

# Oxford Handbooks Online

## **The Evolution of Voice Perception**

Katarzyna Pisanski and Gregory A. Bryant

The Oxford Handbook of Voice Studies

*Edited by Nina Sun Eidsheim and Katherine Meizel*

Print Publication Date: Jul 2019 Subject: Music, Sound Studies

Online Publication Date: Jun 2019 DOI: 10.1093/oxfordhb/9780199982295.013.29

### **Abstract and Keywords**

The human voice is a rich source of information and an important means of interpersonal communication. Beginning with Darwin (1872), nonverbal vocal communication has long interested evolutionary scientists, and in the last quarter century empirical research on voice production and perception from an evolutionary perspective has increased dramatically. One reason for this surge of interest is that behavioral ecologists and evolutionary psychologists have taken advantage of recent technological improvements in acoustic analysis software as well as sound recording and reproduction devices. More importantly, many voice researchers have recognized that the extraction of biologically relevant information from the vocal channel constitutes a set of adaptive problems widely shared across many species. Evolutionary scientists studying human vocal behavior therefore have a rich theoretical framework and an established comparative basis for developing specific research questions. For any vocal species, including humans, we should expect perceptual adaptations designed to process acoustic features of the vocal sounds of conspecifics (other individuals of the same species). An evolutionary approach provides a framework for specifying the nature of these adaptive perceptual problems. In this chapter, the authors describe recent work focusing on human voice perception from an evolutionary perspective and provide examples of the value of this approach for gaining a full understanding of this fundamental aspect of human behavior.

Keywords: behavioral ecology, emotional vocal signals, evolution, evolutionary psychology, interpersonal communication, sexual dimorphism, signal design, speaker intention, vocal attractiveness, voice perception

---

### **(p. 269) Introduction**

THE human voice is a rich source of information and an important means of interpersonal communication. Beginning with Darwin (1872), vocal communication has long interested evolutionary scientists, and in the last quarter century empirical research on voice production and perception from an evolutionary perspective has increased dramatically. One

## The Evolution of Voice Perception

---

reason for this surge of interest is that behavioral ecologists and evolutionary psychologists have taken advantage of recent technological improvements in acoustic analysis software as well as sound recording and reproduction devices. More importantly, many voice researchers have recognized that the extraction of biologically relevant information from the vocal channel constitutes a set of adaptive problems widely shared across many species. Evolutionary scientists studying human vocal behavior therefore have a rich theoretical framework and an established comparative basis for developing specific research questions.

For any vocal species, we should expect perceptual adaptations designed to process acoustic features of the vocal sounds of conspecifics (e.g., other individuals of the same species). Humans are no exception—there is strong evidence that dedicated areas of the human brain, including the middle superior temporal sulcus (STS), are specialized for human voice perception (Belin et al. 2000; Pernet et al. 2015). The human STS is analogous to the vocal perception brain areas in several other species, such as macaques (Petkov et al. 2008). Much like how we learn to process faces and develop face-specific regions in our brains (Kanwisher, McDermott, and Chun 1997), voice-selective areas in humans reliably develop within seven months of birth (Grossmann et al. 2010), and perceptual biases toward speech sounds over similar nonspeech sounds appear as early as an infant's first day of life (Voloumanos and Werker 2004).

**(p. 270)** The comparison of human voice processing to face processing is more than superficial. Belin, Fecteau, and Bedard (2004) suggested that low-level sound features are extracted in the primary auditory cortex and then encoded in the STS for other, functionally distinct tasks, such as extracting speech information, detecting emotions, and identifying the speaker. (See Cornelia Fales's chapter for discussion of timbre's analogy to face processing, where she draws on the same body of scholarship.) The cognitive architecture of face processing follows a similar strategy, with low-level processors feeding into systems solving social identification, affective and related perceptual tasks. It is also not surprising that face and voice processing highly interact, though the details of this multimodal interaction are not well understood (Campanella and Belin 2007). An evolutionary approach provides a framework for specifying the nature of these adaptive perceptual problems. In this chapter, we will describe recent work focusing on voice perception from an evolutionary perspective and will provide examples of the value of this approach for gaining a full understanding of this fundamental aspect of human behavior.

### Levels of Analysis in the Evolutionary Study of Behavior

Tinbergen (1952, 1963) proposed that to understand any animal's behavior, we must answer empirical questions at different levels of analysis. Proximate questions focus on identifying specific causal processes that underlie a behavior. These causal processes often involve physical mechanisms and can include hormonal, neural, and other physiological systems in the body, as well as the vast array of developmental systems that contribute to adult phenotypes. Researchers in various disciplines typically define behavior as the product of numerous mechanisms, but these proximate analyses are often consid-

## The Evolution of Voice Perception

---

ered in the absence of an evolutionary approach (Barkow, Cosmides, and Tooby 1992). Tinbergen emphasized that ultimate questions, which instead focus on identifying the potential fitness consequences of a trait or behavior, must also be asked in order to fully understand a behavior.

The evolutionary analysis of any animal behavior involves the consideration of relevant adaptations underlying a given behavioral trait, as well as the phylogenetic history linking that trait across multiple species over time. Adaptations are evolved solutions to recurrent ecological problems faced by an organism: they are functionally organized, contribute positively to an organism's fitness, and are typically analyzed in terms of design features tailored to specific criteria. In a phylogenetic analysis, the evolutionary history of a given trait is reconstructed through cross-species comparative analysis. Functional and phylogenetic analyses together provide the ultimate explanation for a given behavioral or morphological characteristic.

### Form and Function in Signal Design

Evolutionary behavioral scientists studying human and nonhuman vocal behavior must make a distinction between communicative *signals* and informative *cues* (Maynard Smith and Harper 2003). (p. 271) Signals are defined as adaptive behaviors or structures shaped by selection to influence the behavior of others in a way that is beneficial to the sender (e.g., a vocalizer), and often coevolve with the adaptive responses of the receiver of that signal (e.g., a listener). Typically, signal production must benefit both senders and receivers to be evolutionarily stable. Cues, on the other hand, are any predictive behaviors or structures that influence receivers but were not designed to do so. Receivers might have evolved sensitivities to the predictive information, but the cues did not evolve to have that effect. Although many researchers studying communication have ignored this distinction, focusing instead on proximate levels of analysis involving the physical properties and mechanisms of communicative acts, the role of context in their perception, as well as cultural and developmental factors, both proximate and ultimate levels of analysis are important when seeking a complete understanding of any behavior.

Determining whether a communicative act regularly produced by an organism is a signal or a cue is sometimes difficult. Behaviors often have systematic effects on audiences, some by design and some not. The physical form of a signal can provide important clues regarding its function, and hence a form-fit analysis is often a good place to start when asking whether a trait or behavior evolved for a particular reason (Owren and Rendall 2001). Form-fit analysis investigates whether structural features of a trait show evidence of special design for solving specific adaptive problems (Lauder 1981; Williams 1966). Consider the case of human crying: signal or cue? Crying behavior in infants is clearly related phylogenetically to a variety of mammalian infant vocalizations designed to elicit parental care (for a review see Newman 2007). In humans, crying has particular acoustic features that exploit human audition, and has been shaped by coevolutionary processes reflective of the conflict of interest between senders (e.g., infants) and receivers (e.g., caretakers). Crying elicits investment of targeted listeners in the sender by motivating lis-

## The Evolution of Voice Perception

---

teners to stop the signal, explaining a great deal of its aversive sound characteristics. Crying is clearly a signal in that it evolved to influence caretaker behavior in a manner that benefits both sender and receiver. In order for it to be effective (and affective) and to fulfill its function, crying requires a sound profile that is displeasing to listeners. While there is some evidence that the acoustic characteristics of a cry can communicate information about the particular conditions that triggered the crying (Soltis 2004; Zeifman 2001), it could also be the case that very little information about specific conditions is encoded in the cry signal itself and instead, listeners can infer with some reliability what the infant's needs are based on contextual information alone. Arousal and emotional valence can vary dramatically across cries, which may help in this judgment.

In addition to qualifying as a signal, crying also has byproduct cue value.<sup>1</sup> For example, in ancestral environments, a wailing infant could inform predators about its location and this could have negative fitness consequences not only for the crying infant, but for siblings and parents as well. This heavy potential cost likely figures into the evolutionary dynamics contributing to its effectiveness. A listener wants the crying to stop, not only because it is (proximately) annoying to listen to, but also because in ancestral environments it could be dangerous to broadcast. Cries also potentially reveal information to caretakers (or others) about the crier that is not provided by design. For instance, (p. 272) crying can reveal the infant's health condition that could affect investment decisions by parents. Some researchers have proposed that signaling health information, or vigor, could be one function of crying (Furlow 1997; Lummaa et al. 1998; Soltis 2004), but direct evidence for this proposal is weak. Nevertheless, some basic information is not given by the acoustic structure of cries. For example, cries from infant boys and girls do not systematically differ acoustically, yet listeners will attribute higher-pitched cries to girls and lower-pitched cries to boys. This assumption can affect men's judgments of discomfort (but not women's), as low pitched cries from presumed boys result in greater ratings of discomfort by adult men, revealing sex-stereotype biases in judgment patterns (Reby, Levréro, Gustafsson, and Mathevon 2016).

The example of crying illustrates the complexity of separating byproduct effects of vocal signals from adaptive effects that have been shaped by selection to benefit senders and receivers. In the following section, we will describe research examining the perception of body size and strength through vocal cues. We use the term cue when describing this work because it is not clear that the extracted information (e.g., assessed size or strength) is provided by design, or instead perceivable incidentally as a consequence of source-filter dynamics in vocal production machinery. That said, humans and other animals might advertise or exaggerate their size in various ways by exploiting the relationships between their bodies and their vocal apparatus, and if so, such behaviors should be considered signals in that they are adaptations for sending that specific information.

# Can We Judge Body Size and Strength from the Voice?

Like most terrestrial mammals, human vocalizations are generally produced by pushing air from the lungs up through the closed glottis, vibrating the vocal folds within the larynx, and generating phonation that is subsequently filtered by the supralaryngeal vocal tract. The human supralaryngeal vocal tract (henceforth vocal tract) is comprised of the pharyngeal, oral, and nasal cavities situated above the larynx. This two-stage mode of vocal production produces the two most salient nonverbal characteristics of the human voice: the fundamental frequency (denoted as  $F_0$ ), that together with its harmonics we perceive as voice pitch, and formant frequencies or formants, that are resonances of the vocal tract and affect our perception of voice timbre (Titze 1994). The source-filter theory of speech production (Chiba and Kajiyama 1958; Fant 1960) was key in our understanding that fundamental and formant frequencies are largely independent of one another, and the theory has been central to recent advances in bioacoustics, including research examining vocal communication of body size and threat.

For several decades researchers have been interested in whether the human voice can provide reliable indexical information about the person speaking, in particular, the size (p. 273) of that person's body, their age, and sex (Kreiman and Sidtis 2011). Researchers have also attempted to understand whether the human voice was shaped by selection to advertise these various traits. Indeed, the ability to accurately gauge another individual's body size and strength could be advantageous in many social contexts for the sender or the receiver, or both.

Physical size and strength are key predictors of men's fighting ability. Both traits greatly influence women's mate preferences across diverse cultures, and both have a large impact on an individual's health and life history (Frederick and Haselton 2007; Gallup, White, and Gallup Jr. 2007; Peters 1986; Pisanski and Feinberg 2013). Among humans, body size estimation may also be a necessary precursor for speaker normalization, allowing listeners to recognize speech sounds (e.g., the difference between the vowel sounds "a" and "e") independent of the size of the vocalizer (Patterson, Smith, van Dinther, and Walters 2008). In this section, we will review studies that together suggest that reliable indicators of size and strength are in fact present not only in the visual modality, but also in the acoustic modality, and that listeners are able to estimate size and strength from the human voice alone with some degree of accuracy.

## Perception of Body Size from the Voice

The mammalian larynx and vocal tract typically grow larger along with the rest of the body as an individual develops and matures. More massive vocal folds within a larger larynx will vibrate at a slower rate than will smaller vocal folds, producing a relatively lower voice pitch; however, the length and tension of the vocal folds also affect pitch (Hollien 2014; Titze 2011).<sup>2</sup> Independently of vocal fold dynamics, formant frequencies are in-

## The Evolution of Voice Perception

---

versely related to the length of the vocal tract such that taller people typically have longer vocal tracts and lower formants than do shorter people (Fitch and Giedd 1999; Titze 1994). As a consequence, both fundamental frequency (perceived as voice pitch) and formant frequencies (formants, affecting perception of voice timbre) independently track differences in body size between adults and juveniles. Voice pitch and formants are also sexually dimorphic (lower in males than females) and therefore also track differences in body size between the sexes in a range of mammalian species, including humans (Fitch and Hauser 2003; Rendall, Kollias, Ney, and Lloyd 2005).

