Human vocal communication of body size

HUMAN VOCAL COMMUNICATION OF BODY SIZE

By KATARZYNA PISANSKI, M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2014) McMaster University

(Psychology, Neuroscience & Behaviour) Hamilton, Ontario

TITLE: Human vocal communication of body size

AUTHOR: Katarzyna Pisanski, M.Sc. (University of Lethbridge)

SUPERVISOR: Dr. David R. Feinberg

NUMBER OF PAGES: xix, 229

ABSTRACT

The human voice may convey meaningful information about socially and evolutionarily relevant characteristics of the vocalizer. In turn, listeners may readily evaluate personal characteristics, such as body size, on the basis of nonverbal voice features. Research investigating vocal communication of physical size in humans has focused on two salient and largely independent voice features, fundamental frequency and/or corresponding harmonics (perceived as voice pitch) and formant frequencies (resonance frequencies of the supralaryngeal vocal tract). However, the degree to which fundamental and formant frequencies reliably predict variation in body size controlling for sex and age, and their relative role in the perception or accurate estimation of body size, has to date been unclear. In the current thesis, using meta-analysis, I establish that formants reliably predict variation in men's and women's heights and weights. In contrast, fundamental frequency only weakly predicts men's heights and women's weights. These findings corroborate work on many other mammals whose vocal production, like humans, follows the source-filter model. Despite the lack of a robust physical relationship between fundamental frequency and size within sexes, I further demonstrate that listeners utilize voice pitch to accurately gauge men's relative height. My research suggests that voice pitch indirectly facilitates accurate size assessment by providing a carrier signal (i.e., dense harmonics) for formants. This is the first evidence that pitch does not confound accurate size estimation. Finally, I demonstrate that voices with lowered pitch, but not raised pitch, are perceived as larger when projected from a low than high spatial location.

These results suggest that strong cross-modal perceptual biases linking low pitch to low elevation and large size may, in some contexts, cause errors in size estimation. Taken together, this thesis provides a detailed account of human vocal communication of body size, which can play a meaningful role in sexual and social contexts.

ACKNOWLEDGEMENTS

I offer my profound gratitude to my Ph.D. advisor, David Feinberg, whose knowledge, advice, and unwavering support have been instrumental in my progression from student to scientist. Thank you for believing in me, Dave. I also gratefully acknowledge the contributions and support of past and present members of the Voice Research Lab, especially Cara Tigue, Jill O'Connor, and Paul Fraccaro. This research was made possible through the generous financial support of the Social Sciences and Humanities Research Council of Canada and McMaster University.

A special thank you to Sigal Balshine and Paul Andrews who have become my mentors over the years, and to my committee members Bruce Milliken and Mel Rutherford for their constructive feedback and guidance. More generally, I thank the faculty and students of the truly unique PNB department whose passion for life and science inspired and motivated me everyday. Many ideas and friendships were born out of Friday socials and Psy Jung softball.

Thank you to the faculty outside of McMaster University whom I have had the pleasure to collaborate with and learn from, including Ben Jones, Lisa DeBruine, Greg Bryant, David Reby, Bernhard Fink, Piotr Sorokowski, Emanuel Mora-Macías, and Nancy Estévez Pérez. Now and always I will also acknowledge Drew Rendall, John Vokey, and Louise Barrett whose early influence has guided me over the years.

To my friends, especially Lauren Dobbelaar, Susie Barclay, Marcus Morrisey, Michel Belyk, Lisa Wilson, Erin McCallum, Maria D'Angelo, Adam Reddon, Natalia

Rojas, Matt Pachai, Anna Finkelshtein, Jenn Walsh, Nicole LeBarr, Cody Dey, Michelle Gordon, Ashley Milne, Emilia Farrace, Sjoerd Biesmans, and to all of the strong women of the McMaster WISE initiative, thank you for offering me endless intellectual stimulation but also welcome respites from academia! A special thank you to Mitch LaPointe for countless illuminating conversations and for showing me the true meaning of unconditional love and friendship. Finally, thank you to my dear sisters and brothers in Havana, especially you, Annette Hernández-Abad, who have shared with me a wisdom far beyond your years and an intimacy I did not know was possible.

This thesis is dedicated to my family. To my mother and father, Ewa and Maciej Pisanski, everything in my life and being is a reflection of your love and unwavering support. Each day you remind me that family matters most in life, which is perhaps the greatest life lesson of them all. To my brother and the best person I know, Peter Pisanski, thank you for being my role model while simultaneously, my best friend and biggest fan. Kocham Was.

Pura Vida!

TABLE OF CONTENTS

LIST OF FIGURES

APPENDIX 2

List of Tables

APPENDIX 3

LIST OF ABBREVIATIONS AND SYMBOLS[1]

/ɑ/	International phonetic symbol for vowel sound "ah" (as in father)
/i/	International phonetic symbol for vowel sound "ee" (as in see)
/ɛ/	International phonetic symbol for vowel sound "eh" (as in bet)
/o/	International phonetic symbol for vowel sound "oh" (as in note)
/u/	International phonetic symbol for vowel sound "oo" (as in boot)
ΔF	Formant spacing
c	Speed of sound
CFA	Confirmatory factor analysis
CI	Confidence interval
dB	Decibels
D_f	Formant dispersion
ERB	Equivalent rectangular bandwidth
ES_z	Unweighted standardized effect size
$\overline{ES_z}$	Average weighted standardized effect size
F_0	Fundamental frequency
$F1$	First formant
$F2$	Second formant
$F3$	Third formant

[1] This list is not exhaustive. It contains only those symbols and acronyms used multiple times throughout the thesis and/or those that may be unfamiliar to some readers.

$F4$	Fourth formant
Fn	Mean formant frequency
Hz	Hertz
k	Number of independent samples
LPC	Linear predictive coding
MFF	Geometric mean formant frequency
n	Sample size
N	Pooled sample size
ns	Nonsignificant
PCM	Pulse-code modulation
P_{f}	Formant position
PSOLA	Pitch-synchronous overlap add algorithm
r	Pearson correlation coefficient
$\bar{r}$	Mean weighted correlation coefficient
SPL	Sound pressure level
SWS	Three-formant sine-wave speech
T	Period of a sound wave
VTL	Vocal tract length
(VTL(ΔF))	Apparent vocal tract length (derived from formant spacing)
(VTL(F_i))	Apparent vocal tract length (derived from mean formants)
w	Weight (statistical)

DECLARATION OF ACADEMIC ACHIEVEMENT

I am the primary author and main contributor of the three research papers included in this sandwich thesis. Chapters 2 and 3 reflect two published research articles. Chapter 4 is a research article submitted for publication. For the purpose of this thesis, I have renumbered the pages to follow sequentially and have standardized the use of acronyms and symbols. Supplementary text, figures and tables are also numbered sequentially and are given in the Appendix at the end of the thesis. However, following McMaster University's guidelines for the preparation of sandwich Ph.D. theses, the main figure and table numbers and the reference style of the journal of publication have been retained within each given chapter.

The research presented here is the core product of my doctoral work. In consultation and collaboration with my supervisor David Feinberg, I designed and developed each experiment, created the voice stimuli, conducted the statistical analysis, and wrote the research articles. Below I list the contributions of coauthors for each article as well as the year(s) during which the data were collected.

Pisanski, K., Fraccaro, P.J., Tigue, C.C., O'Connor, J.J.M., Röder, S, Andrews, W.P, Bernhard, F., DeBruine, L.M., Jones, B.C., & Feinberg, D.R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour, 95,* 89-99. Doi: 10.1016/j.anbehav.2014.06.011

Fraccaro, P.J., Tigue, C.C., & O'Connor, J.J.M. — Stimulus collection; voice measurements; manuscript development and revisions

Röder, S., & Bernhard, F., DeBruine, L.M. & Jones, B.C. — Stimulus collection; manuscript revisions

Andrews, W.P. — Statistical analysis; manuscript revisions

Feinberg, D.R. — Design; stimulus collection; manuscript development and revisions

Meta-data collected and analyzed 2012-2014

Pisanski, K., Fraccaro, P.J., Tigue, C.C., O'Connor, J.J.M., & Feinberg, D.R. (2014). Return to Oz: Voice pitch facilitates assessment of men's body size. *Journal of Experimental Psychology: Human Perception & Performance, 40*(4), 1316-1331. Doi: http://dx.doi:10.1037/a0036956

Fraccaro, P.J., Tigue, C.C., & O'Connor, J.J.M. — Stimulus and data collection; manuscript development and revisions

Feinberg, D.R. — Concept and design; statistical analysis; manuscript development and revisions

Experimental data collected 2011-2012

Pisanski, K., O'Connor, J.J.M., Tigue, C.C., & Feinberg, D.R. Low is large: Pitch interacts with spatial cues in size perception (under review). *Submitted July 8, 2014 to Cognition*. 1-21.

O'Connor, J.J.M., Tigue, C.C. — Stimulus collection; manuscript revisions

Feinberg, D.R. — Concept; statistical analysis; manuscript revisions

Experimental data collected 2013

CHAPTER 1: GENERAL INTRODUCTION

In this thesis, I examine relationships between the human voice and body. In Chapter 1, I discuss several functions of acoustic communication and the ubiquitous importance of body size among animals, including humans, from an evolutionary and cognitive perspective. I then review the literature on vocal correlates of physical size and on listeners' voice-based estimates of size. I highlight the holes in this literature and outline the research that will empirically address these open questions in Chapters 2, 3, and 4. In Chapter 5, I discuss the implications, limitations, and future directions of this work.

Acoustic Communication

Animals in a wide variety of taxa communicate using sound, including insects, birds, anurans, fish, and all extant species of mammal (Owren, 2012; Simmons, Popper, Fay, & Gerhardt, 2003). Animal vocalizations play a pivotal role in social communication. Seminal work on mammals, such as red deer and non-human primates, revealed that vocal characteristics frequently determine the outcome of competitive resource contests and mate choice (Clutton-Brock & Albon, 1979; Marler & Tenaza, 1977). In non-sexual scenarios, such as parent-infant or group interactions, vocalizations may function to indicate affective state, identify kin, and maintain social relationships (Marler & Hobbett, 1975; McComb, Reby, Baker, Moss, & Sayialel, 2003; Rendall, Rodman, & Emond, 1996). Vocalizations may also be used to transmit information among individuals, for example through alarm calling in many mammals (e.g., Sherman, 1977) or language in humans.

Nonverbal structural features of the voice may provide information about the vocalizer. In humans, various features of the voice indicate sex (Huber, Stathopoulos, Curione, Ash, & Johnson, 1999; Titze, 1989; Wu & Childers, 1991) and approximate age (Bruckert, Liénard, Lacroix, Kreutzer, & Leboucher, 2006). The human voice can also provide information about an individual's circulating levels of sex and stress hormones (Abitbol, Abitbol, & Abitbol, 1999; Bruckert et al., 2006; Dabbs & Mallinger, 1999; Kirchhübel, Howard, & Stedmon, 2011), social dominance or physical strength (Puts, Apicella, & Cardenas, 2012a; Puts, Hodges, Cardenas, & Gaulin, 2007; Sell et al., 2010), and body size, which is the focus of this thesis. Nonverbal features of the human voice may have been shaped, at least in part, by sexual selection (Darwin, 1871; see also Owren, 2012; Puts, Jones, & DeBruine, 2012b).

Human listeners may in turn use vocal information for a variety of purposes. For example, listeners may use vocal information to discriminate familiar from unfamiliar individuals (Campanella & Belin, 2007; Van Lancker, Kreiman, & Emmorey, 1985) or to assess the quality of potential sexual mates and formidability of potential rivals (Feinberg, 2008; Maynard Smith & Harper, 2003). However, although the voice can provide reliable and socially relevant information about the vocalizer, this does not necessarily imply that listeners can and will *accurately* interpret this information.

In this thesis, I examine vocal indicators of size in humans. I test whether various vocal features reliably correlate with body size in men and women (Chapters 2 and 3) and whether listeners can effectively assess size from the voice (Chapters 3 and 4). The aim of this work is to determine the role that acoustic communication plays in advertising and assessing body size among humans.

Body Size

Body size can have an immense impact on the biology, ecology, and life history of an animal (Peters, 1986). At the behavioural level, body size can explain a good deal of the variation in social and reproductive success among individuals (Andersson, 1994). In most mammals, males are larger than females (Weckerly, 1998) and sexual selection is thought to play a significant role in size dimorphism (Darwin, 1871; but see Fairbairn, Blanckenhorn, & Székely, 2007 for a review of other contributing factors). Relatively larger males are more likely to win resource contests with other males, thus gaining access to higher-quality territories, and are often preferred as mates by females (Bisazza & Marconato, 1988; Cooper & Vitt, 1993; Lindenfors, Gittleman, & Jones, 2007).

Men are taller than women across cultures (Holden & Mace, 1999), although the sexual dimorphism in human body size is smaller than in other great apes (Leigh & Shea, 1995). In North America and Europe, taller men have a higher level of education and earn more money (Judge & Cable, 2004; Peck & Vågerö, 1987; Szklarska, Koziel, Bielicki, & Malina, 2007; Teasdale, Owen, & Sørensen, 1991), are preferred as mates and have more attractive girlfriends (Feingold, 1982; Yancey & Emerson, 2014), and bear more children than do shorter men (Nettle, 2001; Pawlowski, Dunbar, & Lipowicz, 2000). It should be noted, however, that there is variation in the strength and direction of these relationships across cultures (see, e.g., Lee, 1979).

Women generally prefer above-average height in men and prefer that their male mate be taller relative to their own height (Fink, Neave, Brewer, & Pawlowski, 2007; Kurzban & Weeden, 2005; Pawlowski & Koziel, 2002; Swami et al., 2008). However, there is cross-cultural variation in height preferences. Women of the Hadza, Himba, and Datoga tribes living in various regions of Africa show less extreme preferences for male tallness compared to women in North American and Europe (Marlowe, 2004; Sorokowski & Butovskaya, 2012; Sorokowski, Sorokowska, Fink, & Mberira, 2012). The relationships between women's body size and women's social or reproductive success are less clear than they are for men (Monden & Smits, 2008; Sear, 2011). In general, however, it appears that men prefer women of average or below-average height (Pawlowski & Koziel, 2002; Salska et al., 2008; Yancey & Emerson, 2014).

In addition to influencing socioeconomic status, mate preferences, and reproductive success, body size may have a direct impact on health and fertility. For example, shorter individuals may be more likely to die of coronary heart disease, stroke, and respiratory disease than taller individuals (Smith et al., 2000). Among women, very low or very high amounts of body fat may lead to difficulty in conceiving (Zaadstra et al., 1993). Trade-offs between investment in somatic growth versus reproduction may also influence the relationship between size and health or fertility, particularly among women and in regions where resources are scarce (Stearns, 1976).

Advertising and Assessing Body Size

Given the overarching role of body size in predicting dominance and social success and its influence on the outcome of resource contests and mate choice, particularly among men (see Puts, 2010 for review), it may be advantageous in some contexts to advertise or even exaggerate one's own size (Dawkins & Krebs, 1978; Fitch & Hauser, 2003). It is also likely to be advantageous to accurately judge the size of others. Although the advertisement of body size often takes the form of visual displays, such as lateral body positioning in cichlid fish (Carleton, Spady, & Kocher, 2005; Reddon et al., 2011) or feather decorations in many bird species (Hasson, 1991), vocalizations are likewise used by a wide range of species to indicate (or exaggerate) one's size and to judge the size of others (e.g., koalas: Charlton et al., 2011; red deer: Fitch & Reby, 2001; Reby & McComb, 2003; birds: Fitch, 1999; colobus monkeys: Harris, Fitch, Goldstein, & Fashing, 2006; frogs and toads: Ryan, 1988; elephant seals: Sanvito, Galimberti, & Miller, 2007; macaques: Sommers, Moody, Prosen, & Stebbins, 1992; dogs: Taylor, Reby, & McComb, 2010). Vocalizations may be used in conjunction with visual cues to enhance size discrimination (Bro-Jørgensen, 2010) or in place of visual cues when visual information fails (e.g., at longer distances).

In humans there are two key acoustic parameters that may indicate body size, fundamental frequency and formant frequencies, described in detail below.

Mechanisms of Vocal Production

Source-Filter Framework

Like most terrestrial mammals[2], human speech is produced by the larynx (the source) and subsequently filtered by the supralaryngeal vocal tract (the filter; Fant, 1960; Taylor & Reby, 2010). Together this source-filter system determines the acoustic properties of many animal vocalizations, including voiced human speech.

When humans phonate, air expelled from the lungs causes oscillation of the vocal folds within the larynx (see Figure 1 for a diagram of the vocal production apparatus in humans). Longer and heavier vocal folds vibrate at a slower rate than do smaller vocal folds. It is this base rate of vocal fold vibration that determines the fundamental frequency (F_0) of the voice, or the glottal-pulse rate, wherein slower vibrations result in a lower fundamental frequency. Harmonic overtones occur at multiple integers of the fundamental frequency such that a lower fundamental frequency results in more densely spaced harmonics. Fundamental frequency and harmonics are perceived as voice pitch (Kreiman & Sidtis, 2011; Titze, 1994).

[2] Vocal anatomy and mechanisms of speech production are broadly similar in humans and other mammals allowing for meaningful comparisons (see Ghazanfar & Rendall, 2008; Owren, 2012; Taylor & Reby, 2010 for reviews).

In this thesis, I will use the term fundamental frequency when referring to the measured property of the voice, and pitch when referring the perception of that property[3].

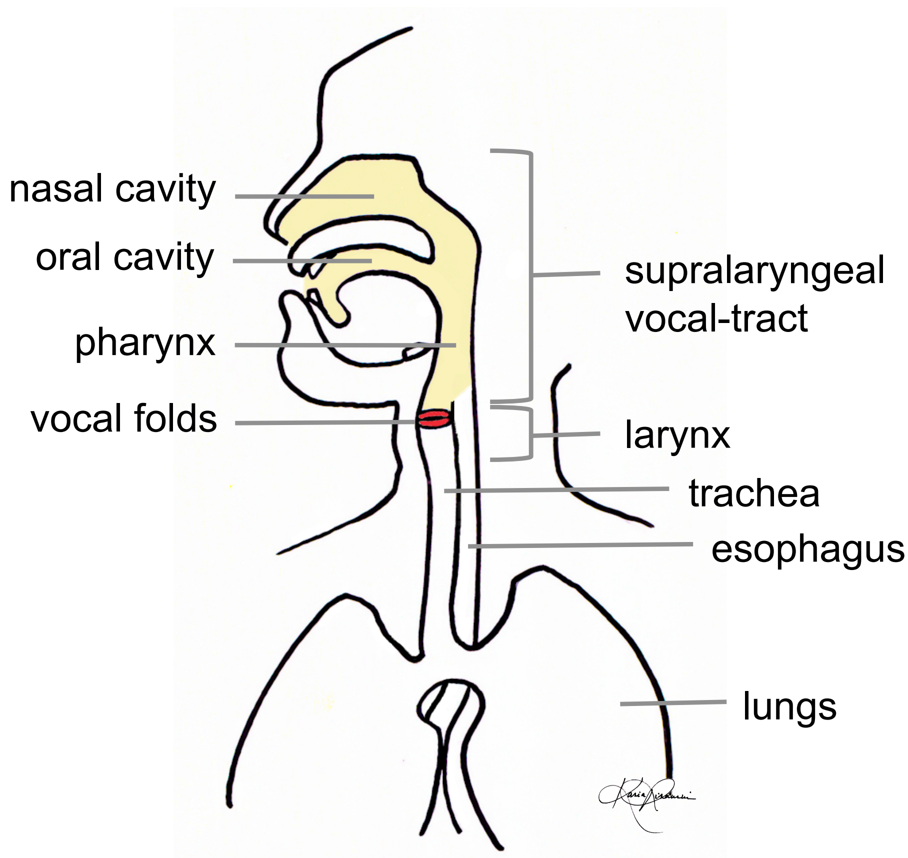

Figure 1. A diagram of the human vocal production apparatus.

[3] Fundamental frequency (F_0) is the lowest frequency of the periodic waveform measured from the vocal signal, representing the base rate of vocal fold vibration (glottal-pulse rate). Pitch typically refers to the perpetual correlate of fundamental frequency and/or corresponding harmonics. The relationship between fundamental frequency and pitch is logarithmic. However, in the range of frequencies that characterize the human voice, fundamental frequency and pitch are roughly equivalent.

From the larynx, the sound travels up through the pharynx and oral and nasal cavities that together comprise the supralaryngeal vocal tract (henceforth vocal tract, see Figure 1), where vocal sounds are filtered. This filtering process, in which certain frequencies are attenuated while others are not, is determined largely by the size and shape of the vocal tract. Thus, the vocal tract acts as a bandpass speech filter producing resonant frequencies referred to as formants (Fant, 1960; Titze, 1994). Formants are generally perceived as the timbre or quality of the voice whereby longer vocal tracts produce lower formants. Manipulations of the tongue, lips, and soft palate can alter which formant frequencies become attenuated and these modifications give rise to different vowel sounds in speech (Lieberman, 1988). This point is illustrated in Figure 2. Spectrograms show a different pattern of formant dispersion for each of the five vowel sounds /ɑ/ /i/ /ɛ/ /o/ and /u/. Note that, for example, /ɑ/ is characterized by closely spaced first and second formants whereas /i/ is characterized by widely spaced first and second formants. Although men's formants are on average lower than are women's (Titze, 1989), the differences in formant spacing among vowel sounds are similar in both sexes.

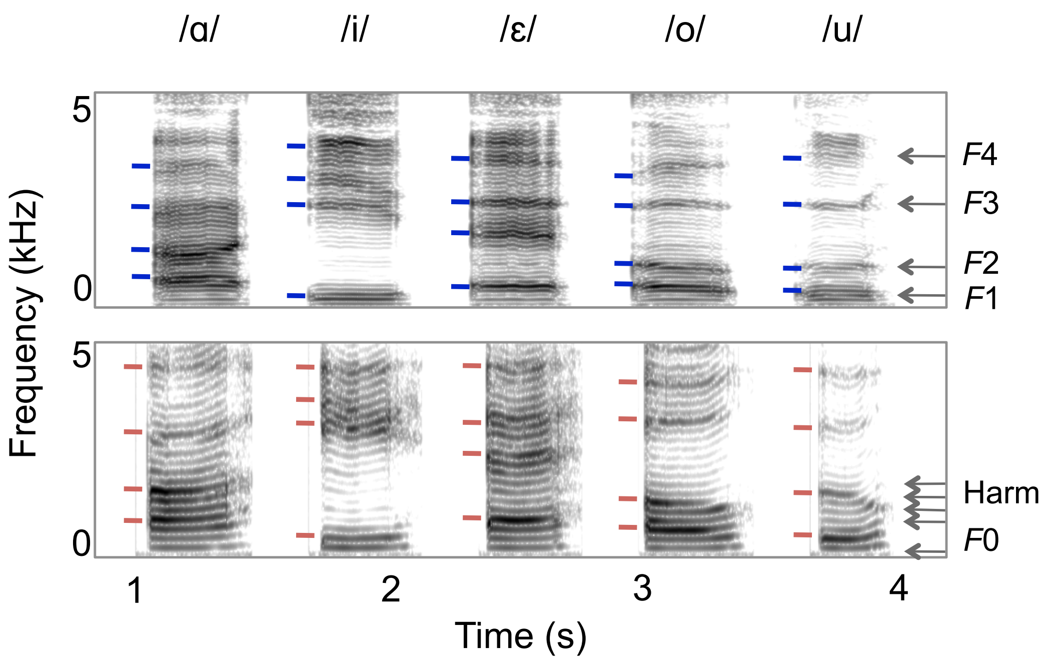

Figure 2. Spectrograms of the vowels /ɑ/ /i/ /ɛ/ /o/ and /u/ (international phonetic symbols) spoken by an adult male (top row) and an adult female (bottom row).

Note. The first to fourth formants ($F1$-$F4$), fundamental frequency (F_0) and harmonics of the fundamental frequency (Harm.) are labeled for the vowel /u/. Formant positions are labeled for each individual vowel with bars positioned to the left of each spectrogram. Note the variation in formant spacing between the male and female voices and among vowels.

Based on the source-filter model of speech production, fundamental and formant frequencies are largely independent (Fant, 1960). Individuals may have any combination of high or low voice pitch paired with high or low formants. However, some studies demonstrate that fundamental and formant frequencies interact to some degree (Kewley-

Port, Li, Zheng, & Neel, 1996; Titze, 2008; Turner, Walters, Monaghan, & Patterson, 2009), affecting vocal production in complex ways that are not yet fully understood.

Hormonal Effects on the Voice

Among humans, vocal fold development and vocal tract length are affected by sex hormones during puberty and in adulthood. The vocal folds contain androgen receptors (Saez & Sakai, 1976). During puberty, an increase in testosterone among males causes the vocal folds to lengthen and widen and causes a corresponding decrease in voice pitch (Harries, Hawkins, Hacking, & Hughes, 1998; Lieberman, McCarthy, Hiiemae, & Palmer, 2001). In females, the vocal folds thicken only slightly at puberty, and the resulting effect on voice pitch is small relative to males (Abitbol et al., 1999). Testosterone is inversely related to pitch between sexes, and likely also within sexes, in adults (Dabbs & Mallinger, 1999; Evans, Neave, Wakelin, & Hamilton, 2008). Hence, injecting a woman with testosterone will cause a marked decrease in her voice pitch that is largely irreversible, and pre-pubertal castration (e.g., among castrati singers in the 15^{th} century) will cause the male voice to maintain feminine pitch qualities into adulthood (Abitbol et al., 1999).

Testosterone also causes the male larynx to descend slightly during puberty. This secondary decent of the larynx elongates the male vocal tract relative to the female vocal tract (Lieberman et al., 2001). To date, however, only one study has found a relationship between higher testosterone levels and lower formants in men (Bruckert et al., 2006). It is not clear how hormones affect the relationships between vocal features and body size.

Vocal Correlates of Body Size

For many years researchers have speculated that fundamental frequency (Darwin, 1871; Morton, 1977; Ohala, 1984) and more recently formant frequencies (Fitch, 1994) may serve as indicators of body size (i.e., may exhibit acoustic allometry). However, researchers have also long debated the degree to which one or the other voice feature can reliably predict size.

According to one prominent theory (Fitch, 1994, 1997, 2000a, 2000b), formant frequencies are hypothesized to predict body size more reliable than fundamental frequency in many animals including humans. The crux of this argument lies in differential anatomical constraints on vocal fold and vocal tract development. In a simplified model, the vocal tract of most vertebrate species can be described as an air-filled tube closed at the laryngeal end and open at the mouth end (Fitch, 1994). Resonant frequencies or formants are heavily constrained by the bony anatomy of the vocal tract and the length of the vocal tract is in turn constrained by the size of the skull (Fitch, 2000b). As a consequence, taller or larger individuals are predicted to have longer vocal tracts than shorter or smaller individuals (Fitch & Giedd, 1999), and thus taller individuals are predicted to produce lower formants.

In contrast to the vocal tract, the larynx and vocal folds that produce fundamental frequency are made up of soft tissue and are relatively more versatile. The larynx grows largely independently of the skull and body (Lieberman et al., 2001). Thus, taller or larger individuals may not always have larger larynges and longer vocal folds than shorter or smaller individuals, and may not produce a relatively lower fundamental frequency. Additionally, some researchers have proposed that it may be easier to manipulate ones fundamental frequency than ones formant frequencies in order to minimize or exaggerate perceived body size in various social contexts (Fitch, 2000a; Morton, 1977), further dissociating pitch from physical size.

Below I describe theoretical and empirical work investigating the relative roles of fundamental and formants frequencies in indicating body size (1) across different species; (2) between men and women or adults and children; and (3) among same-sex adults.

Comparing Across Species

Vocal tract length correlates positively with body size across mammalian species where the differences in size are often substantial (Fitch & Hauser, 2003). For instance, the estimated vocal tract length of the harbour seal *Phoca vitulina* (adults weighing anywhere from 50-170 kg) is several times longer than that of the white-handed gibbon *Hylobates lar* (4-8 kg; Fitch, 2000b). Thus, lower formant frequencies often predict larger body size at the species level. Among primates, for instance, great apes have lower formants than do smaller primate species such as marmosets (Ey, Pfefferle, & Fischer, 2007; Hauser, 1993). It is also the case that larger mammals (August & Anderson, 1987), including larger primates within the family of primates (Ey et al., 2007; Hauser, 1993), have relatively lower fundamental frequencies. However, to my knowledge, no study has tested whether vocal fold length or mass predict body size across species.

Comparing Between Sexes and Age Groups

In many animals, vocal fold thickness and vocal tract length covary with body size between the sexes and between adults and juveniles of the same species, where once again the differences in size are often substantial (see Ey et al., 2007 for review on primates).

Cadaver dissections of humans indicate that men have larynges approximately 20% longer and wider and vocal folds approximately 60% longer than do women, and both men and women have longer vocal folds than do children (Kahane, 1978). Fitch and Giedd (1999), using magnetic resonance imaging to measure the length of the supralaryngeal vocal tract (i.e., from the vocal folds to the lips, see Figure 1), also report significant differences in vocal tract lengths across age-sex groups. The authors report that vocal tract length increased from an average of 10 cm among infants aged two to four to approximately 15 cm among adults aged nineteen to twenty-five. Vocal tract length diverged significantly between the sexes after age fifteen (around the time of puberty and probably due to a secondary descent of the male larynx, Lieberman et al., 2001), resulting in a vocal tract 1-2 cm longer among adult men than women.

Corresponding to these anatomical differences, fundamental and formant frequencies often covary with size across age-sex groups (Ey et al., 2007). Men's fundamental frequency (on average 120 Hz) is approximately half that of women's (on average 210 Hz, see Figure 2) and both are considerably lower than that of pre-pubescent children and adolescents (250-500 Hz). Men's formant frequencies are approximately 15% lower than those of women (Figure 2) and both men and women have considerably lower formants than do children (Bachorowski & Owren, 1999; Hillenbrand, Getty, Clark, & Wheeler, 1995; Huber et al., 1999; Peterson & Barney, 1952; Pisanski & Rendall, 2011; Rendall, Kollias, Ney, & Lloyd, 2005; Titze, 1989).

It should be noted that not all of the variation in fundamental and formant frequencies between men and women or across the lifecycle is tied directly to body size. This is especially true for fundamental frequency. Although men's fundamental frequency is on average 100% lower than that of women's, men's vocal fold length exceeds women's by only 60% (Titze, 1989). Therefore, the morphology of the vocal folds can explain only a portion of the sexual dimorphism in human voice pitch. Men's formants are approximately 15% lower than those of women (Pisanski & Rendall, 2011; Titze, 1989). Although the sex difference in formants scales fairly reliably with the sex difference in vocal tract length (Fitch & Giedd, 1999; Rendall et al., 2005), it still exceeds the mere 10% difference in men's and women's heights. Sex hormone levels can likely explain most of the variation in fundamental and formants frequencies that cannot be explained by size (Abitbol et al., 1999; Bryant & Haselton, 2009; Dabbs & Mallinger, 1999; Harries et al., 1998).

Comparing Within Same-Sex Adults.

The anatomical and structural differences in men's and women's vocal production, along with strong preferences for sexually dimorphic vocal features in the opposite sex (see Feinberg, 2008; Pisanski & Feinberg, 2013 for reviews), suggest that the human voice has undergone sexual selection. This is also the case for physical body size, which is sexually dimorphic and predicts fitness (see section on *Body Size*). If the human voice and body size have been shaped by sexual selection, vocal characteristics may covary with body size in groups of same-sex adults, particularly because this is the level at which potential mates and rivals are typically assessed. We might further predict relationships between the voice and body to be stronger among men than women, because body size may play an important role in male-male competition.

Research on various species suggests, however, that fundamental and formant frequencies are generally only weakly related to vocal anatomy and body size in same-sex adults (Fitch & Hauser, 2003). Additionally, in a wide range of mammalian species whose vocal production mechanisms follow the source-filter model, lower formants predict larger physical size in same-sex adults whereas lower fundamental frequency does not (Charlton et al., 2011; Fitch, 1997; Harris et al., 2006; Reby & McComb, 2003; Sanvito et al., 2007; Vannoni & McElligott, 2008).

Among humans, past work has produced mixed results. Anatomically, there is no strong evidence of a relationship between vocal fold size and body size within sexes (Sawashima et al., 1983). No study has tested whether vocal tract length correlates with size within sexes (but see Fitch & Giedd, 1999). Acoustically, studies have found support both for and against relationships between fundamental or formant frequencies and height or weight in men and women (see González, 2006 for review; see also Tables A1 and A2 in Appendix 2 for a comprehensive summary of past results). Taken together, it remains unclear from past work whether the voice can reliably predict variation in human body size, and whether formants are a more reliable indicator of human size than is fundamental frequency. This research question is addressed in Chapter 2.

Voice-Based Estimates of Body Size

Intimately tied to the question of acoustic allometry is the question of whether and how listeners process vocal information in relation to size. It is often advantageous for individuals to accurately assess the body size of others, and likewise costly to err. This is particularly true in the context of male-male competition and mate-choice in which size discrimination takes place at the within-sex level.

Although vocalizations may aid listeners in accurately judging the size of the vocalizer, vocalizations may likewise confuse size assessment. Consider, for example, the possibility of only a weak relationship between vocal structure and physical size (González, 2006) or the possibility that vocalizers may manipulate various features of the voice to sound smaller or larger than they actually are (Morton, 1977). Perceptual biases may also contribute to erroneous size discrimination (Rendall, Vokey, & Nemeth, 2007).

There are many examples of reliable acoustic communication of body size in the animal kingdom. A classic example is roaring contests in red deer stags, *Cervus elaphus.* Roaring contests occur frequently during the mating season and largely determine the dominance rank and reproductive success of males (Clutton-Brock & Albon, 1979). An extensive research program investigating vocal indicators of size in red deer has revealed that lower formants, but not lower fundamental frequency, reliably predict larger male body size (Reby & McComb, 2003), that males use formants to assess the formidability of rivals in roaring contests (Reby et al., 2005), that females can perceive format shifts in male roars (Charlton, Reby, & McComb, 2007), and that females prefer the roars of males that are manipulated to have lower than higher formants, simulating larger size, regardless of the fundamental frequency of the roar (Charlton, Reby, & McComb, 2008). In fact, many animals including primates appear to be sensitive to variation in formants indicating size in the calls produced by conspecifics (Charlton, Whisson, & Reby, 2013b; Fitch & Fritz, 2006; Sommers et al., 1992).

Past studies on humans have examined the accuracy of listeners' body size estimates from the voice as well as the vocal parameters that listeners use to estimate size. Findings from both lines of research are reviewed below. It should be noted for future reference, however, that surprisingly few studies have combined these two lines of research to test how various vocal parameters affect the accuracy of body size estimation.

Accuracy of Size Estimates from the Voice

Several studies have presented listeners with natural, unmanipulated voices of men and/or women and have tested whether listeners' size estimates reliably correlate with the actual (measured or reported) height or weight of the vocalizer. Early work by Lass and colleagues suggested that listeners were highly accurate in their assessments of both men and women's body size (Lass, Barry, Reed, Walsh, & Amuso, 1979; Lass, Hughes, Bowyer, Waters, & Bourne, 1976; Lass, Kelley, Cunningham, & Sheridan, 1980). However, the statistical methods employed in this work were later revealed to be inappropriate (Cohen, Crystal, House, & Neuburg, 1980). When the data were reanalyzed, only 14% of size estimates correlated with actual height or weight (González, 2003, 2006).

More recent work suggests that listeners are capable of assessing body size from the voice, but that performance is modest and variable. When listeners were asked to report the absolute height or weight of a vocalizer (e.g., in kg and cm), van Dommelen and Moxness (1995) found that estimated height predicted actual height only for men's and not for women's voices, and that men were more accurate than were women at assessing men's size. Collins (2000) and Bruckert et al. (2006) found that women were able to estimate men's weight but not men's height, however, men's actual weight explained only 22% (Collins, 2000) and 16% (Bruckert et al., 2006) of the variance in women's estimates.

Listeners appear to perform better and less variably in forced-choice tasks, where they simply indicate which of two vocalizers is taller. On average, both men and women can correctly identify the taller of two men on 60% of trials (Oliver & González, 2004; Rendall et al., 2007). Accuracy is comparable (Oliver & González, 2004) or slightly lower (Rendall et al., 2007) for assessments of women's relative size. As might be expected, accuracy increases with the relative difference in height between two vocalizers, approaching 90% when the differences in height are substantial (e.g., over 20 cm, Rendall et al., 2007).

Vocal Parameters Used to Estimate Size

Listeners routinely associate relatively low formants and/or relatively low pitch with large body size both between and within sexes. This pattern of results has been demonstrated using a variety of methods and vocal stimuli. For instance, these results have been replicated using rating scales (Feinberg, Jones, Little, Burt, & Perrett, 2005; Pisanski & Rendall, 2011; Sell et al., 2010), a two-alternative forced choice paradigm (Rendall et al., 2007), or absolute size estimates (Bruckert et al., 2006; Collins, 2000); using natural voices (Bruckert et al., 2006; Collins, 2000; Greisbach, 1999; Pisanski & Rendall, 2011; Rendall et al., 2007; Sell et al., 2010; van Dommelen & Moxness, 1995), voices with computer-manipulated frequencies (Feinberg et al., 2005; Pisanski, Mishra, & Rendall, 2012; Pisanski & Rendall, 2011; Rendall et al., 2007), or synthetic voice-like stimuli (Charlton, Taylor, & Reby, 2013a; Irino, Aoki, Kawahara, & Patterson, 2012; Ives, Smith, & Patterson, 2005; Smith & Patterson, 2005); using voices with fundamental and formant frequencies manipulated independently (Feinberg et al., 2005; Pisanski & Rendall, 2011; Rendall et al., 2007) or in conjunction (Pisanski & Rendall, 2011; Smith & Patterson, 2005); and using sets of voices with relatively subtle (Feinberg et al., 2005) to unnaturally large (Smith, Patterson, & Turner, 2005) variation in fundamental or formant frequencies.

