Evolution and Human Behavior 42 (2021) 91–103

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

Can listeners assess men's self-reported health from their voice?

Graham Albert[a,*], Steven Arnocky[b], David A. Puts[c], Carolyn R. Hodges-Simeon[a]

[a] Department of Anthropology, Boston University, Boston, MA, United States
[b] Department of Psychology, Nipissing University, North Bay, Ontario, Canada
[c] Department of Anthropology, The Pennsylvania State University, University Park, PA, United States

A B S T R A C T

Men's voices may provide cues to overall condition; however, little research has assessed whether health status is reliably associated with perceivable voice parameters. In Study 1, we investigated whether listeners could classify voices belonging to men with either relatively lower or higher self-reported health. Participants rated voices for speaker health, disease likelihood, illness frequency, and symptom severity, as well as attractiveness (women only) and dominance (men only). Listeners' were mostly unable to judge the health of male speakers from their voices; however, men rated the voices of men with better self-reported health as sounding more dominant. In Study 2, we tested whether men's vocal parameters (fundamental frequency mean and variation, apparent vocal tract length, and harmonics-to-noise ratio) and aspects of their self-reported health predicted listeners' health and disease resistance ratings of those voices. Speakers' fundamental frequency (f_o) negatively predicted ratings of health. However, speakers' self-reported health did not predict ratings of health made by listeners. In Study 3, we investigated whether separately manipulating two sexually dimorphic vocal parameters—f_o and apparent vocal tract length (*VTL)*—affected listeners' health ratings. Listeners rated men's voices with lower f_o (but not *VTL*) as healthier, supporting findings from Study 2. Women rated voices with lower f_o and *VTL* as more attractive, and men rated them as more dominant. Thus, while both *VTL* and f_o affect dominance and attractiveness judgments, only f_o appears to affect health judgments. Results of the above studies suggest that, although listeners assign higher health ratings to speakers with more masculine f_o, these ratings may not be accurate at tracking speakers' self-rated health.

1. Introduction

The voice of adult humans is conspicuously sexually dimorphic (Pisanski & Bryant, 2016; Puts et al., 2011; Puts et al., 2016). Sex differences largely emerge during puberty with increases in steroid hormones: growth in the length and thickness of the male vocal folds and length of the male vocal tract produce lower fundamental frequency (f_o) and formants, respectively (Abitbol, Abitbol, & Abitbol, 1999; Butler et al., 1989; Fitch & Giedd, 1999; Harries, Hawkins, Hacking, & Hughes, 1998; Harries, Walker, Williams, Hawkins, & Hughes, 1997; Hollien, Green, & Massey, 1994; Markova et al., 2016; Puts et al., 2011; Hodges-Simeon, Gurven, Cárdenas, & Gaulin, 2013). These morphological changes result in men having a deeper, more resonant-sounding voice than women and children (Markova et al., 2016; Titze, 2000). Both sexes attend to these features of the voice (Feinberg, Jones, Little, Burt, & Perrett, 2005; Puts, Hodges, Cárdenas, & Gaulin, 2007), and use them to form interpersonal perceptions. Men's voices manipulated to have lower f_o, and/or formants are rated as more attractive by women in some studies (e.g., Feinberg et al., 2011; Feinberg et al., 2005; Hodges-Simeon et al., 2010; Puts, 2005; but not all: see Shirazi, Puts, & Escasa-Dorne, 2018; Rosenfield, Sorokowska, Sorokowski, & Puts, 2019). Men with lower pitched voices are also perceived as sounding larger, older, stronger, and more dominant (e.g., Bruckert, Liénard, Lacroix, Kreutzer, & Leboucher, 2006; Collins, 2000; Jones, Feinberg, DeBruine, Little, & Vukovic, 2010; Klofstad et al., 2015; Pisanski & Rendall, 2011; Puts, Gaulin, & Verdolini, 2006; Puts et al., 2011; Raine, Pisanski, Bond, Simner, & Reby, 2019; Rosenfield et al., 2019; Sell et al., 2010), which may impact decisions in voting (Klofstad, Anderson, & Peters, 2012), mating (Hodges-Simeon, Gaulin, & Puts, 2011; Hughes, Dispenza, & Gallup Jr., 2004; Puts, 2005; Puts et al., 2007; Saxton, Caryl, & Craig Roberts, 2006), hiring (Mayew, Parsons, & Venkatachalam, 2013), and other social arenas. The goal of the present investigation was to assess whether voice impacts listeners' perceptions of speakers' physical health. As a second aim, we also investigated the extent to which acoustic characteristics of men's voices are related to their self-rated health.

2. Immunocompetence handicapping hypothesis

The reasons why listeners attend to and make decisions based on vocal parameters is an active area of debate (Aung & Puts, 2020; Feinberg, Jones, & Armstrong, 2018; Feinberg, Jones, & Armstrong, 2019;

* Corresponding author: Department of Anthropology, Boston University, 232 Bay State Rd., Boston, MA 02215, United States.
E-mail address: grahama@bu.edu (G. Albert).

https://doi.org/10.1016/j.evolhumbehav.2020.08.001
Received 29 November 2019; Received in revised form 31 July 2020; Accepted 1 August 2020
Available online 6 August 2020

Puts & Aung, 2019), in part because it contributes to our understanding of the evolutionary origins of the conspicuously large sexual dimorphism in human voice pitch (Puts et al., 2016). The immunocompetence handicap hypothesis (ICHH; Folstad & Karter, 1992), based on handicapping models of sexual signaling (Hamilton & Zuk, 1982; Zahavi, 1975, 1977), posits that receivers attend to testosterone (T)-dependent traits in inter- and intrasexual competition because such traits indicate good genes and health (e.g, Gangestad & Simpson, 2000). T promotes mating effort (including the development of secondary sexual characteristics) by shunting energy from the immune system, suggesting that only the healthiest males can afford the immunologic cost of developing such conspicuous traits (Folstad & Karter, 1992; Muehlenbein & Bribiescas, 2005). In addition, infection suppresses T production (Furman, 2014; Simmons & Roney, 2009); therefore, individuals who inherit strong immune systems will, on average, have succumbed to fewer infections and suppressed T production less frequently and for shorter durations, which allows for the development of elaborate ornaments (Foo et al., 2020). From this perspective, T-dependent traits in male mammals may be used to judge quality because, on average, such traits are honest indicators of genetic predictors of health (Folstad & Karter, 1992; Wedekind & Folstad, 1994).

The ICHH has yielded mixed support across species (Alonso-Alvarez, Bertrand, Faivre, Chastel, & Sorci, 2007; Casagrande & Groothuis, 2011; Cox & John-Alder, 2007; Deviche & Cortez, 2005; Edler, Goymann, Schwabl, & Friedl, 2011; Evans, Goldsmith, & Norris, 2000; Fuxjager, Foufopoulos, Diaz-Uriarte, & Marler, 2011; Lea et al., 2018; Lindström, Krakower, Lundström, & Silverin, 2001; Owen-Ashley, Hasselquist, & Wingfield, 2004), including humans. Although some studies have demonstrated relationships between immune upregulation and androgen suppression (Furman, 2014; Muehlenbein, Alger, Cogswell, James, & Krogstad, 2005; Simmons & Roney, 2009), between T-levels and ornament quality (e.g., Dabbs & Mallinger, 1999; Evans, Neave, Wakelin, & Hamilton, 2008; Hodges-Simeon, Gurven, & Gaulin, 2015; Markova et al., 2016; Verdonck, Gaethofs, Carels, & de Zegher, 1999), and between T concentrations and health-based behaviors (Booth, Johnson, & Granger, 1999), few have found linear relationships between T concentrations and immune markers without the presentation of an immune challenge.

One of the most exhaustive studies of the ICHH in humans to date did not reveal significant relationships between immune function and concentrations of multiple androgens, including DHEA, DHT and T (Nowak, Pawłowski, Borkowska, Augustyniak, & Drulis-Kawa, 2018). However, another recent investigation in non-human primates revealed positive relationships between T-dependent behavioral traits and the expression of genes that encode for innate immunity (Lea et al., 2018). The inconsistent relationships between ornament expression, T and immunocompetence could be explained in part by a non-linear relationship between ornament quality and health; that is, once a threshold level of energy is attained for immune function the remainder can be allocated to signal development and expression, reducing the strength of the relationship between health and signal quality (Kokko, 1997).

Although the applicability of the ICHH to signal expression in humans is unclear, this does not negate selection on perceptions of health cues. Rather, direct selection, which focuses on organisms' preferences for and mating biases toward males that are more fecund, provide superior resources, offer more parental care, and/or reduce the female's reproductive costs serves as another plausible explanation for how preferences for cues to health could have evolved (Kokko, Brooks, Jennions, & Morley, 2003). Health may be especially preferred in biparental species such as humans, because choosing a healthy mate increases the likelihood of securing a partner with the capacity to adequately care for offspring during critical developmental periods (Allaire, 1988; Altschuler & Dale, 1999; Drotar, 1994; Thore, 1990; Tybur & Gangestad, 2011). In a similar vein, healthy individuals are less likely to contract diseases and thus less likely to pass such diseases onto their offspring (Gangestad, Haselton, & Buss, 2006). Finally, individuals in poor health are more likely to experience sickness behavior (e.g., lethargy, appetite loss; Adelman & Martin, 2009), making intrasexual status competition less successful. Thus, a direct-benefits explanation points to selection on observers to accurately judge health (Fisher, 1958), whereas the ICHH suggests that at least some of the dimensions on which they make their judgments may be linked to T and heritable immunocompetence.

3. Cues of health

Studies investigating observable cues of health have primarily focused on aspects of the human face (Fink, Grammer, & Matts, 2006; Henderson & Anglin, 2003; Little, McPherson, Dennington, & Jones, 2011; Pound et al., 2014; Roberts, Little, DeBruine, & Petrie, 2017; Zaidi et al., 2019). Indeed, characteristics such as skin coloration (Fink et al., 2006, Henderson et al., 2017; Roberts et al., 2017; Smith, Jones, DeBruine & Little 2009; Roberts et al., 2017), apparent skin texture (Fink, Grammer, & Thornhill, 2001), averageness (Foo, Simmons, & Rhodes, 2017; Jones, 2018) fluctuating asymmetry (Gangestad et al., 2006; Jones et al., 2001), facial adiposity (Henderson, Holzleitner, Talamas, & Perrett, 2016), and sexual dimorphism (Foo et al., 2020; Phalane, Tribe, Steel, Cholo, & Coetzee, 2017; however, see Zaidi et al., 2019) suggest that observers use various aspects of the phenotype to form health impressions of others, some of which are associated with T during development (e.g., Bulygina, Mitteroecker, & Aiello, 2006; Hodges-Simeon et al., 2015; Markova et al., 2016; Roosenboom et al., 2018; Thordarson, Johannsdottir, & Magnusson, 2006; Whitehouse et al., 2015; Verdonck et al., 1999).

4. Current study

In the present study, we examine four dimensions of the voice that have previously been linked, either directly or indirectly, with health and/or T. First, f_o may cue aspects of condition throughout the lifespan (Arnocky, Hodges-Simeon, Ouellette, & Albert, 2018; Furlow, 1997; Hodges-Simeon et al., 2015; Stathopoulos, Huber, & Sussman, 2011). Beginning with infancy, neonatal crying may communicate phenotypic quality to parents (Furlow, 1997; Lummaa, Vuorisalo, Barr, & Lehtonen, 1998; Soltis, 2004). Indeed, infants with higher cry pitch, which is rated as more aversive by parents, tend to be in poorer phenotypic condition and may garner less parental investment than infants with relatively lower cry pitch (Furlow, 1997). Moreover, cry f_o may index infant body size with relatively smaller infants producing cries with higher f_o (Wermke & Robb, 2010). Among adolescent males, those with higher BMI show earlier and steeper descent in f_o (Hodges-Simeon et al., 2013; Hodges-Simeon et al., 2015; Juul, Magnusdottir, Scheike, Prytz, & Skakkebæk, 2007). Among college-aged males, lower f_o was correlated with higher concentrations of secretory immunoglobin-A, a biomarker of mucosal immunity (Arnocky et al., 2018); however, another study found no relationship between vocal masculinity and antibody response to an immune challenge (Skrinda, Krama, Kecko, et al., 2014). Further, studies analyzing the effects of senescence on the human voice have shown changes in f_o (Awan, 2006; Dibazar, Berger, & Narayanan, 2006; Harnsberger, Shrivastav, Brown Jr, Rothman, & Hollien, 2008; Stathopoulos et al., 2011). For example, f_o is used to discriminate individuals in different age groups (Awan, 2006), such that f_o decreases with age in women and men (Awan, 2006; Harnsberger et al., 2008) and then increases in men as they enter old age (Harnsberger et al., 2008; Stathopoulos et al., 2011). Speaker f_o is also used in studies of voice pathology (e.g., Awan, 2006; Stathopoulos et al., 2011).