In a meta-analysis, Pisanski, Fraccaro, Tigue, O'Connor, Röder, Andrews, and Fink et al. (2014) showed that within same-sex groups of adult men and women, formants provide more reliable cues to body size than does voice pitch. This is most likely due to relatively greater anatomical constraints on the development of the vocal tract (which is constrained by skull size) than on the development of the vocal folds, resulting in relatively stronger relationships between formants and body size than between pitch and body size (Fitch 1994; 1997). Indeed, this prediction is also supported by work on several other species whose vocal production follows the source-filter model (for reviews see Ey, Pfeifferle, and Fischer 2007; Fitch and Hauser 2003; Taylor, Charlton, and Reby, 2016; Taylor and Reby 2010). In women, formants further explain a good deal of variation in (p. 274) body shape (e.g., waist-to-hip ratio; Pisanski, Jones, Fink, O'Connor, DeBruine, Röder, and Feinberg 2016).

The dissociation between voice pitch and actual body size among men and women may be tied to a number of factors. For instance, male vocal fold size and subsequent voice pitch are more heavily determined by exposure to testosterone than by body growth (Harries et al. 1998). In fact, individual differences in men's voice pitch in adulthood are predicted by their voice pitch in childhood, and voice pitch in childhood is predicted by voice pitch in infancy (Levrero, Mathevon, Pisanski, Gustafsson & Reby, 2018), suggesting that men's voice pitch could be tied the androgen levels in their mother's uterus (Fouquet et al. 2016). Humans are also capable of voluntarily altering their voice pitch, which is common during regular speech production or singing as well as during everyday social interaction, and may potentially disguise or exaggerate cues to size (Pisanski, Cartei, McGettigan, Raine, and Reby 2016). Indeed, men and women from diverse cultures have been shown to volitionally and spontaneously lower their fundamental frequency (and formant frequencies) when instructed to sound physically larger, and to raise both vocal frequencies when instructed to sound physically smaller (Pisanski, Mora, Pisanski, Reby, Sorokowski, Frackowiak, and Feinberg 2016).

Ignoring for a moment the obvious next question of just how accurately listeners can gauge size from the voice, a large body of work provides compelling evidence of strong systematic perceptual biases on size judgments from the voice. Studies consistently report that listeners from various cultures associate both low voice pitch as well as low formants with large body size between and within sexes (reviewed in Pisanski, Fraccaro, Tigue, O'Connor, and Feinberg 2014). These perceptual associations are only partially grounded in reality because voice pitch is in fact a very weak predictor of body size with-

## The Evolution of Voice Perception

---

in sexes. For this reason, some researchers have suggested that the perceived association between low voice frequencies and large body size represents a deep-rooted and very general perceptual bias linking any low-frequency sound to largeness (Morton 1977; Ohala 1984; Rendall, Vokey, and Nemeth 2007).<sup>3</sup>

Despite the often erroneous perceptual bias linking low voice pitch to large size, several studies suggest that listeners can gauge body size from the voice with accuracy above chance. However, listener's performance is variable and generally modest. Reanalysis of early work by Lass and colleagues indicated that 14 percent of listeners' size estimates correlated with the absolute height or weight of the vocalizers in those studies (González 2003; 2006). When listeners were asked to report absolute height or weight, van Dommelen and Moxness (1995) found that estimated height predicted actual height only for men's and not for women's voices, and that men were more accurate than were women at assessing men's size. Collins (2000) and Bruckert et al. (2006) found that women were able to estimate men's weights but not men's heights, but that men's actual weight explained only 22 percent (Collins 2000) and 16 percent (Bruckert et al. 2006) of the variance in women's estimates.

Forced-choice tasks that involve simply indicating which of two vocalizers is taller elicit comparatively greater accuracy in size estimation than do absolute size estimates. On average, both men and women can correctly identify the taller of two men (p. 275) around 60 percent to 90 percent of the time, depending on the size difference between the men (Oliver and González 2004; Pisanski, Fraccaro, Tigue, O'Connor, and Feinberg 2014; Rendall et al. 2007). Accuracy in size judgments is comparable (Oliver and González 2004) or slightly lower (Rendall et al. 2007) for assessments of women's compared to men's body size.

Even blind persons can assess relative differences in men's (Pisanski, Oleszkiewicz, and Sorokowska 2016) and women's (Pisanski et al. 2017) heights from speech, suggesting that visual experience is not necessary for this capacity. This is further supported by empirical evidence that infants as young as three months of age use vocal cues to gauge body size (Pietraszewski et al. 2017). Importantly, despite a lack of visual experience linking the size of people's bodies with their voices, congenitally blind and late blind adults use the same perceptual rules when assessing body size as do sighted individuals (i.e., associating both low pitch and low formants with largeness even though low pitch does not predict size within-sexes; Pisanski et al. 2017). This provides further support for the hypothesis that body size estimation is based on a deep-rooted and general perceptual bias wherein low frequencies are judged as emanating from large sound sources.

Recent work has employed modern voice synthesis and analysis techniques to highlight the important interplay between fundamental and formant frequencies in the perception of body size from the voice. For example, experiments using natural (Pisanski, Fraccaro, Tigue, O'Connor, and Feinberg 2014) as well as synthetic (Irino et al. 2012) whispered speech (which is largely devoid of pitch information, as whispering involves very little vibration of the vocal folds) indicate that voice pitch is not essential for body size percep-

## The Evolution of Voice Perception

---

tion. However, studies by Charlton, Taylor, and Reby (2013) and Pisanski, Fraccaro, Tigue, O'Connor, and Feinberg (2014) suggest that voice pitch does play a facilitating role in size estimation by providing a dense harmonic spectrum and carrier signal for formants.<sup>4</sup> The fact that harmonics are more densely spaced in men's compared to women's voices may help to explain why size estimates are more accurate with men's than women's voices (see, e.g., Rendall et al. 2007).

As neither fundamental nor formant frequencies account for very much of the variation in body size and shape among same-sex groups of adults (Pisanski, Fraccaro, Tigue, O'Connor, Röder, Andrews, and Fink et al. 2014; Pisanski, Jones, Fink, O'Connor, DeBruine, Röder, and Feinberg 2016), and listeners are only moderately successful in accurately estimating body size from the voice, it appears that vocal indicators of size in humans are weak. Thus, empirical evidence to date suggests that variation in fundamental and formant frequencies is not likely due to strong selection for *honest* vocal indicators of body size and that listeners may not have evolved mechanisms to reliably assess size. Rather, selection may have acted on the human voice in ways that exaggerate actual body size, as appears to be the case in several other mammalian species (Charlton and Reby 2016; Fitch and Hauser 2003). Fundamental and formant frequencies in humans may also function to provide reliable information about other traits that are related to body size and that may be used as proxies of size, such as attractiveness and masculinity or femininity (Pisanski, Mishra, and Rendall 2012; Puts et al. 2016), or, as reviewed in the next section, physical strength.

### (p. 276) Perception of Physical Strength from the Voice

Throughout our evolutionary history, physical strength is likely to have reliably predicted men's ability to accrue resources and gain access to mates, as continues to be the case in many modern populations (Frederick and Haselton 2007; Gallup et al. 2007; Sell, Hone, and Pound 2012). There was probably strong selection pressure on men to advertise their physical strength, potentially through the vocal modality. Strength, although positively related to an individual's body size, is arguably more difficult to assess visually than is size,<sup>5</sup> and predicts fighting ability better than does height or weight (Sell et al. 2009). It follows that listeners may have been selected to rely more heavily on vocal indicators of strength compared to size; however, relatively few studies have examined the vocal communication of strength. A small but growing body of empirical work, reviewed in this section, suggests that the human voice may provide information about physical strength above and beyond information about body size that listeners are able to gauge.

Handgrip strength, upper-body strength, and flexed bicep circumference are common measures of physical strength in empirical studies. Upper-body strength is particularly sexually dimorphic and has been found to predict men's self-reported fighting ability, aggression, and sexual behavior (Gallup et al. 2007; Lassek and Gaulin 2009; Sell et al. 2009; 2010). Sell et al. (2010) showed that listeners could judge men's upper-body strength from the voice alone with above-chance accuracy. Strength estimates were made with voices obtained from four distinct cultures, and were unaffected by language. Al-

## The Evolution of Voice Perception

---

though listeners were able to estimate strength from the voice, the authors were not able to identify which acoustic features were used to make this judgment, as neither fundamental nor formant frequencies predicted the actual physical strength of vocalizers in that study. Recent evidence indicates that listeners can also assess strength from nonverbal vocalizations (i.e., human roars), and that roars may function to maximize impressions of strength compared to speech (Raine, Pisanski, Oleszkiewicz, Simner & Reby, 2018).

Examining potential vocal correlates of strength, Puts, Apicella, and Cardenas (2012) reported a negative relationship between voice pitch and arm strength (a standardized average of hand-grip strength and upper-arm circumference) in a sample of Tanzanian Hadza men. The authors also reported a negative relationship between formant frequencies and arm strength in a sample of Californian men. Hodges-Simeon, Gurven, Puts, and Gaulin (2014) later examined vocal correlates of strength in a peripubertal sample of Tsimane horticulturalists, predicting that vocal indices of size and strength could be more salient in the years spanning puberty (when physical growth and sexual maturation is most rapid) compared to adulthood. In their study, arm strength reliably predicted the fundamental and formant frequencies of males (aged eight to 23) but not of females when controlling for age and body size. Height, adiposity (i.e., fat), and strength together explained most of the variance in men's vocal frequencies (and considerably less of the variance in women's vocal frequencies). Voice pitch predicted variation in men's physical strength above and beyond that which could be explained by body size.

While very few studies have focused directly on vocal communication of physical strength in humans, other work demonstrates that strength-related information (i.e., (p. 277) covariates or proxies of strength) are indeed present in the voice. For example, listeners across many diverse cultures judge individuals with voices of relatively low pitch or low formants as more physically and socially dominant, masculine, and physically larger than individuals with voices of relatively high pitch or formants. In turn, several studies indicate that low pitch or low formants are related to higher facial masculinity, body muscularity, self-reported and other-reported dominance, and levels of circulating testosterone (for reviews, see Pisanski and Feinberg 2013; Puts, Doll, and Hill 2014; Puts et al. 2016). Taken together, the human voice clearly conveys a myriad of useable information related to formidability and threat potential that is likely to have been shaped by sexual selection. The ways in which listeners disentangle or integrate various vocal cues of size, strength, and dominance are less clear.

### Vocal Attractiveness

Researchers have been studying vocal attractiveness for several decades, and the results of this work have broad theoretical as well as practical implications. Studies of voice preferences provide insight into how selection has shaped the human voice, but also inform researchers and the public about perceptual biases, and help to uncover the potential socioeconomic and political implications of vocal stereotyping. For instance, like people with attractive bodies or faces, vocally attractive individuals are often accredited, unduly or not, with positive personality attributes including being perceived as relatively powerful, confident, emotionally stable, intelligent, kind, and socially competent (Kreiman and

## The Evolution of Voice Perception

---

Sidtis 2011). This association between vocal attractiveness and other positive traits has been deemed the vocal attractiveness stereotype (Hughes and Miller 2015; Zuckerman, Hodgins, and Miyake 1990) and stems more broadly from the classical halo effect (Nisbett and Wilson 1977).

Research has focused largely on the relative contributions of fundamental and formant frequencies to voice attractiveness, and for good reason. Voice pitch and formants are sexually dimorphic and reliably predict a host of mate-relevant traits and preferences across mammalian species, including humans, suggesting that these voice features have been most strongly affected by sexual selection. Notably, a growing body of work has begun to highlight the contribution of various other vocal and multimodal traits in the perception of vocal attractiveness, as well as constraints on modulating (i.e., faking) vocal attractiveness.

### Preferences for Sexual Dimorphism in the Voice

The human voice is sexually dimorphic. Men's voice pitch ( $F_0$ ) is on average 120 Hz whereas women's  $F_0$  is almost double that, averaging 210 Hz. Formant frequencies are also typically lower and more closely spaced among men than women (Pisanski, Fraccaro, Tigue, O'Connor, Röder, Andrews, and Fink et al. 2014). What this means, perceptually, is that men's voices sound much "deeper" (lower pitched and more resonant) than do women's. As an example, consider the deep and masculine voice of actor James Earl Jones (the voice of *Star Wars*'s Darth Vader) compared to the much higher (p. 278) and feminine voice of cast member Carrie Fisher (who played Princess Leia). Differential exposure to androgens during puberty in addition to sexual dimorphism in the size of the vocal anatomy can account for some of the sexual dimorphism in fundamental and formant frequencies;<sup>6</sup> however, the sexual dimorphism in  $F_0$  or voice pitch far surpasses what we might expect given the differences in body size and vocal fold mass and length between men and women.

