American Journal of Primatology 37:191–219 (1995)

Vocal Production in Nonhuman Primates: Acoustics, Physiology, and Functional Constraints on "Honest" Advertisement

W. TECUMSEH FITCH[1] AND MARC D. HAUSER[2]
[1]Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island; [2]Departments of Anthropology and Psychology, Program in Neuroscience, Harvard University, Cambridge, Massachusetts

The physiological mechanisms and acoustic principles underlying sound production in primates are important for analyzing and synthesizing primate vocalizations, for determining the range of calls that are physically producible, and for understanding primate communication in the broader comparative context of what is known about communication in other vertebrates. In this paper we discuss what is known about vocal production in nonhuman primates, relying heavily on models from speech and musical acoustics. We first describe the role of the lungs and larynx in generating the sound source, and then discuss the effects of the supralaryngeal vocal tract in modifying this source. We conclude that more research is needed to resolve several important questions about the acoustics of primate calls, including the nature of the vocal tract's contribution to call production. Nonetheless, enough is known to explore the implications of call acoustics for the evolution of primate communication. In particular, we discuss how anatomy and physiology may provide constraints resulting in "honest" acoustic indicators of body size. © 1995 Wiley-Liss, Inc.

Key words: body size, evolution, larynx, primate communication, vocal tract

INTRODUCTION

Vocal production is the process whereby sound is created via movements in sound-producing organs such as the lungs, larynx, nose, and mouth. This paper addresses the acoustic principles and the anatomical and physiological mechanisms underlying this process in nonhuman primates (hereafter "primates"). Although some research has focused specifically on primate vocal production, most of the theoretical basis for understanding production currently derives from the fields of human speech and musical acoustics. Together these fields provide a range of well-researched models presumably capable, with some modifications, of describing many aspects of primate sound production. It will require considerable further

Received for publication March 29, 1994; revision accepted February 7, 1995.

Address reprint requests to Dr. Marc Hauser, Department of Anthropology, Harvard University, 11 Divinity Ave., Cambridge, MA, 02138.

empirical research, however, before similarly detailed models can be constructed to describe the production of primate calls.

For researchers interested in primate communication, an understanding of vocal production is important for a number of reasons. All vocalizations are initially channeled through the production system, which therefore defines the outer limits of acoustic communication by determining what sounds are physically producible. Physiological or acoustic constraints on what is possible will affect the evolution of an entire communication system, creating impediments to be contended with and providing opportunities to be exploited. For example, the primate vocal tract may be unable to modulate sound rapidly enough to "fit" complete thoughts into short-term auditory memory [Lieberman, 1984], limiting primate semantics to the equivalent of relatively simple one-word sentences. Humans have overcome this problem by way of significant anatomical modifications of the vocal tract. On the other hand, the fact that a bird's trachea, unlike a mammal's, is part of its vocal tract, has provided an opportunity for modification of call characteristics presumably unique to the avian class. Birds can increase the length of the vocal tract via tracheal elongation, which appears to have independently evolved a number of times in 56 species of 6 different orders [Clench, 1978; Niemeier, 1979; Johnsgard, 1983]. Thus, the acoustic and anatomical characteristics underlying call production provide the physical context within which other factors (e.g., auditory variables such as frequency resolution, or ecological variables such as foliage density) influence the evolution of a particular species' call system [Green & Marler, 1979; Hauser, in press].

An understanding of vocal production is often a prerequisite for meaningful acoustic analysis or synthesis of primate vocalizations. Several powerful and increasingly available analysis techniques require a basic understanding of production acoustics for their correct use (e.g., linear prediction or LPC, analysis-by-synthesis), and even simple pitch measurements can be made more accurately if production is understood. Further, an increased understanding of call acoustics will allow researchers to synthesize and/or modify vocal signals, enabling them to address a wide range of questions in playback experiments, currently one of the most fruitful means of increasing our understanding of primate communication [Cheney & Seyfarth, 1990].

Research into the neural basis of vocal communication may also profit from better knowledge of peripheral production mechanisms, since the brain controls not the production of frequencies and amplitudes, but the movements of lungs, larynx, tongue, and lips. If research on human speech is any guide, an understanding of the process by which these movements are converted to sound will provide crucial insights into the neural representation of communicative signals [e.g., Liberman & Mattingly, 1985; Lieberman, 1984].

Finally, better understanding of the acoustic and anatomical foundations of primate call production will provide a broader comparative framework for understanding the similarities and differences among anuran, bird, and mammalian communication systems, as well as allowing specific comparisons with human speech and singing. Such a comparative perspective can only enhance our understanding of the evolution of vocal communication in our own and other species [Hauser, in press].

In this paper we provide an introductory tutorial on the acoustics and anatomy underlying vocal production, and a selective review of the literature on vocal production in nonhuman primates. We concern ourselves with peripheral mechanisms, as opposed to the neural basis of vocal production, and we focus on those aspects of primate call production which we feel have been ignored, misinterpreted, or both. In particular, we call attention to the almost complete lack of research on supralaryngeal mechanisms. It is our conviction that most studies of nonhuman primate communication have laid too much emphasis on phonation, and therefore pitch, largely ignoring the filtering properties of the supralaryngeal vocal tract. This oversight may result in an impoverished perspective on the sources of meaningful variation within a vocal repertoire. In the discussion we provide a number of desiderata for future research and finally reexamine some of the acoustic principles described in the first section from a broader evolutionary perspective, paying particular attention to the factors that may or may not constrain vocal signals to be "honest" indicators of a vocalizer's body size.

ACOUSTIC TUTORIAL AND LITERATURE REVIEW

It seems likely that well-established acoustic principles originally developed in connection with either human speech [Fant, 1960] or musical instruments like the clarinet [Benade, 1990] will be adequate to describe the overall acoustic behavior underlying nonhuman primate vocalizations, but our ignorance greatly exceeds our knowledge, and much more research will be necessary before this claim can be confidently made. In this section, we provide a tutorial introduction to the acoustics of vocalization as extrapolated from these fields, noting areas in which our knowledge is particularly limited or where modifications to existing models are likely to be necessary. For the reader interested in more detail, Lieberman and Blumstein [1988] and Sundberg [1987, 1991] provide general, readable overviews of speech and musical acoustics, while the more advanced student will find the classic texts of Morse [1936] and Fant [1960] to provide mathematically detailed treatments of general acoustics and the acoustic theory of speech, respectively.

The Acoustics of the Sound Source: Lungs and Larynx

In general, the lungs provide the airstream for powering mammalian calls. The larynx modulates this airstream, converting this steady, silent flow of air into sound. The lungs are elastic bodies like balloons, and attempt to deflate when fully inflated. The larynx functioned in our earliest tetrapod ancestors as a valve atop the lungs regulating air flow: when the laryngeal valve is fully open (as during quiet breathing) air can move freely in and out of the lungs [Negus, 1949]. When it is tightly closed (as when we prepare to cough), a substantial pressure can build up within the lungs because air flow is prevented. The release of this excess pressure in the form of a cough or sneeze is perhaps the simplest way for the larynx/lung mechanism to produce sound. However, a much wider variety of primate vocalizations result from phonation, an oscillatory laryngeal state between these extremes of opened and closed.

The vocal folds (also called "vocal cords"), which control this laryngeal valve, are stretched from the front to the back of the larynx (Fig. 1). In the front, the folds are fixed to the thyroid cartilage, while the rear portions are attached to the mobile arytenoid cartilages, which can pivot and thus bring the folds together (adduction) or move them apart (abduction), thus changing the size of the glottis (the opening between the folds, through which air flows). If the glottis is loosely closed, an increase in lung pressure will blow them apart. The folds are pulled back together by a combination of two forces: 1) a Bernoulli force created by air flowing between the folds, and 2) the elasticity of the folds themselves [van den Berg, 1958]. These two forces bring the folds together, ready to repeat the cycle again. In this way, an oscillation is set up, creating sound by breaking up the steady air stream from the lungs into a series of puffs. The rate at which the puffs occur, measured in Hertz (abbreviated, Hz: cycles per second), is known as the fundamental frequency of

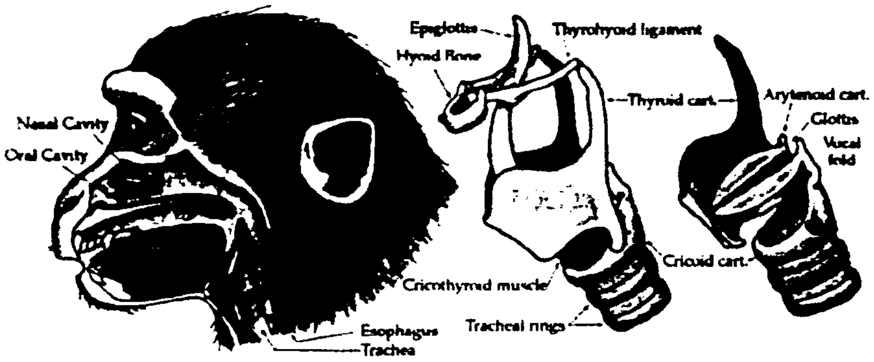

Fig. 1. Schematic side view of primate vocal tract (left); schematic side view of larynx (center); larynx with the thyroid cartilage cut away to reveal the vocal folds and glottis (right).

phonation, or F0. This acoustic parameter represents the primary determinant of the perceived pitch of a vocalization. F0 is controlled by airflow (controlled by the lung-generated pressure beneath the larynx, or "sub-glottal pressure"), and vocal fold tension and vibrating mass [van den Berg, 1968]. In general, longer, thicker folds can produce lower pitches because of their greater mass. Increasing vocal fold tension or subglottal pressure increases F0.

The vocal folds are complex structures: they consist of muscle fibers stretched between the arytenoid and thyroid cartilages (the thyroarytenoid or "vocalis" muscles), which are covered by a mucous membrane whose structure varies considerably among primates. The vocalis muscles directly control the tension of the folds. Muscles external to the folds are also involved in controlling phonation (primarily the cricothyroids, cricoarytenoids, and interarytenoids; these names simply denote the two cartilages to which the muscle attaches). Various combinations of action in these different muscles affect the length (and thus the tension) of the folds, or their thickness (and thus vibratory mass). These changes cause the vocal folds to behave in slightly different ways, making possible different modes or "registers" of phonation. Examples include "fry" mode (a low-pitched, growling sound) and "falsetto" (a very high-pitched, more pure sound). Given the anatomical and physiological similarities in human and primate larynges, the various phonation modes found in humans [described in more detail in van den Berg, 1968; Lieberman & Blumstein, 1988], and perhaps others, are probably available to nonhuman primates. For example, in most nonhuman primates, the portion of the vocal folds bordering the glottis is thin and mechanically independent of the main bulk of the vocal folds formed by the vocalis muscle [Schön Ybarra, in press]. This thin, low-mass portion of the vocal fold, called the "vocal lip," probably behaves like the human vocal fold during falsetto phonation. In human falsetto, the vocal folds are stretched tight (by the contraction of the cricothyroids), and the folds are thin due to a relaxation of the vocalis muscles. This thinness decreases their vibratory mass, as well as the role of the Bernoulli force in closing the glottis. These thin, light folds can vibrate very rapidly, producing a high-pitched, sinusoidal sound perceptually reminiscent of many high-pitched primate calls.

Comparative anatomical and neurophysiological work has documented a number of structural and functional similarities between the human and nonhuman primate larynx [Crelin, 1987; Jürgens, 1990; Negus, 1929, 1949; Sapir et al., 1981; Sutton, 1979]. Thus, for example, Sapir et al. [1981] explored laryngeal function in rhesus macaques by electrically stimulating three muscles, the cricothyroid, geniohyoid, and sternothyroid. Contractions in each of these contributed to an increase in F0, as in humans [Atkinson, 1973]. However, there are a host of interesting anatomical differences in laryngeal anatomy within the primate order, including differences in the structure of the vocal folds (e.g., the vocal lip, described above) and the existence of a wide variety of air sacs of various sizes and shapes opening into various places in the larynx [Negus, 1949; Hill & Booth, 1957; Hast, 1985]. Although these anatomical variations are well-known [see Schön Ybarra, in press, for a recent review], the role they play in structuring call acoustics is almost completely unexplored. Suggestively, Gautier [1971; cited in Schön Ybarra, in press] reported a reduction in sound intensity after puncturing the air sac in *Cercopithecus neglectus*.

