

# HUMAN VOICE IN EVOLUTIONARY PERSPECTIVE

Michael J. Owren

Department of Psychology, Georgia State University  
Atlanta, Georgia 30302

## Introduction

The human voice is a remarkable, multi-faceted instrument that has been studied and discussed by scholars throughout recorded history. Modern scientific study has revealed much about its fundamental properties, such as the physics and physiology of vocal-fold action, the causes and consequences of vocal impairment, and the rich, varied articulatory maneuvers used among the world's many languages. While inquiry has typically been prompted by issues concerning speech communication or vocal performance, work on vocalization in nonhumans is inspiring new questions and insights about the voice from an evolutionary perspective. A major goal in this approach is to understand how and why the human voice has come to have its current, particular form. The premise is that the basic biological forces shaping vocalization in other species have also been important in humans—creating basic commonalities that arguably transcend the many obvious differences that exist between human and nonhuman communication.

This article is intended as an introduction to some of the issues that arise in understanding the voice in evolutionary terms. The *source-filter model* of vocalization will be central throughout, explaining vocal production as a combination of laryngeal energy and vocal-tract resonance. While originally developed in speech science, it is now widely applied to non-human vocalization as well. *Indexical cuing* is a second underlying theme, referring to acoustic aspects of the voice and vocal signals that are correlated with important vocalizer characteristics such as sex, identity, age, and emotional state. Both source-filter production and indexical cuing are deeply rooted in the phylogeny of human vocalization, which becomes clear in reviewing our species' mammalian and primate pasts. Commonalities are especially clear in sex and identity cuing, with sex differences in vocal anatomy and acoustics in particular having inspired a flurry of recent, exciting studies connecting cues from pitch and resonance to vocalizer fitness and reproductive success.

## Source-filter theory

Understanding the voice in comparative perspective begins by examining the physical characteristics of the vocal tract, important features of which are illustrated for humans and nonhuman primates in Fig. 1. Two critical components can be distinguished. First, the *source energy* of vocalization is derived from laryngeal, vocal-fold vibration driven by air flowing from the lungs (*phonation*), or by creating turbulence in the flow by forcing it through a constriction or onto a surface within the tract. In both cases, this source energy excites

“...a flurry of recent, exciting studies connecting cues from pitch and resonance to vocalizer fitness and reproductive success.”

cavities located above the larynx, which make up the *supralaryngeal vocal tract*. Resonances of these cavities are referred to as *formants*, and shape the spectral characteristics of the source energy in accordance with their input-output relation. The overall effect is often referred to as *vocal-tract filtering*, and has long been fundamental to understanding human speech production (Chiba and

Kajiyama, 1941; Fant, 1960; Stevens, 2000). Over the last two decades, however, this two-component, source-filter approach to vocalization has been applied to an ever-increasing range of nonhuman species as well (Taylor and Reby, 2010).

The process involved in producing a complex, tonal sound is also illustrated in the figure using naturally occurring vocalizations from a human male and a female rhesus monkey (*Macaca mulatta*). Each sound is produced by putting the vocal folds in regular, or quasi-periodic, vibratory motion. As the folds are forced apart and come back together, bursts of air emanate from the *glottis*, which is the opening between the folds. The frequency spectrum of glottal air-flow exhibits most energy at the *fundamental frequency* ( $F_0$ ), or base rate of vibration, with energy at corresponding higher harmonics declining exponentially with increasing frequency. The cavities and tissues of each species' supralaryngeal vocal tract can strongly shape glottal waveform components through resonance and anti-resonance effects, which respectively reinforce or damp energy in corresponding frequency regions. The filtering that results mirrors the sizes and shapes of the vocalizer's supralaryngeal vocal-tract cavities. In an adult human male, a relaxed, “neutral” vocal tract is modeled as a uniform, straight tube closed at the glottal end. It is composed of approximately equal-length pharyngeal and oral cavities, with an overall vocal-tract length of about 17 cm measured from glottis to lips. The characteristic frequency spectra of resulting phonated sounds are marked by 4 to 5 prominent spectral peaks in the 0- to 5-kHz range, each of which reflects a formant. In a rhesus monkey, smaller vocal folds and a much shorter supralaryngeal vocal tract produce higher  $F_0$  values and formant frequencies, respectively.

The pattern formed by these peaks can play a major role in determining the auditory quality of a given vocalization. Corresponding effects are routinely evident in many mammals, taking into account differences in overall vocal-tract length and characteristics of individual supralaryngeal cavities. Due to coincidental resemblance to humans in  $F_0$  and vocal-tract length, for example, the chacma baboon (*Papio cynocephalus ursinus*) “grunt” call bears a remarkable resemblance to an unarticulated, human vowel sound (Owren *et*

al., 1997; Rendall *et al.*, 2005). While vocal anatomy can be specialized in particular species, basic principles of production are importantly similar across all mammals. The most important point is that, at least for larger-bodied animals, vocal quality reflects characteristics of both source energy and subsequent vocal-tract filtering. Critical perceptual attributes like pitch, tonality, and other aspects of timbre can all be understood based the combined effects of these two components.

## Origins

**Reptiles and mammals**—Probing the evolutionary history of source-filter production, one might ask if dinosaurs also vocalized using such a system. Films like *Jurassic Park* (1993) and *The Land That Time Forgot* (2009) show them doing exactly that, inasmuch as their sounds are remarkably mammal-like. Such portrayals are only weakly grounded in scientific evidence, however, which consists of little more than finding that certain duck-billed, *Parasaurolophus* dinosaurs had elongated nasal passages forming hollow crests (reviewed by Weishampel, 1997; Isles, 2009). Having ruled out other possible functions, paleontologists have concluded that these crests must have acted as acoustic resonators for vocalization. Unfortunately, there is no evidence as to what the source energy used to excite those cavities might have been.

This intriguing example from dinosaurs does, however, underscore the broader point that, as a group, reptiles have a purely valve-like larynx that cannot also produce sound. Some modern crocodylians, geckos, and tortoises and turtles

do vocalize, but these species represent the exception rather than the rule for reptiles as a whole. In contrast, a sound-producing larynx is ubiquitous among the more than 4,500 extant mammal species. Given that all the current major mammalian groups had already emerged by about 93 million years ago (Bininda-Emonds *et al.*, 2007), laryngeal vocalization must have arisen even earlier—but nonetheless after divergence from the reptile line. When mammals underwent rapid proliferation after the disappearance of dinosaurs about 65 million years ago, they carried that vocalizing larynx along. In fact, one could argue that vocalization is as fundamental to being a mammal as having three middle-ear bones or being homeothermic.

**Primates as mammals**—In spite of fundamental commonalities, mammals do exhibit significant variability in vocal production as well. Differences can occur in both source and filter components, depending on factors such as overall body-size, hearing range, and niche-specific adaptations (Fitch, 2006; Brudzynski, 2010). Vibration frequencies vary widely across species, ranging from infra- to ultra-sonic. Extra-laryngeal vocal sacs can dramatically amplify or attenuate particular frequency regions, and some species even have a mobile larynx that can dramatically increase effective supralaryngeal tract length during sound production.