As a general rule, men and women are attracted to sexual dimorphism in opposite-sex voices. Cross-culturally,<sup>7</sup> women prefer lower pitch and formants compared to average or higher pitch and formants in men's voices (i.e., masculine voices), whereas men often (but not always) prefer higher pitch and formants compared to average or lower pitch and formants in women's voices (i.e., feminine voices; for reviews, see Feinberg 2008; Pisanski and Feinberg 2013).<sup>8</sup> Preferences for sexual dimorphism in opposite-sex voices appear to develop during adolescence around the age at which mate preferences become relevant (Saxton, DeBruine, Jones, Little, and Roberts 2009) and are typically weaker within sexes than between sexes (Babel, McGuire, and King 2014; Jones et al. 2010; Pisanski and Rendall 2011).

Men's preferences for femininity in women's voices and women's preferences for masculinity in men's voices likely evolved under sexual selection as a means of identifying high-quality mates. Higher voice pitch in a woman's voice may be a fairly good indicator that she has relatively high levels of estrogens (Abitbol et al. 1999), which typically indi-

## The Evolution of Voice Perception

---

cate fecundity and therefore reproductive value or fitness (Sherman and Korenman 1975; Venners et al. 2006). Women's voice pitch appears to increase during ovulation (Bryant and Haselton 2009) in conjunction with cyclical changes in estradiol and progesterone levels (Puts et al. 2013), and women's voices are judged by men as most attractive around the time of ovulation (Pipitone and Gallup Jr. 2008) and least attractive around the time of menstruation (Pipitone and Gallup 2012; see also Haselton and Gildersleeve 2011). In addition to often (but not always) being perceived as attractive, high voice pitch in women is closely associated with perceptions of femininity (Feinberg, DeBruine, Jones, and Perrett, 2008; Pisanski and Rendall 2011), youthfulness (Collins and Missing 2003), flirtatiousness (Puts et al. 2011), and sexual interest (Jones et al. 2008). Women's voice pitch can also predict the number of sexual partners they report having had (Hughes, Dispenza, and Gallup Jr. 2004) It should be noted that men sometimes show a preference for relatively lower pitch in women's voices, possibly because low pitch can communicate intimacy, maturity, or confidence.

Among men, relatively low voice pitch or formants generally indicate that a man has higher levels of circulating testosterone compared to a man with a higher-frequency voice (Bruckert et al. 2006; Cartei, Bond, and Reby 2014; Dabbs and Mallinger 1999; Evans et al. 2008). Higher levels of testosterone in men are in turn positively associated with a host of mate-relevant characteristics, including dominance, physical strength and body size, and immune responsiveness (Puts, Doll, and Hill 2014; Rantala et al. 2012; Skrinda et al. 2014). Of course, androgen-dependent physical traits in men such as strength and size are also relevant in male-male competition, and thus men's voice pitch and formants have likely evolved both under intersexual selection (mate choice) and intrasexual selection (competition between same-sex individuals) (Puts, Doll, and Hill 2014).<sup>9</sup> Like women, men's voice pitch predicts reproductive success, although in the case of men the relationship is negative (Apicella, Feinberg, and Marlowe 2007; Puts, Gaulin, and Verdolini 2006).

At the same time, high levels of testosterone in men have been linked to higher levels of infidelity, divorce, aggression, and lower levels of parental and resource investment (Booth and Dabbs 1993; Eisenegger, Haushofer, and Fehr 2011; Mazur and Booth 1998), and women judge men with lower voice pitch as more likely to cheat and less likely to invest in them and their offspring compared to men with higher voice pitch (O'Connor, Re, and Feinberg 2011; O'Connor 2012). As a consequence, high levels of testosterone and low voice pitch or formants in a man's voice may present a trade-off in the context of mate choice for women, wherein women who choose such masculine men as mates may benefit in some ways, but also pay a cost in other ways.

Many researchers have predicted that women's preferences for masculinity in men's traits and behaviors, including men's voices (e.g., low voice pitch or formants), will vary so as to maximize the benefits and minimize the costs of choosing a masculine mate. There is a good deal of empirical evidence to support this prediction. Women's preferences for relatively low voice pitch or low formants are stronger when women judge the vocal attractiveness of a hypothetical short-term versus long-term relationship partner (Feinberg et al. 2012; Puts 2005), particularly among women who attribute low trustwor-

## The Evolution of Voice Perception

---

thiness and dominance to masculine men (Vukovic et al. 2011). Women with high-pitched voices (Vukovic et al. 2010) and those who rate themselves high on attractiveness (Feinberg et al. 2012; Vukovic et al. 2008) also show relatively stronger preferences for low voice pitch than do other women. Multiple studies examining women's preferences for facial, body, and vocal masculinity further show evidence for cyclic shifts, namely, stronger masculinity preferences during the most fertile phase of the menstrual cycle when ovulation risk is highest (for meta-analysis see Gildersleeve, Haselton, and Fales 2014; but see also Wood, Kressel, Joshi, and Louie 2014).<sup>10</sup> These results support the prediction that masculinity preferences may function to increase offspring health (Gangestad and Thornhill 2008; Jones et al. 2013).

In addition to cyclic variation, women's vocal masculinity preferences have been linked directly to changes in women's hormone levels (Feinberg et al. 2006; Pisanski, Hahn, Fisher, DeBruine, Feinberg, and Jones 2014; Puts 2006). Women's vocal masculinity preferences may indeed serve to protect women and their offspring from exposure to pathogens, as masculine men may have stronger immune systems compared to feminine men (Rantala et al. 2012; Skrinda et al. 2014). Studies confirm that higher levels of pathogen disgust sensitivity predict women's preferences for masculinity in men's faces, bodies, and voices (Jones et al. 2013). Women's facial masculinity preferences also correlate positively with population-level pathogen prevalence across nations (DeBruine et al. 2010).<sup>11</sup>

Although vocal femininity is attractive in women's voices and vocal masculinity is attractive in men's voices, it should be noted that extreme sexual dimorphism is unlikely to be attractive. Re et al. (2012) and Saxton, Mackey, McCarty, and Neave (2015) found that women did not prefer relatively lower-pitched men's voices at the extreme lower end of the spectrum (i.e., below 96 Hz). Re et al. noted that extreme low pitch can be (p. 280) indicative of vocal pathology and may therefore be perceived as unattractive. As discussed earlier, low voice pitch among men is also associated with various negative attributes that may be particularly salient in extremely low-pitched voices, such as aggression and infidelity. Although Re et al. (2012) found that men preferred relatively higher-pitched women's voices above the normal range of female voice pitch (i.e., up to 300 Hz, the highest frequency tested), Borkowska and Pawlowski (2011) found that men did not prefer relatively higher pitch in women's voices above a 280 Hz threshold. Men may not prefer voice pitches that fall into the range of adolescent voice pitch (i.e., above 300 Hz), because such voices indicate sexual immaturity. Several studies show that listeners also associate extremely high voice pitch with behavioral immaturity, babyishness, submissiveness, and incompetence (reviewed in Kreiman and Sidtis 2011).

## Other Factors Affecting Vocal Attractiveness

We have thus far highlighted studies of voice attractiveness that have focused largely on sexually dimorphic features of the voice—fundamental and formant frequencies (pitch and formants)—that reliably indicate many evolutionarily relevant traits and behaviors. This large literature provides compelling evidence that pitch and formants have under-

## The Evolution of Voice Perception

---

gone sexual selection and play a meaningful role in mate choice and mate competition across mammalian species, including humans. In addition, judgments of vocal attractiveness hold relevance outside of a sexual context. Vocal attractiveness can, for example, affect perceptions of a vocalizer's personality (Zuckerman, Hodgins, and Miyake 1990), competitiveness for a job opening (Anderson et al. 2014), and political leadership capacity (Anderson and Klofstad 2012; Tigue, Borak, O'Connor, Schandl, and Feinberg 2012). Recent studies have begun to explore vocal attractiveness in a broader social context and suggest that multiple features of the voice play a role.

Babel et al. (2014) examined the relative contributions of harmonics-to-noise ratio, spectral tilt, jitter and shimmer, and speech duration in addition to fundamental and formant frequencies on attractiveness judgments of men and women's voices. Harmonics-to-noise ratio, spectral tilt, and jitter and shimmer are measures of vocal quality that affect perceptions of breathiness, creakiness, and smoothness in the voice and that at some level can be suggestive of vocal pathology, illness, smoking, or alcoholism (Kreiman and Sidtis 2011). The researchers found that breathiness was an attractive quality of women's voices, and one that may be associated with intimacy, whereas male voices with shorter durations of speech were judged as relatively more attractive than those with longer durations. Because men typically speak with shorter durations than do women (Simpson 2009), the authors suggested that this latter finding may in fact represent a preference for sexually dimorphic patterns in opposite sex voices.

Research indicates that men and women with attractive voices are also likely to have attractive faces (Abend et al. 2015; Collins and Missing 2003; Feinberg, DeBruine, Jones, and Little, 2008; Hughes and Miller 2015; Little et al. 2011; O'Connor et al. 2013; Puts et al. 2013; Saxton, Burriss, Murray, Rowland, and Roberts 2009; Skrinda et al. 2014; (p. 281) Wheatley et al. 2014). This finding suggests that faces and voices may develop via similar mechanisms, and therefore provide similar information about an individual's mate quality (Feinberg 2008). It also hints at the possibility that information from one modality may interact with information gathered from another modality such that perceptions of vocal attractiveness may differ in the presence or absence of visual information from the face or body. Indeed, using videos to examine the interaction between vocal and facial indicators of attractiveness, O'Connor et al. (2013) found that men's judgments of women's vocal attractiveness were higher when voices were presented in conjunction with a relatively more feminine face. Men's judgments of women's facial attractiveness were also amplified by the addition of a feminine, high-pitched voice.<sup>12</sup> Little et al. (2013) showed that adaptation to sex-typicality or atypicality in voices affected perceptions of normalcy in faces, and vice versa. Although not yet tested, olfactory information is also likely to amplify or dampen judgments of vocal or facial attractiveness. Given that multiple modalities interact in everyday perception (for review see Spence 2011), multimodal signaling of attractiveness constitutes an important avenue for future work.

Future studies may further investigate the role of voice modulation in vocal attractiveness. There are inherent anatomical constraints on vocal production, such as vocal fold mass and vocal tract length, but manipulation of voice pitch and formants is possible

## The Evolution of Voice Perception

---

within a limited range (see, e.g., Cartei, Cowles, and Reby 2012; Hughes, Mogilski, and Harrison 2014; Puts, Gaulin, and Verdolini 2006). Several studies report that men and women modulate their pitch and other aspects of their voices in response to attractive conversational partners (Anolli and Ciceri 2002; Fraccaro et al. 2011; Hughes, Farley, and Rhodes 2010; Leongómez et al. 2014; Pisanski, Oleszkiewicz, Plachetka, Gmiterek & Reby, 2018); however, evidence as to whether vocal modulation effectively alters perceptions of the vocalizer's attractiveness remains equivocal (Anolli and Ciceri 2002; Fraccaro et al. 2011; Hughes, Farley, and Rhodes 2010; Leongómez et al. 2014; Pisanski et al., 2018).

## Vocal Communication of Affect and Intention

While some aspects of the voice are unavoidably informative, such as formant information revealing body size or honest indicators of one's hormonal profile such as fundamental frequency signaling mate quality, human vocal communication involves many complexities of voice control that interact with language and sociality. One major aspect of communicative behavior in social species, including humans, is the signaling of intention. We define intentional action as planned, goal-directed behavior to achieve a desired future state of affairs. A dog that bares his teeth at another dog is signaling his intention to bite and providing information regarding his investment in something usually obvious in the environment, such as a food source or protection of kin. The baring teeth signal likely evolved from the nonsignaling behavior of moving the lip out of the way of the teeth to prevent injury. The cue of lip movement prior to a bite is predictive and through a ritualization process it may have evolved into a signal, (p. 282) complete with the exaggerated features characteristic of acts shaped in contexts with a substantial conflict of interest between signalers (Krebs and Dawkins 1984).