There are striking differences in vocal frequency range among different species of primates, not accounted for simply by variation in body weight. Thus, for example, although an adult rhesus monkey weighs approximately 8–12 kg, it produces some calls with an F0 of an adult human male (i.e., 150 Hz) and other calls with an F0 characteristic of small Passerine birds (i.e., 6–7 kHz; Hauser & Marler [1993]). Calls produced in this upper frequency range are not generated by means of a whistle-like mechanism, as would be required of humans, but rather, directly involve vocal fold vibration. Support for this claim comes from inspection of spectrographic representations of the call [see Hauser & Marler, 1993], indicating clear evidence of vocal fold vibration (i.e., vertical striations on the spectrogram).

The puffs of air which occur during phonation result in a *harmonic* sound source, so called because its spectrum contains peaks of energy not only at the fundamental frequency, but also at integer multiples of that frequency. These additional peaks of energy are called harmonics (or overtones). The relative amplitudes of the harmonics play a critical role in determining the tonal quality or timbre of the sound. Sounds with weak harmonics (in terms of their amplitude relative to the fundamental) tend to seem "pure" perceptually (similar to a flute) while sounds with many strong harmonics are perceptually buzzy (like a harmonica). As F0 increases, so does the frequency of the harmonics above it. Thus, high-pitched sounds have fewer harmonics within a particular frequency range than do low-pitched sounds.

In addition to the periodic sound of phonation, noisy aperiodic signals may also be produced at the vocal folds as a result of air turbulence. Turbulence has a long-term spectrum closely approximating that of white noise (equal energy at all frequencies) and sounds similar to the static on an untuned radio (or more pleasantly, the rushing of surf or wind in the trees). Such noise can be produced either at the larynx (as in the "h" sound in "hello"), or above the larynx due to a constriction elsewhere in the vocal tract (as in the "f," "s," or "sh" sounds). Coughs or other rapid releases of air pressure also cause turbulent and therefore noisy airflow. Thus, the sound source in mammalian vocalizations can vary from noisy aperiodic sounds (coughs or fricatives) to the tonal sound of phonation with many permutations in between. The vocal repertoires of nonhuman primate species exhibit both noisy signals [e.g., vervet monkeys: Struhsaker, 1967; gorillas: Harcourt et al., 1993] and relatively pure tonal signals [e.g., cotton-top tamarins: Cleveland & Snowdon, 1981].

The ability of several monkey species to generate a broad range of frequencies [Newman, 1985; Zimmerman, 1990; Cleveland & Snowdon, 1981] may indicate subtle but important differences in the neurophysiological substrate and/or anatomy guiding laryngeal function in human and nonhuman primates. Furthermore, anatomical specializations, such as laryngeal air sacs or the modified hyoid bone of the howler monkey *Alouatta* [Kelemen, 1960; Schön Ybarra, in press], may result in unusual phonation modes or other additions to the basic acoustic principles outlined above. Nonetheless, the gross anatomy (both skeletal and muscular) of the larynx does not differ greatly between humans and nonhuman primates [Negus, 1949], and most of what we know about human phonation is likely to apply to other primates: primate phonation is a vibration of the vocal folds, powered by the lungs, which converts a steady air stream to a series of puffs which we can hear as sound. The rate of puffs (F0) is determined by the size of the folds (which varies between animals) and by their tension, length, and active vibratory mass (which can be varied by a single animal via contractions of the laryngeal muscles).

Functional Significance of F0 Production: An Example

An example of an interaction between production acoustics and social communication is the apparent use of pitch contour as a timing cue during vocal interactions. F0 contour, or "pitch contour," is the variation in F0 over the course of an utterance. Although pitch contours of primate vocalizations have received much study [e.g., Biben et al., 1989; May et al., 1988; Newman, 1985; Petersen et al., 1978; reviewed in Stebbins & Sommers, 1992], they have rarely been considered in the context of the organs which produce them.

In human speech, sentences are virtually always produced during the expiratory phase of the respiratory cycle. Over the course of sentence production, the lungs deflate and there is typically a corresponding decrease in subglottal pressure, particularly at the end of the expiration, in anticipation of the negative subglottal pressure necessary to inhale. Unless actively counteracted, this pressure drop leads to a terminal decrease in F0 as the default or "archetypal" pitch contour [Lieberman, 1967; Titze, 1989]. In humans, these physiological facts lead to a relatively consistent correspondence between frequency drop and the termination of a sentence or phrase [Lieberman, 1967; Collier, 1987; Cooper & Sorensen, 1981; Gelfer, 1987; Gelfer et al., 1987]. As a result, in a large number of languages, production of normal declarative sentences is associated with an initially high and terminally low F0. Thus, terminal F0 changes appear to function as communicative "markers" of sentence completion.

The physiological principles governing F0 control suggest that all organisms with lungs and a larynx will show a frequency drop when producing sufficiently long expiratory vocalizations. If frequency declination does characterize nonhuman animal vocal utterances, then groups of callers might use this pitch change to guide the timing of their vocal interactions (as a "conversational cue"). Hauser and Fowler [1991] analyzed vocalizations produced by vervet monkeys (*Cercopithecus aethiops*) and rhesus macaques (*Macaca mulatta*) during situations involving vocal exchanges or turn-taking. In both species, vocal bouts (i.e., strings of 2–3 calls by one individual) were characterized by a gradual drop in F0. Interestingly, young vervets do not exhibit this pattern, and were interrupted more often than adults. When adult rhesus monkeys were interrupted prior to bout termination, they often increased their fundamental frequency following the interruption, a change not observed during uninterrupted bouts.

Thus, the acoustic and physiological principles underlying the control of F0 lead to fundamental frequency declination in both humans and nonhuman primates. In vervets and macaques, terminal F0 drop may serve an important communicative function (as it does in humans). Observations and experiments on additional species will obviously be needed to determine the generality of this pattern.

Acoustics of the Supralaryngeal Vocal Tract

We now return to our tutorial to describe the functioning of the second major part of the production apparatus: the supralaryngeal vocal tract. We have seen that the airstream generated by the lungs is modulated by the larynx into a series of puffs of air or turbulent noise, thereby providing the source of sound for most vocalizations. This "source signal" then radiates through the airways of the pharynx, mouth, and nasal cavities on its way out through the lips and nostrils. These cavities are collectively known as the supralaryngeal vocal tract or simply, the vocal tract (Fig. 1). The vocal tract acts as a filter: it selectively removes certain frequencies from the input source signal. Although the mathematics necessary to accurately model real vocal tracts are complex, and can only practically be calculated by computer programs, important insights can be derived from the consideration of much simpler models. We therefore begin with a mathematical description of the vocal tract as a simple tube full of air. Many of the same basic principles underlie the acoustic behavior of simple tubes and vocal tracts with much more complex shapes. We will describe the acoustics of these more realistic models qualitatively, leaving the difficult underlying mathematics to advanced texts [e.g., Morse, 1936; Fant, 1960].

Consider a simple hard-walled tube of uniform diameter. A sound pressure wave originates at one end (we will call it the glottis end to make the vocal tract analogy explicit), and travels down the length of the tube to reach the lip end after an amount of time dependent on the speed of sound and the length of the tube. When the pulse reaches the lips, some of the energy will radiate out into the air, while some will be reflected back into the tube, to travel in the opposite direction. This reflected energy will arrive back at the glottis, where it will again be reflected. If another glottal pulse happened to be leaving the larynx at precisely that moment, the two pressure waves would coincide, and the reflection would add energy to the newly admitted pulse. In contrast, if the two were out of phase, they would partially cancel. In general, then, certain frequencies will be emphasized or "encouraged" by a tube of a certain length, while others will be attenuated or "discouraged." The tube's preferred frequencies are known as its resonances, and they are largely determined by the length of the tube ("resonance" is the general term from physics; when referring to the vocal tract the basically equivalent term "formant" is used).

The resonances of a tube of a certain length are also determined by the nature of its two endpoints, which mathematically constitute the "boundary conditions" of the system. Certain wavelengths of sound fit into the tube in a way consonant with its boundary conditions, while others do not. The boundary conditions are determined by the tube geometry: a closed end will prevent free air movement and create a region of high pressure, while an open end will allow free movement of air and zero pressure build-up. Thus different types of simple tubes will have different resonance patterns (Fig. 2).

Figure 2 shows the first three resonances of three types of tube: one open at only one end (like the previous example), one open at both ends, and one closed at both ends. The tube with only one end open has a pressure maximum at one end and a zero at the other, and the lowest frequency which can meet these boundary conditions can fit one-fourth of its wavelength in the tube. As a result, such a tube is called a quarter-wavelength resonator. As the diagram shows, other higher frequencies can also meet these boundary conditions, and they are related as odd integer multiples to the lowest resonance. The other two tubes have either highs at both ends or zeros at both ends, and thus behave as half-wave resonators: the lowest resonance has a wavelength twice the length of the tube.

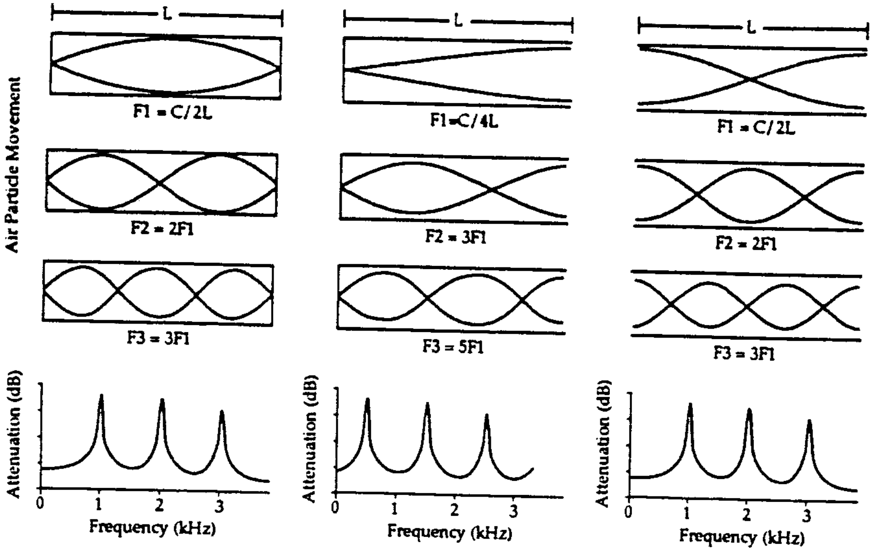

Fig. 2. Resonance frequencies for three types of simple tubes. The series on the left illustrates the first three resonances (indicated by F1, F2, F3) for a tube closed at both ends. The middle series is for a tube closed at one end. The series on the right is for a tube open at both ends. At the bottom is the transfer function for each tube type.

To make these examples more concrete, it is useful to calculate the first three resonances of a simplified model vocal tract 17.5 cm long (approximate length of a human vocal tract). The wavelength of a sound (λ) is inversely proportional to its frequency (f) according to Equation 1:

$$\lambda = \frac{c}{f}, \tag{1}$$

where c is the speed of propagation of sound in air (about 350 m/sec). Thus, for example, the wavelength of a 100 Hz tone ($F\o$ of a deep-voiced human male) is 3.5 m while the wavelength of an 8 kHz marmoset phee is about 4 cm.

During phonation, the glottis remains largely closed, so the quarter-wavelength resonator is the appropriate choice of tubes. As discussed above, the wavelength of the lowest resonance is four times the length of the tube ($L = 17.5$ cm), so $\lambda = 4L = 70$ cm. Combining this with equation 1 above we get Equation 2:

$$f = \frac{c}{\lambda} = \frac{c}{4L} = \frac{35{,}000\frac{\text{cm}}{\text{sec}}}{70_{\text{cm}}} = 500 \text{ Hz}. \tag{2}$$

We can see from Figure 2 that the next resonance is three times this frequency, i.e., $f = 3c/4L = 1{,}500$ Hz. More concisely, the entire set of resonances can be calculated from Equation 3:

$$f = \frac{(2n+1)c}{4L}, \text{ where } n = 0,1,2,3\ldots \tag{3}$$

The graphs at the bottom of Figure 2 are graphical depictions combining all of these resonances. These graphs represent the acoustic *transfer function* for each tube, illustrating which frequencies will be encouraged or discouraged by a tube of a particular conformation. Transfer functions provide a concise and extremely useful picture of the acoustic behavior of any resonant system (including vocal tracts much more complex than those pictured in Fig. 2).

