

Pitch (F_0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry

Drew Rendall, Sophie Kollias, Christina Ney, and Peter Lloyd

Citation: [The Journal of the Acoustical Society of America](#) **117**, 944 (2005);

View online: <https://doi.org/10.1121/1.1848011>

View Table of Contents: <http://asa.scitation.org/toc/jas/117/2>

Published by the [Acoustical Society of America](#)

Articles you may be interested in

[Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques](#)

[The Journal of the Acoustical Society of America](#) **102**, 1213 (1998); 10.1121/1.421048

[Morphology and development of the human vocal tract: A study using magnetic resonance imaging](#)

[The Journal of the Acoustical Society of America](#) **106**, 1511 (1999); 10.1121/1.427148

[Acoustic characteristics of American English vowels](#)

[The Journal of the Acoustical Society of America](#) **97**, 3099 (1998); 10.1121/1.411872

[The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness](#)

[The Journal of the Acoustical Society of America](#) **129**, 2201 (2011); 10.1121/1.3552866

[Physiologic and acoustic differences between male and female voices](#)

[The Journal of the Acoustical Society of America](#) **85**, 1699 (1998); 10.1121/1.397959

[System for Automatic Formant Analysis of Voiced Speech](#)

[The Journal of the Acoustical Society of America](#) **47**, 634 (2005); 10.1121/1.1911939

Pitch (F_0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry^{a)}

Drew Rendall,^{b)} Sophie Kollias, and Christina Ney

Department of Psychology and Neuroscience, University of Lethbridge, Lethbridge, Alberta, T1K 3M4 Canada

Peter Lloyd

Cape Nature Conservation, Western Cape Province, South Africa

(Received 22 July 2004; revised 3 November 2004; accepted 22 November 2004)

Key voice features—fundamental frequency (F_0) and formant frequencies—can vary extensively between individuals. Much of the variation can be traced to differences in the size of the larynx and vocal-tract cavities, but whether these differences in turn simply reflect differences in speaker body size (i.e., neutral vocal allometry) remains unclear. Quantitative analyses were therefore undertaken to test the relationship between speaker body size and voice F_0 and formant frequencies for human vowels. To test the taxonomic generality of the relationships, the same analyses were conducted on the vowel-like grunts of baboons, whose phylogenetic proximity to humans and similar vocal production biology and voice acoustic patterns recommend them for such comparative research. For adults of both species, males were larger than females and had lower mean voice F_0 and formant frequencies. However, beyond this, F_0 variation did not track body-size variation between the sexes in either species, nor within sexes in humans. In humans, formant variation correlated significantly with speaker height but only in males and not in females. Implications for general vocal allometry are discussed as are implications for speech origins theories, and challenges to them, related to laryngeal position and vocal tract length. © 2005 Acoustical Society of America.

[DOI: 10.1121/1.1848011]

PACS numbers: 43.80.Ka, 43.70.Gr

Pages: 944–955

I. INTRODUCTION

Human speech is a multidimensional signal. In addition to conveying the phonetic contrasts essential to language, the speech signal contains cues to myriad nonlinguistic dimensions of a speaker, including sex, age, individual identity, and dialect background (Ladefoged and Broadbent, 1956; Abercrombie, 1967; Bricker and Pruzansky, 1976; Bachorowski and Owren, 1999). Indeed, the latter *indexical* dimensions can intrude on the linguistic function of speech by introducing significant speaker-specific variation into the speech signal that complicates recovery of its linguistic content, requiring some mechanism of “speaker normalization” in listeners. Given the importance of such normalization to effective communication, a central preoccupation of speech science has been uncovering the perceptual and cognitive processes involved, and at the same time also tracing the underlying sources of variation in the speech signal to begin with.

Considerable progress has been made on both fronts, although there remain unresolved issues in each area. For example, in perception, alternative explanatory frameworks stress the importance of either canceling the spurious

speaker-dependent variation to recover canonical phoneme targets, or actively incorporating such variation into the process of phoneme identification given recent evidence that speech comprehension can be facilitated by familiarity with the idiosyncracies of particular talkers’ voices (reviewed in Nearey, 1989; Johnson and Mullenix, 1997).

On the production side, research has focused on speaker-dependent variation in vocal production physiology, particularly differences in the size of critical features of vocal-tract anatomy. There is a solid understanding of the basic relationships here and their implications for the speech signal. For example, many of the basic differences in the speech of children and adults, and within adults between males and females, can be traced to differences in the size of the larynx and vocal-tract cavities that determine voice fundamental frequency (F_0) and formant frequency profiles, respectively (Peterson and Barney, 1952; Hirano *et al.*, 1983; Titze, 1989; Fitch and Giedd, 1999). These differences in turn are to a large extent thought to reflect differences in body size between adults and children, and between males and females (Fitch and Giedd, 1999). However, we are on somewhat shakier ground here (Nordström, 1977). Although some of the differences in voice acoustics (e.g., between children and adults) probably do reflect straightforward differences in vocal anatomical maturation related to differences in general growth and development trajectories, some apparently do not.

For example, the nearly twofold difference in baseline

^{a)}Portions of this work were presented at the First International Conference on Animal Acoustic Communication, cohosted by the Acoustical Society of America and the University of Maryland, College Park, Maryland, July 2003.

^{b)}Electronic mail: d.rendall@uleth.ca

speaking F_0 between adult males and females appears to be out of all proportion to the more modest body-size differences between the sexes, and previous work has also failed to find any correlation between voice F_0 and body size within either sex (Lieberman, 1967; Kunzel, 1989; van Dommelen and Moxness, 1995; Collins, 2000). Similarly, a handful of studies examining formant frequencies (or simply the broader frequency spectrum) in relation to body size variation between and within sexes have yielded weak or mixed results (Lass *et al.*, 1980a, b; van Dommelen and Moxness, 1995; Collins, 2000; Greisbach, 1999; González, 2004). As a result, there is continuing interest in the underlying biology of body-size growth and development that contributes to variability in vocal-tract anatomy and thus to the speaker-specific variation in speech that it is ultimately so important to normalize in speech perception.

There are additional reasons for the interest in potential body-size influences on speech production and voice acoustics. To begin with, it is possible that voice cues to body size, and their underlying vocal anatomical determinants, are themselves subject to specific biological selection pressures. It is certainly true that body size plays a deterministic role in competitive interactions and other social contexts in many species, including humans. As a result, there could be functional value in *advertising* size in various contexts through voice features that reveal, or even exaggerate, vocalizer size (Morton, 1977; Fitch, 1999, 2000). Indeed, there is some broad support for this possibility in the form of demonstrable correlations between different spectral dimensions of the voice (e.g., fundamental frequency, “dominant” frequency/harmonic, and resonance frequencies) and body size in vocalizations used by some anurans and mammals either in aggressive competition among males, or in mate attraction displays, where size matters (e.g., Davies and Halliday, 1978; Reby and McComb, 2003; reviewed in Fitch and Hauser, 2002).

Furthermore, such observations have inspired a recent proposal that a critical feature of human vocal-tract anatomy long held to represent a specific adaptation for language, namely a descended larynx positioned low in the vocal tract by comparison to closely related primate species (Lieberman, 1968; Lieberman *et al.*, 1969), actually represents an adaptation for body-size advertisement (Fitch, 2000; Fitch and Reby, 2001). That is, humans’ descended larynx positions it outside the bony constraints of the skull, allowing it to expand more freely in multiple dimensions and at the same time creates a longer vocal tract. Both characteristics could function to advertise larger body size through vocalizations with lower F_0 and formant frequency profiles, respectively.

To more fully evaluate these ideas and to contribute to our broader understanding of the connections between body-size variation, vocal-tract allometry, and variable speech acoustics, we need a better understanding of the extent to which these features of speech actually track variation in speaker body size. Therefore, in this paper, we undertake analyses of human voice acoustics in relation to several dimensions of speaker body size. We focus specifically on vowel sounds because these manifest clear F_0 and formant

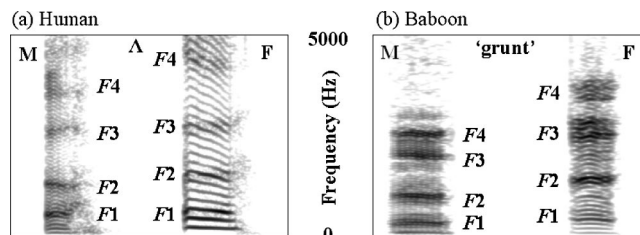


FIG. 1. Spectrograms of (a) human vowels and (b) baboon grunts produced by an adult male (M) and female (F) of each species. The vowel illustrated is that from the word “butt.” Both spectrograms were produced using a 1024-point FFT (approximately 50 ms), a Hanning window, and 90% frame overlap. Note the general similarity in frequency structure between the vowels and grunts, and the similar pattern of F_0 and formant differences between males and females within each species.

profiles that seem to offer the most salient cues to size-related vocal anatomical variation, and we limit our analyses to adults to avoid age-related confounds in body size and vocal-tract anatomy. In addition, where possible, we include comparable data on the grunt vocalizations of adult baboons.