Whenever there is a signal of one's intention, the prospect of deception exists, and a receiver's judgment of the honesty of a given signal must incorporate this possibility. Many theoretical models have been proposed to explain the evolutionary dynamics in the maintenance of signal reliability, and some things are clear. First, signals may not need to be costly to be reliable (Lachmann, Szamado, and Bergstrom 2001), but cost is one way to ensure reliability (Grafen 1990). Second, there are circumstances where some degree of deception can be evolutionarily stable if the benefits of repeated interaction are high enough (Johnstone and Grafen 1993). These aspects of animal signaling have profound implications for our understanding of human vocal behavior, especially as it relates to the role of different vocal systems that play a major part in our vocal repertoire.

Mammalian vocal production is implemented in relatively simple neural and motor circuits that have been conserved across many species (Ackermann, Hage, and Ziegler 2014; Jurgens 2002). In humans, vocal behaviors such as crying, laughter, fear screams, and copulation calls are driven by this emotional vocal system. Our perception of these vocal signals has been shaped by millions of years of evolution where the vocal production machinery is consistent and predictable. Yet humans, unlike any other species, have developed the capacity for language and speech, which has resulted in the development

## The Evolution of Voice Perception

---

of specialized production mechanisms involving fine-grained control over vocal articulators, breathing, and laryngeal musculature (Ghazanfar and Rendall 2008). The volitional control over our vocal speech production, which interfaces in complicated and poorly understood ways with other cognitive systems involved with language in a broad sense, allows us to now produce cries, screams, roars, laughs, and other sounds without the necessary emotional triggers once required of our hominin ancestors, or our primate relatives today.

Researchers examining laughter have explored the implications of the dual pathway model of vocal production in the perception of different kinds of laughs. Informed by work on smiling (e.g., Ekman, Davidson, and Friesen 1990), theorists have proposed that laughter can come in at least two forms that correspond to spontaneous versus volitional control. Different names have been used for these two laugh types, such as voluntary versus involuntary (Ruch and Ekman 2001) and Duchenne versus non-Duchenne (Gervais and Wilson 2005). Bryant and Aktipis (2014) found that judges can distinguish between these laugh types, and there are predictable acoustic differences between them as well, mostly attributable to the role of arousal in the emotional triggering of spontaneous laughter, but also due to the differential role of breath control. McGettigan et al. (2014) have also shown that different brain areas are activated both when producing different types of laughter and when listening to them. Prefrontal areas associated with deliberate control of behavior are implicated in these laughter production mechanisms, as well as the neural system underlying the detection of others' mental states. Laughter is a prime example of how a phylogenetically ancient behavior has been shaped by recent selection in the context of language communication, and now operates at multiple levels in complex human social interaction.

### (p. 283) Emotional Vocal Signals

Intentions can be signaled in any number of ways, including, quite predominantly, through a vocal channel. Researchers examining vocal behavior in humans and nonhumans have often focused on the communication of emotion. Here we define emotions as cognitive programs designed to motivate context-specific, adaptive behavior. Emotional programs include, as part of their design, communicative mechanisms in all modalities, but vocal control is intricately connected to central nervous system structures through the vagus nerve and voice parameters are profoundly affected by changes in emotional motivations (Porges 2001).

Consider a fear scream: in the context of danger, say for example the presence of a predator, emotional systems in the body prepare for rapid action. The body needs glucose for energy as well as increased oxygen intake and the emotion of fear is the program that coordinates the body's needs in this setting. Vocalizations produced under these conditions will often have a sound of urgency. For example, the calls may often be extremely loud, may contain nonlinear acoustic features associated with overblowing the vocal tract constraints (Fitch, Neubauer, and Herzog 2002), and may be characterized by raised pitch and faster temporal properties that together will reveal distress (Banse and Scherer

## The Evolution of Voice Perception

---

1996). The acoustic correlates of fearful vocalizations have a structural form that facilitates the communicative function of signaling urgency and danger (Owren and Rendall 2001). In the nonhuman animal literature, debates continue as to what animal signals mean and whether the notion of functional reference is necessary in our understanding of these signals (e.g., Rendall, Owren, and Ryan 2009). Among humans, however, we can say without a doubt that form-function relationships coexist with linguistic aspects of vocal production that are serving varying functions. These vocal channels co-occur for listeners, and audience members must distinguish between vocal emotional information and other prosodic features helping serve linguistic functions. In our example of a fear vocalization, which can play a role in behaviors such as alarm calls, pain shrieks, and threat displays, we can see how the acoustic form of the call relates to its signaling function. Emotional vocalizations can all be analyzed in this way, and the regularities predict and explain a good deal of universality in how people express themselves vocally. We see the same types of emotional sounds in music across cultures (Juslin and Laukka 2003) as well as in the ways mothers speak to babies (often referred to as infant-directed speech or motherese; Broesch and Bryant 2015; Bryant and Barrett 2007; Fernald 1992).

There has been a long tradition of considering emotional expressions as reflecting internal states that audience members can infer, essentially treating emotional expressions as cues (e.g., Ekman 1997), but this approach makes little sense evolutionarily. Animals, including humans, may pay a large price for giving away internal information for free, as this could provide a means of manipulation for receivers without any benefit for the sender. Signals only evolve in contexts where senders benefit (and typically receivers too; Maynard Smith and Harper 2003), and there is no reason to believe emotional expressions are an exception. Thus, we should consider emotional expressions to be (p. 284) signals, and to be produced strategically with adaptive benefits for senders. One quick point of diversion about the term “strategic.” By this we do not mean volitional, deliberate, or necessarily conscious. Instead, we mean that the behavior is due to selection on the production mechanism that is shaped by a particular and typically beneficial outcome. Vocal emotions are designed to influence listeners in a specific way: they may often be associated with a particular subjective phenomenology; they may be under some conscious control or not; and they are probabilistically suboptimal in any particular circumstance. The strategy of the production of vocal emotions is built into their design.

Voice pitch plays a fundamental role in the sound profiles of many vocal emotional signals. As explained earlier, pitch is the perceptual correlate of  $F_0$  and is determined primarily by the vibration rate of the vocal folds in the larynx (Titze 1994). This vibration rate is determined by both subglottal air pressure as well as laryngeal muscle activity. Because of the inherent relationship between emotional programs and physical arousal in an organism, structural correlates of arousal in communicative signals play a crucial role in their effectiveness. The positivity or negativity of emotions (i.e., valence) is another important dimension that is also signaled through physical features of emotional signals including the acoustic properties of vocalizations. Together, arousal and valence constitute

## The Evolution of Voice Perception

---

a dimensional model of emotion (Russell 1980) that helps researchers make sense of various parameters of different emotional expressions, including vocal signals.

As described earlier, the emotion of fear is designed to motivate a rapid response in an animal (e.g., fight or flight). Emotional programs organize bodily systems in preparation for action. Specifically, fear activates energetic stores making glucose readily available for muscle systems. Respiration capacity is increased, perceptual systems are sharpened, and the animal enters a state that facilitates adaptive fleeing or fighting activity. This suite of preparedness has consequences for physical signals such as vocal expressions. Increased musculature tension results in increased  $F_0$  as well as greater amplitude. Scared animals tend to scream more loudly and faster than they would typically vocalize, with additional nonlinear features (e.g., deterministic chaos, subharmonics) often present in the scream call as a result of excessive airflow pushing through the vocal tract (Fitch et al. 2002). The resulting sound is highly detectable in noisy environments for warning conspecifics of a common threat, such as a predator, but also serves as a warning to enemies for any number of reasons. Highly aroused animals are dangerous, and fear-related vocal signals can help communicate this danger to potential rivals.

In the case of fear, we should expect perceivers to be highly sensitive to relevant acoustic features that can help them predict something important about the environmental context. These sound features end up doing most of the communicative work in rich contexts—a perfect example of the “form follows function” principle in biology (Rendall and Owren 2001). Human sensitivity to fear vocalizations is clearly homologous to other mammals in this regard, and as such, can even affect judgments of very human-specific phenomena such as music. Blumstein et al. (2012) found that acoustic nonlinearities such as deterministic chaos (simulated through music distortion) in musical compositions was associated with greater arousal and negative valence in listeners relative to (p. 285) control versions of the compositions without these nonlinear features. This research was motivated by comparative work showing the incredible similarity in the sound profiles of different species and emotion categories revealing the important role that evolutionary processes have played in shaping human and nonhuman emotion signaling systems (Briefer 2012). Cultural evolutionary processes can result in sound features such as distortion in music being utilized as a compositional tool because it exploits our evolved sensitivity that we share with many mammalian species (Bryant 2013).

## Cross-Cultural Perception of Emotion

There is overwhelming evidence showing that people across diverse cultures can accurately detect emotional signals in faces, and to a lesser extent, voices (Elfenbein and Ambady 2002). People are quite good at identifying vocal emotions, but as in faces, research reveals a within-culture advantage. While some work has examined the cross-cultural perception of emotions in voices in western, and generally industrialized, nations (e.g., Pell et al. 2009; Scherer, Banse, and Wallbott 2001; Thompson and Balkwill 2006), only recently have researchers studied emotional vocal perception in highly disparate, traditional societies. Bryant and Barrett (2008) found that Shuar hunter-horticulturalists could reliably

## The Evolution of Voice Perception

---

identify basic vocal emotions in spoken English sentences. In this experiment, participants listened to spoken sentences that were produced while the speaker was looking at an emotional facial expression, and participants then judged which of two facial expressions the speaker was looking at during the recording of the sentence. Accuracy across different emotion categories varied, with recognition of happiness highest (71 percent) and recognition of disgust at chance. English speaking participants in the United States judged the same English sentences in a content-filtered condition. Results were fairly similar overall, though with some interesting cultural differences across emotional categories. For example, Shuar listeners often judged happy vocalizations as fearful, but did not judge fearful vocalizations as sounding happy. If acoustic similarities between happiness and fear were responsible for these errors, we should expect symmetry in the error pattern. Instead, the apparent bias might be due to different pragmatic display rules across cultures. In Shuar culture, expressions of happiness between strangers might reveal that a speaker is scared or nervous relatively more than in the United States. Within-culture advantages in detecting emotional expressions are often likely due to culture-specific belief patterns that will drive certain biases in listeners' judgments.

Sauter, Eisner, Ekman, and Scott (2010) found that the Himba—a seminomadic group of pastoralists in northern Namibia—were able to reliably recognize nonverbal vocal emotions both within and between cultures. English listeners were also able to recognize vocal emotions produced by Himba speakers. The researchers used a task in which short, read vignettes were followed by pairs of vocal recordings, and listeners reported which vocalization was appropriately tied to the story. Results showed bi-directional recognition for a subset of the emotion categories, most considered “basic” (p. 286) emotions (i.e., anger, disgust, fear, sadness, surprise, and amusement). However, the Himba participants did not choose the appropriate vocalization for most of the positive emotions, such as triumph and sensual pleasure, even though these emotions were recognized within-culture. The authors suggested that affiliative social signals might be subject to culturally specific display rules, and are therefore not as easily transmitted across cultures as are other types of social signals.

More recently, Gendron et al. (2014), using the same task in the same population, found a somewhat different pattern of results that they interpreted as evidence against the cultural universality of emotional expressions. The authors found that when participants judged valence-matched vocal alternatives after hearing a vignette, they did not correctly choose the intended emotion, but when the alternatives were arousal-matched (and therefore differed only in valence), participants correctly choose the intended category better than chance. Gendron et al. suggested possible universality in sensitivity to valence, but not in the recognition of specific emotions. Sauter, Eisner, Ekman, and Scott (2015) reanalyzed their own original data (from Sauter, Eisner, Ekman, and Scott 2010), focusing only on emotion categories in which judges were better than chance by removing some positive emotion trials from their analyses. The pattern of results obtained in both studies was similar regardless whether the distractor matched the target in valence or not. Moreover, the authors pointed out that Gendron et al.'s judges failed to answer correctly in a condition where the emotion alternatives were different in arousal and valence, which affords

## The Evolution of Voice Perception

---

discrimination based on valence, and additionally suggested that Gendron et al.'s participants might not have fully understood the stories.<sup>13</sup>

The debate described here is important because it demonstrates particular difficulties in conducting cross-cultural research. At what point does the explanation of a research procedure teach participants something new about the phenomenon that they otherwise would not know, and does subsequent success in such tasks reveal underlying similarity or acquired knowledge? There is little question that to gather meaningful data in the field, researchers need to ensure participants properly understand the study task. These studies with the Himba provide a great illustration of how tasks requiring subjective interpretation are likely to elicit high variation, and are especially vulnerable to experimenter demand effects and other potential biases.