There is an additional, cross-species similarity to point out, however, which is that all mammalian larynges can evidently produce a range of phonated sounds, including both harmonically structured and noisy versions. Broadly speaking, this division reflects vocal-fold vibration patterns that are either stable and regular, or unstable and chaotic, respec-

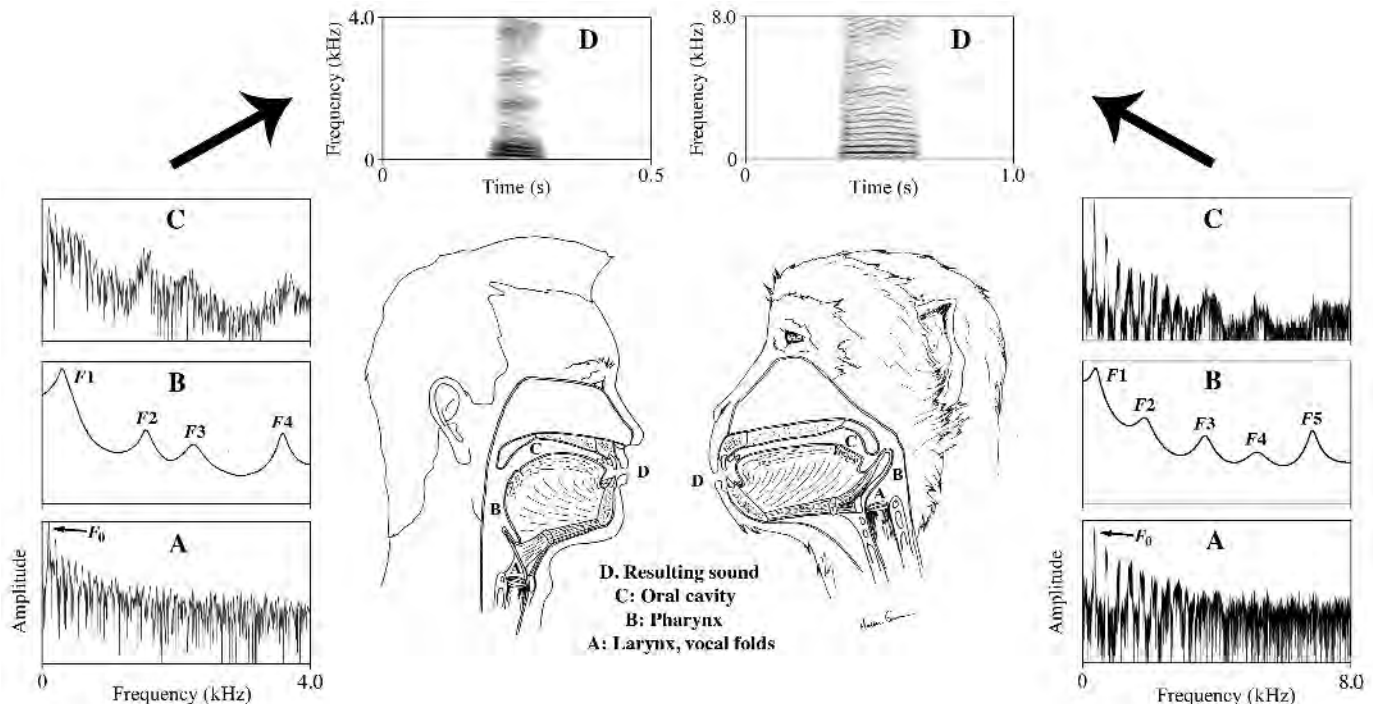


Fig. 1. Schematic views of a human male and a female rhesus monkey vocal tract illustrating the source-filter vocal production process. For both species, panels A and B illustrate source energy frequency spectrum and supralaryngeal transfer function, respectively. Panel C shows the spectrum resulting from combining source and filter, and panel D shows a narrowband spectrogram of the original sound.  $F_0$  refers to the fundamental frequency of the sound, while  $F_1$ – $F_5$  refer to formants. Rhesus monkeys are significantly smaller relative to humans than indicated here, have significantly higher  $F_0$  and formant frequency values. Note that the rhesus vocalization is shown over a wider frequency range. (Drawings by Michael Graham)

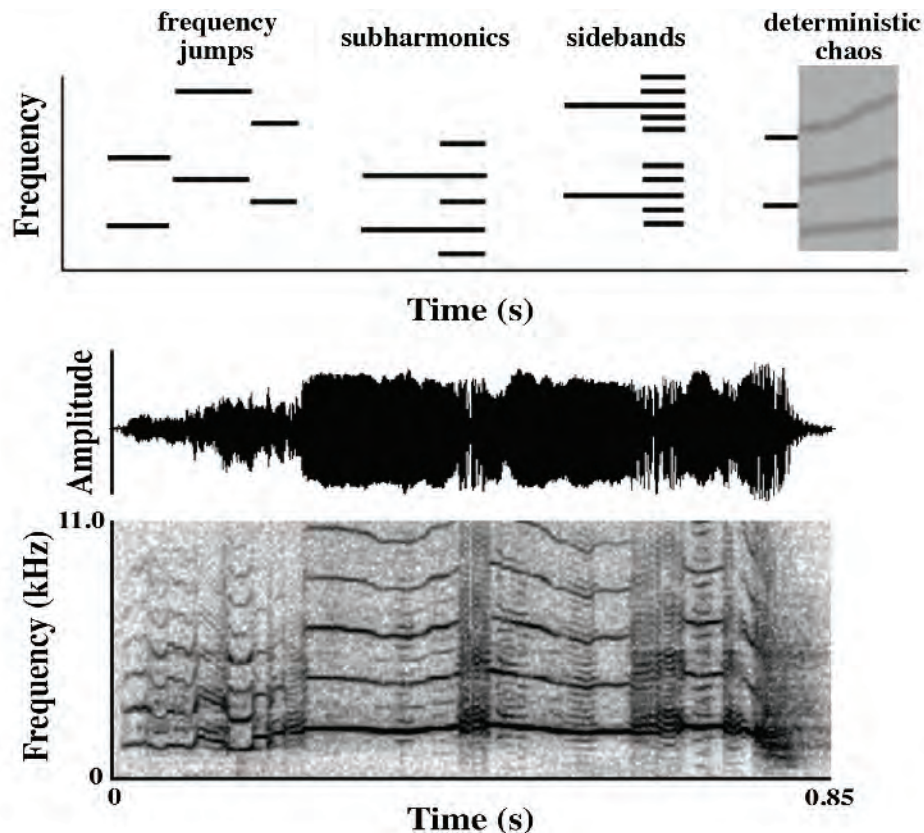


Fig. 2. (top) Schematic depictions of four kinds of “nonlinear phenomena”. Each vocalization begins in stable, harmonic form, then undergoes bifurcation to a different vocal-fold vibration regime. (bottom) A rhesus monkey scream that includes each of the nonlinear phenomena illustrated above. While less than a second in duration, the scream includes at least 22 bifurcations among qualitatively distinct vibration regimes.

tively. A key point is that the two vocal folds influence one another when vibrating, and thereby constitute a coupled, nonlinear-dynamical system. All vocalizations are therefore technically “nonlinear” in nature, with the vocal folds exhibiting characteristic vibration regimes that represent attractor states familiar from classic chaos theory (Wilden *et al.*, 1998; Fitch *et al.*, 2002). It is nonetheless useful to differentiate between *harmonic* vocalizations and *nonlinear phenomena*, illustrated in Fig. 2. The former reflect regular, well-synchronized vibration, while the latter include abrupt frequency jumps, perceptually jarring spectral sidebands believed to be produced by laryngeal amplitude-modulation effects, and viscerally grating deterministic chaos (Riede *et al.*, 2004).