Importantly for our purposes, f_o is also associated with T throughout the life span—during development (Hodges-Simeon et al., 2015; Markova et al., 2016) and adulthood (Aung & Puts, 2020; Cartei, Bond, & Reby, 2014; Dabbs & Mallinger, 1999; Evans et al., 2008; Hodges-Simeon et al., 2020; Puts et al., 2011). In line with the stress-linked ICHH (Rantala et al., 2012), recent research has shown that men with relatively high levels of T and low levels of cortisol have lower f_o (Puts et al., 2016). Vocal attractiveness ratings (which are correlated with masculinity and T; Cartei et al., 2014; Feinberg et al., 2005) have been found to be associated with lower levels of facial fluctuating asymmetry, a putative indicator of developmental instability (Hill et al., 2017; but see Kordsmeyer et al., 2020). Furthermore, the increase in men's f_o with aging occurs with a concomitant decrease in circulating T levels and an increase in frailty (Hyde et al., 2010; Stathopoulos et al., 2011).

Second, we use formant frequencies, which are resonant frequencies of the vocal tract that can be used to estimate vocal tract length based on established algorithms (*VTL*; Kalashnikova, Carignan, & Burnham, 2017; Pisanski et al., 2014; Stevens, 1998). In one study, those with higher self-reported health and mucosal immunity exhibited longer apparent *VTL* (Arnocky et al., 2018); however, the reason for this association is unknown. Although T is both theoretically and empirically associated with f_o (Cartei et al., 2014; Dabbs & Mallinger, 1999; Evans et al., 2008; Puts et al., 2011), and *VTL* is more strongly linked with body size (yet explains only 10% of the variation in height; Pisanski et al., 2014), some evidence suggests that a T-linked, ICHH explanation may be extended to *VTL* as well. Preliminary research shows that exogenous T therapy results in significantly lower *VTL* measures among transgender males compared with cisgender females (Hodges-Simeon et al., 2020; Papp, 2012), which hints that T may influence vocal tract resonance. Although T and formant measures are not correlated in adult males (Arnocky et al., 2018; Puts et al., 2011), T predicts formant changes during puberty (Hodges-Simeon et al., 2015), and may contribute to the male drop in formant measures during this time (Hodges-Simeon et al., 2013).

Third, we measure variation in f_o (f_o-CV; f_o-SD divided by f_o). This is computed by dividing mean f_o.SD by f_o, which affects perceptions of monotonicity. Related work has demonstrated that lower f_o-SD among men is associated with perceptions of masculinity and dominance (Hodges-Simeon et al., 2014), and physical aggression (Puts et al., 2011), as well as measures of mating success (Hodges-Simeon et al., 2011) and status (Leongómez, Mileva, Little, & Roberts, 2017). Lower f_o-SD has been found to correlate with greater mucosal immunity (marginally), self-reported general health, and lower perceived vulnerability to disease (Arnocky et al., 2018). Moreover, f_o-SD increases with age and increasing frailty (Gorham-Rowan & Laures-Gore, 2006). However, research suggests that f_o-CV may be a better measure of variation in f_o than f_o-SD because it is not confounded by speakers mean f_o (cf., Eguchi, 1969; Fouquet, Pisanski, Mathevon, & Reby, 2016). Therefore, we elected to use f_o-CV in the current investigation.

Fourth, we utilize HNR, a measure of vocal quality that affects listeners' perceptions of hoarseness, breathiness, creakiness, and/or smoothness (Gorham-Rowan & Laures-Gore, 2006; Kreiman & Sidtis, 2011; Linville, 1996). HNR is the ratio of energy in harmonic to noise components of a voice. Additive noise appears to be one of the primary measures that listeners rely on when assessing whether a voice is pathological (Kreiman & Gerratt, 2003; Li & Jo, 2004) and is corrected following operations for dysphonia (Parsa & Jamieson, 2001). Severely dysphonic voices have higher noise relative to harmonics and high spectral slopes (Heman-Ackah, Michael, & Goding Jr, 2002; Li & Jo, 2004). Lower HNR has been associated with poor health history, smoking, and alcoholism (Kreiman & Sidtis, 2011). Those experiencing acute illness, such as a sore throat have voices with increased hoarseness, which is caused by edema of the vocal folds (Chan, 2010). Higher HNR is also perceived as more attractive (Babel et al., 2014). Therefore, listeners may perceive voices with higher HNR as sounding healthier.

5. Aims and hypothesis

Despite the links between health and voice parameters, no studies to our knowledge have examined whether individuals can accurately assess health from speakers' voices. Our first aim was to determine if listeners can extract cues of speakers' health from voices alone (Fig. 1A).

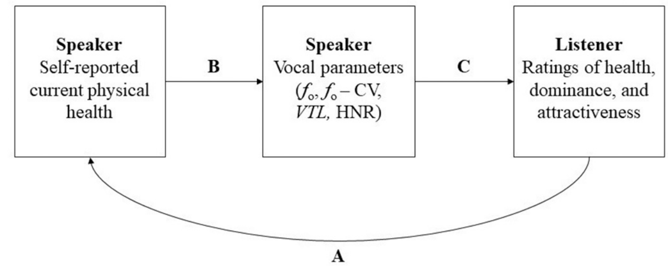

Fig. 1. Summary of research aims.

In Study 1 and 2, we assess whether listeners rate voices belonging to relatively healthier men as being healthier. In the present research, we assess speakers' health using the Health Perception Questionnaire Form II (HPQ II), which has been demonstrated to be a reliable and valid measure of respondents' health (Ware, 1976). Our second aim was to assess whether men with lower, more masculine voices experience better health (Fig. 1B). In Study 2, we tested for relationships between speakers' self-reported health and four vocal parameters, f_o, f_o-CV, HNR, and apparent *VTL*. Aim 3 investigates the relationship between speakers' vocal parameters and listeners' health ratings (Fig. 1C). That is, do listeners use masculine voice characteristics (i.e., lower f_o, f_o-CV, and longer apparent *VTL*) or other aspects of speech (i.e., HNR) to inform their ratings of health? In Studies 2 and 3, we use correlational and experimental designs to determine which speakers are rated as having better health.

6. Study 1

6.1. Methods

6.1.1. Participants

One hundred twenty-nine participants (68 females; aged 18 to 62, $M_{age} = 33.4$, $SD = 9.0$) were recruited through Amazon Mechanical Turk (MTurk), an online sampling technology, to complete a voice rating task and a questionnaire via Qualtrics. Reported ethnicities were Caucasian (68.3%), South Asian (10.3%), Black (8.7%), Asian (5.6%), Latin American (4.8%), Multiple Ethnicities (2.3%), South East Asian (1.6%), and Aboriginal (<1%).

6.1.2. Stimuli and materials

As part of a separate study on health and mating, 108 undergraduate men aged 17 to 29 ($M_{age} = 20.11$, $SD = 2.02$) were recruited from a small university and college in Canada (see Arnocky et al., 2018). All participants provided a voice recording and completed a questionnaire package, which included measures of physical health and disease avoidance behaviors. No participants indicated currently experiencing overt illness symptoms (i.e., current cold or flu).

Voice recordings. In a quiet room, participants were recorded speaking five monophthong vowel sounds (i.e., /eh/, /ee/, /ah/, /oh/, /oo/) into a Samson Meteorite USB Condenser microphone, positioned approximately 20 cm from participants' mouths. Recordings used Goldwave (version 6.10) software in mono with a sampling rate of 44.1 kHz and 16-bit quantization. Voice recordings were saved as high-quality uncompressed wav files.

Acoustic Analysis. Voice recordings were analyzed using Praat voice analysis software version 5.4.22 (Boersma & Weenink, 2014). We used the autocorrelation method to determine the f_o of each vowel sound within the utterance. The pitch floor was set to 60 Hz and the pitch ceiling was set to 300 Hz. To calculate the formant frequencies, vowel sounds were initially extracted (i.e., to exclude sporadic background noise, microphone pops, and accidental fricatives) by the Praat Vocal Toolkit Extract Vowels plug-in (Corretge, 2019). We then used Praat's Burg linear predictive coding algorithm (settings: maximum formant value = 5 kHz; time step = 0.01 s, and window length = 0.025 s). Formant frequencies were obtained for the first four formants of the five vowels. P_f was calculated by averaging standardized measures of F_1-F_4 (formants standardized using within-sex means and standard deviations; Puts et al., 2011). Apparent *VTL* was based on an algorithm provided by Stevens (1998),[1] which is derived from modeling the vocal tract as a uniform tube that is closed at one end (i.e., the vocal folds) and open at the other (i.e., the mouth). See Table 1.

Self-reported health perceptions. Participants completed the 32-item Health Perceptions Questionnaire, Form II (HPQ II; Ware, 1976), which is a frequently used measure of physical health. The HPQ II measures eight facets of perceived health; however, only 3 of these subscales tapped physical health: (1) Current health status, (2) prior health status, and (3) disease resistance. The other 5 subscales (health outlook, worry about health, sick orientation, rejecting the sick role, and attitude toward doctor visits) were oriented toward respondents' personalities and attitudes surrounding health. Previous research has shown that the scales are reliable, valid, and stable over time and across populations (Ware, 1976).

Selection and description of stimuli. We elected to sort voice recordings according to participants' mean scale scores for the *current health status* scale of the HPQ II. The prior health and disease resistance subscales could provide theoretically meaningful ways to test our hypotheses; however, these subscales were less reliable, having lower internal consistency (Prior health $\alpha = 0.53$ and Resistance to disease $\alpha = 0.75$, compared with current health $\alpha = 0.87$), had fewer items (Prior health = 3 and Resistance to disease = 4, compared with current health = 9), and were more dependent on participants' imperfect recollections. Current health status correlated significantly with perceived disease resistance (r (108) = 0.33, $p < .001$), but not prior health (r (108) = 0.11, $p = .26$).

The recordings of the 20 men who scored the highest on current health status and the 20 men who scored the lowest on current health status were inspected for recording quality. After recordings with any background noise were excluded,[2] recordings of the men with the 10 highest and 10 lowest mean scale scores on the current health status scale were selected for use as stimuli. Please see Table 1 for a description of the subjects' acoustic parameters. To ensure equal loudness for stimulus presentation, we digitally manipulated the stimuli to equalize the Root Mean Square (RMS) intensity for all recordings. The average intensity of presentation for the recordings was 58.56 dB (SD = 0.97).

Table 1
Descriptive statistics for voices of men who reported the highest and lowest level of current health status.

Voice Type		N	Min.	Max.	Mean	SD
High Current Health Status	F1	10	435.83	731.67	506.76	88.45
	F2	10	1543.49	1845.30	1667.19	95.40
	F3	10	2445.51	2914.28	2660.03	124.88
	F4	10	3327.31	3837.62	3519.02	168.01
	f_o	10	80.73	129.82	101.22	9.43
	f_o-SD	10	5.32	12.37	9.43	2.47
	D_f	10	963.83	1035.32	1004.09	26.52
	P_f	10	−1.08	2.09	−9.30E-17	0.93
	VTL	10	14.59	17.42	16.34	0.83
Low Current Health Status	F1	10	457.91	516.21	488.52	20.24
	F2	10	1598.07	1738.03	1638.69	45.16
	F3	10	2533.60	2783.71	2649.24	78.64
	F4	10	3384.46	3807.01	3559.80	138.54
	f_o	10	88.15	174.71	114.20	14.54
	f_o-SD	10	6.75	34.31	14.54	9.45
	D_f	10	975.52	1096.93	1023.76	39.43
	P_f	10	−0.85	1.77	9.99E-17	0.79
	VTL	10	15.39	16.91	16.34	0.48

6.1.3. Procedure

The study was approved by the Boston University Institutional Review Board and was conducted online. Previous investigations have demonstrated that online voice rating studies produce findings consistent with those conducted in laboratory settings (e.g., Jones et al., 2008). Participants were instructed to go to a quiet room before beginning the study and wear headphones for the voice rating portion. They completed six blocks of 20 trials of the voice rating task. For each block, participants listened to each of the 20 voices and rated each using a 9-point Likert-type rating scale (e.g. 1 = very unhealthy; 9 = very healthy). In order to gain a more comprehensive understanding of their health perceptions for each speaker, we elected to have listeners' rate speakers on what we perceived to be distinct but related facets of health. Participants were asked to rate how healthy each speaker sounded; how likely each speaker was to get a disease; how frequently each speaker got sick; and how severe the speaker's symptoms were when they did get sick. We expected that general health ratings likely represent listeners' perception of the overall health of the speaker and paralleled the HPQ current health status subscale. Ratings of disease likelihood and illness frequency may tap into perceptions of the vulnerability of speakers' immune system to initial infection and parallel the HPQ disease resistance and prior health subscales. We added an additional question that was not represented by the HPQ; listeners' symptom severity ratings may reflect the perceived strength of the immune system at fighting infection when already sick.