## Recognizing Speaker Intention from the Voice

Much of what people intend to communicate by their speech acts is not contained in the language itself, but must be inferred by listeners based on multiple sources of information (Sperber and Wilson 1995). One central source of information is the nonverbal elements of the voice, and researchers across a variety of disciplines (e.g., psychology, linguistics, computer science, neuroscience, and biology) have explored the ways in which vocal signals are processed by listeners in communicative interactions. Humans seem to be unique in our particular sensitivity to recognizing ostensive intentions to communicate (Scott-Phillips 2014). That is, we can recognize informative intentions (p. 287) (i.e., the content of what one intends to convey), as well as communicative intentions (i.e., the intention to convey anything at all). From a listener's perspective, the perception of intention in the voice requires the separation of vocal signals of emotion from related linguistic and pragmatic prosodic signals (i.e., pitch, loudness, rhythm, and spectral information). It is not currently understood how all of these sources of vocal information interact to convey intention.

Prosodic signals in the voice are intricately tied to the affective and intentional goals of the speaker (Bryant and Fox Tree 2002; Cosmides 1983). Prosody assists listeners in making many distinctions, ranging from focus in lexical items (e.g., using pitch to emphasize meaning), to syntactic disambiguation, all the way up to discourse structure and conversational turns (for a review see Cutler, Dahan, and van Donselaar 1997; Kreiman and Sidtis 2011). General attitudinal information is recognizable fairly quickly in the voice, even without words or pragmatic context. For example, Swerts and Hirschberg (2010) found that listeners could detect whether upcoming speech content would be positive or negative before the linguistic information made it clear. The perception of upcoming "bad" news was most likely driven by a restricted pitch range and a fast speech rate. Speakers use a variety of vocal signaling strategies during discourse, and these vocal patterns often work independently of contextual and linguistic features.

Research on indirect speech such as verbal irony has explored the role of prosody in communicating intention, and many theorists adhere to the notion that voice properties

## The Evolution of Voice Perception

---

uniquely convey types of verbal irony such as sarcasm. Much of the research in this area has relied on vocal recordings of actors (e.g., Anolli, Ciceri, and Infantino 2000; Cheang and Pell 2008). These studies have revealed that speakers tend to lower their voice pitch and produce noisier vocalizations when speaking sarcastically relative to typical speech. Analyses of spontaneous verbal irony are not nearly so consistent, and instead suggest that speakers are highly variable when producing ironic speech (Bryant 2010; Bryant and Fox Tree 2002). Judges also tend to conflate irony with many other affective and intentional categories such as anger, inquisitiveness, and authority even when listening to content-filtered speech in which the words are not available (Bryant and Fox Tree 2005).

Imagine listening to a conversation through a wall. The sound is muffled and words are difficult to understand. The wall essentially acts as a low-pass filter, meaning that frequencies above a certain threshold important for speech are removed. Nevertheless, listeners are typically able to gather a good deal of information about attitudes and intentions from a highly impoverished speech signal that contains little more than  $F_0$ . Theorists have long noticed that many similarities exist across languages and cultures in how speakers use nonverbal voice information, such as pitch, to convey different kinds of meaning (e.g., Ohala 1984). An important use of pitch in everyday speech involves changes at specific moments. These prosodic contrasts play a communicative role in a variety of signals, including in verbal irony (Bryant 2010), conversational turn-taking (Cutler and Pearson 1986), and even nonhuman animal communication (Blumstein et al. 2008; Morton 1977). The nature of detecting affect from prosodic information is a dynamic phenomenon in which moment-to-moment processing can alter many aspects (p. 288) of judgments (Roche, Peters, and Dale 2015). By using new techniques to monitor decision-making processes, such as mouse-tracking and eye movements, researchers can explore the complex dynamics of how listeners interpret prosodic contrasts and other aspects of human's communicative repertoire that are in motion.

Few studies have examined how people can recognize intent across different cultures, but work on infant-directed speech has shown that people are quite good at detecting intentions, especially in speech designed to make intentional information salient to listeners without access to other sources of linguistic information such as words and syntax (Bryant and Barrett 2007; Bryant et al. 2012). In these cross-cultural studies of infant-directed speech, listeners were presented with pairs of recordings of mothers acting out different intentions (e.g., prohibitives, approvals) in a language they do not speak, and then identified, in a forced-choice task, what the correct intention was in the vocalization. Accuracy in this task was well above chance, and listeners could identify whether speakers were intending to talk to a baby or another adult. Participants in these studies were also able to identify intentional information in adult-directed speech, a mode of speech that contains fewer nonverbal signals of intention than infant-directed speech. These results can be interpreted as evidence that intentional information in the voice contains some universal properties, and while we believe this is certainly true to a great extent, important variation also exists in people's ability to judge different categories of spoken language.

# Conclusions

An evolutionary approach to voice perception focuses not only on mechanistic descriptions of vocal communication, but also on phylogenetic and functional explanations of why animals (including humans) attend to and process particular features of the vocal signal. When considering underlying adaptations in both the production and perception side of communicative interactions, it becomes necessary to consider the meaningful distinction between signals and cues. As we have described, many aspects of the voice may serve as cues to various features of the vocalizing animal (e.g., body size), while other aspects of the voice have been shaped by selection to affect or manipulate target listeners in particular ways (e.g., emotional signals). Attending to vocal cues can be highly informative and beneficial; hence, listeners have perceptual specializations that guide this adaptive behavior.

As the study of voice production and perception continues to grow in popularity and to unite seemingly disparate disciplines, some recent developing trends will help the field continue to move forward. In particular, improvements in voice analysis technology have facilitated greater numbers of researchers to use these tools. Perhaps more critically, improvements in the past decade in broadband Internet services and information storage have made large-scale cross-cultural studies much easier to implement (e.g., Bryant et al. 2016). It is now possible to readily and effectively develop and maintain international collaborations that allow researchers worldwide to share stimuli and data (p. 289) efficiently online, including via open-source databases. As in most areas in the behavioral sciences, voice researchers stand to benefit tremendously by implementing cross-cultural analyses. Most research on human behavior, and a good deal of what we think we understand, is rooted in research in industrialized societies on 18- to 23-year-old college students. Henrich, Heine, and Norenzayan (2010) deemed the typical participants in most behavioral research WEIRD (i.e., Western, Educated, Industrialized, Rich, Democratic). These authors make a good case for the notion that WEIRD subjects fall on the extreme end of the spectrum for many measured traits long considered universal, highlighting that more representative participants from around the globe will provide a better understanding of psychological phenomena. Investigations involving traditional, sustenance-based, small-scale societies reveal complexities in cultural variability previously downplayed or simply unnoticed. Voice researchers stand to gain a great deal from conducting their experiments across disparate cultures—the need is there, and the technology is available.

In this chapter we have attempted to present the latest empirical research on the evolution of voice perception, emphasizing how comparative work on nonhuman animals incorporating evolutionary principles can inform efforts to properly characterize human vocal communication, and specifically voice perception. Of course, much work remains. While research exploiting vocal production and processing is currently being conducted at each level of analysis, from low-level physiology to high-level abstract perception, we have a particular interest in examining how vocal communication manifests in naturally occurring contexts. To properly understand the nature of vocal communication, researchers should make every attempt to measure vocal communication outside of the lab (e.g.,

## The Evolution of Voice Perception

---

Pisanski et al., 2018) and with the advent of new technologies for recording and data storage, great strides could be made in the upcoming years. It is now possible to get accurate hormone measures, high-quality voice recordings and images, GPS-assisted location data, time-stamped contextual information, and many other kinds of measures for massive multivariate analyses in which complicated interactions between various sources of influence on vocal production and perception can be examined together. The future is bright for voice research in the evolutionary behavioral sciences.

### Works Cited

Abend, P., Pflüger, L. S., Koppensteiner, M., Coquerelle, M., and Grammer, K. 2015. "The Sound of Female Shape: A Redundant Signal of Vocal and Facial Attractiveness." *Evolution and Human Behavior* 36 (3): 174–81.

Abitbol, J., Abitbol, P., and Abitbol, B. 1999. "Sex Hormones and the Female Voice." *Journal of Voice* 13 (3): 424–46.

(p. 291) Ackermann, H., Hage, S. R., and Ziegler, W. 2014. "Brain Mechanisms of Acoustic Communication in Humans and Nonhuman Primates: An Evolutionary Perspective." *Behavioral and Brain Sciences* 37 (6): 529–46.

Anderson, R. C., and Klofstad, C. A. 2012. "Preference for Leaders with Masculine Voices Holds in the Case of Feminine Leadership Roles." *PloS one* 7 (12): e51216.

Anderson, R. C., Klofstad, C. A., Mayew, W. J., and Venkatachalam, M. 2014. "Vocal Fry May Undermine the Success of Young Women in the Labor Market." *PloS one* 9 (5): e97506.

Anolli, L., and Ciceri, R. 2002. "Analysis of the Vocal Profiles of Male Seduction: From Exhibition to Self-Disclosure." *Journal of General Psychology* 129 (2): 149–69.

Anolli, L., Ciceri, R., and Infantino, M. G. 2000. "Irony as a Game of Implicitness: Acoustic Profiles of Ironic Communication." *Journal of Psycholinguistic Research* 29 (3): 275–311.

Apicella, C. L., Feinberg, D. R., and Marlowe, F. W. 2007. "Voice Pitch Predicts Reproductive Success in Male Hunter-Gatherers." *Biology Letters* 3 (6): 682–4.

Babel, M., McGuire, G., and King, J. 2014. "Towards a More Nuanced View of Vocal Attractiveness." *PloS one* 9 (2): e88616.

Banse, R., and Scherer, K. R. (1996). "Acoustic Profiles in Vocal Emotion Expression." *Journal of Personality and Social Psychology* 70 (3): 614–36.

Barkow, J. H., Cosmides, L., and Tooby, J., eds. 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Belin, P., Fecteau, S., and Bedard, C. 2004. "Thinking the Voice: Neural Correlates of Voice Perception." *Trends in Cognitive Sciences* 8 (3): 129–35.

## The Evolution of Voice Perception

---

- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. 2000. "Voice-Selective Areas in Human Auditory Cortex." *Nature* 403 (6767): 309–12.
- Blumstein, D. T., Bryant, G. A., & Kaye, P. 2012. "The sound of arousal in music is context-dependent." *Biology Letters*, 8(5), 744–47.
- Blumstein, D. T., Richardson, D. T., Cooley, L., Winternitz, J., and Daniel, J. C. 2008. "The Structure, Meaning and Function of Yellow-bellied Marmot Pup Screams." *Animal Behaviour* 76 (3): 1055–64.
- Booth, A., and Dabbs, J. M. 1993. "Testosterone and Men's Marriages." *Social Forces* 72 (2): 463–77.
- Borkowska, B., and Pawlowski, B. 2011. "Female Voice Frequency in the Context of Dominance and Attractiveness Perception." *Animal Behaviour* 82 (1): 55–9.
- Briefer, E. F. 2012. "Vocal Expression of Emotions in Mammals: Mechanisms of Production and Evidence." *Journal of Zoology* 288 (1): 1–20.
- Brosch, T., and Bryant, G. A. 2015. "Prosody in Infant-Directed Speech Is Similar Across Western and Traditional Cultures." *Journal of Cognition and Development* 16 (1): 31–43.
- Bruckert, L., Liénard, J. S., Lacroix, A., Kreutzer, M., and Leboucher, G. 2006. "Women Use Voice Parameters to Assess Men's Characteristics." *Proceedings of the Royal Society Biological Sciences Series B* 273 (1582): 83–9.
- Bryant, G. A. 2010. "Prosodic Contrasts in Ironic Speech." *Discourse Processes* 47 (7): 545–66.
- Bryant, G. A. 2013. "Animal Signals and Emotion in Music: Coordinating Affect Across Groups." *Frontiers in Psychology* 4 (990): 1–13.
- Bryant, G. A., and Aktipis, C. A. 2014. "The Animal Nature of Spontaneous Human Laughter." *Evolution and Human Behavior* 35 (4): 327–35.
- Bryant, G. A., and Barrett, H. C. 2007. "Recognizing Intentions in Infant-Directed Speech: Evidence for Universals." *Psychological Science* 18 (8): 746–51.
- (p. 292)** Bryant, G. A., and Barrett, H. C. 2008. "Vocal Emotion Recognition Across Disparate Cultures." *Journal of Cognition and Culture* 8 (1–2): 135–48.
- Bryant, G. A., Fessler, D. M. T., Fusaroli, R., Clint, E., Aorøe, E., Apicella, C., et al. 2016. "Detecting Affiliation in Colaughter Across 24 Societies." *Proceedings of the National Academy of Sciences* 113 (17): 4682–7.
- Bryant, G. A., and Fox Tree, J. E. 2002. "Recognizing Verbal Irony in Spontaneous Speech." *Metaphor and Symbol* 17 (2): 99–117.