Baboons provide an especially useful comparison here. As primates relatively closely related to humans, they provide a logical opportunity to trace potential continuity in the processes of human vocal production and to help clarify some of the basic vocal anatomical and acoustic-structural relationships involved. Indeed, research on baboon communication has already revealed several productive parallels to speech phenomena. Thus, the most common vocalization produced by baboons—a harmonically rich grunt—is structurally analogous to human vowel sounds, having a clear, stable F_0 and a set of prominent resonance peaks that reflects the combined action of stable vocal-fold vibration and subsequent spectral shaping by the supralaryngeal vocal-tract cavities (Zhinkin, 1963; Andrew, 1976; Richman, 1976). More detailed analyses have shown that the mean F_0 and resonance frequencies of female grunts actually closely match those of vowels produced by English-speaking males, a convergence attributed to a coincidence in vocal-fold and vocal-tract lengths in the two groups (Owren *et al.*, 1997; Rendall, 2003). This work has also shown that the baboons produce different acoustic variants of grunt that, like vowels, differ in spectral properties related to vocal-tract filtering (Owren *et al.*, 1997; Rendall *et al.*, 1999). In addition, sex differences in the F_0 and formant profiles of baboon grunts parallel sex differences in the F_0 and formant profiles of human vowels (Rendall *et al.*, 2004), as illustrated in Fig. 1, suggesting that some similar process of vocalizer normalization might also be important for the perception of different grunt variants in baboons.

Finally, and importantly, although the basic anatomy of vocal production and its functional operation is similar in baboons and humans (Fitch and Hauser, 1995; Owren and Linker, 1995), baboons do not have either a descended larynx or language (in the human sense). Therefore, they represent a very valuable control for these influences on vocal production, allowing us to explore the relationships between vocalizer body size, vocal-tract anatomy, and voice acoustics unconfounded by language and any potentially associated anatomical adaptations.

We begin by examining the F_0 and formant profiles of human vowels and baboon grunts produced by adult males and females of each species and the extent to which sex differences in these acoustic features track body-size differences between the sexes. We follow this with an analysis of size-related F_0 and formant frequency variation among individuals within each sex.

II. BETWEEN-SEX VARIATION IN BODY SIZE AND VOICE ACOUSTICS IN HUMANS AND BABOONS

A. Methods

1. Voice samples and body-size measurements

a. Humans. Matched voice samples and body-size measurements were obtained from undergraduate students recruited from psychology courses at the University of Lethbridge. Students were screened for language and dialect background in an effort to reduce potential acoustic variation associated with such differences across subjects. Only students for whom Canadian English was their first language and lacking strong regional accents were retained in the study. The sample ultimately retained for analysis consisted of a balanced set of 68 students (34 males and 34 females) whose average age was 23 years (range: 18–44 years).

Voice samples were collected individually from subjects in a sound-controlled room in the Laboratory of Comparative Communication and Cognition. Subjects were seated in a comfortable padded chair and were fitted with an adjustable head-worn microphone (AKG C420) connected to a preamplifier (Behringer MX602A) that allowed the recordist to adjust signal recording level appropriately. The preamplifier was connected to a Pentium IV computer through a tunable, 8-pole Butterworth antialias filter (Frequency Devices 900/9L8B). Speech signals were digitized with 16-bit quantization at a sample rate of 22.05 kHz after low-pass filtering at 10 kHz.

Subjects were given a sheet containing the speech material to be recorded. The material emphasized the production of vowel sounds which are the components of speech most likely to reflect size-related variation through variable vocal-fold vibration rates (F_0) and vocal-tract resonance frequencies. The material included a list of isolated vowels, two lists of single-syllable words encompassing a range of different vowels (in either hVd or bVt context), and four short sentences constructed of single-syllable words that replicated many of the vowel sounds present in the word lists. The order of material within each list of isolated vowels, words, and phrases was randomized across subjects to avoid serial order effects. Subjects were instructed to *say* rather than *read* the material in a clear and comfortable voice, pausing deliberately between each item. They were asked to practice the material aloud in order to familiarize them with the content and to get them comfortable speaking in front of the recordist and while wearing the microphone. The practice interval also allowed the recordist to optimize the recording level for each subject. After a variable interval of practice, each subject was recorded producing the material in a single session from beginning to end.

Prior to collecting the speech sample, measurements were taken of each subject's height, weight, and neck circumference (taken at the level of the greatest laryngeal protrusion), and a commercial flatbed scanner was used to obtain an image of each hand, stretched flat on the surface of the scanner. From the scanned images, the length of the third digit was measured in ADOBE PHOTOSHOP®.

b. Baboons. Grunt vocalizations were collected during field research on chacma baboons at two sites in southern Africa. One site was located in the Moremi Game Reserve in northern Botswana, and the other site was in the De Hoop Nature Reserve in the Western Cape Province of South Africa. Baboons at the two sites belong to the same species and subspecies (*Papio hamadryas ursinus*) that is distributed continuously throughout southern Africa (i.e., Namibia, Botswana, Zimbabwe, and South Africa). Subjects were the sexually and physically mature adult males and females of the main study group at each site. Both groups had been studied continuously for several years. As a result, all individuals were known and easily identified. The animals themselves were fully habituated to human observers, allowing high-quality vocal recordings to be made at very close range (0.5–2.0 m). Vocalization recordings were made during a single protracted study period in Botswana (February 1996–March 1997) and during multiple shorter study periods in South Africa (December 2000; May–August 2001; January–May 2002).

Grunt vocalizations were recorded during the course of focused observational research on the baboons which involved consecutive 1-h follows of specific individuals. (For additional details regarding the collection of behavioral data, see Rendall *et al.*, 2000). Baboons grunt in a variety of contexts but especially during dispersed travel and foraging activities and when approaching one another to socialize (Rendall *et al.*, 1999). Recordings were made on Sony type IV metal tapes using a Sony WM-D6C Professional Walkman cassette recorder and a Sennheiser ME 80 directional microphone with K3U power module. Analog field recordings were subsequently digitized with 16-bit accuracy using a sample rate of 22.05 kHz, after low-pass filtering at 10 kHz. The sample ultimately available for acoustic analysis consisted of 1028 grunts from 27 different individuals (11 males and 16 females). Additional details of the grunt sample and subjects studied can be found in Rendall *et al.* (2004).

Body-size measurements of baboons were collected several years earlier (1976–1979) during a multiyear census of the baboon population in the De Hoop Nature Reserve. During this census, animals were captured and anesthetized to collect biometric data including the height (head-crown to tail-base), weight, and length of the right foot of each animal. The sample included complete measurements of 47 adult baboons (>7 years of age), of which 29 were male and 18 were female. These animals were captured in the same area that is now occupied by the main study troop at De Hoop. However, their identities could not be traced to any of the current members of this troop. Hence, the body size and vocalization samples have to be regarded as unmatched.

TABLE I. Means and standard deviations of body-size (cm/kg) and acoustic features (Hz) for humans (vowels) and baboons (grunts) and the degree of sexual dimorphism in each feature. Dimorphism values are given as the ratio of values between the sexes (M:F for body size features; F:M for acoustic features).