While not yet systematically documented, nonlinear phenomena are likely present in every mammalian vocal repertoire, specifically including primates. A critical implication is that the biomechanics of the larynx itself can be primary in determining the qualitatively distinct vocal-types a given species produces (Brown *et al.*, 2003). In other words, whereas the vocalizer’s central nervous system determines global “system parameters” such as sub-glottal air pressure and laryngeal muscle tensions, the larynx itself is the ultimate arbiter of vocal-fold behavior. As in other nonlinear systems, the coupled vocal folds show “exquisite sensitivity” to minor changes in global parameters, with even very small changes potentially producing near-instantaneous bifurcation into

qualitatively different vibratory regimes and associated acoustics.

**Humans as primates**—Overall, it is clear that the human voice has ancient phylogenetic roots. Vocal-tract design is fundamentally similar across mammals, including humans, with corresponding operating principles. As in primate and non-primate mammals alike, the human larynx is a nonlinear-dynamical system whose vibration regimes represent attractor states that give rise to a range of qualitatively different source signals. Any such energy is subsequently shaped by supralaryngeal cavities, including when the source is simply turbulence in the airflow. In the absence of species-specific modifications, supralaryngeal filtering effects are expected to be similar in humans and larger-bodied mammals. Humans are also clearly mammal-like in being endowed with a repertoire of highly heritable, emotion-triggered signals such as spontaneous crying and laughter (Owren and Goldstein, 2008). These sounds emerge in recognizable form very early in life, without apparent need for practice or even to first hear the sounds from others (Owren *et*

*al.*, 2011). Infant crying in particular is marked by chaotic vibration (Mende *et al.*, 1990) resembling that observed in nonhuman primate screaming (Tokuda *et al.*, 2002). Spontaneous, emotion-triggered vocalizations remain important even as the child gains increasing volitional control over sound production and begins to speak.

Humans do have their own specializations, of course, including a thick, highly mobile tongue used to flexibly alter supralaryngeal resonances, and an exceptional degree of volitional control over sound production (Owren *et al.*, 2011). Because supralaryngeal filtering is largely static in nonhuman primates (although see Riede and Zuberbühler, 2003), their vocalizations can be characterized as fundamentally “laryngeal” in nature. In other words, vocal quality is primarily determined by the laryngeal vibration regime involved, which is also the case for spontaneous crying and laughter in humans. In contrast, human speech is marked by a relative paucity of source-energy types—essentially, quasi-periodic phonation versus turbulent noise. In other words, production is importantly “supralaryngeal,” with the tongue, mandible, and lips used to flexibly and dynamically create the many sounds of each different language.

Human vocal-fold structure and response also show important developmental changes (Schweinfurth and Thibeault, 2008; Hartnick *et al.*, 2005). One evident consequence is that the vibration regimes underlying the psyche-shattering shrieks and screams characteristic of young children



become difficult, if not impossible, for adults to produce. Instead, vocal-fold behavior appears to become more stable, centered on regular, synchronized vibration and associated harmonically-structured sounds. In fact, the vocal gymnastics of infants and children would constitute vocal abuse in adults, for whom chronic shouting or screaming can induce vocal-fold nodules and other pathologies (Stemple *et al.*, 2009). Suggestive evidence along these lines is also provided by a recent comparison of tickle-induced laughter in great apes and humans. While all five species produced distinctive-sounding laughter sounds, humans stood out from the others in showing significantly greater regularity in underlying vocal-fold action (Davila Ross *et al.*, 2009). A speculative but logical inference is that human vocal folds show evolutionary modification for more stable response across a range of air pressures and muscle tensions. While arguably losing some flexibility in laryngeal response, adult human voices have become less prone to non-linear phenomena. That change has created a requisitely higher proportion of regular, well-synchronized phonation, which in turn may have promoted the effectiveness of source-filter-based indexical cuing.

### Indexical cuing in the voice

Source-filter theory, laryngeal nonlinearity, and the similarities as well as differences between humans and other mammals create the foundation for understanding vocal indexical cuing. In a sense, all vocalizations must be considered inherently indexical, for instance in simply showing that a vocalizer is present. However, the more important consideration is how indexical cues are affected by the acoustics of a given vocalization. The indexical potency of harmonically structured sounds, in particular, is clearly evident from everyday experience alone. Here, the pitch and timbre of phonated speech allow listeners to immediately discern a talker's sex, identity,

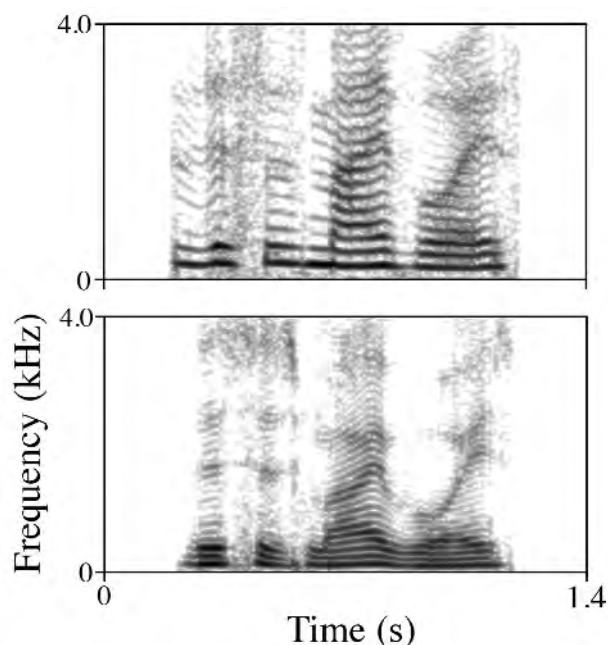


Fig. 3. Narrowband spectrograms of a human female (top) and male (bottom) saying the words “this is my voice.” The lower pitch and resonance in male voices makes formants more distinct and easier to measure than in female voices.

approximate age, and other personal characteristics. These capabilities are traceable to inherent differences in vocal-tract characteristics both among age-sex classes—such as adults versus children and males versus females—and among individuals within each group. For example, phonation allows even potentially subtle differences in vocal-fold size, shape, and tissue properties to be revealed in features such as  $F_0$ , relative noisiness of the glottal signal, and cycle-to-cycle variation in vibration. Thus, humans tested with male versus female voices require fewer than two waveform cycles—each corresponding to a single opening and closing of the glottis—to hear the difference (Owren *et al.*, 2007). Supralaryngeal filtering also contributes strongly to indexical cuing, even as talkers are dynamically altering the pharyngeal and oral cavities for linguistic purposes. Even brief segments of recorded vowel sounds show that details of formant patterning can provide important potential cues to both sex and individual identity (Bachorowski and Owren, 1999).

