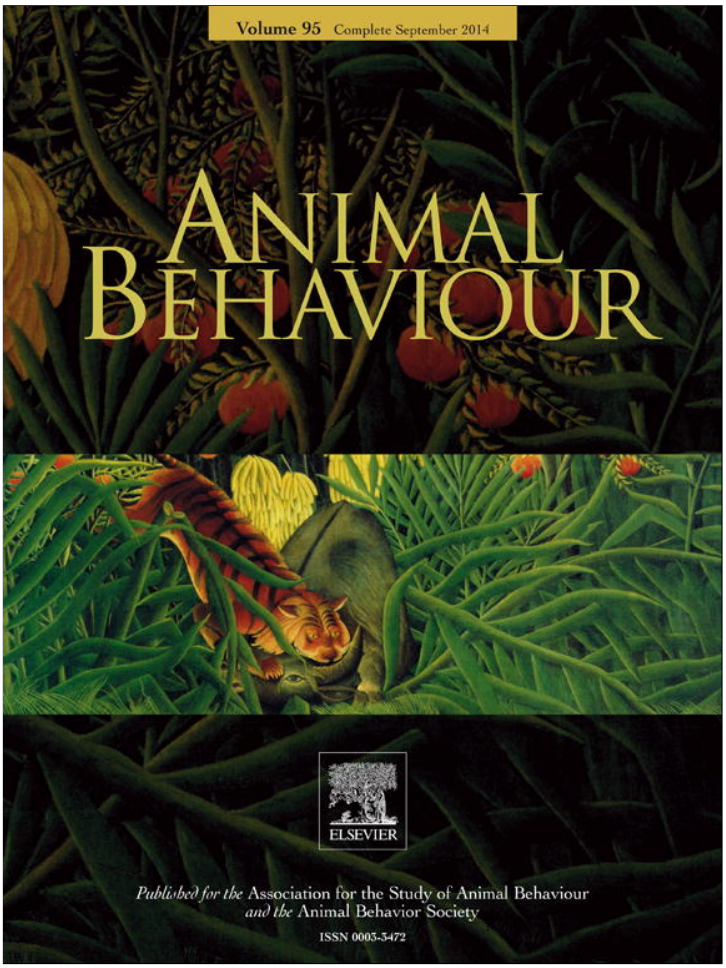
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/authorsrights

Author's personal copy

Animal Behaviour 95 (2014) 89–99

Contents lists available at ScienceDirect

Animal Behaviour

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

Vocal indicators of body size in men and women: a meta-analysis

Katarzyna Pisanski [a], Paul J. Fraccaro [a], Cara C. Tigue [a], Jillian J. M. O'Connor [a], Susanne Röder [b, c, d], Paul W. Andrews [a], Bernhard Fink [b, c], Lisa M. DeBruine [e], Benedict C. Jones [e], David R. Feinberg [a, *]

[a] Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON, Canada
[b] Institute of Psychology, University of Göttingen, Göttingen, Germany
[c] Courant Research Centre Evolution of Social Behavior, University of Göttingen, Göttingen, Germany
[d] Department of General Psychology and Methodology, University of Bamberg, Bamberg, Germany
[e] Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, U.K.

ARTICLE INFO

Article history:
Received 31 January 2014
Initial acceptance 16 April 2014
Final acceptance 4 June 2014
Available online 31 July 2014
MS. number: A14-00100R

Keywords:
bioacoustics
formant
fundamental frequency
height
human
sexual selection
source filter
vocal tract
voice
weight

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 (F0). 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 F0. Individual VTL estimates explained up to 10% of the variance in height and weight, whereas F0 explained less than 2% and correlated only weakly with size within sexes. Statistically reliable size estimates from F0 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.
© 2014 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

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 (F0) 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 F0 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 F0 and formants are decoupled in most vertebrates by showing that changes in F0 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 F0 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).