Humans ^a	Height	Digit-3	Neck	Weight	F_0	$F1$	$F2$	$F3$	$F4$
Male	183.8 [7.5]	3.36 [0.16]	38.4 [2.0]	85.1 [12.6]	113 [17.9]	468 [25.2]	1431 [59.0]	2505 [119.4]	3468 [149.0]
Female	168.3 [6.9]	2.98 [0.13]	32.0 [1.9]	64.4 [8.9]	204 [17.4]	583 [32.3]	1747 [79.9]	2915 [130.3]	4089 [175.1]
Dimorphism	1.09	1.13	1.20	1.32	1.81	1.25	1.22	1.16	1.18
1.20									
Baboons ^b	Height	Foot		Weight	F_0	$F1$	$F2$	$F3$	$F4$
Male	85.0 [4.67]	20.9 [0.79]		29.8 [2.37]	53 [4.43]	332 [41.4]	1062 [44.5]	1878 [107.9]	2603 [194.6]
Female	69.6 [3.22]	17.5 [0.73]		17.3 [1.43]	114 [11.6]	448 [43.3]	1430 [109]	2677 [157]	3407 [232.0]
Dimorphism	1.22	1.19		1.72	2.15	1.35	1.35	1.43	1.31
1.36									

^aMatched samples for body size and acoustic data ($N=34$ males, 34 females). Acoustic data represent the average values across 23 vowel utterances.

^bUnmatched samples for body size ($N=29$ males, 18 females) and acoustic data ($N=11$ males, 16 females).

2. Acoustic analysis

a. Human vowels. Analysis of human voice samples focused on vowels contained in one of the word lists (*beet, bait, bit, bet, bat, butt, boat, boot, book*) and the vowels from several words in the short sentences (“*Put the book on top of the cup.*”; “*Beat the tip of the stack.*”; “*Butt the herd with a stick.*”; “*Pull the rope through the boat.*”). In addition, the unstressed vowel in three instances of the word “*the*” in these phrases was also measured. This vowel was taken to approximate the schwa vowel, produced without emphasis and with a comparatively neutral vocal-tract configuration, making it the most likely source for unbiased F_0 and formant cues to size. In total, 23 vowel sounds were analyzed for each subject.

Acoustic analysis involved obtaining values of the fundamental frequency (F_0) and the frequency of the first four formants ($F1-F4$) from the central, steady-state portion of each vowel, with the exception of the glide vowels in the words “*bait*” and “*boat*.” For these two vowels, analysis was shifted either toward the initial section of the vowel or toward the end to better isolate the portion of interest (i.e., /e/ and /o/, respectively). Acoustic analyses were performed using the PRAAT[©] acoustic software package (version 4.1.9) implemented on the Windows XP Professional Operating System. The F_0 of each vowel was obtained using the pitch-tracking function of PRAAT to first highlight its “pitch” contour. The appropriate portion of each vowel was then circumscribed with interactive cursors and the program queried for the mean pitch (F_0) over that range.

The frequencies of the first four formants of each vowel were obtained using the formant-tracking function of PRAAT to highlight the contour of each formant. In this analysis, the formant-tracking routine was instructed to identify five formants over the range 0–5000 Hz using a frequency analysis window of 25 ms (approximately 500 points). Once again, interactive cursors were used to delimit the appropriate portion of the vowel, and the program was queried for the mean value of each formant over that range. Formant frequency values obtained in this way were independently verified using autocorrelation-based linear predictive coding (LPC). In this analysis, an LPC spectrum was generated from a 25-ms

slice centered on the appropriate region of the vowel. This LPC spectrum was overlaid on a matching sized fast-Fourier transform (FFT) of the same region of the vowel, and the goodness-of-fit of the LPC peaks to emphasized regions in the underlying FFT was examined using 10–14 LPC coefficients. LPC peaks were then compared to formant frequency values obtained through formant tracking.

b. Baboon grunts. The analysis of baboon grunts proceeded in much the same way, except that the analyses were performed using the ENTROPIC SIGNAL PROCESSING SYSTEM WAVES+ software package (ESPS/WAVES+[©] version 5.3) (1997) implemented on the Linux operating system (Red Hat[©] version 6.0). Analyses of these calls have been published previously with detailed descriptions of the methodology used (Rendall *et al.*, 2004). Briefly, the analysis of grunts also focused on obtaining values for the F_0 and frequencies of the first four resonances (or formants) of the calls. F_0 values were obtained from the amplitude waveform of each grunt. Because most grunts exhibited extremely regular periodicity, individual cycles of vocal-fold opening and closing (representing F_0) were clearly visible in the waveform. We measured the period of each cycle of vocal-fold opening and closing from six cycles of vocal-fold vibration from the middle of each grunt, and averaged their inverses to obtain a mean F_0 value (in Hz) for each grunt.

The frequencies of the first four formants of each grunt were identified via autocorrelation-based LPC. An LPC spectrum was generated from a 1024-point (approximately 50 ms) segment from the midpoint of each grunt, using a Hanning window, and 18 coefficients. This LPC spectrum was then overlaid on an independently derived FFT for the same segment of the grunt to ensure the goodness-of-fit of the LPC-derived frequency spectrum. The frequencies of the first four poles of the LPC spectrum were then extracted and retained.

B. Results and discussion

Mean values for body-size and acoustic features of adult males and females of each species are given in Table I. Body-size data for both species were well-conditioned with

bivariate scatterplots revealing no obvious outliers and mean values comparing well with those from previous morphometric studies of humans and baboons, some involving much larger samples (e.g., Hamill *et al.*, 1979; Barrett and Henzi, 1997; McFadden and Bracht, 2003). Male:female dimorphism values confirmed that the bodily dimensions studied obeyed basic biophysical principles. Thus, in humans, sexual dimorphism in height and digit length (both linear dimensions) was approximately 1.10 (i.e., males were 10% “longer” than females). The sex difference in neck circumference (an areal dimension) and weight (a volumetric dimension) were 1.20 and 1.32, respectively, which match almost exactly the values that would be expected based on the square (1.21) and cube (1.33) of the observed linear dimorphism. Although the magnitude of sexual dimorphism in all body dimensions was greater in baboons than it was in humans, the same biophysical relationships were nevertheless observed. Thus, in baboons, sexual dimorphism in linear dimensions (height and foot length) was approximately 1.20, while the difference in volumetric dimension (weight) was 1.72, which is almost exactly the cube of the observed linear dimorphism (1.73).

For humans, mean values for the F_0 and four formants agree well with those reported previously for American English speakers (e.g., Peterson and Barney, 1952; Hillenbrand *et al.*, 1995; Stevens, 1998). For example, Stevens (1998) reported overall means for F_1 – F_3 of 478, 1485, and 2563 Hz for adult males, and 592, 1755, and 2855 Hz for adult females, which match very closely the values obtained here. Means for the F_0 and formant frequencies of baboon grunts also agreed exceptionally well with those reported in the only previous detailed study of their acoustic structure with mean values obtained in the two studies differing by only 1%–4% (Owren *et al.*, 1997).

In both species, then, males are larger than females in each body dimension and they produce sounds with a lower F_0 and lower formant frequencies. The acoustic differences between males and females did not, however, map onto body-size differences between them in a straightforward way. In humans, the familiar sexual dimorphism in F_0 (1.81) greatly exceeded sexual dimorphism in linear (1.10), areal (1.20), and volumetric (1.32) dimensions. The same phenomenon was also found in baboons. The degree of sexual dimorphism in linear (1.20) and volumetric dimensions (1.72) was higher in baboons than it was in humans, but the degree of sexual dimorphism in mean F_0 in baboons (2.15) was higher still.

Thus, in both species, the difference in mean voice pitch between males and females considerably exceeded any difference in body size between the sexes. In humans, the extreme F_0 dimorphism (1.81) is largely, though not completely, accounted for by the large dimorphism in vocal-fold length, which is reported to be 1.50 for the entire fold and 1.60 for the membraneous portion that vibrates (Hirano *et al.*, 1983; Titze, 1989). The remaining 20%–30% excess acoustic dimorphism may be due to an additional sex difference in vocal-fold thickness, which is reported to be approximately 1.20–1.35 (Hirano *et al.*, 1983), or due to other differences in vocal-fold tissue properties (e.g., density,

structuring), although Titze (1989) discounts many of these. Alternatively, it could reflect differences in behavioral control of normative vocal-fold operation between males and females. There are no published data on vocal-fold dimensions in baboons that might provide an anatomical account of the excess F_0 dimorphism in this species.

The connection of formant differences between males and females to body-size differences between them was equivocal. In both species, mean formant differences between the sexes were considerably greater than the sex difference in linear dimensions but smaller than the sex difference in volumetric dimensions. The formant differences best matched the sex differences in observed or expected areal dimensions. The latter outcome might imply that the vocal-tract cavities of males and females differ consistently not just in length but also in a second dimension (e.g., width).