However, indexical cuing can be strongly affected by the nature of the source energy involved. As shown in Fig. 3 for male and female speech, for example, supralaryngeal cues become less evident as  $F_0$  increases. This effect occurs because harmonics occurs at integer multiples of  $F_0$  and raising this basic rate of vibration spaces them further apart. The source spectrum thereby becomes more sparsely populated, with less opportunity for supralaryngeal resonances to create a distinct imprint. Another way to understand this outcome is that formants become less well “sampled” by the source signal, giving the listener less to go on in recovering details of frequency, bandwidth, and amplitude. Some formants may not be sampled at all when  $F_0$ s become very high. Adding some noisiness to otherwise stable vocal-fold vibration can improve the situation, for instance by “filling out” the source spectrum. That effect occurs in breathy phonation in human talkers, as well as in the noisy, but nonetheless regularly phonated “roars” of red deer (*Cervus elaphus*) and other mammals (Taylor and Reby, 2010).

But too much noisiness becomes a liability. Reducing the source energy of speech to noise alone—as in whispering—makes both phonetic and indexical cuing less effective (Tartter, 1991; Katz and Assman, 2001). Deterministic chaos is nonetheless by far the greatest challenge to supralaryngeal cuing. As a general phenomenon, the occurrence of nonlinearity in a voice has been suggested contribute to individual identity signaling (Fitch *et al.*, 2002). Such events might, for example, occur idiosyncratically in particular vocalizers and thereby become compelling cues to their respective identities. Nonlinear vocal phenomena are by nature unstable, however, and therefore not likely to provide as consistent a substrate for indexical cuing as vocalizer-specific vocal-fold properties or supralaryngeal filtering (Rendall, 1996; Owren and Rendall, 2001). Furthermore, informal examination of a variety of chaos-based screams suggests that virtually no source- or resonance-related indexical cuing occurs in such sounds—no matter what species they are from (see Fig. 4). Empirically, direct comparisons of identity signaling in rhesus monkey and baboon vocalizations have shown that harmonically structured sounds are a markedly better vehicle.

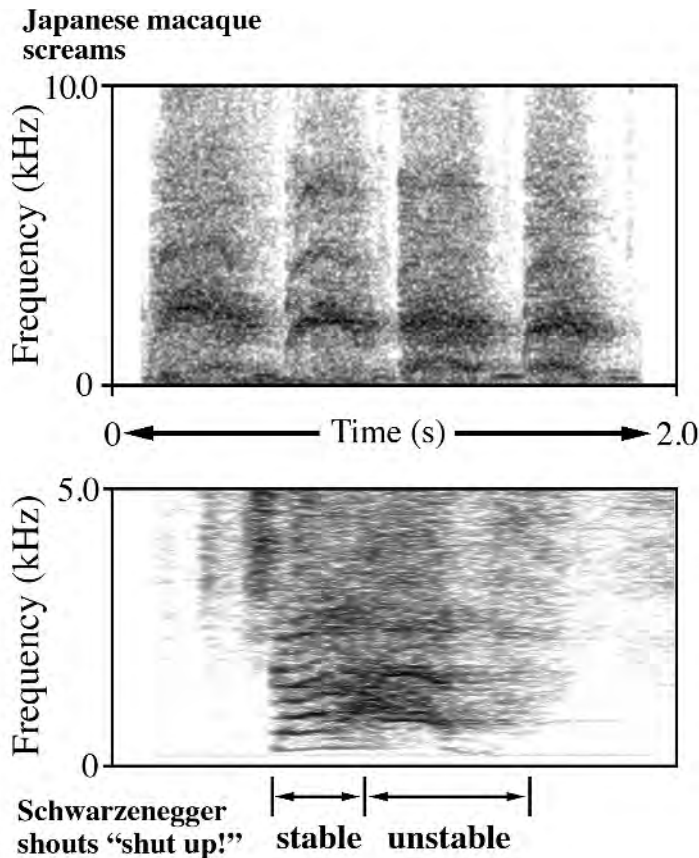


Fig. 4. Narrowband spectrograms of high-arousal, screams produced by an adult female macaque (top), and Arnold Schwarzenegger shouting the words “shut up,” as recorded from the movie *Kindergarten Cop* (1990). The macaque screams deterministic chaos throughout, with no apparent evidence of formant patterning. Schwarzenegger’s shout initially shows regular vocal-fold vibration (the “uh” pattern from “shut”), but then gives way to unstable, irregular, and likely chaotic action. Schwarzenegger’s initially distinctive voice quality is readily apparent in the stable portion, but disappears when the source energy becomes unstable.

Both species have been tested in playback experiments under naturalistic conditions, with adult females hearing either harmonic calls or chaotic screams (Rendall *et al.*, 1996; Rendall *et al.*, 1998; Rendall *et al.*, 2009). These listeners heard sounds from either their own or others’ offspring, or from adult females that were either biological kin or non-kin. Outcomes were unequivocal. When listeners heard harmonically structured calls, their responses clearly depended on the caller’s relationship to them. However, the subjects showed little or no evidence of differentiating among vocalizers when hearing screams. Naïve human listeners tested with rhesus calls in a lab setting were similarly significantly better at discriminating among individual callers when hearing harmonically versus chaotically structured vocalizations (Owren and Rendall, 2003). As yet, there is no ready explanation for the absence of filtering effects in these screams, a puzzle that begs for further investigation.

### Sex differences and sexual selection

Examining the possible impact of sexual selection on male and female voices has become an active and exciting area of research. Sex differences in human vocal characteristics are, of course, so familiar from everyday experience that

they are almost taken for granted. However, when working from an evolutionary perspective, noticing such differences almost reflexively triggers questions about their origin and possible function. In general, sexual selection is proposed to occur when individuals compete for access to opposite-sex mates (*intrasex competition*), or compete to be selected as a mate by members of the opposite sex (*mate-choice competition*). In both cases, one sex may acquire distinctive and unique features that need not have direct counterparts in the other. In humans, examples of these kinds of dimorphisms include body-fat distribution, facial morphology, and beard growth (Boyd and Silk, 2011). In such cases, sexual selection is suspected when differences cannot be readily explained as an artifact of more global dimorphisms, such as in body-size. The next step then becomes to show that the exaggerated features found in one sex or the other play a significant role in intrasex competition, mate-choice competition, or both.

Possible effects of body-size on the voice become important in that primate males are, in fact, larger than females in many species, including humans and all four great ape species. Furthermore, male-female differences in  $F_0$  are common without necessarily exceeding overall dimorphism (Mitani and Gros-Louis, 1995; Ey *et al.*, 2007). However, vocal dimorphisms can be disproportionate as well, which is the case for both  $F_0$  and formants in baboons (Rendall *et al.*, 2004). Outcomes for humans are similar, with adult males being approximately 8% taller and 15-20% heavier than females (Puts, 2010). Laryngeal dimorphism is quite disproportionate, with the vibrating segments of the adult vocal folds being about 60% longer in males than in females, which lower speaking  $F_0$  by approximately 50% (Titze, 1994). Dimorphism in vocal tract length is also disproportionate to height, being about 15–20% greater in males (Fant, 1960; Goldstein, 1980).