* Correspondence: D. R. Feinberg, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada.
E-mail address: feinberg@mcmaster.ca (D. R. Feinberg).

http://dx.doi.org/10.1016/j.anbehav.2014.06.011
0003-3472/© 2014 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Author's personal copy

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 F0 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 F0 to drop (Harries, Hawkins, Hacking, & Hughes, 1998), and directly affects F0 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, F0 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 F0 than smaller individuals (Fitch & Hauser, 2003; Morton, 1977; Titze, 1994). Moreover, because F0 and formants are typically decoupled in vertebrates, larger individuals with correspondingly longer vocal tracts may produce lower formants regardless of F0 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 F0 and formants as honest indicators of size in a wide range of species. This research has generally confirmed that both F0 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, 2007; 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 F0. 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* (Plotsky, Rendall, Riede, & Chase, 2013; Riede & Fitch, 1999). It is of potential interest that both F0 and formants appear to predict the size of male but not female giant pandas, *Ailuropoda melanoleuca* (Charlton, Zhihe, & Snyder, 2009), whereas F0 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 F0 in humans. A large proportion of the variation in F0 and formants among humans can 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). However, 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 F0 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 F0 estimates of size = −0.71 to +0.30; range of r for formant estimates of size = −0.58 and +0.32).

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 (F*n*; 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 the Appendix). 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 F0 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

Author's personal copy

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; Taylor & Reby, 2010; Titze, 1994). 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, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Rendall, Vokey, & Nemeth, 2007) or voice-like stimuli (Charlton, Taylor, & Reby, 2013; Smith & Patterson, 2005). Perceptually, listeners associate low F0 and/or low formants with large body size both between and within sexes (Charlton et al., 2013; van Dommelen & Moxness, 1995; Feinberg, Jones, Little, Burt, & Perrett, 2005; Pisanski et al., 2014; Pisanski, Mishra, & Rendall, 2012; Pisanski & Rendall, 2011; Rendall et al., 2007; Smith & Patterson, 2005). If F0 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 F0–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 F0 and body size in Table S1, and between formants and size in Table S2 (see Supplementary material) including the 24 correlations excluded from the meta-analysis (indicated with an 'x'; see also Fig. S1). 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 F0 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, F0 and formants. Height (mean ± 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 F0 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 et al., 2005; González, 2004; Greisbach, 1999; Pisanski & Rendall, 2011; Rendall et al., 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). Additional details are provided in the Supplementary material.

Statistical Analysis

Detailed methods of analysis of meta-data are described in the Supplementary material and all equations are given in the Appendix. Briefly, we computed a mean weighted correlation coefficient, $\bar{r}$ (Hedges & Olkin, 1985; Lipsey & Wilson, 2001), between F0 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, F0–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 F0 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 F0 and height or weight, and between