A paradox: the perceptual association between pitch and size.

Formants are generally thought to predict physical size among same-sex adults. Therefore the perceptual association between formants and size is not surprising. It is surprising, though, that listeners consistently associate low voice pitch with largeness at the within-sex level, when no such relationship holds in the physical world (see Chapter 2). Multiple theories have been put forth to explain the persistence of a perceptual relationship between pitch and size in the absence of a robust physical relationship, several of which I test in Chapters 3 and 4. These theories are not mutually exclusive.

One possibility is that voice pitch, arguably the most salient feature of the human voice, is associated with size simply because of its perceptual prominence. When pitch and formants are manipulated in opposite directions in *equally* perceptual units (i.e., just-noticeable differences), listeners prioritize formants over pitch in assessments of both size and masculinity (Pisanski & Rendall, 2011). However, in natural speech, listeners may prioritize pitch over formants if pitch is more perceptually salient than are formants (Collins, 2000; Houde & Jordan, 1998; see also Phillips, Scovil, Carmichael, & Hall, 2007). Other researchers have suggested that listeners may overgeneralize sound-size relationships from the levels at which they are reliable, such as between the sexes or among inanimate objects (Grassi, 2005), to levels at which they are no longer reliable, such as within sexes (Collins, 2000; Rendall et al., 2007).

Perhaps the most common explanation for pitch-size associations appeals to cross-modal perceptual biases and is tied to the classic theoretical work of Morton (1977) and Ohala (1984). Morton (1977) proposed that low-frequency sounds are used across the animal kingdom to signal dominance and hostility, whereas high-frequency sounds function to signal subordination and appeasement. As masculinity and large size are perceptually associated (Pisanski et al., 2012), pitch may be associated with size *indirectly* through its association with masculinity or dominance (Pisanski et al., 2012; Rendall et al., 2007). Similarly, Ohala (1984) proposed the existence of a universal 'frequency code' in human speech that links the perception of low frequency sounds to large size across cultures (i.e., a form of sound symbolism, Hinton, Nichols, & Ohala, 1994). Indeed, some studies have revealed apparently non-random associations between speech sounds and their meaning. For example, in the English language, men's names are more likely to contain 'large' sounding phonemes (such as *oh* and *ah*) than are women's names (Pitcher, Mesoudi, & McElligott, 2013; see also Hinton et al., 1994 for a comprehensive review on sound symbolism). In many languages, pitch is also described in spatial terms such as *low* and *high* or *rising* and *falling* (Eitan & Timmers, 2010), although the perceptual relationship between pitch and space and its effect on size perception has not been extensively studied. I explore the perceptual relationship among pitch and spatial orientation in relation to size estimation in Chapter 4.

Does voice pitch confound accurate size assessment?

All of the above hypotheses that attempt to explain the paradoxical perceptual association between low pitch and large size have one thing in common: they assume that voice pitch is *detrimental* to size assessment. Hence, each hypothesis predicts that voice pitch will lower the accuracy of body size estimates because fundamental frequency is not reliably correlated with physical size within sexes. This prediction is absolutely central to the paradox. Yet, to date, this prediction has not been adequately tested. Although past studies demonstrate that listeners associate low frequencies with large size, and that listeners are generally poor judges of body size based on the voice, studies have not established a causal link between the perception and accuracy of size judgments. In fact, only one previous study investigated whether the accuracy of listeners' size assessments varied as a function of the fundamental and formant frequencies of the vocalizer[4] (Rendall et al., 2007).

In that study, Rendall and colleagues (2007) tested how accurately listeners estimated the relative height of men whose fundamental and/or formant frequencies were varied. The results of the study showed that listeners were more accurate in assessing relative height when men had the same fundamental frequency but the taller man had relatively lower formants, than when the taller man had relatively lower formants but higher fundamental frequency relative to the shorter man. From this, Rendall et al. concluded that voice pitch confounds the accuracy of listeners' body size assessments. However, I argue that this conclusion is not supported by their data.

Although Rendall et al. (2007) examined how relative size assessment accuracy varied as a function of the fundamental or formant frequencies of men's voices, the researchers did not examine accuracy for pairs in which the taller man had relatively lower fundamental frequency than the shorter man. More importantly, the researchers did not *directly* test the hypothesis that voice pitch impairs accurate assessment of body size.

[4] Bruckert et al. (2006) and Collins (2000) examined whether women's judgments of men's height and weight were predicted by men's vocal parameters, but did so using principal components of vocal structure that included variables other than fundamental and formant frequencies. Thus, the studies do not provide a test of the relative roles of fundamental and formant frequencies on the accuracy of listeners' size estimates.

A direct test of this hypothesis would require a comparison between size judgments from voices that contain pitch and voices that do not contain pitch, rather than comparisons among stimuli with varying pitch. Using this proposed design, I present the results of a direct test of the confounding pitch hypothesis in Chapter 3.

The Current Thesis

The main goal of this thesis is to examine acoustic communication of body size in humans, both from the perspective of the vocalizer and from the perspective of the listener. The thesis focuses on the relative roles of fundamental and formant frequencies in communicating size. In Chapter 2, I establish relationships between fundamental and formant frequencies and physical body size using large, cross-cultural samples of adult men and women. In Chapter 3, I examine the effect of voice pitch and formants on accurate estimation of men's relative body size. In Chapter 4, I examine how perceptual pitch biases interact with spatial information to affect body size perception. Taken together, this series of experiments provides a detailed account of the reliability of various vocal indicators of physical size, as well as the psychophysical, perceptual, and sociocultural factors that may influence voice-based size estimation in humans.

CHAPTER 2

VOCAL INDICATORS OF BODY SIZE IN MEN AND WOMEN: A META-ANALYSIS

Pisanski, K., Fraccaro, P.J., Tigue, C.C., O'Connor, J.J.M., Röder, S, Andrews, W.P, Bernhard, F., DeBruine, L.M., Jones, B.C., & Feinberg, D.R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour, 95,* 89-99. Doi: 10.1016/j.anbehav.2014.06.011

Preface

Past work has produced highly variable estimates of the relationship between fundamental frequency and physical height or weight, and between formant frequencies and height or weight, within same-sex groups of men or women (see Tables A1 and A2 in Appendix 2 for relationships reported in past studies). In Chapter 2, I use meta-analytical techniques to assess these relationships at the population-level. I also examine how various demographic and methodological factors contribute to variation in voice-size relationships reported by past studies.

This work provides an answer to the long-standing question of whether acoustic allometry is present in humans when sex and age are controlled. Critically, the meta-analysis provides a reliable test of the hypothesis that formant frequencies may predict body size more reliably than fundamental frequency in humans, akin to many other mammals (Fitch, 1994, 1997, 2000a; Fitch & Giedd, 1999). The results of this work may provide some insight into the selection pressures that have operated on the human voice and how these selection pressures may differ between humans and other mammals (Fitch & Hauser, 2003; Taylor & Reby, 2010), including other primates (Ey et al., 2007; Ghazanfar & Rendall, 2008).

Abstract

Animals often use acoustical cues, such as formant frequencies, to assess the size of potential mates and rivals. Reliable vocal cues to size may be under sexual selection. In most mammals and many other vertebrates, formants scale with vocal tract length allometrically and predict variation in size more reliably than fundamental frequency or pitch (F_0). In humans, however, it is unclear from previous work how well voice parameters predict body size independently of age and sex. We conducted a meta-analysis to establish the strength of various voice–size relationships in adult men and women. We computed mean weighted correlations from 295 coefficients derived from 39 independent samples across five continents, including several novel and large cross-cultural samples from previously unpublished data. Where possible, we controlled for sample size, sample sex, mean age, geographical location, study year, speech type and measurement method, and ruled out publication bias. Eleven of 12 formant-based vocal tract length (VTL) estimates predicted men's and women's heights and weights significantly better than did F_0. Individual VTL estimates explained up to 10% of the variance in height and weight, whereas F_0 explained less than 2% and correlated only weakly with size within sexes. Statistically reliable size estimates from F_0 required large samples of at least 618 men and 2140 women, whereas formant-based size estimates required samples of at least 99 men and 164 women. The strength of voice–size relationships varied by sample size, and in some cases sex, but was largely unaffected by other demographic and methodological variables. We confirm here that, analogous to many other vertebrates, formants provide the most reliable vocal cue to size in humans. This finding has important implications for

honest signalling theory and the capacity for human listeners to estimate size from the voice.

Introduction

Among most terrestrial mammals, including humans, the voice is produced by the larynx and subsequently filtered by the supralaryngeal vocal tract (henceforth, vocal tract; Titze, 1994). The vocal folds within the larynx vibrate to produce the fundamental frequency (F_0) and corresponding harmonics that are perceived as voice pitch, whereas formants are the resonant frequencies of the vocal tract. Because of relatively minimal feedback of vocal tract energy on vocal fold vibration, the source-filter model of speech production treats F_0 and formants as anatomically and functionally independent (Fant, 1960; Titze, 1994).

Source-filter theory was originally developed by speech scientists (Fant, 1960; Singh, & Singh, 1976; Titze, 1994), but has since been applied to the study of nonhuman vocalizations (see, e.g., Fitch & Hauser, 1995, 2003; Ohala, 1983; Owren & Bernacki, 1998; Owren, Seyfarth, & Cheney, 1997; Rendall, Owren, & Rodman, 1998; Sommers, Moody, Prosen, & Stebbins, 1992; Taylor & Reby, 2010). Research has confirmed that F_0 and formants are decoupled in most vertebrates by showing that changes in F_0 and formants do not covary in heliox, a mixture of helium and oxygen that transmits sound twice as fast as does air. In a coupled vocal system, heliox causes both F_0 and formants to shift upward, whereas in a decoupled system, only formants shift upward due to a shortened transit time of sound waves traveling up the vocal tract (Hess et al., 2006). Decoupling has been demonstrated in several species of birds, anurans, bats and many mammalian species including humans (see Fitch & Hauser, 2003).

Fundamental and formant frequencies can provide reliable affective and inferential information (e.g., Morton, 1977; Rendall et al., 1998) as well as reliable cues to sex and age. In humans, for instance, both F_0 and formants are typically lower among men than women (Titze, 1989) and lower among adults than prepubescent children (Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). At puberty, testosterone thickens and lengthens boys' vocal folds, causing F_0 to drop (Harries, Hawkins, Hacking, & Hughes, 1998), and directly affects F_0 throughout adulthood (Abitbol, Abitbol, & Abitbol, 1999; Dabbs & Mallinger, 1999; Damrose, 2009). Testosterone may also contribute to the sexual dimorphism in formant frequencies between men and women (Bruckert, Liénard, Lacroix, Kreutzer, & Leboucher, 2006).

In addition to indicating sex and age, F_0 and formants are the two key acoustic parameters that have traditionally been investigated as potential vocal indicators of body size in humans and in mammals more generally (for reviews see: González, 2006; Taylor & Reby, 2010). Here, work on formant production in animals, particularly nonhuman animals, has produced a number of novel and testable hypotheses. Of particular interest to the present study, Fitch (1994, 1997, 2000) proposed that formants could reliably indicate body size in most vertebrates because the vocal tract is constrained by skeletal structures related to body size, which in turn imposes a constraint on resonances of the vocal tract (i.e., formants). Longer vocal tracts are predicted to produce lower and more closely spaced formants. However, because thicker and longer vocal folds vibrate at lower frequencies, it is also possible that, in some species, larger individuals with larger larynges may produce lower F_0 than smaller individuals (Fitch & Hauser, 2003; Morton, 1977; Titze, 1994). Moreover, because F_0 and formants are typically decoupled in vertebrates, larger individuals with correspondingly longer vocal tracts may produce lower formants regardless of F_0 and vice versa (Fitch, 2000; Fitch & Giedd, 1999; Fitch & Hauser, 2003).

Since the first empirical study less than 20 years ago (Fitch, 1997), there has been a surge of research testing the relative roles of F_0 and formants as honest indicators of size in a wide range of species. This research has generally confirmed that both F_0 and formants independently predict variation in body size among individuals of different species, breeds or clades (primates: Hauser, 1993; dogs: Riede & Fitch, 1999; Taylor, Reby, & McComb, 2008; anurans: Gingras, Boeckle, Herbst, & Fitch, 2013; for reviews see: Ey, Pfefferle, & Fischer, 2006; Fitch & Hauser, 2003; Taylor & Reby, 2010) and between males and females of the same species (e.g., in humans and nonhuman primates: Fitch & Giedd, 1999; González, 2006; Rendall, Kollias, Ney, & Lloyd, 2005; Pfefferle & Fischer, 2006).

Studies of a number of mammalian species have found, however, that formants are a better predictor of size within sexes than is F_0. This pattern of results has been observed in studies of rhesus macaques, *Macaca mulatta* (Fitch, 1997), Japanese macaques, Macaca fuscata (Masataka, 1994), colobus monkeys, *Colobus satanas* (Harris, Fitch, Goldstein, & Fashing, 2006), red deer, *Cervus elaphus* (Reby & McComb, 2003), fallow deer, *Dama dama* (Vannoni & McElligott, 2008), koalas, *Phascolarctos cinereus* (Charlton et al., 2011), elephant seals, *Mirounga leonina* (Sanvito, Galimberti, & Miller, 2007), and dogs, *Canis familiaris* (Riede & Fitch, 1999; see Plotsky, Rendall, Riede, & Chase, 2013 for VTL-size relationships is dogs). It is of interest that both F_0 and formants appear to predict the size of male but not female giant pandas, *Ailuropoda melanoleuca* (Charlton, Zhihe, & Snyder, 2009), whereas F_0 is a better predictor of the size of female hamadryas baboons, *Papio hamadryas,* than are formants (Pfefferle & Fischer, 2006).

It is unclear whether formants predict body size within sexes more reliably than does F_0 in humans. A large proportion of the variation in F_0 and/or formants among humans may be attributed to pubertal expression of sex hormones (Abitbol et al., 1999; Harries et al., 1998) and to differences in body size among men, women and children (Fitch & Giedd, 1999; Smith & Patterson, 2005; Turner, Walters, Monaghan, & Patterson, 2009). Thus, when investigating voice–size relationships within age–sex classes, it is not clear from previous work whether any voice parameter reliably predicts variation in human body size. At the within-sex level, adult body size showed no significant physical relationship with F_0 in more than 80% of correlations reported in published studies and no significant relationship with formants in more than 50% of correlations reported in published studies. The strength and direction of reported correlations ranges widely (range of r for F_0 estimates of size = -.71 to +.30; range of r for formant estimates of size = -.58 and +.32).

In addition to controlling for sex and age, a number of demographic and methodological factors may contribute to the variation in reported relationships between the voice and physical body size across studies. These factors may include the size and geographical location of the sample, the length and content of recorded speech materials, or the equipment and techniques used to measure the voice and body size. Proper adjustment of software settings when measuring vocal parameters, particularly formants, as well as the physical properties of the vocal signal may also affect the strength of reported voice–size relationships (Fitch & Fritz, 2006). Finally, the robustness of formant–size relationships is further complicated by differences in the measures used to relate formant structure to vocal tract length or body size across studies. Traditionally, studies examined size in relation to individual formants (i.e., first to fourth formant, $F1$ to $F4$: González, 2004; Greisbach, 1999; Rendall et al., 2005). More recent work has utilized amalgamated measures of formant structure including mean formant frequency (Fn; Pisanski & Rendall, 2011), geometric mean formant frequency (MFF; Smith & Patterson, 2005), formant dispersion (D_f; Fitch, 1997), formant position (P_f, Puts, Apicella, & Cardenas, 2012), formant spacing (ΔF; Reby & McComb, 2003) and apparent vocal tract length derived from mean formants (VTL(F_i); adapted from Fitch, 1997) or from formant spacing (VTL(ΔF); Reby & McComb, 2003; see equations A1–A7 in Appendix 1). Moreover, a variation of the formant-pattern latent variable model uses confirmatory factor analysis (CFA) to relate factor scores of formants to vocal tract length and height within sexes (Turner et al., 2009). These formant-based measures will henceforth be referred to as VTL (vocal tract length) estimates. It is unknown which of these 12 VTL estimates most reliably predicts human body size.

The present study was designed to establish the strength of various relationships between the voice and body size in men and women at the population level. Until now, the nature of these relationships in humans has been unclear and has been a source of extensive and prolonged debate among researchers. We present the results of a meta-analysis of 295 voice–size correlations derived from 39 human adult samples, including several novel and large cross-cultural samples. The 39 samples derive from North and South America, Europe, Asia and Africa. The goals of the meta-analysis were to (1) determine the degree to which F_0 and formants predict the height and weight of either sex, (2) test the relative reliability of 12 VTL estimates in predicting size, (3) assess the effects of age and a number of other demographic or methodological factors on the strength of voice–size relationships, (4) determine the minimum sample sizes required in future studies to obtain reliable estimates of size from the voice, (5) evaluate the effect of sample size and evidence for publication bias in our sample and (6) compute population-level averages of voice and body size parameters.

Human vocal cues may be under sexual selection (Puts, Jones, & DeBruine, 2012) and may honestly indicate mate quality (Maynard Smith & Harper, 2003; Zahavi & Zahavi, 1997). The present study presents the first comprehensive analysis to examine the reliability of vocal cues to size in humans. This work will provide a base for cross-species comparisons that may reveal the adaptive functions of acoustic communication in mammals, and in particular, the relative roles of fundamental and formant frequencies in signalling body size. Although vocal anatomy does not fossilize, it is similar across mammalian species (Fitch & Hauser, 2003; Titze, 1994; Taylor & Reby, 2010). Thus comparative methods may reveal general principles in voice production and perception among mammals that may ultimately help to uncover the functions and evolutionary origins of mammalian vocalizations, including human speech (Ghazanfar & Rendall, 2008).

This meta-analysis will also complement and build on research investigating the ability of human listeners to accurately gauge body size from the voice (Bruckert et al., 2006; Collins, 2000; Pisanski et al., 2014; Rendall et al., 2007) or voice-like stimuli (Charlton et al., 2013; Smith & Patterson, 2005). Perceptually, listeners associate low F_0 and/or low formants with large body size both between and within sexes (Charlton, Taylor, & Reby, 2013; Feinberg, Jones, Little, Burt, & Perrett, 2005; Pisanski et al., 2014; Pisanski, Mishra, & Rendall, 2012; Pisanski & Rendall, 2011; Rendall, Vokey, & Nemeth, 2007; Smith & Patterson, 2005; van Dommelen & Moxness, 1995). If F_0 or formants do not reliably indicate physical size at the within-sex level, these perceptual associations pose somewhat of a paradox (Pisanski et al., 2012, 2014; Pisanski & Rendall, 2011; Rendall et al., 2007).

Methods

Protocols were approved by the McMaster University Research Ethics Board and adhere to the preferred reporting items for systematic reviews and meta-analyses (PRISMA; Moher, Liberati, Tetzlaff, & Altman, 2009).

Literature Review and Study Exclusion Criteria

We performed a systematic online search using Web of Science (v 5.10, December, 2012), identifying publications reporting direct relationships between voice and body size in human adults from the earliest available year (1976) onward. Search terms included, ‘human’, ‘height’, ‘weight’, ‘mass’, ‘body size’, ‘voice/vocal’, ‘speak*’, ‘fundamental’, ‘formant’, ‘frequency’ and ‘correlation*’. We then manually sought additional sources, including those cited in the reference lists of the primary articles and articles published before 1976.

Our search yielded 17 published studies and five unpublished data sets that collectively rendered 319 F_0–size or formant–size correlations for same-sex adults. Twenty-four correlations were excluded from the meta-analysis for one or more of the following reasons: significance levels were given but correlation coefficients were not (n = 10; note that correlations were included whether or not significance levels were given); correlations were derived from combined factors based on principal component analyses (n = 6) or multiple regression (n = 4); voice measures were not mean based (n = 10). We list all published studies and reported correlations between F_0 and body size in Table A1, and between formants and size in Table A2 (see Appendix 2) including the 24 correlations excluded from the meta-analysis (indicated with an ‘x’; see also Fig. A1). In all cases correlations are reported within, not between, sexes (Rendall et al., 2005). The final meta-analysis included results obtained from 39 independent samples (k) of same-sex adults for a total of 295 correlations. Of these, 67 correlations corresponded to relationships between F_0 and body size (34 height; 33 weight) and 228 correlations corresponded to relationships between various formant-based VTL estimates and body size (122 height; 106 weight).

Voice and Body Size Measures from New Samples

To increase the pooled sample size of the meta-analysis, we included correlations derived from a new cross-cultural data set of 700 adults (k = 8 samples from Canada, U.K. and Germany, N = 428 women, 262 men). We measured men's and women's heights, weights, F_0 and formants. Height (mean ± standard deviation, SD: men: 179.34±7.16 cm; women: 166.37±6.93 cm) was measured using metric tape affixed to the wall, and weight (men: 74.96±12.44 kg; women: 63.28±10.77 kg) was measured using an electronic scale.

All acoustic measurements were performed in Praat (Boersma & Weenink, 2013). We measured mean F_0 using Praat's autocorrelation algorithm with a search range of 65–300 Hz for men and 100–600 Hz for women. We measured $F1$–$F4$ using Praat's Burg Linear Predictive Coding (LPC) algorithm with the maximum formant set to 5000 Hz for men and 5500 Hz for women. Formants were first overlaid on a spectrogram and formant number was manually adjusted until the best visual fit of predicted onto observed formants was obtained. These techniques and settings are recommended by the Praat manual (Boersma & Weenink, 2013; see also Styler, 2014) and have been used in many studies examining relationships between formants and physical or perceived body size (e.g., Evans, Neave, & Wakelin, 2006; Feinberg, Jones, Little, Burt, & Perrett, 2005; González, 2004; Greisbach, 1999; Pisanski, & Rendall, 2011; Rendall, Kollias, Ney, & Lloyd, 2005; for technical details see also Owren & Bernacki, 1998). From the mean $F1$–

$F4$ values, we computed Fn, MFF, D_f, P_f, ΔF, VTL(F_i), VTL(ΔF) and T (see equations A1–A7 in Appendix 1). Additional details are provided in Appendix 2.

Statistical Analysis

Detailed methods of analysis of meta-data are described in Appendix 2 and all equations are given in Appendix 1. Briefly, we computed a mean weighted correlation coefficient, $\bar{r}$ (Hedges & Olkin, 1985; Lipsey & Wilson, 2001), between F_0 and height or weight and between 12 different VTL estimates and height or weight, for each sex independently and collapsing across sample sex. This produced a total of 78 $\bar{r}$ measures. We then performed a series of weighted least squares regressions (henceforth meta-regressions) to assess the effects of various demographic and methodological variables on the strength of reported correlations weighted by an index of sample size. Following this, we created funnel plots for the two most frequently reported voice–size relationships with the largest k, F_0–height and $F4$–height, to assess the influence of sample size and to examine possible publication bias among the studies included in our meta-analysis (Egger, Smith, Schneider, & Minder, 1997). We performed statistical power analyses to determine the minimum sample sizes required in future work to reliably predict height and weight from F_0 or formants (where alpha = 0.05 and power = 0.90; Cohen, 1988). Finally, we derived weighted mean voice and body size measures for each sex by pooling values reported by previous work with those derived from our new samples, again weighted by sample size.

Results

Mean Weighted Voice–Size Correlations

Table 1 shows the mean weighted correlation ($\bar{r}$; Lipsey & Wilson, 2001) between F_0 and height or weight, and between each of 12 different formant-based VTL estimates and height or weight, for samples of men and women separately and collapsing across sample sex. Although some VTL estimates predicted men's and women's heights and weights better than did others, all but one VTL estimate ($F1$) predicted size within sexes more reliably than did F_0. Ninety-three per cent (67 of 72) of the derived mean weighted correlations between formants and body size reached statistical significant (p <.05).

Table 1. Mean weighted correlations ($\bar{r}$) among voice and body size measures derived from the meta-analysis

	Men						Women						All adults					
	$\bar{r}$	$SE_{\bar{r}}$	95% CI	p[a]	k[b]	N[c]	$\bar{r}$	$SE_{\bar{r}}$	95% CI	p[a]	k[b]	N[c]	$\bar{r}$	$SE_{\bar{r}}$	95% CI	p[a]	k[b]	N[c]
Voice–height correlations																		
F_0	-.13	.03	-.19, -.07	<.001	21	1119	-.07	.04	-.14, 0	.063	13	795	-.10	.02	-.15, -.06	<.001	34	1914
$F1$	-.13	.05	-.23, -.03	.01	8	395	-.04	.04	-.12, .04	.29	8	637	-.08	.03	-.14, -.02	.015	16	1032
$F2$	-.22	.05	-.31, -.12	<.001	8	395	-.19	.04	-.26, -.11	<.001	8	637	-.20	.03	-.25, -.14	<.001	16	1032
$F3$	-.26	.05	-.35, -.17	<.001	8	395	-.22	.04	-.29, -.14	<.001	8	637	-.24	.03	-.29, -.18	<.001	16	1032
$F4$	-.30	.05	-.39, -.21	<.001	8	395	-.25	.04	-.32, -.18	<.001	8	637	-.27	.05	-.36, -.17	<.001	16	1032
Fn	-.31	.06	-.42, -.20	<.001	4	262	-.22	.04	-.29, -.14	<.001	4	438	-.25	.04	-.32, -.18	<.001	8	700
MFF	-.28	.06	-.39, -.17	<.001	4	262	-.18	.05	-.27, -.08	<.001	4	438	-.22	.04	-.29, -.14	<.001	8	700
D_f	-.18	.04	-.26, -.10	<.001	9	603	-.24	.05	-.33, -.15	<.001	5	500	-.21	.03	-.27, -.15	<.001	14	1103
P_f	-.29	.05	-.37, -.21	<.001	6	470	-.21	.05	-.30, -.11	<.001	4	438	-.25	.03	-.31, -.19	<.001	10	908
ΔF	-.32	.06	-.43, -.21	<.001	4	262	-.22	.05	-.31, -.13	<.001	4	438	-.26	.04	-.33, -.19	<.001	8	700
VTL(F_i)	.24	.06	.12, .35	<.001	4	262	.17	.05	.08, .26	<.001	4	438	.20	.04	.12, .27	<.001	8	700
VTL(ΔF)	.32	.06	.20, .43	<.001	4	262	.23	.05	.14, .32	<.001	4	438	.26	.04	.19, .33	<.001	8	700
CFA	.30	.06	.40,.19	<.001	1	262	-.24	.06	-.34, -.13	<.001	1	326	.27[d]	.04	.34, .19	<.001	2	588

	Men						Women						All adults					
	$\bar{r}$	$SE_{\bar{r}}$	95% CI	p^{a}	k^{b}	N^{c}	$\bar{r}$	$SE_{\bar{r}}$	95% CI	p^{a}	k^{b}	N^{c}	$\bar{r}$	$SE_{\bar{r}}$	95% CI	p^{a}	k^{b}	N^{c}
Voice–weight correlations																		
F_0	-.03	.03	-.09, .03	.30	21	1132	-.14	.04	-.21, -.07	<.001	12	760	-.07	.02	-.12, -.03	.002	33	1892
$F1$	-.15	.06	-.26, -.04	.01	6	315	-.08	.04	-.16, 0	.064	6	553	-.11	.03	-.17, -.04	.002	12	868
$F2$	-.09	.06	-.20, .02	.12	6	315	-.22	.04	-.30, -.14	<.001	6	553	-.17	.03	-.24, -.11	<.001	12	868
$F3$	-.18	.06	-.29, -.07	.002	6	315	-.16	.04	-.24, -.07	<.001	6	553	-.17	.03	-.23, -.10	<.001	12	868
$F4$	-.15	.06	-.26, -.04	.01	6	315	-.24	.04	-.32, -.16	<.001	6	553	-.21	.03	-.27, -.14	<.001	12	868
Fn	-.22	.06	-.33, -.09	<.001	4	259	-.23	.05	-.32, -.14	<.001	4	436	-.22	.04	-.29, -.15	<.001	8	695
MFF	-.24	.06	-.36, -.12	<.001	4	259	-.20	.05	-.29, -.11	<.001	4	436	-.22	.04	-.29, -.14	<.001	8	695
D_f	-.10	.04	-.18, -.02	.016	9	600	-.21	.05	-.29, -.12	<.001	5	498	-.15	.03	-.21, -.09	<.001	14	1098
P_f	-.25	.05	-.33, -.16	<.001	6	467	-.22	.05	-.31, -.13	<.001	4	436	-.23	.03	-.30, -.17	<.001	10	903
ΔF	-.22	.06	-.34, -.10	<.001	4	259	-.21	.05	-.3, -.12	<.001	4	436	-.22	.04	-.29, -.14	<.001	4	695
VTL(F_i)	.20	.06	.07, .31	.002	4	259	.19	.05	.10, .28	<.001	4	436	.19	.06	.07, .31	<.001	8	695
VTL(ΔF)	.22	.06	.10, .34	<.001	4	259	.25	.05	.16, .33	<.001	4	436	.24	.04	.17, .31	<.001	4	695
CFA	.23	.06	.34, .11	<.001	1	254	-.16	.06	-.26, -.05	<.001	1	323	.19[d]	.04	.26, .11	<.001	2	577

F_0: fundamental frequency; $F1$–$F4$: first to fourth formant; Fn: average formant frequency; MFF: geometric mean formant frequency; D_f: formant dispersion; P_f: formant position; ΔF: formant spacing; VTL(F_i): apparent vocal tract length derived from mean formants; VTL(ΔF): apparent vocal tract length derived from formant spacing; CFA: confirmatory factor analysis (factor scores). See Appendix 1 and Appendix 2 for methods used to derive VTL estimates Fn, MFF, D_f, P_f, ΔF, VTL(F_i), VTL(ΔF) and CFA factor scores (equations A1–A7) and to compute meta-statistics $\bar{r}$, $SE_{\bar{r}}$ and 95% confidence intervals, CI (equations A8–A20).

[a] Two-tailed p value, $\alpha = .05$.

[b] k is the number of independent samples used to derive the mean weighted statistics.

[c] N is the pooled sample size for k samples, $\sum n_{i-k}$.

[d] CFA factor scores correlated positively with height and weight for samples of men but negatively for samples of women due to differential factor loadings. We therefore took the absolute value of correlations to derive the CFA $\bar{r}$ across sexes.

Effects of Sex and Mean Age of Sample

We assessed whether sample sex or mean age affected the strength of various voice–size relationships. For F_0–height relationships ($k \leq 34$ samples), a model that included main and interaction terms for sample sex (fixed factor) and mean age (covariate, available for k = 19) was nonsignificant ($F_{3,15} = 0.48$, $p = .71$) and revealed no significant main or interaction effects (all $F < 0.06$, all $p > .80$). The model remained nonsignificant when we dropped the interaction term (sex-by-age: $F_{2,16} = 0.76$, $p = .49$) or either main effect term (sex: $F_{1,17} = 0.31$, $p = .58$; age: $F_{1,32} = 0.99$, $p = .33$). Thus, neither sample sex nor mean age significantly affected the strength of F_0–height relationships, although we did observe a trend toward stronger F_0–height relationships in samples of men than women (see Table 1). For F_0–weight relationships ($k \leq 33$), a model that included both main and interaction terms for sample sex and mean age (available for k = 18) approached significance ($F_{3,14} = 2.65$, $p = .09$, $R^2 = 0.36$) but revealed no significant main or interaction effects (all $F < 1.63$, all $p > .22$). A model that contained only sample sex as a fixed factor was significant ($F_{1,31} = 4.21$, $p = .049$, $R^2 = 0.12$). Thus, sample sex but not mean age affected the strength of F_0–weight relationships, where F_0 predicted women's weight significantly better than men's weight.

Additional analyses were performed to assess the effect of sample sex on correlations reported for each of 12 VTL estimates ($F1$–$F4$, Fn, MFF, D_f, P_f, ΔF, VTL(F_i), VTL(ΔF), CFA formant scores) and height or weight (k = 8–16). We did not examine the potential effect of mean age on formant–size relationships due to the small number of samples for which age information was available. No model regressing sample sex on formant–height (all F <1.62, p >.22) or formant–weight relationships (all F <2.05, p >.18) reached statistical significance. Thus, sample sex did not affect the strength of formant–height or formant–weight relationships for any of the 12 VTL estimates.

Effects of Other Demographic and Methodological Factors

In addition to sample size, sex and age, several other factors varied across the samples included in our meta-analysis that may have influenced the strength of correlations reported by past studies. Five key additional factors were study year (from 1972 to 2013), geographical location of sample (11 levels: Argentina; Bolivia; Canada; England; Germany; Lebanon; Poland; Scotland; Spain; Tanzania; U.S.A.), type of speech stimulus measured (six levels: one vowel; five vowels; word; sentence; paragraph; a combination of these), acoustic hardware/software used for voice measurement and analysis (seven levels: *FFI* F_0 *indicator*, see Hollien & Harrington, 1977; *Kay Elemetric VISI Pitch*, see Kay Elemetrics, 1996; *MDVP Multi-Dimensional Voice Program*, see Kay Elemetrics, 1993 and Kent, Vorperian, & Duffy, 1999; *Signalize*, see Keller, 1993; *Praat*, see Boersma, 2001 and Boersma & Weenink, 2013; custom software; not specified) and the method of height or weight measurement (three levels: direct measure; self-report; not specified). We assessed the effects of these five variables on F_0–height and F_0–weight relationships only because there were too few samples (k = 4–16), and thus not enough variation in these factors, among studies examining formant–size relationships. There were also too few studies to assess statistically the effect of the technique (e.g., software parameters) used to measure formants (but see Discussion). Study year was included as a covariate, and any other variable as a fixed factor, in independent regressions. Sample sex was included as an additional fixed factor in regressions on F_0–weight relationships.

For F_0–height relationships ($k \leq 34$), the effect of the method used to measure height approached significance ($F_{2,31} = 2.50$, $p = .098$, $R^2 = 0.14$). Thus, although measured and self-reported mean height for men or women differed by only 0.12 and 1.48 cm, respectively, F_0 predicted height marginally better when height was measured using metric tape ($k = 29$, $r = -.17$) than self-report ($k = 4$, $r = .01$). There were no significant effects of study year ($F_{1,32} = 1.58$, $p = .22$), sample location ($F_{10,23} = 1.21$, $p = .34$), speech stimulus ($F_{5,28} = 0.99$, $p = .44$), or voice measurement ($F_{4,29} = 1.25$, $p = .31$). For F_0–weight relationships ($k \leq 33$), no model reached statistical significance (all $F < 2.25$, all $p > 0.10$), although an interaction between sample location and sample sex approached significance ($F_{5,16} = 2.83$, $p = .051$). There were no main or interaction (with sample sex) effects of study year, speech stimulus, voice measurement, or method of weight measurement (all $F < 1.96$, all $p > .16$).

Effect of Sample Size and Tests for Publication Bias

Figure 1 shows funnel plots for F_0–height and $F4$–height relationships (those with the largest k) among samples of men and women. In all cases, the plots revealed greater variance in effect sizes (x-axis, representing the strength of voice–size correlations) among smaller samples than among larger samples (y-axis). There were five outliers among samples of men. One study, falling to the right of the upper confidence interval (Künzel, 1989), reported a positive F_0–height correlation among samples of men that was stronger than would be expected given the average (negative) effect size and the size of the study's sample. Two samples falling to the left of the lower confidence interval (Graddol & Swann, 1983: Canada1) reported stronger than expected negative F_0–height correlations among men (see Fig. 1c). Two samples reflected an unusually strong negative (Rendall et al., 2005) or positive (UK2) $F4$–height correlation among men (see Fig. 1b). However, because these outliers were distributed fairly symmetrically in each plot and/or carried little weight, removing these outliers did not change our average weighted effect sizes or confidence intervals (see Appendix 2). There were no outliers among samples of women. Finally, the plot depicting $F4$–height relationships in men and women was symmetrical; there were an equal number of data points falling to either side of the average weighted effect size for each sex (Fig. 1b). However, the plots depicting F_0–height relationships in men (Fig. 1c) and women (Fig. 1d) revealed a slight skew to the left.