Figure 3 illustrates the effect of combining the laryngeal source signal with the transfer function of the simplified vocal tract from the middle of Figure 2. The source signal has an $F\o$ of 150 Hz, so its spectrum contains energy at that frequency and at each harmonic above it, with gradually decreasing amplitude. The vocal tract transfer function has been continued into a higher frequency range, so now the first five resonances are visible (in accordance with Equation 3). Because there is a peak in the transfer function at 500 Hz, any energy input from the larynx at a frequency near 500 Hz will emerge at the lips with high amplitude. However, energy at 1,000 Hz (1 kHz) falls in a valley of the transfer function, and will accordingly be heavily attenuated when passed through the tube.

It is important to realize that the transfer function is not the same as the output spectrum. The output spectrum is the input signal *as filtered by* the transfer function, so the spectrum of the output signal depends crucially on both the transfer function and the input signal. If the input signal is a sine tone (energy at a single frequency), the output spectrum will accordingly have only a single frequency, and will reveal very little about the transfer function; if the input is white noise (with energy at all frequencies), the output spectrum will closely resemble the transfer function. Thus, the widely spaced harmonics of a high-pitched source functionally sample the transfer function of the vocal tract at relatively sparse intervals, and thus may reveal less about it than a low frequency or noisy source [Ryalls & Lieberman, 1982]. This point is particularly important when performing acoustic analyses of high-pitched calls typical of smaller primates, since the source harmonics may be too sparse to provide much information about the vocal tract transfer function. One way around this measurement problem is to excite the vocal tract with an artificial source of low-frequency sound, such as the artificial larynges used by laryngectomy patients [as in Sundberg, 1975]. Such a device, simply placed against the skin of the neck and buzzed, would allow a very accurate determination of the transfer function in captive or tame animals.

Although the tube model we have examined so far obviously involves drastic simplifications, it illustrates a number of important points which hold for the more complex shapes of real vocal tracts. Because we are dealing with propagation down the length of the tube, bends or curves make little acoustic difference (this is true as long as the diameter of the tube is small relative to the wavelength of the sound in question, probably a safe assumption for most primate calls). Similarly, within broad limits, the diameter of the tube has little effect on its resonant frequencies (increasing tube diameter slightly increases the acoustically effective length of the tube; see Morse [1936] for more on this “end correction”). The critical factors determining the transfer function for simple tubes of constant diameter are their length and boundary conditions. Boundary conditions are controlled by the size of the glottal or lip openings. Vocal tract length can be modified in two ways: it can be increased at the lip end by tensing the orbicularis oris muscle and thus rounding and protruding the lips, or at the glottis end by raising or lowering the larynx (using the laryngeal strap muscles: the sternothyroid and sternohyoid). In either case, the effect of lengthening the vocal tract is the same: it lowers all the resonant frequencies of the tube.

Actual primate vocal tracts, of course, are not tubes of constant diameter. Instead, their diameter varies along their length. We can model a real tract as a

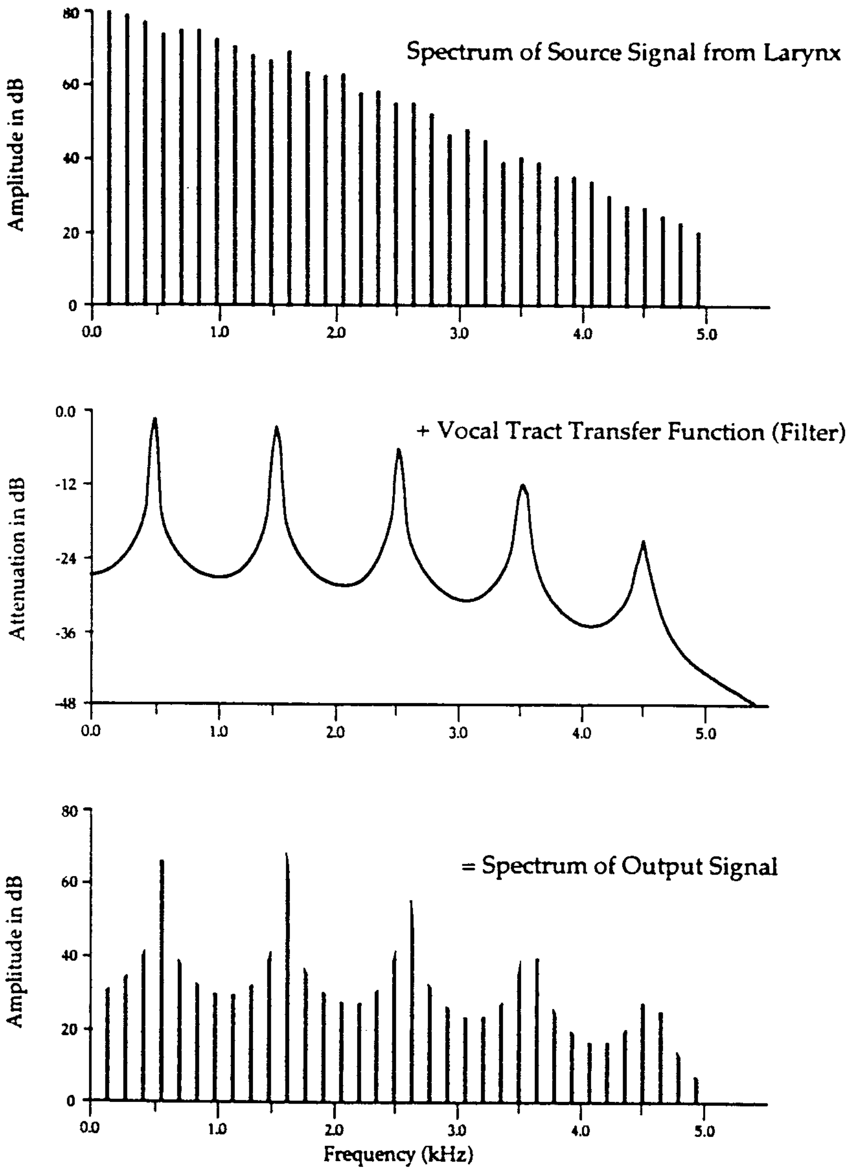

Fig. 3. The spectrum (with frequency [kHz] on the x-axis, amplitude [dB] on the y-axis) of the laryngeal source signal (top); the transfer function for a model vocal tract of length 17.5 cm (amplitude attenuation [dB] on the y-axis) showing the first five resonances (middle); and the result of filtering the source signal through the vocal tract filter (bottom).

connected series of many short lengths of tube which vary in diameter. The specific shape of each cross-section is unimportant, because the cross-section's acoustic properties are almost completely determined by its overall area and not its detailed shape. Thus the resonant frequencies of a vocal tract can be quite accurately determined by modeling the tract as a series of cylindrical tube sections, whose cross-sectional area varies as a function of distance from the glottis.

As a pulse of pressure travels from the glottis up the tube, it passes between sections of varying diameter. At each transition between sections, part of the pulse will be transmitted onward, and part reflected backward. The greater the change in area from one section to the next (technically, the greater the impedance difference), the more reflection will occur. These reflections create resonant frequencies in the tube. Calculation of the transfer function for these more elaborate tubes requires complex mathematics beyond the scope of this article [see Fant, 1960; and Morse, 1936] and practically speaking is performed by computer programs. However, it is important to note that, in general, the resonant frequencies of a real vocal tract will *not* be evenly spaced in the frequency domain (in contrast to those of the simple tube of constant diameter considered above).

Formant frequency patterns play a primary role in human speech [Lieberman, 1984; Lieberman & Blumstein, 1988; Remez et al., 1981], and the vocal tract shapes corresponding to most of the different sounds of human speech are well documented [Baer al., 1991; Fant, 1960; Nearey, 1978]. Production of many of these sounds requires a vocal tract configuration unique to humans: a nearly 90° connection between the pharyngeal and oral cavities [Crelin, 1987; Lieberman et al., 1969; Lieberman, 1984; Negus, 1949]. In nonhuman primates, the pharyngeal cavity is short and positioned high in the neck. Typically (during quiet breathing), the epiglottis and larynx are raised and sealed into the back of the nasal cavity. This anatomical difference means that there is virtually no bend in the nonhuman vocal tract between larynx and lips. Although the bend in the human vocal tract is acoustically unimportant (as discussed earlier), it is anatomically crucial, because it functionally partitions the tract into two independent sections, allowing the tongue body two degrees of freedom. The human tongue can move up or down (modifying the size of the oral cavity) or forward and backward (modifying the pharyngeal cavity).

A set of muscles encircling the pharynx, the pharyngeal constrictors, can shrink the pharynx and pull the tongue back (as in the vowel [a] in "father"). The tongue can be pulled forward by the genioglossus muscle to make vowels like [i] (as in "beet"). The tongue can be pulled backwards and up in the vocal cavity to make vowels like [u] (as in "boot") using the styloglossus muscle. Although the same muscles exist in other primates, they cannot have the same acoustic effect due to the difference in vocal tract anatomy. It is has been suggested that these anatomical differences are critical for understanding why nonhuman primates cannot produce human speech, and more importantly, why they are constrained with regard to the range of sounds and sound contrasts that can be produced [Borden & Harris, 1984; Lieberman, 1984, 1985, 1991; Lieberman et al., 1969].

At present this claim is difficult to assess given the paucity of empirical data on supralaryngeal function in nonhuman primates. The types of muscular maneuvers available, the cross-sectional area functions possible, and their acoustic consequences remain largely mysterious for nonhuman primates. The only explicit attempt to assess these factors is the pioneering study by Lieberman et al. [1969]. Here, the acoustics of the macaque vocal tract were studied by physically exploring the anatomical range of potential articulations in anesthetized animals, followed by computer modeling of the acoustics of these observed vocal tract shapes. They found the range of vowels producible by the macaque to be quite limited with respect to humans. In particular, the standard nonhuman primate vocal tract appears incapable of producing the "point vowels" [a], [i], and [u], described above, which are universal in human language [Maddieson, 1984].

Although the standard primate vocal tract is limited in the formant frequency patterns it can produce, several studies suggest that formant *perception* is highly developed in nonhuman primates. Sommers et al. [1992] reported that Japanese macaques (*Macaca fuscata*) showed greater sensitivity to small changes in formant frequencies than did humans in some studies. Moody et al. [1990] summarize results for perception of formants in Japanese macaques.

The Nasal Tract

In addition to the oral and pharyngeal cavities, the nasal cavities and sinuses can play an important role in the acoustics of the vocal tract [House & Stevens, 1956; Fant, 1960; Dang & Honda, 1994]. Because these cavities are mostly encased in bone, their acoustic properties are relatively static during a given call (although they can change significantly with the swelling of nasal mucosa and due to nasal congestion). The nasal passage can be modeled as a tube of varying cross-sectional diameter, like the oral cavity described above. The nasal tract stretches from the nostrils back to the "oropharyngeal port," which is where the pharynx branches into the nasal and oral cavities. The coupling between these cavities is controlled by a moveable flap of tissue called the velum which (in humans at least) can close off the nasal tract (for a purely oral sound), assume a neutral position which allows sound to propagate through both cavities (in nasal vowels), or close off the mouth (velar nasals like the "ng" sound at the end of "song") resulting in a sound that propagates through the nose alone [Fant, 1960]. Because most primates keep the larynx locked into the nasal cavity during breathing [Lieberman, 1984], it seems likely that at least some calls follow this third pattern, and thus show little spectral change over the course of the call.

The acoustic behavior of the combined oral and nasal cavities is complex, and depends on which of the above options are utilized [see Fant, 1960; or Dang & Honda, 1994, for more detail]. The distribution of these three vocal tract options is unknown in nonhuman primates. In humans, if both the oral and nasal tract are open (both at the lips and nostrils and at the velar port), F1 (i.e., the first formant frequency) is weakened (relative to oral sounds), nasal formants are added, and "anti-resonances" or zeros (valleys in the transfer function) are introduced (nasalized vowels, as in the French "bon" are produced this way). Although articulatory data were lacking, Hauser [1992] found that some macaques consistently produce coos which sound nasal, and have acoustic features consistent with a nasalized articulation. The male proboscis monkey, *Nasalis larvatus*, has a greatly enlarged bulbous nose which apparently functions in vocalization: during "honking" vocalizations, the long drooping nose stiffens slightly with each "honk" [Napier & Napier, 1985]. To firmly establish that the nasal cavity is involved in filtering the laryngeal source, it would be necessary to record nasal airflow, or block the nasal cavity and measure the resulting changes in spectral structure [e.g., see Suthers et al., 1988, for an example with bats].

