Published in Journal of Human Evolution, Vol. 48, Issue 1, 2005, p. 85-96
which should be used for any reference to this work

Vocal production mechanisms in a non-human primate: morphological data and a model

Tobias Riede[a,*], Ellen Bronson[b,1], Haralambos Hatzikirou[a],
Klaus Zuberbühler[c]

[a] *Institut für Theoretische Biologie, Humboldt-Universität zu Berlin, Invalidenstrasse 43, 10115 Berlin, Germany*
[b] *The Baltimore Zoo, Druid Hill Park, Baltimore, MD, USA*
[c] *School of Psychology, University of St. Andrews, St Andrews, Scotland, UK*

Abstract

Human beings are thought to be unique amongst the primates in their capacity to produce rapid changes in the shape of their vocal tracts during speech production. Acoustically, vocal tracts act as resonance chambers, whose geometry determines the position and bandwidth of the formants. Formants provide the acoustic basis for vowels, which enable speakers to refer to external events and to produce other kinds of meaningful communication. Formant-based referential communication is also present in non-human primates, most prominently in Diana monkey alarm calls. Previous work has suggested that the acoustic structure of these calls is the product of a non-uniform vocal tract capable of some degree of articulation. In this study we test this hypothesis by providing morphological measurements of the vocal tract of three adult Diana monkeys, using both radiography and dissection. We use these data to generate a vocal tract computational model capable of simulating the formant structures produced by wild individuals. The model performed best when it combined a non-uniform vocal tract consisting of three different tubes with a number of articulatory manoeuvres. We discuss the implications of these findings for evolutionary theories of human and non-human vocal production.

Keywords: Cercopithecinae; formants; vocal tract; motor pattern; *Cercopithecus diana*

---

* Corresponding author. Tel.: +49 30 20938652.

E-mail addresses: tobiasriede@web.de (T. Riede), kz3@st-and.ac.uk (K. Zuberbühler).

[1] Present Address: Smithsonian's National Zoological Park, 3001 Connecticut Ave., NW, Washington D.C. 20008, USA.

Introduction

Adult male Diana monkeys (*Cercopithecus diana*) utter predator-specific alarm calls in response to two of their predators, the crowned eagle (*Stephanoaetus coronatus*) and the leopard (*Panthera pardus*), hereafter 'eagle alarm calls' and 'leopard alarm calls'. Nearby monkeys respond appropriately to these calls, without having seen the predator, suggesting that the calls contain information about the external event, that is, the predator type present (Zuberbühler, 2000a, 2003). In previous work, we have documented that some prominent acoustic features of Diana monkey alarm calls are best conceptualised as formants, the resonance frequencies of the vocal tract (Riede and Zuberbühler, 2003a,b). Formants are the acoustic product of a series of bandpass filters that shape the sound primarily produced by actions of the vocal folds in the larynx and emitted from the mouth. The filtering process takes place in the entire vocal tract, which consists of the pharynx, as well as the laryngeal, nasal, and oral cavities. The cross-sectional diameters and length of the vocal tract determines the location of the formants' acoustic energy. During speech production in humans the shape of the vocal tract is changing constantly and rapidly, due to precise movements of the various articulators, such as the lips, tongue, jaw, or larynx.

For many years, it has been the default assumption that mammalian vocal tracts, including those of non-human primates, resemble a uniform or flared tube during vocalization (Lieberman, 1968; Lieberman et al., 1969; Shipley et al., 1991). In a uniform cylindrical vocal tract the resonance frequencies are expected to appear as odd numbered multiples of the first resonance, and all resonances are evenly spaced. The straight line in Fig. 1 represents the different combinations of the first and second formants, which would be expected under the uniform tube assumption for different vocal tract lengths and of a given diameter. More recently, various studies challenged this view, by suggesting that some animal vocalisations are the product of non-uniform vocal tracts (e.g. Owren et al., 1997). For example, we have previously demonstrated that the location of the first (F1) and second (F2) formant in Diana monkey alarm calls cannot be explained by a uniform vocal tract but must be the result of a more complex vocal tract geometry (Riede and Zuberbühler, 2003b, see dots in Fig. 1).

Second, in previous work we have shown that Diana monkey leopard and eagle alarm calls differ most prominently in the fine structure of the formants albeit showing very little variability in the fundamental frequency (Riede and Zuberbühler, 2003a,b). In particular, the first and second formant (higher formants are only rarely detectable in the spectrum) of Diana monkey leopard alarm calls exhibit a threefold stronger decrease in frequency compared to the spectrographically less modulated formants in the eagle alarm calls (Fig. 2), suggesting that these animals are able to adjust the shape of their vocal tracts during vocal behaviour independent of the laryngeal source.

These observations have lead to the hypothesis that Diana monkeys possess some control over the shape of their vocal tract, and that they employ this ability to communicate about some important environmental events. In humans, acts of active changes in vocal tract shape during vocalization are termed articulation, the result of movement of the tongue, the mandible, the lips, the larynx and so on. For example, narrowing the lip aperture causes a lowering of the formants (Stevens, 1999, p. 284), an articulatory manoeuvre that can also be observed in Diana monkeys, especially when producing leopard alarm calls (Riede and Zuberbühler, 2003 a,b). Evidence for articulation caused by mouth opening also exists from studies of the vocal behaviour of rhesus monkeys (*Macaca mulatta*, Hauser et al., 1993; Hauser and Schön-Ybarra, 1994) and domestic cats (*Felis catis*, Shipley et al., 1991). A second articulatory manoeuvre that could explain the formant behaviour in Diana monkey alarm calls is the lowering of the larynx (Story et al., 1996; Fitch, 2000; Fitch and Reby, 2001). Acoustically, lowering the larynx causes an equal lowering of all formants due to vocal tract elongation. As both articulatory manoeuvres have previously been described in animal communication, it is likely that they are promising candidates to explain the formant acoustics of Diana monkey alarm calls as well. Yet, the relative amount of variation explained by them and whether or not the