This possibility is plausible for baboons where the larynx is positioned high in the vocal tract such that the supralaryngeal vocal tract is defined primarily by the oral cavity, which in turn is determined by the species’ projected face, or muzzle. The muzzle of adult males is longer than that of adult females but it is also broader, reflecting the extreme sexual dimorphism in canine tooth size in this species. This sex difference in muzzle size could yield vocal-tract cavities that differ in both length and width between the sexes, although detailed measurements of muzzle and oral cavity dimensions would be needed to test this possibility properly. Furthermore, the match between observed formant dimorphism in baboons (1.36) and the expected dimorphism in areal dimensions in this species (1.44) was not perfect, which might indicate that two-dimensional differences in the vocal-tract cavities of males and females are not uniform along the entire muzzle, perhaps because the muzzle tapers towards the tip (which it does). Alternatively, it could simply mean that the notion of two-dimensional variation is not relevant at all. Instead, the excess formant dimorphism beyond what would be expected based on the difference in linear dimensions between males and females might be behavioral in origin.

In humans, the average sexual dimorphism in formants (1.20) exactly matched the sex difference in areal body dimensions (1.20), but it is not so obvious how the vocal-tract cavities of human males and females might differ consistently in two dimensions to support this outcome (cf. Apostol *et al.*, 2004). It is possible that some of the formant difference between the sexes reflects behavioral effects similar to those just proposed (Lieberman, 1986). In fact, there is some evidence from the speech of children of increased lip protrusion in boys compared to girls (Sachs *et al.*, 1972; Goldstein, 1980). However, the sex difference in formants observed in this study was also very similar to the baseline difference in vocal-tract length that is reported to characterize adult males and females, which is estimated at 15%–20% (Fant, 1966; Goldstein, 1980). A disproportionate share of this difference is proposed to lie in the pharyngeal region, with males having an oral cavity that is roughly similar in length to that of females but a pharyngeal cavity that is appreciably longer (Fant, 1966; Goldstein, 1980; Lieberman, 1986; Fitch and Giedd, 1999).

Consistent with the latter proposal, there was consider-

TABLE II. Means and standard deviations of $F1-F4$ for adult males (M) and females (F) for different vowels from individual words spoken in bVt context, and the degree of formant dimorphism (F:M) between the sexes for each vowel.

	F1			F2			F3			F4		
Word	M	F	F:M	M	F	F:M	M	F	F:M	M	F	F:M
<i>Beet</i>	242	290	1.20	2259	2801	1.24	3023	3437	1.14	3630	4261	1.17
	[20.1]	[38.2]		[143.6]	[133.3]		[169.5]	[174.5]		[206.7]	[208.5]	
<i>Bit</i>	434	530	1.22	1851	2266	1.22	2630	3069	1.17	3586	4226	1.18
	[36.2]	[70.8]		[128.8]	[166.0]		[134.8]	[141.8]		[198.1]	[276.3]	
<i>Bait</i>	407	456	1.12	2011	2497	1.24	2653	3131	1.18	3539	4170	1.18
	[32.1]	[33.2]		[104.7]	[139.0]		[117.9]	[159.1]		[141.2]	[249.0]	
<i>Bet</i>	589	754	1.28	1711	2046	1.20	2562	2983	1.16	3602	4238	1.18
	[48.3]	[54.7]		[101.0]	[143.4]		[147.2]	[162.8]		[194.7]	[280.3]	
<i>Bat</i>	731	951	1.30	1550	1819	1.17	2466	2850	1.16	3541	4215	1.19
	[62.5]	[73.4]		[87.2]	[137.1]		[142.0]	[233.8]		[205.7]	[281.8]	
<i>Boat</i>	386	459	1.19	859	993	1.16	2438	2848	1.17	3289	3880	1.18
	[28.3]	[39.9]		[70.7]	[91.7]		[159.8]	[198.6]		[202.0]	[259.0]	
<i>Boot</i>	286	364	1.27	1091	1303	1.19	2251	2729	1.21	3254	3979	1.22
	[29.0]	[39.3]		[119.4]	[179.1]		[134.9]	[148.2]		[217.3]	[279.8]	
<i>Book</i>	455	550	1.21	1037	1272	1.23	2432	2813	1.16	3353	3915	1.17
	[35.5]	[74.0]		[73.6]	[143.3]		[160.8]	[171.8]		[205.9]	[217.6]	
<i>Butt</i>	625	822	1.32	1257	1595	1.27	2507	2868	1.14	3505	4116	1.17
	[48.9]	[68.8]		[75.2]	[115.2]		[166.3]	[180.6]		[210.6]	[220.4]	
<i>The</i>	437	516	1.18	1340	1669	1.25	2486	2897	1.17	3486	4135	1.19
	[47.1]	[51.1]		[70.3]	[104.2]		[135.3]	[132.1]		[177.5]	[226.6]	

able variation in the magnitude of formant dimorphism across different formants and different vowel sounds. This variation is reported in Table II and illustrated in Fig. 2. Table II lists mean frequency values of $F1-F4$ for different vowels produced by males and females and the magnitude of

formant dimorphism in each of them. The latter dimorphism values for particular formants and particular vowels deviated considerably from the overall mean formant dimorphism of 1.20, ranging from a low of 1.12 to a high of 1.32. Predictably the variation across different vowels was greater in $F1$ and $F2$ than in $F3$ and $F4$.

The variation specifically in $F1$ and $F2$ is illustrated in Fig. 2, which plots the mean frequency values of $F1$ and $F2$ for different vowels produced by males and females. The values for each sex are interconnected to capture the overall formant space occupied by males and females and to illustrate how these spaces are shifted relative to each other along the $F1-F2$ axes.

For only two vowels (/I/ and /U/) and the grunts of baboons did the formant differences between males and females reflect equal proportional differences in $F1$ and $F2$. For the remaining vowels studied, the sex differences reflected disproportionate differences either in $F1$ (/E/, /æ/, /Λ/, /o/, /u/), or in $F2$ (/i/, /e/, /ə/). This pattern of primarily non-proportional formant variation between sexes in humans is what Fant (1966), termed “non-uniform F -pattern scaling,” and it is consistent with the proposal that the difference in vocal-tract length between males and females is not evenly distributed across the vocal tract but rather is concentrated more in certain regions, such as the pharynx (see also Whiteside, 2001).

III. WITHIN-SEX VARIATION IN BODY SIZE AND VOICE ACOUSTICS IN HUMANS

To examine whether the variable F_0 and formant frequency profiles of adult males and females might track body-size variation within each sex, we undertook additional body-size acoustic comparisons within males and females, respectively.

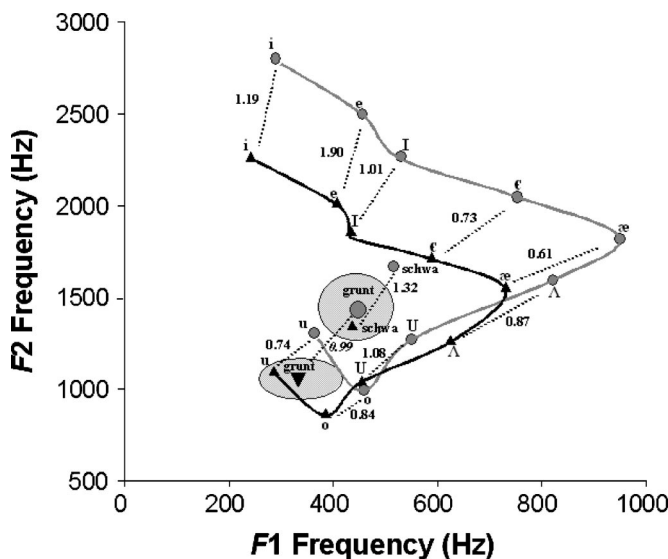


FIG. 2. Comparisons of the formant space occupied by human vowels and baboon grunts and the extent to which sex differences within each species reflect disproportionate shifts in either $F1$ or $F2$. Small black triangles and gray circles represent the mean frequency values of $F1$ and $F2$ for each of the ten vowels produced by human males and females, respectively. Solid lines connect and encircle the vowels for each sex to illustrate their relative placement in $F1-F2$ vowel space. Dashed lines connecting the same vowel in males and females, and the slope values next to them, indicate the extent to which the sex difference reflects a disproportionate shift in either $F1$ (<1.0), or $F2$ (>1.0), or a proportional shift in both formants (1.0). The larger black triangle and gray circle represent the mean frequency values of $F1$ and $F2$ for grunts produced by adult male and female baboons. Because the baboons did not produce the equivalent of different vowels, shaded circles are used to illustrate the range of variation observed in grunt formants for each sex.