If the mouth is closed at the lips and coupled to the nasal cavity (as in the [m] sound), it acts as an enclosed resonant cavity which shifts the nasal resonances and introduces anti-resonances. This articulation appears to be used in the macaque "girney," a call which sounds remarkably speech-like, and appears to be used conversationally among female macaques while swapping or inspecting each others' infants [Hauser & Fowler, 1991]. The girney shows substantial fluctuations in the strongest resonant frequency, which appear to be related to jaw movements that rapidly change the shape of the sealed oral cavity.

In humans, the nasal passages and sinuses vary considerably in shape and asymmetry between individuals [Dang & Honda, 1994], suggesting that effects of the nasal tract could be used as cues for individual recognition. This may be especially true for primates, since the nasal passages in most species are much larger and more complex than those of humans.

Combining Source and Filter: Feedback and Non-Feedback Systems

Our discussion thus far has treated the operation of the glottal source and vocal tract filter as independent systems. But, there are at least two qualitatively different ways a system composed of larynx and vocal tract can behave. Either one (or both) may be operative in the vocal repertoires of a given nonhuman primate species. If the fundamental frequency of the source is near a resonant frequency of the filter, the two will interact: it may be much easier to maintain vibration at a resonant frequency than otherwise. Such a system, where the fundamental frequency is affected or even determined by the resonant frequencies of the system, can be called a "feedback" system [Sundberg, 1991]. Most wind instruments (e.g., clarinets, trumpets, oboes, tubas, etc.) work in this manner: the pitch of the instrument is controlled by the length of the resonant tube (this is particularly clear in the slide trombone). The flared shape found at the end of these instruments (the bell) is optimized to reflect sound back into the tube, thus augmenting the feedback and strengthening the coupling.

The second case is characteristic of human speech and is described as a non-feedback or non-interactive system. If the lowest resonances are considerably higher in frequency than F0, there will be little or no interaction between source and tract, and the F0 of the laryngeal source will be determined solely by factors intrinsic to the larynx/lung system (e.g., the tension on, and size of, the vocal folds or the subglottal pressure). This represents a non-feedback system, which exhibits independence of source and filter. In addition to the human voice, accordions and harmonicas fall into this category.

The source-filter theory of speech [Fant, 1960; Müller, 1848] takes as its starting point the independence of laryngeal source and vocal tract filter (for a typical male human phonating at 120 Hz, the lowest formant is around 250 Hz, so this assumption is valid). In this situation, the vocal tract acts simply as a filter modifying the spectrum of the glottal source, and changes in this filter have no significant effects on F0 or the glottal waveform. In speech, the resonant peaks in the transfer function are referred to as formants (numbered F1, F2, etc.), a term which connotes the independence of source and filter (unlike the term resonance, which suggests the possibility of their interdependence).

The source-filter theory has proven very successful as a model of the speech production process. There are, however, some difficulties with the assumption of source/tract independence in speech [see Flanagan, 1968; Bickley & Stevens, 1986; Klatt & Klatt, 1990]. Several different types of interaction between the glottal source and F1 have been identified. During the open portion of the phonation cycle, the trachea can be coupled to the supralaryngeal vocal tract and modify its spectral shape. Similarly, the opening and closing of the glottis changes the boundary conditions of the entire vocal tract, and thus affects the formants in a pitch-dependent way. Nonetheless, these findings all represent minor modifications of the source/filter theory as laid out in Fant [1960], which is currently widely accepted as *the* theory of speech production.

In singing, F0 is commonly much higher than in speech, increasing the likelihood of source-tract interactions. Professional soprano singers can be observed to vary the first formant (via mouth opening) to coincide with F0 [Sundberg, 1987], resulting in a thousandfold increase in radiated energy over normal speech. Tenors appear to avoid this strategy of "formant tuning," perhaps to maintain a male quality in the voice [Titze et al., 1994]. Schutte and Miller [1986] measured supra- and sub-glottal pressure in sopranos singing high in their range (i.e., near 1 kHz), and found clear evidence for acoustic coupling between the trachea and vocal tract: during part of the phonatory cycle, air actually streamed back towards the lungs. Whether this results in source/tract coupling is still unclear [see Sundberg 1979, 1987, for further discussion]; it is clear, however, that this "back pressure" greatly increases the loudness and subjective ease of vocalization. A similar vocal strategy may be adopted by primates during particularly high or loud calls, some of which are important in regulating spacing between groups [Waser, 1977], or in the echolocation calls of bats used for prey detection, recognition, and capture [Hartley & Suthers, 1988; Suthers, 1988; Suthers & Fattu, 1973].

Given the wide range of F0s and vocal tract lengths seen in nonhuman primates, the assumption of source/tract independence may not hold for all primate calls. The extent to which the vocalizations of other primates (and other mammals) exhibit feedback between source and filter has unfortunately received little empirical study, and even the most basic issues remain unresolved. While the wind instrument model predicts enhanced vibration at resonant frequencies [Benade, 1990], some theoretical models suggest that it will be difficult for the vocal folds to vibrate at the same frequency as F1 due to "stalling" of the folds [e.g., Ishizaka & Flanagan, 1972; van den Berg, 1968]. Sundberg [1979] suggests that the effect of source/tract coupling in singing will depend upon the frequencies involved and the register of phonation (e.g., falsetto or modal), with high fundamentals showing positive coupling.

Addressing this issue requires, at a minimum, simultaneous measures of laryngeal and vocal tract functioning in a variety of calls and species. The ideal data set would require imaging the moving vocal tract (via X-ray cinematography or FMRI), and electroglottography (which provides a trace of the vocal fold movements), along with detailed acoustic measurements of both F0 and vocal tract resonances as they change through time. However, considerable insights may be obtainable from acoustic measurements alone.

Bauer [1987] made a good start in this direction with his exploration of the relationship between changes in the configuration of the supralaryngeal tract and call structure. He performed a frame-by-frame analysis of a film sequence of a male chimpanzee's vocal transition from submissive "screams" to aggressive "waahbarks," and measured the vertical distance between the teeth and lips at the center of the mouth. Both lip and teeth opening were positively correlated with F0; lip opening, however, accounted for the most significant proportion of the variation. Specifically, screams were associated with lip retraction and a relative increase in F0. Waahbarks, in contrast, were associated with lip protrusion and a relative decrease in F0. Based on these analyses, Bauer concluded that changes in F0 were the result of changes in vocal tract configuration.

There are, however, potential problems with Bauer's interpretation of the data. If source and tract are coupled, the magnitude of change in F0 (maximum = ~800 Hz) is unlikely to result from changes in lip opening alone. The acoustic differences between the chimpanzee's scream and waahbark are more likely to be due to changes in the length of the vocal tract—as measured by changes from lip retraction to protrusion—rather than the changes in mouth opening which Bauer measured. For both screams and waahbarks, the vertical distance between the lips can be the same in both retracted and protruded positions. Furthermore, the changes in F0 reported by Bauer [1987] could be due to unmeasured changes in laryngeal muscle tension or subglottal pressure rather than changes in vocal tract shape, and not the result of source/tract coupling. Simultaneous measurements of F0 and vocal tract resonances are required to make these explanations less ambiguous.

A series of more recent studies explored the effects of articulation on the acoustic structure of rhesus monkey vocalizations under semi-natural conditions [Hauser, 1992; Hauser et al., 1993; Hauser & Schön Ybarra, 1994]. Hauser et al. [1993] studied the relationship between mandibular position and call morphology for a small sample of individuals producing coos, a call used in a variety of affiliative contexts. Regression analyses revealed that changes in mandibular position were highly correlated with changes in the first resonance frequency, but not with changes in the fundamental frequency. Thus, at least during coo production, the filtering effects of the macaque supralaryngeal tract appear to operate independently of the laryngeal source.

Hauser and Schön Ybarra [1994] examined the role of lip configuration in vocal tract acoustics by administering xylocaine to the upper and lower lips. The study involved two captive but unrestrained rhesus monkeys who, following injections of xylocaine, were allowed to interact with their social group. One individual produced a relatively large sample of coos, whereas the other individual produced a majority of "noisy screams" [Gouzoules et al., 1984]. During the production of coos, the lips are normally protruded, which lengthens the vocal tract. In contrast, the lips are normally retracted during the production of noisy screams, thereby shortening the tract. By blocking these lip movements, the xylocaine treatment leads to coos produced with a shorter vocal tract (relative to the normal, elongated, tract) and to noisy screams produced with a relatively longer vocal tract. If source and filter are independent, length of the vocal tract will have minimal effects on F0 and call duration, but will result in significant changes in the resonance frequencies of the call (a longer vocal tract resulting in lowered resonance frequencies).

Acoustic analyses of coos revealed that xylocaine had no effect on fundamental frequency or call duration, but was associated with a significant increase in resonant frequencies, as predicted. For the noisy screams, there were no detectable differences in call morphology between calls produced with and without xylocaine. This could result from the changes being too small to be statistically significant, or to a compensatory articulation (e.g., raising the larynx) which negated the lip effect. At present, it is not possible to distinguish between these alternatives. These experiments are similar to those performed with "bite blocks," which are objects placed in the mouth to disrupt the normal patterns of human speech. Such experiments have provided significant insights into the neuromotor organization of human speech [Lindblom et al., 1979], and could be further explored in nonhuman primates [see discussion in Hauser & Schön Ybarra, 1994].

In general then, the few existing observations are consistent with the idea that the vocalizations of nonhuman primates thus far studied are produced with the larynx and vocal tract behaving independently, and can thus, like human speech, be modeled as non-feedback systems. However, Bauer's [1987] data and the wide range of fundamental frequencies observed in nonhuman primates suggest the possibility of source-tract interactions. Such interactions would necessitate acoustic models, as well as techniques of analysis and synthesis, quite different from those appropriate for human speech. The field is clearly in dire need of more data on this subject.

Summary

To summarize our discussion of the acoustics of vocal communication, call production in primates and other mammals originates at the larynx, which converts a steady stream of air from the lungs into a series of puffs of air (phonation), or into turbulent noise. This signal is known as the glottal source. Because the gross anatomy of the larynx is relatively stable across primate species, much of our current understanding of human laryngeal function will probably be applicable to other primates. However, variation in larynx dimensions, the presence of laryngeal air sacs and the vocal lip, and great differences in the fundamental frequency of phonation may result in significant quantitative differences in the glottal source across primate species.

The next step in the production pathway involves the transmission of the glottal signal through the vocal tract (including both oral and nasal branches), the length and shape of which determine a set of resonant frequencies (its absolute volume and overall curvature are relatively unimportant). The vocal tract can modify the glottal signal in one of two ways. If the resonances of the vocal tract are near the fundamental frequency of the source, the behavior of the larynx may be influenced by the vocal tract resonances. This could be called a feedback vocal tract, and its acoustic behavior would be similar to that of many wind instruments and perhaps of humans singing high notes. Alternatively, if the resonances are considerably higher in frequency than F0, the source and tract will be essentially independent, and the vocal tract will simply act as a filter which lets energy pass through at its resonances (formants). Then, F0 will be determined exclusively by the larynx and lungs. Such a non-feedback model applies to normal human speech and some wind instruments (e.g., harmonica).

DISCUSSION

Generalizations about the mechanisms underlying nonhuman primate vocal production must be treated cautiously due to the paucity of current knowledge. To fill this gap, it is necessary to obtain considerable information about the range of inter-specific, intra-specific, and intra-individual variability in call morphology, as well as about the articulatory gestures used to produce them. For example, how flexible (or stereotyped) are the gestures used to produce a given call by a given individual? If calls result from highly stereotyped production routines, a single call per individual might potentially represent an adequate sample for acoustic analysis. It seems much more likely, however, based on human data, that there is considerable variability between exemplars of the "same" call, and that primates may have the ability to use different articulatory gestures to obtain the same acoustic target. More broadly, although there is suggestive evidence of dialectal variation in red-bellied tamarins [Maeda & Masataka, 1989] and chimpanzees [Mitani et al., 1992], far too few studies have explicitly looked for geographic variation in call structure between populations of the same species to determine whether such variation is typical or atypical of the nonhuman primates.