Author's personal copy

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																		
F0	−0.13	0.03	−0.19, −0.07	<0.001	21	1119	−0.07	0.04	−0.14, 0	0.063	13	795	−0.10	0.02	−0.15, −0.06	<0.001	34	1914
F1	−0.13	0.05	−0.23, −0.03	0.01	8	395	−0.04	0.04	−0.12, 0.04	0.29	8	637	−0.08	0.03	−0.14, −0.02	0.015	16	1032
F2	−0.22	0.05	−0.31, −0.12	<0.001	8	395	−0.19	0.04	−0.26, −0.11	<0.001	8	637	−0.20	0.03	−0.25, −0.14	<0.001	16	1032
F3	−0.26	0.05	−0.35, −0.17	<0.001	8	395	−0.22	0.04	−0.29, −0.14	<0.001	8	637	−0.24	0.03	−0.29, −0.18	<0.001	16	1032
F4	−0.30	0.05	−0.39, −0.21	<0.001	8	395	−0.25	0.04	−0.32, −0.18	<0.001	8	637	−0.27	0.05	−0.36, −0.17	<0.001	16	1032
F*n*	−0.31	0.06	−0.42, −0.20	<0.001	4	262	−0.22	0.04	−0.29, −0.14	<0.001	4	438	−0.25	0.04	−0.32, −0.18	<0.001	8	700
MFF	−0.28	0.06	−0.39, −0.17	<0.001	4	262	−0.18	0.05	−0.27, −0.08	<0.001	4	438	−0.22	0.04	−0.29, −0.14	<0.001	8	700
D_f	−0.18	0.04	−0.26, −0.10	<0.001	9	603	−0.24	0.05	−0.33, −0.15	<0.001	5	500	−0.21	0.03	−0.27, −0.15	<0.001	14	1103
P_f	−0.29	0.05	−0.37, −0.21	<0.001	6	470	−0.21	0.05	−0.30, −0.11	<0.001	4	438	−0.25	0.03	−0.31, −0.19	<0.001	10	908
ΔF	−0.32	0.06	−0.43, −0.21	<0.001	4	262	−0.22	0.05	−0.31, −0.13	<0.001	4	438	−0.26	0.04	−0.33, −0.19	<0.001	8	700
VTL(F_i)	0.24	0.06	0.12, 0.35	<0.001	4	262	0.17	0.05	0.08, 0.26	<0.001	4	438	0.20	0.04	0.12, 0.27	<0.001	8	700
VTL(ΔF)	0.32	0.06	0.20, 0.43	<0.001	4	262	0.23	0.05	0.14, 0.32	<0.001	4	438	0.26	0.04	0.19, 0.33	<0.001	8	700
CFA	0.30	0.06	0.40, 0.19	<0.001	1	262	−0.24	0.06	−0.34, −0.13	<0.001	1	326	0.27[d]	0.04	0.34, 0.19	<0.001	2	588
Voice–weight correlations																		
F0	−0.03	0.03	−0.09, 0.03	0.30	21	1132	−0.14	0.04	−0.21, −0.07	<0.001	12	760	−0.07	0.02	−0.12, −0.03	0.002	33	1892
F1	−0.15	0.06	−0.26, −0.04	0.01	6	315	−0.08	0.04	−0.16, 0	0.064	6	553	−0.11	0.03	−0.17, −0.04	0.002	12	868
F2	−0.09	0.06	−0.20, 0.02	0.12	6	315	−0.22	0.04	−0.30, −0.14	<0.001	6	553	−0.17	0.03	−0.24, −0.11	<0.001	12	868
F3	−0.18	0.06	−0.29, −0.07	0.002	6	315	−0.16	0.04	−0.24, −0.07	<0.001	6	553	−0.17	0.03	−0.23, −0.10	<0.001	12	868
F4	−0.15	0.06	−0.26, −0.04	0.01	6	315	−0.24	0.04	−0.32, −0.16	<0.001	6	553	−0.21	0.03	−0.27, −0.14	<0.001	12	868
F*n*	−0.22	0.06	−0.33, −0.09	<0.001	4	259	−0.23	0.05	−0.32, −0.14	<0.001	4	436	−0.22	0.04	−0.29, −0.15	<0.001	8	695
MFF	−0.24	0.06	−0.36, −0.12	<0.001	4	259	−0.20	0.05	−0.29, −0.11	<0.001	4	436	−0.22	0.04	−0.29, −0.14	<0.001	8	695
D_f	−0.10	0.04	−0.18, −0.02	0.016	9	600	−0.21	0.05	−0.29, −0.12	<0.001	5	498	−0.15	0.03	−0.21, −0.09	<0.001	14	1098
P_f	−0.25	0.05	−0.33, −0.16	<0.001	6	467	−0.22	0.05	−0.31, −0.13	<0.001	4	436	−0.23	0.03	−0.30, −0.17	<0.001	10	903
ΔF	−0.22	0.06	−0.34, −0.10	<0.001	4	259	−0.21	0.05	−0.3, −0.12	<0.001	4	436	−0.22	0.04	−0.29, −0.14	<0.001	4	695
VTL(F_i)	0.20	0.06	0.07, 0.31	0.002	4	259	0.19	0.05	0.10, 0.28	<0.001	4	436	0.19	0.06	0.07, 0.31	<0.001	8	695
VTL(ΔF)	0.22	0.06	0.10, 0.34	<0.001	4	259	0.25	0.05	0.16, 0.33	<0.001	4	436	0.24	0.04	0.17, 0.31	<0.001	4	695
CFA	0.23	0.06	0.34, 0.11	<0.001	1	254	−0.16	0.06	−0.26, −0.05	<0.001	1	323	0.19[d]	0.04	0.26, 0.11	<0.001	2	577

F0: fundamental frequency; F1–F4: first to fourth formant; F*n*: 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 and Supplementary material for methods used to derive VTL estimates F*n*, 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, $\alpha = 0.05$.

[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.

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 F0. Ninety-three per cent (67 of 72) of the derived mean weighted correlations between formants and body size reached statistical significant ($P < 0.05$).