A. Methods

Methods for the collection and analysis of body size and acoustic data were the same as those described previously except that they were limited to humans for whom we had matched body size and acoustic data from the same individuals.

We used regression analysis to test the extent to which variation in the F_0 and formant frequencies of individual males and females was associated with variation in three body-size dimensions (height, weight, and neck circumference). We conducted two related regression analyses, one using individuals' mean values for all 23 vowels produced, and the other using individuals' means on only the schwa vowels (from the three instances of the unstressed word "the" in the short phrases).

For the formant frequency component of both analyses, we first calculated for each vowel the mean frequency difference between successive formants (i.e., $[(F_2 - F_1) + (F_3 - F_2) + (F_4 - F_3)]/3$). This yields a derived variable summarizing the average spacing between formants. Fitch has previously labeled this variable "formant dispersion" (FD) and has proposed that it should be an especially good index of overall vocal tract length and thus also body size (Fitch, 1997, 1999, 2000). Therefore, in the first set of analyses, we regressed this variable, in combination with F_0 , on subjects' height, weight, and neck circumference.

We followed this with separate regressions of the combination of F_0 and the individual formant frequencies (i.e., F_1 – F_4) on subjects' height, weight, and neck circumference. We undertook this second analysis for two reasons. First, we wanted to test the extent to which any significant effects observed for the FD variable might be attributed to the differential contributions of specific formants. Second, we were especially concerned to test this because, although increasingly widely accepted and used in the literature, the algorithm for calculating FD, as proposed by Fitch, effectively reduces to the frequency difference between F_4 and F_1 . Because F_1 is also the least variable formant in human vowels, the resulting FD values are potentially determined primarily only by F_4 . Therefore, we ran a second series of regressions using the individual formant frequencies to be sure that all four formants were included and equally weighted in analyses. We also reran all regression analyses after applying a log transformation to the data as is often appropriate for tests of allometric relationships. Results using log-transformed data did not differ from those obtained using the original (untransformed) data. Therefore, for ease of presentation and interpretation, we present only the results based on the original data.

B. Results and discussion

Results of regression analyses are given in Table III. For females, multiple regressions of F_0 and FD on height, weight, and neck circumference yielded no significant overall effects at an alpha level of 0.01. This was true both for regressions using subjects' mean values across all 23 vowel sounds produced and for regressions using subjects' mean values for only the three schwa vowels. Multiple regressions

using F_0 and the individual formant frequencies (i.e., F_1 – F_4) also failed to identify any significant effects.

For males, multiple regressions using the combination of F_0 and FD yielded a statistically significant overall effect for height but not for either weight or neck circumference. This pattern was the same for regressions using subjects' means for all vowels and using their means for only the schwa vowels. In both vowel-type comparisons, the regression of acoustic features on subjects' weights yielded P values that might be regarded as approaching significant levels (0.068 and 0.084). However, in both cases the adjusted R^2 values were low ($\text{adj-}R^2 = 0.105$ and 0.093 , respectively) by comparison to the same analyses on height ($\text{adj-}R^2 = 0.306$ and 0.270 , respectively). In the case of height, the statistically significant effects observed both for all vowels and for only the schwa vowels were due to FD but not F_0 .

Follow-up regression tests in males using subjects' F_0 and the frequencies of each of the first four formants again yielded a statistically significant overall effect for height but not either weight or neck circumference in both vowel-type comparisons. For all vowels combined, the result was attributable to the effect of only F_4 . For the schwa vowels, the result was attributable to the effects of the F_0 and all four formant frequencies, with the strongest effect being for F_4 .

These variable effects across males and females and within males across the different acoustic features are illustrated in Fig. 3, which provides scatterplots comparing the height of individual males and females to their respective F_0 and formant frequency values for the case of schwa vowels.

Overall, there were no obvious relationships between voice acoustics and body size within adult females, whether height, weight, or neck circumference. In males, the effects were clear only for height and were strongest for formants, particularly F_4 . Effects for males were slightly better for schwa vowels than for the average of all vowels produced, although the effects were clear for both. The latter outcome is not too surprising given that subjects' formant values averaged across all vowels were very similar to those for their schwa vowels, suggesting that speakers' average and neutral vocal-tract configurations were similar.

IV. GENERAL DISCUSSION AND IMPLICATIONS

A. Relationships between body size, vocal-tract anatomy, and voice acoustics

Comparisons of body size and voice acoustics in adult male and female humans and baboons point to significant deviations from strictly neutral vocal allometry. In both species, the F_0 difference between males and females exceeded any measured difference in body size between them. In humans, much of the excess F_0 dimorphism can be traced directly to a dimorphism in vocal-fold length that itself greatly exceeds the body size differences between males and females (Hirano *et al.*, 1983; Titze, 1989). The same is probably true of the excess F_0 dimorphism in baboons, although there are no systematic data on vocal-fold length in this species to confirm it. It was also the case in humans that F_0 variation did not track body-size variation within either sex. There was some indication that it might track height in males, but the

TABLE III. Results of multiple regression tests of F_0 and FD (or F_0 and individual formant frequencies, $F1-F4$) on body size dimensions (height, weight, and neck circumference) within (a) females and (b) males. Results are presented for analyses using the average of all vowels combined as well as for analyses using the average of only the schwa vowels.

(a) Females	Acoustic features ^a	Body size	<i>Adj-R</i> ²	<i>F</i>	<i>P</i>
<i>F</i> ₀ + FD					
All vowels		Height	0.000	0.45	0.639
		Weight	0.031	1.51	0.237
		Neck	0.092	2.61	0.090
Schwa vowels		Height	0.000	0.71	0.499
		Weight	0.053	2.15	0.134
		Neck	0.037	1.41	0.259
<i>F</i> ₀ + <i>F</i> 1 – <i>F</i> 4					
All vowels		Height	0.000	0.80	0.562
		Weight	0.017	1.11	0.377
		Neck	0.083	1.58	0.199
Schwa vowels		Height	0.000	0.81	0.554
		Weight	0.036	1.24	0.318
		Neck	0.069	1.47	0.231
<hr/>					
(b) Males	Acoustic features ^a	Body size	<i>Adj-R</i> ²	<i>F</i>	<i>P</i>
<i>F</i> ₀ + FD					
All vowels	FD ^b	Height	0.306	8.28	0.001
		Weight	0.105	2.94	0.068
		Neck	0.079	2.41	0.106
Schwa vowels	FD ^b <i>F</i> ₀ ^c	Height	0.270	7.11	0.003
		Weight	0.093	2.68	0.084
		Neck	0.069	2.22	0.126
<i>F</i> ₀ + <i>F</i> 1 – <i>F</i> 4					
All vowels	<i>F</i> 4 ^c	Height	0.339	4.38	0.005
		Weight	0.161	2.26	0.075
		Neck	0.082	1.59	0.196
Schwa vowels	<i>F</i> 4 ^b <i>F</i> ₀ , <i>F</i> 1, <i>F</i> 2 ^d <i>F</i> 3 ^c	Height	0.510	7.86	0.000
		Weight	0.089	1.64	0.182
		Neck	0.029	1.20	0.336

^aAcoustic features in each significant regression.

^b P value < 0.01.

^c P value < 0.10.

^d P value < 0.05.

relationship was weak. Several previous studies have similarly failed to find robust body-size effects for F_0 within sexes both in humans and in some other primate and mammal species (Kunzel, 1989; McComb, 1991; Masataka, 1994; van Dommelen and Moxness, 1995; Collins, 2000).

One functional explanation for the mismatch between F_0 and body size in both species follows the body-size exaggeration hypothesis (Morton, 1977; Fitch, 1999, 2000), namely that it reflects deceptive use of F_0 to signal a body size that is larger (or smaller) than reality. It is certainly true in humans that speakers can dynamically adjust voice pitch in different circumstances to project an acoustic “image” of larger or smaller size (or their social correlates, greater or lesser threat), sometimes to extreme extents such as when talking to babies or pets. However, for this to account for the differences in normative pitch profiles and larynx size between males and females would require a sustained evolutionary history of successful deception of listeners in spite of perfectly good visual cues to a speaker’s true size, which seems unlikely.