## The Evolution of Voice Perception

---

- Bryant, G. A., and Fox Tree, J. E. 2005. "Is There an Ironic Tone of Voice?" *Language and Speech* 48 (3): 257-77.
- Bryant, G. A., and Haselton, M. G. 2009. "Vocal Cues of Ovulation in Human Females." *Biology Letters* 5 (1): 12-15.
- Bryant, G. A., Liénard, P., and Barrett, H. C. 2012. "Recognizing Infant-Directed Speech Across Distant Cultures: Evidence from Africa." *Journal of Evolutionary Psychology* 10 (2): 147-59.
- Campanella, S., and Belin, P. 2007. "Integrating Face and Voice in Person Perception." *Trends in Cognitive Sciences* 11 (12): 535-43.
- Cartei, V., Bond, R., and Reby, D. 2014. "What Makes a Voice Masculine: Physiological and Acoustical Correlates of Women's Ratings of Men's Vocal Masculinity." *Hormones and Behavior* 66 (4): 569-76.
- Cartei, V., Cowles, H. W., and Reby, D. 2012. "Spontaneous Voice Imitation Abilities in Adult Speakers." *PloS one* 7 (2): e31353.
- Charlton, B. D., and Reby, D. 2016. "The Evolution of Acoustic Size Exaggeration in Terrestrial Mammals." *Nature Communications* 7: 12739.
- Charlton, B. D., Taylor, A. M., and Reby, D. 2013. "Are Men Better than Women at Acoustic Size Judgements?" *Biology Letters* 9 (4): 20130270.
- Cheang, H. S., and Pell, M. D. 2008. "The Sound of Sarcasm." *Speech Communication* 50 (5): 366-81.
- Chiba, T., and Kajiyama, M. 1958. *The Vowel: Its Nature and Structure*. Tokyo: Phonetic Society of Japan.
- Collins, S. A. 2000. "Men's Voices and Women's Choices." *Animal Behaviour* 60 (6): 773-80.
- Collins, S. A., and Missing, C. 2003. "Vocal and Visual Attractiveness Are Related in Women." *Animal Behaviour* 65 (5): 997-1004.
- Cosmides, L. 1983. "Invariances in the Acoustic Expression of Emotion during Speech." *Journal of Experimental Psychology: Human Perception and Performance* 9 (6): 864-81.
- Cutler, A., Dahan, D., and Van Donselaar, W. 1997. "Prosody in the Comprehension of Spoken Language: A Literature Review." *Language and Speech* 40 (2): 141-201.
- Cutler, A., and Pearson, M. 1986. "On the Analysis of Prosodic Turn-taking Cues." *Intonation in Discourse*, edited by Catherine Johns-Lewis, 139-55. London: Croom Helm.
- Dabbs, J. M., and Mallinger, A. 1999. "High Testosterone Levels Predict Low Voice Pitch among Men." *Personality and Individual Differences* 27 (4): 801-4.

## The Evolution of Voice Perception

---

Darwin, C. 1872. *The Expression of the Emotions in Man and Animals*. London: John Murray.

DeBruine, L. M., Jones, B. C., Crawford, J. R., Welling, L. L., and Little, A. C. 2010. "The Health of a Nation Predicts Their Mate Preferences: Cross-cultural Variation in Women's Preferences for Masculinized Male Faces." *Proceedings of the Royal Society B: Biological Sciences* 277 (1692): 2405-10.

Eisenegger, C., Haushofer, J., and Fehr, E. 2011. "The Role of Testosterone in Social Interaction." *Trends in Cognitive Sciences* 15 (6): 263-71.

(p. 293) Ekman, P. 1997. "Should We Call It Expression or Communication?" *Innovation: The European Journal of Social Science Research* 10 (4): 333-44.

Ekman, P., Davidson, R. J., and Friesen, W. V. 1990. "The Duchenne Smile: Emotional Expression and Brain Physiology II." *Journal of Personality and Social Psychology* 58 (2): 342-53.

Elfenbein, H. A., and Ambady, N. 2002. "On the Universality and Cultural Specificity of Emotion Recognition: A Meta-Analysis." *Psychological Bulletin* 128 (2): 208-35.

Evans, S., Neave, N., Wakelin, D., and Hamilton, C. 2008. "The Relationship Between Testosterone and Vocal Frequencies in Human Males." *Physiology and Behavior* 93 (4-5): 783-8.

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 (4): 253-67.

Fant, F. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

Feinberg, D., DeBruine, L., Jones, B., Little, A., O'Connor, J., and Tigue, C. 2012. "Women's Self-perceived Health and Attractiveness Predict their Male Vocal Masculinity Preferences in Different Directions across Short- and Long-Term Relationship Contexts." *Behavioral Ecology and Sociobiology* 66 (3): 413-18.

Feinberg, D. R. 2008. "Are Human Faces and Voices Ornaments Signaling Common Underlying Cues to Mate Value?" *Evolutionary Anthropology* 17 (2): 112-18.

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 (4): 233-41.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., and Perrett, D. I. 2008. "The Role of Femininity and Averageness of Voice Pitch in Aesthetic Judgments of Women's Voices." *Perception* 37 (4): 615-23.

## The Evolution of Voice Perception

---

Feinberg, D. R., Jones, B. C., Law-Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., et al. 2006. "Menstrual Cycle, Trait Estrogen Level, and Masculinity Preferences in the Human Voice." *Hormones and Behavior* 49 (2): 215–22.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., and Perrett, D. I. 2005. "Manipulations of Fundamental and Formant Frequencies Influence the Attractiveness of Human Male Voices." *Animal Behaviour* 69 (3): 561–8.

Fitch, W. T. 1994. "Vocal Tract Length Perception and the Evolution of Language." PhD diss., Brown University.

Fitch, W. T. 1997. "Vocal Tract Length and Formant Frequency Dispersion Correlate with Body Size in Rhesus Macaques." *Journal of the Acoustical Society of America* 102 (2 Pt 1): 1213–22.

Fitch, W. T. 2000. "The Evolution of Speech: A Comparative Review." *Trends in Cognitive Sciences* 4 (7) 258–67.

Fitch, W. T., and Giedd, J. 1999. "Morphology and Development of the Human Vocal Tract: A Study Using Magnetic Resonance Imaging." *Journal of the Acoustical Society of America* 106 (3): 1511–22.

Fitch, W. T., and Hauser, M. 2003. "Unpacking 'Honesty': Vertebrate Vocal Production and the Evolution of Acoustic Signals." In *Acoustic Communication*, 65–137. New York: Springer.

Fitch, W. T., Neubauer, J., and Herzog, H. 2002. "Calls Out of Chaos: The Adaptive Significance of Nonlinear Phenomena in Mammalian Vocal Production." *Animal Behaviour* 63 (3): 407–18.

Fouquet, M., Pisanski, K., Mathevon, M., and Reby, D. 2016. "Seven and Up: Individual Differences in Male Voice Fundamental Frequency Emerge before Puberty and Remain Stable Throughout Adulthood." *Royal Society Open Science* 3 (10): 160395.

Fraccaro, P. J., Jones, B. C., Vukovic, J., Smith, F. G., Watkins, C. D., Feinberg, D. R., et al. 2011. "Experimental Evidence that Women Speak in a Higher Voice Pitch to Men They Find Attractive." *Journal of Evolutionary Psychology* 9 (1): 57–67.

(p. 294) Frederick, D. A., and Haselton, M. G. 2007. "Why Is Muscularity Sexy? Tests of the Fitness Indicator Hypothesis." *Personality and Social Psychology Bulletin* 33 (8): 1167–83.

Furlow, B. F. 1997. "Human Neonatal Cry Quality as an Honest Signal of Fitness." *Evolution and Human Behavior* 18 (3): 175–93.

Gallup, A. C., White, D. D., and Gallup Jr., G. G. 2007. "Handgrip Strength Predicts Sexual Behavior, Body Morphology, and Aggression in Male College Students." *Evolution and Human Behavior* 28 (6): 423–9.

## The Evolution of Voice Perception

---

Gangestad, S. W., and Thornhill, R. 2008. "Human Oestrus." *Proceedings of the Royal Society B-Biological Sciences* 275 (1638): 991-1000.

Gendron, M., Roberson, D., and Barrett, L. F. 2015. "Cultural Variation in Emotion Perception Is real: A Response to Sauter, Eisner, Ekman, and Scott (2015)." *Psychological Science* 26 (3): 357-9.

Gendron, M., Roberson, D., van der Vyver, J. M., and Barrett, L. F. 2014. "Cultural Relativity in Perceiving Emotion from Vocalizations." *Psychological Science* 25 (4): 911-20.

Gervais, M., and Wilson, D. S. 2005. "The Evolution and Functions of Laughter and Humor: A Synthetic Approach." *Quarterly Review of Biology* 80 (4): 395-430.

Ghazanfar, A. A., and Rendall, D. 2008. "Evolution of Human Vocal Production." *Current Biology* 18 (11): R457-60.

Gildersleeve, K., Haselton, M. G., and Fales, M. R. 2014. "Do Women's Mate Preferences Change Across the Ovulatory Cycle? A Meta-Analytic Review." *Psychological Bulletin* 140 (5): 1205-59.

González, J. 2003. "Estimation of Speakers' Weight and Height from Speech: a Re-analysis of Data from Multiple Studies by Lass and Colleagues." *Perceptual and Motor Skills* 96 (1): 297-304.

González, J. 2006. "Research in Acoustics of Human Speech Sounds: Correlates and Perception of Speaker Body Size." *Recent Research Development in Applied Physics* 9: 1-15.

Grafen, A. 1990. "Biological Signals as Handicaps." *Journal of Theoretical Biology* 144 (4): 517-46.

Grossmann, T., Oberecker, R., Koch, S. P., and Friederici, A. D. 2010. "The Developmental Origins of Voice Processing in the Human Brain." *Neuron* 65 (6): 852-8.

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." *Journal of Laryngology and Otology* 112 (5): 451-4.

Haselton, M. G., and Gildersleeve, K. 2011. "Can Men Detect Ovulation?" *Current Directions in Psychological Science* 20 (2): 87-92.

Henrich, J., Heine, S. J., and Norenzayan, A. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33 (2-3): 61-83.

Hodges-Simeon, C. R., Gurven, M., Puts, D. A., and Gaulin, S. J. 2014. "Vocal Fundamental and Formant Frequencies are Honest Signals of Threat Potential in Peripubertal Males." *Behavioral Ecology* 25 (4): 984-8.

Hollien, H. 2014. "Vocal Fold Dynamics for Frequency Change." *Journal of Voice* 28 (4): 395-405.

## The Evolution of Voice Perception

---

Hufschmidt, C., Weege, B., Röder, S., Pisanski, K., Neave, N., and Fink, B. 2015. "Physical Strength and Gender Identification from Dance Movements." *Personality and Individual Differences* 76: 13-17.

Hughes, S. M., Dispenza, F., and Gallup Jr., G. G. 2004. "Ratings of Voice Attractiveness Predict Sexual Behavior and Body Configuration." *Evolution and Human Behavior* 25 (5): 295-304.

Hughes, S. M., Farley, S. D., and Rhodes, B. C. 2010. "Vocal and Physiological Changes in Response to the Physical Attractiveness of Conversational Partners." *Journal of Nonverbal Behavior* 34 (3): 155-67.

(p. 295) Hughes, S. M., and Miller, N. E. 2015. "What Sounds Beautiful Looks Beautiful Stereotype: The Matching of Attractiveness of Voices and Faces." *Journal of Social and Personal Relationships* 33 (7): 984-96.

Hughes, S. M., Mogilski, J. K., and Harrison, M. A. 2014. "The Perception and Parameters of Intentional Voice Manipulation." *Journal of Nonverbal Behavior* 38 (1): 107-27.

Irino, T., Aoki, Y., Kawahara, H., and Patterson, R. D. 2012. "Comparison of Performance with Voiced and Whispered Speech in Word Recognition and Mean-Formant-Frequency Discrimination." *Speech Communication* 54 (9): 998-1013.

Johnstone, R. A., and Grafen, A. 1993. "Dishonesty and the Handicap Principle." *Animal Behaviour* 46 (4): 759-64.

Jones, B. C., Boothroyd, L., Feinberg, D. R., and DeBruine, L. M. 2010. "Age at Menarche Predicts Individual Differences in Women's Preferences for Masculinized Male Voices in Adulthood." *Personality and Individual Differences* 48 (7): 860-3.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., and Vukovic, J. 2008. "Integrating Cues of Social Interest and Voice Pitch in Men's Preferences for Women's Voices." *Biology Letters* 4 (2): 192-4.