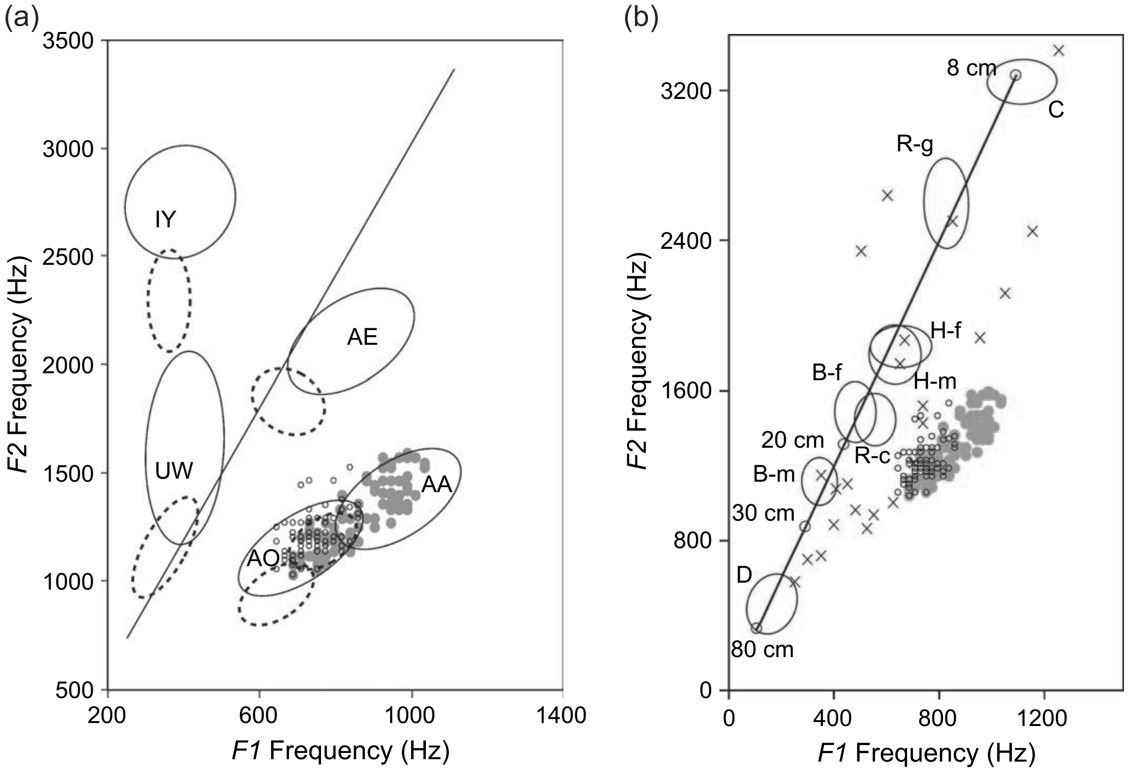

Fig. 1. First (F1) and second (F2) formant chart for human vowels and for animal vocalisations. Fig. 1a. Frequency values of F1 and F2 from 50 eagle alarm calls (black open circles; measured in 10 calls from each of 5 adult male Diana monkeys) and for 50 leopard alarm calls (gray shaded area; measured in 10 calls from each of 5 adult male Diana monkeys). In each call F1 and F2 were measured at the beginning of the call, in the middle and at the end. Formant data were taken from Riede and Zuberbühler (2003b). The monkey data are plotted within the American-English F1/F2 vowel space mapping (redrawn from data of Lee et al., 1999). Human formant data are for adult males (dashed ellipses) and for 10-12 year old children (solid ellipses) (Table 2 and Fig. 4 in Lee et al., 1999). Human vowel data were recorded from subjects saying the target words *bead* /IY/, *bat* /AE/, *pot* /AA/, *ball* /AO/, *boot* /UW/ (Lee et al., 1999). The straight line represents the expected F1 and F2 values for uniform/cylindrical tubes of certain length, according to the equation $Fn = (2n - 1)c/4L$; (n = n-th formant, c = speed of sound; L = length of the vocal tract). Fig. 1b. Same formant (F1, F2) data for Diana monkeys as in Fig. 1a. The open ellipses represent ranges for formant data for male (B_m) and female (B_f) chacma baboon grunts (redrawn from Owren et al., 1997; Rendall et al., 2004), and for human laughter (man: H_m, woman: H_f) (redrawn from Bachorowski and Owren, 2001), for rhesus monkey coo calls (R_c) and grunt calls (R_g) (estimated after data from Fitch 1997; Rendall et al., 1998) and for male Red Deer roars (D) (estimated after data from Fitch and Reby, 2001). The straight line, as in Fig. 1a, represents the expected F1 and F2 values for uniform/cylindrical tubes of certain lengths. 4 lengths (8 cm, 20 cm, 30 cm and 80 cm) are indicated by open circles. Note that axis scaling is different in the two diagrams. Formants in domestic dog growls (crosses) (data from Riede and Fitch, 1999) show substantial variability because each data point represents a different breed which show high variability in head shape and thus vocal tract length (from Yorkshire terrier with 8 cm vocal tract length to Rottweiler with approx. 23 cm vocal tract length).

acoustic effects they generate are sufficient remains unknown and needs to be investigated.

Here, we provide measurements of vocal tract length and shape from three anaesthetised Diana monkeys using lateral radiographs. We additionally present results of the dissection of one male specimen. We then use these findings to generate the most likely computational model of a Diana monkey vocal tract to test how much formant change each of several changes in vocal tract geometry can account for, and whether or not a single change is sufficient for producing the formant characteristics observed in the natural alarm calls.

