species-specific vocalisations, implying that formants could play a role in vocal communication between canids.

We thank Professor L. Brunberg for providing the opportunity to gather the data. We thank Günter Tembrock and Hanspeter Herzel for discussions and Brad Story, Michael Owren and two anonymous reviewers for helpful comments on an earlier version of the manuscript. This work was supported by a grant from the Land Berlin (NaFöG) to T.R. and by an NIH postdoctoral fellowship to W.T.F. (NIH grant IT32DC00038).

## References

- Andrew, R. J.** (1976). Use of formants in the grunts of baboons and other nonhuman primates. *Ann. N.Y. Acad. Sci.* **280**, 673–693.
- Baer, T., Gore, J. C., Gracco, L. C. and Nye, R. W.** (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *J. Acoust. Soc. Am.* **90**, 799–828.
- Baru, A. V.** (1975). Discrimination of synthesized vowels [a] and [i] with varying parameters (fundamental frequency, intensity, duration and number of formants) in dog. In *Auditory Analysis and Perception of Speech* (ed. G. Fant and M. A. A. Tatham), pp. 91–101. New York: Academic Press.
- Bogert, B. P. and Peterson, G. E.** (1957). The acoustics of speech. In *Handbook of Speech Pathology* (ed. L. E. Travis), pp. 109–173. New York: Appleton-Century-Crofts.