Jones, B. C., Feinberg, D. R., Watkins, C. D., Fincher, C. L., Little, A. C., and DeBruine, L. M. 2013. "Pathogen Disgust Predicts Women's Preferences for Masculinity in Men's Voices, Faces, and Bodies." *Behavioral Ecology* 24 (2): 373-9.

Jurgens, U. 2002. "Neural Pathways Underlying Vocal Control." *Neuroscience and Biobehavioral Reviews* 26 (2): 235-8.

Juslin, P. N., and Laukka, P. 2003. "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?" *Psychological Bulletin* 129 (5): 770.

Kanwisher, N., McDermott, J., and Chun, M. M. 1997. "The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception." *Journal of Neuroscience* 17 (11): 4302-11.

## The Evolution of Voice Perception

---

Krebs, J. R., and Dawkins, R. 1984. "Animal Signals: Mind-Reading and Manipulation." In *Behavioral Ecology: An Evolutionary Approach*, edited by J. R. Krebs and N. B. Davies, 380–402. Oxford: Blackwell.

Kreiman, J., and Sidtis, D. 2011. *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Hoboken, NJ: John Wiley and Sons.

Lachmann, M., Szamado, S., and Bergstrom, C. T. 2001. "Cost and Conflict in Animal Signals and Human Language." *Proceedings of the National Academy of Sciences* 98 (23): 13189–94.

Lassek, W. D., and Gaulin, S. J. 2009. "Costs and Benefits of Fat-Free Muscle Mass in Men: Relationship to Mating Success, Dietary Requirements, and Native Immunity." *Evolution and Human Behavior* 30 (5): 322–8.

Lauder, G. V. 1981. "Form and Function: Structural Analysis in Evolutionary Morphology." *Paleobiology* 7 (4): 430–42.

Leongómez, J. D., Binter, J., Kubicová, L., Stolařová, P., Klapilová, K., Havlíček, J., et al. 2014. "Vocal Modulation during Courtship Increases Proceptivity even in Naive Listeners." *Evolution and Human Behavior* 35 (6): 489–96.

Levrero, F., Mathevon, N., Pisanski, K., Gustafsson, E., and Reby, D. 2018. "The pitch of babies' cries predicts their voice pitch at age 5." *Biology Letters* 14 (7): 20180065.

Lieberman, D., McCarthy, R., Hiiemae, K., and Palmer, J. 2001. "Ontogeny of Postnatal Hyoid and Larynx Descent in Humans." *Archives of Oral Biology* 46 (2): 117–28.

Little, A. C., Connely, J., Feinberg, D. R., Jones, B. C., and Roberts, S. C. 2011. "Human Preference for Masculinity Differs According to Context in Faces, Bodies, Voices, and Smell." *Behavioral Ecology* 22 (4): 862–8.

(p. 296) Little, A. C., Feinberg, D. R., DeBruine, L. M., and Jones, B. C. 2013. "Adaptation to Faces and Voices: Unimodal, Cross-Modal, and Sex-Specific Effects." *Psychological Science* 24 (11): 2297–305.

Lummaa, V., Vuorisalo, T., Barr, R. G. and Lehtonen, L. 1998. "Why Cry? Adaptive Significance of Intensive Crying in Human Infants." *Evolution and Human Behavior* 19 (3): 193–202.

Marcinkowska, U. M., Kozlov, M. V., Cai, H., Contreras-Garduño, J., Dixson, B. J., Oana, G. A., et al. 2014. "Cross-Cultural Variation in Men's Preference for Sexual Dimorphism in Women's Faces." *Biology Letters* 10 (4): 20130850.

Mazur, A., and Booth, A. 1998. "Testosterone and Dominance in Men." *Behavioral and Brain Sciences* 21 (03): 353–63.

Maynard Smith, J., and Harper, D. 2003. *Animal Signals*. Oxford: Oxford University Press.

## The Evolution of Voice Perception

---

Morton, E. S. 1977. "On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds." *American Naturalist* 111 (981): 855-69.

Newman, J. D. 2007. "Neural Circuits Underlying Crying and Cry Responding in Mammals." *Behavioural Brain Research* 182 (2): 155-65.

Nisbett, R. E., and Wilson, T. D. 1977. "The Halo Effect: Evidence for Unconscious Alteration of Judgment." *Journal of Personality and Social Psychology* 34 (4): 250-6.

O'Connor, J. J. M., Fraccaro, P. J., Pisanski, K., Tigue, C. C., and Feinberg, D. R. 2013. "Men's Preferences for Women's Femininity in Dynamic Cross-Modal Stimuli." *PloS one* 8 (7): e69531.

O'Connor, J. J. M., Fraccaro, P. J., Pisanski, K., Tigue, C. C., O'Donnell, T. J., and Feinberg, D. R. 2014. "Social Dialect and Men's Voice Pitch Influence Women's Mate Preferences." *Evolution and Human Behavior* 35 (5): 368-75.

O'Connor, J. J. M., Re, D. E., and Feinberg, D. R. 2011. "Voice Pitch Influences Perceptions of Sexual Infidelity." *Evolutionary Psychology* 9 (1): 64-78.

O'Connor, J. J. M., Fraccaro, P. J., and Feinberg, D. R. 2012. "The Influence of Male Voice Pitch on Women's Perceptions of Relationship Investment." *Journal of Evolutionary Psychology* 10 (1): 1-13.

Ohala, J. J. 1984. "An Ethological Perspective on Common Cross-language Utilization of F0 of Voice." *Phonetica* 41 (1): 1-16.

Oliver, J. C., and González, J. 2004. "Percepción a través de la Voz de las Características Físicas del Hablante: Identificación de la Estatura a partir de una Frase o una Vocal." *Revista de Psicología General y Aplicada* 57 (1): 21-34.

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 Anthropology: Issues, News, and Reviews* 10 (2): 58-71.

Patterson, R. D., Smith, D. R., van Dinther, R., and Walters, T. C. 2008. "Size Information in the Production and Perception of Communication Sounds." In *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay, 43-75. New York: Springer.

Pell, M. D., Monetta, L., Paulmann, S., and Kotz, S. A. 2009. "Recognizing Emotions in a Foreign Language." *Journal of Nonverbal Behavior* 33 (2): 107-20.

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., and Belin, P. 2015. "The Human Voice Areas: Spatial Organization and Inter-individual Variability in Temporal and Extra-temporal Cortices." *NeuroImage* 119: 164-74.

## The Evolution of Voice Perception

---

Peters, R. H. 1986. *The Ecological Implications of Body Size*, Volume 2. Cambridge: Cambridge University Press.

(p. 297) Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. 2008. "A Voice Region in the Monkey Brain." *Nature Neuroscience* 11 (3): 367-74.

Pietraszewski, D., Wertz, A. E., Bryant, G. A., and Wynn, K. 2017. "Three-Month-Old Human Infants Use Vocal Cues of Body Size." *Proceedings of the Royal Society Biological Sciences Series B* 284 (1856): 20170656.

Pipitone, R. N., and Gallup Jr., G. G. 2008. "Women's Voice Attractiveness Varies Across the Menstrual Cycle." *Evolution and Human Behavior* 29 (4): 268-74.

Pipitone, N. R., and Gallup, G. G. 2012. "The Unique Impact of Menstruation on the Female Voice: Implications for the Evolution of Menstrual Cycle Cues." *Ethology* 118 (3): 281-91.

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., and Reby, D. 2016. "Voice Modulation: A Window into the Origins of Human Vocal Control?" *Trends in Cognitive Sciences* 20 (4): 304-18.

Pisanski, K., and Feinberg, D. R. 2013. "Cross-Cultural Variation in Mate Preferences for Averageness, Symmetry, Body Size, and Masculinity." *Cross-Cultural Research* 47 (2): 162-97.

Pisanski, K., Feinberg, D., Oleszkiewicz, A., and Sorokowska, A. 2017. "Voice Cues Are Used in a Similar Way by Blind and Sighted Adults When Assessing Women's Body Size." *Scientific Reports* 7 (1): 10329.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., and Feinberg, D. R. 2014. "Return to Oz: Voice Pitch Facilitates Assessments of Men's Body Size." *Journal of Experimental Psychology: Human Perception and Performance* 40 (4): 1316-31.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J., Röder, S., Andrews, P., Fink, B., et al. 2014. "Vocal Indicators of Body Size in Men and Women: A Meta-Analysis." *Animal Behaviour* 95: 89-99.

Pisanski, K., Hahn, A. C., Fisher, C. I., DeBruine, L. M., Feinberg, D. R., and Jones, B. C. 2014. "Changes in Salivary Estradiol Predict Changes in Women's Preferences for Vocal Masculinity." *Hormones and Behavior* 66 (3): 493-7.

Pisanski, K., Jones, B. C., Fink, B., O'Connor, J. J., DeBruine, L. M., Röder, S., and Feinberg, D. R. 2016. "Voice Parameters Predict Sex-specific Body Morphology in Men and Women." *Animal Behaviour* 112: 13-22.

Pisanski, K., Mishra, S., and Rendall, D. 2012. "The Evolved Psychology of Voice: Evaluating Interrelationships in Listeners' Assessments of the Size, Masculinity, and Attractiveness of Unseen Speakers." *Evolution and Human Behavior* 33 (5): 509-19.

## The Evolution of Voice Perception

---

Pisanski, K., Mora, E. C., Pisanski, A., Reby, D., Sorokowski, P., Frackowiak, T., and Feinberg, D. R. 2016. "Volitional Exaggeration of Body Size through Fundamental and Formant Frequency Modulation in Humans." *Scientific Reports* 6: 34389.

Pisanski, K., Oleszkiewicz, A., Plachetka, J., Gmiterek, M., and Reby, D. 2018. "Voice pitch modulation in human mate choice." *Proceedings of the Royal Society B: Biological Sciences* 285 (1893): 20181634.

Pisanski, K., Oleszkiewicz, A., and Sorokowska, A. 2016. "Can Blind Persons Accurately Assess Body Size from the Voice?" *Biology Letters* 12 (4): 20160063.

Pisanski, K., and Rendall, D. 2011. "The Prioritization of Voice Fundamental Frequency or Formants in Listeners' Assessments of Speaker Size, Masculinity, and Attractiveness." *Journal of the Acoustical Society of America* 129 (4): 2201–12.

Porges, S. W. 2001. "The Polyvagal Theory: Phylogenetic Substrates of a Social Nervous System." *International Journal of Psychophysiology* 42 (2): 123–46.

Puts, D. A. 2005. "Mating Context and Menstrual Phase Affect Women's Preferences for Male Voice Pitch." *Evolution and Human Behavior* 26 (5): 388–97.

(p. 298) Puts, D. A. 2006. "Cyclic Variation in Women's Preferences for Masculine Traits—Potential Hormonal Causes." *Human Nature* 17 (1): 114–27.

Puts, D. A., Apicella, C. L., and Cardenas, R. A. 2012. "Masculine Voices Signal Men's Threat Potential in Forager and Industrial Societies." *Proceedings of the Royal Society Biological Sciences Series B* 279 (1728): 601–9.

Puts, D. A., Bailey, D. H., Cárdenas, R. A., Burriss, R. P., Welling, L. L., Wheatley, J. R., et al. 2013. "Women's Attractiveness Changes with Estradiol and Progesterone across the Ovulatory Cycle." *Hormones and Behavior* 63 (1): 13–19.

Puts, D. A., Barndt, J. L., Welling, L. L., Dawood, K., and Burriss, R. P. 2011. "Intrasexual Competition among Women: Vocal Femininity Affects Perceptions of Attractiveness and Flirtatiousness." *Personality and Individual Differences* 50 (1): 111–15.

Puts, D. A., Doll, L. M., and Hill, A. K. 2014. "Sexual Selection on Human Voices." In *Evolutionary Perspectives on Human Sexual Psychology and Behavior*, edited by Viviana A. Weekes-Shackelford and Todd K. Shackelford, 69–86. New York: Springer.

Puts, D. A., Gaulin, S. J., and Verdolini, K. 2006. "Dominance and the Evolution of Sexual Dimorphism in Human Voice Pitch." *Evolution and Human Behavior* 27 (4): 283–96.

Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., and Jablonski, N. G. 2016. "Sexual Selection on Male Vocal Fundamental Frequency in Humans and Other Anthropoids." *Proceedings of the Royal Society Biological Sciences Series B* 283 (1829): 20152830.

## The Evolution of Voice Perception

---

Raine, J., Pisanski, K., Oleszkiewicz, A., Simner, J., and Reby, D., 2018. "Human listeners can accurately judge strength and height relative to self from aggressive roars and speech." *iScience* 4: 273–280.