Methods

Radiography

Two adult males and one adult female Diana monkey (*C. diana diana*) from the Baltimore Zoo

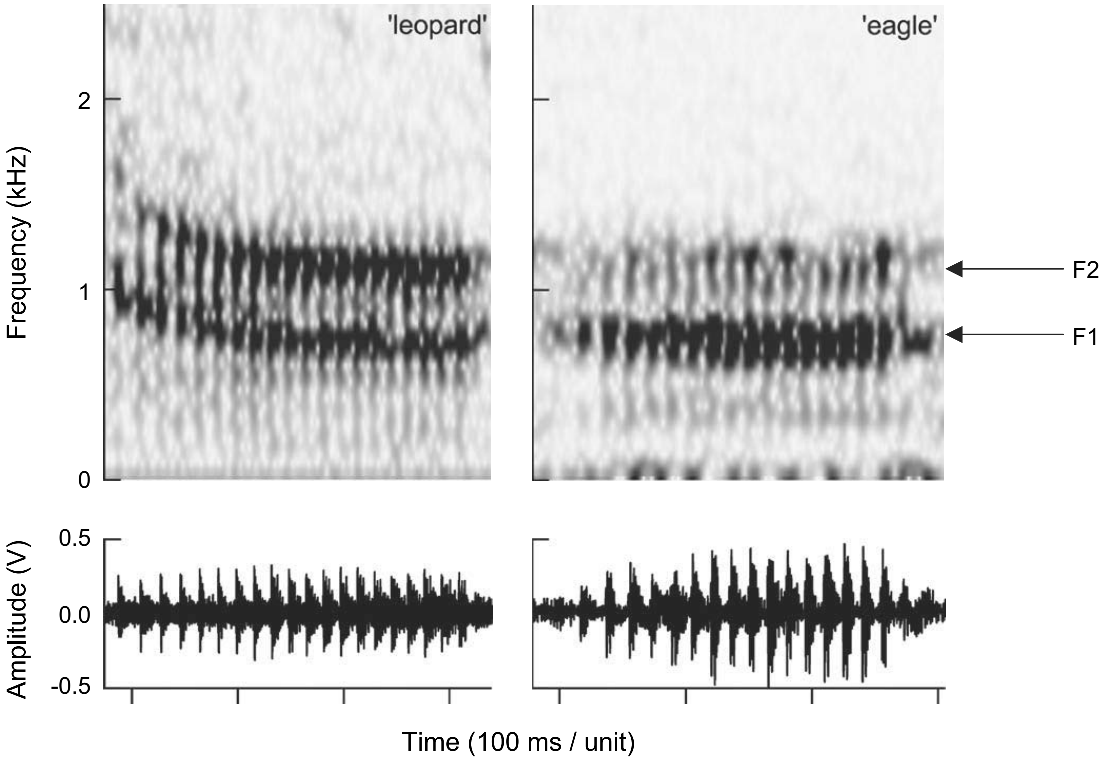

Fig. 2. Spectrogram and time series of a leopard and an eagle alarm call uttered by a male Diana monkey. Note that the first (F1) and the second (F2) formant behave differently in the two calls. There is a downward modulation at the beginning of the leopard call but not in the eagle alarm call.

served as subjects. In November 2002, the animals were anaesthetised with Ketamine (9.5 mg/kg) or a combination of Tiletamine-Zolazepam (4.3-5 mg/kg) via intramuscular injection as part of their routine annual medical examination. Each animal was then placed in lateral recumbence on a radiographic table. Lateral images of the head-neck region were taken, one with a relaxed head position and another with a straightened head position. A one-centimetre lead reference square was positioned at the midsaggital level of the head for calibration. The vocal tract length (VTL) was determined from tracings of the X-ray images using a Microtek MRS-600Z scanner and Scion Image Beta 4.0 (www.scioncorp.com) for measuring. Image clarity was sufficient to delineate the outlines of the oral vocal tract. The midpoint of the thyrohyoid bone, which was always visible, was used as the point of origin for all vocal tract length measurements. The thyrohyoid appeared on radiographs just cranial to the glottis. Thus, our estimate was slightly smaller than the actual vocal tract length but in a manner consistent across individuals. We then drew 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 (Fig. 3). We used this line as an estimate of the length of the vocal tract using the calibration squares as a reference. Skull length was defined as the distance between the front of the incisors and the external occipital protuberance of the occipital bone. Studies in dogs have shown that these measurements are reliably replicable (Riede and Fitch, 1999). We estimated the shape of the vocal tract by measuring the rostroventral distance along the laryngeal, pharyngeal, and oral cavities, from the glottis to the lips.

Dissection

Portions of a carcass from an adult male Diana monkey were obtained from the Baltimore Zoo.

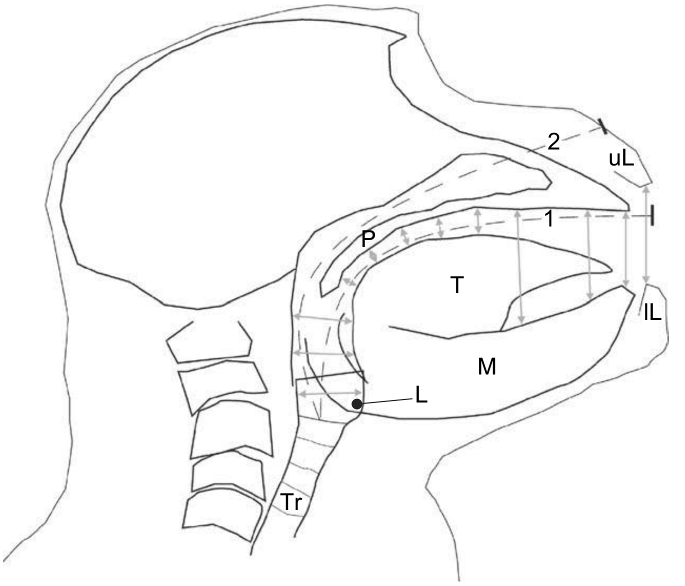

Fig. 3. Schematic drawing of the head-neck region of a Diana monkey with details as seen in the dissection as well as in the lateral x-ray. T - tongue, Tr - trachea, uL - upper lip, lL - lower lip, L - larynx, P - palate, dashed line 1 - oral vocal tract length, dashed line 2 - nasal vocal tract length, arrows indicate the dorso-ventral distances of the oral vocal tract.