In humans, vocal-tract development proceeds along similar trajectories in males and females until puberty, at which point boys famously show marked laryngeal growth (Titze, 1994; Harries *et al.*, 1998). Physical changes include lengthening and thickening of the vocal folds, effects triggered by increases in circulating sex steroid levels—particularly testosterone. Both masculinizing and feminizing effects are classically hormone-related, with dimorphism resulting from differential tissue growth in one sex or the other (Dixson, 2009). In the male voice, the process can occur in as little as a year, but can also take up to five years. The larynx also shows a pubertal growth spurt in girls, but much more modestly. The vocal-tract also grows longer during this period, with male puberty being associated with a process of secondary laryngeal descent. This laryngeal lowering thereby lengthens the pharynx, ultimately positioning the male larynx a full vertebra’s distance below its female counterpart (Fitch and Giedd, 1999). Overall, then, evidence from both male and female anatomy and vocal acoustics are indicative of sexual selection in human vocal production.

### Intrasex competition

Within-sex competition is common among mammals, most frequently between males (Puts, 2010). As a rule, the larger individual wins in male-male contests, with many

encounters being resolved before escalation to violence. Vocalizations often play a key role in such cases, with intimidation through vocal signaling of size believed to be a critical factor (Bradbury and Vehrencamp, 2011). Results from a number of nonhuman species are consistent with this view, for instance demonstrating correlations among vocalizer body-size, vocal-tract length, and formant frequencies, as well as listeners' sensitivity to vocalizer resonance cues (Fitch, 1997; Fitch and Fritz, 2005; Harris *et al.*, 2006; Reby and McComb, 2003; Riede and Fitch, 1999). Using signaling to influence contest outcomes necessarily creates selection pressure for exaggeration, however, in this case of apparent size. In one extreme case, red deer males have been shown to lower their larynx more than 30 cm when vocalizing (Fitch and Reby, 2001). Females are indeed affected by resulting resonance cues, but do not show this effect themselves.

If voice-related intrasex competition also occurs in humans, it is thus reasonable to expect that males will be most affected. Male vocal characteristics in particular should be correlated with overall body-size, but vocalizers may also exaggerate those cues. Between age-sex classes, at least, it is clear that key vocal characteristics are significantly correlated with body-size. Both  $F_0$  and formants are lower in adults than in children (Hirano *et al.*, 1983; Hollien *et al.*, 1994), and in adult males than in adult females (Hillenbrand *et al.*, 1995; Rendall *et al.*, 2005). Human listeners are also sensitive to these differences, and use them to identify vocalizers as men, women, or children (Coleman, 1976; Owren *et al.*, 2007). However, correlations between voice and body-size are much weaker within age-sex class—including in adult males. In fact, there may be no relationship between  $F_0$  and body-size in either males or females (Rendall *et al.*, 2007). There is stronger evidence of a reliable correlation between vocal-tract length and body-size, but the degree of correlation is again modest, and not entirely consistent across studies. The picture is also complicated by the fact that human listeners are not very good at judging vocalizer body size (Collins, 2000; van Dommelen and Moxness, 1993; González, 2004). Furthermore, judgments tend to be based on  $F_0$  differences, which is the less-reliable cue (Rendall *et al.*, 2007). Formants do predominate when stimuli are equated for discriminability on the two dimensions (Pisanski and Rendall, 2011), but

with the caveat that naturally occurring resonance differences between the sexes are significantly smaller than pitch differences. In other words, equating for discriminability means presenting formant cues that are arguably proportionately larger than the  $F_0$  cues.

Overall, then, results concerning intrasex competition based on body-size signaling are mixed. On the one hand, it is clear that disproportionate sex differences do exist for both  $F_0$  and formants. Furthermore,  $F_0$  cues sway listener judgments for both male and female vocalizers, while it is specifically male versions that are exaggerated. A smaller, but detectable effect is also present for formants, most often in male voices. On the other hand, within-group correlations between vocal characteristics and body-size are uncertain for  $F_0$  and modest for formants. Furthermore, listeners are paradoxically more swayed by vocal pitch, which is almost entirely unreliable.

One possible explanation for these seemingly contradictory outcomes is that reliable body-size cues are unnecessary—vocalizers may instead be capitalizing on the strong, global relationship between physical size and both pitch and resonance that exists in the world at large. In other words, because a strong relationship exists between the size of an object or animal and associated pitch and resonance cues in the world at large, listeners are swayed even by unreliable vocal cues (Rendall *et al.*, 2004).  $F_0$  cues may also be easiest to exaggerate, as the human larynx grows more or less perpendicularly to the body axis, and can protrude from the neck without disturbing other tissue (Fitch, 2000; Fitch and Hauser, 2002). As the pharynx grows along the body axis and oral cavity length is likely constrained by mandible and tooth geometry, vocal-tract length remains more proportional to body-size as a whole. One might even argue that  $F_0$  cues have become exaggerated to the point of unreliability in human males, with formant cues differing only in being affected to a lesser degree. However, that account leaves unexplained why listeners would be differentially sensitive to the less reliable cue. A second, quite different argument is that  $F_0$  cuing is more accurate than hitherto believed. In this view, relying on height and weight differences importantly underestimates male-female differences. Specifically, human males have 60% more lean muscle-mass than females, and 80% greater mus-

## Loudspeakers and room acoustics carry the message



[www.odeon.dk](http://www.odeon.dk)





cle mass in the arms (Puts, 2010). To test that idea, researchers estimated male “threat potential” by measuring height, bicep size, hand strength, salivary testosterone level, and inherent aggressiveness. Outcomes showed stronger correlations between  $F_0$  and formants, and size, strength, and testosterone level than previously reported for either height or weight (Puts *et al.*, 2011; see also Sell *et al.*, 2010). The resulting argument is that the male voice does provide important, reliable cues to vocalizer competitive capabilities, and that listeners are responding reasonably to those cues.

### Mate-choice competition

Recent studies have also addressed the related question of whether vocal characteristics play an important role in mate-choice competition—here expecting both sexes to show such effects. The basic approach has been to ask listeners to rate the relative attractiveness of a variety of male and female voices, with pitch and resonance again being the critical variables. Testing females, it is common to find a preference for masculinized voices—meaning those with lower vocal pitch and resonances (Feinberg, 2008; Jones *et al.*, 2010; Pisanski and Rendall, 2011). From a mate-choice perspective, these characteristics may represent hormone-related ornamentation that has emerged precisely due to being attractive to females (Feinberg, 2008). Although the relationship is modest, male salivary testosterone levels have indeed been found to be inversely correlated with both  $F_0$  and formant frequencies (Dabbs and Mallinger, 1999; Bruckert *et al.*, 2006; Evans *et al.*, 2008). Other evidence supporting an influence of mate-choice selection includes a statistical correlation between these vocal characteristics and both number of children fathered (Apicella *et al.*, 2007) and number of sexual encounters reported (Hodges-Simeon *et al.*, 2011). Finally, listeners are more likely to expect infidelity from males with masculinized voices (O'Connor, 2011), which are also preferred more by females when approaching ovulation than at non-fertile times within the menstrual cycle (Feinberg *et al.*, 2006).