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 F0–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 ($F_{3,15} = 0.48$, $P = 0.71$) and revealed no significant main or interaction effects (all $F < 0.06$, all $P > 0.80$). The model remained nonsignificant when we dropped the interaction term (sex-by-age: $F_{2,16} = 0.76$, $P = 0.49$) or either main effect term (sex: $F_{1,17} = 0.31$, $P = 0.58$; age: $F_{1,32} = 0.99$, $P = 0.33$). Thus, neither sample sex nor mean age significantly affected the strength of F0–height relationships, although we did observe a trend toward stronger F0–height relationships in samples of men than women (see Table 1). For F0–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 ($F_{3,14} = 2.65$, $P = 0.09$, $R^2 = 0.36$) but revealed no significant main or interaction effects (all $F < 1.63$, all $P > 0.22$). A model that contained only sample sex as a fixed factor was significant ($F_{1,31} = 4.21$, $P = 0.049$, $R^2 = 0.12$). Thus, sample sex but not mean age affected the strength of F0–weight relationships, where F0 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, F*n*, 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 $F < 1.62$, $P > 0.22$) or formant–weight relationships (all $F < 2.05$, $P > 0.18$) 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

Author's personal copy

levels: one vowel; five vowels; word; sentence; paragraph; a combination of these), acoustic hardware/software used for voice measurement and analysis (seven levels: FFI F0 indicator (Hollien & Harrington, 1977); Kay Elemetrics VISI Pitch (Kay Elemetrics, 1996); MDVP Multi-Dimensional Voice Program (Kay Elemetrics, 1993; Kent, Vorperian, & Duffy, 1999); Signalize (Keller, 1993); Praat (Boersma, 2001; 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 F0–height and F0–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 F0–weight relationships.

For F0–height relationships ($k \leq 34$), the effect of the method used to measure height approached significance ($F_{2,31} = 2.50$, $P = 0.098$, $R^2 = 0.14$). Thus, although measured and self-reported mean height for men or women differed by only 0.12 and 1.48 cm, respectively, F0 predicted height marginally better when height was measured using metric tape ($k = 29$, $r = -0.17$) than self-report ($k = 4$, $r = 0.01$). There were no significant effects of study year ($F_{1,32} = 1.58$, $P = 0.22$), sample location ($F_{10,23} = 1.21$, $P = 0.34$), speech stimulus ($F_{5,28} = 0.99$, $P = 0.44$), or voice measurement ($F_{4,29} = 1.25$, $P = 0.31$). For F0–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 ($F_{5,16} = 2.83$, $P = 0.051$). 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 $P > 0.16$).

Effect of Sample Size and Tests for Publication Bias

Figure 1 shows funnel plots for F0–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 F0–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 F0–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 Supplementary material). 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