- Carterette, E., Shipley, C. and Buchwald, J.** (1979). Linear prediction theory of vocalization in cat and kitten. In *Frontiers in Speech Communication Research* (ed. B. Lindblom and S. Ohman), pp. 245–257. New York: Academic Press.
- Clark, A. R. and Brace, A. H.** (1996). (eds) *The International Encyclopedia of Dogs*. London: Batsford.
- Clutton-Brock, J.** (1995). Origins of the dog: domestication and early history. In *The Domestic Dog* (ed. J. Serpell), pp. 7–20. Cambridge: Cambridge University Press.
- Darwin, C.** (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.
- Dewson, J. H.** (1964). Speech sound discrimination by cats. *Science* **141**, 555–556.
- Fant, G.** (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.
- 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’ advertisement. *Am. J. Primatol.* **37**, 191–219.
- Fujimura, O. and Lindqvist, J.** (1970). Sweep-tone measurements of vocal tract characteristics. *J. Acoust. Soc. Am.* **49**, 541–548.
- Gaskell, C. J.** (1974). The radiographic anatomy of the pharynx and the larynx of the dog. *J. Small Anim. Pract.* **14**, 89–100.
- Gaunt, A. S. and Nowicki, S.** (1998). Sound production in birds: Acoustic and physiology revisited. In *Animal Acoustic Communication* (ed. S. L. Hopp, M. J. Owren and C. S. Evans), pp. 291–321. Berlin: Springer.
- Hauser, M. D., Evans, C. S. and Marler, P.** (1993). The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations. *Anim. Behav.* **45**, 423–433.
- Hendricks, J. C.** (1992). Brachycephalic airway syndrome. *Vet. Clin. North Am.: Small Animal Practice* **22**, 1145–1153.
- Herrmann, L.** (1890). Phonophotographische Untersuchungen. *Pflügers Arch.* **47/48**, 13.
- Hienz, R. D. and Brady, J. V.** (1988). The acquisition of vowel discrimination by nonhuman primates. *J. Acoust. Soc. Am.* **84**, 186–194.
- Hienz, R. D., Sachs, M. B. and Sinnott, J. M.** (1981). Discrimination of steady-state vowels by blackbirds and pigeons. *J. Acoust. Soc. Am.* **70**, 699–706.
- Kent, R. D.** (1978). Imitation of synthesized vowels by preschool children. *J. Acoust. Soc. Am.* **63**, 1193–1198.
- Kent, R. D.** (1979). Isovowel lines for the evaluation of vowel formant structure in speech disorders. *J. Speech Hearing Disorders* **44**, 513–521.
- Kneller, S. K.** (1994). The larynx, the pharynx and the trachea. In *Textbook of Veterinary Diagnostic Radiology* (ed. S. K. Kneller and D. E. Thrall), pp. 227–233. Philadelphia: W. B. Saunders.
- Lieberman, P. H., Klatt, D. H. and Wilson, W. H.** (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* **164**, 1185–1187.
- Markel, J. D. and Gray, A. H.** (1976). *Linear Prediction of Speech*. New York: Springer-Verlag.
- McComb, K. E.** (1988). Roaring and reproduction in red deer *Cervus elaphus*. PhD thesis, University of Cambridge.
- Miller, M. E.** (1964). *Anatomy of the Dog*. Philadelphia: Saunders.
- Morey, D. F.** (1992). Size, shape and development in the evolution of the domestic dog. *J. Archaeol. Sci.* **19**, 181–204.
- Nowicki, S.** (1987). Vocal tract resonances in oscine bird sound production: evidence from bird songs in a helium atmosphere. *Nature* **325**, 53–55.
- Ohala, J. J.** (1984). An ethological perspective on common cross-language utilization of F of voice. *Phonetica* **41**, 1–16.
- Owren, M. J.** (1990). Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*). *J. Acoust. Soc. Am.* **101**, 2951–2963.
- 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 Bernacki, R. H.** (1998). Applying Linear Predictive Coding (LPC) to frequency-spectrum analysis of animal acoustic signals. In *Animal Acoustic Communication* (ed. S. L. Hopp, M. J. Owren and C. S. Evans), pp. 129–162. Berlin: Springer.
- Owren, M. J. and Rendall, D.** (1997). An affect-conditioning model of nonhuman primate vocal signaling. In *Perspectives in Ethology*, vol. 12, *Communication* (ed. D. Owings, M. Beecher and N. Thompson), pp. 299–346. Town: Publisher.
- Owren, M. J., Seyfarth, R. M. and Cheney, D. L.** (1997). The acoustic features of the vowel-like grunt calls in chacma baboons (*Papio cynecephalusursinus*): Implications for production processes and function. *J. Acoust. Soc. Am.* **101**, 2951–2963.
- Peterson, G. E. and Barney, H. L.** (1952). Control methods used in a study of vowels. *J. Acoust. Soc. Am.* **24**, 175–184.
- Ryalls, J. H. and Lieberman, P.** (1982). Fundamental frequency and vowel perception. *J. Acoust. Soc. Am.* **72**, 1631–1634.
- Shipley, C., Carterette, E. C. and Buchwald, J. S.** (1991). The effect of articulation on the acoustical structure of feline vocalization. *J. Acoust. Soc. Am.* **89**, 902–909.
- Sommers, M. S., Moody, D. B., Prosen, C. A. and Stebbins, W. C.** (1992). Formant frequency discrimination by Japanese macaques (*Macaca fuscata*). *J. Acoust. Soc. Am.* **91**, 3499–3510.
- Stevens, K. N. and House, A. S.** (1955). Development of a quantitative description of vowel articulation. *J. Acoust. Soc. Am.* **27**, 484–493.
- Story, B. H., Titze, I. R. and Hoffman, E. A.** (1996). Vocal tract area functions from magnetic resonance imaging. *J. Acoust. Soc. Am.* **100**, 537–554.
- Suthers, R. A.** (1994). Variable asymmetry and resonance in the avian vocal tract: a structural basis for individually distinct vocalizations. *J. Comp. Physiol. A* **175**, 457–466.
- Titze, I. R.** (1994). *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice-Hall.
- Vila, C., Savolainen, P., Maldonado, J. E., Amorim, I. R., Rice, J. E., Honeycutt, R. L., Crandell, K. A., Lundberg, J. and Wayne, R. K.** (1997). Multiple and ancient origins of the domestic dog. *Science* **276**, 1687–1689.
- Westneat, M. W., Long, J. H., Hoese, W. and Nowicki, S.** (1993). Kinematics of bird song: functional correlation of cranial movements and acoustic features in sparrows. *J. Exp. Biol.* **182**, 147–171.
- Wykes, P. M.** (1991). Brachycephalic airway obstructive syndrome. *Probl. Vet. Med.* **3**, 188–197.