Males show approximately converse preferences, as might be expected. For instance, many studies have revealed a preference for higher-pitched female voices (Apicella and Feinberg, 2009; Pisanski and Rendall, 2011). This effect may be traceable to pitch as a fertility cue, with females being most fertile and having highest speaking  $F_0$  values in early adulthood (Stathopoulos *et al.*, 2011). Males have also been found to prefer voices of females recorded when close to ovulation (Pipitone and Gallup, 2008), a point in the menstrual cycle that is also associated with increased vocal pitch (Bryant and Haselton, 2009)—although not uniquely so (Fischer *et al.*, 2011). Other evidence includes increased pitch among females when believing they are communicating with more masculinized and attractive males (Fraccaro *et al.*, 2011), and women with higher-pitched voices are deemed more likely to exhibit infidelity (O'Connor, 2011). There are again complications, of course, but perhaps fewer than for intrasex competition in voice. For example, not all studies have found more masculine or feminine voices to be the most attractive. In at least one case, mid-range or aver-

age voices in the opposite sex have been the most attractive for both male and female listeners (Hughes *et al.*, 2010). The same work reported that all participants tended to speak at lower pitches when interacting with an attractive partner. While consistent with previously reported female preferences, the result is inconsistent with other findings for males. There is also disagreement as to whether  $F_0$  and formants work separately (Hodges-Simeon *et al.*, 2010; Jones *et al.*, 2010) or synergistically (Feinberg *et al.*, 2008; Feinberg *et al.*, 2011).

### Conclusions

This review has moved quickly and lightly over a variety of topics, each of which deserves much more thorough treatment. Nonetheless, the evidence covered underscores the fact that the human voice does have a long evolutionary history and has been importantly shaped through shared phylogeny with other species. Vocal-fold action and vocal-tract resonance have emerged as recurring themes, equally applicable to vocal production in humans and nonhuman mammals and creating substantive evolutionary connections between the two. However, it has also become apparent that hominin evolution also brought important changes. Understanding those changes raises questions that comparisons to other primates and mammals alone may not fully address. Yet, combining clues from other species with evidence of novel human vocal characteristics may ultimately prove to be an effective means of shedding further light on hominin evolution overall. Three issues will be briefly followed up in closing, including the evident weakness of correlations between human vocal characteristics and physical features such as overall body-size, possible changes in vocal-fold stability over hominin evolution, and the intertwining of indexical and phonetic cuing in speech—the most unique of human vocalizations.

**Sexual selection and the voice**—Understanding the role of sexual selection on the voice is a recent undertaking, and progress has been rapid. The overall approach has been validated not only by finding evidence of sexual-selected vocal effects in both sexes, but also by the fact that outcomes are predictably somewhat different in males versus females. However, it is also difficult to avoid the feeling that an important piece of the puzzle is still missing. Correlations between vocal and physical characteristics are too weak, for example, and it is not satisfying to invoke global correlations from the world at large to explain an apparently illusory relationship between pitch and size. Another possibility is that those human mating decisions have become sufficiently complex over evolutionary time that vocal characteristics have lost an important link to physical characteristics that they once had. However, a more compelling explanation may emerge through more substantive recasting of key vocalizer traits as a combination of physical and psychological characteristics, such as threat potential. Overall, understanding sexual selection effects in the human voice has some surprising, but interesting, complexities that may require imagination and re-thinking to untangle.

**Vocal-fold stability**—The question of whether vocal-fold response characteristics changed over the course of hominin evolution has broader potential implications than one might first imagine. For example, available comparative data indicate that vocal-fold composition can vary across primates and it furthermore appears that developmental modifications known to occur in humans are correlated with changes in acoustic output. Combining these two kinds of information can help illuminate relationships between vocal-fold morphology and how vocalizations were being used, thereby shedding new light on the adaptive changes occurring in hominins. The specific suggestion made here is that vocal action became more stable, especially in adults. This change would be natural to connect to increasing reliance on vowel-like sounds, for instance in association with the evolution of speech. However, greater vocal-fold stability may have emerged earlier to facilitate indexical cuing in the context of increasingly complex hominin social groups and relationships. While not directly connected to the emergence of speech, such changes may have helped set the stage for this development. Detailed knowledge of the relationship between vocal-fold structure and communicative function could be key in resolving such questions.

**Indexical and phonetic cuing**—A final note concerns the interplay of indexical and phonetic cuing. Historically, speech scientists have struggled to separate the two in seeking to find invariant physical features that distinguish indi-

vidual phonemes in the face of the acoustical variability occurring across talkers (as well as other factors). The upshot has been that cues to the phonetic content of speech may not exist independently of a given talker's personal, vocal characteristics. This conflation of the phonetic and indexical is understandable in that both flow simultaneously from the same source and filter system during speech. If so, however, then understanding phonetic cuing requires getting a handle on indexical cuing as well—which in turn brings the evolutionary perspective into the picture. In a sense, sexual selection effects in the voice have already had a notable effect on how phonetic features are understood. In the early years of modern acoustic phonetics, researchers focused mainly on adult male talkers, importantly because their speech revealed prominent and easily measured formants. Formant measurement was notably more difficult in speech from adult females or children, so much so that some came to consider the sound-spectrographic technology being used to be inherently “sexist.” There is some truth to that charge, particularly as it was later found that female speech is actually the more intelligible. As discussed earlier, finding prominent, well-defined formants in males versus other talkers is straightforwardly due to the forces of sexual selection lengthening their vocal folds and supralaryngeal vocal tracts. One can only wonder if history might have unfolded differently in the study of speech had the role of evolutionary forces on the voice itself been realized from the beginning.**AT**