Heterosexual women also rated the attractiveness of each speaker for both short-term and long-term romantic relationships (a short-term romantic relationship was defined as a single date, one-night stand or a brief affair and a long-term relationship as steady dating or marriage). We elected to have female listeners rate both short-term and long-term attractiveness of speakers because previous research has demonstrated that relationship context affects women's preferences for masculine characteristics including voices (e.g., Feinberg et al., 2012), such that women may perceive more masculine males to be less likely to invest in long-term relationships and more likely to engage in infidelity (O'Connor, Pisanski, Tigue, Fraccaro, & Feinberg, 2014; O'Connor, Re, & Feinberg, 2011). All men rated physical and social dominance. A physically dominant man was defined as one who would probably win a fistfight against the average man, whereas a socially dominant man was defined as someone who tells other people what to do, is respected, influential, and often a leader (Puts et al., 2006). Both the presentation of all voice recordings within each block and the block presentation order was fully randomized.

6.2. Data screening

Listeners' ratings of each voice recording across all conditions were included as the dependent variable for the study. We then computed the Mahalanobis distance statistic to screen for multivariate outliers, which could bias the results of our multivariate analysis. All assumptions for multivariate analysis (i.e., multi-collinearity, normality, linearity, and homogeneity of variance) were met.

[1] *Note.* The algorithm for computing apparent *VTL* by Stevens (1998) is correlated with the measure of apparent *VTL* created by Reby and McComb (2003) and used by Pisanski et al. (2014) at $r = 0.80$. The results of Study 2, in which apparent VTL is used as a predictor and an outcome variable, are unaffected by choice of apparent *VTL* algorithm.

[2] *Note:* this results in the removal of $n = 7$ from the men who reported having relatively good health and $n = 7$ from the men who reported having relatively poor health.

6.3. Analytic plan

Analysis was conducted in R (version 3.6.2; R core team). We used the *lmer* function from the package, *lme4* (Bates et al., 2015) to conduct mixed effects linear regressions with Maximum Likelihood estimation. For all regressions, listener and speaker identity were random effects. The category of the voice, higher or lower self-reported health, was the fixed effect. Participants' ratings of speakers' health, disease resistance, illness frequency and symptom severity were the dependent variables in four separate linear mixed effects models. We conducted two linear mixed effects models in which women's attractiveness ratings served as the dependent variable and two linear mixed effects models in which men's dominance ratings served as the dependent variable. To assess the amount of variance explained by our fixed and random effects, we computed Pseudo R^2 using the *r.squaredGLMM* function of from the Multi-Model Inference (*MuMIn*) package (Barton & Barton, 2019).

6.4. Results

The fixed effect of health status approached but did not achieve conventional levels of statistical significance for listeners' health (b = 0.42, SE = 0.23, t = 1.84, p = .08), disease likelihood, (b = 0.33, SE = 0.17 t = 1.93, p = .07) and illness frequency ratings, (b = 0.36, SE = 0.19 t = 1.86, p = .08). In contrast, listeners' symptom severity ratings did not approach conventional levels of statistical significance (b = 0.22, SE = 0.16 t = 1.37, p = .19). Health status did not significantly predict women's ratings of speakers' attractiveness for either long-term (b = 0.51, SE = 0.32 t = 1.60, p = .13) or short-term relationships (b = 0.42, SE = 0.25, t = 1.70, p = .11). However, men assigned significantly higher physical (b = 0.86, SE = 0.25, t = 3.49, p < .001) and social dominance ratings (b = 0.62, SE = 0.18, t = 3.45, p < .001) to voices of relatively healthier men. The total proportion of variance explained by both the fixed and random effects for our models range between 13.21% to 36.51%; however, when only the fixed effect of health status was considered, the total proportion of variance explained ranged between 0.34% and 4.83%. Please see Table S1 for a report of the Pseudo R^2 for the fixed and random effects in all models and Table S2 for the variance and standard deviations of the random effects.

6.5. Discussion

Men assigned greater physical and social dominance ratings to the voices of male speakers' in the highest decile of self-reported health, but listeners did not assign significantly higher health-based ratings to these voices. However, their ratings of health, disease likelihood and illness frequency approached conventional levels of statistical significance. Unexpectedly, women did not rate the voices of healthier men as sounding more attractive; however, men did rate the voices of men in the highest decile of health as sounding more dominant.

In Study 2, we improve upon and extend Study 1 in several ways. First, we use a larger sample of speakers (n = 108), allowing us to look for relationships between listener-rated health and self-reported health among a more representative sample of male speakers, rather than restricting our sample to the top and bottom decile of speakers. Second, given the large number of vocal samples, we sought to simplify the raters' task by reducing the dimensions on which they assessed the voices from 6 to 2. In Study 1, the rating categories (i.e., current health, disease likelihood, illness frequency, and symptom severity) represented different but overlapping components of health; however, in Study 2, we updated our rating categories to tap a state/trait distinction. Specifically, *current health* reflects whether the speaker is currently ill (i.e., a state description), whereas *disease resistance* reflects whether the speaker is vulnerable to illness in general (i.e., a trait description). Third we included additional acoustic measures, f_o-CV and HNR, in our analysis (Kreiman & Gerratt, 2003; Li & Jo, 2004). Finally, we included measures of speakers' self-reported health so that we could test for relationships between acoustic parameters and self-reported health (Fig. 1B). Previous investigations have found that listeners do not rely on jitter or shimmer when classifying a voice as pathological (Kreiman & Gerratt, 2003), thus we elected not to include measures of jitter and shimmer.

7. Study 2

7.1. Methods

7.1.1. Participants

One hundred fifty-six participants (80 females; aged 19 to 35, M_{age} = 24.21, SD_{age} = 2.44) were recruited in the same way as Study 1. Reported ethnicities were: Caucasian (54.5%), South Asian (21.2%), Multiple Ethnicities (9.9%), Black (9.6%), Asian (7.1%), Latin American (6.4%), South East Asian and Arab West/Asian (<1%).

7.1.2. Stimuli and materials

The stimuli for the study were the voice recordings of all 108 men from Arnocky et al. (2018). Please see Study 1 for a description of the sample, voice recording procedure, and acoustic analysis. The average intensity of presentation for the recordings was 56.53 dB (SD = 4.71).

7.1.3. Procedure

Recruitment procedures and participant instructions for completing the voice rating experiment were the same as in Study 1. Participants completed two blocks of 108 trials of the voice-rating task. For each block, participants listened to all 108 voices and rated them using a 9-point Likert-type scale. They assessed the current health and the disease resistance of each speaker. Prior to each block, participants were provided with definitions of current health and disease resistance. We defined a person with good current health as someone who "feels healthy and is judged to be healthy by doctors, meaning they are not currently ill", and a person with high disease resistance was described as someone who "has a body that resists illness well, so they get sick less often than other people". We generated these definitions based on our inspection of the item content of the current health and disease resistance subscales of the HPQ-II (Ware, 1976). Both the presentation of all voice recordings within each block and the block presentation order was fully randomized.

7.1.4. Data analysis

Data screening procedures were the same as in Study 1. We computed means for the respondent's ratings of current health and disease resistance for all 108 recordings. Because of the high correlation between listeners' mean current health and disease resistance (r(108) =0.89, p < .001), we computed the average of these variables and used it as a dependent variable in our regression analysis.

7.1.5. Analytic plan

We conducted a multiple linear regression in which we entered f_o, f_o-CV, HNR and apparent *VTL* as predictors for participants health-based ratings. We expected f_o, and f_o-CV to negatively predict health ratings, and apparent *VTL* and HNR to positively predict health ratings. Additionally, we conducted regressions evaluating the relationship between speakers' self-reported health with their vocal characteristics, as well as regressions evaluating relationships between speakers' self-reported health and listeners health estimates. For all regressions, we used the *lm* function in *stats* package.

7.2. Results

7.2.1. Speakers' f_o predicted listeners' health ratings

First, we entered the four acoustic variables (f_o, apparent *VTL*, f_o-CV, HNR) into a multiple regression as predictors of listeners' health-based ratings. Results of this linear regression (Table S3) indicated that f_o (b = − 0.006, β = − 0.22, SE = 0.002, t = −2.12, p = .04) was a significant negative predictor of listeners' health-based ratings (see Fig. 2). Whereas, neither f_o-CV($b = -0.50$, $\beta = 0.10$, SE $=0.56$, $t = 0.90$, $p = .37$), HNR ($b = -0.41$, $\beta = -0.08$, SE $= 0.58$, $t = -0.70$, $p = .49$) nor apparent *VTL* ($b = -0.04$, $\beta = -0.09$, SE $=0.05$, $t = -0.87$, $p = .38$), were significant predictors of listeners' health-based ratings.

7.2.2. Speakers' self-reported health did not predict acoustic characteristics

Next, we conducted four multiple regressions in which we tested if men's self-reported current health and their disease resistance predicted f_o, apparent *VTL*, f_o-CV, or HNR. Please see Models 10 to Model 17 in Table S3.

7.2.3. Speaker's self-reported health did not predict listeners' health ratings

Finally, we conducted a multiple linear regression in which men's self-reported current health and disease resistance were entered as predictors of listeners' ratings of speakers' health. Neither men's self-reported current health ($b = 0.07$, $\beta = 0.11$, SE $= 0.07$, $t = 0.96$, $p = .33$) nor self-reported disease resistance ($b = -0.07$, $\beta = -0.11$, SE $= 0.07$, $t = -0.95$, $p = .34$) predicted listeners' health ratings. The proportion of variance explained by the predictors was less than 1% (Table S3, Model 18).

7.3. Discussion

Overall, results from Study 2 do not support the hypothesis that listeners can judge speakers' health from men's voice recordings. Interestingly, listener-rated health was associated with lower f_o; however, these effects were relatively small. The relationship between speaker's f_o and self-reported disease resistance trended in the expected direction. The findings from Study 2 suggest that f_o is not a robust signal of health but may bias listeners to assign higher health to those with lower f_o. The primary purpose of Study 3 was to bolster the findings of Study 2 by investigating if experimentally manipulating f_o or apparent *VTL* affected listener ratings of speakers' health, attractiveness, or dominance.

8. Study 3

8.1. Methods

8.1.1. Participants

Eighty-seven participants (44 females) aged 19 to 54, $M_{age} = 32.95$, $SD_{age} = 7.60$) completed a voice rating task and a questionnaire. The study was administered via Qualtrics and participants were recruited through MTurk. Reported ethnicities were Caucasian (71.8%), South Asian (10.6%), Black (8.2%), Asian (7.1%), Latin American (1.2%), South East Asian (1.1%), and Multiple Ethnicities (<1%).

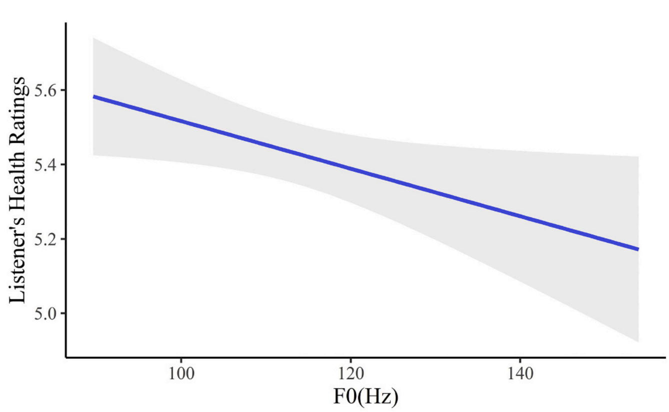

Fig. 2. Line graph depicting the relationship between listener's health ratings (y-axis) with speakers f_o (x-axis).