Another possibility is that F_0 variation is not about size

per se, but instead reflects physiological dimensions more closely aligned with sex variation (i.e., relative “maleness” and “femaleness”) that are only indirectly related to body size. It is known that there are androgen receptors in the larynx of several species including baboons and humans, that the onset of voice-pitch changes in males during adolescence parallels the onset of surging testosterone independently of body growth spurts, and that the F_0 of adult males correlates with salivary testosterone levels (Saez and Martin, 1976; Aufdemorte *et al.*, 1983; Hollien *et al.*, 1994; Dabbs and Mallinger, 1999). Female sex hormones are also known to influence vocal-fold dynamics in adult women (e.g., Abitbol *et al.*, 1999). Taken together, such findings suggest that much of the F_0 variation both between and within sexes might be tracking (and signaling) the active products of sexual physiology—that can vary continuously within as well as between sexes and that have a role in, but do not strictly determine, body size.

What is clear is that the growth programs affecting larynx size (and resulting F_0 profiles) are at least partially dissociated from the more general growth programs that affect

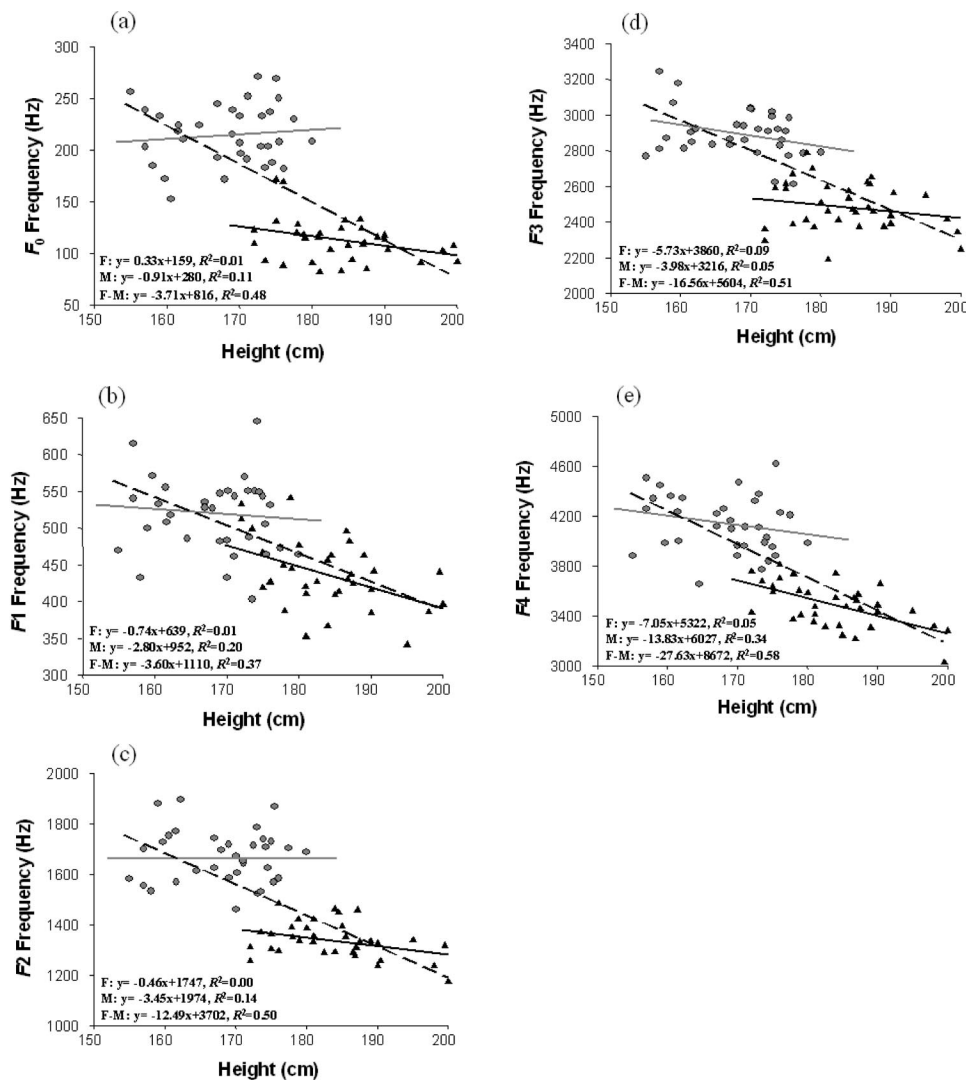


FIG. 3. Scatterplots illustrating the relationship between height and (a) the F_0 and (b)–(e) the formant frequencies of schwa vowels produced by human males (black triangles) and females (gray circles). In each scatterplot, a regression equation and R^2 value is given for separate analyses conducted within each sex (F=female; M=male). In each scatterplot, a third regression line (dashed), regression equation, and R^2 value is included that reflects a between-sex analysis (F–M) and illustrates the statistical artifacts that arise when the data are collapsed across males and females. Predictably, the regression slopes are much steeper and the R^2 values are much higher for the between-sex analyses than they are for those conducted within sexes.

true overall body size in both species. By extension, the hypertrophied larynx and disproportionately deep F_0 of human males compared to females is evidently not uniquely facilitated by humans' descended larynx that positions it outside the bony constraints of the skull. The same, indeed more extreme, phenomenon of disproportionately low F_0 was found in male baboons despite the positioning of their larynx high in the vocal tract and within the bony confines of the skull. Thus, it appears that the general capacity for dramatic laryngeal growth is neither entirely constrained by a location within the confines of the skull, nor necessarily facilitated only by descent of the larynx outside the skull boundaries.

In view of past mixed results on the scaling of F_0 with body size, Fitch (1997) proposed that formants should scale far more reliably with overall body size than does F_0 . The argument is that formants reflect the size of a speaker's vocal-tract cavities which are constrained by the size and shape of the surrounding bony anatomy (i.e., skull size and shape; neck length), and that these in turn are determined by

general growth processes affecting overall body size. This is a wholly intuitive proposal for which Fitch has provided some supporting evidence in humans, in one other species of monkey, and in domestic dogs where he has reported very strong correlations (e.g., $r = 0.90$) between body size (height or weight) and either voice formants or vocal tract length measurements obtained from imaging techniques (Fitch, 1997; Fitch and Giedd, 1999; Riede and Fitch, 1999). However, in each case, the data were collapsed across both sexes and all age classes (i.e., lumping together data from infant, juvenile, adolescent, and adult males and females), or across different dog breeds, which artificially inflates the size of the correlations. As a result, the very high correlation coefficients obtained seriously overestimated the extent to which variation in vocal-tract length among subjects was due to their body-size differences *per se* as opposed to their marked sex and age differences (or breed differences in the case of dogs).

Nevertheless, the results of the present study partially

support Fitch's general proposal, even though the relationships are not as clear as might be expected. In baboons, sex differences in formants were broadly consistent with body-size dimorphism: males produced lower formants and were larger bodied than females. The formant dimorphism also approximated expected dimorphism in areal dimensions between the sexes, which is plausible given the sex differences in both the length and width of the muzzle in baboons, although more detailed data on muzzle dimensions as well as matched body size and acoustic data on the same individuals are needed to test this properly.

In humans, sex differences in formants were also broadly consistent with body-size dimorphism, but the formant differences were substantially greater than would be expected based on the height difference between males and females. However, the formant dimorphism did exactly match the difference in overall vocal-tract length that has been reported for adult males and females (Fant, 1966; Goldstein, 1980). This outcome tends to confirm the link between observable formant profiles and vocal-tract dimensions generally. However, it also confirms that the growth programs that determine vocal-tract cavity dimensions (as with the larynx) can be at least partially decoupled from those that determine overall body size. This result may not be too surprising considering the diversity of pressures that can influence the growth of various bony structures of the skull and face impinging on vocal-tract anatomy quite independently of body size. As a result, it may be fairly common to find only a rather loose connection between adult vocal-tract dimensions and body size *per se*.

This point is further confirmed by analyses within sexes in humans, where formant variation was correlated only with height and only in males and not in females, demonstrating that, however intuitive they might seem, the relationships between body size, vocal-tract dimensions, and voice acoustics are not obligatory.

The fact that formants did correlate with height in males, however, provides some support for Fitch's hypothesis. Indeed, this asymmetric outcome suggests that males but not females might be under specific pressure to provide reliable voice cues to body size. This is certainly consistent with predictions from evolutionary theory about the particular importance of body size in social and sexual competition for male as compared to female mammals. It is also consistent with the finding that the sex difference in pharyngeal cavity length in humans originates at puberty, with additional pharyngeal extension occurring in males at this stage just when body size cuing in the service of social and sexual competition is likely to become especially important (Fitch and Giedd, 1999; Lieberman *et al.*, 2001). Of course, it remains to be established whether the observed formant cues to body size are actually perceptually salient to listeners as would be required by the hypothesis. This is an especially important caveat given that the strongest formant effects obtained here were for F_4 , which was the highest formant measured and therefore necessarily also the weakest and potentially least salient formant to listeners.