Rantala, M. J., Moore, F. R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., et al. 2012. "Evidence for the Stress-Linked Immunocompetence Handicap Hypothesis in Humans." *Nature Communications* 3: 694.

Re, D. E., O'Connor, J. J. M., Bennett, P. J., and Feinberg, D. R. 2012. "Preferences for Very Low and Very High Voice Pitch in Humans." *PloS one* 7 (3): e32719.

Reby, D., Levréro, F., Gustafsson, E., and Mathevon, N. 2016. "Sex Stereotypes Influence Adults' Perception of Babies' Cries." *BMC Psychology* 4 (19): 1–12.

Rendall, D., Kollias, S., Ney, C., and Lloyd, P. 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." *Journal of the Acoustical Society of America* 117 (2): 944–55.

Rendall, D., Owren, M. J., and Ryan, M. J. 2009. "What Do Animal Signals Mean?" *Animal Behaviour* 78 (2): 233–40.

Rendall, D., Vokey, J. R., and Nemeth, C. 2007. "Lifting the Curtain on the Wizard of Oz: Biased Voice-based Impressions of Speaker Size." *Journal of Experimental Psychology: Human Perception and Performance* 33 (5): 1208–19.

Roche, J. M., Peters, B., and Dale, R. 2015. "'Your Tone Says It All': The Processing and Interpretation of Affective Language." *Speech Communication* 66: 47–64.

Ruch, W., and Ekman, P. 2001. "The Expressive Pattern of Laughter." In *Emotion, Qualia, and Consciousness*, edited by A. Kaszniak, 426–43. Tokyo: Word Scientific.

Russell, J. A. 1980. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology* 39: 1161–78.

Ryalls, J. H., and Lieberman, P. 1982. "Fundamental Frequency and Vowel Perception." *Journal of the Acoustical Society of America* 72 (5): 1631–4.

Sauter, D. A., Eisner, F., Calder, A. J., and Scott, S. K. 2010. "Perceptual Cues in Nonverbal Vocal Expressions of Emotion." *Quarterly Journal of Experimental Psychology* 63 (11): 2251–72.

(p. 299) Sauter, D. A., Eisner, F., Ekman, P., and Scott, S. K. 2010. "Cross-Cultural Recognition of Basic Emotions through Nonverbal Emotional Vocalizations." *Proceedings of the National Academy of Sciences* 107 (6): 2408–12.

Sauter, D. A., Eisner, F., Ekman, P., and Scott, S. K. 2015. "Emotional Vocalizations are Recognized across Cultures Regardless of the Valence of Distractors." *Psychological Science* 26 (3): 354–6.

## The Evolution of Voice Perception

---

Saxton, T. K., Burriss, R. P., Murray, L. K., Rowland, H. M., and Roberts, S. C. 2009. "Face, Body, and Speech Cues Independently Predict Judgments of Attractiveness." *Journal of Evolutionary Psychology* 7 (1): 23-35.

Saxton, T. K., DeBruine, L. M., Jones, B. C., Little, A. C., and Roberts, S. C. 2009. "Face and Voice Attractiveness Judgments Change during Adolescence." *Evolution and Human Behavior* 30 (6): 398-408.

Saxton, T. K., Mackey, L. L., McCarty, K., and Neave, N. 2015. "A Lover or a Fighter? Opposing Sexual Selection Pressures on Men's Vocal Pitch and Facial Hair." *Behavioral Ecology* 27 (2): 512-9.

Scherer, K. R., Banse, R., and Wallbott, H. 2001. "Emotion Inferences from Vocal Expression Correlate across Languages and Cultures." *Journal of Cross-Cultural Psychology* 32 (1): 76-92.

Scott-Phillips, T. 2014. *Speaking Our Minds: Why Human Communication is Different, and How Language Evolved to Make it Special*. London: Palgrave MacMillan.

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." *Proceedings of the Royal Society Biological Sciences Series B* 277 (1699): 3509-18.

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., and Gurven, M. 2009. "Human Adaptations for the Visual Assessment of Strength and Fighting Ability from the Body and Face." *Proceedings of the Royal Society B: Biological Sciences* 276 (1656): 575-84.

Sell, A., Hone, L. S., and Pound, N. 2012. "The Importance of Physical Strength to Human Males." *Human Nature* 23 (1): 30-44.

Sherman, B. M., and Korenman, S. G. 1975. "Hormonal Characteristics of the Human Menstrual Cycle throughout Reproductive Life." *Journal of Clinical Investigation* 55 (4): 699-706.

Simpson, A. P. 2009. "Phonetic Differences Between Male and Female Speech." *Language and Linguistics Compass* 3 (2): 621-40.

Skrinda, I., Krama, T., Kecko, S., Moore, F. R., Kaasik, A., Meija, L., et al. 2014. "Body Height, Immunity, Facial and Vocal Attractiveness in Young Men." *Naturwissenschaften* 101 (12): 1017-25.

Soltis, J. 2004. "The Signal Functions of Early Infant Crying." *Behavioral and Brain Sciences* 27 (4): 443-58.

Spence, C. 2011. "Crossmodal Correspondences: A Tutorial Review." *Attention, Perception, and Psychophysics* 73 (4): 971-95.

## The Evolution of Voice Perception

---

Sperber, D., and Wilson, D. 1995. *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.

Swerts, M., and Hirschberg, J. 2010. "Prosodic Predictors of Upcoming Positive or Negative Content in Spoken Messages." *Journal of the Acoustical Society of America* 128 (3): 1337-45.

Taylor, A. M., Charlton, B. D., and Reby, D. 2016. "Vocal Production by Terrestrial Mammals: Source, Filter, and Function." In *Vertebrate Sound Production and Acoustic Communication*, edited by R. A. Suthers, W. T. Fitch, R. R. Fay, and A. N. Popper, 229-59. Heidelberg: Springer.

Taylor, A. M., and Reby, D. 2010. "The Contribution of Source-Filter Theory to Mammal Vocal Communication Research." *Journal of Zoology* 280 (3): 221-36.

(p. 300) Thompson, W., and Balkwill, L. L. 2006. "Decoding Speech Prosody in Five Languages." *Semiotica* 158 (1/4): 407-24.

Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., and Feinberg, D. R. 2012. "Voice Pitch Influences Voting Behavior." *Evolution and Human Behavior* 33 (3): 210-6.

Tinbergen, N. 1952. "Derived Activities: Their Causation, Biological Significance, Origin and Emancipation during Evolution." *Quarterly Review of Biology* 27 (1): 1-32.

Tinbergen, N. 1963. "On Aims and Methods of Ethology." *Zeitschrift für Tierpsychologie* 20 (4): 410-33.

Titze, I. R. 1994. *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall.

Titze, I. R. 2011. "Vocal Fold Mass Is Not a Useful Quantity for Describing F0 in Vocalization." *Journal of Speech, Language and Hearing Research* 54 (2): 520-2.

van Dommelen, W. A., and Moxness, B. H. 1995. "Acoustic Parameters in Speaker Height and Weight Identification: Sex-Specific Behaviour." *Language and Speech* 38 (3): 267-87.

Vouloumanos, A., and Werker, J. F. 2007. "Listening to Language at Birth: Evidence for a Bias for Speech in Neonates." *Developmental Science* 10 (2): 159-64.

Vukovic, J., Feinberg, D. R., Jones, B. C., DeBruine, L. M., Welling, L. L. M., Little, A. C., et al. 2008. "Self-Rated Attractiveness Predicts Individual Differences in Women's Preferences for Masculine Men's Voices." *Personality and Individual Differences* 45 (6): 451-6.

Vukovic, J., Jones, B. C., DeBruine, L., Feinberg, D. R., Smith, F. G., Little, A. C., et al. 2010. "Women's Own Voice Pitch Predicts their Preferences for Masculinity in Men's Voices." *Behavioral Ecology* 21 (4): 767-72.

Vukovic, J., Jones, B. C., Feinberg, D. R., DeBruine, L. M., Smith, F. G., Welling, L. L., et al. 2011. "Variation in Perceptions of Physical Dominance and Trustworthiness Predicts Indi-

## The Evolution of Voice Perception

---

vidual Differences in the Effect of Relationship Context on Women's Preferences for Masculine Pitch in Men's Voices." *British Journal of Psychology* 102 (1): 37–48.

Wheatley, J. R., Apicella, C. A., Burriss, R. P., Cárdenas, R. A., Bailey, D. H., Welling, L. L. M., and Puts, D. A. 2014. "Women's Faces and Voices are Cues to Reproductive Potential in Industrial and Forager Societies." *Evolution and Human Behavior* 35 (4): 264–71.

Williams, G. C. 1966. *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.

Wood, W., Kressel, L., Joshi, P. D., and Louie, B. 2014. "Meta-Analysis of Menstrual Cycle Effects on Women's Mate Preferences." *Emotion Review* 6 (3): 229–49.

Zeifman D. M. 2001. "An Ethological Analysis of Human Infant Crying: Answering Tinbergen's Four Questions." *Developmental Psychobiology* 39 (4): 265–85.

Zuckerman, M., Hodgins, H., and Miyake, K. 1990. "The Vocal Attractiveness Stereotype: Replication and Elaboration." *Journal of Nonverbal Behavior* 14 (2): 97–112.

### Notes:

(1.) It is worth noting here that all signals have potential cue value in that they might inform intended and unintended perceivers in ways that are not part of the evolved design (Maynard Smith and Harper 2003).

(2.) Larger vocal folds vibrate at a slower rate than do smaller vocal folds, resulting in a relatively lower fundamental frequency and perceived pitch; however, regardless of mass, fundamental frequency increases when the vocal folds are stretched and become tenser.

(3.) Low pitch is perceptually associated with large physical size and high pitch with small size regardless whether the sounds are pure or complex tones, musical passages, or vocalizations (see, e.g., Bien, ten Oever, Goebel, and Sack 2012; Evans and Treisman 2010).

(4.) Harmonic density facilitates more accurate speech recognition via the same mechanism (Ryalls and Lieberman 1982).

(5.) There is, nevertheless, evidence that people are able to accurately assess strength from photographs of faces and bodies (Sell et al. 2009) and even from videos of avatars synthesized from male dancers (Hufschmidt et al. 2015).

(6.) During puberty, an increase in testosterone among males increases the mass of the vocal folds and causes a drop in voice pitch (Harries et al. 1998; Lieberman, McCarthy, Hiiemae, and Palmer 2001) that is much greater than the pubertal drop in pitch among females (Abitbol, Abitbol, and Abitbol 1999). The male larynx also descends slightly during puberty, elongating the male vocal tract relative to the female vocal tract thereby lowering formants (Fitch and Giedd 1999; Lieberman et al. 2001).

## The Evolution of Voice Perception

---

(7.) Although preferences for sexual dimorphism in the voice have been documented in many cultures, the degree to which listeners prefer sexual dimorphism in potential mates has been shown to vary cross-culturally as a function of various evolutionary relevant factors, such as pathogen prevalence (see Pisanski and Feinberg 2013).

(8.) Although the just-noticeable differences or discrimination thresholds in voice pitch and formant perception are approximately 5 percent for vowel sounds (Pisanski and Rendall 2011), studies that manipulate the voice using computer software indicate that larger differences, in the order of 10 percent from baseline or greater, are typically required to affect judgments of vocal attractiveness (Feinberg, Jones, Little, Burt, and Perrett 2005; Pisanski and Rendall 2011) as supported by psychoacoustic research (Re, O'Connor, Bennett, and Feinberg 2012).

(9.) See also Borkowska and Pawlowski (2011) and Puts et al. (2011) for empirical evidence in support of intrasexual competition in women.

(10.) Women taking hormonal contraceptives may not show systematic variation in their vocal masculinity preferences and are typically excluded from studies examining cyclic shifts in women's preferences (Gildersleeve et al. 2014).

(11.) Men's facial femininity preferences in women appear to correlate positively, rather than negatively, with the health of a given nation (Marcinkowska et al. 2014). The potential ultimate function of this relationship is unclear.

(12.) In another study, O'Connor, Fraccaro and Feinberg (2014) showed that women's perceptions of men's vocal attractiveness interacted with sociolinguistic cues to men's socioeconomic status.

(13.) Unlike Gendron et al., Sauter, Eisner, Ekman, and Scott (2010) tested participants for their comprehension of vignettes by asking participants to explain the stories, and only included those individuals who demonstrated full understanding. Gendron, Roberson, and Barrett (2015) responded by claiming that confirmation of vignette understanding with repeated explanation amounts to category learning, and thus undermines attempts at measuring actual universals.

### **Katarzyna Pisanski**

postdoctoral research fellow at the University of Sussex, Brighton

### **Gregory A. Bryant**

Professor of Communication, University of California, Los Angeles