The specimen was collected in August 2003 and was deep-frozen within 36 hours post-mortem. Unfortunately, the thorax had been severely damaged during pathological examination, thus the dimension of the airsac cavity could not be completely investigated. The specimen was then cut in the median plane and each half was photographed.

Modelling the vocal tract

A popular vocal tract model used for simulation of vowel sounds is the "Wave-Reflection Analogue" (or "Wave Digital Filter") (Smith, 1992, 1998). The model divides the three-dimensional vocal tract into a finite number of cylinders of certain length. Reflection coefficients are calculated at each cylinder junction, based on the relative areas of adjoining sections. In this model, waves propagate through the system as a function of the reflection coefficients, which determine the incident and reflected components of the pressure waves at each junction at each step in time. This has the effect of transporting a wavefront from section to section. We chose the commonly used practice of setting the reflection coefficients at the sound source and at the mouth opening at 1 (e.g. Olesen, 1995, p. 29). The Wave-Reflection-Analogue is an attractive method for acoustic modelling of the vocal tract because computations are performed serially in time-synchrony with the acoustic wave propagation (Story et al., 1996). Computations are efficient because the equations describing the wave propagation take on a digital filter structure in their final form.

Results

Radiography and dissection

Vocal tract length

The oral vocal tract lengths in the three radiographed individuals were 9.7 and 10.5 cm (males A and B) and 8.9 cm (female). The oral vocal tract length in the dissected individual (male B) was 10.1 cm (Table 1).

Vocal tract shape

During vocalization, the animals' head is lifted so the mouth points forward. We observed that in this head position, the epiglottis was touching the soft palate. The vocal tract had a non-uniform shape caused by a major constriction separating the front oral cavity from the laryngo-pharyngeal

Table 1
Morphological data of study animals

name/sex	body mass (kg)	age (years)	oral vocal tract length (cm)	nasal vocal tract length (cm)	skull length (cm)	oral vocal tract length (cm) - dissected
Digit/male B	8.1	>20	10.51	11.59	12.51	10.1
Dennis/male A	9.3	8	9.71	9.88	11.71	N/A
Di/female	4.6	17	8.89	9.03	11.32	N/A

cavity. The constriction which is built by the tongue and the palate, extends over 2 to 4 cm. Fig. 4 summarises radiograph measurements of rostroventral distances in the vocal tract as a function of the distance from the glottis.

Larynx

The length of the vocal folds in the freshly dissected specimen was 11 mm on both sides. Cercopithecines possess airsacs branching off from the larynx, located subcutaneously in the throat and ventral thorax area, and these structures are very likely to affect the acoustic output (Gautier, 1971). In our specimen the air sac projected medially between the thyroid cartilage and the epiglottis toward the subcutaneous area of the neck. However, because of severe damage to the thorax of the specimen we were not able to reconstruct the dimensions of the air sac.

Modelling the vocal tract

Our goal was to simulate the formant behaviour as observed in natural alarm calls, particularly to account for the formant transitions that characterise the leopard alarm calls. Is a single vocal tract adjustment, such as the movement of larynx, mouth, or jaw, sufficient to achieve the naturally observed formant behaviour?

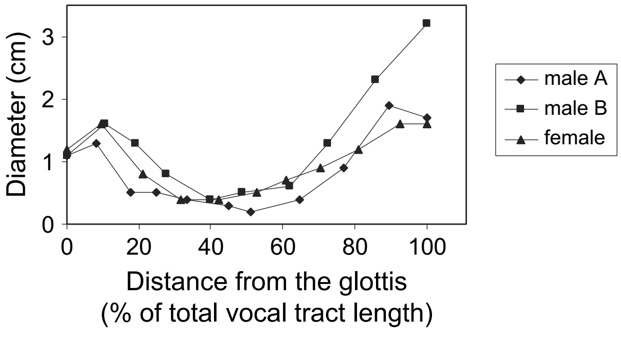

Fig. 4. Vocal tract shape measured as dorso-ventral distance of the oral vocal tract (as indicated in Fig. 3) in a distance from the glottis, measured in 3 monkeys, 2 males and 1 female.

Vocal tract non-uniformity

Formants in Diana monkey alarm calls are the likely product of a non-uniform vocal tract (Riede and Zuberbühler, 2003a,b), in the simplest case a two-tube model. A two-tube vocal tract can produce vowels with a low second formant, like the /a/ vowel, as demonstrated for instance by Stevens (1972). The lowest F2 values are achieved if both segments are of the same length. However, in our case a two-tube model, created by a narrow laryngo-pharyngeal tube attached to a wider oral tube, was unable to generate the formant patterns in the natural Diana monkey alarm calls, mainly due to the insufficient lowering of the second formant (see calculations in the next paragraph). Interspersing a third narrow connecting tube between the laryngo-pharyngeal back tube and the frontal oral tube is known to lower the second formant dramatically while the first formant is mildly increased (e.g. Stevens, 1972). Finally, adding a small fourth tube will allow for separate lip aperture and variable mouth opening, hereby accommodating natural observations.