## References

- Apicella, C. L., and Feinberg, D. R. (2009). “Voice pitch alters mate-choice-relevant perception in hunter-gatherers,” *Proc. Royal Soc. B* **276**, 1077–1082.
- Apicella, C. L., Feinberg, D. R., and Marlowe, F. W. (2007). “Voice pitch predicts reproductive success in male hunter-gatherers,” *Biology Lett.* **3**, 682–684.
- Bachorowski, J.-A., and Owren, M. J. (1999). “Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech,” *J. Acoust. Soc. Am.* **106**, 1054–1063.
- Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L., and Purvis, A. (2007). “The delayed rise of present-day mammals,” *Nature* **446**, 507–512.
- Boyd, R., and Silk, J. (2011). *How Humans Evolved* (W. W. Norton, New York).
- Bradbury, J. W., and Vehrencamp, S. L. (2011). *Principles of Animal Communication* (Sinauer Associates, New York).
- Brown, C. H., Alipour, F., Berry, D. A., and Montequin, D. (2003). “Laryngeal biomechanics and vocal communication in the squirrel monkey (*Saimiri boliviensis*),” *J. Acoust. Soc. Am.* **113**, 2114–2126.
- Bruckert, L., Liénard, J.-S., Lacrois, A., Kreutzer, M., and Leboucher, G. (2006). “Women use voice parameters to assess men's characteristics,” *Proc. Royal Soc. B* **273**, 83–89.
- Brudzynski, S. (editor) (2010), *Handbook of Mammalian Vocalization, Volume 19: An Integrative Neuroscience Approach* (Elsevier Science, Oxford, UK).
- Bryant, G. A., and Haselton, M. G. (2009). “Vocal cues of ovulation in human females,” *Current Biol.* **5**, 12–15.
- Chiba, T., and Kajiyama, J. (1941). *The Vowel: Its Nature and Structure* (Tokyo: Tokyo-Kaiseikan).
- Coleman, R. O. (1976). “A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness of the voice,” *J. Speech and Hearing Res.* **19**, 168–180.
- Collins S. (2000). “Men's voices and women's choices,” *Animal Behaviour* **60**, 773–780.
- Dabbs, Jr. J. M., and Mallinger, A. (1999). “High testosterone levels predict low voice pitch among men,” *Personality and Individual Differences* **27**, 801–804.
- Davila Ross, M., Owren, M. J., and Zimmermann, E. (2009). “Reconstructing the phylogeny of laughter in great apes and humans,” *Current Biol.* **19**, 1106–1111.
- Dixson, A. F. (2009). *Sexual Selection and the Origins of Human Mating Systems* (Oxford University Press, New York).
- Evans, S., Neave, N., Wakelin, D., and Hamilton, C. (2008). “The relationship between testosterone and vocal frequencies in human males,” *Physiology & Behavior* **93**, 783–788.
- Ey, E., Pfefferle, D., and Fischer, J. (2007). “Do age- and sex-related variations reliably reflect body size in non-human primate vocalizations? A review,” *Primates* **48**, 253–267.
- Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, The Hague).
- Feinberg, D. R. (2008). “Are human faces and voices ornaments signaling common underlying cues,” *Evolutionary Anthropol.* **17**, 112–118.



- Feinberg, D. R., DeBruine, L. M., Jones, B. C., and Little, A. C. (2008). "Correlated preferences for men's facial and vocal masculinity," *Evolution and Human Behavior* **29**, 233–241.
- Feinberg, D. R., Jones, B. C., DeBruine, L. M., O'Connor, J. J. M., Tigue, C. C., and Borak, D. J. (2011). "Integrating fundamental and formant frequencies in women's preferences for men's voices," *Behavioral Ecology* **22**, 1320–1325.
- Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., Hillier, S. G., and Perrett, D. I. (2006). "Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice," *Hormones and Behavior* **49**, 215–222.
- Fischer, J., Semple, S., Fickenscher, G., Jurgens, R., Kruse, E., Heistermann, M., and Amir, O. (2011). "Do women's voices provide cues of the likelihood of ovulation? The importance of sampling regime," *PLoS One* **6**, e24490.
- Fitch, W. T. (1997). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," *J. Acoust. Soc. Am.* **102**, 1213–1222.
- Fitch, W. T. (2000). "The evolution of speech: A comparative review," *Trends in Cognitive Sciences* **4**, 258–267.
- Fitch, W. T. (2006). "Production of vocalizations in mammals," in *Encyclopedia of Language and Linguistics, 2nd Ed., Vol. 10*, edited by K. Brown (Elsevier Science, Oxford, UK), pp. 115–121.
- Fitch, W. T., and Fritz, W. B. (2006). "Rhesus macaques spontaneously perceive formants in conspecific vocalizations," *J. Acoust. Soc. Am.* **120**, 2132–2141.
- Fitch, W. T., and Giedd, J. (1999). "Morphology and development of the human vocal tract. A study using magnetic resonance imaging," *J. Acoust. Soc. Am.* **106**, 1511–1522.
- Fitch, W. T., and Hauser, M. D. (2002). "Unpacking 'honesty': Vertebrate vocal production and the evolution of acoustic signals," in *Acoustic Communication*, edited by A. M. Simmons, R. R. Fay, and A. N. Popper (Springer Verlag, New York), pp. 65–137.
- Fitch, W. T., Neubauer, J., and Herzel, H. (2002). "Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production," *Animal Behaviour* **63**, 407–418.
- Fitch, W. T., and Reby, D. I. (2001). "The descended larynx is not uniquely human," *Proc. Royal Soc. B* **268**, 1669–1675.
- Fraccaro, P. J., Jones, B. C., Vukovic, J., Smith, F. G., Watkins, C. D., Feinberg, D. R., Little, A. C., and DeBruine, L. M. (2011). "Experimental evidence that women speak in a higher voice pitch to men they find attractive," *J. Evolutionary Psychol.* **9**, 57–67.
- Goldstein, U. G. (1980). "An articulatory model for the vocal tracts of growing children," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- González, J. (2004). "Formant frequencies and body size of a speaker: A weak relationship in adult humans," *J. Phonetics* **32**, 277–287.
- Harries, M., Hawkins, S., Hacking, J., and Hughes, I. (1998). "Changes in the male voice at puberty: Vocal fold length and its relationship to the fundamental frequency of the voice," *J. Laryngol. and Otology* **112**, 451–454.
- Harris, T. R., Fitch, W. T., Goldstein, L. M., and Fashing, P. I. (2006). "Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass," *Ethology* **112**, 911–920.
- Hartnick, C. J., Rehbar, R., and Prasad, V. (2005). "Development and maturation of the pediatric human vocal fold lamina propria," *The Laryngoscope* **115**, 4–15.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Hirano, M., Kurita, S., and Nakashima, T. (1983). "Growth, development and aging of human vocal folds," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 22–43.
- Hodges-Simeon, C. R., Gaulin, S. J. C., and Puts, D. A. (2010). "Different vocal parameters predict perceptions of dominance and attractiveness," *Human Nature* **21**, 406–427.
- Hodges-Simeon, C. R., Gaulin, S. J. C., and Puts, D. A. (2011). "Voice correlates of mating success in men: Examining 'Contests' versus 'mate choice' modes of sexual selection," *Archives of Sexual Behavior* **40**, 551–557.
- Hollien, H., Green, R., and Massey, K. (1994). "Longitudinal research on adolescent voice change in males," *J. Acoust. Soc. Am.* **96**, 2646–2654.
- Hughes, S. M., Farley, S. D., and Rhodes, B. C. (2010). "Vocal and physiological changes in response to the physical attractiveness of conversational partners," *J. Nonverbal Behavior* **34**, 155–167.
- Isles, T. E. (2009). "The socio-sexual behaviour of extant archosaurs: Implications for understanding dinosaur behaviour," *History of Biol.* **21**, 139–214.
- Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., and Vukovic, J. (2010). "A domain-specific opposite-sex bias in human preferences for manipulated voice pitch," *Animal Behaviour* **79**, 57–62.
- Katz, W. F., and Assmann, P. F. (2001). "Identification of children's and adults' vowels: Intrinsic fundamental frequency, fundamental frequency dynamics, and presence of voicing," *J. Phonetics* **29**, 23–51.
- Mende, W., Herzel, H., and Wermke, K. (1990). "Bifurcation and chaos in newborn infant cries," *Physics Lett. A* **145**, 418–425.
- Mitani, J. C., and Gros-Louis, J. (1995). "Species and sex differences in the screams of chimpanzees and bonobos," *Int. J. Primatology* **16**, 393–411.
- O'Connor, J. J. M. (2011). "Voice pitch influences perceptions of sexual infidelity," *Evolutionary Psychol.* **9**, 64–78.
- Owren, M. J., Amoss, R. T., and Rendall, D. (2011). "Two organizing principles of vocal production: Implications for nonhuman and human primates," *Am. J. of Primatology*, **73**, 530–544.
- Owren, M. J., Berkowitz, M., and Bachorowski, J.-A. (2007). "Listeners judge talker sex more efficiently from male than from female vowels," *Perception & Psychophysics* **69**, 930–941.
- Owren, M. J., and Goldstein, M. H. (2008). "The babbling-scaffold hypothesis: Subcortical primate-like circuitry helps teach the human cortex how to talk," in *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication*, edited by D. K. Oller and U. Griebel (MIT Press, Cambridge, MA) pp. 169–192.
- Owren, M. J., and Rendall, D. (2001). "Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling," *Evolutionary Anthropol.* **10**, 58–71.
- Owren, M. J., and Rendall, D. (2003). "Salience of caller identity in rhesus monkey (*Macaca mulatta*) coo and screams: Perceptual experiments with human listeners," *J. Comparative Psychol.* **117**, 380–390.
- Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (1997). "The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cynocephalus ursinus*): Implications for production processes and functions," *J. Acoust. Soc. Am.* **101**, 2951–2963.