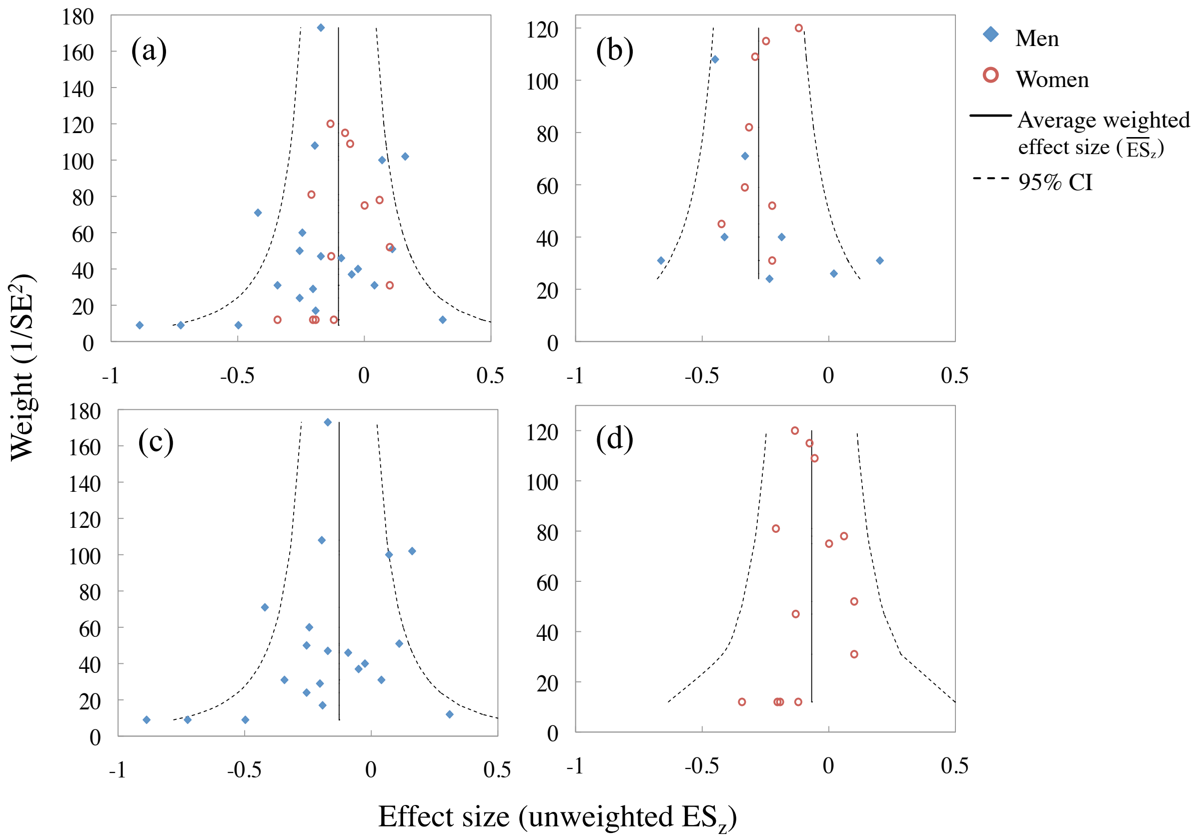

Figure 1. Funnel plots used to assess the effect of sample size and to detect publication bias in reported relationships between (a) F_0 and height in samples of either sex (k = 34 samples, $\overline{ES}_z$ = -0.10), (b) $F4$ and height in samples of either sex (k = 16; $\overline{ES}_z$ = -0.27), or (c) F_0 and height in samples of men (k = 21; $\overline{ES}_z$ = -0.13) and (d) F_0 and height in samples of women (k = 13; $\overline{ES}_z$ = -0.07).

Note. The unweighted standardized effect size of each reported relationship is plotted on the x-axis (ES$_z$; see equation A8) and the inverse of its standard error squared (an index of sample size) is plotted on the y-axis. The solid centre line represents the average weighted standardized effect size across all studies ($\overline{\text{ES}}_z$; see equation A10) and the dotted lines represent the lower and upper 95% confidence intervals corresponding to the y-axis. F_0–height effect sizes are plotted separately for samples of men (c) and women (d) because of a sex difference in $\overline{\text{ES}}_z$ values for this relationship. See Appendix 2 for additional details.

These distributions of data provide no strong evidence of publication bias for any reported voice–size relationships (Egger et al., 1997). Although relatively more studies reported negative F_0–height relationships stronger than the average weighted effect size (i.e., to the left of the average) than F_0–height relationships closer to 0 (i.e., to the right of the average), in both sexes this asymmetry was small. Moreover, removal of outliers or of studies with the smallest samples had no effect. It is important to note that the funnel-like distribution of the data in these plots indicates that sample size can account for the majority of the variation in correlations reported by past studies, wherein studies with larger samples (higher on the y-axis) typically show less variation in the strength of reported correlations than do studies with smaller samples (Egger et al., 1997; see Appendix 2 for additional details).

Minimum Required Sample Sizes for Future Work

Table A3 (Appendix 2) shows the comprehensive results of statistical power analyses that establish the minimum sample sizes required for future work to reliably predict height and weight from F_0 or formant-based VTL estimates (given the $\overline{r}$ ’s that we obtained) with an alpha level of .05 and a conservative power level of .90 (Cohen, 1988). To summarize, samples of at least 618 men and 2140 women are required to predict height from mean F_0. Notably, the average sample sizes of past studies were 55 men and 39 women with no sample exceeding 176 men or 81 women (see Table A1). In contrast, much smaller samples are required to confidently predict height or weight from formants. For instance, height can be predicted with any one of the four strongest VTL estimates (for men: ΔF, VTL(ΔF), *F*4, CFA; for women: *F*4, D_f, VTL(ΔF), CFA) using samples of as few as 99–113 men or 164–195 women.

Our own pooled sample size was large enough that we can be statistically confident in the mean weighted F_0–height correlation reported for men and in all formant–height correlations reported for both sexes, but it was not large enough to confidently predict the F_0–height relationship for women; for this we would require a sample of at least 2140 women in a single study. See Table 1 for all pooled sample sizes used in the calculation of mean weighted correlations (*N*).

Mean Weighted Voice and Body Size Measures

Table 2 shows population-level weighted means and standard deviations of F_0, *F*1–*F*4, *Fn*, MFF, D_f, P_f, ΔF, VTL(ΔF), VTL(F_i), height and weight for men and women. These averages are based on pooled sample sizes of 78–1291 same-sex adults from a wide range of geographical regions. Measured and self-reported height and weight differed by 0.12 cm and 1.88 kg, respectively, among men, and by 1.48 cm and 0.77 kg, respectively, among women.

Table 2. Population-level weighted means ($\overline{X}$) and standard deviations (SD) of voice and body size measures for each sex

	Men		Women	
	Weighted $\overline{X} \pm$ SD	N^{a}	Weighted $\overline{X} \pm$ SD	N^{a}
Voice parameter				
F_0 (Hz)	119.25±16.54	1170	209.74 ±19.53	610
$F1$ (Hz)	452.34±34.67	504	521.24 ±51.59	472
$F2$ (Hz)	1477.03±82.05	504	1840.42±145.8	472
$F3$ (Hz)	2535.55±110.62	504	3012.57±175.4	472
$F4$ (Hz)	3462.34±153.35	504	4099.27±199.96	472
Fn (Hz)	2015.99±90.58	262	2293.55±107.03	438
MFF (Hz)	1589.07±69.85	262	1847.60±94.14	438
D_f (Hz)	1065.84±55.47	520	1194.79±67.56	438
P_f (Z(Hz))	- 0.69±0.34	262	0.30±0.39	438
ΔF (Hz)	1010.66±46.09	262	1187.16±66.29	438
VTL(F_i) (cm)	18.75±1.57	470	15.61±0.88	438
VTL(ΔF) (cm)	17.35±0.79	262	14.78±0.84	438

		Men		Women	
		Weighted $\overline{X} \pm$ SD	N[a]	Weighted $\overline{X} \pm$ SD	N[a]
Body size parameter					
Height (cm)	Measured	178.06±6.79	919	165.96±6.87	664
	Self-report	177.94±6.23	362	167.44±5.1	126
	All[b]	178±6.58	1334	165.96±6.64	871
Weight (kg)	Measured	75.69±11.83	919	62.08±9.94	664
	Self-report	73.13±9.92	362	60.7±8.3	78
	All[b]	75.01±11.05	1291	61.47±9.48	823

F_0: fundamental frequency (voice pitch); $F1$–$F4$: first to fourth formant; Fn: mean formant frequency; MFF: geometric mean formant frequency; D_f: formant dispersion; P_f: formant position; ΔF: formant spacing; VTL(F_i): apparent vocal tract length derived from mean formants; VTL(ΔF): apparent vocal tract length derived from formant spacing (see equations A1–A7). Voice and body size measures are mean based.

[a] N is the pooled sample size for k samples, $\sum n_{i-k}$.

[b] González (2007) did not specify the method of body measurement; therefore, reported height and weight from this study were included only in the 'all' category (52 men, 81 women).

Discussion

Formants Predict Height and Weight Better Than Does Fundamental Frequency

Given considerable inconsistencies in prior reports, the goal of this work was to establish population-level estimates of the strength of various voice–size relationships in adult humans of each sex. The most commonly investigated and heavily debated vocal correlate of human body size is pitch (i.e., fundamental frequency, F_0). Past estimates of F_0–size relationship effect sizes within sexes have a broad range, from strongly negative, to null, to strongly positive. The results of our meta-analysis resolve this debate. The analysis revealed only weak or marginal negative relationships between F_0 and height or weight within either sex. In fact, F_0 accounted for less than 2% of the variance in height or weight within sexes, whereas individual formant-based VTL estimates could explain upwards of 10% of the variance in height or weight within sexes.

Ultimately these results are in line with the proposal that formants are more constrained by anatomical structures related to body size and may therefore predict size more reliably than does F_0 in many mammals (Fitch, 1994, 1997, 2000; Fitch & Giedd, 1999; Fitch & Hauser, 2003). The length and dimensions of the mammalian vocal tract (and resultant formants) are constrained by the bony skull and an individual's height. This has been confirmed by radiographic analysis in rhesus macaques (Fitch, 1997), dogs (Plotsky et al., 2013) and humans (Fitch & Giedd, 1999). In contrast, the larynx is made up of soft tissue and is positioned lower in the vocal tract (Lieberman, McCarthy, Hiiemae, & Palmer, 2001; Titze, 1994). Laryngeal size, movement and resultant F_0 are therefore proposed to vary largely independent of skull and body size (Fitch, 2000) and F_0 may even be modulated to disguise or exaggerate size (Fitch, 2000; Morton, 1977). The strong influence of pubertal and circulating testosterone on F_0 in human males may further contribute to the discordance between men's F_0 and physical size (Dabbs & Mallinger, 1999; Harries et al., 1998). In short, although F_0 can reliably predict variation in size between men and women or adults and children (Hillenbrand et al., 1995; Rendall et al., 2005; Titze, 1989), we confirm here that F_0 is not a reliable predictor of height or weight among same-sex human adults.

Our results parallel those reported for many other groups of mammals in which physical size within sexes is predicted more reliably by formants than by F_0, including several species of ungulates, pinnipeds, elephants, carnivores, marsupials and primates (for reviews see: Ey et al., 2007; Fitch & Hauser, 2003; Taylor & Reby, 2010). These cross-species commonalities in vocal cues to size have been linked to a common mode of vocal production in mammals resulting from a two-stage source-filter process (Taylor & Reby, 2010). However, there are some notable exceptions. Voice pitch appears to predict size as reliably as or more reliably than do formants in male giant pandas (Charlton et al., 2009) and female hamadryas baboons (Pfefferle & Fischer, 2006). Like humans (Fitch & Giedd, 1999; Lieberman et al., 2001), some deer and gazelle species are known to have descended larynges (Fitch & Reby, 2001; Frey et al., 2008; Frey, Volodin, Volodina, Soldatova, & Juldaschev, 2011; McElligott, Birrer, & Vannoni, 2006). A descended larynx effectively elongates the vocal tract and lowers formants, and may have evolved for deceptive size exaggeration (Fitch, 1999; Fitch & Reby, 2001*)*. However, due to physiological limits on vocal tract elongation (e.g., in red and fallow deer, the sternal attachment of the sternothyroid muscle acts to constrain laryngeal descent; see Fitch & Reby, 2001), formants may correlate with size even in species with descended larynges. Indeed, this is true for red and fallow deer (Taylor & Reby, 2010) and the current Chapter demonstrates that this is also true for humans. Further comparative work is needed to identify the sensory, evolutionary and environmental factors that could have caused vocal signalling systems to diverge in mammals. However, some of the reported cross-species differences in vocal cues to size may be related to voice measurement, an issue that is discussed in greater detail below.

We found that reliable F_0 -based estimates of size require sample sizes 4–10 times larger than do formant-based VTL estimates. This provides further support for the relative robustness of formant cues to size. Because all previous work has fallen short of these minimum sample sizes, this finding also highlights the necessity for future work to utilize substantially larger samples in order to obtain statistically reliable predictions of body size from the voice.

The Relative Reliability of 12 Different VTL Estimates of Body Size

Our meta-analysis included 12 different formant-based VTL estimates used in past studies to estimate physical size in humans and other mammals (Bruckert et al., 2006;

Charlton et al., 2009, 2011, 2013; Collins, 2000; Collins & Missing, 2003; Evans et al., 2006; Ey et al., 2007; Feinberg et al., 2005; 2008; Fitch, 1997; 2000; Gingras et al., 2003; González 2006, 2007; Graddol & Swann, 1983; Hamdan et al., 2012; Harris et al., 2006; Pfefferle, & Fischer, 2006; Pisanski & Rendall, 2011; Pisanski et al., 2014; Plotsky et al., 2013; Puts, Apicella, & Cardenas, 2012; Reby & McComb, 2003; Rendall et al., 2005, 2007; Riede & Fitch, 1999; Sanvito et al., 2007; Sell et al., 2010; Smith & Patterson, 2005; Taylor et al., 2010; Turner et al., 2009; Vannoni & McElligott, 2008; Wyman et al., 2012). In addition to the individual formants $F1$–$F4$, we sought to determine which of the amalgamated measures, Fn, MFF, D_f, P_f, ΔF, VTL(F_i), VTL(ΔF), and factor scores from a latent variable CFA model best predicted height and weight within sexes in humans. These latter, more recent measures of formant structure may, arguably, be more reliable predictors of VTL and size than are individual formants because they control for variation across formants and within individuals (Greisbach, 1999).

We found that, among $F1$–$F4$, the higher the individual formant the better it predicted body size. This may be due to the role of $F1$ and $F2$ in speech perception, wherein the relative positions of these lower formants shift more across speech sounds than do the positions of higher formants. This suggests that higher formants may, as a consequence, provide more consistent and reliable information about the dimensions of an individual's vocal tract (Greisbach, 1999; Peterson & Barney, 1952). Differences in the predictive power of individual formants may also be tied to differences in how these formants relate to various cavities of the vocal tract across development and in adulthood

(Fant, 1960, p. 121; Fitch & Giedd, 1999; Turner et al., 2009).

Each of the formant-based VTL estimates predicted variation in body size to a similar degree as did all other VTL estimates, barring *F*1, which was a relatively poor predictor of size. However, the relative predictive power of each VTL estimate varied to some extent for height or weight and among samples of men or women. Thus, ΔF and VTL(ΔF) best predicted men's height, followed closely by *Fn, F*4 and CFA, whereas P_f best predicted men's weight, followed by MFF. Among women, *F*4 best predicted women's height, followed by D_f, CFA and VTL(ΔF), whereas VTL(ΔF) best predicted women's weight, followed by *F*4 and *Fn*.

Theoretical Implications for Vocal Size Communication in Humans

Our results suggest that the relationship between F_0 and body size is likely too weak to have strongly influenced the evolution and perceptions of sex differences in human vocal features, particularly given that stable ancestral (and indeed modern-day) social groups are predicted to have contained only 150–200 individuals (Dunbar, 1993; Hamilton, Milne, Walker, Burger, & Brown, 2007). Similarly, with social groups this small, humans may have difficulty estimating the relative size of potential mates or rivals from vocal cues alone. Indeed, human listeners can correctly identify the taller of two men from their voices only about 60% of the time (González, 2006; Pisanski et al., 2014; Rendall et al., 2007). The sexual dimorphism in human height and mass is small relative to that of other great apes (Leigh & Shea, 1995; Lindenfors, Gittleman, & Jones, 2007).

Sexual dimorphism in human size is also much smaller than the sexual dimorphism in human F_0 (Rendall et al., 2005).

Taken together this evidence suggests that variation in human F_0, and to some degree in formants, is not likely due to strong sexual selection on vocal indicators of physical size (or at least not honest indicators of size). Likewise, listeners may not have evolved to reliably assess body size from the voice. Instead, F_0 and formants may more reliably indicate related traits such as attractiveness, masculinity, dominance and threat than physical size per se (reviewed in: Feinberg, 2008; Puts, Jones, & DeBruine, 2012), and may covary with or contribute to the perception of these dimensions in both men and women. This possibility is further supported by evidence that F_0 (Dabbs & Mallinger, 1999; Evans, Neave, Wakelin, & Hamilton, 2008) and formants (Bruckert et al., 2006) can predict circulating levels of testosterone in men.

It is important to note that multiple factors are likely to affect the degree to which either F_0 or formants correlate with physical size. We cannot, for instance, rule out the possibility that relationships between voice and physical size are obscured by voice modulation (see, e.g., Cartei, Cowles, & Reby, 2012; Fraccaro et al., 2013; Puts, Gaulin, & Verdolini, 2006). Indeed, selection may have favoured individuals who were able to exaggerate their size by manipulating F_0 or formants (Fitch 1999, 2000; Fitch & Reby, 2001; Morton, 1977). Moreover, we cannot be certain whether relationships between voice and physical size, or perception thereof, are a direct product of selection pressure on the vocal apparatus for the purpose of communicating size (Ey et al., 2007; Fitch, 2000), or whether the evolution of vocal cues to body size is simply a by-product of allometric scaling or perceptual bias.

In the absence of a strong physical relationship, the strong perceptual association between F_0 and size poses a paradox. This is particularly true in light of evidence that listeners are generally poor at accurately estimating size from the voice (reviewed in Pisanski et al., 2014; Pisanski & Rendall, 2011). Several hypotheses have been proposed to resolve this paradox (see, e.g., Morton, 1977; Ohala, 1983; Rendall et al., 2007). Most recently, studies have revealed that F_0 can in fact facilitate the perception (Charlton et al., 2013) and accurate estimation (Pisanski et al., 2014) of men's body size. In part this is because voices with relatively low F_0 have more closely spaced harmonics than do voices with higher F_0 (Ryalls & Lieberman, 1982). As a result, lower-pitched voices provide a denser carrier signal for formants from which to gauge size (see Pisanski et al., 2014, for additional discussion). Further explanation for the discrepancy between physical and perceived vocal cues relative to size in humans, and how exactly this discrepancy affects accurate size estimation, is an important avenue for future work.

Demographic and Methodological Influences

Age did not affect the strength of F_0–size relationships in meta-regressions. This is likely because the samples included in our meta-analysis largely if not entirely comprised postpubertal and premenopausal adults (mean ± SD age = 22.4±4.8; range 17–

36). Although F_0 and formants gradually decrease with age across an individual's lifetime (Bruckert et al., 2006; Evans et al., 2006; Feinberg et al., 2005; Hillenbrand et al., 1995), these voice changes are most pronounced during puberty (around age 13) and, for women, during menopause (around age 50; Abitbol et al., 1999; Boulet & Oddens, 1996; Hollien, Green, & Massey, 1994). Thus, body size and voice parameters are not predicted to undergo substantial changes in the age range studied in our meta-analysis. Note, however, that mean age data were available for only approximately 60% of the samples in our meta-analysis and, for some meta-regressions (e.g., formants on height and weight), too few to assess its effects reliably.

Sample sex did not significantly affect the strength of correlations between F_0 and height or between formants and height or weight in our meta-analysis. However, F_0 predicted variation in weight substantially better among samples of women than among samples of men. Low F_0 significantly predicted heavier body weight in women, but we found no relationship between F_0 and weight in men. Our results corroborate those of Vukovic, Feinberg, DeBruine, Smith and Jones (2010), who reported a significant negative correlation between women's F_0 and factor scores derived from a principal component analysis of women's weight, body mass index, percentage body fat, waist and hip circumference and waist-to-hip ratio. Considering the complex interplay among fat distribution, sex hormones and the female voice (Abitbol et al., 1999; Evans, Hoffman, Kalhhoff, & Kissebah; 1983; Friedman, 2011; Hughes, Dispenza, & Gallup, 2004; Singh, 1993), the relationship between F_0 and women's body weight or shape may reflect individual differences in women's sex hormone levels rather than body size per se.

Relationships between F_0 and height were marginally but not significantly stronger in studies that measured height directly rather than through self-report. Other methodological and demographic variation across samples included in the meta-analysis, such as study year, country, speech stimulus and type of acoustic analysis, had no effect on the strength of voice–size correlations. Thus, vocal cues to size appear to be cross-culturally stable, but future studies should investigate the strength of these relationships in other human populations. The strength of vocal estimates of size did not vary as a function of the program used for acoustic analysis or whether the speech stimulus measured was a vowel, word, sentence, paragraph, or any combination of these.

Techniques used to measure and analyse vocal parameters may account for some portion of the variation in size estimates reported in past studies. This is particularly true for formant analysis where improper techniques can cause errors in measurement (see, e.g., Turner et al., 2009; Fitch & Fritz, 2006; but see Burris, Vorperian, Fourakis, Kent, & Bolt, 2014). It is important to adjust acoustic software settings to the signal. For instance, in analyses of human speech in Praat (Boersma & Weenink, 2013), F_0 search ranges and maximum formant settings should be set lower for men's voices than for women's and children's voices, and the number of formants selected in the 'formant settings' window should be adjusted for best visual fit (see Methods). Dynamic range in Praat is set to 50 dB by default for formant tracking and should be adjusted depending on the signal-to-noise ratio of the signal. This changes the contrast in a spectrogram and can improve the visbility of formants to the formant algorithm and human user. These and other techniques for appropriate measurement and manipulation of a variety of vocal parameters are given in Fitch and Fritz (2006), Owren (2008) and Owren and Bernacki (1998), as well as in Praat manuals available online (Boersma & Weenink, 2013; Styler, 2014).

Variation in vocal measurement may also arise purely as a result of physical differences among speech signals. For example, because men's voices are characterized by more densely spaced harmonics than are women's and children's voices (Hildebrand et al., 1995; Titze, 1989), it is often easiest to measure formants from men's voices. Likewise, although the generalization of source-filter theory from human to nonhuman vocal production has facilitated measurement and resynthesis of animal calls, many types of animal vocalizations do not lend themselves to F_0 and formant measurement. This is true for chaotic vocalizations and those in which F_0 is higher than $F1$ (Fitch & Fritz, 2006). Certain preconditions must be met when analysing animal calls using software, such as Praat, designed for human speech analysis (see Owren & Bernacki, 1998). Taking these many factors into consideration, employing standardized protocols for vocal analysis and reporting software parameters is paramount.

Despite the appreciable variation in voice–size correlations reported across past studies (see Tables A1 and A2), we did not find evidence of strong systematic publication bias. Funnel plots demonstrated that the majority of the variation in past work can be attributed to sample size, with substantially greater variation in voice–size relationships reported by smaller studies. This, along with the results of our power analyses, again highlights the necessity for future studies to use considerably larger sample sizes to relate voice to body size reliably at the within-sex level in humans.

Conclusions

Following a long and ongoing history of disagreement among voice researchers (González, 2003, 2007; Graddol & Swann, 1983; Hollien & Jackson, 1973; Künzel, 1989; Lass & Brown, 1978; Morton, 1977; Puts, Apicella, et al., 2012; Rendall et al., 2007), fundamental frequency (F_0) was found to correlate only weakly, if at all, with body size within either sex of human adults. In contrast, almost all VTL estimates correlated strongly with men's and women's heights and weights. Individually, four VTL estimates proved to be particularly robust independent predictors of body size in same-sex samples of as few as 99 men or 164 women. These were VTL(ΔF) and ΔF (Reby & McComb, 2003), Fn (Pisanski & Rendall, 2011) and P_f (Puts, Apicella, et al., 2012). Currently we are investigating the strength of relationships among additional features of the human voice, such as minimum and maximum F_0, jitter, shimmer, harmonics-to-noise ratio and additional indices of body size, including circumferences of the hips, waist and chest, waist-to-hip ratio, chest-to-hip ratio and body mass index.

An integrated account of human vocal communication is imperative if we are to understand the selection pressures that have shaped the voice and language, and that continue to shape the social communication and mating systems of humans (Feinberg, 2008; Fitch, 2000; Puts, Jones, et al., 2012). Research on voice production should be merged with research on voice perception in order to understand how vocal signals have been shaped by the complex interplay between sender and receiver (Dawkins & Krebs, 1978; Owren, Rendall, & Ryan, 2010; Zahavi & Zahavi, 1997). Similarly, because the vocal production apparatus is similar across mammalian species (Fitch & Hauser, 2003; Taylor & Reby, 2010; Titze, 1994), cross-species comparisons will be essential to advance theory and research in animal bioacoustics and behavioural ecology. Indeed, recent comparative work on mammalian acoustic communication guided by the source-filter framework has generated several novel and testable hypotheses (Ey et al., 2007; Fitch & Hauser, 2003; Gingras et al., 2013; Taylor & Reby, 2010). The present finding that formants relate to size more reliably than does voice pitch in a number of mammalian species, including our own, provides a strong base from which to make further predictions about the mechanisms that drive or constrain voice production and perception in mammals and that contribute to variation across species.

References

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice, 13*, 424–446.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*, 341–345.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer*. [Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/.

Boulet, M.J., & Oddens, B.J. (1996). Female voice changes around and after the menopause: an initial investigation. *Maturitas, 23*, 15–21.

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 B: Biological Sciences, 273*, 83–89.

Burris, C., Vorperian, H.K., Fourakis, M., Kent, R.D., & Bolt, D.M. (2014). Quantitative and descriptive comparison of four acoustic analysis systems: vowel measurements. *Journal of Speech, Language, and Hearing Research*, *57,* 26–45.

Cartei, V., Cowles, H.W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PLoS One, 7*, e31353.

Charlton, B.D., Ellis, W.A.H., McKinnon, A.J., Cowin, G.J., Brumm, J., & Nilsson, K. (2011). Cues to body size in the formant spacing of male koala (*Phascolarctos* *cinereus*) bellows: honesty in an exaggerated trait. *Journal of Experimental Biology, 214*, 3414–3422.

Charlton, B.D., Taylor, A.M., & Reby, D. (2013). Are men better than women at acoustic size judgments? *Biology Letters*, *9*, 20130270.

Charlton, B.D., Zhihe, Z., & Snyder, R.J. (2009). The information content of giant panda, *Ailuropoda melanoleuca*, bleats: acoustic cues to sex, age and size. *Animal Behaviour*, *78*, 893–898.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: L. Erlbaum.

Collins, S.A. (2000). Men's voices and women's choices. *Animal Behaviour, 60*, 773–780.

Collins, S.A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65*, 997–1004.

Dabbs, J.M., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences, 27*, 801–804.

Damrose, E.J. (2009). Quantifying the impact of androgen therapy on the female larynx. *Auris Nasus Larynx, 36*, 110–112.

Dawkins, R., & Krebs, J.R. (1978). Animal signals: information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 282–309). Oxford, U.K.: Blackwell.

Dunbar, R.I.M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences, 16*, 681–693.

Egger, M., Smith, G.D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.

Evans, D.J., Hoffman, R.G., Kalhhoff, R.K., & Kissebah, A.H. (1983). Relationship of androgenic activity to body fat topography, fat cell morphology, and metabolic aberrations in premenopausal women. *Journal of Clinical Endocrinology and Metabolism, 57*, 304–310.

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*, 160–163.

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior, 93*, 783–788.

Ey, E., Pfefferle, D., & Fischer, J. (2007). Do age- and sex-related variations reliably reflect body size in non-human primate vocalizations? A review. *Primates*, *48,* 253–267.

Fant, F. (1960). *Acoustic theory of speech production*. The Hague, The Netherlands:

Mouton.

Feinberg, D.R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary Anthropology, 17*, 112–118.

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*, 561–568.

Fitch, W.T. (1994). *Vocal tract length perception and the evolution of language* (Doctoral dissertation). Providence, RI: Brown University.

Fitch, W.T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America, 102*, 1213–1222.

Fitch, W.T. (1999). Acoustic exaggeration of size in birds via tracheal elongation: comparative and theoretical analyses. *Journal of Zoology, 248*, 31–48.

Fitch, W.T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences, 4*, 258–267.

Fitch, W.T., & Fritz, J.B. (2006). Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *Journal of the Acoustical Society of America, 120,* 2132–2141.

Fitch, W.T., & Giedd, J. (1999). Morphology and development of the human vocal tract:

a study using magnetic resonance imaging. *Journal of the Acoustical Society of America, 106*, 1511–1522.

Fitch, W.T., & Hauser, M.D. (1995). Vocal production in nonhuman primates: acoustics, physiology, and functional constraints on 'honest' advertisement. *American Journal of Primatology, 37,* 191–219.

Fitch, W.T., & Hauser, M. (2003). Unpacking 'honesty': vertebrate vocal production and the evolution of acoustic signals. In A. Simmons, A. N. Popper, & R. R. Fay (Eds.), *Acoustic communication* (pp. 65–137). New York, NY: Springer.

Fraccaro, P.J., O'Connor, J.J., Re, D.E., Jones, B.C., DeBruine, L.M., & Feinberg, D.R. (2013). Faking it: deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour, 85*, 127–136.

Frey, R., Gebler, A., Olson, K.A., Odonkhuu, D., Fritsch, G., Batsaikhan, N., & Stuermer, I.W. (2008). Mobile larynx in Mongolian gazelle: retraction of the larynx during rutting barks in male Mongolian gazelle (*Procapra gutturosa* Pallas, 1777). *Journal of Morphology*, *269,* 1223–1237.

Frey, R., Volodin, I., Volodina, E., Soldatova, N.V., & Juldaschev, E.T. (2011). Descended and mobile larynx, vocal tract elongation and rutting roars in male goitred gazelles (*Gazella subgutturosa* Güldenstaedt, 1780). *Journal of Anatomy*, *218*, 566–585.

Friedman, A.D. (2011). Adiposity, estrogen, and voice: the opera has just begun.

Menopause, 18, 723–724.

Ghazanfar, A.A., & Rendall, D. (2008). Evolution of human vocal production. *Current Biology*, *18*, R457–R460.

Gingras, B., Boeckle, M., Herbst, C., & Fitch, W.T. (2013). Call acoustics reflect body size across four clades of anurans. *Journal of Zoology, 289*, 143–150.

González, J. (2003). Estimation of speakers' weight and height from speech: a re-analysis of data from multiple studies by Lass and colleagues. *Perceptual and Motor Skills, 96*, 297–304.

González, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics, 32*, 277–287.

González, J. (2006). Research in acoustics of human speech sounds: correlates and perception of speaker body size. *Recent Research Development in Applied Physics, 9*, 1–15.

González, J. (2007). Correlations between speakers' body size and acoustic parameters of voice. *Perceptual and Motor Skills, 105*, 215.

Graddol, D., & Swann, J. (1983). Speaking fundamental frequency: some physical and social correlates. *Language and Speech, 26*, 351–366.

Greisbach, R. (1999). Estimation of speaker height from formant frequencies. *Forensic*

Linguistics, 6, 265–277.

Hamdan, A.L., Al-Barazi, R., Tabri, D., Saade, R., Kutkut, I., & Sinno, S. (2012). Relationship between acoustic parameters and body mass analysis in young males. *Journal of Voice, 26*, 144–147.

Hamilton, M.J., Milne, B.T., Walker, R.S., Burger, O., & Brown, J.H. (2007). The complex structure of hunter–gatherer social networks. *Proceedings of the Royal Society B: Biological Sciences*, *274*, 2195–2203.

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. *Journal of Laryngology and Otology, 112*, 451–454.

Harris, T.R., Fitch, W.T., Goldstein, L.M., & Fashing, P.J. (2006). Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass. *Ethology, 112*, 911–920.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hess, D.R., Fink, J.B., Venkataraman, S.T., Kim, I.K., Myers, T.R., & Tano, B.D. (2006). The history and physics of heliox. *Respiratory Care*, *51*, 608–612.

Hillenbrand J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099.

Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *Journal of the Acoustical Society of America, 96*, 2646–2654.

Hollien, H., & Harrington, W. (1977). Fundamental frequency indicator (FFI). *Occasionally*, *2*, 4–6.

Hollien, H., & Jackson, B. (1973). Normative data on the speaking fundamental frequency characteristics of young adult males. *Journal of Phonetics, 1*, 117–120.

Hauser, M. (1993). The evolution of nonhuman primate vocalizations: effects of phylogeny, body weight, and social context. *American Naturalist*, *142*, 528–542.

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*, 295–304.

Irino, T., Aoki, Y., Kawahara, H., & Patterson, R.D. (2012). Comparison of performance with voiced and whispered speech in word recognition and mean-formant-frequency discrimination. *Speech Communication, 54*, 998–1013.

Kay Elemetrics (1993). *Multi-Dimensional Voice Program (MDVP).* [Computer program]. Pine Brook, NJ: Kay Elemetrics.

Kay Elemetrics (1996). *Visi-Pitch II Model 3300 instruction manual*. Lincoln Park, NJ: Kay Elemetrics.

Keller, E. (1993). *Signalize. Signal analysis for speech arid sound*. Charlestown, MA:

Network Technology.

Kent, R.D., Vorperian, H.K., & Duffy, J.R. (1999). Reliability of the Multi-Dimensional Voice Program for the analysis of voice samples of subjects with dysarthria. *American Journal of Speech-Language Pathology*, 8, 129.

Künzel, H.J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica, 46*, 117–125.

Lass, N.J., & Brown, W.S. (1978). Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies. *Journal of the Acoustical Society of America, 63*, 1218–1220.

Latinus, M., McAleer, P., Bestelmeyer, P.E., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, *23,* 1075–1080.

Leigh, S.R., & Shea, B.T. (1995). Ontogeny and the evolution of adult body size dimorphism in apes. *American Journal of Primatology*, *36*, 37–60.

Lieberman, P. (1968). Primate vocalizations and human linguistic ability. *Journal of the Acoustical Society of America, 44*, 1574–1584.

Lieberman, D.E., McCarthy, R.C., Hiiemae, K.M., & Palmer, J.B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology, 46*, 117–128.

Lindenfors, P., Gittleman, J.L., & Jones, K.E. (2007). Sexual size dimorphism in mammals. In D. J. Fairbairn, W. U. Blanckenhorn & T. Szeley (Eds.), *Sex, size and gender roles: Evolutionary studies of sexual size dimorphism* (pp. 16–26). Oxford, U.K.: Oxford University Press.

Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. New York, NY: Sage.

Masataka, N. (1994). Lack of correlation between body size and frequency of vocalisations in young female Japanese macaques (*Macaca fuscata*). *Folia Primatol*ogica, *63*, 115–118.

Majewski, W., Hollien, H., & Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males. *Phonetica, 25*, 119–125.

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford, U.K.: Oxford University Press.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine, 151*, 264–269.

Morton, E.S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, *111,* 855–869.

Ohala, J.J. (1983). Cross-language use of pitch: an ethological view. *Phonetica, 40*, 1–18.

Owren, M.J. (2008). GSU Praat tools: scripts for modifying and analyzing sounds using

Praat acoustics software. *Behavior Research Methods*, *40*, 822–829.

Owren, M.J., & Bernacki, R.H. (1998). Applying linear predictive coding (LPC) to frequency-spectrum analysis of animal acoustic signals. In S. L. Hopp, M. J. Owren & C. S. Evans (Eds.), *Animal acoustic communication: Sound analysis and research methods* (pp. 130–162). New York, NY: Springer.

Owren, M.J., Rendall, D., & Ryan, M.J. (2010). Redefining animal signaling: influence versus information in communication. *Biology and Philosophy*, *25*, 755–780.

Owren, M.J., Seyfarth, R., & Cheney, D. (1997). The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): implications for production processes and functions. *Journal of the Acoustical Society of America, 101*, 2951–2963.

Patterson, R. D., Smith, D. R. R., van Dinther, R., & Walters, T. C. (2008). Size information in the production and perception of communication sounds. In W. A. Yost, Popper, A. N., & R. R. Fay, (Eds.), *Auditory perception of sound sources* (pp. 43–75). New York, NY: Springer.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175–184.

Pfefferle, D., & Fischer, J. (2006). Sounds and size: identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*. *Animal*

Behaviour, *72*, 43–51.

Pisanski, K., Fraccaro, P.J., Tigue, C.C., O'Connor, J.J.M., & Feinberg, D.R. (2014). Return to Oz: Voice pitch facilitates assessment of men's body size. *Journal of Experimental Psychology: Human Perception & Performance, 40*(4), 1316-1331. Doi: http://dx.doi:10.1037/a0036956

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,* 509–519.

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

Plotsky, K., Rendall, D., Riede, T., & Chase, K. (2013). Radiographic analysis of vocal tract length and its relation to overall body size in two canid species. *Journal of Zoology, 291*, 76–86.

Puts, D.A., Apicella, C.L., & Cardenas, R.A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 279*, 601–609.

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*, 283–296.

Puts, D.A., Jones, B.C., & DeBruine, L.M. (2012). Sexual selection on human faces and voices. *Journal of Sex Research*, *49*, 227–243.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour, 65*, 519–530.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F-0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America, 117*, 944–955.

Rendall, D., Owren, M. J., & Rodman, P.S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *Journal of the Acoustical Society of America, 103*, 602–614.

Rendall, D., Vokey, J.R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1208–1219.

Riede, T., & Fitch, W.T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). *Journal of Experimental Biology, 202*, 2859–2867.

Sanvito, S., Galimberti, F., & Miller, E.H. (2007). Vocal signalling of male southern elephant seals is honest but imprecise. *Animal Behaviour, 73*, 287–299.

Sell, A., Bryant, G.A., Cosmides, L., Tooby, J., Sznycer, D., & von Rueden, C. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences, 277*, 3509–3518.

Singh, D. (1993). Adaptive significance of female physical attractiveness: role of waist-to-hip ratio. *Journal of Personality and Social Psychology, 65*, 293.

Singh, S., & Singh, K. (1976). *Phonetics: Principles and practices*. Baltimore, MD: University Park Press.

Smith, D.R., & Patterson, R.D. (2005). The interaction of glottal-pulse rate and vocal tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America, 118*, 3177–3186.

Sommers, M.S., Moody, D.B., Prosen, C.A., & Stebbins, W.C. (1992). Formant frequency discrimination by Japanese macaques (*Macaca fuscata*). *Journal of the Acoustical Society of America*, *91*, 3499–3510.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*, 1046–1055.

Styler, W. (2014). *Using Praat for linguistic research*. Version 1.4.5. Retreived from http://savethevowels.org/praat.

Taylor, A.M., Reby, D., & McComb, K. (2010). Size communication in domestic dog, *Canis familiaris*, growls. *Animal Behaviour*, *79*, 205–210.