8.1.2. Stimuli/materials

Voice recordings were sorted according to participants' mean scores for the current health status subscale of the HPQ II. We selected the highest quality recordings from four men whose current health status scale scores fell closest to the median ($M_{\text{current health}} = 3.73$)

Using Praat's Pitch Synchronous Overlap and Add (*PSOLA*) algorithm (Boersma & Weenink, 2014), we created lowered and raised f_o versions of each of the four voice recordings of the men speaking the five monophthong vowel sounds. Lowered f_o versions of each recording were created by lowering the f_o of the original speaker's voice by 0.5 equivalent rectangular bandwidths (ERBs), or approximately 20 Hz. Raised f_o versions of each of the four voice recordings were created by raising the f_o of the original speaker's voice by 0.5 ERBs (Feinberg et al., 2005).

We used Praat to manipulate apparent *VTL* (cf., Feinberg et al., 2011). Two versions of each of the original voice recordings were created by altering the apparent *VTL* by 10%. We altered vocal tract qualities by resampling the sound by ±10% of the original rate as a way to raise or lower all frequencies by 10%. Then, we overrode the new sampling rate with the original sampling rate to restore the duration of the recording to its original value, while leaving all frequencies raised or lowered (Feinberg et al., 2011). This manipulation produced a significant difference in the third, $t(3) = -8.90$, $p = .003$, and the fourth, $t(3) = -3.62$, $p = .036$, formants, but not the first or second formants ($ps > 0.05$). The mean apparent *VTL* of voice recordings with increased *VTL* was 16.88 cm, whereas the mean apparent *VTL* of voice recordings with decreased *VTL* was 15.32 cm, $t(3) = 17.11$, $p < .001$. As in Study 1, we equalized the Root Mean Square (RMS) intensity for all recordings to ensure that they were presented with equal loudness. The average intensity of presentation for the recordings was 63.74 dB (SD = 0.07). See Table 2 for pitch and formant characteristics of manipulated voice

Table 2
Descriptive statistics men's voice recordings with lowered and raised f_o and apparent VTL.

		N	Min.	Max.	Mean	SD
Masculinized	F1	4	506.72	556.89	531.730	21.31
	F2	4	1405.90	1536.80	1484.16	58.77
	F3	4	2546.80	2614.70	2584.27	34.43
	F4	4	3295.50	4030.60	3639.40	328.33
	f_o	4	78.00	119.00	94.41	10.05
	f_o-*SD*	4	6.77	13.51	10.05	2.80
	D_f	4	923.69	1163.91	1035.89	110.587
	P_f	4	−0.35	0.72	6.52E-16	0.49
	VTL	4	15.61	17.22	16.55	0.75
Feminized	F1	4	494.08	545.93	529.76	24.35
	F2	4	1409.74	1543.45	1488.89	63.67
	F3	4	2530.30	2601.80	2571.36	35.12
	F4	4	3119.00	4328.90	3616.00	533.62
	f_o	4	84.14	166.96	130.94	14.56
	f_o-*SD*	4	12.50	18.10	14.56	2.66
	D_f	4	857.69	1261.40	1028.75	178.44
	P_f	4	−0.40	0.76	−1.87E-15	0.52
	VTL	4	15.10	17.90	16.65	1.18
Increased *VTL*	F1	4	476.52	568.13	522.97	40.99
	F2	4	1371.19	1675.40	1514.93	157.78
	F3	4	2413.70	2565.30	2486.48	67.19
	F4	4	3231.70	4000.20	3542.28	328.42
	f_o	4	80.20	164.12	102.48	13.43
	f_o-*SD*	4	7.94	21.14	13.43	5.98
	D_f	4	887.86	1165.71	1006.43	118.47
	P_f	4	−0.24	0.40	- 3.47E-16	0.30
	VTL	4	16.39	17.15	16.88	0.34
Reduced *VTL*	F1	4	537.93	576.22	566.14	18.81
	F2	4	1549.43	1663.50	1605.230	51.67
	F3	4	2804.90	2858.50	2831.43	22.37
	F4	4	3745.50	4074.50	3885.03	153.42
	f_o	4	93.11	153.71	124.14	13.97
	f_o-*SD*	4	11.27	18.88	13.97	3.59
	D_f	4	1056.76	1166.09	1106.29	52.94
	P_f	4	−0.71	0.78	−2.03E-15	0.67
	VTL	4	14.87	15.60	15.320	0.32

stimuli.

8.1.3. Procedure

After reporting age, ethnicity, sex, sexual orientation, and relationship status, participants listened to the 16 voices one at a time and rated how healthy each of the speakers sounded, how likely each speaker was to get a disease, how frequently each speaker got sick, and how severe the symptoms of each speaker were when they did get sick. Heterosexual women rated the attractiveness of each speaker for a short-term and long-term romantic relationship. All men, regardless of sexual orientation, rated physical and social dominance.

8.1.4. Data screening

Data screening procedures were the same as Study 1. Due to an experimenter error, the incorrect question was presented to listeners for two of the trials within the block in which listeners rated speakers' disease likelihood and were therefore excluded. These two trials were both of men's voice recordings with raised f_o.

8.1.5. Analytic plan

The analytic plan was like Study 1. We conducted mixed effects linear regressions using *lmer* function from the package, *lme4* (Bates et al., 2015). We used listener and speaker identity as random effects in all models. Voice manipulation (either manipulated f_o or apparent VTL)

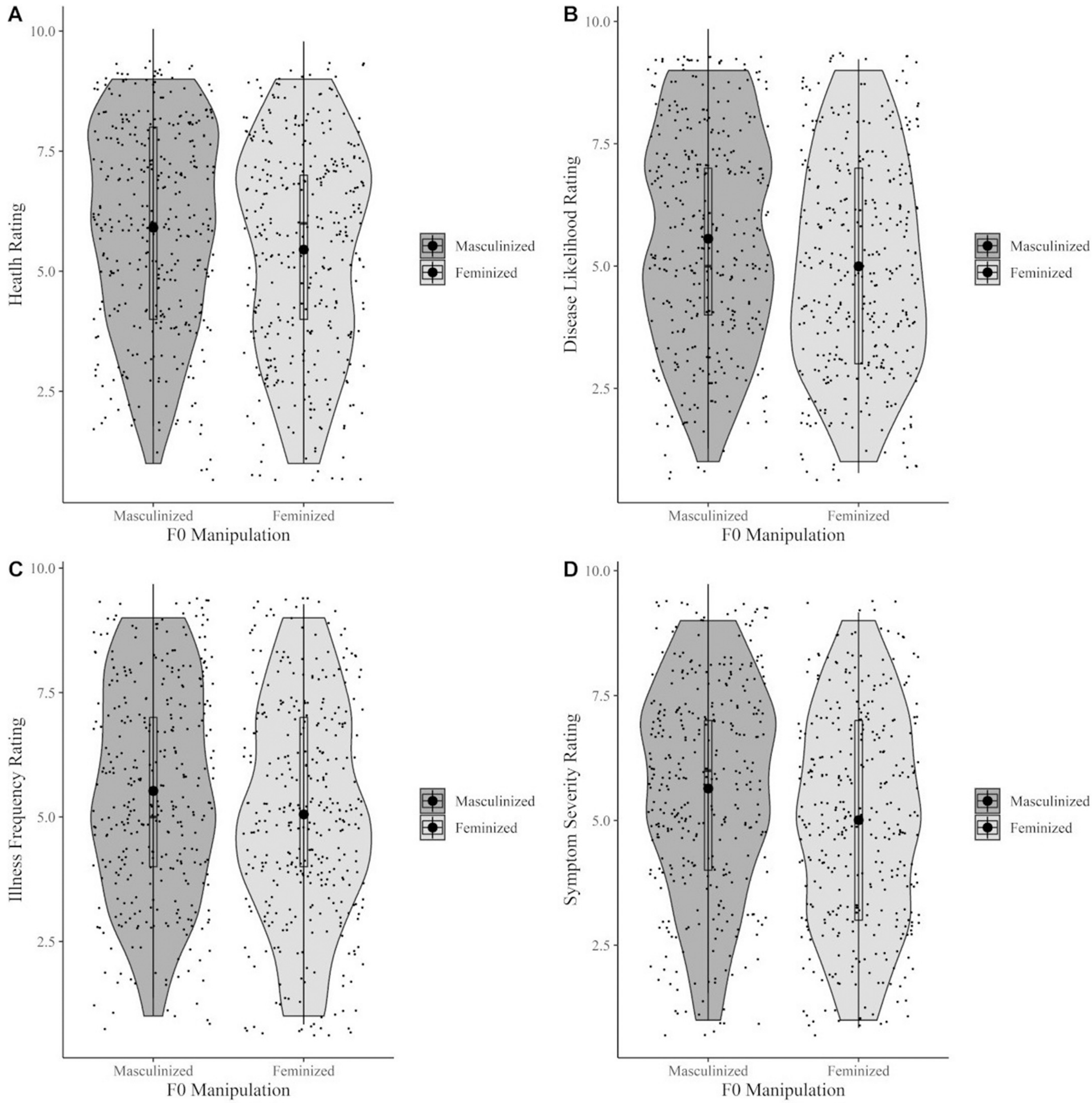

Fig. 3. Violin plot with nested boxplot, comparing listeners' health-based ratings for men's voices with raised and lowered fo.

was the fixed effect. Participants' ratings of speakers' health, disease resistance, illness frequency and symptom severity were the dependent variables. We conducted four linear mixed effects models testing whether f_o or apparent *VTL* manipulation effected heterosexual women's attractiveness ratings. Another four linear mixed effects models testing whether these manipulations effected men's dominance ratings. We computed Pseudo R^2 in the same way as in Study 1.

8.2. Results

8.2.1. f_o manipulation

Listeners rated lowered f_o versions of the recordings as sounding healthier ($b = 0.46$, SE = 0.13, $t = 3.44$, $p < .001$), less likely to get a disease ($b = 0.56$, SE = 0.14, $t = 3.97$, $p < .001$), to experience illness less frequently ($b = 0.47$, SE = 0.14, $t = 3.50$, $p < .001$), and to experience less severe symptoms when ill ($b = 0.63$, SE = 0.15, $t = 4.35$, $p <$ .001) (see Fig. 3). In line with previous research (e.g., Feinberg, DeBruine, Jones, & Perrett, 2008; Puts, 2005), heterosexual women rated these voices as more attractive for both long-term ($b = 1.31$, SE = 0.45, $t = 2.89$, $p = .02$) and short-term relationships ($b = 1.19$, SE = 0.28, $t = 6.57$, $p = .02$). Men assigned higher physical ($b = 1.24$, SE = 0.43, $t = 2.89$, $p < .001$) and social dominance ($b = 1.04$, SE = 0.18, $t =$ 5.72, $p < .001$) scores to men's voices with lowered f_o, replicating the results of previous research (e.g., Puts et al., 2006; Puts et al., 2007). The total proportion of variance explained by both fixed and random effects for our models ranged from 19.09% to 39.76%; whereas when only the fixed effect of f_o manipulation was considered, the total proportion of variance explained ranged from 1.11% to 9.54%. Please see Table S4 for the Pseudo R^2 for fixed and random effects in all models and Table S5 for variance and standard deviations of random effects.