- Pipitone, R. N., and Gallup Jr., G. G. (2008). "Women's voice attractiveness varies across the menstrual cycle," *Evolution and Human Behavior* **29**, 268–274.
- Pisanski, K., and Rendall, D. (2011). "The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness," *J. Acoust. Soc. Am.* **129**, 2201–2212.
- Puts, D. A. (2010). "Beauty and the beast: mechanisms of sexual selection in humans," *J. Human Evolution* **31**, 157–175.
- Puts, D. A., Apicella, C. L., and Cárdenas, R. A. (2011). "Masculine voices signal men's threat potential in forager and industrial societies," *Proc. Royal Soc. B*. doi: 10.1098/rspb.2011.0829.
- Reby, D., and McComb, K. (2003). "Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer," *Animal Behaviour* **65**, 519–530.
- Rendall, D. (1996). *Social Communication and Vocal Recognition in Free-Ranging Rhesus Monkeys*. unpublished doctoral dissertation, University of California-Davis.
- Rendall, D., Emond, R. E., and Rodman, P. S. (1996). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," *Animal Behaviour* **51**, 1007–1015.
- Rendall, D., Lloyd, P., Kollias, S., and Ney, C. (2005). "Pitch ( $F_0$ ) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," *J. Acoust. Soc. Am.* **117**, 944–955.
- Rendall, D., Notman, H., and Owren, M. J. (2009). "Asymmetries in the individual distinctiveness and maternal recognition of infant contact calls and distress screams in baboons," *J. Acoust. Soc. Am.* **125**, 1792–1805.
- Rendall, D., Owren, M. J., and Rodman, P. S. (1998). "The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations," *J. Acoust. Soc. Am.* **103**, 602–614.
- Rendall, D., Owren, M. J., Weerts, E., and Hienz, R. D. (2004). "Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners," *J. Acoust. Soc. Am.* **115**, 411–421.
- Rendall, D., Vokey, J. R., and Nemeth, C. (2007). "Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size," *J. of Experimental Psychol.: Human Perception and Performance* **33**, 1208–1219.
- Riede, T., Owren, M. J., and Clark Arcadi, A. (2004). "Nonlinear acoustics in the pant-hoot vocalizations of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos," *Am. J. Primatology* **64**, 277–291.
- Riede, T., and Fitch, W. T. (1999). "Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*)," *J. Experimental Biol.* **202**, 2859–2867.
- Riede, T., and Zuberbühler, K. (2003). "The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization," *J. Acoust. Soc. Am.* **114**, 1132–1142.
- Schweinfurth, J. M., and S. L. Thibeault, S. L. (2008). "Does hyaluronic acid distribution in the larynx relate to the newborn's capacity for crying?" *Laryngoscope* **118**, 1692–1699.
- Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., Krauss, A., and Gurven, M. (2010). "Adaptations in humans for assessing physical strength from the voice," *Proc. Royal Soc. B* **277**, 3509–3518.
- Stathopoulos, E. T., Huber, J. E., and Sussman, J. E. (2011). "Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age," *J. Speech, Language, and Hearing Res.* **54**, 1011–1021.
- Stemple, J. C., Glaze, L. E., and Klaben, B. G. (2009). *Clinical Voice Pathology: Theory and Practice, 4th Ed.* (Plural Publishing, San Diego).
- Stevens, K. N. (2000). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
- Tartter, V. C. (1991). "Identifiability of vowels and speakers from whispered syllables," *Perception & Psychophysics* **49**, 365–371.
- Taylor, A. M., and Reby, D. (2010). "The contribution of source-filter theory to mammal vocal communication research," *J. Zoology* **280**, 221–236.
- Titze, I. R. (1994). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
- Tokuda, I., Riede, T., Neubauer, J., Owren, M. J., and Herzel, H. (2002). "Nonlinear prediction of irregular animal vocalizations," *J. Acoust. Soc. Am.* **111**, 2908–2919.
- van Dommelen, W. A., and Moxness, B. H. (1993). "Acoustic parameters in speaker height and weight identification: Sex-specific behaviour," *Language and Speech* **38**, 267–287.
- Weishampel, D. B. (1997). "Dinosaurian cacophony," *Bioscience* **47**, 150–159.
- Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (1998). "Subharmonics, biphonation, and deterministic chaos in mammalian vocalization," *Bioacoustics* **9**, 171–196.



Michael J. Owren received his B.A. in Psychology from Reed College, and his Ph.D. in Animal Behavior from Indiana University, Bloomington. He is a Fellow of the Acoustical Society of America (2011), as well as the Association for Psychological Science (2011). Following a National Research Service Award (NRSA) postdoctoral fellowship jointly sponsored at the University of Pennsylvania and the University of California-Davis, Michael held faculty positions at the University of Colorado at Denver, Reed College, and Cornell University. He is currently an Associate Professor in Psychology at Georgia State University, where he is also Chair of the Cognitive Sciences program. Broadly stated, his research focuses on the role of acoustic structure and communicative function in the evolution of both human and nonhuman vocalizations, particularly including primate calls, emotion-triggered human vocalizations such as spontaneous laughter, and human speech. Michael has been a Consulting Editor for *Psychological Science* and the *Journal of Comparative Psychology*, and is currently an Associate Editor for *Emotion Review* and the *Journal of the Acoustical Society of America* (Bioacoustics).