Taylor, A.M., & Reby, D. (2010). The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology, 280*, 221–236.

Titze, I.R. (1989). Physiological and acoustic differences between male and female voices. *Journal of the Acoustical Society of America, 85*, 1699–1707.

Titze, I.R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Turner, R.E., Walters, T.C., Monaghan, J.J.M., & Patterson, R.D. (2009). A statistical, formant-pattern model for segregating vowel type and vocal tract length in developmental formant data. *Journal of the Acoustical Society of America, 125*, 2374.

van Dommelen, W.A., & Moxness, B.H. (1995). Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and Speech, 38*, 267–287.

Vannoni, E., & McElligott, A.G. (2008). Low frequency groans indicate larger and more dominant fallow deer (*Dama dama*) males. *PLoS One, 3,* e3113.

Vukovic, J., Feinberg, D.R., DeBruine, L., Smith, F.G., & Jones, B.C. (2010). Women's voice pitch is negatively correlated with health risk factors. *Journal of Evolutionary Psychology, 8,* 217–225.

Wyman, M.T., Mooring, M.S., McCowan, B., Penedo, M.C.T., Reby, D., & Hart, L.A. (2012). Acoustic cues to size and quality in the vocalizations of male North

American bison, *Bison bison*. *Animal Behaviour*, *84,* 1381–1391.

Zahavi, A., & Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle.* Oxford, U.K.: Oxford University Press.

CHAPTER 3

RETURN TO OZ: VOICE PITCH FACILITATES ASSESSMENTS OF MEN'S BODY SIZE

Pisanski, K., Fraccaro, P.J., Tigue, C.C., O'Connor, J.J.M., & Feinberg, D.R. (2014). Return to Oz: Voice pitch facilitates assessment of men's body size. *Journal of Experimental Psychology: Human Perception & Performance, 40*(4), 1316-1331. Doi: http://dx.doi:10.1037/a0036956

Preface

Having established that formants reliably predict variation in men's and women's heights and weights, whereas fundamental frequency does not (see Chapter 2), in Chapter 3, I test the hypothesis that pitch (the acoustic correlate of fundamental frequency and/or corresponding harmonics) confounds accurate size assessment (González, 2006; Pisanski et al., 2012; Rendall et al., 2007). To directly test this hypothesis, I examine whether listeners are more accurate in relative assessments of men's body size with voices that contain pitch (regular, voiced modal speech), than voices of the same men that do not contain pitch (whispered speech and synthesized three-formant sine-wave speech). As a control, I confirm that the *perceptual* association between low formants and large size is unaffected by the absence of pitch from the voice. Finally, I test whether voice pitch has an indirect, facilitating effect on the accuracy of listeners' size assessments. This work presents novel findings that help us to understand why the perceptual association between pitch and size exists in the absence of a reliable physical relationship.

Abstract

Listeners associate low voice pitch (fundamental frequency and/or harmonics) and formants (vocal tract resonances) with large body size. Although formants reliably predict size within sexes, pitch does not reliably predict size in groups of same-sex adults. Voice pitch has therefore long been hypothesized to confound within-sex size assessment. Here we performed a knockout test of this hypothesis using whispered and three-formant sine-wave speech devoid of pitch. Listeners estimated the relative size of men with above-chance accuracy from voiced, whispered, and sine-wave speech. Critically, although men's pitch and physical height were unrelated, the accuracy of listeners' size assessments increased in the presence rather than absence of pitch. Size assessments based on relatively low pitch yielded particularly high accuracy (70-80%). Results of Experiment 2 revealed that amplitude, noise, and signal degradation of unvoiced speech could not explain this effect; listeners readily perceived formant shifts in manipulated whispered speech. Rather, in Experiment 3, we show that the denser harmonic spectrum provided by low pitch allowed for better resolution of formants, aiding formant-based size assessment. These findings demonstrate that pitch does not confuse body size assessment as has been previously suggested, but instead facilitates accurate size assessment by providing a carrier signal for vocal tract resonances.

Introduction

The irony of the so-called great and powerful Oz, a now infamous character starring in the 1939 classic film the *Wizard of Oz* (Fleming, 1939), is that despite his low, ominous voice, the Wizard was revealed to be a fairly small-bodied, entirely ordinary man. Indeed, a number of empirical studies have revealed a consistent bias in listeners to associate low-frequency voices with larger perceived body size both between and within sexes (Feinberg, Jones, Little, Burt, & Perrett, 2005; Pisanski & Rendall, 2011; Rendall, Vokey, & Nemeth, 2007; Smith & Patterson, 2005; van Dommelen & Moxness, 1995). As noted by Rendall et al. (2007), the conundrum here is that, not unlike our perception of the Wizard, these perceived associations are often misleading. Not all voice features are thought to provide reliable information about body size at every level of analysis (e.g., among same-sex individuals; for reviews see González, 2006; Kreiman & Sidtis, 2011; Patterson, Smith, van Dinther, & Walters, 2008).

Vocal Indicators of Body Size

Two features of the voice, formant frequencies (formants) and fundamental frequency (pitch), have traditionally been proposed to relate to body size among mammals. The source-filter theory of speech production treats the two voice features as largely anatomically and functionally independent (Fant, 1960; Lieberman & Blumstein, 1988; Titze, 1994). Formant frequencies are resonances of the supralaryngeal vocal tract associated with the percept of timbre, wherein larger individuals with longer vocal tracts typically have lower formants than do smaller individuals (Fitch, 1997, 2000a, 2000b;

Fitch & Giedd, 1999). Fundamental frequency (F_0) and corresponding harmonics (i.e., glottal-pulse rate) are related to the length and tension of the vocal folds and are perceived as voice pitch (Lieberman & Blumstein, 1988; Titze, 1994). Across primate species (Ey, Pfefferle, & Fischer, 2007; Hauser, 1993) and within anuran species (Gingras, Boeckle, Herbst, & Fitch, 2013; Ryan, 1988) larger individuals with larger larynges typically have lower voice pitch. Hence, both vocal tract resonances (i.e., formants) and pitch can independently predict size variation between or within many animal species (for reviews see Fitch & Hauser, 2003; Kreiman & Sidtis, 2011; Taylor & Reby, 2010).

Among humans, however, both formants and pitch independently predict the substantial variation in body size (e.g., height) between adults and children or between sexes (Peterson & Barney, 1952; Titze, 1989), but only formants reliably predict size within adults of the same sex. Indeed, most studies that have examined the relationship between formants and body size in humans report a significant negative relationship, even when sex and age are controlled for (Bruckert, Liénard, Lacroix, Kreutzer, & Leboucher, 2006; Evans, Neave, & Wakelin, 2006; González, 2004, 2006; Greisbach, 1999; Puts, Apicella, & Cardenas, 2012; Rendall, Kollias, Ney, & Lloyd, 2005; Sell et al., 2010; but see Collins 2000). In contrast, most studies that have examined the relationship between voice pitch and body size report no significant relationship when sex and age are controlled for (Bruckert et al., 2006; Collins, 2000; González, 2004, 2007; Hamdan et al., 2012; Hollien & Jackson, 1973; Künzel, 1989; Lass & Brown, 1978; Majewski, Hollien, & Zalewski, 1972; Sell et al., 2010; van Dommelen & Moxness, 1995; but see Collins & Missing, 2003; Evans et al., 2006; Graddol & Swann, 1983; Puts et al., 2012).

There are a number of possible, non-mutually exclusive explanations for the lack of a robust physical relationship between voice pitch and size within groups of same-sex adults in humans, relative to the more reliable relationship between formants and size (see, e.g., González, 2006; Kreiman & Sidtis, 2011; Rendall et al., 2007). Among these is the proposition that formants, unlike pitch, are closely tied to and constrained by anatomical structures related to body size and may as a consequence predict size more reliably than pitch (Fitch, 1997). Indeed, formants are related to the length and dimensions of the vocal tract that are constrained by an individual's skull and body size (Fitch & Giedd, 1999; Fitch, 2000a, 2000b). Conversely, pitch is produced by the vocal folds within the larynx that is made up of soft tissue and that develops independently of body size (Lieberman, McCarthy, Hiiemae, & Palmer, 2001). Vocal fold development and voice pitch are instead largely influenced by exposure to testosterone. At puberty, testosterone thickens and lengthens boys' vocal folds causing voice pitch to drop (Harries, Hawkins, Hacking, & Hughes, 1998; Lee, Potamianos, & Narayanan, 1999), and there continues to be a negative relationship between circulating levels of testosterone and pitch in adult men (Dabbs & Mallinger, 1999; Evans, Neave, Wakelin, & Hamilton, 2008). Most of the variation in voice pitch across individuals is therefore tied to developmental differences and to sexual dimorphism, whereas pitch and size are largely unrelated within age-sex classes.

It should be noted that although formants predict size both between and within age-sex classes, the relationship is nevertheless considerably weaker among same-sex adults. This is because there is far less variation in size among same-sex than opposite-sex adults and because vocal tract length and height are not perfectly correlated (see, e.g., Fitch & Giedd, 1999; Patterson et al., 2008).

Voice pitch and formants have been shown to have independent effects on listeners' perceptions of body size (Pisanski, Mishra, & Rendall, 2012; Pisanski & Rendall, 2011; Rendall et al., 2007) but may also interact to affect perceptions of size (Smith & Patterson, 2005). When plotted on log-log coordinates, this interaction takes the form of an ellipse, wherein pitch has a linear effect on size perception for voices whose formants represent those of adult men and women (see Patterson et al., 2008 for a detailed discussion). Voice pitch and formants have also been shown to interact in a similar manner to affect speech perception more generally (e.g., relative syllable recognition; Vestergaard, Fyson, & Patterson, 2009, 2011).

Perception of Body Size from the Voice

What is perhaps most perplexing about listeners' voice-based perceptions of body size is that they do not always map onto what we know about the physical relationships between the voice and size. Despite the lack of a robust physical relationship between pitch and size among adult men or women, listeners consistently associate low voice pitch, in addition to low formants, with perceived largeness even at the within-sex level (Feinberg et al., 2005; Pisanski, Mishra, & Rendall, 2012; Pisanski & Rendall, 2011; Smith & Patterson, 2005; van Dommelen & Moxness, 1995). The puzzling perceptual association between pitch and size among same-sex adults has been termed a *misattribution bias* (Rendall et al., 2007), presumably driven by erroneous overgeneralization of sound-size relationships (González, 2006; Rendall et al., 2007). Indeed, voice pitch has long been thought to reduce the accuracy of voice-based size assessment and to explain why listeners are generally poor at accurately estimating body size from speech (Bruckert et al., 2006; Collins, 2000; Greisbach, 1999; Rendall et al., 2007). Based on this hypothesis, voice pitch is predicted to interfere with accurate size assessment.

Although previous work has established *perceptual* relationships among voice pitch, formants, and body size, few studies have examined how the *accuracy* of listeners' size assessments might vary as a function of these two voice features. Rendall et al. (2007) found that listeners assessed men's relative body size using formants more accurately when pitch was matched between vocalizers than when the shorter male in the pair had relatively lower pitch. Although this work suggests that listeners use formants as well as voice pitch to assess size within sexes, there has been no direct test of how the accuracy of listeners' assessments is affected when pitch is present or absent from the acoustical signal[5]. Such empirical investigations will be fundamental to understanding the degree to which voice pitch or formants are honest and reliable indicators of size among humans at the within-sex level, and whether the perceptual association of low pitch with large size is erroneous for same-sex adults.

[5] Earlier work performed by Lass, Kelley, Cunningham and Sheridan (1980) examining size estimation from unvoiced speech has been widely discredited due to erroneous statistical analysis (see González, 2003).

If voice pitch confounds accurate body size assessment, as has previously been suggested, we would expect accuracy to increase when pitch cues are absent. To directly test the effect of voice pitch on size assessment accuracy within sexes, we examined listeners' accuracy among speech types where pitch cues were present versus entirely absent. In unvoiced, whispered speech, vocal folds do not produce periodic pulses; therefore, F_0 and the perception of pitch are absent but formants are present. Likewise, three-formant sine-wave speech (SWS), while intelligible (Remez, Rubin, Pisoni, & Carrell, 1981), contains only three time-varying sinusoids matching the frequency pattern of the first three formants ($F1$ - $F3$) of the original voice recording.

Experiment 1: Does Voice Pitch Confound Size Assessment Accuracy?

In Experiment 1, we used whispered speech and SWS to examine listeners' ability to assess relative body size from natural or abiological speech devoid of pitch, and compared this to listeners' accuracy from natural modal speech (regular voiced register) that contains pitch (Figure 1; listen also to Audio S1-S4). We did this to directly test whether voice pitch does, in fact, confound the accuracy of listeners' body size assessments. We additionally examined how size assessment accuracy varied as a function of the relative height between men and the relative pitch and/or formant structure of men's voices.

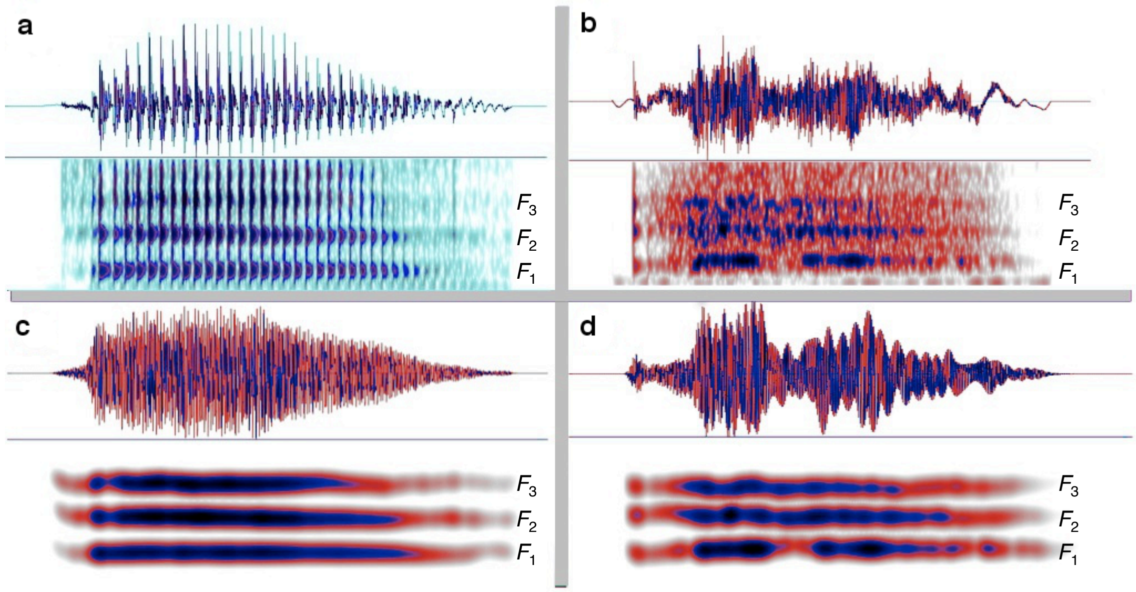

Figure 1. Four types of speech stimuili. Amplitude waveforms (top of each panel) and broadband spectrograms (bottom of each panel) illustrating the vowel /ɛ/ (spoken by a stimulus male, age 18, 187 cm tall, modal /ɛ/ F_0 = 100 Hz, $F1$ - $F3$ = 706, 1809, 2848 Hz) for: (a) modal speech; (b) whispered speech; (c) modal SWS (synthesized from *a*); and (d) whispered SWS (synthesized from *b*).

Note. In each panel, the x-axis represents time (0-0.33 s) and the y-axis represents changes in air pressure (waveform) and frequency (spectrogram; 0-4 kHz) over time. Modal speech contains voice pitch (as can be observed from the glottal-pulses present in panel *a* only) whereas whispered and SWS do not. $F1$ - $F3$ represent the first three formants. All speech stimuli used in Experiments 1-3 consisted of the five English monophthong vowels, /ɑ/, /i/, /ɛ/, /o/, and /u/. Listen also to Audio S1-S4.

Participants

Seventy-seven women (age: 18.8±1.7 years) were recruited from the psychology undergraduate research pool at McMaster University in Hamilton, Ontario, Canada to provide voice-based assessments of men's body size. All participants received partial course credit and provided informed consent. Each listener was randomly assigned to assess the relative size of one of four male groups, each containing 15 different male-male vocalizer pairs (group 1: n=20 listeners; group 2: n=19; group 3: n=20; group 4: n=18, see Materials and Table A4 in Appendix 3 for information regarding the vocalizer groups and pairs).

Materials

Male sample characteristics.

Thirty men (age: 19±2.7) were recruited from the psychology undergraduate research pool at McMaster University to provide voice recordings for use as stimuli in Experiments 1-3. We measured men's height directly using metric tape while blind to the acoustic properties (e.g., pitch and formants) of the men's voices. The average height of the men in our sample (M±SD height: 178.9±6.3 cm) compares well with that of the general population of Canadian men (where M=175 cm, Shields, Gorber, Janssen, & Tremblay, 2011). The range of heights in our sample (167-193 cm) is comparable to the ranges in past studies that have assessed relationships between the voice and size in men (e.g., Bruckert et al., 2006; Evans et al., 2006; González, 2004, 2007; Hamdam et al., 2012; Hollien & Jackson, 1973; Majewski et al., 1972; Künzel, 1989). The mean difference in height between men in stimulus pairs used to assess the accuracy of listeners' relative size assessments was 7.42±5.58 cm and ranged from 0-21 cm (see Table A4). The acoustic properties of men's voices (see *Voice measurement* and Table A5) agree well with those of previous samples of English speaking men (Bachorowski & Owren, 1999; Pisanski & Rendall, 2011; Puts, Apicella, & Cardenas, 2012).

Voice recording.

We recorded men's voices in both voiced (modal) and unvoiced (whispered) registers in an anechoic sound-controlled booth using a Sennheiser MKH 800 condenser microphone with a cardioid pick-up pattern and at an approximate distance of 5-10 cm. Speech recordings were of the five Canadian English monophthong vowels, /ɑ/ as in "father", /i/ as in "see", /ɛ/ as in "bet", /o/ as in "note", and /u/ as in "boot" (to listen, access Audio S1 & S2). Several previous studies have used this sequence of vowel sounds in isolation (Bruckert et al., 2006; Collins, 2000; Collins & Missing, 2003; Feinberg et al., 2005) or embedded in a single-syllable phrase containing one or more consonants (Pisanski et al., 2012; Pisanski & Rendall, 2011; Rendall et al., 2007) to assess listeners' voice-based perceptions of body size (see Ives, Smith, & Patterson, 2005 for discrimination thresholds in size perception from vowels and syllable phrases). Audio was digitally encoded with an M-Audio Fast Track Ultra interface at a sampling rate of 96 kHz and 32-bit amplitude quantization, and stored onto a computer as PCM WAV files using Adobe Soundbooth CS5 version 3.0.

Voice measurement.

All acoustic measurements were performed in Praat (Boersma & Weenink, 2013) and taken from the central, steady-state portion of each vowel. For modal speech, we measured mean F_0 and perceived pitch using Praat's autocorrelation algorithm with a search range set to 65-300 Hz. Perceived pitch was measured in three scales: semitones (re 1 Hz), mel, and equivalent rectangular bandwidth (ERB). The latter two scales are quasi-logarithmic. Tables A5 and A6 provide summary statistics of pitch measures.

For both modal and whispered speech, formants $F1$ - $F4$ were measured using the Burg Linear Predictive Coding (LPC) algorithm. Formants were first overlaid on a spectrogram and manually adjusted until the best visual fit of predicted onto observed formants was obtained. This method of formant measurement has been used by a number of studies examining the relationship between formants and physical or perceived body size (Evans et al., 2006; Feinberg et al., 2005; González, 2004; Greisbach, 1999; Pisanski & Rendall, 2011; Rendall et al., 2005, 2007). Although the LPC method has been criticized for potentially reporting a harmonic of the fundamental in the place of $F1$ in measurements of modal speech (Turner, Walters, Monaghan, & Patterson, 2009), we confirmed that this was not the case for our own formant measurements. We did this by normalizing formant frequencies to F_0 (formant frequency/F_0) and plotting their frequency of occurrence (see Turner et al., 2009, their Figure 6). In our voice sample, formants that were integer multiples of F_0 were not more common than were other formant values, indicating that our formant measurements showed no systematic bias or error. Mean $F1$ - $F3$ values for synthesized modal and whispered SWS were equivalent to those of the corresponding natural speech. Table A5 provides summary statistics of formant $F1$ - $F4$ measures.

In addition to $F1$ - $F4$, we computed several measures of formant structure that have previously been used to assess the relationship between formants and body size among humans and other species. For all derivations, n is the total number of formants measured (n=4) and F_i is the frequency of the ith formant in Hz. Average formant frequency, Fn (Pisanski & Rendall, 2011) is given by:

$$F_n = \frac{\sum_{i=1}^{n} F_i}{n} \qquad \text{(equation 1)}$$

Apparent vocal tract length, VTL (adapted from Fitch, 1997) is given by:

$$VTL(F_i) = \frac{\sum_{i=1}^{n} (2i-1)\left(c/4F_i\right)}{n} \qquad \text{(equation 2)}$$

where i refers to the formant number and c is the speed of sound in a uniform tube with one end closed, c=35000 cm/s. Formant position, P_f (Puts et al., 2012) is given by:

$$P_f = \frac{\sum_{i=1}^{n} F'_i}{n} \qquad \text{(equation 3)}$$

where F'_i is the standardized ith formant.

Formant dispersion, D_f (Fitch, 1997; Fitch & Giedd, 1999) is given by:

$$D_f = \frac{\sum_{i=1}^{n}(F_{i+1} - F_i)}{n-1} \qquad \text{(equation 4)}$$

Geometric mean formant frequency, MFF (Irino, Aoki, Kawahara, & Patterson, 2012; Ives et al., 2005; Smith & Patterson, 2005) is given by:

$$MFF = \left(\prod_{i=1}^{n} F_i\right)^{1/n} \qquad \text{(equation 5)}$$

Like previous work (Jovičić, 1998), we found that the absolute values of $F1$ - $F4$ were in some cases higher for whispered than for modal vowels (see Table A5). Critically, however, whispering did not significantly affect the *relative* difference in formants between paired vocalizers and thus was unlikely to affect listener's *relative* assessments of size. In both modal and whispered speech, the taller male in the pair always had lower $F1$ - $F4$ values relative to the shorter male, and relative differences in $F1$, $F2$, and $F4$ were not statistically different between modal and whispered speech (see Table A7).

Editing and creation of speech type stimuli.

Copies of each original voice recording (30 modal, 30 whispered) were edited using Praat (Boersma & Weenink, 2013). Vowels were flanked by 250 ms of silence resulting in modal and whispered voice stimuli that were 3.46±0.41 and 3.59±0.37

seconds in duration, respectively. A copy of each modal and whispered voice stimulus was additionally re-synthesized into three-formant sine-wave speech (SWS) in Praat (Boersma & Weenink, 2013; see Figure 1). This type of minimal speech synthesis involves creating three time-varying sinusoids that match the frequency pattern of the first three formant frequencies ($F1$ - $F3$) of the original voice recording (Remez et al., 1981). Synthesized SWS is devoid of the majority of the acoustic information present in natural and voiced human speech, including F_0 and its harmonics. Nevertheless, the formant information embedded in the sinusoidal pattern is sufficient to elicit the percept of intelligible speech (i.e., vowel sounds, Remez et al., 1981; to listen, access Audio S3 & S4). All stimuli were amplitude normalized to 70 dB RMS SPL and played back at constant amplitude within participants.

Pairing of speech stimuli.

To pair speech stimuli, the 30 male vocalizers were first pseudo-randomly paired four separate times (groups 1 - 4), ensuring only that each pairing occurred no more than once among all four groups, resulting in 60 unique male-male pairs (15 pairs per group). Table A4 summarizes the means and ranges of height differences between men. Height differences did not differ significantly across the four groups (one-way ANOVA: $F_{3,56}$=0.067, p=.997). Speech stimuli were then paired within each speech type (e.g., modal-modal), resulting in 60 stimulus pairs per speech type and 240 in total.

Procedure

All experiments were approved by the McMaster Research Ethics Board and comply with the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct. The experiment was conducted in the Voice Research Laboratory at McMaster University. Each listener was randomly assigned to assess the relative size of one of four male groups, for all four speech types, resulting in a total of 60 size assessment trials per participant. Voices were presented to participants in a private room via a custom computer interface and through Sennheiser HD-280 PRO headphones.

On each trial, listeners were presented with two men's voices of the same speech type (modal, whispered, modal SWS, or whispered SWS). Voices were played consecutively, prompted by the participant selecting the 'play' button for the individual file. After listening to each voice in the pair, participants were asked to select which of the two voices belonged to the taller man by selecting the corresponding button on the screen. Participant responses automatically loaded the next trial. Trials were blocked by speech type. The presentation order of blocks, of paired voice stimuli within each block, and of voice stimuli within each pair was fully randomized. The order of vowels in each voice stimulus was always /ɑ/, /i/, /ɛ/, /o/, and /u/. Following the experiment, participants provided their age.

Results and Discussion

We coded a response as correct ('1') if the speech stimulus chosen was that of the taller man's in the pair and otherwise as incorrect ('0'). Trials on which there was no difference in height between men in a pair (n=3 pairs) or in which the difference in height was negligible (i.e., ≤0.5 cm, n=3 pairs) were not included in analyses; however, their removal did not affect the general pattern or statistical significance of our results. The effect of speech type on listeners' accuracy scores did not vary as a function of the male group that they were assigned to assess (no significant group-by-speech-type interactions in repeated measures analysis of variance [rmANOVA]: all F<1.4, all p>.249). Hence, for statistical analyses, size assessment accuracy scores were averaged across listeners from groups 1 - 4 for each speech type. We confirmed that the data were normally distributed for all speech types (Shapiro-Wilk, df = 77, modal: W=0.969, p=.06; whispered: W=0.971, p=0.074; modal SWS: W=0.989, p=.724; whispered SWS: W=0.973, p=.102).

The effects of speech type and men's relative height on size assessment accuracy.

Listeners performed above chance (>0.5 proportion correct, two-tailed one-sample t-tests) in assessments of size from modal speech (t_{76}=7.69, p<.001), whispered speech (t_{76}=2.66, p=.01), modal SWS (t_{76}=2.05, p=.044), and whispered SWS (t_{76}=3.24, p=.002; see Figure 2a). After controlling for multiple comparisons, size assessment accuracy remained significantly above chance for modal speech, whispered speech, and whispered SWS (Bonferroni correction, α/n = .0125, two-tailed). Thus, listeners extracted reliable size information from speech whether or not voice pitch was present and whether voice stimuli were natural or synthesized.

If it were true that voice pitch confounds accurate size assessment, listeners' accuracy would be expected to be higher for synthesized SWS and natural whispered speech, both of which are devoid of pitch cues, relative to modal speech that contains natural pitch. In fact, accuracy in listeners' size assessments was higher for natural than for synthesized (SWS) stimuli (repeated measures analysis of variance, rmANOVA, $F_{1,73}$=6.38, p=.014), and higher for modal than for whispered speech among natural but not among synthesized voices ($F_{1,73}$=6.63, p=.012). Planned paired-sample *t*-tests (two-tailed) revealed that listeners' accuracy was significantly better for modal speech than for all other speech types (modal vs. whispered: t_{76}=2.89, p=.005; modal vs. modal SWS: t_{76}=3.7, p<.001; modal vs. whispered SWS: t_{76}=3.07, p=.003), whereas accuracy did not differ among the other three speech types (all $|t|$<0.979, all p>.33; Figure 2a). Thus, listeners were *more* rather than less accurate in size assessments when pitch cues were present, despite the absence of a reliable physical relationship between pitch and height in this sample of men (see Table A6).

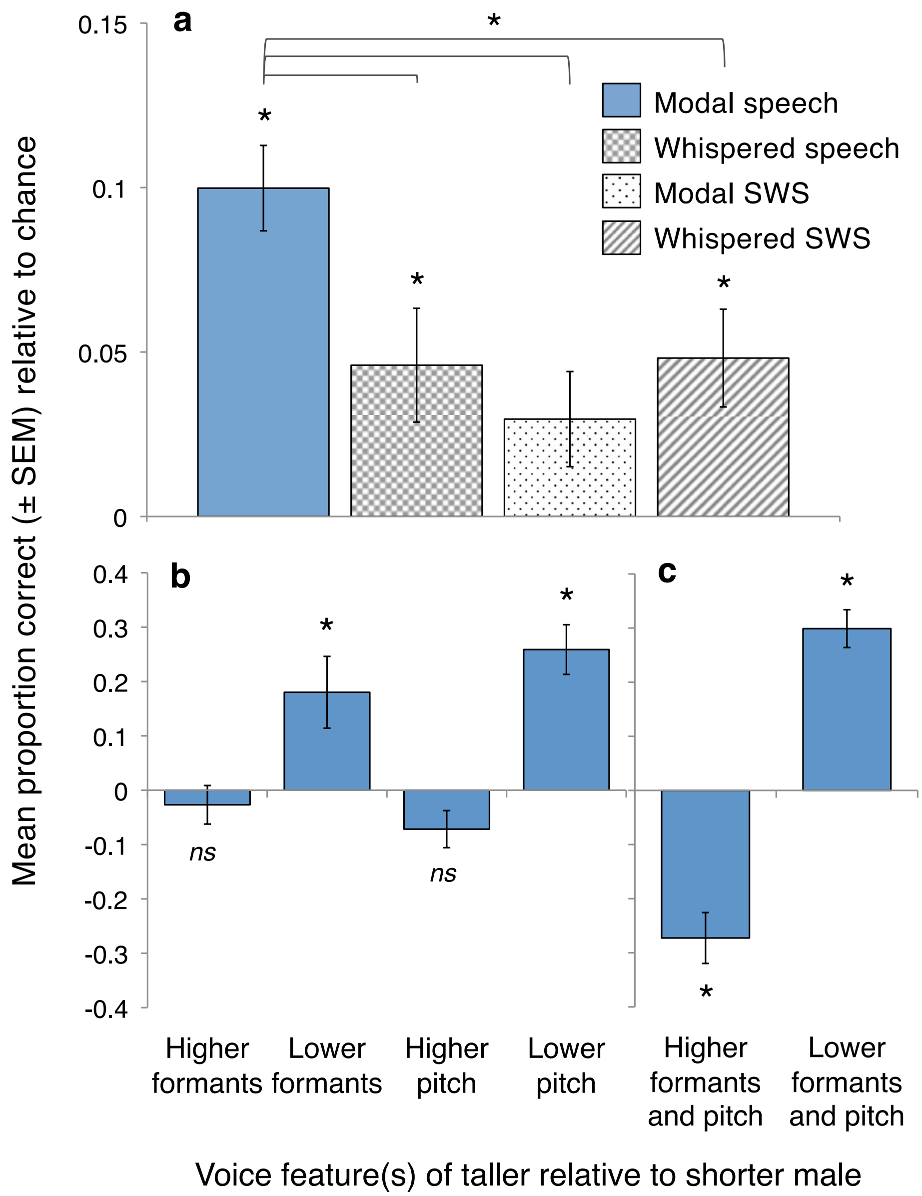

Figure 2. Accuracy in listeners' body size assessments (Experiment 1). Mean proportion correct size assessment (± standard error mean) as a function of (a) speech type and (b-c) the formants and/or voice pitch of the taller male relative to the shorter male in a vocalizer pair for the modal speech condition.

Note. Mean proportion correct refers to the average proportion of trials in which the taller male vocalizer was correctly identified relative to chance. Thus, the y-axis represents the difference between participant's mean accuracy and chance accuracy, where values above 0 indicate above-chance performance, and values below 0 indicate below-chance performance. (a) * $p<.0125$ Bonferroni correction, one-sample t-tests comparing accuracy for each speech type against chance*,* and paired-sample t-tests comparing accuracy among speech types; all tests two-tailed, n=77 listeners). (b-c) * $p<.05$; *ns* $p>.05$, two-tailed one-sample t-tests; see Table 1 for additional details).

Regression analyses showed that accuracy generally improved as the difference in height between men increased (see Figures 3 and 4). Linear regression indicated that size assessment accuracy increased with the difference in height between male vocalizers for modal speech ($F_{1,51}$=4.486, p=.039, β=0.284, R^2=0.081), whispered speech ($F_{1,50}$= 5.714, p=.021, β=0.320, R^2=0.10), and modal SWS conditions ($F_{1,51}$=6.323, p=.015, β=0.332, R^2=0.11). The linear regression slope for whispered SWS was not significant ($F_{1,50}$=1.471, p=.231, β =- 0.169, R^2=0.03; Figure 3). The results of our linear regression for modal speech are similar to those reported by Rendall et al. (2007). If voice pitch confounds size assessment, we might expect that the slope of the linear regression of accuracy on relative height (see Figure 3) would be steeper for modal speech than for whispered speech or SWS. This is because, in order to counteract the apparently erroneous cues to size provided by pitch, listeners might require greater differences in height (and indeed, in formants) between men to assess body size accurately when pitch cues are present than when pitch cues are absent. However, we did not find any evidence of this. In fact, linear regression slopes were comparable among modal (β=.284), whispered (β=.320), and modal SWS (β=.332) conditions.

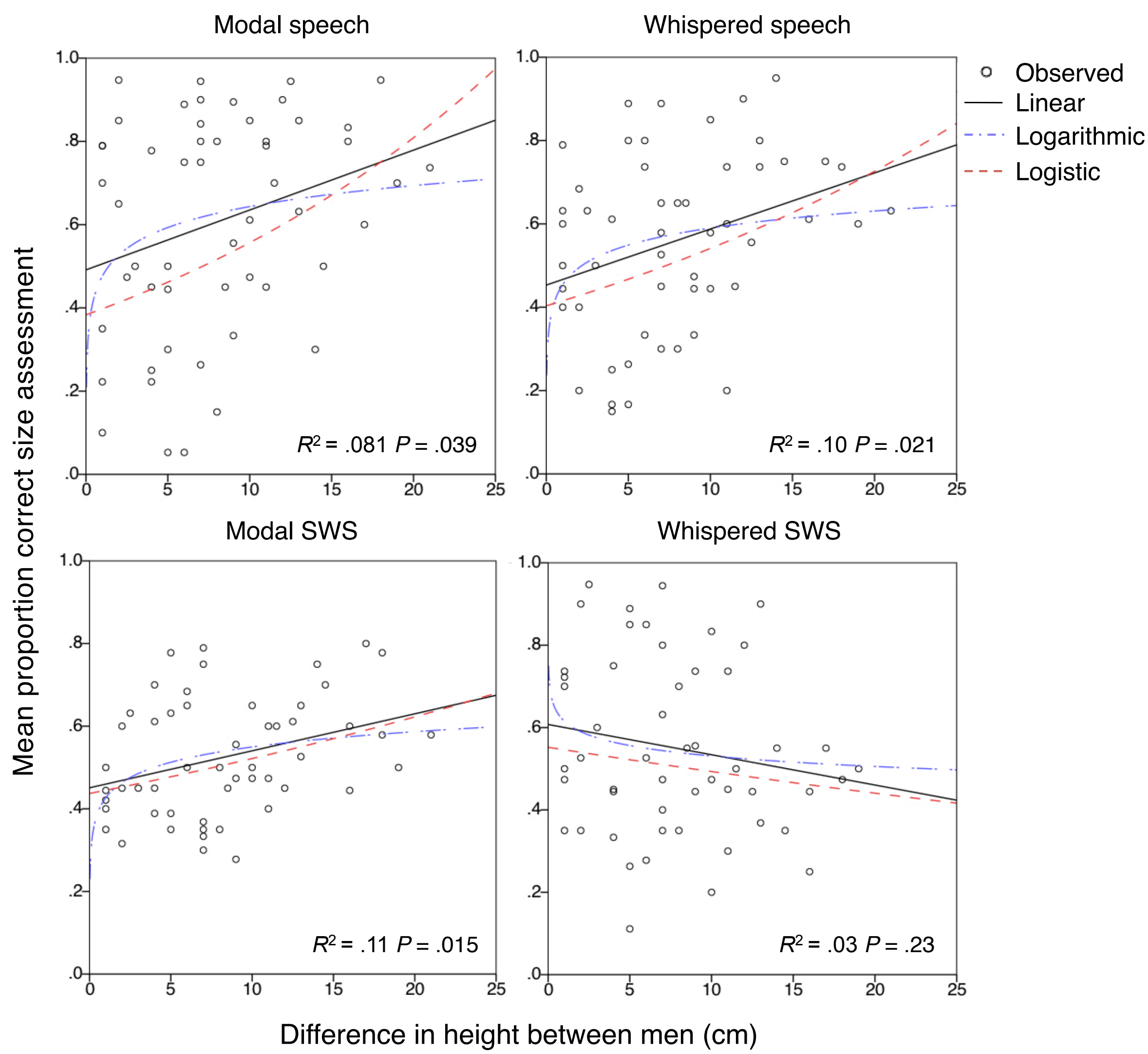

Figure 3. Accuracy in listeners' body size assessments (Experiment 1) as a function of men's relative height (taller - shorter male) for each speech type.

Note. Each data point represents the mean proportion correct size assessment of all listeners for a given vocalizer pair. A total of six high-leverage outliers (1-2 pairs per regression, resulting in n=52-54 pairs) were identified (Cook's $D > 4/n$ or .075) and removed from each respective analysis, but this did not affect the direction of the regressions. Size assessment accuracy increased with the difference in height between men with modal speech, whispered speech, and modal SWS conditions (best fitting model: linear regression, $p<.05$). The slope for whispered SWS was not statistically significant (p=.23).