B. Implications for theories of speech origins

These issues also bear centrally on theories of speech origins. One of the most popular hypotheses here is Lieberman's (1968, 1984) proposal that human speech and language hinge on a descended larynx. Lieberman has argued that a larynx that lies low in the vocal tract by comparison to closely related primate species creates a uniquely large pharyngeal cavity which in turn greatly expands the range of formant contrasts that can be produced and that are so central to modern language. On this view, the origins of language can be dated in human prehistory by the appearance of a descended larynx in the human fossil record. This is a compelling thesis and an exceptionally productive one because it provides a potential anatomical landmark for recognizing language behavior in our ancestors which is otherwise difficult to infer from only fossilized remains.

There have been numerous challenges to the details of Lieberman's proposal (e.g., Falk, 1975; Boe *et al.*, 2002), but the most global of these has come recently from Fitch (2000). Fitch argues that a descended larynx lengthens the vocal tract, thereby lowering the formant frequencies and signaling larger body size, with attendant advantages in social competition. Although a descended larynx could still facilitate language, through the mechanism proposed by Lieberman, Fitch argues that the original pressure driving laryngeal descent was not actually language but body-size cuing because formant-related language benefits accrue only once the larynx is fully descended, and it is highly unlikely that it descended all at once in a single step. In contrast, the formant-related body-size-cuing benefits of a descended larynx accrue throughout the descent process (an important requirement of sustained directional evolutionary change) with an incrementally lower larynx yielding incrementally lower formant frequencies which advertise progressively larger body size. What keeps the system honest throughout the descent process (an additional important requirement of sustained change) is the correlation between voice formants, vocal-tract length, and actual body size, which at the lower limits is imposed by the impossibility of additional laryngeal descent without serious risk of choking.

In short, humans' descended larynx reflects a history of sustained selection for reliable body-size cuing, and its descended position was only secondarily co-opted for a language function. If true, this would also mean that a descended larynx in the fossil record is no longer a reliable marker of language ability in prehumans as Lieberman proposes.

Fitch has provided some indirect support for his proposal in the phenomenon of permanent and dramatic tracheal elongation in certain bird species that are argued to benefit specifically from body-size exaggeration, and in the dynamic laryngeal retraction, or descent, that occurs during male vocal courtship displays in some deer species that yields lower formant frequencies that nevertheless correlate with male size (Fitch, 1999; Fitch and Reby, 2001; Reby and McComb, 2003).

Some of our own results for humans showing a correlation between formant frequencies and body size in males are also consistent with Fitch's proposal. However, the lack of

any correlations in human females necessarily weakens it. To account for this difference between males and females, one could argue that the relationship between body size and vocal-tract length also holds for adult females but that it is not actually manifest in female formant profiles because women behaviorally modify vocal production in ways that obscure the correlation to voice formants (cf. Whiteside, 2001). However, the fact that the average formant profiles observed for males and females in this study nicely matches reports of the average difference in vocal-tract length between them tends to discount this possibility.

Alternatively, one could argue that the relationships between body size, vocal-tract length, and voice formants used to hold for females in addition to males, but that the relationships in females broke down at some point in human prehistory when the larynx descended to a point where the additional choking risk to females exceeded any modest benefits that they gained through reliable size signaling. Or, it could perhaps be argued that the body-size advantages never applied to females but that female laryngeal descent still occurred because it was leashed to laryngeal descent occurring in males for purposes of body-size cuing. Neither of these two possibilities is inherently parsimonious, however.

Clearly, there are problems with both general scenarios for the evolution of human vocal anatomy and resulting voice acoustics and the particular functional selective pressures that have shaped them. However, their important ramifications both for theories of speech origins and for our understanding of the anatomical bases of variable speech acoustics and the perceptual processes that must effectively normalize them make them critical issues to resolve with future research. In doing this, we will probably need to more fully integrate a number of additional hypotheses for the evolution of human vocal-tract anatomy, some of which are unrelated to either speech or body-size cuing functions *per se*.

For example, Owren (1996) has proposed an "acoustic signature" hypothesis of human laryngeal descent, according to which the larynx descended as a part of a compensatory response to facial foreshortening occurring during human evolution (Lieberman, 1998). This response functioned to maintain a constant vocal-tract length and thereby preserve coevolved systems for the production and perception of formant-related cues to individual and kin identity that are hypothesized to have been of critical importance to social communication in our primate ancestors (Rendall, 1996). Alternatively, Nishimura (2003; Nishimura *et al.*, 2003) has offered a two-step model of laryngeal descent that gives priority to the swallowing (deglutition) function of the larynx and the tightly coordinated hyo-laryngeal dynamics needed to accomplish this, an idea that derives additional support from recent longitudinal data on the trajectory of laryngeal descent that occurs during early human development which also appears to be optimized for efficient deglutition rather than linguistic competence (Lieberman *et al.*, 2001). Ultimately, these and other scenarios will need to be thoroughly integrated if we are to arrive at a comprehensive account of the evolution of human vocal-tract anatomy and voice acoustics and their ramifications for systems of speech production and perception.

ACKNOWLEDGMENTS

We are grateful to the Office of the President and the Department of Wildlife and National Parks in the Republic of Botswana and to the Department of Home Affairs and the Cape Nature Conservancy in South Africa, as well as to R. Seyfarth and D. Cheney, and P. Henzi and L. Barrett in these two countries, respectively. We also thank Professor Anders Löfqvist and anonymous reviewers for comments which improved the manuscript. Research was approved by the Animal Welfare and Human Subjects Research Committees of the University of Lethbridge, and was supported by grants to D.R. from the Natural Sciences and Engineering Research Council of Canada, the L.S.B. Leakey Foundation, and the University of Lethbridge.

- Abercrombie, D. (1967). *Elements of General Phonetics* (Aldine, Chicago).
- Abitbol, J., Abitbol, P., and Abitbol, B. (1999). "Sex hormones and the female voice," *J. Voice* **13**, 424–446.
- Andrew, R.J. (1976). "Use of formants in the grunts of baboons and other nonhuman primates," *Ann. N.Y. Acad. Sci.* **280**, 673–693.
- Apostol, L., Perrier, P., and Bailly, G. (2004). "A model of acoustic interspeaker variability based on the concept of formant-cavity affiliation," *J. Acoust. Soc. Am.* **115**, 337–351.
- Aufdemorte, T.B., Sheridan, P.J., and Holt, G.R. (1983). "Autoradiographic evidence of sex steroid receptors in the laryngeal tissues of the baboon (*Papio cynocephalus*)," *Laryngoscope* **93**, 1607–1611.
- Bachorowski, J.-A., and Owren, M.J. (1999). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech," *J. Acoust. Soc. Am.* **106**, 1054–1063.
- Barrett, L., and Henzi, S.P. (1997). "An interpopulation comparison of body weight in chacma baboons," *S. Afr. J. Sci.* **93**, 436–438.
- Boë, L.-J., Heim, J.-L., Honda, K., and Maeda, S. (2002). "The potential Neandertal vowel space was as large as that of modern humans," *J. Phonetics* **30**, 465–484.
- Bricker, P.D., and Pruzansky, S. (1976). "Speaker recognition," in *Contemporary Issues in Experimental Phonetics*, edited by N.J. Lass (Academic, New York), pp. 295–326.
- Collins, S.A. (2000). "Men's voices and women's choices," *Anim. Behav.* **60**, 773–780.
- Dabbs, Jr., J.M., and Mallinger, A. (1999). "High testosterone levels predict low voice pitch among men," *Personality Individ. Diff.* **27**, 801–804.
- Davies, N.B., and Halliday, T.R. (1978). "Deep croaks and fighting assessment in toads, *Bufo bufo*," *Nature (London)* **274**, 683–685.
- ESPS/WAVES+© (1997). Entropic Research Laboratory, Inc., 600 Pennsylvania Ave. S.E. Suite 202, Washington, D.C., 20003.
- Falk, D. (1975). "Comparative anatomy of the larynx in man and the chimpanzee: Implications for language in Neanderthal," *Am. J. Phys. Anthropol.* **43**, 123–132.
- Fant, G. (1966). "A note on vocal tract size factors and non-uniform F-pattern scaling," *Speech Transactions Laboratory Quarterly Progress and Status Report* **4**, 22–30.
- 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. (1999). "Acoustic exaggeration of size in birds by tracheal elongation: Comparative and theoretical analyses," *J. Zool.* **248**, 31–49.
- Fitch, W.T. (2000). "The evolution of speech: A comparative review," *Trends Cogn. Sci.* **4**, 258–267.
- Fitch, W.T., and Giedd, J. (1999). "Morphology and development of the human vocal tract. A study using magnetic resonance imaging," *J. Acoust. Soc. Am.* **106**, 1511–1522.
- 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**, 179–190.
- Fitch, W.T., and Hauser, M.D. (2002). "Unpacking 'honesty': Vertebrate vocal production and the evolution of acoustic signals," in *Acoustic Communication*, edited by A.M. Simmons, R.R. Fay, and A.N. Popper (Springer, New York), pp. 65–137.
- Fitch, W.T., and Reby, D. (2001). "The descended larynx is not uniquely human," *Proc. R. Soc. London, Ser. B* **268**, 1669–1675.