8.2.2. Apparent VTL manipulation

Listeners rated recordings with longer apparent VTL as sounding less likely to get a disease ($b = -0.40$, SE = 0.15, $t = -2.73$, $p = .01$). These recordings were not rated as sounding healthier ($b = -0.07$, SE = 0.15, $t = -0.45$, $p = .65$) or as belonging to individuals who experienced illness less frequently ($b = -0.09$, SE = 0.15, $t = -0.62$, $p = .45$) or experienced less severe symptoms when ill ($b = -0.27$, SE = 0.16, $t = -1.68$, $p = .09$). Women rated these recordings as sounding more attractive for both long-term ($b = 0.95$, SE = 0.37, $t = 2.56$, $p = .03$) and short-term ($b = 1.33$, SE = 0.46, $t = 2.88$, $p = .02$) relationships. Men assigned significantly higher physical ($b = 2.02$, SE = 0.21, $t = 9.69$, $p <$ .001) and social dominance ($b = 1.44$, SE = 0.21, $t = 6.91$, $p < .001$) ratings to recordings of men's voices with longer apparent VTL, replicating the results of previous research. The total proportion of variance explained by both fixed and random effects for our models ranged from 4.57% to 38.54%; whereas when only the fixed effect of health status was considered, the total proportion of variance explained ranged from 0.04% to 17.59%. Please see Table S6 for a report of Pseudo R^2 for fixed and random effects in all models and Table S7 for variance and standard deviations of the random effects.

8.3. Discussion

We found that manipulating f_o affected listeners' perceptions of health, such that they assigned higher health-based ratings to lowered voice recordings. In line with previous investigations, heterosexual women rated men's voices with lowered f_o as more attractive. Men rated men's voices with lowered f_o as sounding more dominant. The same pattern did not emerge for the apparent *VTL* manipulation. Women rated recordings of men's voices with increased apparent *VTL* as more attractive, and men rated these recordings as sounding more dominant.

9. General discussion

Researchers have posited that women's preferences for men's lower-pitched voices reflect a preference for mates in good health (e.g., Feinberg, 2008; Jones et al., 2010), whereas men use the voices of other men to assess speakers' condition, and therefore threat potential (Sell et al., 2010; Hodges-Simeon et al., 2013; Hodges-Simeon et al., 2015; Puts, et al., 2012). However, to our knowledge, no study has investigated if listeners can accurately assess speakers' health from their voices. This reflects a critical gap in the literature. Based on the ICHH (Folstad & Karter, 1992; Hamilton & Zuk, 1982; Kokko, 1997; Kokko et al., 2003; Muehlenbein & Bribiescas, 2005; Wedekind & Folstad, 1994), and direct-benefits explanations (Able, 1996; Tybur & Gangestad, 2011; Allaire, 1988; Altschuler & Dale, 1999; Drotar, 1994; Thore, 1990; Tybur & Gangestad, 2011; Roberts & Little, 2008), we would expect listeners' to be able to judge health accurately, and to use T-linked characteristics of speakers' voices in their assessments.

The results of the above investigation are mixed. In Study 1, listeners did not assign higher health ratings to voice recordings provided by men with higher current health. Moreover, women did not rate these men's recordings as sounding more attractive. However, men rated these voice recordings as sounding more dominant, which may suggest that male listeners perceive the voice of men with higher self-reported health as belonging to physically stronger men. This may suggest that listeners' dominance assessments may index the health and condition of the speaker; however, future research is required to test this hypothesis. Because our sample of speakers represented the extremes of self-reported health (i.e., the top and bottom 10%), we followed up with a second study using a larger, more representative sample of men. We tested for linear relationships between speakers' acoustic characteristics, listeners' health ratings, and speakers' self-reported current health and disease resistance ratings. Study 2 also failed to show a relationship between self- and listener-reported health. Together these findings suggest that the voice may not reveal information about actual health in young men, at least when listeners are asked to make health-based ratings. Men did, however, rate the voices of men with better self-reported health as sounding more dominant, suggesting that male listeners' may base their dominance assessments on speakers' condition.

Our second aim was to assess the relationship between speakers' self-reported health and various acoustic dimensions of their voice (see Fig. 1B). Although not statistically significant, the relationship between men's f_o and self-reported disease resistance trended in the predicted direction, such that those with lower f_o reported higher resistance to disease. We found no other association between speakers' voice parameters and their health. Our third aim was to assess if listeners use aspects of the voice in their perceptions of health (see Fig. 1C). Both Study 2 (using a correlational design) and Study 3 (using an experimental design) showed that lower f_o was significantly associated with health ratings. We found a significant negative relationship between speakers f_o and listeners' health ratings, suggesting that men with lower-than-average f_o are perceived by listeners as healthier; however, the portion of variance in listeners' ratings of health explained by f_o was small and future investigations will be needed to replicate this finding. Longer apparent VTL, lower f_o-CV and higher HNR did not predict higher health ratings. Thus, while both formants and f_o affect dominance and attractiveness judgments, only f_o appears to consistently affect health judgments.

Why f_o—but not *VTL*—should affect health perceptions is unclear. While an "attractiveness halo" for f_o may partially explain these findings, the lack of relationship between *VTL* and health assessments—given its effects on attractiveness and dominance—requires additional explanation. Signaling theory predicts that selection should act not only on signalers, but also on receivers to detect, on average, fitness-relevant information from conspecifics (Grafen, 1990; Laidre & Johnstone, 2013; Maynard-Smith, J.,& Price, G. R., 1973; Searcy & Nowicki, 2005; Zahavi, 1975, 1977). The source-filter theory of speech production (Fant, 1960) suggests that the vocal folds and tract length are largely independent and, by extension, may provide different types of information to receivers. The vocal folds (and therefore f_o) are more closely associated with T than is the vocal tract (and therefore the formants; Markova et al., 2016). For instance, whereas exogenous T therapy partially affects vocal tract resonances among transgender males, f_o shows almost complete masculinization (Hodges-Simeon et al., 2020). Further, f_o is more sexually dimorphic in humans than in other extant apes (Puts et al., 2016) and, in humans, is more strongly affected by aging than are formants (Eichhorn, Kent, Austin, & Vorperian, 2018; Reubold, Harrington, & Kleber, 2010).

Estimates of *VTL*, on the other hand, more closely map onto body size. Because the sex difference in stature is between 7 and 8% (Gaulin & Boster, 1985; Gray & Wolfe, 1980), the pubertal descent of the larynx likely contributes to the approximately 15% difference between adult males and females in formant frequencies (Fant, 1960; Fitch & Giedd, 1999). Formant-based measures have been inconsistently related to speakers' height (Bruckert et al., 2006; Pisanski et al., 2014; Pisanski & Bryant, 2016) and weight (Evans, Neave, & Wakelin, 2006; Pisanski & Bryant, 2016), but are better predictors of variation in speaker body-size than is f_o (Pisanski et al., 2014). However, formant-based measures explain only approximately 10% of the variation in height within sexes (Pisanski et al., 2014). Additional variation may be explained by T exposure during development (Hodges-Simeon et al., 2013; Hodges-Simeon et al., 2015), though more research is needed.

10. Limitations and future directions

First, we used one self-reported measure, current health (Ware, 1976), with several subscales, to characterize speakers' physical health. It is unclear, however, whether speakers' self-reported health is a reliable measure of their actual health, or if health can be quantified on a single continuum. Error may be introduced during memory retrieval processes and/or with self-presentation bias. Furthermore, men may be inaccurate at knowing or reporting their actual physical health, and future investigations may utilize biomarkers of immunity (e.g., Arnocky et al., 2018). Previous research has provided evidence of a link between masculine morphology in men and their immunocompetence (Arnocky et al., 2018; Foo et al., 2020; Phalane et al., 2017; Rantala et al., 2012; Skrinda et al., 2014; Zaidi et al., 2019). Although earlier investigations have relied on a single measure of immunity (Rantala et al., 2012; Skrinda et al., 2014) when exploring relationships between secondary sexual characteristics and health, recent work has begun to use multiple measures (Foo et al., 2020). For example, Foo et al. (2020) found that aspects of adolescent males' antibacterial and cell-mediated immunity predicted 3D measures of their facial masculinity in adulthood. Like Foo et al., future investigations testing for relationships between sexually dimorphic vocal parameters and measures of health could benefit from obtaining multiple objective measures of physical health and grouping them using dimension reduction techniques, such as principle component analysis, before testing whether these components predict sexual dimorphic vocal parameters.

Future studies would also benefit from utilizing additional theoretically informed measures. For example, one of our subscales, perceived disease resistance, should be a better means of testing the hypotheses generated from the ICHH than other subscales, because it may reflect participants' overall perception of the strength of their immune system. Study 2 showed a marginally significant association between f_o and disease resistance. The failure of this relationship to reach statistical significance suggests that our sample may be under-powered to detect this (likely small) effect, and future investigations should therefore gather data from larger samples.

Second, the ICHH posits that, because T is immunosuppressive, only males whose immune system is strong enough to withstand its immune-suppressive effects can develop robust secondary sexual characteristics. Critically, in order to test ICHH-based hypotheses, it is necessary to quantify health during sexual maturation and then to measure variation in secondary sexual characteristics in adulthood. Currently, most studies testing the ICHH quantify secondary sexual characteristics and immune function following sexual maturation (e.g., Arnocky et al., 2018, Skrinda et al., 2014; Rantala et al., 2012; but see Foo et al., 2020; Hodges-Simeon et al., 2015; and Hodges-Simeon et al., 2020 for exceptions). Therefore, future investigations should use a longitudinal design and measure immune function in adolescence and sexually dimorphic vocal parameters with the attainment of sexual maturation.

Third, and related to the previous two limitations, both listeners and speakers were drawn exclusively from a healthy, WEIRD (Western, educated, industrialized, rich and democratic) population (Gurven & Lieberman, 2020; Henrich et al., 2010). Individuals from WEIRD populations are in better health than individuals living in environments with high pathogen loads and little to no access to modern medicine or health infrastructure (DeBruine, Jones, Crawford, Welling, & Little, 2010; Gangestad & Buss, 1993; Gurven, Kaplan, Winking, Finch, & Crimmins, 2008; Gurven & Lieberman, 2020). In addition, it may be the case that, after a sufficient amount of energy is allocated to immune function, the remainder can be allocated to signal quality, obscuring the link between signal quality and condition in populations with low immunological burden (Kokko, 1997). Therefore, future investigations should obtain voice recordings, health, and hormonal datafrom speakers living in a wide range of environments (e.g., Hodges-Simeon et al., 2015; Hodges-Simeon et al., 2019). For example, women from WEIRD populations may be less sensitive to cues of health in prospective mates than women living in environments with high pathogen load. Indeed, women living in environments with higher pathogen loads rate masculinized faces and voices as more attractive than women living in environments with access to modern medicine (Penton-Voak, Jacobson, & Trivers, 2004; Little, Apicella, & Marlowe, 2007; but see Brooks et al., 2011 and Scott et al., 2014).

Fourth, in future investigations we will seek to improve upon the stimuli in several ways. Using standardized sentences, such as the rainbow passage (Fant, 1960), rather than vowels, may provide listeners with more acoustic information, which they can use to make their health assessments. In the current investigation, speakers sat 20 cm from a Samson Meteorite USB Condenser microphone when making their recordings. In order to standardize speakers' distance in future investigations, we will use a head-mounted microphone. Although the apparent VTL shift of 10% has been demonstrated to be perceptible (the just-noticeable-difference of formant shifts is around 6%; Pisanski & Rendall, 2011), this resulted in a 1.5 cm apparent VTL change, which is within the average length for young men and longer than the average length for young women (Fitch & Giedd, 1999). It is possible that our manipulation of apparent *VTL* may have been too subtle to produce an effect on health ratings. Future studies should manipulate f_o and VTL by similar perceptual amounts (see Puts et al., 2007)

Fifth, we examined only four acoustic parameters and experimentally manipulated only two. Future investigations may benefit by covering a broader set of acoustic variables. Just as health and immunocompetence are multidimensional and difficult to capture using single measures, vocal indicators of health may be many. A wide variety of illness and disease symptoms include dysphonia, the catchall term for auditory-perceptual symptoms of voice disorders (Chan, 2010; Heman-Ackah et al., 2002; Li & Jo, 2004; Millqvist et al., 2008; Parsa & Jamieson, 2001). Acoustic measurements are just one component of dysphonia (American Speech-Language-Hearing Association (ASHA), 2020), but offer a desirable level of objectivity. In addition to those parameters used in the present research, speech rate, jitter, shimmer, spectral tilt and cepstral peak prominence (CPP) have been used in dysphonia indices (e.g., Maryn, De Bodt, Barsties, & Roy, 2014; Wuyts et al., 2000), and often worsen with age and poor health. For example, speakers who are perceived as old produce speech at a significantly slower rate (Harnsberger et al., 2008) and speak with increased levels of jitter and shimmer (Linville, 1996; Orlikoff, 1990). Moreover, older men with chronic disease, such as atherosclerosis, have a greater range of jitter values than healthy older men (Orlikoff, 1990). CPP, a measure of relative amplitude, has been demonstrated to be a useful acoustic parameter for the classification of pathological voices (Castellana, Carullo, Corbellini, & Astolfi, 2018), and may be more reliable than jitter and shimmer (Heman-Ackah et al., 2003).