We additionally fitted an inverse cumulative distribution function to our data using the probit model. The probit model is preferred for analyzing a binary response variable obtained from a two-alternative forced-choice task because it produces estimated probabilities of likelihood that are constrained between 0 and 1, and allows for the effect of the independent variable to vary across different values of the dependent variable (Long, 1997). Akin to the results of the linear regressions, the probit model showed that the probability of correct size assessment increased with the relative difference in height between male vocalizers for modal speech (estimated increase in accuracy Z-score for every cm difference in height ± SE = 0.20±0.007, p=.005), whispered speech (0.25±0.007, p=.001), and modal SWS (0.016±0.007, p=.028), but not for whispered SWS (0.003±0.008, p=.73; Figure 4).

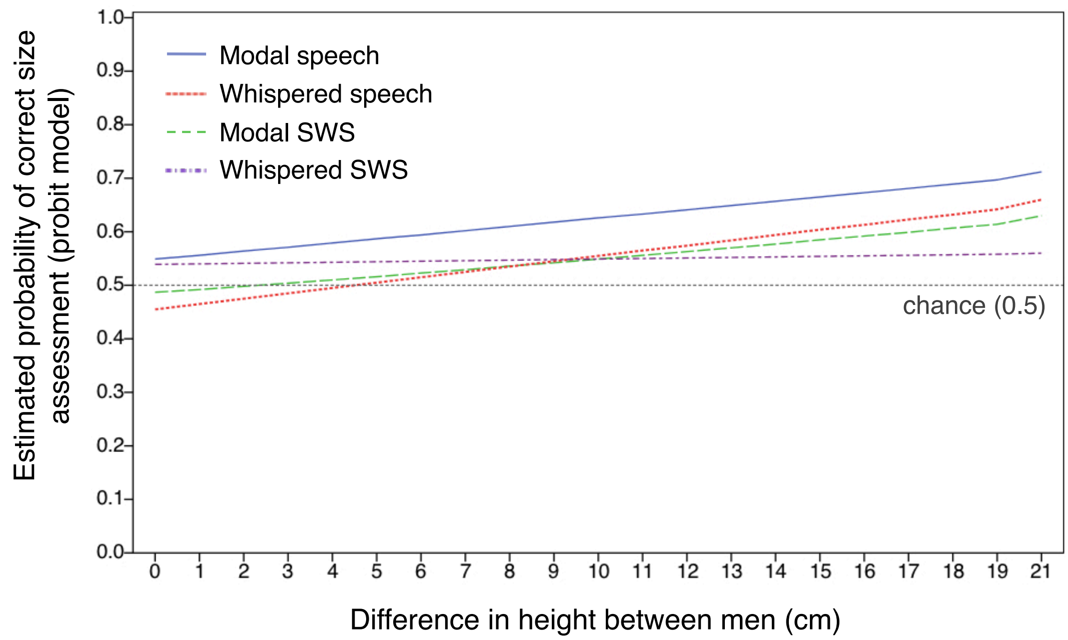

Figure 4. Probit inverse cumulative distribution functions fitted to listeners' size assessment accuracy scores (Experiment 1) as a function of men's relative height (taller - shorter male) for each speech type. Estimated probability of correct size assessment increased with the difference in height between men with modal speech, whispered speech, and modal SWS conditions (p<.05), but not with whispered SWS (p=.73).

Noting the linear shape of these psychometric functions (see Figure 4), we can see that the effect of men's relative height on listeners' accuracy was effectively constant across height differences. Additionally, listeners performed around chance levels when the differences in height between men were close to 0, and improved as the difference in height increased, but listeners did not approach the upper limit of optimal performance (i.e., 0.95-1 proportion correct) and thus the functions do not asymptote near 1. For this reason, the linear regression models provided a better fit to our data than did the probit models (Pearson goodness-of-fit tests for probit: modal: χ^2=71.02, p<.001; whispered: χ^2=58.94, p<.001; modal SWS: χ^2=22.35, p=.27; whispered SWS: χ^2=30.2, p=.036).

The effects of men's relative formants and pitch on size assessment accuracy.

We examined the degree to which formants or pitch predicted actual relative height between men using multiple derivations of either voice feature, including: average formant frequency (Fn, as given by equation 1); apparent vocal tract length (VTL, equation 2); formant position (P_f, equation 3); formant dispersion (D_f, equation 4); geometric mean formant frequency (MFF, equation 5); fundamental frequency (F_0); and perceived voice pitch in semitone, mel, and equivalent rectangular bandwidth (ERB) scales (see *Voice measurement* for additional details). Formant measures were significantly correlated with one another (two-tailed bivariate regressions: $|r|$ modal = .56–.999, all p<.01, Table A8), as were voice pitch measures ($|r|$ = .993–.999, all p<.01, Table A9). Regardless of the measure used, formants predicted relative height better than did voice pitch (Table A6).

Nevertheless, listener's size assessment accuracy with modal speech was predicted both by differences in men's formants (ANOVA, Fn: $F_{1,53}$=24.95, p<.001) and differences in men's voice pitch (F_0: $F_{1,53}$=54.42, p<.001), together accounting for 62% (Adjusted R^2) of the variance in accuracy ($Fn + F_0$: $F_{1,53}$=41.64, p<.001). Relatively lower formants or lower pitch in the taller vocalizer independently facilitated accuracy in size assessment, resulting in above-chance performance, whereas higher formants or higher pitch did not, resulting in chance performance. The facilitating effects of lower formants or pitch were significantly greater than the null effects of higher formants or pitch (independent two-tailed *t*-tests, *Fn*: t_{52}=2.76, *p*=.01; F_0: t_{52}=5.85, *p*<.001). In other words, both voice features independently aided accurate size assessment, and neither independently confounded accurate size assessment. Notably, lower pitch in the taller vocalizer resulted in a mean accuracy of at least 70% across all trials (Table 1).

Table 1. Mean proportion correct size assessment from modal speech as a function of the relative voice features of the vocalizer pair

Voice Feature(s) of Taller Relative to Shorter Male	Proportion Correct[b] (M±SD)	*df*[c]	*t*[c]	*p*[c]
Independent Effects[a]				
Higher Formants	0.47±0.30	20	-0.405	.690
Lower Formants	0.68±0.20	32	5.078	<.001
Higher Pitch	0.43±0.23	25	-1.567	.130
Lower Pitch	0.76±0.18	27	7.59	<.001
Combined Effects				
Higher Formants and Higher Pitch	0.23±0.15	9	-5.79	<.001
Higher Formants and Lower Pitch	0.70±0.22	10	2.985	.014
Lower Formants and Higher Pitch	0.55±0.18	15	1.179	.257
Lower Formants and Lower Pitch	0.80±0.14	16	8.56	<.001

a. Effect of one voice feature while controlling for the other. Formant measure = *Fn*, mean formant frequency (see equation 1); Pitch measure = F_0, mean fundamental frequency.

[b.] Proportion correct size assessment (means ± standard deviations) based on a total of 54 vocalizer pairs whose difference in height exceeded 0.5 cm.

[c.] Two-tailed one-sample *t*-tests against chance (0.50).

The combined effects of formants and pitch on accuracy were cumulative (Table 1; Figure 2c). When the taller vocalizer had *both* lower formants and lower voice pitch, accuracy reached its highest level (80%). Likewise, only when the taller vocalizer had *both* higher formants and higher voice pitch did accuracy fall significantly below chance (23%). This pattern of results suggests that listeners may have shifted their criterion for relative size assessment, particularly on trials in which the frequency differences in the pitch and formants of vocalizers' voices were congruent. Thus, there appears to have been a consistent response bias toward correctly choosing the taller male when his pitch and formants were both lower relative to the shorter males, but a consistent bias toward incorrectly choosing the shorter male when his pitch and formants were both relatively higher. This possible response bias is not as apparent on trials in which the frequency differences in the pitch and formants of vocalizers were incongruent. This supports the previous conclusion that the effects of voice pitch and formants on size perception were cumulative.

Experiment 2: Does the Absence of Voice Pitch Hinder Formant-Based Size Perception?

If voice pitch must be present in the acoustical signal for listeners to be able to extract body size information from the formant frequencies of the voice, then removing pitch from the voice might in fact impair size assessment. If this were true, it may explain the results of Experiment 1, wherein listeners were less accurate in size assessments when pitch cues were absent than when pitch cues were present. We tested this possibility in Experiment 2 by examining how the presence or absence of voice pitch affected listeners' perceptions of size from modal and whispered voices with manipulated formants.

Modal speech with lowered compared to raised formants is typically associated with larger perceived size when controlling for voice pitch (Feinberg et al., 2005; Pisanski et al., 2012; Smith & Patterson, 2005). However, it is not known whether lowered formants in natural whispered speech, where pitch cues are entirely absent, will elicit analogous perceptions of size among same-sex adults. A recent psychoacoustic study by Irino et al. (2012) has shown that the just-noticeable difference in size perception from formants is the same (~5%) for synthesized voiced speech and synthesized whispered speech, suggesting that formant shifts in natural voices would also be equally perceivable regardless whether pitch is present or absent.

Participants

For Experiment 2, a new group of 40 women (age: 19.38±2.62) and 18 men (age: 21.17±5.15) was recruited from Conestoga College's nursing undergraduate research pool. All participants received partial course credit and provided informed consent.

Materials

The voices used in Experiment 2 (5 modal and 5 whispered from the same 5 males) were randomly drawn from the pool of speech stimuli used for Experiment 1. The formant component of men's modal and whispered speech was raised or lowered by 10% from baseline using Praat, holding F_0 and harmonics constant in the case of modal speech (Boersma & Weenink, 2013; Feinberg et al., 2005). These manipulations were performed using resampling override and Pitch-Synchronous Overlap Add (PSOLA) algorithms to return pitch to its original value (now a standard feature in Praat; Boersma & Weenink, 2013). The magnitude of these manipulations corresponded to approximately two times the just-noticeable difference in formant perception from vowels (Pisanski & Rendall, 2011), similar to that used in previous work examining the effects of manipulated formants on size perception (Pisanski & Rendall, 2011), and was representative of a large portion of the natural variation in formants among men (Peterson & Barney, 1952; Pisanski & Rendall, 2011; Lee et al., 1999). We paired raised-formant with lowered-formant speech stimuli within vocalizers and within each speech type resulting in a total of 10 voice pairs (5 modal-modal and 5 whispered-whispered). Thus, both voice stimuli within a pair originated from the same man, such that the only difference between the stimuli was in their formants (raised versus lowered).

Procedure

The experiment was completed online. Previous research using an analogous procedure has shown that listeners' voice-based assessments of men are the same whether collected online or in the laboratory (Feinberg et al., 2011). Before beginning the experiment, all participants consented to wearing headphones for the duration of the experiment. Each participant then completed a total of 10 trials. On each trial, participants were presented with a single pair of voices (raised-formant vs. lowered-formant) matched for speech type. Akin to Experiment 1, voices were played consecutively, prompted by the participant selecting the 'play' button for the individual file. After listening to each voice in the pair, participants were asked to select which of the two voices belonged to the taller man by selecting the corresponding button on the screen. Participant responses automatically loaded the next trial. The presentation of voice pairs was blocked by speech type. The order of speech type was counter-balanced between participants (whispered followed by modal, or, modal followed by whispered) and the presentation order of voice stimuli within each voice pair was fully randomized (raised-formant voice played first, or, lowered-formant voice played first). The order of vowels in each voice stimulus was always /ɑ/, /i/, /ɛ/, /o/, and /u/. Following the experiment, participants provided their age and sex.

Results and Discussion

Responses were coded as '1' if the voice chosen was that with lowered formants and otherwise as '0'. We then calculated the proportion of trials on which listeners associated relatively lower formants with larger size by averaging responses across trials and participants within each condition. Because the data were heavily skewed, and not normally distributed for either speech type (Shapiro-Wilk, $df = 58$, modal: W=0.811, *p*<.001, skewness = -0.94; whispered: *W*=0.862, *p*<.001, skewness = -0.73), we used nonparametric statistical tests to analyze listeners' responses (all two-tailed).

For both modal (*M*±SD =75±29%) and whispered speech (71±29%), listeners associated relatively lower formants with larger body size on approximately three quarters of all trials, and significantly above chance (one-sample binomial tests vs. 0.5, *n*=58, *p*<.001). We found no significant effect of speech type (Wilcoxon Signed Rank, *n*=58: *Z*=-1.36, *p*=.17) or of listener sex (Mann-Whitney, *n*=58; modal: *U*=286, *p*=.19; whispered: *U*=285, *p*=.20) on formant-based perceptions of size. Consistent with the psychoacoustic work of Irino et al. (2012), the absence of voice pitch did not hinder listeners' ability to extract size information from the formant frequencies of the voice. Thus, the absence of pitch cannot explain the results of Experiment 1.

Experiment 3: Does Harmonic Density of Voice Pitch Facilitate Formant-Based Size Perception?

The absence of voice pitch did not hinder formant-based size perception (Experiment 2). However, the presence of voice pitch, and in particular, the density of harmonics sampling the formant envelope, may facilitate formant perception and thereby increase the accuracy of listeners' size assessments. More densely spaced harmonics in a low-pitched voice (see spectrogram in Figure 5 for illustration) have been shown to enhance the salience of corresponding formant frequencies and to aid in vowel perception (Assmann & Nearey, 2008; Ryalls & Lieberman, 1982) and in the perception of size information from synthetic tones (Charlton, Taylor, & Reby, 2013). If this is also true for formant-based size perception from natural human voices, it may explain why listeners in Experiment 1 performed better from speech containing pitch, and in particular, better when the taller male's voice pitch was relatively lower than higher (Table 1; Figure 2b and c). We tested the Harmonic Density Hypothesis in a third experiment. We predicted that, if denser harmonics enhance formant detection in natural speech and improve the accuracy of size perception, accuracy would be relatively higher in the lowered-pitch condition.

Participants

For Experiment 3, a new group of 120 women was recruited from the psychology undergraduate research pool at McMaster University. All participants received partial course credit and provided informed consent. Each participant was randomly assigned to a raised-pitch (n=60, age: 20±2.6) or lowered-pitch condition (n=60, age: 19±2.19).

Materials

The 60 modal speech stimulus pairs used in Experiment 3 were identical to those used in Experiment 1, except that for Experiment 3, the F_0 of the modal stimulus pairs was manipulated using Praat's PSOLA algorithm (Boersma & Weenink, 2013) holding formants constant. The voice F_0 of all speech stimuli was either raised or lowered by adding or subtracting 0.5 ERBs of the baseline F_0 for speech stimuli used in the raised-pitch and lowered-pitch conditions, respectively (note that both vocalizers within each pair received the same pitch manipulation). The ERB scale controls for discrepancies between F_0 and perceived pitch, where one ERB is roughly equivalent to a 20 Hz absolute

F_0 manipulation of a voice with a mean F_0 of 120 Hz, or to roughly 3 semitones. Thus, our manipulations resulted in a mean F_0 difference of about 40 Hz between the raised-pitch and lowered-pitch groups of male vocalizer pairs (for example, see spectrogram in Figure 5). This difference is greater than the just-noticeable difference in voice pitch perception (Pisanski & Rendall, 2011). Critically, the magnitude of our F_0 manipulation (a 40 Hz difference between conditions) was exactly analogous to the degree of natural variation in pitch among men in Experiment 1, where men's natural voice pitch ranged from 90.4 – 129.8 Hz.

Procedure

The experimental procedure for Experiment 3 was identical to Experiment 1 except that each participant assessed the relative size of all 60 male pairs (pairs from groups 1- 4 inclusive), in either the raised-pitch or lowered-pitch condition. Once again, listeners were presented with two men's voices and asked to select which of the two voices on each trial belonged to the taller man.

Results and Discussion

We confirmed that the data were normally distributed for both conditions (Shapiro-Wilk, $df = 30$, raised-pitch: W=0.959, p=.30; lowered-pitch: W=0.967, p=.45). Figure 5 illustrates that listeners performed significantly better in the lowered-pitch condition where harmonics were denser (M±SD = 60.25±6.59% correct) than in the raised-pitch condition where harmonics were sparser (57.72±5.93%; one-way ANOVA,

$F_{1,\,118}$=4.893, *p*=.029). Thus, as predicted, we found that harmonic density facilitated accurate voice-based size perception.

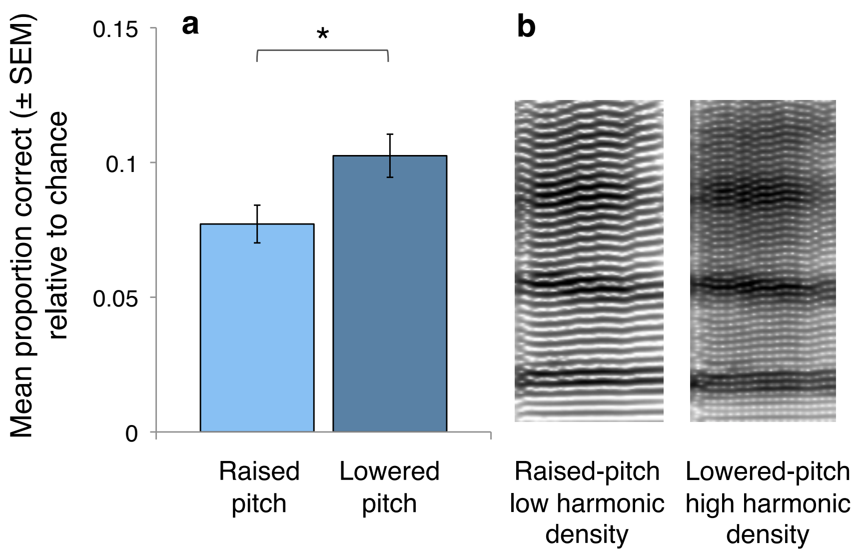

Figure 5. Harmonic Density Hypothesis. (a) Mean proportion correct size assessment was significantly higher for the lowered-pitch condition (high harmonic density) than for the raised-pitch condition (low harmonic density; **p*=.029, two-tailed one-way ANOVA, *n*=120). (b) Narrowband spectrograms depicting a stimulus male's voice (vowel /ɛ/) with raised-pitch (F_0=122 Hz) and lowered-pitch (F_0=83 Hz).

Note. Mean proportion correct (± standard error mean) refers to the average proportion of trials in which the taller male vocalizer was correctly identified relative to chance, where values above 0 indicate above-chance performance.

However, its effect size (Cohen's *d*=0.41) was one-quarter the strength of what is to be expected by the asymmetrical gains-to-losses in accuracy reported in the modal condition of Experiment 1 (where Cohen's *d*=1.62). That is, the gains in accuracy observed in Experiment 1 on trials in which the taller vocalizer had relatively lower voice pitch were considerably greater than any losses in accuracy that resulted from his having relatively higher voice pitch. Moreover, despite a comparable pitch range in the two Experiments (40 Hz), the ratio of gains-to-losses in accuracy between lower and higher voice pitch observed in Experiment 1 was, on average, on the order of 25% (Table 1). This is much greater than the ratio of gains-to-losses observed in Experiment 3. Hence, while harmonic density certainly plays a role, it cannot fully explain the facilitating role of voice pitch in body size assessment.

General Discussion

The results of these experiments provides evidence that human listeners are more accurate in voice-based assessments of men's relative body size when pitch cues are present than when pitch cues are absent. This is true despite the lack of a robust, direct relationship between voice pitch and men's body size in our sample of men. This finding, in addition to the finding that size assessment accuracy increases with the harmonic density of human speech, provides support for an indirect, facilitating role of voice pitch in body size assessment.

We took several measures to ensure that our results were not due to distorted, degraded, or noisy whispered speech. First, we confirmed that modal and whispered formants were significantly correlated and that the formant differences between men (relative $F1$ - $F4$) were largely the same for modal and whispered speech. Second, we presented all stimuli at the same normalized amplitude to each participant. Pilot testing indicated that listeners did not perceive any differences in the loudness of modal and whispered speech. Moreover, if perceived loudness had been lower for whispered than for all other speech types, accuracy for whispered speech may have been lower than for SWS, but this was not the case. Third, listeners proved capable of assessing relative size from whispered speech above chance, confirming that size information was preserved in unvoiced speech, and performed no worse than from modal SWS where sound was periodic and without aperiodic noise in the signal.

The results of Experiment 2 further demonstrate that the absence of voice pitch does not in itself reduce listeners' ability to extract size information from formants, and that this cannot explain our findings. Whispering affects accuracy but not the perceptual association between low frequencies and perceived height. Indeed, recent work has found that the just-noticeable difference for formant-based size perception is the same (~5%) for synthesized modal and synthesized whispered speech (Irino et al., 2012). Whereas Irino and colleagues have shown that synthetic whispered speech supports size assessment over a wide range of formants representing men, women, and children, the results of Experiments 1 and 2 of the current study show that relative size can be gleaned from natural whispered speech even among same-sex adults, where the differences in formants among individuals are considerably smaller.

The results of Experiment 3 indicate that accurate size assessment is, at least in part, tied to the density of the spectral sample. The denser sampling provided by lower voice pitch appears to increase the salience of corresponding formant frequencies, aiding listeners in extracting reliable formant-based (i.e., vocal tract or size) information from the voice. This finding is in line with earlier work that has shown that formant-based vowel perception is more accurate with natural voices with a lower than higher pitch (150 Hz range in F_0, Ryalls & Lieberman, 1982), and that listeners are more likely to associate downward shifts in formant spacing with larger perceived body size from synthesized tones of a lower than higher pitch (310 Hz range in F_0, Charlton et al., 2013). It is important to highlight that our work provides the first evidence that low voice pitch not only improves formant perception, but also results in more accurate within-sex size assessment from human speech with natural formants. Moreover, we show that this is true even when the differences in voice pitch represent the natural degree of variation found among same-sex adults. The current study contributes to a growing body of literature that has found that voice pitch and formants interact in complex ways to affect voice perception (Feinberg et al., 2011; Patterson et al., 2008; Smith & Patterson, 2005; Vestergaard et al., 2009, 2011).

Although the relationships among formants, vocal tract length, and height are relatively weak within sexes compared to between sexes (Fitch & Giedd, 1999; González, 2004, 2007; Patterson et al., 2008; Rendall et al., 2005), the current set of experiments and those by Rendall et al. (2007) show that listeners can nevertheless assess the relative size of same-sex adults from natural voiced speech by attending to differences in formants. It is not known, however, whether listeners preferentially attend to variation in certain formants more than in others. On one hand, the relative positions of $F1$ and $F2$ shift constantly in continuous speech within individuals, facilitating vowel perception, whereas $F3$ and $F4$ remain more stable and may as a consequence provide more reliable information about vocal tract length (Greisbach, 1999). On the other hand, neuroimaging studies have shown that normalization processes that compensate for individual differences in vocal tract length during vowel perception occur early in the processing of speech sounds and mainly involve the lower formants, $F1$ - $F3$ (Monahan & Idsardi, 2010; Sjerps, Mitterer & McQueen, 2011). Thus, $F4$ is relatively inconsequential for size normalization. Finally, Fant (1960, p. 121) noted variation in the relation of different formants to different physical dimensions of the oral and nasal cavities and supralaryngeal vocal tract. Variation in anatomical constraints on different formants might additionally affect their relative reliability as cues to size. Taken together, it is possible that certain formants may indicate size more reliably than others, but it is unclear which formants. It is important to note that listeners in Experiment 1 estimated relative size from three-formant SWS containing only three sine waves corresponding to formants $F1$ - $F3$. Listeners' accuracy from SWS was above chance and no different than from whispered speech, which contained higher formants, indicating that the lower formants are sufficient for size assessment. Nevertheless, because $F4$ may contain additional size information, future studies should investigate whether size assessment accuracy is relatively higher for four-formant SWS than for three-formant SWS.

Because we were ultimately interested in the processes and mechanisms that individuals use to assess relative body size, particularly of same-sex individuals and in everyday life, we designed our Experiments to reflect the natural difficulty of size discrimination at this level. The differences in men's formants and voice pitch reflected the natural degree of variation in the general population of men. Likewise, the range of heights in our sample of male vocalizers (167-193 cm) was analogous to the ranges reported in numerous other studies that have examined voice-based estimation of men's body size from natural speech stimuli (Bruckert et al., 2006; Collins, 2000; Rendall et al., 2007) or the relationship between formants and physical height (Bruckert et al., 2006; Collins, 2000; Evans et al., 2006; González, 2004, 2007; Graddol & Swann, 1983; Hamdan et al., 2012; Hollien & Jackson, 1973; Künzel, 1989; Majewski et al., 1972; van Dommelen & Moxness, 1995).

While the design of our study was intended to increase the ecological validity and generalizability of our results, a potential drawback of the design is that, for some male voice pairs, the differences in formants or height may have been too small for listeners to perceive. Psychoacoustic studies have shown that the just-noticeable difference for formant-based size discrimination from isolated vowel sounds is approximately 7-8%, slightly higher than the just-noticeable difference for size discrimination from vowels paired with consonants in single-syllable phrases (4-6%) or vowels embedded in words (5%; Irino et al., 2012; Ives et al., 2005; Smith & Patterson, 2005). These studies used synthesized voices that were scaled to represent uniform differences in formants across a wide range of apparent vocal tract lengths, many representing body sizes beyond the natural range of the general population (i.e., very short children or very tall adults). Nevertheless, the just-noticeable difference in formant perception reported for natural human speech (5-6% for vowels in words; Pisanski & Rendall, 2011) is consistent with the thresholds established using synthetic speech-like sounds. Also consistent with these thresholds are the results of Rendall et al. (2007), who reported that listeners were unable to accurately assess the relative height of men from natural voiced speech when differences in height between the two men were less than 10 cm.

The present study provides the foundation for future work to investigate whether harmonic density improves accuracy of size estimation between age-sex classes (e.g., between men and women or children and adults) in which the relative formants and heights of paired vocalizers are likely to consistently exceed perceptual discrimination thresholds. The effect of harmonic density on size perception may be greater in populations exhibiting a wider range of pitch and formants. Interestingly, psychoacoustic studies show that in the range of vocal tract lengths representing small children, the influence of voice pitch decreases with vocal tract length more rapidly with low-pitched than with high-pitched synthetic voices (discussed in Patterson et al., 2008). Future work should also examine whether the role of harmonic density in size perception is greater for syllable phrases, words, and longer stretches of speech than it is for sequences of isolated vowel sounds.

We argue that it is unlikely that humans are better at assessing size from the voice when both pitch and formants are present simply because this is typical in everyday life or because listeners have more experience with modal speech than with SWS or whispered speech. First, the results of Experiment 2 demonstrate that formant-based size *perception* is similar for modal and whispered speech. Listeners do not appear to have any difficulty associating formant manipulations with size in whispered speech despite their relative lack of familiarity with this type of speech, but rather show deficits only in their ability to assess size accurately from whispers. This is further evidenced by an equivalent just-noticeable difference in size perception between modal and whispered speech (Irino et al., 2012; although the just-noticeable difference in size perception for SWS may be analogous to that for modal and whispered speech, there is currently no research to support this). Second, expertise and familiarity cannot explain why only relatively *lower* voice pitch facilitated accuracy from modal speech, especially if taller and shorter individuals are equally likely to have relatively higher voice pitch as they were in our sample. Listeners are not likely to have learned from experience to associate low pitch with large size within sexes because they would not have experienced this association any more frequently than they would have experienced the association between low pitch and small size or high pitch and large size. Finally, simply using a strategy involving consistently applying a learned perceptual rule or heuristic that “low is large” (Morton, 1977; Pisanski et al., 2012; see also Chapter 4) would likely not result in above-chance accuracy because low voice pitch is not directly related to size within sexes.

Future work should nevertheless explore other viable explanations for the facilitating role of voice pitch in body size perception above and beyond that which can be explained by increased spectral sampling or harmonic density. It may, for instance, be the case that the relationship between voice pitch and size within sexes is present but is statistically weak requiring large samples (Puts et al., 2012) or non-linear statistics (Fitch & Giedd, 1999; Turner et al., 2009). There may also be additional pitch-related features of the human voice that can provide reliable information about size but that have yet to be thoroughly investigated (e.g., jitter and shimmer, González, 2007; Hamdan et al., 2012).

Conclusions

We tested a number of hypotheses to empirically address the proposal that voice pitch confounds within-sex size estimation from the human voice. We overturn this common belief. Despite having no reliable *direct* relationship to men's height, we show that the presence of voice pitch in the vocal signal nevertheless increases accurate size perception, in part by providing a strong carrier signal for formants. Determining the physical and perceptual mechanisms (misattributions or otherwise) underlying vocal indices of body size is essential to understanding signaler-receiver psychology as well as, more broadly, the origins and functions of animal vocalizations.

References

Assmann, P. F., & Nearey, T. M. (2008). Identification of frequency-shifted vowels. *Journal of the Acoustical Society of America, 124*(5), 3203-3212. doi:10.1121/1.2980456

Bachorowski, J. A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America, 106*(2), 1054-1063.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.2.15) [Software]. Available from www.praat.org

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 B: Biological Sciences, 273*(1582), 83-89. doi:10.1098/rspb.2005.3265

Charlton, B.D., Taylor, A.M., & Reby, D. (2013). Are men better than women at acoustic size judgements? *Biology Letters, 9*, 20130270. doi:10.1098/rsbl.2013.0270

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

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65*, 997-1004. doi: 10.1006/anbe.2003.2123

Dabbs, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences, 27*(4), 801-804. doi:10.1016/s0191-8869(98)00272-4

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. doi: 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. doi: 10.1016/j.physbeh.2007.11.033

Ey, E., Pfefferle, D. & Fischer, J. (2007). Do age-and sex-related variations reliably reflect body size in non-human primate vocalizations? A review. *Primates, 48*, 253-267.

Fant, F. (1960). *Acoustic theory of speech production*. Holland: Mouton, The Hague.

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. doi: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*, 561-568. doi:10.1016/j.anbehav.2004.06.012

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America, 102*(2 Pt 1), 1213-1222. doi:10.1121/1.421048

Fitch, W. T. (2000a). Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry. *Zoology, 103*, 40-58.

Fitch, W. T. (2000b). The evolution of speech: A comparative review. *Trends in Cognitive Sciences, 4*(7), 258-267. doi: 10.1016/s1364-6613(00)01494-7

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

Fitch, W. T., & Hauser, M. (2003). Unpacking "honesty": Vertebrate vocal production and the evolution of acoustic signals. *Acoustic Communication, 16*, 65-137. doi:10.1007/0-387-22762-8_3

Fleming, V. (1939). The Wizard of Oz [Motion picture]. Los Angeles, CA: Metro-Goldwyn-Mayer Studios Inc.

Gingras, B, Boeckle, M, Herbst, C. T., & Fitch, W. T. (2013). Call acoustics reflect body size across four clades of anurans. *Journal of Zoology, 289*(2), 143-150. doi:10.1111/j.1469-7998.2012.00973.x

González, J. (2003). Estimation of speakers' weight and height from speech: a re-analysis of data from multiple studies by Lass and colleagues. *Perceptual and Motor Skills, 96*(1), 297-304. doi:10.2466/pms.2003.96.1.297

González, J. (2004). Formant frequencies and body size of speaker: A weak relationship in adult humans. *Journal of Phonetics, 32*(2), 277-287.

González, J. (2006). Research in acoustics of human speech sounds: Correlates and perception of speaker body size. *Recent Research Developments in Applied Physics, 9*, 1-15.

González, J. (2007). Correlations between speakers' body size and acoustic parameters of voice. *Perceptual and Motor Skills, 105*(1), 215. doi:10.2466/PMS.105.5.215-220

Graddol, D., & Swann, J. (1983). Speaking fundamental frequency: Some physical and social correlates. *Language and Speech, 26*(4), 351-366. doi: 10.1177/002383098302600403

Greisbach, R. (1999). Estimation of speaker height from formant frequencies. *International Journal of Speech Language and Law, 6*(2), 265-277.

Hamdan, A. L., Al-Barazi, R., Tabri, D., Saade, R., Kutkut, I., Sinno, S. (2012). Relationship between acoustic parameters and body mass analysis in young males. *Journal of Voice, 26*(2), 144-147. doi:10.1016/J.Jvoice.2011.01.011

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. *Journal of Laryngology and Otology, 112*(5), 451-454. doi:10.1017/S0022215100140757

Hauser, M. D. (1993). The evolution of nonhuman primate vocalizations: Effects of phylogeny, body weight, and social context. *American Naturalist*, 528-542. doi:10.1086/285553.

Hollien, H., & Jackson, B. (1973). Normative data on the speaking fundamental frequency characteristics of young adult males. *Journal of Phonetics, 1*(2), 117-120.

Irino, T., Aoki, Y., Kawahara, H., & Patterson, R. D. (2012). Comparison of performance with voiced and whispered speech in word recognition and mean-formant-frequency discrimination. *Speech Communication 54*, 998-1013. doi:10.1016/j.specom.2012.04.002

Ives, D. T., Smith, D. R. R, & Patterson, R. D. (2005). Discrimination of speaker size from syllable phrases. *Journal of the Acoustical Society of America, 118*, 3816-3822. doi: 10.1121/1.2118427

Jovičić, S. T. (1998). Formant feature differences between whispered and voiced sustained vowels. *Acta Acustica united with Acustica, 84*(4), 739-743.

Kreiman, J., & Sidtis, D. (2011). Physical characteristics and the voice: Can we hear what a speaker looks like? In J. Kreiman & D. Sidtis (Eds.), *Foundations of voice* *studies: An interdisciplinary approach to voice production and perception* (pp. 110-155). West Sussex, UK: Wiley-Blackwell.

Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica, 46*(1-3), 117-125.

Lass, N. J., Kelley D. T, Cunningham, C. M., & Sheridan, K. J. (1980). A comparative study of speaker height and weight identification from voiced and whispered speech. *Journal of Phonetics, 8*(2), 195-204.

Lass, N. J., & Brown, W. S. (1978). Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies. *Journal of the Acoustical Society of America, 63*(4), 1218-1220. doi: 10.1121/1.381808

Lee, S., Potamianos, A., & Narayanan, S., (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America, 105*, 1455–1468. doi: 10.1121/1.426686

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology, 46*(2), 117-128.

Lieberman, P., Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge, MA: Cambridge University Press.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables* (Vol. 7). Thousand Oaks, CA: Sage Publications.

Majewski, W., Hollien, H., & Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males. *Phonetica, 25*(2), 119-125. doi: 10.1159/000259375

Monahan, P. J., & Idsardi, W. J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalisation. Language and cognitive processes, 25(6), 808-839. doi: 10.1080/01690965.2010.490047

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, 855-869.

Patterson, R. D., Smith, D. R. R., van Dinther, R., & Walters, T. C. (2008). Size information in the production and perception of communication sounds. In W. A. Yost, Popper, A. N., & Fay, R. R. (Eds.), *Auditory perception of sound sources* (pp. 43-75). New York, NY: Springer.

Peterson, G. E, & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175-184. doi:10.1121/1.1906875

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. doi: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.

Journal of the Acoustical Society of America, 129(4), 2201-2212. doi:10.1121/1.3552866

Puts, D. A., Apicella, C. L., & Cardenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 279*(1728), 601-609. doi:10.1098/rspb.2011.0829

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*(4497), 947-949. doi:10.1126/science.7233191

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F-0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America, 117*(2), 944-955. doi: 10.1121/1.1848011

Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1208-1219. doi:10.1037/0096-1523.33.5.1208

Ryalls, J. H., & Lieberman, P. (1982). Fundamental frequency and vowel perception. *Journal of the Acoustical Society of America, 72*(5), 1631-1634. doi:10.1121/1.388499

Ryan, M. J. (1988). Constraints and patterns in the evolution of anuran acoustic communication In B. Fritzsch (Ed.), *The evolution of the amphibian auditory system* (pp. 637-677). Hoboken, NJ: John Wiley & Sons, Inc.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences, 277*(1699), 3509-3518. doi: 10.1098/rspb.2010.0769

Shields, M., Gorber, S.C., Janssen, I., & Tremblay, M.S. (2011). Bias in self-reported estimates of obestity in Canadian health surveys: An update on correction equations for adults. *Health Reports, 22*(3), 35-45.

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Listening to different speakers: On the time-course of perceptual compensation for vocal tract characteristics. *Neuropsychologia*, 49(14), 3831-3846. doi: 10.1016/j.neuropsychologia.2011.09.044

Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America, 118*(5), 3177-3186. doi:10.1121/1.2047107

Taylor, A. M., & Reby, D. (2010). The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology, 280*(3), 221-236. doi:10.1111/J.1469-7998.2009.00661.X

Titze, I.R. (1989). Physiological and acoustic differences between male and female voices. *Journal of the Acoustical Society of America, 85*, 1699-1707. doi:10.1121/1.397959

Titze, I.R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Turner, R. E., Walters, T. C., Monaghan, J. J. M, & Patterson, R. D. (2009). A statistical, formant-pattern model for segregating vowel type and vocal tract length in developmental formant data. *Journal of the Acoustical Society of America, 125*(4), 2374-2386. doi: 10.1121/1.3079772

van Dommelen, W. A, & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: Sex-specific behaviour. *Language and Speech, 38*(3), 267-287. doi:10.1177/002383099503800304

Vestergaard, M. D., Fyson, N. R. C., & Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *Journal of the Acoustical Society of America*, *125*(2) 1114–1124. doi: 10.1121/1.3050321

Vestergaard, M. D., Fyson, N. R. C., & Patterson, R. D. (2011). The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening, *Journal of the Acoustical Society of America, 130*(1), 429-439. doi: 10.1121/1.3596462

CHAPTER 4

LOW IS LARGE: PITCH INTERACTS WITH SPATIAL CUES IN SIZE PERCEPTION

Pisanski, K., O'Connor, J.J.M., Tigue, C.C., & Feinberg, D.R. Low is large: Pitch interacts with spatial cues in size perception (under review). *Submitted July 8, 2014 to Cognition*. 1-21.