There is thus ample room for improvement in our knowledge of the most basic and descriptive aspects of primate vocalizations. Despite this need for basic research, a theoretical perspective is likely to sharpen debate, encourage experimentation, and broaden the appeal of the results. We believe that our understanding of certain issues in primate vocal production is complete enough for some theoretical points to be developed in more detail. In the next section, therefore, we outline a theoretical perspective which will hopefully help provide a richer framework within which to interpret future data on vocal production.

Audiomotor Coevolution and "Truth in Advertising"

Most of the sounds produced by the vocal organs are signals to other animals (possible exceptions include coughing and bat echolocation pulses). Such vocaliza-signaler. tions have typically evolved to communicate in ways that are adaptive to the Similarly, perceptual systems typically evolve so as to optimize the reception of information adaptive to the perceiver [Marler, 1955, 1967; Green & Marler, 1979]. In many instances the selective forces acting on sender and perceiver may be at cross purposes, memorably dubbed "manipulation" and "mind-reading" respectively, by Krebs and Dawkins [1984]. For example, an advertising male is often under selection to manipulate potential mates into perceiving him as healthier and stronger than he really is, while perceivers will be selected to accurately perceive his actual quality.

Krebs and Dawkins [1984] suggested that such conflicts of interest between signaler and perceiver might lead to escalatory "arms races," and thus play a potent role in the evolution of communication systems (in any sensory modality). A cycle would typically start with the existence of some latent relationship between an important quality (size, territory quality, state of arousal, etc.) and a perceptual cue (color, loudness, fur sheen, etc.). Pioneering "mind readers" develop perceptual mechanisms which enable them to make use of this information for their own benefit. For example, let us say that an attacking animal needs to expose its claws before an attack, so that there is a reliable correlation between claw-baring and striking. A "mind-reader" who retreats upon noticing its rival exposing its claws can thus evade damaging attacks.

Once such a perceptual mechanism is in place, however, it may pave the way for manipulation by signalers. In our example, the attacker could bare its claws even when not intending to attack, scaring its rivals away without the need for a fight. The signal (claw-baring) has become decoupled from its original referent (attacking), and claw-baring has become an effective "bluff." Now there will be selection on perceivers, either to simply ignore claw-baring, or to distinguish the "real" claw-baring, which precedes an attack, from "fake" claw-baring, which doesn't.

Many signals may be open to bluff in this manner. For a signal to be stable evolutionarily, there should be a reliability component which prevents such cheating [Zahavi, 1977; Clutton-Brock & Albon, 1979; Davies & Halliday, 1978; Krebs & Dawkins, 1984]. Zahavi [1975, 1977] proposed that reliability is typically maintained via a "handicap principle," suggesting that a signal must be difficult to perform (i.e., costly relative to the individual's condition) if it is to remain reliable over evolutionary time. This is not the case, however, if anatomical or physiological constraints impose reliability. Constraints that prevent manipulative modification of the signal, or maintain it within manageable limits, could enforce a form of "truth in advertising," ensuring that such a physically constrained signal continues to provide useful information for perceivers.

It is also possible for a perceptual mechanism to be maintained despite its occasional manipulation, due to its enduring value in other more general contexts; this is just the sort of perceptual mechanism that manipulators are looking for. For example, one important cue to the size of an object is the visual arc its image occupies on the retina at a certain distance (other, more sophisticated cues depend on binocular vision, comparison with other objects, etc.). Piloerection in mammals ("raising the hackles") and feather erection in birds are widespread behaviors that take advantage of this perceptual mechanism, allowing signalers to increase their apparent body size with no increase in actual mass. Many or most mammalian and avian species make use of this manipulative trick, often during aggressive interactions [Darwin, 1872; Fitch, 1994, and references therein], but we nonetheless continue to use visual arc as a cue to size. Apparently the disadvantage resulting from occasional manipulation of the perceptual mechanism equating visual arc and size is offset by the broad usefulness of this mechanism in estimating the size of trees, rocks, food, and other objects.

Acoustic Cues to Body Size: Laryngeal Cues

In the remainder of this discussion we will apply the "mind-reading/manipulation" perspective sketched above to the evolution of vocal signals that provide cues to the vocalizer's body size. Body size is an extremely important characteristic of vertebrates influencing a wide variety of factors including basic metabolism [Schmidt-Nielsen, 1984], fecundity [Brown et al., 1993], and competitive or mating success [Darwin, 1871; Wiley, 1974; Brown & Maurer, 1986]. It seems likely that some aspects of vocalizations will provide information about body size. The dimensions of the sound-producing organs directly influence their acoustic behavior, as discussed earlier. If the size of these organs is closely tied to body size, the sounds an animal makes will contain cues to its body size. The importance of size in social interactions should eventually lead to perceivers who make use of these cues. Furthermore, the covert nature of sound (the fact that a vocalizer can be heard while unseen) may make it particularly suited for manipulation. We might expect differences in "mind-reading" and "manipulative" acoustic strategies in species that live in dense environments with reduced visual contact (e.g., rainforests) as opposed to more open environments where interactors could rely more heavily on visual cues. In the next section we discuss the factors which contribute to the reliability (or manipulability) of various possible vocal cues to body size.

A commonly cited example of a body size cue is the purported link between large body size and low F0 suggested by Darwin [1872] and elaborated upon by Morton [1977, 1982], Ohala [1983, 1984], and Davies and Halliday [1978]. Gouzoules and Gouzoules [1990, p. 325] wrote that "the generality of the relationship between body size and the frequency-related acoustic features of calls is widely accepted." For a size/pitch relationship to exist, two conditions must be satisfied. Larger animals must have larger vocal folds, and larger vocal folds must result in lower fundamental frequencies. As we discussed above, the second condition holds, with the proviso that larger folds don't necessarily make lower sounds: bass singers have longer folds than sopranos [Sawashima et al., 1983], but by increasing tension and switching to falsetto they can still produce tones higher than a soprano's lowest note. Nonetheless, the lowest note either singer can produce is closely tied to his or her vocal fold length.

Surprisingly, support for the second condition (a link between total body size and vocal fold length within a species) appears to be much more tenuous. Despite popular belief, numerous studies have failed to find any relationship between body size and voice pitch in humans [Lass & Brown, 1978; Künzel, 1989; Sawashima et al., 1983]. Vocal fold length and voice F0 are strongly correlated with total larynx size [Williams & Eccles, 1990]. Male humans undergo an enlargement of the larynx at puberty, with the result that adult males have significantly longer vocal folds, and thus lower pitch, than adult females. However, this pitch difference is far greater than the relatively modest size difference between men and women; within either gender vocal fold length shows a correlation to neck circumference, not body size [Sawashima et al., 1983].

The larynx is a soft cartilaginous organ suspended at the top of the trachea in the neck, so growth of the larynx is not particularly constrained by the dimensions of the rest of the body. Thus the larynx and vocal folds can show an increase in size vastly disproportionate to any increase in body size (as occurs in human males at puberty). This pattern has been carried to an extreme in the hammerhead bat, *Hypsignathus monstrosus*, in which the male's larynx is about the size of its head and occupies most of its body cavity (the female's larynx is one-third this size; Kingdon [1974]). This runaway evolution is probably accounted for by the fact that the species shows a lek-type mating system where females choose males by their call, and 6% of males perform 79% of the matings [Bradbury, 1977].

Thus, vocal fold length, the main determinant of the lowest attainable pitch, does not appear to be tightly coupled to body size. Moreover, voice pitch is an easy cue to "fake" via an increase in larynx size disproportionate to any increase in body size. The result is that, in most of the species which have been examined, there is no relationship between body size and voice pitch. Although some anurans (toads and frogs) [Davies & Halliday, 1978; Ryan, 1988] show the expected negative correlation between body size and pitch, several other anuran species do not [Wilczynski et al., 1993], and a relatively low percentage of variance in call pitch is generally explained by body size [Gerhardt, 1982]. In general, male frogs are smaller than females, but some have larger larynges and lower voices [McClelland & Wilczynski, 1989]. McComb [1991] studied roaring in red deer and showed that a) there is no association between size and pitch in males, and b) females do not prefer low-pitched roars. In primates, Gouzoules and Gouzoules [1990] showed that a discriminant function could perform better than chance at assigning a screaming macaque to its weight class based on a number of spectral parameters, but they reported no correlation between body weight and F0. Finally, as mentioned above, studies of humans have consistently failed to find a correlation between body size and F0 in adults [Lass & Brown, 1978; Künzel, 1989; Sawashima et al., 1983]. Thus the majority of species that have been examined show no correlation between body size and voice pitch.

Despite this lack of correlation, human listeners still use pitch as a cue to adult human body size [Fitch, 1994], and the widespread occurrence of low-pitched growls during aggressive interactions [Morton, 1977, 1982; Hauser, 1993a] suggests that other species might do the same. Why should such an apparently inaccurate mechanism remain in place? Perhaps pitch provides a valid cue when a broader range of body sizes is considered. For example, the vocalizations of the young in many species are higher pitched than those of adults [e.g., Peterson & Barney, 1952], suggesting that pitch might provide a size cue for immature animals. Another possibility is that F0 may still provide a reliable indicator of body size *between* species. Hauser [1993a] examined the relationship between body weight and pitch in nonhuman primates, analyzing several hundred vocalizations from 43 species (23 genera, including the Prosimii, Platyrrhini, Catarrhini, and Hominoidea). In general, body weight accounts for a significant proportion of the variation in vocal pitch among species. Thus, the largest species within the Primate order (e.g., gorillas, chimpanzees) produce relatively lower pitched calls than the smallest species (e.g., bushbabies, marmosets). There are, however, important exceptions to this general pattern, indicating that for some genera, body weight is clearly not the most important factor underlying the range of frequencies used. For example, within the genus *Macaca*, most variation in vocal pitch may be attributable to habitat differences rather than average body size. In any case, the pitch and size differences between species, or between immature conspecifics, may justify the continued existence of a perceptual mechanism that equates pitch and body size, thereby providing for continuing manipulation of this mechanism during interactions between adults within a given species.

Morton [1977, 1982] suggested that the putative (or primitive) relationship between pitch and body size leads to a set of "motivation-structure rules" whereby aggressive vocalizations have lower pitches than submissive calls. This pattern appears to hold in at least some primates (most species within the Catarrhini produce low frequency vocalizations in the context of aggression, but this is not consistently the case for the Prosimii, Platyrrhini, or Hominoidea [Hauser, 1993a]). However, in species where an accurate estimate of male quality plays an important role in determining female fitness, such as the polygynous red deer mentioned earlier, selection may be strong enough to eliminate the body size-pitch connection, resulting in red deer females who ignore pitch in their assessment of males [McCoomb, 1991].

An examination of the relationship between body size and vocal pitch within a particular nonhuman primate species would be relatively easy with captive animals (F0 measurements of low-pitched vocalizations such as growls combined with measures of body weight and length). Determining whether animals perceive an association between low pitch and large body size is difficult to assess directly (requiring operant techniques), but could be investigated via playback experiments in species where large males are preferred by females (or avoided by other males).

In summary, although early vertebrates may have exhibited a close coupling between body size and voice pitch, and an inverse correlation between F0 and body size is found in some anurans, pitch does not provide a reliable cue to the body size of adults in most vertebrates which have been studied. In fact, the anatomical location and structure of the larynx makes this cue eminently "fakeable," in that the length of the vocal folds can easily be increased without a concomitant increase in body size, as occurs in the males of many species and is carried to a startling extreme by the huge larynx of the hammerhead bat. We now consider some other potential acoustic cues to body size that may not be as easy to "fake" as fundamental frequency.

Acoustic Cues to Body Size: Supralaryngeal and Other Cues

Unlike the larynx, the pharyngeal and oral portions of the supralaryngeal vocal tract are firmly ensconced in the bones of the skull, jaw, and spinal column. Thus, the size of the vocal tract should be tightly constrained by that of the entire head. Because the length of the vocal tract determines its resonant frequencies (as described earlier), the lowest formants producible by an individual could provide a relatively robust and unfakable cue to its body size. Formant frequencies are perceived with high accuracy by humans (they are one of the most important cues in speech perception) and by macaques [Sommers et al., 1992], and presumably by some "talking" birds, who base their vocalizations on the formant patterns of speech [Klatt & Stefanski, 1974].