The formant values for (a) natural alarm calls and (b) different tube dimensions of the model are given in Table 2. Our anatomical results indicated a VTL in adult Diana monkeys of about 10 cm. However, in a uniform cylindrical vocal tract the first formant would be expected to be much higher than measured in Diana monkey alarm calls, while the second formant would be much lower (Fig. 1). A 10 cm uniform *1-tube model* (1 cm diameter)

Table 2
Formant values in natural calls and in the computational modal

Natural vocalisations (Leopard alarm calls)	F_1 (Hz)	F_2 (Hz)
beginning of call	942	1407
middle of call	783	1199
end of call	795	1227

#	tube dimensions (diameter/length in cm)	F_1 (Hz)	F_2 (Hz)
1	A 1.0/10.0 - B 1.0/0.0 - C 1.0/0.0 - D 1.0/0.0	754	2264
2	A 1.0/5.0 - B 1.0/0.0 - C 3.0/5.0 - D 1.0/0.0	1391	2110
3	A 1.0/0.88 - B 0.3/3.0 - C 2.4/6.0 - D 2.4/0.0	949	1489
4	A 1.0/0.88 - B 0.3/3.0 - C 2.4/5.75 - D 1.5/0.25	945	1403
5	A 1.0/0.88 - B 0.3/3.0 - C 2.4/5.75 - D 0.71/0.25	950	1390
6	A 1.0/1.27 - B 0.3/3.0 - C 2.4/5.75 - D 0.71/0.25	805	1370
7	A 1.0/1.27 - B 0.3/3.0 - C 2.0/5.75 - D 0.71/0.25	782	1194
8	A 1.0/1.23 - B 0.3/3.0 - C 2.0/5.75 - D 0.76/0.25	795	1227

Formant values for natural leopard alarm calls were determined from monkey vocalisations recorded in the Taï forest, Ivory Coast (see Riede and Zuberbühler, 2003b). Formant calculations are based on different tube model dimensions as well as mouth openings. tube dimensions: #1 - uniform tube 1 cm diameter and 10 cm length; #2 to #5 - multi-tube-approximations (tubes A to D see Fig. 5) with respective dimensions (diameter/length in cm); #6 to #8 - articulatory manoeuvres.

resulted in the following formants: F1 = 754 Hz, F2 = 2264 Hz (Table 2, #1), that is, a 23% lower F1, and a 61% higher F2 compared to the measurements at the beginning of natural leopard alarm calls (Table 2). A 10 cm non-uniform *2-tube model* generates the following formants: F1 = 1391 Hz, F2 = 2110 Hz (Table 2, #2), that is now a 48% higher F1 and still a 50% higher F2 compared to the measurements at the beginning of natural leopard alarm calls. A 10 cm non-uniform *3-tube model* generates substantially smaller deviations from the expected values (F1 = 949 Hz, F2 = 1489 Hz), although F2 still deviates by approximately 6% (Table 2, #3). A 10 cm non-uniform *3-tube model* with a very short 4th tube, achieved by cutting off a short segment from tube C, to simulate the lip aperture, lead to formant values that resembled the natural conditions most closely at the beginning of the leopard call (F1 = 945 Hz, F2 = 1403 Hz, Table 2, #4).

Our anatomical and observational data combined with the physics of tube resonance acoustics therefore suggest a 3-tube model with a short fourth tube attached to the end as the best model to explain Diana monkey vocal behaviour (Fig. 5).

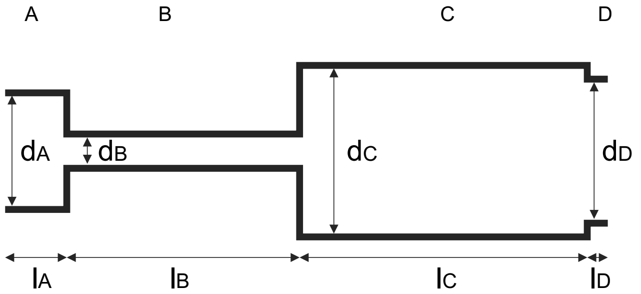

Fig. 5. Schematic drawing of the 3-tube-approximation of the monkey's vocal tract indicating the lengths (l_A, l_B, l_C, l_D) and diameter dimensions (d_A, d_B, d_C, d_D) used for the calculation. Tube D represents the mouth opening, and is either as wide as tube C in diameter ($d_C = d_D$) or narrows down ($d_C < d_D$) in order to simulate the closing of the lip aperture.

Articulation

Leopard alarm calls are the likely product of articulation, as evidenced by the two formants undergoing a dramatic downward modulation during the first half of a call (Fig. 2) (Riede and Zuberbühler, 2003b). In our model, closing the mouth corresponded with a lowering of the tube D diameter, leading to a decrease in the first two formant frequencies. However, with this manoeuvre the second formant decreased more strongly than the first formant, unlike in the natural call (Table 2, #5). This suggested that Diana monkeys engaged in additional articulatory manoeuvres while producing leopard alarm calls. To accommodate for these findings we added the following two articulatory manoeuvres to the model which brought its formant values to a very good match with those of the natural calls: (a) elongation of tube A by about 4% (corresponding to a laryngeal lowering of about 4 mm; see Table 2, #6), (b) narrowing of tube C by about 17% (corresponding to a rising of the mandible by about 4 mm, leading to narrowing of the frontal cavity accompanying mouth closing; see Table 2, #7). Finally, in leopard alarm calls the first and second formants increase slightly from the middle of the call to the end of the call (Table 2). Our model could simulate this effect by slight adjustments in the mouth opening and the larynx position (Table 2, #8).

Discussion

In this study we were interested in the vocal tract structure and motor patterns underlying alarm call production in wild Diana monkeys. We have collected a set of anatomical data, which allowed us to estimate the length and shape of the vocal tract in this species with satisfactory accuracy. We used these measurements in combination with a number of theoretical considerations to construct a simple computational model of the Diana monkey vocal tract that simulated the natural calls remarkably well. The model consisted of a three-tube non-uniform vocal tract, with an additional short forth tube, which could engage in three types of simultaneous articulatory manoeuvres to change the overall shape of the vocal tract: mouth closing, jaw movement, larynx movement.

Vocal tract shapes change considerably during all sorts of activities, such as food intake (see for instance Lieberman and McCarthy, 1999) and still X-ray images from dead or anaesthetized animals may not necessarily capture the entire range of vocal tract mobility during vocalization. To address questions of motor patterns underlying vocal behaviour, the ideal approach is to observe the movement of the articulators in the animal's vocal tract during vocalisation, using cine-radiographic or related techniques (e.g. Fitch, 2000; Riede et al., 2004). For a number of reasons this approach was unsuitable here. First, there are no other species known which produce these acoustic features and could serve as a comparison. Second, Diana monkeys are a highly endangered species, which excludes any kind of risky or invasive research. Computational modelling thus may be the best approach to identify likely mechanisms of vocal production and articulation.