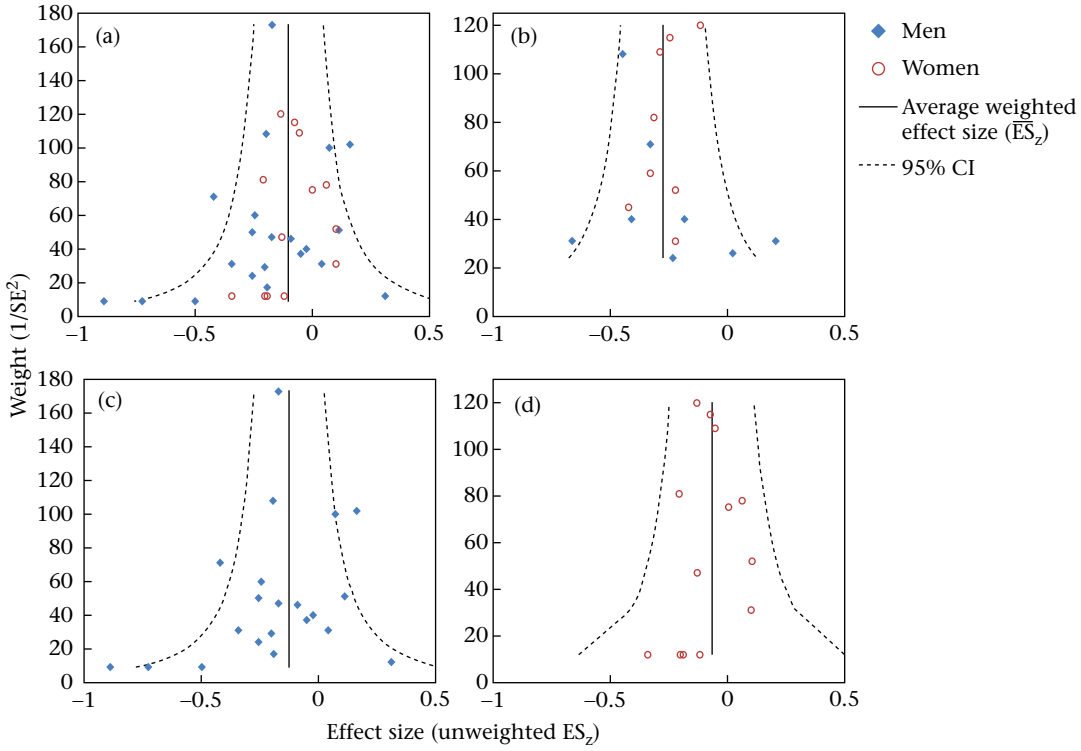


Figure 1. Funnel plots used to assess the effect of sample size and to detect publication bias in reported relationships between (a) F0 and height in samples of either sex ($k = 34$ samples, $\overline{ES}_z = -0.10$), (b) F4 and height in samples of either sex ($k = 16$; $\overline{ES}_z = -0.27$), or (c) F0 and height in samples of men ($k = 21$; $\overline{ES}_z = -0.13$) and (d) F0 and height in samples of women ($k = 13$; $\overline{ES}_z = -0.07$). 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 ($\overline{ES}_z$; see equation A10) and the dotted lines represent the lower and upper 95% confidence intervals corresponding to the Y axis. F0–height effect sizes are plotted separately for samples of men (c) and women (d) because of a sex difference in $\overline{ES}_z$ values for this relationship. See Supplementary material for additional details.

Author's personal copy

each sex (Fig. 1b). However, the plots depicting F0–height relationships in men (Fig. 1c) and women (Fig. 1d) revealed a slight skew to the left.

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 F0–height relationships stronger than the average weighted effect size (i.e. to the left of the average) than F0–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 Supplementary material for additional details).

Minimum Required Sample Sizes for Future Work

Supplementary Table S3 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 F0 or formant-based VTL estimates (given the $\bar{r}s$ that we obtained) with an alpha level of 0.05 and a conservative power level of 0.90 (Cohen, 1988). To summarize, samples of at least 618 men and 2140 women are required to predict height from mean F0. 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 S1). 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), F4, CFA; for women: F4, 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 F0–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 F0–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 F0, F1–F4, F*n*, 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
Weighted means and standard deviations of voice and body size measures for each sex

		Men		Women	
		Weighted $\bar{X}$±SD	N[a]	Weighted $\bar{X}$±SD	N[a]
Voice parameter					
F0 (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
F*n* (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
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

F0: fundamental frequency (voice pitch); F1–F4: first to fourth formant; F*n*: 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, F0). Past estimates of F0–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 F0 and height or weight within either sex. In fact, F0 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 F0 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 F0 are therefore proposed to vary largely independent of skull and body size (Fitch, 2000) and F0 may even be modulated to disguise or exaggerate size (Fitch, 2000; Morton, 1977). The strong influence of pubertal and circulating testosterone on F0 in human males may further contribute to the discordance between men's F0 and physical size (Dabbs & Mallinger, 1999; Harries et al., 1998). In short, although F0 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 F0 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 F0, including several species of ungulates, pinnipeds, elephants, carnivores, marsupials and primates (for reviews see: Ey, Pfefferle, & Fischer, 2007; Fitch & Hauser, 2003; Taylor

Author's personal copy

& 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 appear 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). 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 F0-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, DeBruine, Jones, & Perrett, 2008; Feinberg et al., 2005; Fitch, 1997, 2000; Gingras et al., 2013; 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, et al., 2012; Reby & McComb, 2003; Rendall et al., 2005, 2007; Riede & Fitch, 1999; Sanvito et al., 2007; Sell et al., 2010; Smith & Patterson, 2005; Taylor, Reby, & McComb, 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, F*n*, 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 F1, 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 F*n*, F4 and CFA, whereas P_f best predicted men's weight, followed by MFF. Among women, F4 best predicted women's height, followed by D_f, CFA and VTL(ΔF), whereas VTL(ΔF) best predicted women's weight, followed by F4 and F*n*.