- Goldstein, U.G. (1980). "An articulatory model for the vocal tracts of growing children," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- González, J. (2004). "Formant frequencies and body size of a speaker: A weak relationship in adult humans," *J. Phonetics* **32**, 277–287.
- Greisbach, R. (1999). "Estimation of speaker height from formant frequencies," *Forensic Linguistics* **6**, 265–277.
- Hamill, P.V.V., Drizd, T.A., Johnson, C.L., Reed, R.B., Roche, A.F., and Moore, W.M. (1979). "Physical growth: National Center for Health Statistics percentiles," *Am. J. Clin. Nutr.* **32**, 607–629.
- Hillenbrand, J., Getty, L.A., Clark, M.J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Hirano, M., Kurita, S., and Nakashima, T. (1983). "Growth, development and aging of human vocal folds," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D.M. Bless and J.H. Abbs (College-Hill, San Diego), pp. 22–43.
- Hollien, H., Green, R., and Massey, K. (1994). "Longitudinal research on adolescent voice change in males," *J. Acoust. Soc. Am.* **96**, 2646–2654.
- Johnson, K., and Mullenix, J.W. (1997). *Talker Variability in Speech Processing* (Academic, New York).
- Kunzel, H.J. (1989). "How well does average fundamental frequency correlate with speaker height and weight," *Phonetica* **46**, 117–125.
- Ladefoged, P., and Broadbent, D.E. (1957). "Information conveyed by vowels," *J. Acoust. Soc. Am.* **39**, 98–104.
- Lass, N.J., Phillips, J.K., and Bruchey, C.A. (1980b). "The effect of filtered speech on speaker height and weight identification," *J. Phonetics* **8**, 91–100.
- Lass, N.J., Kelley, D.T., Cunningham, C.M., and Sheridan, K.J. (1980a). "A comparative study of speaker height and weight identification from voices and whispered speech," *J. Phonetics* **8**, 195–204.
- Lieberman, D.E. (1998). "Sphenoid shortening and the evolution of modern human cranial shape," *Nature (London)* **393**, 158–162.
- Lieberman, P. (1967). *Intonation, Perception and Language* (MIT Press, Cambridge, MA).
- Lieberman, P. (1968). "Primate vocalizations and human linguistic ability," *J. Acoust. Soc. Am.* **44**, 1574–1584.
- Lieberman, P. (1984). *The Biology and Evolution of Language* (Harvard University Press, Cambridge, MA).
- Lieberman, P. (1986). "Some aspects of dimorphism and human speech," in *Sexual Dimorphism in Living and Fossil Primates*, edited by M. Pickford and B. Chiarelli (Il Sedicesimo, Italy), pp. 67–75.
- 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.
- Lieberman, D.E., McCarthy, R.C., Hiiemae, K.M., and Palmer, J.B. (2001). "Ontogeny of postnatal hyoid and larynx descent in humans," *Arch. Oral Biol.* **46**, 117–128.
- Masataka, N. (1994). "Lack of correlation between body size and frequency of vocalizations in young female Japanese macaques (*Macaca fuscata*)," *Folia Primatol.* **63**, 115–118.
- McComb, K.E. (1991). "Female choice for high roaring rates in red deer, *Cervus elaphus*," *Anim. Behav.* **41**, 79–88.
- McFadden, D., and Bracht, M.S. (2003). "The relative lengths and weights of metacarpals and metatarsals in baboons," *Horm. Behav.* **43**, 347–355.
- Morton, E.S. (1977). "On the occurrence and significance of motivation-structural rules in some birds and mammal sounds," *Am. Nat.* **111**, 855–869.
- Nearey, T. (1989). "Static, dynamic, and relational properties in vowel perception," *J. Acoust. Soc. Am.* **85**, 2088–2113.
- Nishimura, T. (2003). "Comparative morphology of the hyo-laryngeal complex in anthropoids: Two steps in the evolution of the descent of the larynx," *Primates* **44**, 41–49.
- Nishimura, T., Mikami, A., Suzuki, J., and Matsuzawa, T. (2003). "Descent of the larynx in infant chimpanzees," *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6930–6933.
- Nordström, P.-E. (1977). "Female and infant vocal tracts simulated from male area functions," *J. Phonetics* **5**, 81–92.
- Owren, M.J. (1996). "An 'acoustic signature' model of speech evolution," *J. Acoust. Soc. Am.* **99**, 2558.
- Owren, M.J., and Linker, C.D. (1995). "Some analysis methods that may be useful to acoustic primatologists," in *Current Topics in Primate Vocal Communication*, edited by E. Zimmermann, J.D. Newman, and U. Jurgens (Plenum, New York), pp. 1–27.
- Owren, M.J., Seyfarth, R.M., and Cheney, D.L. (1997). "The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cynocephalus ursinus*): Implications for production processes and functions," *J. Acoust. Soc. Am.* **101**, 2951–2963.
- Peterson, G.E., and Barney, H.L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Reby, D., and McComb, K. (2003). "Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags," *Anim. Behav.* **65**, 519–530.
- Rendall, D. (1996). "Social communication and vocal recognition in free-ranging rhesus monkeys (*Macaca mulatta*)," Ph.D. dissertation, University of California-Davis.
- Rendall, D. (2003). "Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons," *J. Acoust. Soc. Am.* **113**, 3390–3402.
- Rendall, D., Cheney, D.L., and Seyfarth, R.M. (2000). "Proximate factors mediating 'contact' calls in adult female baboons and their infants," *J. Comp. Psychol.* **114**, 36–46.
- Rendall, D., Owren, M.J., Weerts, E., and Hienz, R.J. (2004). "Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners," *J. Acoust. Soc. Am.* **115**, 411–421.
- Rendall, D., Seyfarth, R.M., Cheney, D.L., and Owren, M.J. (1999). "The meaning and function of grunt variants in baboons," *Anim. Behav.* **57**, 583–592.
- Richman, B. (1976). "Some vocal distinctive features used by gelada monkeys," *J. Acoust. Soc. Am.* **60**, 718–724.
- Riede, T., and Fitch, W.T. (1999). "Vocal tract length and acoustics of vocalization in the domestic dog, *Canis familiaris*," *J. Exp. Biol.* **202**, 2859–2869.
- Sachs, J., Lieberman, P., and Erikson, D. (1972). "Anatomical and cultural determinants of male and female speech," in *Language Attitudes: Current Trends and Prospects*, edited by R.W. Shuy and R.W. Fasold (Georgetown University Press, Washington D.C.), pp. 74–84.
- Saez, S.J., and Martin, P.M. (1976). "Androgen receptors in human pharyngo-laryngeal mucosa and pharyngo-laryngeal epithelium," *J. Steroid Biochem.* **7**, 919–921.
- Stevens, K.N. (1998). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
- Titze, I.R. (1989). "Physiologic and acoustic differences between male and female voices," *J. Acoust. Soc. Am.* **85**, 1699–1707.
- van Dommelen, W.A., and Moxness, B.H. (1995). "Acoustic parameters in speaker height and weight identification: Sex-specific behaviour," *Lang. Speech* **38**, 267–287.
- Whiteside, S.P. (2001). "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," *J. Acoust. Soc. Am.* **110**, 464–478.
- Zhinkin, N.I. (1963). "An application of the theory of algorithms to the study of animal speech," in *Acoustic Behaviour of Animals*, edited by R.G. Busnel (Elsevier, Amsterdam), pp. 132–180.