Vocal tract length

Earlier interpretations of the formant patterns in Diana monkeys relied on simple skull measurements (Riede and Zuberbühler, 2003b). In this study, we were able to provide more accurate data using radiography and dissection. The vocal tract length in 3 anaesthetised and 1 dead Diana monkeys ranged between 9 and 11 cm. The true vocal tract length during vocalisation might be slightly different due to protrusion of the lips or temporary up-and-down-movements of the larynx. Laryngeal lowering during vocalisation has been reported in piglets (*Sus scrofa*), domestic dogs (*Canis familiaris*), goats (*Capra hircus*), cotton-top tamarins (*Saguinus oedipus*) (Fitch, 2000), and red deer (*Cervus elaphus*; Fitch and Reby, 2001). We therefore concluded that changes in vocal tract length, caused by mild larynx lowering, are a mechanism likely to take place during Diana monkey alarm calling.

Vocal tract shape

Non-uniformity is an important characteristic of the human vocal tract (e.g. Story et al., 1996). It has been a tacit assumption that non-human primate vocalisations are the result of a uniform tube-like or a flared tube-like vocal tract, resulting in F1 and F2 to be situated along the straight line in Fig. 1. Our data from Diana monkeys show that this is not true for all non-human primates. Our anatomical data indicate that the larger frontal oral cavity and the smaller laryngo-pharyngeal cavity are separated by a constriction, the *Isthmus faucium*, which is built by the tongue and the palate. This kind of non-uniformity has been reported in a baboon (Zhinkin, 1963, p. 162), in cotton-top tamarins (Fig. 2 in Fitch, 2000), and it is further supported by formant data in domestic dogs (Riede and Fitch, 1999; Fig. 1b this study). Non-uniformity of this kind is not reported for the vocal tracts of cats and pigs (cats: Shipley et al., 1991; pigs: Fitch, 2000).

Speech production

In 3-month old infants, the larynx and the root of the tongue begin to descend into the pharyngeal cavity (Laitman et al., 1977; George, 1978). As a consequence, the tongue becomes increasingly mobile and serves as the anterior wall of the pharynx, an arrangement that permits prominent alternations in the cross-sectional area and shape of the vocal tract. In older infants and adult humans, temporary deviations from uniformity are normal during vowel production (Fig. 1b). The formant characteristics of the vowels are the result of unique shapes of the vocal tract (Titze, 1994; Story et al., 1996). In non-human primates, the larynx is normally positioned higher in the pharynx (Keleman, 1948, 1969; Laitman et al., 1977) although a laryngeal descence during early ontogeny has been described in chimpanzees (Nishimura et al., 2003). This position can be altered by specific muscles (Negus, 1949), but non-human primates are thought to make only restricted use to change the shape of their vocal tracts during vocalization (Lieberman and McCarthy, 1999), similar to humans when producing a *schwa* vowel (Lieberman, 1968). Here we demonstrated that the vocal tract in Diana monkeys during leopard alarm call production is likely to undergo a complex motor pattern not described for a non-human primate before.

Modelling the Diana monkey vocal tract

As such, the model was able to explain two acoustic 'anomalies' in the Diana monkey alarm calls, that is, the close proximity of the first and second formant in both alarm call types (explained by a 3-tube approach) and the prominent formant downward modulation in the leopard alarm calls (explained by the simultaneous movement of at least 3 articulators).

It was assumed earlier that the non-human mammalian vocal tract cannot create vowels as /u/ or /o/ because the production of these vowels requires an extreme constriction in the vocal tract. As mentioned above, studies in other non-human primates and the present study suggest that this constriction is not uniquely human. Moreover, a modelling approach of the Neandertal vocal tract suggests that this early hominid was equally capable of producing the full range of vowels due to a similar constriction (Boe et al., 2002), despite the fact that their vocal tract were characterised by a higher laryngeal position than in the modern humans.

Our model showed that one single change of vocal tract shape alone could not account for the formant lowering in leopard alarm calls. It was the combination of an initially non-uniform vocal tract and the several simultaneous articulatory manoeuvres that provided the best match between simulated and real data. Direct observations of vocalising Diana monkeys suggested that males not only narrow the lip aperture but also raise the mandible. Our model incorporated this fact by simulating changing the lip aperture and mandible movement (i.e. narrowing diameter of tubes C and D; Fig. 5). It also addressed the effect of larynx descending and its effect on lowering the first and second formant.

Modelling approaches rely heavily on anatomical information in order to produce meaningful output. This is because the same or similar formant patterns can be obtained with a range of different vocal tract configurations (see for example discussion in Espy-Wilson et al., 2000). In this study, we only incorporated assumptions that were based on empirical evidence provided by our anatomical findings or on facts provided by earlier investigations.

Three important questions were not addressed in this study but await further investigation: (a) the possibility of nasalization and (b) the acoustic role of air sacs. The potential acoustic effects have been discussed elsewhere (Riede and Zuberbühler, 2003b). (C) The possibility of tongue movement has been demonstrated during swallowing (Hiiemae et al., 1995) but is unknown to take place during vocalization. Before incorporating these structures into models further anatomical investigations will be necessary.