Theoretical Implications for Vocal Size Communication in Humans

Our results suggest that the relationship between F0 and body size is likely too weak to have strongly influenced the evolution and perception 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 F0 (Rendall et al., 2005).

Taken together this evidence suggests that variation in human F0, 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, F0 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, et al., 2012). This possibility is further supported by evidence that F0 (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 F0 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 F0 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 F0 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 F0 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 F0 have more closely spaced harmonics than do voices with higher F0 (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.

Author's personal copy

Demographic and Methodological Influences

Age did not affect the strength of F0–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 F0 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 F0 and height or between formants and height or weight in our meta-analysis. However, F0 predicted variation in weight substantially better among samples of women than among samples of men. Low F0 significantly predicted heavier body weight in women, but we found no relationship between F0 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 F0 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 F0 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 F0 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. Fitch & Fritz, 2006; Turner et al., 2009; 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), F0 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 visibility 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 (Hillenbrand 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 F0 and formant measurement. This is true for chaotic vocalizations and those in which F0 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 S1 and S2), 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 (F0) 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), F*n* (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 F0, 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

Author's personal copy

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.

Acknowledgments

We thank Diana Borak, Michael Burt, R. Elisabeth Cornwell, Lisa De Couto, Anthony Little, Fiona Moore, David Perrett, Miriam Law Smith and Michael Stirrat for assistance with the collection of voice recordings and body size measurements, Stephanie Wu for assistance with voice measurements and Greg Bryant for insightful suggestions. Portions of this work were supported by grants from the Social Sciences and Humanities Research Council of Canada, the Canadian Foundation for Innovation, the European Research Council (Starting Grant 282655, OCMATE), Ontario's Ministry of Research and Innovation (Early Researcher Award Program ER11-08-084) and the German Research Foundation.

Supplementary Material

Supplementary material for this article is available, in the online version, at http://dx.doi.org/10.1016/j.anbehav.2014.06.011.

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.

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.

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

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.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 268*(1477), 1669–1675.

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., et al. (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.

Author's personal copy

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.
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.
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.
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, 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 Primatologica, 63*, 115–118.
Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford, U.K.: Oxford University Press.
McElligott, A. G., Birrer, M., & Vannoni, E. (2006). Retraction of the mobile descended larynx during groaning enables fallow bucks (*Dama dama*) to lower their formant frequencies. *Journal of Zoology, 270*, 340–345.
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.
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*. Advance online publication. Retrieved from 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.
Ryalls, J. H., & Lieberman, P. (1982). Fundamental frequency and vowel perception. *Journal of the Acoustical Society of America, 72*(5), 1631–1634.
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.
Styler, W. (2014). *Using Praat for linguistic research*. Version 1.4.5.. Retrieved from: http://savethevowels.org/praat
Taylor, A. M., & Reby, D. (2010). The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology, 280*, 221–236.
Taylor, A. M., Reby, D., & McComb, K. (2008). Human listeners attend to size information in domestic dog growls. *Journal of the Acoustical Society of America, 123*, 2903–2909.
Taylor, A. M., Reby, D., & McComb, K. (2010). Size communication in domestic dog, *Canis familiaris*, growls. *Animal Behaviour, 79*, 205–210.
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.
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.

APPENDIX

Equations

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

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

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

Author's personal copy

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

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

$$D_j = \frac{\sum_{i=1}^{n}(F_{i+1} - F_i)}{n-1} \tag{A3}$$

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

$$P_f = \frac{\sum_{i=1}^{n} F_i'}{n} \tag{A4}$$

where F_i' is the standardized ith 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 $(2i-1)/2$ 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

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

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

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

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

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

$$\mathrm{SE}_{\mathrm{ES}_Z} = \frac{1}{\sqrt{n-3}} \tag{A9}$$

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:

$$\overline{\mathrm{ES}}_Z = \frac{\sum(w_i \mathrm{ES}_{Z_i})}{\sum w_i} \tag{A10}$$

$$\mathrm{SE}_{\overline{\mathrm{ES}}_Z} = \frac{1}{\sqrt{\sum w_i}} \tag{A11}$$

where w_i is the weight corresponding to the ith sample,

$$w_i = \frac{1}{\mathrm{SE}^2_{\mathrm{ES}_Z}} \tag{A12}$$

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

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

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

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

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

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

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

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

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

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

$$\frac{1}{\mathrm{SE}^2_{r_i}} \tag{A19}$$

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.