Although no studies that we are aware of have specifically examined the relationship between vocal tract length and body size, Dmitriev and Kiselev [1979] described a negative correlation between vocal tract length and formants in humans, as predicted by acoustic theory, and Sachs et al. [1972] found a negative correlation between vowel formant frequencies and height. These data are consistent with a positive correlation between vocal tract length and height in humans. Furthermore, Fitch [1994] showed that human listeners judge long vocal tracts to signify large bodies. He asked subjects to listen to vowels and then rate the body size of speakers, and found a clear positive correlation between vocal tract length and body size rating. The vowels were generated by a computer model of the vocal tract, so that vocal tract length could be precisely controlled and all potentially contaminating variables eliminated.

A number of otherwise inexplicable phenomena in primate communication make sense if animals use formant frequencies to gauge body size. Although the maximum length of the vocal tract may be constrained by skeletal features, vocal tract length is adjustable via 1) retraction or protrusion of the lips, and 2) raising or lowering of the larynx. If perceivers make use of vocal tract length when gauging body size, manipulative signalers might be expected to maximize their vocal tract length when attempting to project a large body (e.g., during aggression) and reduce vocal tract length when trying to seem small (e.g., during appeasement). The primate bared teeth display ("grimace" or "fear grin") seems inexplicable as a visual signal: why should a display which exposes the teeth (including, in most primates, a rather fearsome set of canines) be used almost universally as a signal of appeasement, submission, and friendliness [Andrew, 1963; Hauser, 1993b]? If animals, like humans, judge voices produced with short vocal tracts as coming from smaller speakers, the acoustic effect of lip retraction will be to decrease the apparent size of the speaker, an effect consistent with a message of appeasement [Ohala, 1980, 1983, 1984]. The fact that bare-teeth displays are so common among mammals, and appear to represent one of the most ancient communicative signals [van Hoof, 1972], and that all canids (i.e., foxes, wolves, coyotes, and dogs) use a "grin" involving retraction of the mouth corners to indicate submission [Fox & Cohen, 1977], suggests that the use of vocal tract length in body size judgments may be widespread in mammals.

A second prediction is that dominant animals in aggressive situations should attempt to lengthen the vocal tract via larynx lowering or lip protrusion [Fitch, 1994]. Although larynx height is difficult to measure, lip protrusion is readily visible in many observational contexts and is thus well suited to an initial empirical approach to this question. Fitch [1994] found that saki monkeys (*Pithecia pithecia*) protruded their lips during aggressive vocalizations. That this articulatory maneuver was designed to increase apparent body size was suggested by a positive correlation between lip protrusion and piloerection of the sakis' thick coat of fur, which increases their visually apparent body size. Epple [1967] described a "protruded lips face" in the tamarin *Saguinus geoffroyi* which functions as an aggressive threat. Van Hooff [1967, p. 18], in his general review of primate facial displays, described the "tense-mouth face" in which the "mouth corners are brought forward. As a result the mouth often looks like a narrow slit." This display is usually performed by the dominant animal immediately preceding an attack, and is associated with a low-pitched bark in at least some species (chimpanzees and baboons). Kaufman [1967] reports a macaque threat display in which a high-ranking male "would thrust his face close to the other's face and protrude his lips. The threatened monkey responded by sitting very still and giving a fear grin." All canids share a characteristic threat face, described by Fox and Cohen [1977, p. 734] as an "aggressive pucker or mouth-closed, lips forward expression." Finally, bears protrude their lips during aggressive interactions [Pruitt & Burghardt, 1977]. In all these cases, therefore, modification of vocal tract length is consistent with a perceptual mechanism which equates long vocal tracts (and hence low formants) with large bodies.

It is important to note that we do not predict that all calls produced with lip protrusion are aggressive. For example, we might expect courtship vocalizations to be produced with vocal tract elongation if large body size is important in mate choice. It also seems likely that an attacker about to bite will bare its teeth, reducing vocal tract length. In canids preparing to attack, the growl (a threat vocalization made with fully protruded lips) turns to a snarl which exposes the canines but *not* the rear teeth, maintaining the maximum tract length compatible with biting (the submissive "grin" exposes all teeth and thus shortens the tract much more). Thus, the use of vocal tract length as a cue to body size may result in different behavior in different species, although we expect lip protrusion to be common during threat vocalizations.

Furthermore, vocal tract length will obviously have an acoustic effect only during vocalization (although during close contact aggression or courtship, quiet vocalizations such as grunts, whispers, or heavy breathing may be adequate). Because impulsive sources (coughs or clicks), noisy sources (hissing, wheezing), and low-frequency phonation (growls) have energy spread densely over a wide frequency region, they provide a sound source which delineates the vocal tract transfer function in sharp relief. In contrast, high-pitched phonation has energy concentrated in a few widely spaced frequency bands, and thus provides much less information about vocal tract length or configuration. We thus expect calls for which vocal tract length is an important variable to be produced with low-pitched, noisy, or impulsive sources.

The long nose of the proboscis monkey *Nasalis larvatus* may represent another example of vocal tract elongation, in this case via lengthening of the nasal vocal tract. Extremely long noses are seen only in the males, and during threat vocalizations ("honking"), the long drooping nose stiffens slightly with each "honk" [Napier & Napier, 1985]. Because the length of the nasal tract (unlike the oral tract) is unconstrained by skull size, and can thus be elongated simply by growing the cartilaginous portion of the nose, the long nose of this species may represent an adaptation for "faking" large body size in this forest-dwelling species.

Supralaryngeal cues to body size may be important in nonmammalian communication as well. "Dominant frequency," which plays a crucial role in the communication system of hylid frogs, may represent a vocal tract resonance. The dominant frequency is the primary determinant of female mate choice, and is related not to laryngeal size or vocal fold volume, but to the size of the head and body of the vocalizing frog [Wilczynski et al., 1993]. In birds, vocal tract length is not anatomically constrained [Nowicki & Marler, 1988; Nowicki, et al., 1992], and vocal tract elongation is widespread [56 species of 6 different orders: Neimeier, 1979; Johnsgard, 1983; Clench, 1978]. This may be yet another example of acoustic "fakery" to project a large body size.

In summary, a significant body of data suggests that the acoustic effects of vocal tract length may play an important role as a cue to body size throughout the primate order, if not for all vertebrates. Because of the anatomy of the supralaryngeal vocal tract in mammals, this cue may be less susceptible to "cheating" than pitch. Unfortunately, even less data are available to evaluate this idea than for the linkage between F0 and body size described in the previous section. Perhaps because there is no general appreciation amongst primatologists of the acoustic effect of lip protrusion, it is rare to find anything more than an informal mention of lip protrusion in the primate literature. We hope that this paper encourages primatologists to incorporate measures of vocal tract resonances into future studies of call acoustics. Consideration of vocal tract length may also aid interpretation of situations in which the meaning of the signal is unclear. Should a mother comforting a disturbed infant retract the lips, in order to seem small and unthreatening, or protrude them so as to seem large, powerful, and reassuring?

Other vocal adaptations may be related to body size estimation as well. Lung volume, which is presumably tightly constrained by the volume of the thoracic cavity, may provide an acoustic cue to body size through its influence on any one of several acoustic variables. The air contained within the bronchi and trachea creates subglottal resonant frequencies: the larger the air volume, the lower the resonance. These resonances could affect call acoustics during vocalizations in which the glottis was open, or leaky [Cranen & Boves, 1987], such as vigorous exhalations or coughs. Another correlation may exist for some calls between lung volume and call intensity or duration.

Laryngeal air sacs may represent means of "faking" large body size by lowering F0 or formants, or increasing call duration. Laryngeal air sacs are found in many mammals, but attain their greatest morphological diversity and development in primates [Negus, 1949]. It is thus surprising that virtually no research has investigated the role of primate air sacs in respiration or vocalization. While Negus [1949] argued that most air sacs provide a reserve of air during strenuous activity, many authors [e.g., Hill & Booth, 1957; Kelemen, 1960; Napier & Napier, 1985; Schön Ybarra, in press] find his argument unconvincing and suggest that air sacs play an important acoustic role, an opinion we share. Haimoff [1983] suggested that male siamangs use the air sac to suppress the fundamental during duetting. Gautier [1971; cited in Schön Ybarra, in press] reported a reduction in sound intensity after puncturing the air sac in *Cercopithecus neglectus*. Thus air sacs may increase the efficiency of sound transmission via impedence matching [Kelemen, 1960; Hartley & Suthers, 1988]. Finally, inflated air sacs could act as Helmholtz resonators [Morse, 1981; Fant, 1960] if the passage between the sac and larynx remained open. If this is the case, inflatable sacs may act as tunable resonators, with the volume of the sac controlling the resonant frequency, and fully inflated sacs would result in an extremely low-frequency resonance consistent with large projected body size.

If lung volume (a normal determinant of call duration) serves as a cue to body size in some species, augmentation of the respiratory volume available in the lungs with "reserves" from air sacs might serve to increase the projected body size. For example, the gular pouches of siamangs could perhaps serve as air reservoirs, and thus allow individuals to increase call duration. Of course, inflated air sacs also alter the visual appearance of the animal, and may thus function as visual cues as well [Hill & Booth, 1957]. Additionally, calling rate or call duration may indicate the vocalizer's immediate energy levels, and thus provide more useful cues to its strength and fighting ability than static body size cues [Clutton-Brock & Albon, 1979]. Unfortunately, there are no data currently available to distinguish among these interesting hypotheses.

We have devoted much of this discussion to a consideration of acoustic cues to body size. This is both because we find the topic intrinsically interesting and important, and because all of its elements are open to empirical test (i.e., the strength of the correlations between acoustic variables and size, the perceptual abilities necessary to sense the variables, and productive abilities to vary them). An approach that provides concrete, testable hypotheses about phenomena as diverse as the large nose of the proboscis monkey, lip protrusion in sakis, the primate "fear grin" and the human smile provides a valuable framework for increasing our understanding of vocal production and primate anatomy. More generally, though, our examination of acoustic cues to body size illustrates the ways in which consideration of the acoustics and physiology of vocal production could yield rich insights into the evolution of communication. The rapid increase in availability of desktop computers with sound processing capabilities, and the decrease in their prices, puts the requisite analysis techniques within the reach of virtually any interested researcher. We believe that a basic knowledge of the acoustics of vocal production is highly relevant to many aspects of primate biology, and we hope that this review encourages primatologists to integrate a consideration of the mechanisms of vocal production into their future studies of anatomy and communication.

CONCLUSIONS

1. Although much remains to be discovered, the basic acoustic principles underlying the production of primate calls are well understood.
2. The lungs produce the source of power for vocalizations, a pressurized airstream. This is modulated by the larynx to produce a source of sound. In phonation, this source is a periodic series of air puffs, which repeat at a rate known as the fundamental frequency (F0). Other sound sources are impulsive (e.g., coughs and clicks) or noisy (hisses, whispers).
3. The sound source is channeled through the supralaryngeal vocal tract, a resonant tube of air including the pharyngeal, oral, and nasal cavities. The length and some aspects of the shape of this tube determine the frequencies of resonances (called formants in human speech). The vocal tract might directly influence laryngeal behavior (resulting in a feedback system, like many wind instruments), or be relatively independent of the larynx, acting simply as a filter that removes certain frequencies from the laryngeal source (as in human speech).
4. Krebs and Dawkins [1984] suggested that selection may lead to communicative signals designed to "manipulate" perceivers. We discuss primate vocal production from this perspective, focusing on the various possible acoustic cues to body size. A wide variety of anatomical modifications and behavioral patterns can be seen as adaptations for modifying a vocalizer's projected size, within the limits of their anatomy.
5. The anatomical configuration of the larynx and vocal tract constrain primate communication systems in different ways. Converging evidence suggests that the size of the larynx is relatively independent of total body size, making F0 a poor cue to body size for adults within most species.
6. The resonant frequencies of the vocal tract may provide a more "honest" cue to body size, because they are determined by skull size and mandible length, and thus may be more tightly coupled to body size. Although a number of observations are consistent with the idea that animals use resonant frequencies as cues to body size, no research has directly evaluated this hypothesis.