Sixth, the obtained rating patterns could reflect a halo effect. Past research has demonstrated that listeners' rate men's voices manipulated to have lower f_o as sounding larger, more masculine, and more attractive (Feinberg et al., 2005; Pisanski, Mishra, & Rendall, 2012; Pisanski & Rendall, 2011). Listeners' ratings of speakers' health, disease likelihood, illness frequency and symptom severity, may not be independent from each other or men's ratings of speakers' dominance or women's ratings of speakers' attractiveness (Pisanski et al., 2012; Pisanski & Rendall, 2011). Future investigations should attempt to disentangle the extent to which listeners' health ratings of male speakers' voices reflect a tendency to assign positive characteristics to voices with lower than average f_o.

11. Conclusions

Many researchers have assumed that low-pitched men's voices are judged to be attractive and dominant because they cue the speakers' superior condition. However, to our knowledge no study has assessed whether individuals can accurately assess speakers' health from their voices. Overall, our results suggest that individuals may rely on f_o of voices when judging speakers' health, even though f_o does not reliably cue the self-reported health of speakers. This research adds to vocal attractiveness and dominance research by demonstrating that voices with experimentally lowered f_o are rated as sounding healthier. We replicate the finding that women rate these voices with lowered f_o as more attractive and men rate them as more dominant. Future investigations should move beyond a single self-reported measure of health toward a more comprehensive quantification of physical health. Researchers should continue to investigate listeners' abilities to assess health from speakers' voices in order to determine if good genes models are sufficient to explain these preferences.

References

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice, 13*(3), 424–446. https://doi.org/10.1016/S0892-1997(99)80048-4.

Able, D. J. (1996). The contagion indicator hypothesis for parasite-mediated sexual selection. *Proceedings of the National Academy of Sciences, 93*(5), 2229–2233. https://doi.org/10.1073/pnas.93.5.2229.

Adelman, J. S., & Martin, L. B. (2009). Vertebrate sickness behaviors: Adaptive and integrated neuroendocrine immune responses. *Integrative and Comparative Biology, 49*(3), 202–214. https://doi.org/10.1093/icb/icp028.

Allaire, S. (1988). How a chronically ill mother manages. *The American Journal of Nursing, 88*, 47–49. https://doi.org/10.2307/3425826.

Alonso-Alvarez, C., Bertrand, S., Faivre, B., Chastel, O., & Sorci, G. (2007). Testosterone and oxidative stress: The oxidation handicap hypothesis. *Proceedings of the Royal Society B: Biological Sciences, 274*(1611), 819–825. https://doi.org/10.1098/rspb.2006.3764.

Altschuler, J., & Dale, B. (1999). On Being an Ill Parent. *Clinical Child Psychology and Psychiatry*, 23–37. https://doi.org/10.1177/1359104599004001003.

American Speech-Language-Hearing Association (ASHA). (2020). Voice disorders. Retrieved from https://www.asha.org/Practice-Portal/Clinical-Topics/Voice-Disorders/.

Arnocky, S., Hodges-Simeon, C., Ouellette, D., & Albert, G. (2018). Do men with more masculine voices have better immunocompetence? *Evolution and Human Behavior, 39*(6), 602–610. https://doi.org/10.1016/j.evolhumbehav.2018.06.003.

Aung, T., & Puts, D. (2020). Voice pitch: A window into the communication of social power. *Current Opinion in Psychology, 33*, 154–161. https://doi.org/10.1016/j.copsyc.2019.07.028.

Awan, S. N. (2006). The aging female voice: Acoustic and respiratory data. *Clinical Linguistics & Phonetics, 20*(2–3), 171–180. https://doi.org/10.1080/02699200400026918.

Babel, M., McGuire, G., & King, J. (2014). Towards a more nuanced view of vocal attractiveness. *PloS one, 9*(2), e88616. https://doi.org/10.1371/journal.pone.0088616.

Barton, K., & Barton, M. K. (2019). Package "MuMIn". *R Package Version, 1*(6).

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … Bolker, M. B. (2015). Package "lme4". *Convergence, 12*(1), 2.

Boersma, P., & Weenink, D. (2014). *Praat: Doing phonetics by computer [Computer program]. Version 5.4.22*.

Booth, A., Johnson, D. R., & Granger, D. A. (1999). Testosterone and men's health. *Journal of Behavioral Medicine, 22*(1), 1–19. https://doi.org/10.1023/A:1018705001117.

Brooks, R., Scott, I. M., Maklakov, A. A., Kasumovic, M. M., Clark, A. P., & Penton-Voak, I. S. (2011). National income inequality predicts women's preferences for masculinized faces better than health does. *Proceedings of the Royal Society B: Biological Sciences, 278*(1707), 810–812. https://doi.org/10.1098/rspb.2010.0964.

Bruckert, L., Liénard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society of London B: Biological Sciences, 273*(1582), 83–89. https://doi.org/10.1098/rspb.2005.3265.

Bulygina, E., Mitteroecker, P., & Aiello, L. (2006). Ontogeny of facial dimorphism and patterns of individual development within one human population. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 131*(3), 432–443. https://doi.org/10.1002/ajpa.20317.

Butler, G. E., Walker, R. F., Walker, R. V., Teague, P., Riad-Fahmy, D., & Ratcliffe, S. G. (1989). Salivary testosterone levels and the progress of puberty in the normal boy. *Clinical Endocrinology, 30*(5), 587–596. https://doi.org/10.1111/j.1365-2265.1989.tb01431.x.

Cartei, V., Bond, R., & 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), 569–576. https://doi.org/10.1016/j.yhbeh.2014.08.006.

Casagrande, S., & Groothuis, T. G. (2011). The interplay between gonadal steroids and immune defence in affecting a carotenoid-dependent trait. *Behavioral Ecology and Sociobiology, 65*(10), 2007–2019. https://doi.org/10.1007/s00265-011-1210-6.

Castellana, A., Carullo, A., Corbellini, S., & Astolfi, A. (2018). Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel. *IEEE Transactions on Instrumentation and Measurement, 67*(3), 646–654. https://doi.org/10.1109/TIM.2017.2781958.

Chan, T. V. (2010). The patient with sore throat. *Medical Clinics, 94*(5), 923–943. https://doi.org/10.1016/j.mcna.2010.06.001.

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour, 60*(6), 773–780. https://doi.org/10.1006/anbe.2000.1523.

Corretge, R. (2019). Praat vocal toolkit. http://www.praatvocaltoolkit.com.

Cox, R. M., & John-Alder, H. B. (2007). Increased mite parasitism as a cost of testosterone in male striped plateau lizards Sceloporus virgatus. *Functional Ecology, 21*(2), 327–334. https://doi.org/10.1111/j.1365-2435.2007.01251.x.

Dabbs, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences, 27*(4), 801–804. https://doi.org/10.1017/S0191-8869(98)00272-4.

DeBruine, L. M., Jones, B. C., Crawford, J. R., Welling, L. L., & 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), 2405–2410. https://doi.org/10.1098/rspb.2009.2184.

Deviche, P., & Cortez, L. (2005). Androgen control of immunocompetence in the male house finch, Carpodacus mexicanus Müller. *Journal of Experimental Biology, 208*(7), 1287–1295. https://doi.org/10.1242/jeb.01531.

Dibazar, A. A., Berger, T. W., & Narayanan, S. S. (2006). Pathological voice assessment. *International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, 2006*, 1669–1673. https://doi.org/10.1109/IEMBS.2006.259835.

Drotar, D. (1994). Impact of parental health problems on children: Concepts, methods, and unanswered questions. *Journal of Pediatric Psychology, 19*(5), 525–536. https://doi.org/10.1093/jpepsy/19.5.525.

Edler, R., Goymann, W., Schwabl, I., & Friedl, T. W. (2011). Experimentally elevated testosterone levels enhance courtship behaviour and territoriality but depress acquired immune response in red bishops Euplectes orix. *Ibis, 153*(1), 46–58. https://doi.org/10.1111/j.1474-919X.2010.01075.x.

Eguchi, S. (1969). Development of speech sounds in children. *Acta Oto-Laryngologica, 257*, 1–51.

Eichhorn, J. T., Kent, R. D., Austin, D., & Vorperian, H. K. (2018). Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice, 32*(5), 644–e1. https://doi.org/10.1016/j.jvoice.2017.08.003.

Evans, M. R., Goldsmith, A. R., & Norris, S. R. (2000). The effects of testosterone on antibody production and plumage coloration in male house sparrows (Passer domesticus). *Behavioral Ecology and Sociobiology, 47*(3), 156–163. https://doi.org/10.1007/s002650050006.

Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. *Biological Psychology, 72*(2), 160–163. https://doi.org/10.1016/j.biopsycho.2005.09.003.

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior, 93*(4–5), 783–788. https://doi.org/10.1016/j.physbeh.2007.11.033.

Fant, G. (1960). *Acoustic theory of speech production*. Hague: Mouton & Co.

Feinberg, D. R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary Anthropology, 17*(2), 112–118. https://doi.org/10.1002/evan.20166.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., Little, A. C., O'Connor, J. J., & Tigue, C. 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), 413–418. https://doi.org/10.1007/s00265-011-1287-y.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception, 37*(4), 615–623. https://doi.org/10.1068/p5514.

Feinberg, D. R., Jones, B. C., & Armstrong, M. M. (2018). Sensory exploitation, sexual dimorphism, and human voice pitch. *Trends in Ecology & Evolution, 33*(12), 901–903. https://doi.org/10.1016/j.tree.2018.09.007.

Feinberg, D. R., Jones, B. C., & Armstrong, M. M. (2019). No evidence that men's voice pitch signals formidability. *Trends in Ecology & Evolution, 34*(3), 190–192. https://doi.org/10.1016/j.tree.2018.12.014.

Feinberg, D. R., Jones, B. C., DeBruine, L. M., O'connor, J. J. M., Tigue, C. C., & Borak, D. J. (2011). Integrating fundamental and formant frequencies in women's preferences for men's voices. *Behavioral Ecology, 22*(6), 1320–1325. https://doi.org/10.1093/beheco/arr134.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69*(3), 561–568. https://doi.org/10.1016/j.anbehav.2004.06.012.

Fink, B., Grammer, K., & Matts, P. J. (2006). Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior, 27*(6), 433–442. https://doi.org/10.1016/j.evolhumbehav.2006.08.007.

Fink, B., Grammer, K., & Thornhill, R. (2001). Human (Homo sapiens) facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology, 115*(1), 92–99. https://doi.org/10.1037/0735-7036.115.1.92.

Fisher, R. A. (1958). *The genetical theory of natural selection* (2d ed.). New York: Dover Publications.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 106*(3), 1511–1522. https://doi.org/10.1121/1.427148.

Folstad, I., & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *American Naturalist, 139*(3), 603–622. https://doi.org/10.2307/2462500.

Foo, Y. Z., Simmons, L. W., Perrett, D. I., Holt, P. G., Eastwood, P. R., & Rhodes, G. (2020). Immune function during early adolescence positively predicts adult facial sexual dimorphism in both men and women. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2020.02.002.

Foo, Y. Z., Simmons, L. W., & Rhodes, G. (2017). Predictors of facial attractiveness and health in humans. *Scientific Reports, 7*, 39731. https://doi.org/10.1038/srep39731.

Fouquet, M., Pisanski, K., Mathevon, N., & 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. https://doi.org/10.1098/rsos.160395.

Furlow, B. F. (1997). Human neonatal cry quality as an honest signal of fitness. *Evolution and Human Behavior, 18*, 175–193. https://doi.org/10.1016/S1090-5138(97)00006-8.

Furman, D. (2014). Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination. *Proceedings of the National Academy of Sciences of the United States of America, 111*(2), 866–874. https://doi.org/10.1073/pnas.1321060111.

Fuxjager, M. J., Foufopoulos, J., Diaz-Uriarte, R., & Marler, C. A. (2011). Functionally opposing effects of testosterone on two different types of parasite: Implications for the immunocompetence handicap hypothesis. *Functional Ecology, 25*(1), 132–138. https://doi.org/10.1111/j.1365-2435.2010.01784.x.