Evolutionary implications

At least three implications for theories of speech evolution emerge from this study. First, a recent study on Neanderthal vocal tracts suggests that a low larynx alone is not a sufficient anatomical prerequisite for producing the full range of vowels (Boe et al., 2002). Our findings on Diana monkey vocal behaviour corroborate this conclusion. Second, articulatory effects within a single utterance are very important in human speech, generated by the rhythmic mandible-generated open-close alternation of the mouth (MacNeilage and Davis, 2000). Modern languages share certain patterns of consonant-vowel co-occurrences, the result of biomechanical constraints associated with mandibular movements (MacNeilage, 1998;

MacNeilage and Davis, 2000). Our data show that these patterns are also present in non-human primate vocal production. Diana monkeys appear to engage in the same rhythmic mandible-generated open-close alternation of the mouth during a leopard alarm call bout. Our modelling data suggest that articulation in the leopard alarm calls takes place by narrowing the lip aperture accompanied with raising the mandible. Since leopard alarm calls are uttered in bouts, the basic pattern would be raising (during vocalisation) and lowering (between calls) of the mandible. Third, two utterances in humans have been considered most similar to non-human vocalisation, because of their non-verbal characteristics, laughing and infant crying. Laughing is acoustically a highly variable utterance (Bachorowski and Owren, 2001) without consonants, which demonstrated a persistent lack of articulation effects in supralaryngeal filtering (Bachorowski and Owren, 2001; Owren and Bachorowski, 2003). Infant crying is acoustically very variable, but the acoustic variability can be ascribed to source characteristics (Zeskind and Collins, 1987; Mende et al., 1990), suggesting that articulation plays no major role (Lieberman et al., 1971). Diana monkey leopard alarm calls do not resemble either of these human utterances but overlap substantially with the /a/ vowel and the /o/ vowel F1/F2-range of a 10 to 12 years old child with a similar vocal tract length (Fig. 1a).

In most other non-human species, however, formant frequencies do not vary appreciably within single calls (e.g. Chacma baboons *Papio hamadrayas ursinus*: Owren et al., 1997). However, articulatory effects within a single utterance have been described for domestic cats and for red deer. In domestic cats, jaw movement appear to be the most powerful articulator in 'meow' vocalisations, responsible for the formant change within a single call (Shipley et al., 1991). In red deer, formant variability has been ascribed to extreme lowering of the larynx during a single call (Fitch and Reby, 2001). In this respect Diana monkeys are a first example of animal vocalisations where acoustic and anatomical data suggest non-uniformity of the vocal tract. Some investigators have argued that primates and non-primates can prominently manipulate the position of the lips (Hauser, 1992; Hauser et al., 1993; Hauser and Schön-Ybarra, 1994), larynx (Fitch, 2000; Fitch and Reby, 2001) and tongue (Hiiemae et al., 1995), thus changing the length, shape, and thus presumably resonances of the vocal tract. However, available formant data in other species suggest more or less uniform vocal tract conditions (Fig. 1b). In sum, our study shows that a number of vocal tract adaptations important in human speech production are also present in a non-human primate. In Diana monkeys, these are a non-uniform vocal tract and three different articulation mechanisms caused by movement of the mandible, the lips, and the larynx. Moreover, Diana monkeys utilise these mechanisms to produce two types of vocalisations, which function to communicate important events in the environment.

One remaining question about the Diana monkey alarm call system is whether formant variability is perceptually relevant for conspecific recipients. So far, playback experiments have shown that leopard and eagle alarm calls are discriminated by conspecifics (Zuberbühler, 2000a), by other monkeys such as Campbell's monkeys (*Cercopithecus campbelli*) (Zuberbühler, 2000b), and even by birds (Rainey et al., 2004a,b). It remains unknown which acoustic parameters recipients attend to when making these discriminations, although formants are very likely candidates (Hienz and Brady, 1988; Owren, 1990a,b; Sommers et al., 1992).

Acknowledgement

We thank the Baltimore Zoo Chimp Forest and hospital staff for their support in collecting data and performing examinations on the three captive animals. The investigation of three Diana monkeys was approved by the Baltimore Zoo Institutional Animal Care and Use Committee. We also thank Jack Bradbury, Cornell University, for technical support and Sue Anne Zollinger, Indiana University, for help in the preparation of the figures. T. Riede was supported by a fellowship within the Postdoctoral Programme of the German Academic Exchange Service (DAAD).

We are most grateful for the comments of Michael Owren, Brad Story and two reviewers.

References

Bachorowski, J.A., Owren, M.J., 2001. Not all laughs are alike: voiced but not unvoiced laughter elicits positive affect in listeners. Psychol. Sci. 12, 252–257.

Boe, J.-L., Heim, J.-L., Honda, K., Maeda, S., 2002. The potential Neandertal vowel space was as large as that of modern humans. J. Phonetics 30, 465–484.

Espy-Wilson, C.Y., Boyce, S.E., Jackson, M., Alwan, A., 2000. Acoustic modeling of American English /r/. J. Acoust. Soc. Am. 108, 343–356.

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., 2000. The phonetic potential of nonhuman vocal tracts: comparative cineradiographic observations of vocalizing animals. Phonetica 57, 205–218.

Fitch, T.W., Reby, D., 2001. The descended larynx is not uniquely human. Proc. R. Soc. Lond. B 268, 1669–1675.

Gautier, J.P., 1971. Etude morphologique et fonctionnelle des annexes extra-laryngées des cercopithecinae; liason avec les cris d'espacement. Biol. Gabon. 7, 229–267.

George, S.L., 1978. A longitudinal and cross-sectional analysis of the growth of the postnatal cranial base angle. Am. J. Phys. Anthropol. 49, 171–178.

Hauser, M., 1992. Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: a learned mode of production? J. Acoust. Soc. Am. 91, 2175–2179.

Hauser, M.D., Evans, C.S., Marler, P., 1993. The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. Anim. Behav. 45, 423–433.

Hauser, M.D., Schön-Ybarra, M., 1994. The role of lip configuration in monkey vocalization: experiments using xylocaine as a nerve block. Brain Lang. 46, 232–244.

Hienz, R.D., Brady, J.V., 1988. The acquisition of vowel discrimination by nonhuman primates. J. Acoust. Soc. Am. 84, 186–194.

Hiiemae, K.M., Hayenga, S.M., Reese, A., 1995. Patterns of tongue and jaw movement in a cinefluorographic study of feeding in the macaque. Arch. Oral. Biol. 40, 229–246.

Keleman, G., 1948. The anatomical basis of phonation in the chimpanzee. J. Morphol. 82, 229–257.

Keleman, G., 1969. Anatomy of the larynx and the anatomical basis of vocal performance. Chimpanzee 1, 165–186.