Preface

In addition to associating low voice pitch and low formants with large body size (see Chapter 3), studies have also shown that listeners associate low-frequency sounds with largeness more generally. This is true even if the low-frequency sounds are abiological such as synthesized or musical tones or sounds produced by large objects colliding (Charlton et al., 2013a; Grassi, 2005; Eitan & Timmers, 2010; Huron, Kinney, & Precoda, 2006). Additionally, low pitch is associated with low spatial locations and high pitch with high spatial locations (Pratt, 1930; Walker, 1987). Many researchers have hypothesized that perceptual associations between low vocal frequencies and large size may represent a general bias that affects perception in multiple domains (Morton, 1977;

Ohala, 1984; Parise, Knorre, & Ernst, 2014; Peña, Mehler, & Nespor, 2011; Pisanski et al., 2012; Walker et al., 2010).

To date only one study has demonstrated that people conceptually associate physically high elevations with largeness (Eitan & Timmers, 2010). This association is likely to be grounded in reality, because the voices of relatively taller people will project from higher elevations. However, because people in western culture also associate low pitch with low elevation and largeness, size perception may represent a cross-modal incongruence between *high* and *large* in the spatial modality, and *low* and *large* in the auditory modality.

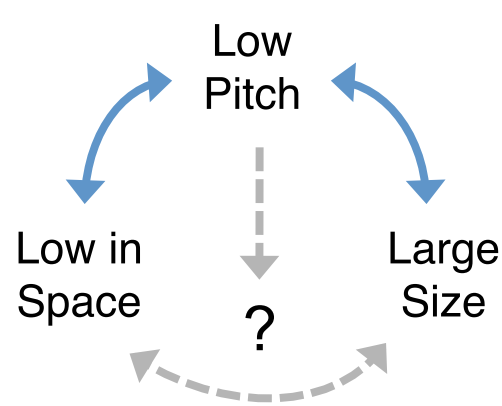

Figure 1. Perceptual associations among pitch, size, and spatial location. Solid lines represent perceptual relationships established by previous studies. Dashed lines represent relationships that to date have not been thoroughly investigated, and which I test in Chapter 4.

In Chapter 4, I test whether low voice pitch causes listeners to erroneously associate low spatial locations with large size (see Figure 1 above). The results of Chapter 4 provide novel insight into how perceptual biases, particularly those that may be misleading, can affect body size estimation. Thus, in conjunction with the findings presented in Chapters 2 and 3, this work represents yet another step toward understanding why the perceptual association between pitch and body size exists at the within-sex level.

Abstract

Incongruent cues from different sensory modalities pose a challenge for multimodal integration. Although taller objects emit sounds from higher elevations, low-pitched sounds are mapped cross-modally both to large size and to low elevation. Here we examine this cross-modal perceptual incongruence between *high* and *large* in the spatial modality, and *low* and *large* in the auditory modality. We measured listeners' judgments of human body size with pitch-manipulated voices projected from a high versus low, and right versus left, spatial location. Following the Pratt Effect in sound localization, wherein low-pitched tones projected from high elevations are perceived as originating low to the ground, we predicted that low voice pitch would influence vertical but not horizontal spatial cues to size. Indeed, lowered-pitch voices were judged as larger when originating from physically low than high spaces, however size perception of raised-pitch voices was not affected by elevation. Voices were also judged as larger when originating from the right than left regardless of pitch. These effects were present for judgment's of both men's and women's voices. Our findings provide novel evidence for a multidimensional spatial mapping of pitch that is generalizable to voice pitch, and demonstrate that cross-modal incongruence between pitch-size and pitch-height associations can cause errors of multisensory integration in size perception.

Introduction

Cross-modal sensory perception is fundamental for developing and calibrating normal multisensory integration (Seilheimer, Rosenberg, & Angelaki, 2014 for review). Although it is generally studied as a low-level perceptual phenomenon, multisensory integration is likely to involve both low-level and high-level neurocognitive mechanisms (Campanella & Belin, 2007; Seilheimer et al., 2014). Auditory pitch elicits strong cross-modal associations with both spatial location and size that affect performance in perceptual, cognitive, and attentional tasks (see Marks, 2000; Spence & Deroy, 2013 for reviews) and that have been observed across cultures (Diffloth, 1994; Ohala, 1984; Ultan, 1978) and in infants as young as four months of age (Peña, Mehler, & Nespor, 2011; Roffler & Butler, 1968; Walker et al., 2010). Indeed, in many languages and in musical discourse, pitch is regularly described relative to space and size using words such as *high* and *low*, *falling* and *rising*, *heavy* and *light,* or *thick* and *thin* (Ashley, 2004; Dolscheid, Shayan, Majid, & Casasanto, 2013).

In western English-speaking culture, low pitch is perceived as originating vertically lower in physical space whereas high pitch is mapped to higher elevations, commonly known as frequency-elevation mapping[6] (Eitan & Timmers, 2010; Evans & Treisman, 2010; Mudd, 1963; Parise, Knorre, & Ernst, 2014). Low pitch is also perceptually associated with large physical size and high pitch with small size regardless whether the sounds are pure or complex tones, musical passages, or vocalizations (Bien,

[6] Although weaker than its vertical correspondence, low pitch is sometimes also mapped horizontally, but the direction of the mapping varies across studies (see Eitan & Granot, 2006).

ten Oever, Goebel, & Sack, 2012; Evans & Treisman, 2010; Feinberg, Jones, Little, Burt, & Perrett, 2005; Huron, Kinney, & Precoda, 2006; Morton, 1977; Ohala, 1984; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014).

In the context of vocal cues to body size, voices of taller individuals will project from spatially higher elevations than those of shorter individuals. To our knowledge only one other study has tested for an analogous perceptual association. Corresponding to the physical relationship, Eitan and Timmers (2010) found that participants associated spatially high objects with large size in a task that involved pairing antonyms (such as large-small) with either high or low elevation. However, participants also associated low pitch with large size. Taken together this poses a cross-modal incongruence (Spence, 2011) between *high* and *large* in the spatial modality, and *low* and *large* in the auditory modality. No previous study has examined which of these cross-modal correspondences prevail when spatial cues and pitch cues to size conflict.

The binding of cross-modally incongruent cues can cause errors in multimodal integration and perception (Spence, 2011). A well-known example of the cross-modal binding problem is the Pratt Effect (Pratt, 1930), in which the perceived spatial location of pure tones is determined more by their pitch than their spatial elevation. Low-pitched tones projected from high elevations are perceived as originating low to the ground. Hence, auditory pitch cues override spatial cues in sound localization. The Pratt Effect has been replicated many times (Bregman & Steiger, 1980; Cabrera, Ferguson, Tilley, & Morimoto, 2005; Morimoto & Aokata, 1984; Roffler & Butler, 1968; Trimble, 1934).

Pitch has also been shown to affect performance in visuospatial tasks that do not involve explicit sound localization, for example, responding to high-pitched tones with *down* responses on a keyboard (Rusconi, Kwan, Giordano, Umilta, & Butterworth, 2006). Low-pitched sounds appear to affect spatial perception more than relatively high-pitched sounds (Eitan & Granot, 2006; Eitan & Timmers, 2010).

In the present study we tested for the first time whether pitch interacts with spatial cues in size perception. We measured listeners' perceptions of human body size with pitch-manipulated voices projected from a high versus low, and right versus left, spatial location. Following the effect of auditory pitch on sound localization we hypothesized that low pitch may override spatial height cues in the perception of body size. Thus, we predicted that listeners would associate vertically high spatial cues with largeness for voices with raised-pitch (where high pitch is mapped to high elevations, *congruent* to the mapping of large size to high elevations), but vertically low spatial cues with largeness for those same voices with lowered-pitch (where low pitch is mapped to low elevations, *incongruent* to the mapping of large size to high elevations). In contrast, we predicted that pitch would have no effect on size perception along the horizontal axis.

Methods

Participants

Forty-six participants (mean age: 19.5±1.6 years, all female) took part in the experiment as raters, and 10 different participants (mean age: 18±0.3 years, 5 males, 5 females) provided voice recordings. All participants were recruited from the psychology undergraduate research pool at McMaster University, provided informed consent, and received partial course credit for their participation.

Auditory Stimuli

We recorded voices in an anechoic sound-controlled booth using a Sennheiser MKH 800 condenser microphone with a cardioid pick-up pattern. Content-neutral recordings were of the five English monophthong vowels /ɑ/, /i/, /ɛ/, /o/, and /u/. Audio was digitally encoded with an M-Audio Fast Track Ultra interface at a sampling rate of 96 kHz and 32-bit amplitude quantization, and stored onto a computer as PCM WAV files using Adobe Soundbooth CS5 version 3.0.

Pitch was raised or lowered by 10% from baseline using the Pitch-Synchronous Overlap Add (PSOLA) algorithm in Praat version 5.2.15 (Boersma & Weenink, 2013). This resulted in two versions (raised-pitch and lowered-pitch) of each original voice. The PSOLA method alters one voice feature (e.g., pitch) while leaving other features unaltered (Moulines & Charpentier, 1990). Our pitch manipulation corresponded to approximately two times the just-noticeable difference in voice pitch perception (Pisanski & Rendall, 2011; Re, O'Connor, Bennett, & Feinberg, 2012) and size perception from voice pitch (Smith & Patterson, 2005). Pitch-manipulationed men's voices (M±SD, raised-pitch: 122±4 Hz; lowered-pitch: 99±3 Hz) and women's voices (raised-pitch: 238±1 Hz; lowered-pitch: 194±1 Hz) spanned natural ranges for each sex (Titze, 1989). The sound pressure level (SPL) of each of the 20 stereo identically-channelled voice stimuli was amplitude normalized to 70 dB using the root–mean squared method (see also *Sound-Pressure Level Check*) and each voice stimulus was panned 100% left or right in Praat.

Sound-Speaker Array And Stimulus Playback

Four sound-speakers (19 cm Bose, Companion 2 series II multi media speaker system, Canada) were positioned parallel on the wall in a symmetrical array (see Fig. 2). The center of the array was positioned parallel with the center of the participants head adjacent the array at a distance of 213 cm (7 feet) and 0° elevation and azimuth, such that voices projected along the vertical axis would be perceived as coming from above or below (16° elevation, 0° azimuth) and voices projected along the horizontal axis would be perceived as coming from the left or right (16° azimuth, 0° elevation) of the participant's head. Head position was standardized using a mounted chin rest. The four sound-speaker symmetrical array allowed us to manipulate the spatial location of voices along one axis while holding the other constant. This design limited localization cues to either the azimuth or elevation planes reducing the potential for sound-source confusion (Middlebrooks & Green, 1991), and ensured that the distance from the participant and the sound-source was the same for all four sound-speakers.

Auditory stimuli were played back through a computer via a THX TruStudio Pro high-definition Sound Blaster at a sampling rate of 96 kHz and 24-bit DAC resolution (Creative Technologies Ltd., Model SB1095, Singapore). The voices were played from only one of two channels in each condition (vertical condition: high-low channel; horizontal condition: left-right channel).

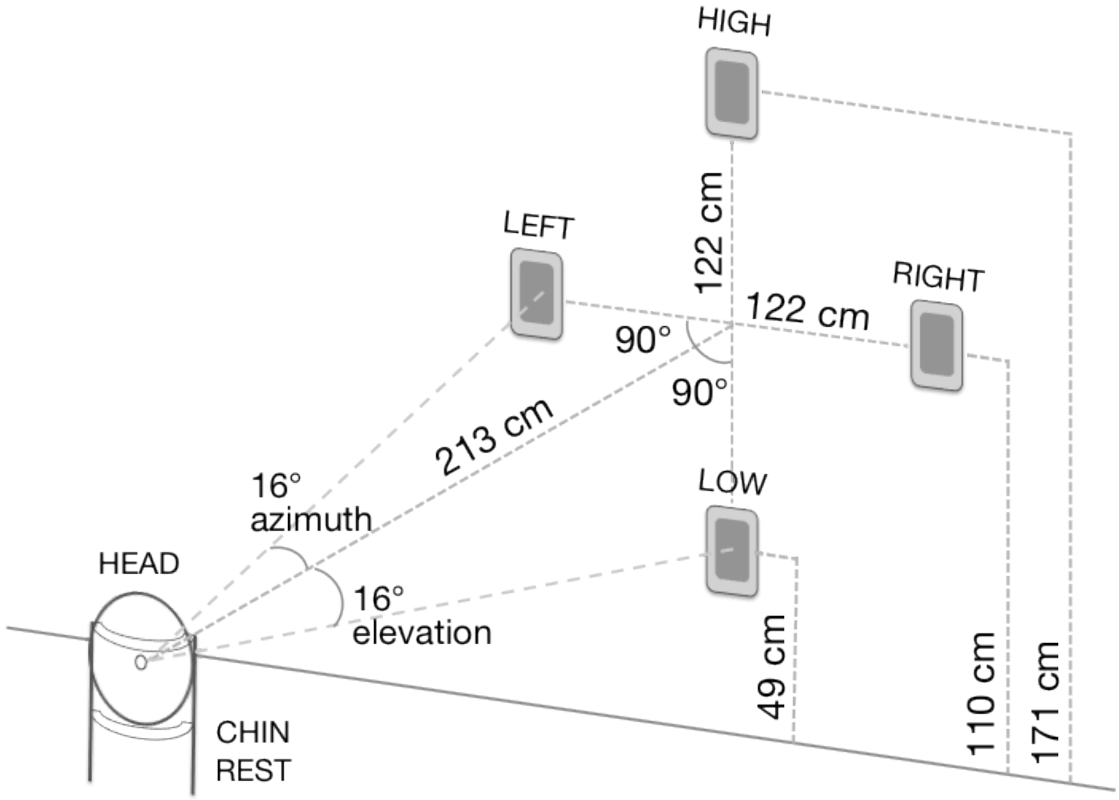

Figure 2. Four sound-speaker array.

Sound-Pressure Level Check

Sound localization errors are minimized (3-5º) when the SPL of auditory stimuli is 70 dB or higher (Davis & Stephens, 1974). Thus, a sound level meter (Brüel & Kjær, Type 2239, Denmark) was used to test the free-field SPL of voice stimuli projected from each of the four sound-speakers at the location of the chin rest. Average free-field SPL was 71.02 dB. Because louder sounds may be perceived as lower in pitch (Davis & Stephens, 1974; Wier, Jesteadt, & Green, 1977) and as originating from a larger source

(Walker, 1987), we confirmed that there were no perceivable differences in free-field SPL (i.e., differences were <1 dB) between lowered-pitch (70.94 dB) and raised-pitch (71.1 dB) voices.

Procedure

The experiment was approved by the McMaster Research Ethics Board. It was conducted in a private room in the Voice Research Laboratory at McMaster University. Each participant completed the same protocol. Before beginning the experiment, participants' heads were positioned in a chin rest and their stool height was adjusted. Once comfortable, participants were instructed to leave their head in the chin rest for the entirety of the experiment. They were then instructed that they would hear a series of voices and that their task would be to rate the body size of each person speaking on a scale from 1 ("very small") to 7 ("very large"). The experimenter then initiated the experimental protocol on the computer and left the room.

Auditory stimuli were projected directly from the speakers mounted on the wall (free-field). Participants were presented with a single voice on each trial and completed 40 trials in each of two conditions: a *vertical condition* in which each voice stimulus (10 raised-pitch, 10 lowered-pitch) was projected once from the high and once from the low sound-speaker, and a *horizontal condition* in which each voice stimulus was projected once from the left and once from the right sound-speaker. The experimenter implemented one channel change manually in between conditions, which participants did not observe. Experimenters were not visible or audible during testing. Condition order was counter-balanced between participants and the presentation of voice stimuli within each condition was randomized, including sound-speaker side. Participants inputted size judgments and their age into a custom computer interface using the numeric keypad of a keyboard.

Analysis

Average body size estimates were calculated separately for each subject for each condition. We ran a repeated measures analysis of variance (dependent variable: body size estimate [coded 1-7]; within-subject factors: *pitch manipulation* [raised-pitch, lowered-pitch], *axis* [vertical, horizontal], and *sound-speaker side* [high, low or left, right] and *sex of voice* [male, female]). We then created contrast variables by subtracting size judgments between speakers on the same axis (high-low; left-right) and used planned paired-sample and one-sample *t*-tests to examine significant effects revealed by the omnibus model (dependent variable: contrast variable; alpha .05; 95% confidence intervals (CI) given).

Results

We found a significant three-way interaction among *pitch manipulation, axis,* and *sound-speaker side* ($F_{1,45}$=8.54, p=.005, η_ρ^2=.16). As illustrated in Fig. 3, size estimates between high and low sound-speakers differed for the two pitch manipulations (t_{45}=2.88, p=.006; 95% CI=0.06–0.36). This pattern was observed for judgments of both men's and women's voices (no *sex of voice* effect: t_{45}=0.89, p=.40; 95% CI=-0.18–0.44). Thus, collapsing across sex of voice, listeners associated vertically low spatial cues with largeness for lowered-pitch voices (t_{45}=2.94, p=.005; 95% CI=0.05–0.25) but not for raised-pitch voices (t_{45}=-1.03, p=.31; 95% CI=-0.19–0.06).

In contrast, size estimates between left and right sound-speakers did not vary as a function of pitch manipulation (t_{45}=-1.22, p=.23; 95% CI=-0.25–0.06; see Fig. 3). Again this pattern was observed for judgments of both men's and women's voices (t_{45}=1.1, p=.28; 95% CI=-0.14–0.48). Thus, collapsing across sex of voice and pitch manipulation, listeners associated horizontally rightward spatial cues with largeness (t_{45}=3.94, p<.001; 95% CI=0.07–0.21). There was a significant interaction among *sex of voice, axis,* and *sound-speaker side* ($F_{1,45}$=6.19, p=.017, η_ρ^2=.12), but sex of voice did not affect size estimates along the vertical (t_{45}=1.84, p=.073; 95% CI=-0.02–0.32) or horizontal axes (t_{45}=1.73, p=.091; 95% CI=-0.02–0.28).

We found several main effects that corroborate past work (see, e.g., Eitan & Timmers, 2010; Feinberg et al., 2005; Pisanski, Mishra, & Rendall, 2012): an effect of *pitch manipulation,* where lowered-pitch voices (M±SD =4.42±0.82) were judged as larger than were raised-pitch voices (3.5±0.93; $F_{1,45}$=414.97, p<.001, η_ρ^2=.90); *sex of voice,* where men's voices (4.39±0.82) were judged as larger than were women's voices (3.52±0.9; $F_{1,45}$=76.5, p<.001, η_ρ^2=.63); and *sound-speaker side,* where voices projected from the low (4.03±1.23) and right (4.07±1.24) sound-speakers were judged as larger than were voices projected from the high (3.99±1.23) and left (3.97±1.23) sound-speakers ($F_{1,45}$=9.82, p=.003, η_ρ^2=.18).

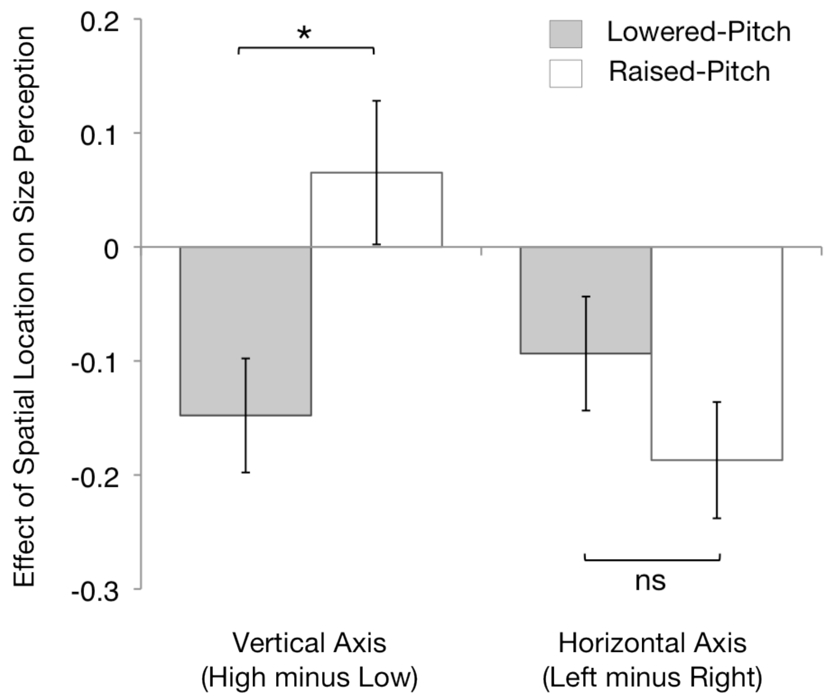

Figure 3. Mean difference ± SEM in size perception between sound-speaker locations on the vertical (high-low) or horizontal (left-right) axis.

Note. Dark and light bars represent size estimates for lowered-pitch and raised-pitch voices, respectively, collapsing across sex of voice. Values below 0 indicate larger size estimates for voices originating spatially low than high (vertical axis) or spatially right than left (horizontal axis)(* p=.006; ns = p>.05).

Discussion

In everyday perception we are faced with the challenge of integrating cues across multiple and often incongruent modalities. Here, we present the first evidence that the spatial location of the sound-source of a human voice alters the perceived size of the vocalizer in both the vertical and horizontal dimensions. However, as predicted by the Pratt Effect (Pratt, 1930), we show that auditory pitch influences the association between elevation and size. Lowered-pitch voices were judged as larger when originating from physically low than high spaces whereas raised-pitch voices were not. In contrast, regardless of pitch, voices were judged as larger when originating from the right than left. The magnitude of these effects was analogous for judgments of men's and women's voices.

This work offers insight into the nature and scope of cross-modal associations among voice pitch, space, and size. It demonstrates that pitch affects elevation perception not only in sound localization (Bregman & Steiger, 1980; Morimoto & Aokata, 1984; Roffler & Butler, 1968; Trimble, 1934), but also in an indirect spatial task involving size estimation. Moreover, this finding indicates that the cross-modal correspondence between *low* and *large* (in the auditory modality) can be stronger than the incongruent correspondence between *low* and *small* (in the spatial modality).

Our study does not support the hypothesis that frequency-elevation mapping is functionally adaptive such that it aids in locating the source of a sound (Parise et al., 2014; Stumpf, 1883; Walker et al., 2010). Here, low voice pitch caused listeners to associate low spatial verticality with large body size. This does not reflect the true nature of the relationship in humans. Voices of taller people project from higher elevations than do those of shorter people, and pitch does not reliably predict body size within sexes (for meta-analysis, see Pisanski et al., 2014). Indeed, in the absence of conflicting pitch cues, listeners appropriately associate high elevations with largeness and low elevations with smallness (Eitan & Timmers, 2010). Interestingly, smaller animals may sometimes emit sounds from higher elevations than do larger animals (e.g., insects and birds compared to large terrestrial mammals). Future studies may use the current paradigm to examine how space and pitch interact to affect voice-based size perception of non-human animals.

Low pitch affected size perception along the vertical axis more than did high pitch. Past studies also report strong associations between low pitch and low elevation but weak or no associations between high pitch and high elevation (Eitan & Granot, 2006; Eitan & Timmers, 2010). Thus our findings provide additional support in refutation of the *directional symmetry hypothesis* (Eitan & Granot, 2006), and suggest that low frequencies may elicit stronger cross-modal correspondences than high frequencies. This pitch asymmetry also suggests that our findings cannot be explained by low-level interactions (e.g., that low-frequency sounds transmit better from lower than higher space, Morton, 1977). However, these interpretations may be treated with caution because although we controlled for the difference between physical and perceived frequency in the auditory domain using a logarithmic scale, our treatment of height was linear. Thus lower elevations may be perceived as further in magnitude from the azimuth (i.e., the participant's head) than higher elevations.

Ours is the first study to test for a horizontal association between size and space and the first of its kind using human voices. Although some studies report horizontal pitch biases, low pitch is sometimes associated with the right and sometimes with the left (Eitan & Granot, 2006; Eitan & Timmers, 2010; Mudd, 1963; Stewart, Walsh, & Frith, 2004). In the present study listeners consistently associated the right with large size. Horizontal size estimates may reflect semantic or numeric coding. For example, large numbers are associated with rightward responses in various cognitive tasks (Dehaene, Bossini, & Giraux, 1993). Another possibility is that our finding reflects hemispheric specialization that develops through experience. This is supported by evidence of reversed horizontal biases in people from cultures that read right-to-left (Maass & Russo, 2003) and among musicians (Stewart et al., 2004). Furthermore, because 90-95% of men and women are right-handed, the right sides of their bodies are typically stronger and larger than the left, which may contribute to a right-large bias (Günther, Bürger, Rickert, Crispin, & Schulz, 2008; Roy, Ruff, & Plato, 1994; Schell, Johnston, Smith, & Paolone, 1985).

Our study demonstrates that spatial correspondences with pitch, typically studied with tones and music (Ashley, 2004; Eitan & Timmers, 2010), generalize to human voice pitch. Whereas previous studies have used broad pitch ranges spanning 200-8000 Hz in pure tone stimuli (see, e.g., Cabrera et al., 2005; Mudd, 1963; Parise et al., 2014; Pratt, 1930), we demonstrate that frequency-elevation mapping is elicited with subtle pitch differences of only 20-40 Hz at pitch centers of 100-250 Hz. This supports the hypothesis that pitch biases in body size perception may reflect a broader sound-size symbolism (Morton, 1977; Ohala, 1984). We demonstrate here that pitch biases in spatial perception are highly general and may contribute to errors in multisensory integration in a wide range of contexts, including body size estimation.

References

Ashley, R. (2004). Musical pitch space across modalities: Spatial and other mappings through language and culture. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.) *Proceedings of the 8th international conference on music perception and cognition* (pp. 64-71). Adelaide, Austrailia: Causal Productions.

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage, 59*(1), 663-672.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.2.15).

Bregman, A. S., & Steiger, H. (1980). Auditory streaming and vertical localization: Interdependence of "what" and "where" decisions in audition. *Perception & Psychophysics, 28*(6), 539-546.

Cabrera, D., Ferguson, S., Tilley, S., & Morimoto, M. (2005). Recent studies on the effect of signal frequency on auditory vertical localization. *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display,* 1-8.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11*(12), 535-543.

Davis, R., & Stephens, S. (1974). The effect of intensity on the localization of different acoustical stimuli in the vertical plane. *Journal of Sound and Vibration, 35*(2), 223-229.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371.

Diffloth, G. (1994). *i*: big, *a*: small. In J. J. Ohala, L. Hinton & J. Nichols (Eds.), *Sound symbolism* (pp. 107-113). Cambridge, MA: Cambridge University Press

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch psychophysical evidence for linguistic relativity. *Psychological Science, 24*(5), 613-621.

Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners' images of motion. *Music Perception: An Interdisciplinary Journal, 23*(3), 221-248.

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition, 114*(3), 405-422.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 6.

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*, 561-568.

González, J. (2006). Research in acoustics of human speech sounds: Correlates and perception of speaker body size. *Recent Research Development in Applied Physics, 9*, 1-15.

Günther, C. M., Bürger, A., Rickert, M., Crispin, A., & Schulz, C. U. (2008). Grip strength in healthy caucasian adults: reference values. *The Journal of Hand Surgery, 33*(4), 558-565.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review, 1*(3), 170-177.

Maass, A., & Russo, A. (2003). Directional Bias in the Mental Representation of Spatial Events Nature or Culture? *Psychological Science, 14*(4), 296-301.

Marks, L. E. (2000). Synesthesia. In E. Cardeña, S. J. Lynn & S. Krippner (Eds.), *Varieties of anomalous experience: examining the scientific evidence* (pp. 121-149). Washington, DC: American Psychological Association.

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology, 42*(1), 135-159.

Morimoto, M., & Aokata, H. (1984). Localization cues of sound sources in the upper hemisphere. *Journal of the Acoustical Society of Japan, 5*(3), 165-173.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist, 111*, 855-869.

Moulines, E., & Charpentier, F. (1990). Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9*, 453-467.

Mudd, S. (1963). Spatial stereotypes of four dimensions of pure tone. *Journal of Experimental Psychology, 66*(4), 347-352.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of *F*0 of voice. *Phonetica, 41*(1), 1-16.

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences, 111*(16), 6104-6108.

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science, 22*(11), 1419-1421.

Pisanski, K., Fraccaro, P.J., Tigue, C.C., O'Connor, J.J.M., & Feinberg, D.R. (2014). Return to Oz: Voice pitch facilitates assessment of men's body size. *Journal of Experimental Psychology: Human Perception & Performance, 40*(4), 1316-1331.

Pisanski, K., Fraccaro, P.J., Tigue, C.C., O'Connor, J.J.M., Röder, S, Andrews, W.P, Bernhard, F., DeBruine, L.M., Jones, B.C., & Feinberg, D.R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour, 95,* 89-99.

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.

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

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology, 13*(3), 278-285.

Re, D. E., O'Connor, J. J. M., Bennett, P. J., & Feinberg, D. R. (2012). Preferences for very low and very high voice pitch in humans. *PloS ONE, 7*(3), e32719.

Roffler, S. K., & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *Journal of the Acoustical Society of America, 43*, 1255-1259.

Roy, T. A., Ruff, C. B., & Plato, C. C. (1994). Hand dominance and bilateral asymmetry in the structure of the second metacarpal. *American Journal of Physical Anthropology, 94*(2), 203-211.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition, 99*(2), 113-129.

Schell, L. M., Johnston, F. E., Smith, D. R., & Paolone, A. M. (1985). Directional asymmetry of body dimensions among white adolescents. *American Journal of Physical Anthropology, 67*(4), 317-322.

Seilheimer, R. L., Rosenberg, A., & Angelaki, D. E. (2014). Models and processes of multisensory cue combination. *Current Opinion in Neurobiology, 25*, 38-46.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America, 118*(5), 3177-3186.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971-995.

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition, 22*(1), 245-260.

Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception & Psychophysics, 66*(2), 183-195.

Stumpf, C. (1883). *Tonpsychologie (Vol. 1)*. Leipzig, Germany: Hirzel.

Titze, I. R. (1989). Physiological and acoustic differences between male and female voices. *Journal of the Acoustical Society of America, 85*, 1699-1707.

Trimble, O. C. (1934). Localization of sound in the in the anterior-posterior and vertical dimensions of "auditory space". *British Journal of Psychology, 24*(3), 320-334.

Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Eds.), *Universals of human language Vol. 2*: *Phonology* (pp. 527-568). Stanford, CA: Stanford University Press.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., et al. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21-25.

Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Perception & Psychophysics, 42*(5), 491-502.

Wier, C. C., Jesteadt, W., & Green, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America, 61*, 178-184.

CHAPTER 5: GENERAL DISCUSSION

Relationships among physical body size, the voice, and perceptions of size from the voice are multifaceted and complex. Biological, psychological, and cultural factors can influence these relationships. In this thesis, I examined vocal production and perception in relation to men and women's body size. This work provides evidence that reliable information about body size is present in the human voice and that listeners can accurately assess size from the voice at the within-sex level. However, this work also highlights that relationships between the voice and body size in humans are weak within sexes and that perceptual pitch biases can be misleading in the context of body size estimation. In what follows, I expand briefly on the findings, implications, and future directions of each empirical chapter individually.

Formants Predict Size Better Than Does Fundamental Frequency

In Chapter 2, in a meta-analysis, I showed that fundamental frequency significantly predicted height among men but not women, and weight among women but not men. However, very large sample sizes were required for fundamental frequency to reliably predict height or weight at the within-sex level. In contrast, a variety of formant-based measures independently and reliably predicted height and weight in both men and women with much smaller minimum sample sizes. Based on the results of the meta-analysis, future studies should use VTL(ΔF) (Reby & McComb, 2003) to estimate men's height and women's weight, *F*4 to estimate women's height, and P_f (Puts et al., 2012a) to estimate men's weight from formants.

The results of the meta-analysis provide the most conclusive evidence to date in support of the hypothesis that formants are a more reliable predictor of body size, and in particular of height, than is pitch in humans (Fitch, 1994, 2000a, 2000b; Fitch & Giedd, 1999). A similar pattern of results has been demonstrated in a number of other mammals (see Fitch & Hauser, 2003 for review). The finding that formants, unlike pitch, reliably predict height and weight in men and women suggests that selection could have operated on vocal tract length and/or formant frequencies to reliably communicate body size. However, the relationship between formants and size could be maintained even in the absence of selection pressure, for example, due to anatomical constraints on vocal tract length (Fitch, 2000a, b). As no more than 10% of the variation in height and 6% of the variation in weight among men or women was accounted for by any given vocal measure, the findings of Chapter 2 corroborate previous suggestions that vocal indicators of physical size in humans are weak even when sex and age are controlled for (González, 2004, 2006; Rendall et al., 2005).

As noted in Chapter 2, the possible theoretical implications of this work must be considered with extreme caution. Human vocal and brain tissues do not fossilize, which makes it difficult to discern when and why changes in human vocal anatomy took place (Fitch, 2000a). It is equally difficult to determine whether vocal indicators of body size are a direct product of selection on the vocal apparatus for the purpose of communicating size or simply a byproduct, tracking variation in the size of a vocal apparatus that was shaped by evolution for another purpose (Dawkins & Krebs, 1978; Ghazanfar & Rendall, 2008; Lieberman, 1968). Moreover, selection may operate on the voice to convey other socially relevant information in ways that may obscure vocal size communication. Selection may, for instance, operate on vocal indicators of masculinity (e.g., via testosterone's effect on vocal fold mass and pitch: Evans et al., 2008) or on indexical properties of the voice that allow for individual recognition (e.g., individual variation in formants: Owren, 1996).

The results of Chapter 2 indicate that sample size can account for the majority of the variation in results reported by past studies investigating relationships between the voice and body. This finding has important practical implications. According to power analyses, future studies need to use samples of no fewer than 99 men and 164 women to estimate height from formants, or 618 men and 2140 women to estimate height from fundamental frequency, with *statistical* confidence. To date, only one study has employed an appropriately large sample size to estimate size from formants (although not large enough to estimate size from fundamental frequency; n=176 men; Puts et al., 2012a), whereas several studies have utilized samples of only 15 adults or fewer (Graddol & Swann, 1983; Lass & Brown, 1978; van Dommelen & Moxness, 1995).

Cross-cultural research is important for a comprehensive understanding of human behaviour (Henrich, Heine, & Norenzayan, 2010; Sue, 1999). In Chapter 2, I examined vocal correlates of body size in studies with samples from a range of cultures. The meta-analysis included 34 independent samples of adults from the countries Argentina, Bolivia, Canada, England, Germany, Lebanon, Poland, Scotland, Spain, Tanzania, and the United States of America. Chapter 2 revealed that relationships between fundamental frequency and physical height or weight were similar across these regions. The major implication of this finding is that vocal indicators of size in humans may be largely unaffected by ecological, environmental, and social factors. However, more research is needed in a wider range of cultures and geographic regions to test this prediction. For example, future work may further examine the reliability of vocal indicators of body size in populations that are unusually short (e.g., pygmies of Central Africa whose average heights typically range from 130 to 145 cm; Perry & Dominy, 2009) or unusually tall (e.g., see Collins 2000 for a study on men and women in the Netherlands, who measure on average 185 and 171 cm tall, respectively; Schönbeck et al., 2012).

Future studies may further examine the relative roles of fundamental and formants frequencies in communicating size across the lifespan. To date, corroborating the results of Chapter 2, a couple of studies indicate a significant negative relationship between formants and body size in children and adolescents controlling for sex and age (ages 4-16: Perry, Ohde, & Ashmead, 2001), but no reliable relationship between fundamental frequency and body size in children controlling for sex and age (ages 5-11: Glaze, Bless, Milenkovic, & Susser, 1988). However, Hollien (1994) reported a significant negative relationship between fundamental frequency and body size in boys during, but not before or after, puberty.

Future studies may also examine relationships among features of the voice other than fundamental frequency and formants, and indices of size other than height or weight, that to date have been examined in only a small number of studies with mixed results

(Bruckert et al., 2006; Collins, 2000; Collins & Missing, 2003; Evans, Neave, & Wakelin, 2006; González, 2007; Hamdan et al., 2012; Hollien & Jackson, 1973).

Listeners Use Both Formants and Pitch to Accurately Estimate Body Size

In Chapter 2, I established that fundamental frequency does not strongly predict the physical height or weight of men and women. However, the results of Chapter 3 demonstrate that listeners nevertheless utilize both pitch and formants to estimate body size. Although this finding is not novel (Bruckert et al., 2006; Collins, 2000; Feinberg et al., 2005; Pisanski et al., 2012; Pisanski & Rendall, 2011; Smith & Patterson, 2005; van Dommelen & Moxness, 1995), Chapter 3 presents the first evidence that pitch does not lower the accuracy of listeners' body size estimates.

Analogous to previous studies (Oliver & González, 2004; Rendall et al., 2007), listeners in Chapter 3 correctly identified the taller of two men approximately 60% of the time from regular modal speech. Listeners could also assess the relative size of men from natural whispered speech and synthetic three-formant sine-wave speech on a proportion of trials significantly greater than what would be expected by chance alone, despite the absence of voice pitch. Thus, Chapter 3 illustrates that listeners can accurately estimate men's relative body size from the voice in the presence and absence of voice pitch, as well as from minimal formant information.

Critically, the first of three experiments in Chapter 3 showed that listeners were significantly more accurate in assessments of men's relative body size with modal speech that contained fundamental frequency/pitch, than with whispered and sine-wave speech that did not contain pitch. This was true despite the absence of a physical relationship between fundamental frequency and height in the male voice sample. The second experiment confirmed that this effect was not due to a degraded whispered speech signal because the *perceptual* association between low formants and large size was unaffected by whispering. The third experiment showed that the facilitating effect of pitch on size estimation was, at least in part, related to harmonic density. High harmonic density may increase the salience of formants and allow listeners to more readily extract size-related information from the voice.