ACKNOWLEDGMENTS

This work was supported by an NSF pre-doctoral fellowship and a research grant from Apple Computer, Inc., to W.T.F.; and an NSF Young Investigator Award to M.D.H. We thank Philip Lieberman, Eric Nicolas, and four anonymous reviewers for comments on an earlier version of this manuscript.

REFERENCES

Andrew, R.J. Evolution of facial expressions. SCIENCE 142:1034–1041, 1963.

Atkinson, J.R. ASPECTS OF INTONATION IN SPEECH: IMPLICATIONS FROM AN EXPERIMENTAL STUDY OF FUNDAMENTAL FREQUENCY. Ph.D. thesis, University of Connecticut, 1973.

Baer, T.; Gore, J.C.; Gracco, L.C.; Nye, R.W. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 90(2):799–828, 1991.

Bauer, H.R. Frequency code: orofacial correlates of fundamental frequency. PHONETICA 44:173–191, 1987.

Benade, A.H. FUNDAMENTALS OF MUSICAL ACOUSTICS. New York, Dover Publications, Inc., 1990.

Biben, M.; Symmes, D.; Bernhards, D. Contour variable in vocal communication between squirrel monkey mothers and infants. DEVELOPMENTAL PSYCHOBIOLOGY 22(6):617–631, 1989.

Bickley, C.; Stevens, K. Effect of a vocal tract constriction on the glottal source: Experimental and modeling studies. JOURNAL OF PHONETICS 14:373–382, 1986.

Borden, G.J.; Harris, K.S. SPEECH SCIENCE PRIMER: PHYSIOLOGY, ACOUSTICS, AND PERCEPTION OF SPEECH. Baltimore, MD, Williams & Wilkins, 1984.

Bradbury, J.W. Lek behavior in the hammer-headed bat. ZEITSCHRIFT FÜR TIERPSYCHOLOGY 45:225–255, 1977.

Brown, J.H.; Marquet, P.A.; Taper, M.L. Evolution of body size: consequences of an energetic definition of fitness. AMERICAN NATURALIST 142:573–584, 1993.

Brown, J.H.; Maurer, B.A. Body size, ecological dominance and Cope's rule. NATURE 324:248–250, 1986.

Cheney, D.L.; Seyfarth, R.M. HOW MONKEYS SEE THE WORLD: INSIDE THE MIND OF ANOTHER SPECIES. Chicago, Chicago University Press, 1990.

Clench, M.H. Tracheal elongation in birds-of-paradise. CONDOR 80:423–430, 1978.

Cleveland, J.; Snowdon, C.T. The complex vocal repertoire of the adult cotton-top tamarin, *Saguinus oedipus oedipus*. ZEITSCHRIFT FÜR TIERPSYCHOLOGIE 58:231–270, 1981.

Clutton-Brock, T.H.; Albon, S.D. The roaring of red deer and the evolution of honest advertising. BEHAVIOUR 69:145–170, 1979.

Collier, R. F0 declination: The control of its setting, resetting, and slope. Pp. 403–421 in LARYNGEAL FUNCTION IN PHONATION AND RESPIRATION. T. Baer; C. Sasaki; K. Harris, eds. Boston, College-Hill Publishers, 1987.

Cooper, W.; Sorensen, J. FUNDAMENTAL FREQUENCY IN SENTENCE PRODUCTION. Berlin, Springer-Verlag, 1981.

Cranen, B.; Boves, L. On subglottal formant analysis. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 81(3):734–746, 1987.

Crelin, E. THE HUMAN VOCAL TRACT. New York, Vantage Press, 1987.

Dang, J.; Honda, K. Morphological and acoustical analysis of the nasal and the paranasal cavities. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 96(4):2088–2100, 1994.

Darwin, C. THE DESCENT OF MAN AND SELECTION IN RELATION TO SEX. London, John Murray, 1871.

Darwin, C. THE EXPRESSION OF THE EMOTIONS IN MAN AND ANIMALS. London, John Murray, 1872.

Davies, N.B.; Halliday, T.R. Deep croaks and fighting assessment in toads *Bufo bufo*. NATURE 274:683–685, 1978.

Dmitriev, L.; Kiselev, A. Relationship between the formant structure of different types of singing voices and the dimension of supraglottal cavities. FOLIA PHONIATRICA 31:238–241, 1979.

Epple, G. Vergleichende Untersuchungen über Sexual- und Sozialverhalten der Krallenaffen (Hapalidae). FOLIA PRIMATOGICA 7:37–65, 1967.

Fant, G. ACOUSTIC THEORY OF SPEECH PRODUCTION. The Hague, Mouton & Co., 1960.

Fitch, W.T. VOCAL TRACT LENGTH AND THE EVOLUTION OF LANGUAGE. Ph.D. thesis, Brown University, 1994.

Flanagan, J.L. Source-system interaction in the vocal tract. ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 155:9–17, 1968.

Fox, M.W.; Cohen, J.A. Canid communication. Pp. 728–748 in HOW ANIMALS COMMUNICATE. T.A. Sebeok, ed. Bloomington, Indiana University Press, 1977.

Gautier, J.P. Etude morphologique et fonctionnelle des annexes extra-laryngées des cercopithecinae; liaison avec les cris d'espacement. BIOLOGY GABONICA 7(2):230–267, 1971.

Gelfer, C.A. SIMULTANEOUS PHYSIOLOGICAL AND ACOUSTIC STUDY OF FUNDAMENTAL FREQUENCY DECLINATION. Ph.D. thesis, New York, CUNY, 1987.

Gelfer, C.E.; Harris, K.S.; Baer, T. Controlled variables in sentence intonation. Pp. 422–435 in LARYNGEAL FUNCTION IN PHONATION AND RESPIRATION. T. Baer; C. Sasaki; K. Harris, eds. Boston, College-Hill Publishers, 1987.

Gerhardt, H.C. Sound pattern recognition in some North American tree-frogs (Anura: Hylidae): Implications for mate choice. AMERICAN ZOOLOGIST 22:581–595, 1982.

Gouzoules, H.; Gouzoules, S. Body size effects on the acoustic structure of pigtail macaque (*Macaca nemestrina*) screams. ETHOLOGY 85:324–334, 1990.

Gouzoules, S.; Gouzoules, H.; Marler, P. Rhesus monkey (*Macaca mulatta*) screams: Representational signaling in the recruitment of agonistic aid. ANIMAL BEHAVIOUR 32:182–193, 1984.

Green, S.; Marler, P. The analysis of animal communication. Pp. 73–158 in SOCIAL BEHAVIOR AND COMMUNICATION, HANDBOOK OF BEHAVIORAL NEUROBIOLOGY, VOL. 3. P. Marler; J. Vandenbergh eds. New York, Plenum Press, 1979.

Haimoff, E.H. Occurrence of anti-resonance in the song of the siamang (*Hylobates syndactylus*.) AMERICAN JOURNAL OF PRIMATOLOGY 5:249–256, 1983.

Harcourt, A.H.; Stewart, K.; Hauser, M.D. The social use of vocalizations by gorillas.

I. Social behaviour and vocal repertoire. BEHAVIOUR 124:89–122, 1993.

Hartley, D.J.; Suthers, R.A. The acoustics of the vocal tract in the horseshoe bat, *Rhinolophus hildebrandti*. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 84:1201–1213, 1988.

Hartley, R.S.; Suthers, R.A. Airflow and pressure during canary song: Direct evidence for mini-breaths. JOURNAL OF COMPARATIVE PHYSIOLOGY [A] 165: 15–26, 1989.

Hast, M. Comparative anatomy of the larynx: Evolution and function. Pp. in VOCAL FOLD PHYSIOLOGY: BIOMECHANICS, ACOUSTICS AND PHONATORY CONTROL. I.R. Titze; R.C. Scherer, eds. Denver, Denver Center for the Performing Arts, 1985.

Hauser, M.D. Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: A learned mode of production? JOURNAL OF THE ACOUSTIC SOCIETY OF AMERICA 91:2175–2179, 1992.

Hauser, M.D. The evolution of nonhuman primate vocalizations: Effects of phylogeny, body weight and social context. AMERICAN NATURALIST 142(3):528–542, 1993a.

Hauser, M.D. Right hemisphere dominance for production of facial expression in monkeys. SCIENCE 261:475–477, 1993b.

Hauser, M.D. THE EVOLUTION OF COMMUNICATION. Cambridge, MA, MIT Press, in press.

Hauser, M.D.; Evans, C.S.; Marler, P. The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations. ANIMAL BEHAVIOUR 45:423–433, 1993.

Hauser, M.D.; Fowler, C. Declination in fundamental frequency is not unique to human speech: Evidence from nonhuman primates. JOURNAL OF THE ACOUSTIC SOCIETY OF AMERICA 91:363–369, 1991.

Hauser, M.D.; Marler, P. Food-associated calls in rhesus macaques (*Macaca mulatta*). I. Socioecological factors influencing call production. BEHAVIORAL ECOLOGY 4:194–205, 1993.

Hauser, M.D.; Schön Ybarra, M. The role of lip configuration in monkey vocalizations: Experiments using xylocaine as a nerve block. BRAIN AND LANGUAGE 46:232–244, 1994.

Hill, W.C.O.; Booth, A.H. Voice and larynx in African and Asiatic Colobidae. JOURNAL OF THE BOMBAY NATURAL HISTORY SOCIETY 54:309–321, 1957.

House, A.S.; Stevens, K.N. Analog studies of the nasalization of vowels. JOURNAL OF SPEECH AND HEARING DISORDERS 21:218–232, 1956.

Ishizaka, K.; Flanagan, J.L. Synthesis of voiced sounds from a two-mass model of the vocal cords. THE BELL SYSTEM TECHNICAL JOURNAL 51:1233–1268, 1972.

Johnsgard, P. CRANES OF THE WORLD. Bloomington, Indiana University Press, 1983.

Jürgens, U. Vocal communication in primates. Pp. 51–76 in NEUROBIOLOGY OF COMPARATIVE COGNITION. R.P. Kesner; D.S. Olton, eds. Hillsdale, NJ, Lawrence Erlbaum Associates, 1990.

Kaufmann, J.H. Social relations of adult males in a free-ranging band of rhesus monkeys. Pp. in SOCIAL COMMUNICATION AMONG PRIMATES. S.A. Altmann, ed. Chicago, University of Chicago Press, 1967.

Kelemen, G. The vocal organ of the howling monkey (*Alouatta palliata*). JOURNAL OF MORPHOLOGY, 107(2):123–140, 1960.

Kingdon, J. EAST AFRICAN MAMMALS: VOL II PART A: INSECTIVORES AND BATS. New York, Academic Press, 1974.

Klatt, D.H.; Klatt, L.C. Analysis, synthesis, and perception of voice quality variations among female and male talkers. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 87:820–857, 1990.

Klatt, D.H.; Stefanski, R.A. How does a mynah bird imitate human speech? JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 55(4):822–832, 1974.

Krebs, J.R.; Dawkins, R. Animal signals: Mind-reading and manipulation. Pp. 380–402 in BEHAVIOURAL ECOLOGY. J.R. Krebs; N.B. Davies, eds. Sunderland, MA, Sinauer Associates Inc., 1984.

Künzel, H.J. How well does average fundamental frequency correlate with speaker height and weight? PHONETICA 46:117–125, 1989.

Lass, N.J.; Brown, W.S. Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 63(4):1218–1220, 1978.

Liberman, A.M.; Mattingly, I.G. The motor theory of speech perception revised. COGNITION 21:1–36, 1985.

Lieberman, P. INTONATION, PERCEPTION AND LANGUAGE. Cambridge, MA, MIT Press, 1967.

Lieberman, P. THE BIOLOGY AND EVOLUTION OF LANGUAGE. Cambridge, MA, Harvard University Press, 1984.

Lieberman, P. The physiology of cry and speech in relation to linguistic behavior. Pp. 29–57 in INFANT CRYING. B.M. Lester; C.F.Z. Boukydis, eds. New York, Plenum Press, 1985.

Lieberman, P. UNIQUELY HUMAN. Cambridge, MA, Harvard University Press, 1991.