Laitman, J.T., Crelin, E.S., Conlogue, J., 1977. The function of the epiglottis in monkey and man. Yale J. Biol. Med. 50, 43–48.

Lee, S., Potamianos, A., Narayanan, S., 1999. Acoustics of children's speech. Developmental changes of temporal and spectral parameters. J. Acoust. Soc. Am. 105, 1455–1468.

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

Lieberman, P., Klatt, D.H., Wilson, W.A., 1969. Vocal tract limitations on the vocal repertoires of rhesus monkey and other non-human primates. Science 164, 1185–1187.

Lieberman, P., Harris, K.S., Wolff, P., Russell, L.H., 1971. Newborn infant cry and non-human primate vocalizations. J. Speech Hear. Res. 14, 718–727.

Lieberman, D.E., McCarthy, R.C., 1999. The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. J. Hum. Evol. 36, 487–517.

MacNeilage, P.F., 1998. The frame/content theory of evolution of speech production. Behav. Brain Sci. 21, 499–511.

MacNeilage, P.F., Davis, B.L., 2000. On the origin of internal structure of word forms. Science 288, 527–531.

Mende, C., Herzel, H., Wermke, K., 1990. Bifurcations and chaos in newborn infant cries. Phys. Lett. 145 (A), 418–424.

Negus, C., 1949. The comparative anatomy and physiology of the larynx. Grune and Stratton, New York.

Nishimura, T., Mikami, N., Suzuki, J., Matsuzawa, T., 2003. Descent of the larynx in chimpanzee infants. Proc. Natl Acad. Sci. 100, 6930–6933.

Olesen, M., 1995. *A speech production model including the nasal cavity*. Aalborg (DK).

Owren, M.J., 1990a. Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*). I. Natural calls. J. Comp. Psychol. 104, 20–28.

Owren, M.J., 1990b. Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*). II. Synthetic calls. J. Comp. Psychol. 104, 29–40.

Owren, M.J., Seyfarth, R.M., Cheney, D.L., 1997. The acoustic features of vowel-like grunt vocalization in chacma baboons (*Papio cyncephalus ursinus*). Implications for production processes and function. J. Acoust. Soc. Am. 101, 2951–2963.

Owren, M.J., Bachorowski, J.A., 2003. Reconsidering the evolution of nonlinguistic communication: the case of laughter. J. Nonverbal Behav. 27, 183–200.

Rainey, H.J., Zuberbühler, K., Slater, P.J.B., 2004a. Hornbills can distinguish between primate alarm calls. Proc. R. Soc. Lond. B 271, 755–759.

Rainey, H.J., Zuberbühler, K., Slater, P.J.B., 2004b. The responses of black-casqued hornbills to predator vocalisations and primate alarm calls. Behaviour 141, 1263–1277.

Rendall, D., Owren, M.J., 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.

Rendall, D., Owren, M.J., Weerts, E., Hienz, R.D., 2004. Sex differences in the acoustic structure of vowel-like grunt vocalization of baboons and their perceptual discrimination by baboon listeners. J. Acoust. Soc. Am. 115, 411–421.

Riede, T., Fitch, T., 1999. Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). J. Exp. Biol. 202, 2859–2867.

Riede, T., Zuberbühler, K., 2003a. Pulse register phonation in Diana monkey alarm calls. J. Acoust. Soc. Am. 113, 2919–2926.

Riede, T., Zuberbühler, K., 2003b. The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. J. Acoust. Soc. Am. 114, 1132–1142.

Riede, T., Beckers, G., Blevins, W., Suthers, R., 2004. Inflation of the esophagus and vocal tract filtering in Ring doves. J. Exp. Biol 207, 4025–4036.

Shipley, C., Carterette, E.C., Buchwald, J.S., 1991. The effect of articulation on the acoustical structure of feline vocalization. J. Acoust. Soc. Am. 89, 902–909.

Smith, J.O., 1992. Physical modeling using digital waveguides. Comput. Music J. 16, 73–87.

Smith, J.O., 1998. Principles of digital waveguide models. In: Kahrs, M., Brandenburg, K. (Eds.), Applications of Digital Signal Processing to Audio and Acoustics. Kluwer Academic Publisher, Boston, Massachusetts, USA, pp. 417–466.

Sommers, M.S., Moody, D.B., Prosen, C.A., Stebbins, W.C., 1992. Formant frequency discrimination by Japanese macaques (*Macaca fuscata*). J. Acoust. Soc. Am. 91, 3499–3510.

Stevens, K.N., 1972. Quantal nature of speech. In: David Jr., E.E., Denes, P.B. (Eds.), Human Communication: A Unified View. McGraw-Hill, New York.

Stevens, K.N., 1999. Acoustic Phonetics. MIT Press, Cambridge, Massachusetts.

Story, B., Titze, I., Wong, D., 1996. A Simplified Model for Simulation and Transformation of Speech. In: Proceedings of IEEE International Joint Symposia on Intelligence and Systems, Rockville, MD, pp. 320–327.

Titze, I.R., 1994. Principles of Voice Production. Prentice-Hall, Englewood Cliffs, New Jersey.

Zeskind, P.S., Collins, V., 1987. Pitch of infant crying and caregiver responses in a natural setting. Infant Behav. Dev. 10, 501–504.

Zhinkin, N.I., 1963. An application of the theory of algorithms to the study of animal-speech-methods of vocal intercommunication between monkeys. In: Busnel, R.G. (Ed.), Acoustic Behavior of Animals. Elsevier, Amsterdam.

Zuberbühler, K., 2000a. Referential labeling in Diana monkeys. Animal Behaviour 59, 917–927.

Zuberbühler, K., 2000b. Interspecific semantic communication in two forest primates. Proc. R. Soc. Lond. B 267, 713–718.

Zuberbühler, K., 2003. The effects of natural and sexual selection on the evolution of guenon loud calls. In: Glenn, M., Cords, M. (Eds.), The Guenons. Plenum Press, New York.