Chapter 3 demonstrates that voice pitch facilitates rather than confounds the accuracy of listeners' size estimates. However, the harmonic density effect could only explain a portion of the facilitating role of voice pitch in body size estimation. Moreover, as illustrated in Chapter 4, perceptual pitch biases may still lead to inaccurate estimation of body size in some contexts. Additional justifications for the perceptual association between pitch and body size are considered below (see, e.g., Conclusion).

Many studies that have examined voice-based size estimation have used a cross-sex design. These studies have typically examined women's perceptions of men's size (Collins, 2000; Evans et al., 2006; Feinberg et al., 2005). Likewise in Chapter 3, Experiments 1 and 3, voices were recorded from males and all listeners were female. Although some voice perception studies report no effect of sex (González, 2003; Pisanski & Rendall, 2011; Re, O'Connor, Bennett, & Feinberg, 2012; Rendall et al., 2007; see also Chapter 3, Experiment 2), other studies have found that men are better than women at assessing size from the voice (Charlton et al., 2013a; van Dommelen & Moxness, 1995) or that listeners of both sexes are more accurate in estimating the size of men compared to women (Rendall et al., 2007; van Dommelen & Moxness, 1995). Future studies should use vocalizers and listeners of both sexes to further test the possibility of sex-specificity in accurate size estimation.

Cross-Modal Perceptual Biases Influence Size Estimation

Many researchers have suggested that pitch is linked to size via a general perceptual bias (Grassi, 2005; Hinton, Nichols, & Ohala, 1994; Morton, 1977; Ohala, 1983, 1984; Pitcher et al., 2013; Rendall et al., 2007; Ultan, 1978; Walker et al., 2010). This type of association is also referred to as sound-size symbolism (Ultan, 1978). The association between pitch and size does appear general. In many cultures, low-pitched sounds are associated with largeness not only in body size perception (Ohala, 1983, 1984), but also in the perception of the size of shapes (Marks, 2004), inanimate objects (Grassi, 2005; Marks, 2004), and musical scales (Huron et al., 2006). In reality, as illustrated in Chapters 2 and 3, pitch and body size are not reliably related at the within-sex level. Therefore, some researchers have suggested that the perceptual association between pitch and size may explain why listeners associate low-pitched voices with large physical body size (González, 2006; Pisanski et al., 2012; Rendall et al., 2007).

In addition to large size, low pitch is perceptually associated with low spatial orientation in many (Eitan & Timmers, 2010; Evans & Treisman, 2010; Mudd, 1963; Parise et al., 2014) although not all cultures (Ashley, 2004). Here as well, researchers have suggested that the association between low pitch and low elevation is general and functional (Parise et al., 2014; Stumpf, 1883; Walker et al., 2010). As illustrated in Chapter 4, Figure 1, no previous study has tested whether pitch influences the perception of size in space.

The goal of Chapter 4 was to investigate whether voice pitch affects the ability of listeners to correctly use vertical spatial cues to gauge body size. The results of this work indicated that in the absence of conflicting pitch information, listeners were more likely to associate high versus low elevation with large size, corroborating past work (Eitan & Timmers, 2010). However, listeners associated low elevation with large size for low-pitched voices only. This effect may be due to the strong perceptual bias linking low pitch to low elevation and large size (see Chapter 4, Figure 1). Listeners associated right spatial locations with large body size irrespective of voice pitch.

Chapter 4 adds to a large body of literature on cross-modal correspondences (for reviews, see: Marks, 2000; Seilheimer, Rosenberg, & Angelaki, 2014; Spence, 2011; Spence & Deroy, 2013). However, this work presents the first evidence that pitch affects size perception in vertical space and the first evidence of an association between perceived largeness and horizontal spatial location. The key implication of the results of Chapter 4 is that pitch biases may lead to errors in size estimation in some contexts. Thus, while some researchers have suggested that associations between pitch and size or spatial location may be functionally adaptive (Morton, 1977; Parise et al., 2014; Peña et al., 2011; Stumpf, 1883; Walker et al., 2010), for instance facilitating sound localization in natural scenes (Parise et al., 2014), the results of Chapter 4 indicate that the mapping of low pitch to low elevation is not likely to facilitate accurate body size estimation.

Some studies report an effect of musical training on cross-modal pitch correspondences (see, e.g., Stewart, Walsh, & Frith, 2004; Walker, 1987). Future studies might examine whether musicians show a stronger association between low elevation and large size for low-pitched voices for than do non-musicians, and whether pianists associate the left rather than the right with largeness. Future studies may also examine whether low-pitched voices are percieved as larger at lower elevations across cultures. For example, comparisons might be made between cultures that typically associate low pitch with lowness in space and large size (e.g., Canada) and cultures that associate low pitch with highness in space and small size (e.g., the Manza of Central Africa; see Ashley, 2004 for review). Comparisons between musicians and non-musicians or across cultures will help to clarify the degree to which cross-modal pitch biases are socially acquired or hard-wired at birth (see Seilheimer et al., 2014 and Spence and Deroy, 2013 for recent debates on this topic). Finally, a natural extension of Chapter 4 will be to examine whether cross-modal incongruency between the *visual* and acoustic domains causes errors in size estimation.

Conclusion

The main findings from each empirical chapter of this thesis are, Chapter 2: formants predict variation in physical height and weight among both men and women but fundamental frequency is only weakly related to height among men and weight among women; Chapter 3: despite the lack of a robust physical relationship between fundamental frequency and size, voice pitch facilitates accurate body size assessment by providing a carrier signal for formants; Chapter 4: perceptual biases linking low-pitched voices to largeness can interact with spatial information to influence and potentially confuse vocal estimates of body size.

A key goal of this thesis was to gain insight into the paradoxical perceptual association between low voice pitch and large body size within sexes. My results suggest that there may be several non-mutually-exclusive explanations for this association. Chapter 2 suggests that the association may reflect a direct albeit very weak relationship between fundamental frequency and height in large samples of men or fundamental frequency and weight in large samples of women. Chapter 3 suggests that the association may reflect an indirect relationship between pitch and size, wherein low-pitched voices enhance the salience of formant-based indicators of body size. Finally, Chapter 4 suggests that the association is also likely to reflect a cross-modal perceptual bias that links low frequency sounds to perceptions of largeness and low elevation in a wide array of contexts.

This thesis examined vocal communication of body size in humans. Body size is a strong candidate for investigations of vocal social indicators namely because size can influence social and reproductive outcomes in many species (Peters, 1986), and in many traditional and modern human cultures (see, e.g., Courtiol, Raymond, Godelle & Ferdy, 2010; Judge & Cable, 2004; Monden & Smits, 2008; Yancey & Emerson, 2014; but see also Sear, 2011; Sear & Marlowe, 2009). However, research in the past twenty years has uncovered a complex web of information in nonverbal features of the human voice. The human voice can convey emotion and meaning (Gobl & Ni Chasaide, 2003), but far beyond this, the voice can provide reliable information about a variety of physical, endocrinological, and psychosocial qualities that may be evolutionarily relevant (for reviews, see: Feinberg, 2008; Kreiman & Sidtis, 2011; Owren, 2012; Pisanski & Feinberg, 2013). A person's voice can also have real societal consequences, affecting, for example, whether that person is hired for a job (Anderson, Klofstad, Mayew, & Venkatachalam, 2014; DeGroot & Motowidlo, 1999) or earns votes as a political leader (Tigue, Borak, O'Connor, Schandl, & Feinberg, 2012). It is now becoming increasingly evident that the human voice functions not only as the carrier of language, but also a useful social tool.

References

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice, 13*(3), 424-446.

Anderson, R. C., Klofstad, C. A., Mayew, W. J., & Venkatachalam, M. (2014). Vocal fry may undermine the success of young women in the labor market. *PLoS ONE, 9*(5), e97506.

Andersson, M. (1994). *Sexual selection*. New Jersey: Princeton University Press.

Ashley, R. (2004). Musical pitch space across modalities: Spatial and other mappings through language and culture. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.) *Proceedings of the 8th international conference on music perception and cognition* (pp. 64-71). Adelaide, Austrailia: Causal Productions.

August, P. V., & Anderson, J. G. T. (1987). Mammal sounds and motivation-structural rules: a test of the hypothesis. *Journal of Mammology, 68*, 1-9.

Bachorowski, J. A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America, 106*(2), 1054-1063.

Bisazza, A., & Marconato, A. (1988). Female mate choice, male-male competition and parental care in the river bullhead. *Cuttos gobio L.* (Pisces, Cottidae). *Animal Behaviour, 36*, 1352-1360.

Bro-Jørgensen, J. (2010). Dynamics of multiple signalling systems: Animal communication in a world in flux. *Trends in Ecology & Evolution, 25*(5), 292-300.

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 B: Biological Sciences, 273*(1582), 83-89.

Bryant, G. A., & Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biology Letters, 5*(1), 12-15.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11*(12), 535-543.

Carleton, K. L., Spady, T. C., & Kocher, T. D. (2005). Visual communication in East African cichlid fishes: Diversity in a phylogenetic context. . In F. Ladich, S. P. Collin, P. Moller & B. G. Kapoor (Eds.), *Communication in fishes* (pp. 481-511). Enfield, NH: Science Publishers Inc.

Charlton, B. D., Ellis, W. A. H., McKinnon, A. J., Cowin, G. J., Brumm, J., Nilsson, K., et al. (2011). Cues to body size in the formant spacing of male koala (*Phascolarctos cinereus*) bellows: Honesty in an exaggerated trait. *Journal of Experimental Biology, 214*(20), 3414-3422.

Charlton, B. D., Reby, D., & McComb, K. (2007). Female perception of size-related formant shifts in red deer, *Cervus elaphus*. *Animal Behaviour, 74*, 707-714.

Charlton, B. D., Reby, D., & McComb, K. (2008). Effect of combined source (*F*0) and filter (formant) variation on red deer hind responses to male roars. *Journal of the Acoustical Society of America, 123*(5), 2936-2943.

Charlton, B. D., Taylor, A. M., & Reby, D. (2013a). Are men better than women at acoustic size judgements? *Biology Letters, 9*(4).

Charlton, B. D., Whisson, D. A., & Reby, D. (2013b). Free-ranging male koalas use size-related variation in formant frequencies to assess rival males. *PLoS ONE, 8*(7), e70279.

Clutton-Brock, T. H., & Albon, S. D. (1979). The roaring of red deer and the evolution of honest advertisement. *Behaviour*, 145-170.

Cohen, J. R., Crystal, T. H., House, A. S., & Neuburg, E. P. (1980). Weighty voices and shaky evidence: A critique. *Journal of the Acoustical Society of America, 68*, 1884-1886.

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour, 60*, 773-780.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65*, 997-1004.

Cooper, W. E., & Vitt, L. J. (1993). Female mate choice of large male broad-headed skinks. *Animal Behaviour, 45*, 683-693.

Courtiol, A., Raymond, M., Godelle, B., & Ferdy, J. (2010). Mate choice and human stature: Homogamy as a unified framework for understanding mating preferences. *Evolution, 64*, 2189-2203.

Dabbs, J., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences, 27*(4), 801-804.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London, UK: John

Murray.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: information or manipulation. *Behavioural Ecology: An Evolutionary Approach, 2*, 282-309.

DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology, 84*(6), 986.

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition, 114*(3), 405-422.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 6.

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.

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.

Ey, E., Pfefferle, D., & Fischer, J. (2007). Do age-and sex-related variations reliably reflect body size in non-human primate vocalizations? A review. *Primates, 48*(4), 253-267.

Fairbairn, D. J., Blanckenhorn, W. U., & Székely, T. (2007). *Sex, size, and gender roles:*

evolutionary studies of sexual size dimorphism. London, UK: Oxford University Press.

Fant, F. (1960). *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.

Feinberg, D. R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary Anthropology, 17*(2), 112-118.

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*, 561-568.

Feingold, A. (1982). Do taller men have prettier girlfriends? *Psychological Reports, 50*, 810-815.

Fink, B., Neave, N., Brewer, G., & Pawlowski, B. (2007). Variable preferences for sexual dimorphism in stature (SDS): Further evidence for an adjustment in relation to own height. *Personality and Individual Differences, 43*(8), 2249-2257.

Fitch, W. (1999). Acoustic exaggeration of size in birds via tracheal elongation: comparative and theoretical analyses. *Journal of Zoology, 248*(1), 31-48.

Fitch, W.T. (1994). *Vocal tract length perception and the evolution of language* (Doctoral thesis). Providence, RI: Brown University.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America, 102*(2 Pt 1), 1213-1222.

Fitch, W. T. (2000a). The evolution of speech: a comparative review. *Trends in Cognitive Sciences, 4*(7), 258-267.

Fitch, W. T. (2000b). Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry. *Zoology, 103*, 40-58.

Fitch, W. T., & Fritz, J. B. (2006). Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *Journal of the Acoustical Society of America, 120*(4), 2132-2141.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America, 106*(3), 1511-1522.

Fitch, W. T., & Hauser, M. (2003). Unpacking "honesty": Vertebrate vocal production and the evolution of acoustic signals. *Acoustic Communication, 16*, 65-137.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human, *268*(1477), 1669-1675.

Ghazanfar, A. A., & Rendall, D. (2008). Evolution of human vocal production. *Current Biology, 18*(11), R457-R460.

Glaze, L. E., Bless, D. M., Milenkovic, P., & Susser, R. D. (1988). Acoustic characteristics of children's voice. *Journal of Voice, 2*(4), 312-319.

Gobl, C., & Ni Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication, 40*(1), 189-212.

González, J. (2003). Estimation of speakers' weight and height from speech: a re-analysis of data from multiple studies by Lass and colleagues. *Perceptual and Motor Skills, 96*(1), 297-304.

González, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics, 32*(2), 277-287.

González, J. (2006). Research in acoustics of human speech sounds: Correlates and perception of speaker body size. *Recent Research Development in Applied Physics, 9*, 1-15.

González, J. (2007). Correlations between speakers' body size and acoustic parameters of voice. *Perceptual and Motor Skills, 105*(1), 215.

Graddol, D., & Swann, J. (1983). Speaking fundamental frequency: Some physical and social correlates. *Language and Speech, 26*(4), 351-366.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics, 67*(2), 274-284.

Greisbach, R. (1999). Estimation of speaker height from formant frequencies. *Forensic Linguistics, 6*(2), 265-277.

Hamdan, A. L., Al-Barazi, R., Tabri, D., Saade, R., Kutkut, I., Sinno, S., Nassar, J. (2012). Relationship between acoustic parameters and body mass analysis in young males. *Journal of Voice, 26*(2), 144-147.

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. *Journal of Laryngology and Otology, 112*(5), 451-454.

Harris, T. R., Fitch, W. T., Goldstein, L. M., & Fashing, P. J. (2006). Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass. *Ethology, 112*(9), 911-920.

Hasson, O. (1991). Sexual displays as amplifiers: Practical examples with an emphasis on feather decorations. *Behavioural Ecology, 2*(3), 189-197.

Hauser, M. D. (1993). The evolution of nonhuman primate vocalizations: Effects of phylogeny, body weight, and social context. *American Naturalist*, 528-542.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences, 33*(2-3), 111-135.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099-3111.

Hinton, L., Nichols, J., & Ohala, J. J. (1994). *Sound symbolism*: Cambridge, MA: Cambridge University Press.

Holden, C., & Mace, R. (1999). Sexual dimorphism in stature and women's work: A phylogenetic cross-cultural analysis. *American Journal of Physical Anthropology, 110*(1), 27-45.

Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males *Journal of the Acoustical Society of America, 96*(5), 2646-2654.

Hollien, H., & Jackson, B. (1973). Normative data on the speaking fundamental frequency characteristics of young adult males. *Journal of Phonetics, 1*(2), 117-

120.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279*(5354), 1213-1216.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *Journal of the Acoustical Society of America, 106*(3), 1532-1542.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *1*(3), 170-177.

Irino, T., Aoki, Y., Kawahara, H., & Patterson, R. D. (2012). Comparison of performance with voiced and whispered speech in word recognition and mean-formant-frequency discrimination *Speech Communication 54*, 998-1013.

Ives, D. T., Smith, D. R., & Patterson, R. D. (2005). Discrimination of speaker size from syllable phrases. *Journal of the Acoustical Society of America, 118*, 3816-3822.

Judge, T. A., & Cable, D. M. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *Journal of Applied Psychology, 89*(3), 428-441.

Kahane, J. C. (1978). A morphological study of the human prepubertal and pubertal larynx. *American Journal of Anatomy, 151*(1), 11-19.

Kewley-Port, D., Li, X., Zheng, Y., & Neel, A. T. (1996). Fundamental frequency effects on thresholds for vowel formant discrimination. *Journal of the Acoustical Society of America, 100*(4), 2462-2470.

Kirchhübel, C., Howard, D. M., & Stedmon, A. W. (2011). Acoustic correlates of speech when under stress: Research, methods and future directions. *International Journal of Speech, Language & the Law, 18*(1), 75-98.

Kreiman, J., & Sidtis, D. V. L. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*: West Sussex, UK: Wiley-Blackwell.

Kurzban, R., & Weeden, J. (2005). HurryDate: Mate preferences in action. *Evolution and Human Behavior, 26*, 227-244.

Lass, N. J., Barry, P. J., Reed, R. A., Walsh, J. M., & Amuso, T. A. (1979). The effect of temporal speech alterations on speaker height and weight identification. *Language and Speech, 22*(2), 163-171.

Lass, N. J., & Brown, W. S. (1978). Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies. *Journal of the Acoustical Society of America, 63*(4), 1218-1220.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America, 59*(3), 675-678.

Lass, N. J., Kelley D. T, Cunningham, C. M., & Sheridan, K. J. (1980). A comparative study of speaker height and weight identification from voiced and whispered speech. *Journal of Phonetics, 8*(2), 195-204.

Lee, R. B. (1979). *The !Kung San: Men, women, and work in a foraging society*.

Cambridge, MA: Cambridge University Press.

Leigh, S. R., & Shea, B. T. (1995). Ontogeny and the evolution of adult body size dimorphism in apes. *Journal of Primatology, 36*, 37-60.

Lieberman, D., McCarthy, R., Hiiemae, K., & Palmer, J. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology, 46*(2), 117-128.

Lieberman, P. (1968). Primate vocalizations and human linguistic ability. *Journal of the Acoustical Society of America, 44*(6), 1574-1584.

Lieberman, P., & Blumstein, S.E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge, MA: Cambridge University Press.

Lindenfors, P., Gittleman, J. L., & Jones, K. E. (2007). Sexual size dimorphism in mammals. In D. J. Fairbairn;, W. U. Blanckenhorn; & T. Szekely (Eds.), *Sex, size and gender roles: evolutionary studies of sexual size dimorphism* (pp. 19-26). Oxford, UK: Oxford University Press.

Marks, L. E. (2000). Synesthesia. In E. Cardeña, S. J. Lynn & S. Krippner (Eds.), *Varieties of anomalous experience: examining the scientific evidence* (pp. 121-149). Washington, DC: American Psychological Association.

Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 85-105). Cambridge, MA: MIT Press.

Marler, P., & Hobbett, L. (1975). Individuality in a long-range vocalization of wild chimpanzees. *Zeitschrift für Tierpsychologie, 38*(1), 97-109.

Marler, P., & Tenaza, R. (1977). Signaling behavior of apes with special reference to vocalization. In T. A. Sebeok (Ed.), *How animals communicate* (pp. 965-1033). Bloomington, IN: Inidiana University Press.

Marlowe, F. W. (2004). Mate preferences among Hadza hunter-gatherers. *Human Nature, 15*, 365-376.

Maynard Smith, J., & Harper, D. (2003). *Animal signals.* Oxford, UK: Oxford University Press.

McComb, K., Reby, D., Baker, L., Moss, C., & Sayialel, S. (2003). Long-distance communication of acoustic cues to social identity in African elephants. *Animal Behaviour, 65*(2), 317-329.

Monden, C. W. S., & Smits, J. (2008). Maternal height and child mortality in 42 developing countries. *American Journal of Human Biology, 21*, 305-311.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist, 111*, 855-869.

Mudd, S. (1963). Spatial stereotypes of four dimensions of pure tone. *Journal of Experimental Psychology, 66*(4), 347.

Nettle, D. (2001). Height and reproductive success in a cohort of British men. *Human Nature, 13*, 473-491.

Ohala, J. J. (1983). Cross-language use of pitch: an ethological view. *Phonetica, 40*(1), 1-18.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of

*F*0 of voice. *Phonetica, 41*(1), 1-16.

Oliver, J. C., & González, J. (2004). Percepción a través de la voz de las características físicas del hablante: identificación de la estatura a partir de una frase o una vocal. *Revista de Psicología General y Aplicada, 57*(1), 21-34.

Owren, M. J. (1996). An "acoustic-signature" model of vowel evolution. *Journal of the Acoustical Society of America, 99*(4), 2558-2574.

Owren, M. J. (2012). Human voice in evolutionary perspective. *Acoustics Today, 7*(4), 24-34.

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences, 111*(16), 6104-6108.

Pawlowski, B., Dunbar, R. I., & Lipowicz, A. (2000). Tall men have more reproductive success. *Nature, 403*(6766), 156.

Pawlowski, B., & Koziel, S. (2002). The impact of traits offered in personal advertisements on response rates. *Evolution and Human Behavior, 23*, 139-149.

Peck, A. M., & Vågerö, D. H. (1987). Adult body height and childhood socioeconomic group in the Swedish population. *Journal of Epidemiology and Community Health, 41*, 333-337.

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science, 22*(11), 1419-1421.

Perry, G. H., & Dominy, N. J. (2009). Evolution of the human pygmy phenotype. *Trends* *in Ecology & Evolution, 24*(4), 218-225.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America, 109*(6), 2988-2998.

Peters, R. H. (1986). *The ecological implications of body size* (Vol. 2). Cambridge, MA: Cambridge University Press.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175-184.

Phillips, D. P., Scovil, S. J., Carmichael, M. E., & Hall, S. E. (2007). Adaptation of central pitch-specific mechanisms. *Perception, 36*, 918-930.

Pisanski, K., & Feinberg, D. R. (2013). Cross-cultural variation in mate preferences for averageness, symmetry, body size, and masculinity. *Cross-Cultural Research, 47*(2), 162-197.

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.

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

Pitcher, B. J., Mesoudi, A., & McElligott, A. G. (2013). Sex-biased sound symbolism in

English-language first names. *PLoS ONE, 8*(6), e64825.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology, 13*(3), 278.

Puts, D. A. (2010). Beauty and the beast: mechanisms of sexual selection in humans. *Evolution and Human Behaviour, 31*, 157-175.

Puts, D. A., Apicella, C. L., & Cardenas, R. A. (2012a). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 279*(1728), 601-609.

Puts, D. A., Hodges, C. R., Cardenas, 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.

Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012b). Sexual selection on human faces and voices. *Journal of Sex Research, 49*(2-3), 227-243.

Re, D. E., O'Connor, J. J. M., Bennett, P. J., & Feinberg, D. R. (2012). Preferences for very low and very high voice pitch in humans. *PloS ONE, 7*(3), e32719.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour, 65*, 519-530.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences,* *272*(1566), 941-947.

Reddon, A. R., Voisin, M. R., Menon, N., Marsh-Rollo, S. E., Wong, M. Y., & Balshine, S. (2011). Rules of engagement for resource contests in a social fish. *Animal Behaviour, 82*(1), 93-99.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F-0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America, 117*(2), 944-955.

Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour, 51*(5), 1007-1015.

Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1208-1219.

Ryan, M. J. (1988). Constraints and patterns in the evolution of anuran acoustic communication In B. Fritzsch (Ed.), *The evolution of the amphibian auditory system* (pp. 637-677). Hoboken, NJ: John Wiley & Sons, Inc.

Saez, S., & Sakai, F. O. (1976). Androgen receptors in human pharyngo-laryngeal mucosa and pharyngo-laryngeal epithelioma. *Journal of Steroid Biochemistry, 7*(11), 919-921.

Salska, I., Frederick, D. A., Pawlowski, B., Reilly, A. H., Laird, K. T., & Rudd, N. A. (2008). Conditional mate preferences: Factors influencing preferences for height.

Personality and Individual Differences, 44, 203-215.

Sanvito, S., Galimberti, F., & Miller, E. H. (2007). Vocal signalling of male southern elephant seals is honest but imprecise. *Animal Behaviour, 73*, 287-299.

Sawashima, M., Hirose, H., Honda, K., Yoshioka, H., Hibi, S., Kawase, N., Yamada, M. (1983). Stereoendoscopic measurement of the laryngeal structure. In D. Bless & G. Abbs (Eds.), *Vocal fold physiology: Contemporary research & clinical issues* (pp. 264-276). New York, NY: College-Hill.

Schönbeck, Y., Talma, H., van Dommelen, P., Bakker, B., Buitendijk, S. E., HiraSing, R. A., van Buuren, S. (2012). The world's tallest nation has stopped growing taller: The height of Dutch children from 1955 to 2009. *Pediatric research, 73*, 371-377.

Sear, R. (2011). Height and reproductive success: Is bigger always better? In C. Störmer & K. P. Willführ (Eds.), *Homo Novus—A human without illusions* (pp. 127-143). New York, NY: Springer

Sear, R., & Marlowe, F. (2009). How universal are human mate choices? Size does not matter when Hadza foragers are choosing a mate. *Biology Letters*, 5, 606-609.

Seilheimer, R. L., Rosenberg, A., & Angelaki, D. E. (2014). Models and processes of multisensory cue combination. *Current Opinion in Neurobiology, 25*, 38-46.

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 B: Biological Sciences, 277*(1699), 3509-3518.

Sherman, P. W. (1977). Nepotism and the evolution of alarm calls. *Science, 197*(4310), 1246-1253.

Simmons, A. M., Popper, A. N., Fay, R. R., & Gerhardt, H. C. (2003). *Acoustic communication* (Vol. 16): New York, NY: Springer.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America, 118*(5), 3177-3186.

Smith, D. R., Patterson, R. D., & Turner, R. E. (2005). The processing and perception of size information in speech sounds. *Journal of the Acoustical Society of America, 117*(1), 305-318.

Smith, G. D., Hart, C., Upton, M., Hole, D., Gillis, C., Watt, G., Hawthorne, V. (2000). Height and risk of death among men and women: Aetiological implications of associations with cardiorespiratory disease and cancer mortality. *Journal of Epidemiology and Community Health, 54*(2), 97-103.

Sommers, M. S., Moody, D. B., Prosen, C. A., & Stebbins, W. C. (1992). Formant frequency discrimination by Japanese macaques (*Macacafuscata*). *Journal of the Acoustical Society of America, 91*(6), 3499-3510.

Sorokowski, P., & Butovskaya, M. (2012). Height preferences in humans may not be universal: Evidence from the Datoga people of Tanzania. *Body Image, 9*(4), 510-516.

Sorokowski, P., Sorokowska, A., Fink, B., & Mberira, M. (2012). Variable preferences for sexual dimorphism in stature (SDS) might not be universal: Data from a semi nomad population (Himba) in Namibia. *Journal of Cross-Cultural Psychology, 43*, 32-37.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971-995.

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition, 22*(1), 245-260.

Stearns, S. C. (1976). Life-history tactics: A review of the ideas. *Quarterly Review of Biology, 51*, 3-47.

Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception & Psychophysics, 66*(2), 183-195.

Stumpf, C. (1883). *Tonpsychologie (Vol. 1)*. Leipzig, Germany: Hirzel.

Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist, 54*(12), 1070.

Swami, V., Furnham, A., Balakumar, N., Williams, C., Canaway, K., & Stanistreet, D. (2008). Factors influencing preferences for height: A replication and extension. *Personality and Individual Differences, 45*(5), 395-400.

Szklarska, A., Koziel, S., Bielicki, T., & Malina, R. M. (2007). Influence of height on attained level of education in males at 19 years of age. *Journal of Biosocial Science, 39*, 575-582.

Taylor, A., Reby, D., & McComb, K. (2010). Size communication in domestic dog, *Canis familiaris*, growls. *Animal Behaviour, 79*(1), 205-210.

Taylor, A. M., & Reby, D. (2010). The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology, 280*(3), 221-236.

Teasdale, T. W., Owen, D. R., & SØrensen, T. I. (1991). Intelligence and education levels in adult males at the extreme of stature. *Human Biology, 63*, 19-30.

Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior, 33*(3), 210-216.

Titze, I. R. (1989). Physiological and acoustic differences between male and female voices. *Journal of the Acoustical Society of America, 85*, 1699-1707.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory. *Journal of the Acoustical Society of America, 123*(5), 2733-2749.

Turner, R. E., Walters, T. C., Monaghan, J. J., & Patterson, R. D. (2009). A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. *Journal of the Acoustical Society of America, 125*, 2374-2386.

Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Eds.), *Universals of human language Vol. 2*: *Phonology* (pp. 527-568). Stanford, CA: Stanford University Press.

van Dommelen, W. A., & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and Speech, 38*(3), 267-287.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices. *Journal of Phonetics, 13*(1), 19-38.

Vannoni, E., & McElligott, A. G. (2008). Low frequency groans indicate larger and more dominant fallow deer (*Dama dama*) males. *PloS one, 3*(9), e3113.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21-25.

Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Perception & Psychophysics, 42*(5), 491-502.

Weckerly, F. W. (1998). Sexual-size dimorphism: influence of mass and mating systems in the most dimorphic mammals. *Journal of Mammalogy*, 33-52.

Wu, K., & Childers, D. G. (1991). Gender recognition from speech. Part I: Coarse analysis. *Journal of the Acoustical Society of America, 90*(4), 1828-1840.

Yancey, G., & Emerson, M. O. (2014). Does height matter? An examination of height preferences in romantic coupling. *Journal of Family Issues*. Retrieved from http://jfi.sagepub.com/content/early/2014/01/12/0192513X13519256

Zaadstra, B. M., Seidell, J. C., Van Noord, P., te Velde, E. R., Habbema, J., Vrieswijk, B., et al. (1993). Fat and female fecundity: prospective study of effect of body fat distribution on conception rates. *British Medical Journal, 306*(6876), 484-487.

APPENDIX

Appendix 1: Equations for Chapter 2

Formant-Based VTL Estimates

For all derivations of formant structure, n is the total number of formants measured (here, $n = 4$) and F_i is the frequency of the ith formant in Hertz (Hz). Average formant frequency (in Hz; Pisanski & Rendall, 2011) is given by

$$F_n = \frac{\sum_{i=1}^{n} F_i}{n} \qquad \text{(A1)}$$

Geometric mean formant frequency (in Hz; Smith & Patterson, 2005) is given by

$$MFF = \left(\prod_{i=1}^{n} F_i \right)^{1/n} \qquad \text{(A2)}$$

Formant dispersion (in Hz; Fitch, 1997) is given by

$$D_f = \frac{\sum_{i=1}^{n} (F_{i+1} - F_i)}{n-1} \qquad \text{(A3)}$$

Formant position (in Z(Hz); Puts et al., 2012a) is given by

$$P_f = \frac{\sum_{i=1}^{n} F'_i}{n} \qquad \text{(A4)}$$

where F'_i is the standardized *i*th formant. Formant spacing (in Hz; Reby & McComb, 2003) was estimated by calculating the best fit for equation (A5) to the mean frequency of each formant for each vocalizer, where

$$F_i = \frac{(2i-1)}{2}\Delta F \tag{A5}$$

and where *i* refers to the formant number. We did this by plotting mean formant frequencies for each individual vocalizer against (2*i* - 1)/2 increments of formant spacing, where ΔF is equal to the slope of the linear regression line with an intercept set to 0 (see Reby & McComb, 2003).

Apparent vocal tract length derived from mean formants (in cm; adapted from Fitch, 1997; see also Titze 1994) is given by

$$VTL(F_i) = \frac{\sum_{i=1}^{n}(2i-1)(c/4F_i)}{n} \tag{A6}$$

where *i* refers to the formant number and *c* is the speed of sound in a uniform tube with one end closed, i.e., the vocal tract (c = 33 500 cm/s).

Apparent vocal tract length derived from formant spacing (in cm; Reby & McComb, 2003) is given by

$$VTL(\Delta F) = \frac{c}{2\Delta F} \tag{A7}$$

where c = 35 000 cm/s, and ΔF was estimated using the method described above. Finally, using our previously unpublished data, we performed a confirmatory factor analysis (CFA) with a maximum-likelihood estimator to derive a formant-based latent variable. In each model the observed variables were T (the period of the wave for each of $F1$–$F4$, where $T_i = 1/F_i$), computed separately for each vocalizer and each vowel ('ah'/ɑ/, 'ee'/i/, 'e'/ɛ/, 'oh'/o/ and 'oo'/u/). A separate factor analysis was performed on each sex (N = 264 males, 326 females). The latent variable resulting from this model reflects the shared variance across all formants within each vowel, approximately 90% of which is related to vocal tract length, and thus produces a factor score for each vocalizer that corresponds to his or her height (Turner et al., 2009). Factor scores were averaged across the five vowels for each vocalizer.

Mean Weighted Correlations

We derived $\bar{r}$ as follows. For each of the 295 voice–size correlations, we first calculated a standardized effect size and its standard error using Fisher's r-to-z transformation, as given by

$$ES_Z = 0.5\ln\left(\frac{1+r}{1-r}\right) \tag{A8}$$

$$SE_{ES_Z} = \frac{1}{\sqrt{n-3}} \tag{A9}$$

where r is the original reported Pearson correlation coefficient and n is the size of the corresponding sample. Next, we weighted each effect size by an index of its sample size, giving more weight to effects derived from larger samples, and averaged across weighted effect sizes for each type of voice–size relationship.

The average weighted standardized effect size and its standard error are given by:

$$\overline{ES}_Z = \frac{\sum\left(w_i ES_{Z_i}\right)}{\sum w_i} \tag{A10}$$

$$SE_{\overline{ES}_Z} = \frac{1}{\sqrt{\sum w_i}} \tag{A11}$$

where w_i is the weight corresponding to the ith sample,

$$w_i = \frac{1}{SE_{ES_Z}{}^2} \tag{A12}$$

Finally, we transformed the average weighted effect sizes back to Pearson correlation coefficients using Fisher's z-to-r transformation to obtain mean weighted correlations, $\bar{r}$, and corresponding standard errors as given by

$$\bar{r} = \frac{e^{2\overline{ES}_Z} - 1}{e^{2\overline{ES}_Z} + 1} \tag{A13}$$

$$SE_{\bar{r}} = \frac{e^{2SE_{\overline{ES}_Z}} - 1}{e^{2SE_{\overline{ES}_Z}} + 1} \tag{A14}$$

Confidence Intervals For Mean Weighted Correlations

Lower and upper 95% confidence intervals for mean weighted correlations were first estimated from average standardized weighted effect sizes, as given by

$$\overline{ES}_{z(lower)} = \overline{ES} - Z_{(1-\alpha)}\left(SE_{\overline{ES}_Z}\right) \tag{A15}$$

$$\overline{ES}_{Z(upper)} = \overline{ES} + Z_{(1-\alpha)}\left(SE_{\overline{ES}_Z}\right) \tag{A16}$$

where $Z_{(1-\alpha)}$ = 1.96.

These confidence intervals were then transformed using Fisher's *z*-to-*r* transformation to correspond with $\bar{r}$, as given by

$$\bar{r}_{lower} = \frac{e^{2\overline{ES}_{Z\,lower}} - 1}{e^{2\overline{ES}_{Z\,lower}} + 1} \tag{A17}$$

$$\bar{r}_{upper} = \frac{e^{2\overline{ES}_{Z\,upper}} - 1}{e^{2\overline{ES}_{Z\,upper}} + 1} \tag{A18}$$

Meta-Regressions

In each meta-regression model, we treated the reported correlation coefficient for each type of voice–size relationship as the dependent variable, weighted by

$$\frac{1}{SE_{r_i}^{\ 2}} \tag{A19}$$

where r_i is the Pearson correlation reported by the ith sample, and

$$\mathrm{SE} = 1 \big/ \sqrt{n_i - 3} \tag{A20}$$

where n_i is the size of the corresponding ith sample.

Appendix 2: Supplementary Material for Chapter 2

Statistical Analysis of Meta-Data

Mean weighted correlations and meta-regressions.

The mean weighted correlation ($\bar{r}$) is a reliable index of a population-level relationship controlling for sample size (Lipsey & Wilson, 2001). We computed $\bar{r}$ for 26 types of voice–size relationships, for each sex independently and collapsing across sample sex, for a total of 78 $\bar{r}$ measures. For each $\bar{r}$, we computed its standard error, its significance against a table of critical values (alpha = .05, two tailed) and 95% confidence intervals. See equations A8–A18 in the Appendix for details. In addition, we performed a series of weighted least squares regressions (henceforth, meta-regressions). In each model, we treated the reported correlation coefficient as the dependent variable weighted by the inverse of its standard error squared (an index of sample size, see equations A19–A20). We included additional variables as fixed factors or covariates (see Results in main text for details).

Power analysis.

We performed statistical power analyses to determine the minimum sample sizes required to reliably obtain the mean weighted correlations that we derived, given an alpha of .05 (the probability of a false positive or type I error) and a conservative power level of .90 (the probability of a true positive), where beta is therefore .10 (the probability of a type II error; Cohen, 1988).

Funnel plots.

To assess the influence of sample size and to examine possible publication bias in our sample, we created funnel plots for the two most frequently reported voice–size relationships with the largest *k*: F_0–height and $F4$–height. The likelihood of detecting bias using this method is high (Egger, Smith, Schneider, & Minder, 1997). Voice–size relationships reported by studies with smaller sample sizes (lower on the y-axis) are expected to vary more than those reported by larger studies, resulting in more scatter at the base of the plot. Thus, where publication bias is absent, this expected pattern in the spread of effect sizes will result in a symmetrical, inverted funnel-shaped plot. In contrast, an asymmetrical funnel is suggestive of publication bias in the direction of the skew (Egger et al., 1997). Data points that fall outside of the confidence intervals are flagged as outliers and potential sources of bias.