Gangestad, S. W., & Buss, D. M. (1993). Pathogen prevalence and human mate preferences. *Ethology and Sociobiology, 14*, 89–96. https://doi.org/10.1016/0162-3095(93)90009-7.

Gangestad, S. W., Haselton, M. G., & Buss, D. M. (2006). Evolutionary foundations of cultural variation: Evoked culture and mate preferences. *Psychological Inquiry, 17*(2), 75–95. https://doi.org/10.1207/s15327965pli1702_1.

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences, 23*(4), 573–587.

Gaulin, S., & Boster, J. (1985). Cross-cultural differences in sexual dimorphism: Is there any variance to be explained? *Ethology and Sociobiology, 6*(4), 219–225. https://doi.org/10.1016/0162-3095(85)90014-7.

Gorham-Rowan, M. M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders, 39*(3), 171–184. https://doi.org/10.1016/j.jcomdis.2005.11.005.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology, 144*(4), 517–546. https://doi.org/10.1016/S0022-5193(05)80088-8.

Gray, J. P., & Wolfe, L. D. (1980). Height and sexual dimorphism of stature among human societies. *American Journal of Physical Anthropology, 53*(3), 441–456. https://doi.org/10.1002/ajpa.1330530314.

Gurven, M., Kaplan, H., Winking, J., Finch, C., & Crimmins, E. M. (2008). Aging and inflammation in two epidemiological worlds. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 63*(2), 196–199. https://doi.org/10.1093/gerona/63.2.196.

Gurven, M. D., & Lieberman, D. E. (2020). WEIRD bodies: Mismatch, medicine and missing diversity. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2020.04.001.

Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: A role for parasites? *Science, 218*(4570), 384–387. https://doi.org/10.1126/science.7123238.

Harnsberger, J. D., Shrivastav, R., Brown, W. S., Jr., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice, 22*(1), 58–69. https://doi.org/10.1016/j.jvoice.2006.07.004.

Harries, M., Hawkins, S., Hacking, J., & Hughes, I. (1998). Changes in the male voice at puberty: Vocal fold length and its relationship to the fundamental frequency of the voice. *The Journal of Laryngology & Otology, 112*(05), 451–454.

Harries, M. L. L., Walker, J. M., Williams, D. M., Hawkins, S., & Hughes, I. A. (1997). Changes in the male voice at puberty. *Archives of Disease in Childhood, 77*(5), 445–447. https://doi.org/10.1136/adc.77.5.445.

Heman-Ackah, Y. D., Michael, D. D., Baroody, M. M., Ostrowski, R., Hillenbrand, J., Heuer, R. J., … Sataloff, R. T. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *The Annals of Otology, Rhinology, and Laryngology, 112*(4), 324–333.

Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S., Jr. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16*(1), 20–27. https://doi.org/10.1016/S0892-1997(02)00067-X.

Henderson, A. J., Holzleitner, I. J., Talamas, S. N., & Perrett, D. I. (2016). Perception of health from facial cues. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 371*(1693). https://doi.org/10.1098/rstb.2015.0380.

Henderson, A. J., Lasselin, J., Lekander, M., Olsson, M. J., Powis, S. J., Axelsson, J., & Perrett, D. I. (2017). Skin colour changes during experimentally-induced sickness. *Brain, Behavior, and Immunity, 60*, 312–318. https://doi.org/10.1016/j.bbi.2016.11.008.

Henderson, J. J., & Anglin, J. M. (2003). Facial attractiveness predicts longevity. *Evolution and Human Behavior, 24*(5), 351–356. https://doi.org/10.1016/S1090-5138(03)00036-9.

Henrich, J., Heine, S. J., & Norenzayan, A (2010). Most people are not WEIRD. *Nature, 466*(7302), 29. https://doi.org/10.1038/466029a.

Hill, A. K., Cárdenas, R. A., Wheatley, J. R., Welling, L. L. M., Burriss, R. P., Claes, P., … Puts, D. A. (2017). Are there vocal cues to human developmental stability? Relationships between facial fluctuating asymmetry and voice attractiveness. *Evolution and Human Behavior, 38*(2), 249–258. https://doi.org/10.1016/j.evolhumbehav.2016.10.008.

Hodges-Simeon, C. R, Asif, S., Gurven, M., Blackwell, A. D., & Gaulin, S. J. (2019). Testosterone is positively and estradiol negatively associated with mucosal immunity in Amazonian adolescents. *American Journal of Human Biology, 31*(5), e23284. https://doi.org/10.1002/ajhb.23284.

Hodges-Simeon, C. R., Gaulin, S. J., & Puts, D. A. (2011). Voice correlates of mating success in men: Examining "contests" versus "mate choice" modes of sexual selection. *Archives of Sexual Behavior, 40*(3), 551–557. https://doi.org/10.1007/s10508-010-9625-0.

Hodges-Simeon, C. R., Grail, G., Albert, G., Groll, M. T., Stepp, C. E., Carré, J. M., & Arnocky, S. (2020). *Is voice lowering most important in masculine gender presentation? Insights from transgender men's voice change after hormone replacement therapy*. Under review.

Hodges-Simeon, C. R., Gurven, M., Cárdenas, R. A., & Gaulin, S. J. C. (2013). Voice change as a new measure of male pubertal timing: A study among Bolivian adolescents. *Annals of Human Biology, 40*, 209–219. https://doi.org/10.3109/03014460.2012.759622.

Hodges-Simeon, C. R., Gurven, M., & Gaulin, S. J. C. (2015). The low male voice is a costly signal of phenotypic quality among Bolivian adolescents. *Evolution and Human Behavior, 36*(4), 294–302. https://doi.org/10.1016/j.evolhumbehav.2015.01.002.

Hodges-Simeon, C. R., Gurven, M., Puts, D. A., & Gaulin, S. J. (2014). Vocal fundamental and formant frequencies are honest signals of threat potential in peripubertal males. *Behavioral Ecology, 25*(4), 984–988 (doi:1093/beheco/aru081).

Hodges-Simeon, Carolyn R. H, Gaulin, Steven J. C., Puts, David A., et al. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature, 21*(4), 406–427. https://doi.org/10.1007/s12110-010-9101-5.

Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *The Journal of the Acoustical Society of America, 96*(5), 2646–2654. https://doi.org/10.1121/1.411275.

Hughes, S. M., Dispenza, F., & Gallup, G. G., Jr. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior, 25*(5), 295–304. https://doi.org/10.1016/j.evolhumbehav.2004.06.001.

Hyde, Z., Flicker, L., Almeida, O. P., Hankey, G. J., McCaul, K. A., Chubb, S. P., & Yeap, B. B. (2010). Low free testosterone predicts frailty in older men: The health in men study. *The Journal of Clinical Endocrinology & Metabolism, 95*(7), 3165–3172. https://doi.org/10.1210/jc.2009-2754.

Jones, A. L. (2018). The influence of shape and colour cue classes on facial health perception. *Evolution and Human Behavior, 39*(1), 19–29. https://doi.org/10.1016/j.evolhumbehav.2017.09.005.

Jones, B. C., DeBruine, L. M., Perrett, D. I., Little, A. C., Feinberg, D. R., & Smith, M. J. L. (2008). Effects of menstrual cycle phase on face preferences. *Archives of Sexual Behavior, 37*(1), 78–84. https://doi.org/10.1007/s10508-007-9268-y.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour, 79*(1), 57–62. https://doi.org/10.1016/j.anbehav.2009.10.003.

Jones, B. C., Little, A. C., Penton-Voak, I. S., Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Facial symmetry and judgements of apparent health: Support for a "good genes" explanation of the attractiveness–symmetry relationship. *Evolution and Human Behavior, 22*(6), 417–429. https://doi.org/10.1016/S1090-5138(01)00083-6.

Juul, A., Magnusdottir, S., Scheike, T., Prytz, S., & Skakkebæk, N. E. (2007). Age at voice break in Danish boys: Effects of pre-pubertal body mass index and secular trend. *International Journal of Andrology, 30*(6), 537–542. https://doi.org/10.1111/j.1365-2605.2007.00751.x.

Kalashnikova, M., Carignan, C., & Burnham, D. (2017). The origins of babytalk: Smiling, teaching or social convergence? *Royal Society Open Science, 4*(8), 170306. https://doi.org/10.1098/rsos.170306.

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences, 279*(1738), 2698–2704. https://doi.org/10.1098/rspb.2012.0311.

Klofstad, Casey A., Anderson, Rindy C., & Nowicki, Stephen (2015). Perceptions of Competence, Strength, and Age Influence Voters to Select Leaders with Lower-Pitched Voices. *PloS one, 10*(8), e0133779. https://doi.org/10.1371/journal.pone.0133779.

Kokko, H. (1997). Evolutionarily stable strategies of age-dependent sexual advertisement. *Behavioral Ecology and Sociobiology, 41*(2), 99–107. https://doi.org/10.1007/s002650050369.

Kokko, H., Brooks, R., Jennions, M. D., & Morley, J. (2003). The evolution of mate choice and mating biases. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*(1515), 653–664. https://doi.org/10.1098/rspb.2002.2235.

Kordsmeyer, T. L., Thies, Y. T., Ekrami, O., Stern, J., Schild, C., Spoiala, C., … Penke, L. (2020). *No evidence for an association between facial fluctuating asymmetry and vocal attractiveness in men and women*.

Kreiman, J., & Gerratt, B. R. (2003). Jitter, shimmer, and noise in pathological voice quality perception. *ISCA Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis*. https://doi.org/10.1121/1.4780067.

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons.

Laidre, M. E., & Johnstone, R. A. (2013). Animal signals. *Current Biology, 23*(18), R829–R833. https://doi.org/10.1016/j.cub.2013.07.070.

Lea, A. J., Akinyi, M. Y., Nyakundi, R., Mareri, P., Nyundo, F., Kariuki, T., … Tung, J. (2018). Dominance rank-associated gene expression is widespread, sex-specific, and a precursor to high social status in wild male baboons. *Proceedings of the National Academy of Sciences, 115*(52), E12163–E12171. https://doi.org/10.1073/pnas.1811967115.

Leongómez, J. D., Mileva, V. R., Little, A. C., & Roberts, S. C. (2017). Perceived differences in social status between speaker and listener affect the speaker's vocal characteristics. *PLoS One, 12*(6), Article e0179407. https://doi.org/10.1371/journal.pone.0179407.

Li, T., & Jo, C. (2004). Discrimination of severely noisy pathological voice with spectral slope and HNR. In *, 3. Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP'04. 2004* (pp. 2218–2221). IEEE.

Lindström, K. M., Krakower, D., Lundström, J. O., & Silverin, B. (2001). The effects of testosterone on a viral infection in greenfinches (Carduelis chloris): An experimental test of the immunocompetence-handicap hypothesis. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 268*(1463), 207–211. https://doi.org/10.1098/rspb.2000.1352.

Linville, S. E. (1996). The sound of senescence. *Journal of Voice, 10*(2), 190–200. https://doi.org/10.1016/S0892-1997(96)80046-4.

Little, A. C., Apicella, C. L., & Marlowe, F. W. (2007). Preferences for symmetry in human faces in two cultures: data from the UK and the Hadza, an isolated group of hunter-gatherers. *Proceedings of the Royal Society B: Biological Sciences, 274*(1629), 3113–3117. https://doi.org/10.1098/rspb.2007.0895.

Little, A. C., McPherson, J., Dennington, L., & Jones, B. C. (2011). Accuracy in assessment of self-reported stress and a measure of health from static facial information. *Personality and Individual Differences, 51*(6), 693–698. https://doi.org/10.1016/j.paid.2011.06.010.

Lummaa, V., Vuorisalo, T., Barr, R. G., & Lehtonen, L. (1998). Why cry? Adaptive significance of intensive crying in human infants. *Evolution and Human Behavior, 19*(3), 193–202. https://doi.org/10.1016/S1090-5138(98)00014-2.

Markova, D., Richer, L., Pangelinan, M., Schwartz, D. H., Leonard, G., Perron, M., … Paus, T. (2016). Age-and sex-related variations in vocal-tract morphology and voice acoustics during adolescence. *Hormones and Behavior, 81*, 84–96. https://doi.org/10.1016/j.yhbeh.2016.03.001.