Lieberman, P.; Blumstein, S.E. SPEECH PHYSIOLOGY, SPEECH PERCEPTION, AND ACOUSTIC PHONETICS. Cambridge, UK, Cambridge University Press, 1988.

Lieberman, P.; Klatt, D.H.; Wilson, W.H. Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates. SCIENCE 164:1185–1187, 1969.

Lindblom, B.; Lubker, J.; Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. JOURNAL OF PHONETICS 7:147–161, 1979.

Maddieson, I. PATTERNS OF SOUNDS. Cambridge, Cambridge University Press, 1984.

Maeda, T.; Masataka, N. Locale-specific behavior of the tamarin (*Saguinus l. labiatus*). ETHOLOGY 75:25–30, 1989.

Marler, P. Characteristics of some animal calls. NATURE 176:6–7, 1955.

Marler, P. Animal communication signals. SCIENCE 157:769–774, 1967.

Marler, P. Primate vocalizations: Affective or symbolic? Pp. 14–32 in PROGRESS IN APE RESEARCH. G. Bourne, ed. New York, Academic Press, 1978.

Marler, P.; Evans, C.S.; Hauser, M.D. Animal signals? Reference, motivation or both? Pp. 66–86 in NONVERBAL VOCAL COMMUNICATION: COMPARATIVE AND DEVELOPMENTAL APPROACHES. H. Papousek; U. Jürgens; M. Papousek, eds. Cambridge, UK, Cambridge University Press, 1992.

May, B.J.; Moody, D.B.; Stebbins, W.C. The significant features of Japanese monkey coo sounds: a psychophysical study. ANIMAL BEHAVIOUR 36:1432–1444, 1988.

McClelland, B.; Wilcyzynski, W. Release call characteristics of male and female *Rana pipiens*. COPEIA 1989:1045–1049, 1989.

McCoomb, K.E. Female choice for high roaring rates in red deer, *Cervus elaphus*. ANIMAL BEHAVIOUR 41:79–88, 1991.

Mitani, J.C.; Hasegawa, T.; Gros-Louis, J.; Marler, P.; Byrne, R. Dialects in wild chimpanzees? AMERICAN JOURNAL OF PRIMATOLOGY 27:233–244, 1992.

Moody, D.B.; Stebbins, W.C.; May, B.J. Auditory perception of communication signals by Japanese monkeys. Pp. 311–344 in COMPARATIVE PERCEPTION: COMPLEX PERCEPTION. W.C. Stebbins; M.A. Berkley eds. New York, John Wiley and Sons, Inc., 1990.

Morse, P.M. VIBRATION AND SOUND. Woodbury, NY, Acoustical Society of America, 1936.

Morton, E.S. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. AMERICAN NATURALIST 111:855–869, 1977.

Morton, E.S. Grading, discreteness, redundancy, and motivational-structural rules. Pp. 183–212 in ACOUSTIC COMMUNICATION IN BIRDS, VOL. 1. D. Kroodsma; E.H. Miller, eds. New York, Academic Press, 1982.

Müller, J. THE PHYSIOLOGY OF THE SENSES, VOICE AND MUSCULAR MOTION WITH MENTAL FACULTIES (W. Baly, Translator). London, Walton and Maberly, 1848.

Napier, J.R.; Napier, P.H. THE NATURAL HISTORY OF THE PRIMATES. Cambridge, MA, MIT Press, 1985.

Nearey, T. PHONETIC FEATURES FOR VOWELS. Bloomington, Indiana University Linguistics Club, 1978.

Negus, V.E. THE MECHANISM OF THE LARYNX. St. Louis, C.V. Mosby Company, 1929.

Negus, V.E. THE COMPARATIVE ANATOMY AND PHYSIOLOGY OF THE LARYNX. New York, Hafner Publishing Company, 1949.

Newman, J.D. Squirrel monkey communication. Pp. 99–126 in HANDBOOK OF SQUIRREL MONKEY RESEARCH. L.A. Rosenblum; C.L. Coe, eds. New York, Plenum, 1985.

Niemeier, M.M. STRUCTURAL AND FUNCTIONAL ASPECTS OF VOCAL ONTOGENY IN *GRUS CANADENSIS* (GRUIDAE: AVES). Ph.D. thesis, University of Nebraska, 1979.

Nowicki, S.; Marler, P. How do birds sing? MUSIC PERCEPTION 5:391–426, 1988.

Nowicki, S.; Westneat, M.; Hoese, W. Birdsong: motor function and the evolution of communication. SEMINARS IN THE NEUROSCIENCES 4:385–390, 1992.

Ohala, J.J. Abstract: The acoustic origin of the smile. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA Suppl.1 68: S33, 1980.

Ohala, J.J. Cross-language use of pitch: An ethological view. PHONETICA 40:1–18, 1983.

Ohala, J.J. An ethological perspective on common cross-language utilization of F0 of voice. PHONETICA 41:1–16, 1984.

Petersen, M.R.; Beecher, M.D.; Zoloth, S.R.; Moody, D.B.; Stebbins, W.C. Neural lateralization of species-specific vocalizations by Japanese macaques. SCIENCE 202: 324–326, 1978.

Peterson, G.E.; Barney, H.L. Control methods used in a study of vowels. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 24(2):175–184, 1952.

Pruitt, C.H.; Burghardt, G.M. Communication in terrestrial carnivores: Mustelidae, Procyonidae and Ursidae. Pp. 767–793 in HOW ANIMALS COMMUNICATE. T.A. Sebeok, ed. Bloomington, Indiana University Press, 1977.

Remez, R.E.; Rubin, P.E.; Pisoni, D.B.; Carrell, T.D. Speech perception without traditional speech cues. SCIENCE 212:947–950, 1981.

Ryalls, J.H.; Lieberman, P. Fundamental frequency and vowel perception. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 72(5):1631–1634, 1982.

Ryan, M.J. Constraints and patterns in the evolution of anuran acoustic communication. Pp. in THE EVOLUTION OF THE AMPHIBIAN AUDITORY SYSTEM. B. Fritzch; M.J. Ryan; W. Wilczynski; T.E. Hetherington; W. Walkowiak, eds. New York, Wiley, 1988.

Sachs, J.; Lieberman, P.; Erickson, D. Anatomical and cultural determinants of male and female speech. Pp. in LANGUAGE ATTITUDES: CURRENT TRENDS AND PROSPECTS. R. Shuy; R. Fasold, eds. Washington, Georgetown University Press, 1972.

Sapir, S.; Campbell, C.; Larson, C. Effect of geniohyoid, cricothyroid and sternothyroid muscle stimulation on voice fundamental frequency of electrically elicited phonation in rhesus macaque. LARYNGOSCOPE 91: 457–468, 1981.

Sawashima, M.; Hirose, H.; Honda, K.; Yoshioka, H.; Hibi, S.R.; Kawase, N.; Yamada, M. Stereoendoscopic measurement of the laryngeal structure. Pp. in VOCAL FOLD PHYSIOLOGY: CONTEMPORARY RESEARCH AND CLINICAL ISSUES. D.M. Bless; J.H. Abbs, eds. New York, College-Hill, 1983.

Schmidt-Nielsen, K. SCALING: WHY IS ANIMAL SIZE SO IMPORTANT? New York, Cambridge University Press, 1984.

Schön Ybarra, M. A comparative approach to the nonhuman primate vocal tract: Implications for sound production. in FRONTIERS IN PRIMATE VOCAL COMMUNICATION. E. Zimmerman, ed. New York, Plenum Press, in press.

Schutte, H.K.; Miller, D.G. The effect of Fo/F1 coincidence in soprano high notes on pressure at the glottis. JOURNAL OF PHONETICS 14:385–392, 1986.

Sommers, M.S.; Moody, D.B.; Prosen, C.A.; Stebbins, W.C. Formant frequency discrimination by Japanese macaques (*Macaca fuscata*). JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 91(6):3499–3510, 1992.

Stebbins, W.C.; Sommers, M.S. Evolution, perception and the comparative method. Pp. 211–228 in THE EVOLUTIONARY BIOLOGY OF HEARING. D.B. Webster; R.F. Fay; A.N. Popper, eds. New York, Springer-Verlag, 1992.

Struhsaker, T.T. Auditory communication among vervet monkeys (*Cercopithecus aethiops*). Pp. 281–324 in SOCIAL COMMUNICATION AMONG PRIMATES. S.A. Altmann, ed. Chicago, Chicago University Press, 1967.

Sundberg, J. Formant technique in a professional female singer. ACUSTICA 32:89–96, 1975.

Sundberg, J. Formants and fundamental frequency control in singing: An experimental study of coupling between vocal tract and voice source. SPEECH TRANSACTIONS LABORATORY QUARTERLY PROGRESS AND STATUS REPORT 1979:65–78, 1979.

Sundberg, J. THE SCIENCE OF THE SINGING VOICE. DeKalb, Northern Illinois University Press, 1987.

Sundberg, J. THE SCIENCE OF MUSICAL SOUNDS. New York, Academic Press, Inc., 1991.

Suthers, R.A. The production of echolocation signals by bats and birds. Pp. 23–45 in ANIMAL SONAR: PROCESSES AND PERFORMANCE. P.E. Nachtigall; P.W.B. Moore, eds. New York, Plenum Press, 1988.

Suthers, R.A.; Fattu, J.M. Mechanisms of sound production in echolocating bats. AMERICAN ZOOLOGIST 13:1215–1226, 1973.

Suthers, R.A.; Hartley, D.J.; Wenstrup, J.J. The acoustic role of tracheal chambers and nasal cavities in the production of sonar pulses by the horseshoe bat, *Rhinolophus hildebrandti*. JOURNAL OF COMPARATIVE PHYSIOLOGY [A] 162:799–813, 1988.

Sutton, D. Mechanisms underlying vocal control in nonhuman primates. Pp. 45–68 in NEUROBIOLOGY OF SOCIAL COMMUNICATION IN PRIMATES. H. Steklis; M. Raleigh, eds. New York, Academic Press, 1979.

Titze, I.R. On the relation between subglottal pressure and fundamental frequency in phonation. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 85:901–906, 1989.

Titze, I.R.; Mapes, S.; Story, B. Acoustics of the high tenor voice. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 95(2):1133–1142, 1994.

van den Berg, J. Myoelastic-aerodynamic

theory of voice production. JOURNAL OF SPEECH AND HEARING RESEARCH 1:227–44, 1958.

van den Berg, J. Sound production in isolated human larynges. ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 155:18–27, 1968.

van Hoof, J.A.R.A.M. A comparative approach to the phylogeny of laughter and smiling. Pp. 12–53 in NONVERBAL COMMUNICATION. R.A. Hinde, ed. Cambridge, UK, Cambridge University Press, 1972.

van Hooff, J.A.R.A.M. The facial displays of the Catarrhine monkeys and apes. Pp. 7–68 in PRIMATE ETHOLOGY. D. Morris, ed. London, Weidenfeld and Nicolson, 1967.

Waser, P.M. Experimental playbacks show vocal mediation of avoidance in a forest monkey. NATURE 255:56–58, 1977.

Wilczynski, W.; McClelland, B.E.; Rand, A.S. Acoustic, auditory and morphological divergence in three species of neotropical frog. JOURNAL OF COMPARATIVE PHYSIOLOGY 172:425–438, 1993.

Wiley, R.H. Evolution of social organization and life history patterns among grouse. QUARTERLY REVIEW OF BIOLOGY 49: 201–227, 1974.

Williams, R.G.; Eccles, R. A new clinical measure of external laryngeal size which predicts the fundamental frequency of the larynx. ACTA OTOLARYNGOLICA (STOCKHOLM) 110:141–148, 1990.

Zahavi, A. Mate selection: A selection for a handicap. JOURNAL OF THEORETICAL BIOLOGY 53:205–214, 1975.

Zahavi, P. Reliability in communication systems and the evolution of altruism. Pp. 253–259 in EVOLUTIONARY ECOLOGY. B. Stonehouse; C. Perrins, eds. London, Macmillan Press, 1977.

Zimmerman, E. Differentiation of vocalizations in bushbabies (Galginae, Prosimiae, Primates) and their significance for assessing phylogenetic relationships. ZEITSCHRIFT FUR ZOOLOGISCHE SYSTEMATIC UND EVOLUTIONS-FORSCHUNG 28:217–239, 1990.