We calculated the average weighted standardized effect size ($\overline{\mathrm{ES}}_z$, see equation A10) for the centre summary line using fixed effects meta-analysis (which is the preferred method; see, e.g., Sterne & Egger, 2001) and 95% confidence intervals for each voice–size relationship and each sex separately. The $\overline{\mathrm{ES}}_z$ for F_0–height relationships was substantially stronger for samples of men (-.13) than women (-.07); therefore, we created additional F_0–height funnel plots for each sex separately to avoid overestimating or underestimating potential skew in the distribution of the data. The average unweighted effect size of $F4$–height relationships did not differ substantially between the sexes (-.27); therefore, both sexes were plotted together.

To examine the influence of outliers, each outlier was removed sequentially beginning with the outlier with the smallest weight in each plot. Each removal was followed by recalculation of the $\overline{\text{ES}}_z$ and confidence intervals (remaining outliers always remained outside of the confidence intervals following this recalculation). The removal of both outliers in Fig. 1b, representing F4–height relationships in men and women, had no effect on the $\overline{\text{ES}}_z$ (before removal: -.277; after removal: -.273). The removal of all three outliers in Fig. 1c, representing F_0–height relationships in men, had no effect on the $\overline{\text{ES}}_z$ (before removal: -.127; after removal: -.129). Thus, because these outliers were distributed fairly symmetrically in each plot and/or carried little weight, removing these outliers did not change our average weighted effect sizes or confidence intervals.

Mean weighted voice and body size measures.

We derived weighted mean voice and body size measures for each sex by pooling values reported by previous work with values derived from our new samples and weighting the values by their corresponding sample size. We converted all available size measures to the metric scale (cm and kg) and retained all voice measures in the Hertz scale except for P_f, which is represented on a standardized scale of Hertz (Z(Hz)), and VTL(F_i) or VTL(ΔF), which are reported in centimetres. We then multiplied each measure by its corresponding sample size, summed across samples, and divided the summed value by the pooled sample size (*N*) for each given measure and sex. We did this separately for samples in which body size was measured directly or was self-reported to examine whether the two methods of measurement produced different mean height or weight values.

Analysis of Previously Unpublished Data

We used several new data sets of voice and body size measures of adult men and women to increase the pooled sample size of the meta-analysis. We measured men's and women's heights, weights, F_0 and formants.

Sample characteristics.

The data sets derived from eight samples (*k*) of vocalizers and a total of 700 (*N*) adults: four samples from Scotland (UK1: n = 43 men, 123 women; UK2: n = 34 men, 112 women), three from Canada (Canada1: n = 74 men, 118 women; Canada2: n = 111 men) and one from Germany (Germany: n = 85 women).

Voice measures.

Methods of voice measurement and calculation were identical for all eight samples of vocalizers and were performed in Praat (Boersma & Weenink, 2013). For all samples, voice recordings were of the five vowels, /ɑ/, /i/, /ɛ/, /o/ and /u/ (International Phonetic Alphabetic notation), except for the UK1 sample, for which the vowels were /eI/, /i/, /aI/, /o/ and /ju/. Measurements were taken from the steady-state portion of each of five isolated vowels per vocalizer, averaged within vocalizers, and then within sex and sample to obtain mean values. See main text for additional details.

VTL estimates.

From the mean $F1$–$F4$ values we computed Fn, MFF, D_f, P_f, ΔF, VTL(F_i), VTL(ΔF) and T (the period of the formant wave, for use in the confirmatory factor analysis, CFA). Equations and methods used to compute these respective VTL estimates are given in the Appendix (see equations A1–A7 in Appendix 1).

To date, the most widely used of these measures is D_f (Fitch, 1997), which calculates the mean distance between successive formants (see, e.g., Bruckert, Liénard, Lacroix, Kreutzer, & Leboucher, 2006; Collins, 2000; Evans, Neave & Wakelin, 2006; Feinberg, Jones, Little, Burt, & Perrett, 2005; Fitch, 1997; González, 2004; Puts, Apicella, & Cardenas, 2012). However, D_f has been criticized for mathematically omitting differences related to $F2$ and $F3$ (Pisanski, Mishra, & Rendall, 2012; Puts et al., 2012; Rendall, Kollias, Ney & Lloyd, 2005). Formant spacing, ΔF, is a measure of the mean distance between formants that is derived by relating mean formant frequencies to increments of formant spacing. This method has been used successfully to predict variation in body mass among red deer stags (Reby & McComb, 2003) as well as to investigate formant modulation in men and women (Cartei, Cowles, & Reby, 2012).

More recent formant measures used to estimate size in humans include mean formant frequency, Fn, an average of the lower formants (Pisanski & Rendall, 2011), geometric mean formant frequency, MFF (Irino, Aoki, Kawahara, & Patterson, 2012; Smith & Patterson, 2005) and formant position, P_f, an average of the lower formants standardized using between-sex values (Puts et al., 2012). The measures VTL(F_i) and VTL(ΔF) are estimates of vocal tract length derived from mean formant frequencies and formant spacing, respectively, that incorporate the speed of sound in the vocal tract[7] (Fitch, 1994, 1997; Reby & McComb, 2003). Finally, formant pattern latent models (here a confirmatory factor analysis, CFA) have been shown to control for vowel-based variation in formants as well as noise related to potential measurement error, where such error is present, and to produce factor scores that scale with vocal tract length and overall body size (Turner, Walters, Monaghan, & Patterson, 2009).

[7] Studies have used various speeds of sound (*c*) to estimate mammalian VTLs from formants, including 335 m/s (in macaques: Fitch, 1997), 340 m/s (in humans: see Patterson, Smith, van Dinther & Walters, 2008) and 350 m/s (in red deer: Reby & McComb, 2003). The speed of sound in air is around 335–345 m/s, but in the warm and damp mammalian vocal tract, the speed of sound is estimated to be slightly faster, about 350 m/s (Titze, 1994). Thus, 350 m/s is likely the most appropriate value to use when estimating VTL.

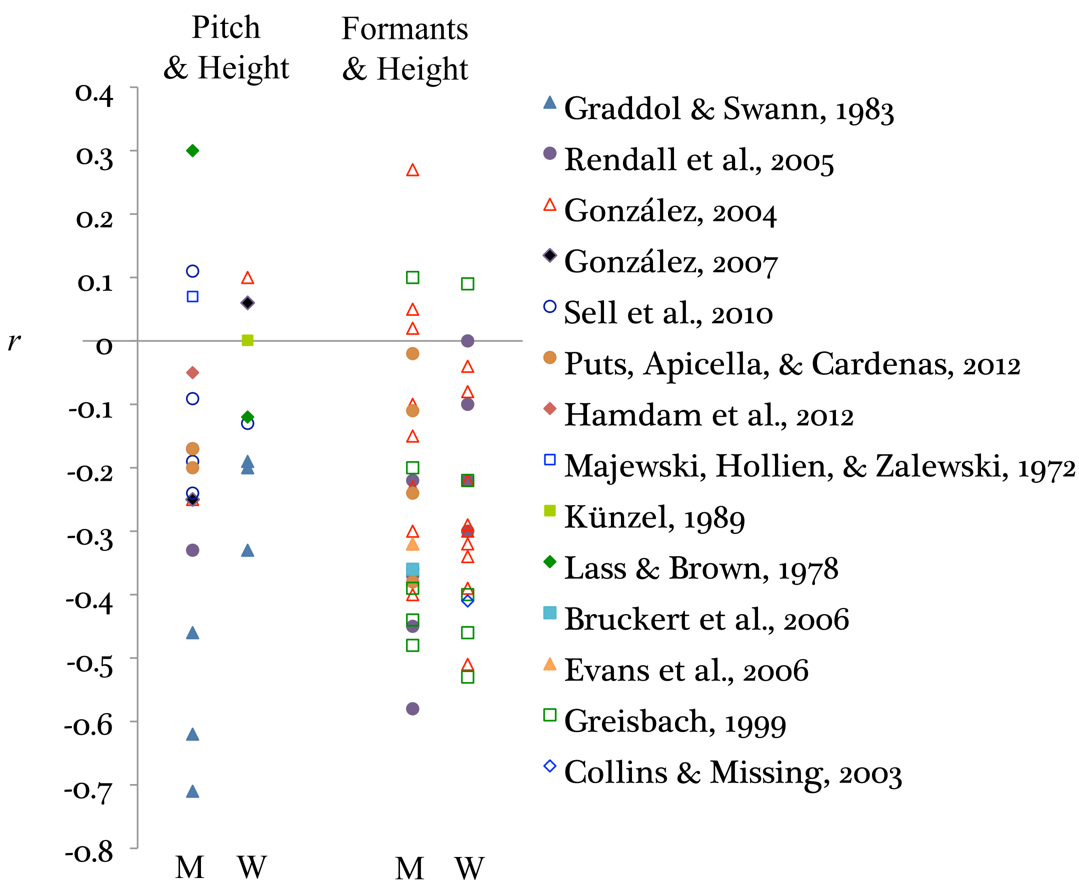

Figure A1. Bivariate Pearson correlations (r) between pitch (F_0) and height, or formants (any formant-based estimate of vocal tract length, VTL) and height, within each sex as reported by past studies (13 studies, k = 40 same-sex samples; M = samples of men, W = samples of women).

Note. A correlation of 0 indicates no relationship. The figure demonstrates the wide variation in the direction and strength of previously reported voice–size relationships. See Tables A1 and A2 for additional details regarding each study sample (e.g., significance levels of correlations; country, size and mean age of samples; and average voice and size measures of men and women)

Table A1. Published studies measuring relationships between voice pitch (F_0) and height or weight within sexes

Study	Location	*n*	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean F_0 (range) in Hz	Speech type	*r*	*p*[b]	Voice–size relationship[c]	Meta[d]
Men											
Majewski, Hollien, and Zalewski (1972)	Poland	103	21.7	174	68	138	Paragraph	.07	ns	F_0–height	+
			(17–28)	(160–1856)	(52–90)	(93–195)		-.02	ns	F_0–weight	+
Hollien and Jackson (1973)	U.S.A.	157	20.3	178	75	129	Paragraph	?	ns	F_0–height	x
			(18–26)	(161–197)	(53–118)	(93–178)		?	ns	F_0–weight	x
Lass and Brown (1978)	U.S.A.	15	~20.8	180	75	120	Sentence	.30	ns	F_0–height	+
			(~18–25)	(172–188)	(61–101)	(84–150)		.16	ns	F_0–weight	+
Graddol and Swann (1983)	England	12	(~26–39)	178	72	114	Paragraph (neutral)	-.71	**	F_0–height	+
				(168–185)	(63–80)	(98–136)		-.31	ns	F_0–weight	+
						122	Paragraph (social)	-.46	ns	F_0–height	+
						(111-136)		-.35	ns	F_0–weight	+
						86	Vowel	-.62	*	Low F_0–height	x
						(68–111)	(lowest F_0)	-.26	ns	Low F_0–weight	x
Künzel (1989)	Germany	105	(~19–61)	178	78	116	Paragraph	.16	ns	F_0–height	+
				(166–195)				.07	ns	F_0–weight	+
van Dommelen and	Norway	15	26	184	81	114	Paragraph	?	ns	F_0–height	x

Study	Location	n	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean F_0 (range) in Hz	Speech type	r	p[b]	Voice–size relationship[c]	Meta[d]
Moxness (1995)	Norway	15	(23–32)	(171–191)	(63–114)	114	and words	?	ns	F_0–height	x
								?	ns	F_0–weight	x
Collins (2000)	Netherlands	34	22.4 (18–30)	(164–205)	(51–89)		Vowels	?	ns	PC[e](harmonic spacing)–PC(size)	x
González (2004)	Spain	27	(~20–30)	177	76		Vowels	-.25	ns	F_0–height	+
				(160–189)	(62–102)			-.22	ns	F_0–weight	+
Rendall, Kollias, Ney, and Lloyd (2005)	Canada	34	~23 (~18–44)	184	85	113	Vowels (isolated and in words)	-.33	?	F_0–height	+
Bruckert, Liénard, Lacroix, Kreutzer, and Leboucher (2006)	France	26	24.2	(165–190)	(60–90)		Vowels	?	ns	PC(pitch)–PC(size)	x
Evans, Neave, and Wakelin (2006)	England	50	29.1	180	81	107	Vowels	?[f]	ns	F_0–height	x
			(18–68)	(165–195)	(3–111)	(76–160)		-.34[f]	*	F_0–weight	x
González (2007)	Spain	53	21.6	177	75		Vowel	-.25	ns	F_0–height	+
			(20-27)	(162–197)	(52–102)			-.24	ns	Min. F_0–height	x
								-.26	ns	Max. F_0–height	x
								-.14	ns	F_0–weight	+
								-.09	ns	Min. F_0–weight	x

Study	Location	*n*	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean F_0 (range) in Hz	Speech type	*r*	*p*[b]	Voice–size relationship[c]	Meta[d]
								-.16	ns	Max. F_0–weight	x
Sell,Bryant, Cosmides, Tooby, Sznycer, and von Rueden (2010)	U.S.A.	63	18.7 (18–22)				Sentence	-.24	ns	F_0–height	+
								.01	ns	F_0–weight	+
		50	20.2				Sentence	-.17	ns	F_0–height	+
			(18–31)					-.02	ns	F_0–weight	+
	U.S.A.	54	19.9				Vowels	.11	ns	F_0–height	+
			(18–23)					.18	ns	F_0–weight	+
	Bolivia	49	35.8				Sentence	-.09	ns	F_0–height	+
			(19–68)					.19	ns	F_0–weight	+
	Argentina	20	34.8				Sentence	-.19	ns	F_0–height	+
			(15–71)					-.13	ns	F_0–weight	+
Puts, Apicella, and Cardenas (2012)	U.S.A.	176	20.1	179	79	111	Sentence	-.17	*	F_0–height	+
			(18–26)	(159–197)	(53–132)			.01	ns	F_0–weight	+
	Tanzania	32		162.5	51	115	Word	-.20	ns	F_0–height	+
				(146–179)	(39–67)			-.08	ns	F_0–weight	+
Hamdan, Al-	Lebanon	40	24.2	179	80	120	Vowel	-.05	ns	F_0–height	+

Study	Location	n	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean F_0 (range) in Hz	Speech type	r	p[b]	Voice–size relationship[c]	Meta[d]
Barazi, Tabri, Saade, Kutkut, and Sinno (2012)	Lebanon	40	(18–40)	(163–196)	(59–110)	(84–158)	Vowel	-.05	ns	F_0–height	+
								-.12	ns	F_0–weight	+
Women											
Lass and Brown, (1978)	U.S.A.	15	~20.8	168	59	197	Sentence	-.12	ns	F_0–height	+
			(~18–25)	(156–183)	(44–75)	(153–217)		-.26	ns	F_0–weight	+
Graddol and Swann (1983)	England	15	(~26–39)	166	60	203	Paragraph (neutral)	-.33	ns	F_0–height	+
				(158–174)	(52–70)	(174–233)		.06	ns	F_0–weight	+
						208	Paragraph (social)	-.19	ns	F_0–height	+
						(188–238)		-.01	ns	F_0–weight	+
						153	Vowel (lowest F_0)	-.20	ns	F_0–height	+
						(126–183)		.19	ns	F_0–weight	+
Künzel (1989)	Germany	78	(~19–61)	167	61	211	Paragraph	.001	ns	F_0–height	+
				(155–179)				.16	ns	F_0–weight	+
van Dommelen and Moxness, (1995)	Norway	15	25.5	169	63	216	Paragraph and words	?	ns	F_0–height	+
			(20–36)	(155–180)	(51–72)			?	ns	F_0–weight	+
Collins and Missing (2003)	Spain	30	21.4 (19–26)	(156–178)	(47–95)		Vowels	-.37	*	PC(harmonic spacing)–PC(size)	x
González (2004)	Spain	55	(~20–30)	164	58		Vowels	.1	ns	F_0–height	+

Study	Location	n	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean F_0 (range) in Hz	Speech type	r	p[b]	Voice–size relationship[c]	Meta[d]
González (2004)		55	(~20–30)	(150–187)	(41–80)		Vowels	.1	ns	F_0–height	+
								-.08	ns	F_0–weight	+
Rendall et al. (2005)	Canada	34	~23 (~18–44)	168	64.4	204	Vowels (isolated and in words)	.1	?	F_0–height	+
González (2007)	Spain	81	21.6 (20–27)	164 (150–187)	57 (40–80)		Vowels	.06	ns	F_0–height	+
								.00	ns	Min. F_0–height	x
								.08	ns	Max. F_0–height	x
								-.17	ns	F_0–weight	+
								-.2	ns	Min. F_0–weight	x
								-.05	ns	Max. F_0–weight	x
Sell et al. (2010)	U.S.A.	50	18.8 (18–22)				Sentence	-.13	ns	F_0–height	+
								-.23	ns	F_0–weight	+

Studies are listed in order of year published for each sample sex. *n:* sample size; F_0: fundamental frequency.

[a] Where '~' is indicated, age (mean and range) was collapsed across samples of men and women.

[b] Statistical significance of correlation coefficient (r): *$p < .05$; **$p < .01$; ? = not given; ns = nonsignificant.

[c] Voice and size parameters are mean based unless otherwise specified.

[d] Studies included in the meta-analysis are indicated by '+'; those not included are indicated by 'x' (see main text for exclusion criteria).

[e] PC: principal component measures were used in regression.

[f] Partial correlation controlling for age (where age and F_0 were correlated, r = .43).

Table A2. Published studies measuring relationships between formants (VTL estimates) and height or weight within sexes

Study	Location	*n*	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean formant measure	Speech type	*r*	*p*[b]	Voice–size relationship[c]	Meta[d]
Men											
Greisbach (1999)	Germany	43	(19–40)	182.1			Vowel	.10	?	*F*1–height	+
				(165–200)				-.20	?	*F*2–height	+
								-.44	**	*F*3–height	+
								-.39	**	*F*4–height	+
								-.46	**	*F*3+*F*4–height	x
Collins (2000)	Nether-lands	34	22.4 (18–30)	(164–205)	(51–89)		Vowels	?	ns	PC[e](F_i, D_f)–PC(size)	x
González (2004)	Spain	27	(~20–30)	177	76		Vowels	-.10	ns	*F*1–height	+
				(160–189)	(62–102)			-.4	*	*F*2–height	+
								-.32	ns	*F*3–height	+
								-.23	ns	*F*4–height	+

Study	Location	*n*	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean formant measure	Speech type	*r*	*p*[b]	Voice–size relationship[c]	Meta[d]
								-.20	ns	$F1$–weight	+
								-.10	ns	$F2$–weight	+
								-.18	ns	$F3$–weight	+
								-.33	ns	$F4$–weight	+
		29	(~20–30)	177	74		Paragraph	-.15	ns	$F1$–height	+
				(162–189)	(62–102)			-.30	ns	$F2$–height	+
								.27	ns	$F3$–height	+
								.02	ns	$F4$–height	+
								.05	ns	D_f–height	+
								.19	ns	$F1$–weight	+
								.21	ns	$F2$–weight	+
								.06	ns	$F3$–weight	+
								.33	ns	$F4$–weight	+
								.32	ns	D_f–weight	+

Study	Location	*n*	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean formant measure	Speech type	*r*	*p*[b]	Voice–size relationship[c]	Meta[d]
Rendall, Kollias, Ney, and Lloyd (2005)	Canada	34	~23 (~18–44)	184	85	$F1$–$F4$ = 468, 1431, 2505, 3468 (in Hz)	Vowels (isolated and in words)	-.45	?	$F1$–height	+
								-.37	?	$F2$–height	+
								-.22	?	$F3$–height	+
								-.58	?	$F4$–height	+
Bruckert, Liénard, Lacroix, Kreutzer, and Leboucher (2006)	France	26	24.2	(165–190)	(60–90)		Vowels	-.36	†	PC($F1$, $F4$, D_f)–PC(size)	x
Evans, Neave, and Wakelin (2006)	England	50	29.1 (18–68)	180 (165–195)	81 (63–111)	D_f = 1652 (in Hz)	Vowels	-.32	*	D_f–height	+
								-.43	**	D_f–weight	+
Sell, Bryant, Cosmides, Tooby, Sznycer, and von Rueden (2010)	U.S.A.	54	19.9 (18–23)				Vowels	-.36	*	D_f–height	+
								.04	ns	D_f–weight	+

Study	Location	n	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean formant measure	Speech type	r	p[b]	Voice–size relationship[c]	Meta[d]
Puts, Apicella, and Cardenas (2012)	U.S.A.	176	20.1 (18–26)	179 (159–197)	79 (53–132)	$F1$–$F4$ = 434, 1456, 2442, 3390; D_f = 986 (in Hz); P_f = -0.85 (in Z(Hz))	Sentence	-.02	ns	D_f–height	+
								-.24	**	P_f–height	+
								-.01	ns	D_f–weight	+
								-.23	**	P_f–weight	+
	Tanzania	32		163 (146–179)	51 (39–67)	$F1$–$F4$ = 418, 1286, 2611, 3591; D_f = 1059 (in Hz); P_f = - 0.80 (in Z(Hz))	Word	-.11	ns	D_f–height	+
								-.38	*	P_f–height	+
								-.04	ns	D_f–weight	+
								-.39	*	P_f–weight	+
Women											
Greisbach (1999)	Germany	48	(~19–40)	169 (153–183)			Vowel	.09	?	$F1$–height	+
								-.22	?	$F2$–height	+
								-.46	**	$F3$–height	+

Study	Location	*n*	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean formant measure	Speech type	*r*	*p*[b]	Voice–size relationship[c]	Meta[d]
								-.40	**	$F4$–height	+
								-.49	**	$F3$+$F4$–height	x
Collins and Missing (2003)	Spain	30	21.4 (19–26)	(156–178)	(47–95)		Vowels	-.41	*	PC($F4$, $F5$, D_f) –height	x
González (2004)	Spain	55	(~20–30)	164	58		Vowels	-.08	ns	$F1$–height	+
				(150–187)	(41–80)			-.51	**	$F2$–height	+
								-.29	*	$F3$–height	+
								-.22	ns	$F4$–height	+
								-.08	ns	$F1$–weight	+
								-.47	**	$F2$–weight	+
								-.24	ns	$F3$–weight	+
								-.24	ns	$F4$–weight	+
		62	(~20–30)	163	58		Paragraph	-.04	ns	$F1$–height	+
				(150–187)	(40–80)			-.30	*	$F2$–height	+

Study	Location	n	Mean age[a] (range) in years	Mean height (range) in cm	Mean weight (range) in kg	Mean formant measure	Speech type	r	p[b]	Voice–size relationship[c]	Meta[d]
								-.34	**	$F3$–height	+
								-.32	**	$F4$–height	+
								-.39	**	D_f–height	+
								-.27	*	$F1$–weight	+
								-.34	**	$F2$–weight	+
								-.16	ns	$F3$–weight	+
								-.19	ns	$F4$–weight	+
								-.06	ns	D_f–weight	+
Rendall et al. (2005)	Canada	34	~23 (~18–44)	168	64	$F1$–$F4$ = 583, 1747, 2915, 4089 (in Hz)	Vowels (isolated and in words)	-.10	?	$F1$–height	+
								.00	?	$F2$–height	+
								-.30	?	$F3$–height	+
								-.22	?	$F4$–height	+

Studies are listed in order of year published for each sample sex. n: sample size; $F1$–$F4$: first to fourth formant; D_f: formant dispersion; P_f: formant position.

[a] Where '~' is indicated, age (mean and range) was collapsed across samples of men and women.

[b] Statistical significance of reported correlation coefficient (*r*): †p=.06; *p<.05; **p<.01; ? = not given; ns = nonsignificant.

[c] Voice and size parameters are mean based unless otherwise specified.

[d] Studies included in the meta-analysis are indicated by '+'; those not included are indicated by 'x' (see main text for exclusion criteria).

[e] PC: principal component measures were used in regression.

Table A3. Power analysis: minimum sample sizes required to reliably predict height or weight from F_0 or VTL estimates

	Men		Women		All adults	
	$\bar{r}$	Minimum *n*	$\bar{r}$	Minimum *n*	$\bar{r}$	Minimum *n*
Voice–height correlations						
$\boldsymbol{F_0}$	**-0.13**	**618**	**-0.07**	**2140**	**-0.10**	**1047**
$F1$	-0.13	618	-0.04	6563	-0.08	1638
$F2$	-0.22	213	-0.19	287	-0.20	259
$F3$	-0.26	151	-0.22	213	-0.24	178
$F4$	-0.30	113	-0.25	164	-0.27	140
Fn	-0.31	105	-0.22	213	-0.25	164
MFF	-0.28	130	-0.18	320	-0.22	213
D_f	-0.18	320	-0.24	178	-0.21	234
P_f	-0.29	121	-0.21	234	-0.25	164
ΔF	-0.32	99	-0.22	213	-0.26	151
VTL(F_i)	0.24	178	0.17	360	0.20	259
VTL(ΔF)	0.32	99	0.23	195	0.26	151
CFA	0.30	113	-0.24	178	0.27	140
Voice–weight correlations						
$\boldsymbol{F_0}$	**-0.03**	**11 671**	**-0.14**	**532**	**-0.07**	**2140**
$F1$	-0.15	463	-0.08	1638	-0.11	864
$F2$	-0.09	1293	-0.22	213	-0.17	360
$F3$	-0.18	320	-0.16	406	-0.17	360
$F4$	-0.15	463	-0.24	178	-0.21	234
Fn	-0.22	213	-0.23	195	-0.22	213

	Men		Women		All adults	
	$\bar{r}$	Minimum *n*	$\bar{r}$	Minimum *n*	$\bar{r}$	Minimum *n*
MFF	-0.24	178	0.20	259	-0.22	213
D_f	-0.10	1047	-0.21	234	-0.15	463
P_f	-0.25	164	-0.22	213	-0.23	195
ΔF	-0.22	213	-0.21	234	-0.22	213
VTL(F_i)	0.20	259	0.19	287	0.19	287
VTL(ΔF)	0.22	213	0.25	164	0.24	178
CFA	0.23	195	-0.16	406	0.19	287

n: sample size; $\bar{r}$: mean weighted correlation coefficient; F_0: fundamental frequency; *F*1–*F*4: first to fourth formant; *Fn*: average formant frequency; MFF: geometric mean formant frequency; D_f: formant dispersion; P_f: formant position; ΔF: formant spacing; VTL(F_i): apparent vocal tract length derived from mean formants; VTL(ΔF): apparent vocal tract length derived from formant spacing; CFA: confirmatory factor analysis (factor scores). This table shows the results of power analysis used to assess the minumum sample sizes required to obtain, with statistical confidence, the corresponding mean weighted correlations reported here and in Table 1, where alpha = .05 (the probability of a false positive or type I error) and power = .90 (the probability of a true positive). Substantially larger sample sizes are required to reliably assess body size from F_0 (shown in bold) than from formant-based VTL estimates.

References

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* (Version 5.2.15). [Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/.

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 B: Biological Sciences, 273*, 83–89.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: L. Erlbaum.

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour, 60*, 773–780.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65*, 997–1004.

Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PLoS One*, *7*, e31353.

Egger, M., Smith, G.D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.

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*, 160–163.

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*, 561–568.

Fitch, W. T. (1994). *Vocal-tract length perception and the evolution of language* (Doctoral dissertation). Providence, RI: Brown University.

Fitch, W. T. (1997). Vocal-tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America, 102*, 1213–1222.

González, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics, 32*, 277–287.

González, J. (2007). Correlations between speakers' body size and acoustic parameters of voice. *Perceptual and Motor Skills, 105*, 215.

Graddol, D., & Swann, J. (1983). Speaking fundamental frequency: Some physical and social correlates. *Language and Speech, 26*, 351–366.

Greisbach, R. (1999). Estimation of speaker height from formant frequencies. *Forensic Linguistics, 6*, 265–277.

Hamdan, A. L., Al-Barazi, R., Tabri, D., Saade, R., Kutkut, I., & Sinno, S. (2012). Relationship between acoustic parameters and body mass analysis in young males. *Journal of Voice, 26*, 144–147.

Hollien, H., & Jackson, B. (1973). Normative data on the speaking fundamental frequency characteristics of young adult males. *Journal of Phonetics, 1*, 117–120.

Irino, T., Aoki, Y., Kawahara, H., & Patterson, R.D. (2012). Comparison of performance with voiced and whispered speech in word recognition and mean-formant-frequency discrimination. *Speech Communication, 54*, 998–1013.

Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica, 46*, 117–125.

Lass, N. J., & Brown, W. S. (1978). Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies. *Journal of the Acoustical Society of America, 63*, 1218–1220.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. New York, NY: Sage.

Majewski, W., Hollien, H., & Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males. *Phonetica, 25*, 119–125.

Patterson, R. D., Smith, D. R. R., van Dinther, R., & Walters, T. C. (2008). Size information in the production and perception of communication sounds. In W. A. Yost, Popper, A. N., & R. R. Fay, (Eds.), *Auditory perception of sound sources* (pp. 43–75). New York, NY: Springer.

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*, 509–519.

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

Puts, D. A., Apicella, C. L., & Cardenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 279*, 601–609.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour, 65*, 519–530.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F-0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America, 117*, 944–955.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., & von Rueden, C. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences, 277*, 3509–3518.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America, 118*, 3177–3186.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*, 1046–1055.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Turner, R. E., Walters, T. C., Monaghan, J. J. M., & Patterson, R. D. (2009). A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. *Journal of the Acoustical Society of America, 125*, 2374.

van Dommelen, W. A., & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and Speech, 38*, 267–287.

Appendix 3: Additional Data Tables for Chapter 3

Table A4. Height differences between men in vocalizer pairs used in Experiments 1 and 3 of Chapter 3

Group	Male Pairs	Height Difference Between Men in Pair		
		Mean ± SD (cm)[a]	Std. error (cm)	Range (cm)
1	1-15	7.87±6.13	1.58	0-19
2	16-30	7.6±6.22	1.61	0.5-21
3	31-45	7.06±5.05	1.30	0-16
4	46-60	7.13±5.37	1.39	0-18
All (1-4)	1-60	7.42±5.58	0.72	0-21

a. Relative height between men did not differ significantly across the four groups of pairs (one-way ANOVA, two-tailed: $F_{3,56}$=0.067, p=.997).

Table A5. Frequencies of men's voices used as stimuli in Chapter 3

Vowel n[a]	$F1$ (Hz)	$F2$ (Hz)	$F3$ (Hz)	$F4$ (Hz)	F_0 (Hz)[c]
Modal Speech					
/ɑ/	720±63	1228±88	2496±195	3600±200	115±13
n	30	30	30	30	30
/i/	297±39	2245±179	3030±179	3638±197	114±13
n	30	30	30	29	30
/ɛ/	579±66	1744±142	2535±160	327±212	111±13
n	30	30	30	30	30
/o/	480±46	974±67	2434±157	3273±175	109±12
n	30	30	30	30	30
/u/	338±34	1070±159	2246±266	3272±187	108±12
n	30	30	30	30	30
All[b]	483±29	1453±78	2549±118	3461±132	112±12
Whispered Speech					
/ɑ/	945±77	1435±154	2622±143	3584±188	
n	30	30	30	30	
/i/	438±77	2387±199	3018±222	3763±258	
n	13	28	27	28	
/ɛ/	766±86	1892±185	2653±146	3615±189	
n	30	30	30	29	
/o/	702±89	1121±96	2582±164	3379±200	
n	24	28	28	28	
/u/	438±73	1236±219	2469±191	3384±219	
n	20	29	29	29	
All[b]	724±104	1617±135	2666±133	3548±157	

$F1$ - $F4$ = first to fourth formant; F_0 = fundamental frequency; Hz = hertz. Means ± standard deviations of the first four formants and the fundamental frequency (voice pitch) of each of five vowels measured from men's modal or whispered speech.

[a.] *n* = number of men's voices included in calculating the mean voice measure.

[b.] Averaged across all five vowels.

[c.] Whispered speech does not contain F_0.

Table A6. Average measures of formant structure and voice pitch, and the degree to which each predicted relative height among vocalizer pairs in Experiments 1 and 3 of Chapter 3

	Mean	SD	Predictive of relative height[a]
Formants			
Fn	1986.4 Hz	65.73 Hz	61.1% (33 of 54)
VTL	17.80 cm	0.68 cm	63% (34 of 54)
P_f	-0.85 F'	0.32 F'	61.1% (33 of 54)
D_f	992.87 Hz	44.64 Hz	70.4% (38 of 54)
MFF	1576 Hz	58.27 Hz	63% (34 of 54)
Pitch			
F_0	112.07 Hz	11.83 Hz	51.9% (28 of 54)
Pitch (semitones)	81.6 cents	2.11 cents	55.6% (30 of 54)
Pitch (mel)	102.55 mel	12.14 mel	53.7% (29 of 54)
Pitch (ERB)	3.34 ERB	0.37 ERB	53.7% (29 of 54)

Fn = average formant frequency; VTL = apparent vocal tract length; P_f = formant position; D_f = formant dispersion; MFF = geometric mean formant frequency (see equations 1-5); F_0 = fundamental frequency; ERB = equivalent rectangular bandwidth; Hz = hertz; F' = standardized formant. Means ± standard deviations of voice measures taken from n = 30 men.

[a.] The percentage (and number) of vocalizer pairs in which the voice measure predicted the relative difference in height between men (i.e., the taller man had relatively lower formants [or longer VTL] or lower pitch than did the shorter man), based on 54 vocalizer pairs whose difference in height exceeded 0.5 cm.

Table A7. Average formant differences in modal and whispered speech between men in vocalizer pairs used in Experiments 1 and 3 of Chapter 3

Formant	Speech type	Mean difference[a] ± SD (Hz)	t[b]	p[b]
$F1$	Modal	-3.56±36.25	1.486	.143
	Whispered	-25.1±143.4		
$F2$	Modal	-38.43±89.32	0.069	.945
	Whispered	-40.22±177.48		
$F3$	Modal	-23.39±168.78	2.316	.024*
	Whispered	-80.95±166.77		
$F4$	Modal	-79.76±173.99	-0.436	.665
	Whispered	-68. 31±218.85		

$F1$ - $F4$ = first to fourth formant; Hz = hertz. Means ± standard deviations of the relative formants (first to fourth) of men in vocalizer pairs taken from either modal or whispered speech (n = 54 pairs), and results of paired sample t-tests comparing relative formants between modal and whispered speech types.

a. Differences in men's $F1$ - $F4$ within each pair were calculated by subtracting the F_i of the shorter male in the pair from the corresponding F_i of the taller male in the pair, such that mean differences below 0 reflect relatively lower F_i in the taller male.

b. Results of paired sample t-tests (df=53, two-tailed) indicate that men's relative $F1$, $F2$, and $F4$ were no different for modal and whispered speech. Differences in $F3$ between men were greater for whispered than for modal speech but were in the predicted direction (the taller man had lower $F3$ than the shorter man). Thus, any differences in relative formants across speech types could only have improved listeners' accuracy from whispered speech and whispered SWS compared to modal speech, and therefore cannot account for listeners' poorer performance from whispered and SWS compared to modal speech in Experiment 1. *p<.05, two-tailed paired sample t-test.

Table A8. Relationships among five different measures of formant structure taken from men's modal or whispered speech in Chapter 3

Speech type		Modal				Whispered				
	Formant measure	VTL	P_f	D_f	MFF	*Fn*	VTL	P_f	D_f	MFF
Modal	*Fn*	-.909**	.925**	.847**	.93**	.467**	-.401*	.447*	.555**	.335†
	VTL		-.998**	-.575**	-.999**	-.433*	.391*	-.420*	-.421*	-.345†
	P_f			.601**	.994**	.440*	-.394*	.427*	.439*	.319†
	D_f				.56**	.468**	-.404*	.445*	.610**	.350†
	MFF					.43**	-.392*	.42*	.42*	.347†
Whispered	*Fn*						-.969**	.998**	.456*	.958**
	VTL							-.979**	-.306†	-.992**
	P_f								.410*	.973**
	D_f									.237

Fn = average formant frequency; VTL = apparent vocal tract length; P_f = formant position; D_f = formant dispersion; MFF = geometric mean formant frequency (see equations 1-5). Pearson correlation coefficients (*r*) are given in each cell, *n* = 30 men per cell.

** p<.01, * p<.05, † p<.1, two-tailed bivariate Pearson correlations.

Table A9. Relationships among four different measures of voice pitch taken from men's modal speech in Chapter 3

Pitch measure	F_0	semitones	mel	ERB
F_0		.988**	.999**	.998**
semitones			.993**	.995**
mel				.999**
ERB				

ERB = equivalent rectangular bandwidth; F_0 = fundamental frequency.

Pearson correlation coefficients (r) are given in each cell, n = 30 men per cell.

** p<.01, two-tailed bivariate Pearson correlations.

Appendix 4: Audio S1-S4 Captions for Chapter 3

Audio S1. Modal speech: Natural voiced speech originating from the same adult male as Audio S2-S4 (age 18, 187 cm tall, average modal F_0 = 103 Hz, $F1 - F4$ = 532, 1600, 2650, 3392 Hz, respectively). The male is speaking a series of five English monophthong vowels, /ɑ/, /i/, /ɛ/, /o/, and /u/. See also Figure 1.

Audio S2. Whispered speech: Natural unvoiced speech originating from the same adult male as Audio S1, S3, and S4. The male is speaking a series of five English vowels.

Audio S3. Modal SWS: Synthesized modal three-formant sine-wave speech originating from the same adult male as Audio S1, S2, and S4. The male is speaking a series of five English vowels. Synthesized from natural modal speech.

Audio S4. Whispered SWS: Synthesized whispered three-formant sine-wave speech originating from the same adult male as Audio S1-S3. The male is speaking a series of five English vowels. Synthesized from natural whispered speech.