Maryn, Y., De Bodt, M., Barsties, B., & Roy, N. (2014). The value of the acoustic voice quality index as a measure of dysphonia severity in subjects speaking different languages. *European Archives of Oto-Rhino-Laryngology, 271*(6), 1609–1619.

Mayew, W. J., Parsons, C. A., & Venkatachalam, M. (2013). Voice pitch and the labor market success of male chief executive officers. *Evolution and Human Behavior, 34*(4), 243–248. https://doi.org/10.1016/j.evolhumbehav.2013.03.001.

Maynard-Smith, J.,& Price, G. R.. (1973). The logic of animal conflict. *Nature, 246*(5427), 15. https://doi.org/10.1038/246015a0.

Millqvist, E., Bende, M., Brynnel, M., Johansson, I., Kappel, S., & Ohlsson, A. C. (2008). Voice change in seasonal allergic rhinitis. *Journal of Voice, 22*(4), 512–515. https://doi.org/10.1016/j.jvoice.2006.12.003.

Muehlenbein, M. P., Alger, J., Cogswell, F., James, M., & Krogstad, D. (2005). The reproductive endocrine response to plasmodium vivax infection in Hondurans. *The American Journal of Tropical Medicine and Hygiene, 73*(1), 178–187. https://doi.org/10.4269/ajtmh.2005.73.178.

Muehlenbein, M. P., & Bribiescas, R. G. (2005). Testosterone-mediated immune functions and male life histories. *American Journal of Human Biology, 17*, 527–558. https://doi.org/10.1002/ajhb.20419.

Nowak, J., Pawłowski, B., Borkowska, B., Augustyniak, D., & Drulis-Kawa, Z. (2018). No evidence for the immunocompetence handicap hypothesis in male humans. *Scientific Reports, 8*(1), 7392. https://doi.org/10.1038/s41598-018-25694-0.

O'Connor, J. J., Pisanski, K., Tigue, C. C., Fraccaro, P. J., & Feinberg, D. R. (2014). Perceptions of infidelity risk predict women's preferences for low male voice pitch in short-term over long-term relationship contexts. *Personality and Individual Differences, 56*, 73–77. https://doi.org/10.1016/j.paid.2013.08.029.

O'Connor, J. J., Re, D. E., & Feinberg, D. R. (2011). Voice pitch influences perceptions of sexual infidelity. *Evolutionary Psychology, 9*(1) (doi:147470491100900109).

Orlikoff, R. F. (1990). The relationship of age and cardiovascular health to certain acoustic characteristics of male voices. *Journal of Speech, Language, and Hearing Research, 33*(3), 450–457. https://doi.org/10.1044/jshr.3303.450.

Owen-Ashley, N. T., Hasselquist, D., & Wingfield, J. C. (2004). Androgens and the immunocompetence handicap hypothesis: Unraveling direct and indirect pathways of immunosuppression in song sparrows. *The American Naturalist, 164*(4), 490–505. https://doi.org/10.1086/423714.

Papp, V. (2012). *The female-to-male transsexual voice: Physiology vs. performance in production (doctoral dissertation)*.

Parsa, V., & Jamieson, D. G. (2001). Acoustic discrimination of pathological voice. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2001/027.

Penton-Voak, I. S., Jacobson, A., & Trivers, R. (2004). Populational differences in attractiveness judgements of male and female faces: Comparing British and Jamaican samples. *Evolution and Human Behavior, 25*(6), 355–370. https://doi.org/10.1016/j.evolhumbehav.2004.06.002.

Phalane, K. G., Tribe, C., Steel, H. C., Cholo, M. C., & Coetzee, V. (2017). Facial appearance reveals immunity in African men. *Scientific Reports, 7*(1), 1–9. https://doi.org/10.1038/s41598-017-08015-9.

Pisanski, K., & Bryant, G. A. (2016). *The evolution of voice perception*. Oxford, UK: Oxford University Press.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J., Röder, S., Andrews, P. W., … Feinberg, D. R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour, 95*, 89–99. https://doi.org/10.1016/j.anbehav.2014.06.011.

Pisanski, K., Mishra, S., & 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), 509–519. https://doi.org/10.1016/j.evolhumbehav.2012.01.004.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America, 129*(4), 2201–2212. https://doi.org/10.1121/1.3552866.

Pound, N., Lawson, D. W., Toma, A. M., Richmond, S., Zhurov, A. I., & Penton-Voak, I. S. (2014). Facial fluctuating asymmetry is not associated with childhood ill-health in a large British cohort study. *Proceedings of the Royal Society B: Biological Sciences, 281*(1792). https://doi.org/10.1098/rspb.2014.1639.

Puts, D. A. (2005). Mating context and menstrual phase affect women's preference for male voice pitch. *Evolution and Human Behavior, 26*(5), 388–397. https://doi.org/10.1016/j.evolhumbehav.2005.03.001.

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2011). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 279*(1728), 601–609. https://doi.org/10.1098/rspb.2011.0829.

Puts, D. A., & Aung, T. (2019). Does Men's voice pitch signal formidability? A reply to Feinberg et al. *Trends in Ecology & Evolution, 34*(3), 189–190. https://doi.org/10.1016/j.tree.2018.12.004.

Puts, D. A., Gaulin, S. J., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior, 27*(4), 283–296. https://doi.org/10.1016/j.evolhumbehav.2005.11.003.

Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., … Jablonski, N. G. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings of the Royal Society B: Biological Sciences, 283*(1829). https://doi.org/10.1098/rspb.2015.2830.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28*(5), 340–344. https://doi.org/10.1016/j.evolhumbehav.2007.05.002.

Raine, J., Pisanski, K., Bond, R., Simner, J., & Reby, D. (2019). Human roars communicate upper-body strength more effectively than do screams or aggressive and distressed speech. *PLoS One, 14*(3). https://doi.org/10.1371/journal.pone.0213034.

Rantala, M. J., Moore, F. R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., & Krams, I. (2012). Evidence for the stress-linked immunocompetence handicap hypothesis in humans. *Nature Communications, 3*, 694. https://doi.org/10.1038/ncomms1696.

Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication, 52*(7–8), 638–651. https://doi.org/10.1016/j.specom.2010.02.012.

Roberts, S. C., & Little, A. C. (2008). Good genes, complementary genes and human mate preferences. *Genetica, 134*(1), 31–43. https://doi.org/10.1007/s10709-008-9254-x.

Roberts, S. C., Little, A. C., DeBruine, L. M., & Petrie, M. (2017). Discrimination of attractiveness and health in Men's faces: The impact of color cues and variation in relation to sex and age of rater. *Adaptive Human Behavior and Physiology, 3*(4), 401–411. https://doi.org/10.1007/s40750-017-0081-0.

Roosenboom, J., Indencleef, K., Lee, M. K., Hoskens, H., White, J. D., Liu, D., … Feingold, E. (2018). SNPs associated with testosterone levels influence human facial morphology. *Frontiers in Genetics, 9*, 497. https://doi.org/10.3389/fgene.2018.00497.

Rosenfield, K. A., Sorokowska, A., Sorokowski, P., & Puts, D. A. (2019). Sexual selection for low male voice pitch among Amazonian forager-horticulturists. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2019.07.002.

Saxton, T. K., Caryl, P. G., & Craig Roberts, S. (2006). Vocal and facial attractiveness judgments of children, adolescents and adults: *The ontogeny of mate choice. Ethology, 112*(12), 1179–1185. https://doi.org/10.1111/j.1439-0310.2006.01278.x.

Scott, I. M., Clark, A. P., Josephson, S. C., Boyette, A. H., Cuthill, I. C., Fried, R. L., … Honey, P. L. (2014). Human preferences for sexually dimorphic faces may be evolutionarily novel. *Proceedings of the National Academy of Sciences, 111*(40), 14388–14393. https://doi.org/10.1073/pnas.1409643111.

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication: Reliability and deception in signaling systems*. Princeton University Press.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Von Rueden, C., … Gurven, M. (2010). Adaptations in humans for assessing physical strength from the

voice. *Proceedings of the Royal Society of London B: Biological Sciences*. https://doi.org/10.1098/rspb.2010.0769.
Shirazi, T. N., Puts, D. A., & Escasa-Dorne, M. J. (2018). Filipino women's preferences for male voice pitch: Intra-individual, life history, and hormonal predictors. *Adaptive Human Behavior and Physiology, 4*(2), 188–206. https://doi.org/10.1007/s40750-018-0087-2.
Simmons, Z. L., & Roney, J. R. (2009). Androgens and energy allocation: Quasi-experimental evidence for effects of influenza vaccination on men's testosterone. *American Journal of Human Biology, 21*(1), 133–135. https://doi.org/10.1002/ajhb.20837.
Skrinda, I., Krama, T., Kecko, S., et al. (2014). Body height, immunity, facial and vocal attractiveness in young men. *Naturwissenschaften, 101*(12), 1017–1025. https://doi.org/10.1007/s00114-014-1241-8.
Smith, F. G., Jones, B. C., DeBruine, L. M., & Little, A. C. (2009). Interactions between masculinity–femininity and apparent health in face preferences. *Behavioral Ecology, 20*(2), 441–445. https://doi.org/10.1093/beheco/arn141.
Soltis, J. (2004). The signal functions of early infant crying. *Behavioral and Brain Sciences, 27*(4), 443–458. https://doi.org/10.1017/S0140525X0400010X.
Stathopoulos, E. T., Huber, J. E., & Sussman, J. E. (2011). Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4-93 years of age. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2010/10-0036.
Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: The MIT Press.
Thordarson, A., Johannsdottir, B., & Magnusson, T. E. (2006). Craniofacial changes in Icelandic children between 6 and 16 years of age–a longitudinal study. *The European Journal of Orthodontics, 28*(2), 152–165. https://doi.org/10.1093/ejo/cji084.
Thore, S. E. (1990). Mothers with chronic illness: A predicament of social construction. *Health Care for Women International, 11*(2), 209–221. https://doi.org/10.1080/07399339009515889.
Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior, 27*(2), 131–144. https://doi.org/10.1016/j.evolhumbehav.2005.06.001.
Titze, I. R. (2000). *Principles of voice production (second printing)*. Iowa City, IA: National Center for Voice and Speech.
Tybur, J. M., & Gangestad, S. W. (2011). Mate preferences and infectious disease: Theoretical considerations and evidence in humans. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 366*(1583), 3375–3388. https://doi.org/10.1098/rstb.2011.0136.
Verdonck, A., Gaethofs, M., Carels, C., & de Zegher, F. (1999). Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *The European Journal of Orthodontics, 21*(2), 137–143. https://doi.org/10.1093/ejo/21.2.137.
Ware, J. E. (1976). Scales for measuring general health perceptions. *Health Services Research, 11*(4), 396–415.
Wedekind, C., & Folstad, I. (1994). Adaptive or nonadaptive immunosuppression by sex hormones? *The American Naturalist, 143*(5), 936–938. https://doi.org/10.1086/285641.
Wermke, K., & Robb, M. P. (2010). Fundamental frequency of neonatal crying: Does body size matter? *Journal of Voice, 24*(4), 388–394. https://doi.org/10.1016/j.jvoice.2008.11.002.
Whitehouse, A. J., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., … Eastwood, P. (2015). Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. *Proceedings of the Royal Society B: Biological Sciences, 282*(1816), 20151351.
Wuyts, F. L., Bodt, M. S. D., Molenberghs, G., Remacle, M., Heylen, L., Millet, B., … Heyning, P. H. V. D. (2000). The dysphonia severity index: An objective measure of vocal quality based on a multiparameter approach. *Journal of Speech, Language, and Hearing Research, 43*(3), 796–809.
Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal of Theoretical Biology, 53*(1), 205–214. https://doi.org/10.1016/0022-5193(75)90111-3.
Zahavi, A. (1977). The cost of honesty: Further remarks on the handicap principle. *Journal of Theoretical Biology, 67*(3), 603–605. https://doi.org/10.1016/0022-5193(77)90061-3.
Zaidi, A. A., White, J. D., Mattern, B. C., Liebowitz, C. R., Puts, D. A., Claes, P., & Shriver, M. D. (2019). Facial masculinity does not appear to be a condition-dependent male ornament and does not reflect MHC heterozygosity in humans. *Proceedings of the National Academy of Sciences, 116*(5), 1633–1638. https://doi.org/10.1073/pnas.1808659116.