

# Journal of Experimental Psychology: Human Perception and Performance

## **Return to Oz: Voice Pitch Facilitates Assessments of Men's Body Size**

Katarzyna Pisanski, Paul J. Fraccaro, Cara C. Tigue, Jillian J. M. O'Connor, and David R. Feinberg

Online First Publication, June 16, 2014. <http://dx.doi.org/10.1037/a0036956>

### CITATION

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., & Feinberg, D. R. (2014, June 16). Return to Oz: Voice Pitch Facilitates Assessments of Men's Body Size. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. <http://dx.doi.org/10.1037/a0036956>

# Return to Oz: Voice Pitch Facilitates Assessments of Men's Body Size

Katarzyna Pisanski, Paul J. Fraccaro, Cara C. Tigue, Jillian J. M. O'Connor, and David R. Feinberg  
McMaster University

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 3-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.

*Keywords:* speech perception, height, fundamental frequency, formant, harmonic density

*Supplemental materials:* <http://dx.doi.org/10.1037/a0036956.supp>

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).

---

Katarzyna Pisanski, Paul J. Fraccaro, Cara C. Tigue, Jillian J. M. O'Connor, and David R. Feinberg, Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada.

This research was supported by grants to David R. Feinberg from the Social Sciences and Humanities Research Council of Canada, the Canadian Foundation for Innovation, and the Early Researcher Award Program of the Ontario Ministry of Research and Education. Portions of this work were presented at the 49th annual meeting of the Animal Behavior Society, New Mexico, June 2012; the XXI Biennial Conference on Human Ethology, Vienna, Austria, August 2012; and Behaviour, 2013, a joint meeting of the 33rd International Ethological Conference (IEC) and the Association for the Study of Animal Behaviour (ASAB), NewcastleGateshead, United Kingdom, August 2013. We thank Philip Lieberman, David Reby, Drew Rendall, Michael Ryan, and three peer reviewers for their insightful comments, as well as Paul Andrews, Lisa DeBruine, Bernhard Fink, Benedict Jones, and Susanne Röder for helpful feedback and stimulating discussions, all of which substantially improved the manuscript.

Correspondence concerning this article should be addressed to David R. Feinberg, Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4L8, Canada. E-mail: feinberg@mcmaster.ca

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, nonmutually 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, 2000a, 2000b; Fitch & Giedd, 1999). 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 interesting 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.<sup>1</sup> 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.

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 ( $F_1$ – $F_3$ ) of the original voice recording.

<sup>1</sup> 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).

## 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; see also Supplemental Materials, Audio S1–S4). We did this to directly test whether voice pitch does, in fact, confound the accuracy of listeners' body size assessments. We also 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.

### Participants

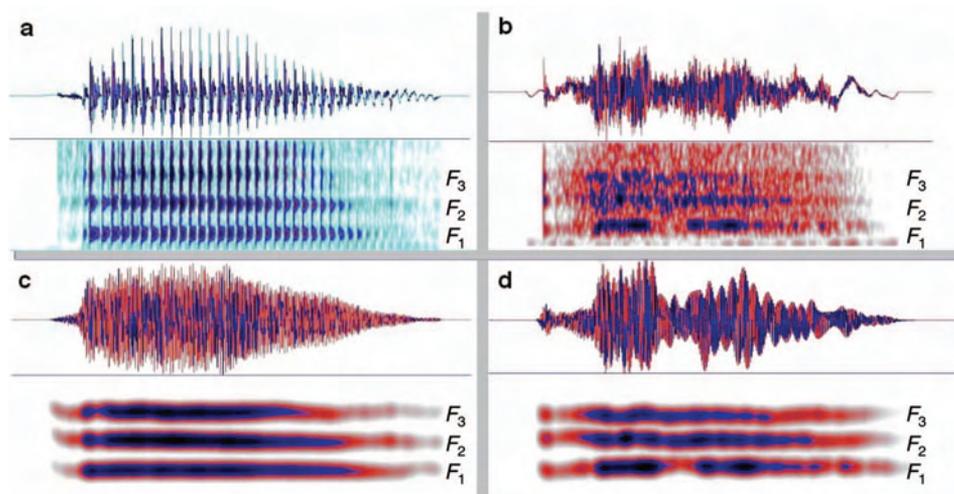
Seventy-seven women (age =  $18.8 \pm 1.7$  years) were recruited from the psychology undergraduate research pool at McMaster University in Hamilton, Ontario, Canada, to provide voice-based assessments of men's body size. All participants received partial course credit and provided informed consent. 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 A1 [in Appendix] for information regarding the vocalizer groups and pairs).

### Materials

**Male sample characteristics.** Thirty men (age =  $19 \pm 2.7$  years) 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 ( $M \pm SD$  height:  $178.9 \pm 6.3$  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; Hamdan 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  $7.42 \pm 5.58$  cm and ranged from 0–21 cm (Table A1). The acoustic properties of men's voices (see Voice Measurement and Table A2) 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, /a/ as in “father,” /i/ as in “see,” /e/ as in “bet,” /o/ as in “note,” and /u/ as in “boot” (to listen, access Supplemental Materials, Audio S1 and 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



**Figure 1.** Four types of speech stimuli. Amplitude waveforms (top of each panel) and broadband spectrograms (bottom of each panel) illustrating the vowel /e/ (spoken by a stimulus male, age 18, 187 cm tall, modal /e/  $F_0 = 100$  Hz,  $F_1$ – $F_3 = 706, 1809, 2848$  Hz) for (a) modal speech, (b) whispered speech, (c) modal SWS (synthesized from a), and (d) whispered SWS (synthesized from b). 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.  $F_1$ – $F_3$  represent the first three formants. All speech stimuli used in Experiments 1–3 consisted of the five English monophthong vowels, /a/, /i/, /e/, /o/, and /u/. See also Supplementary Materials, Audio S1–S4. The color version of this figure appears in the online article only.

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 A2 and A3 provide summary statistics of pitch measures.

For both modal and whispered speech, formants  $F_1$ – $F_4$  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  $F_1$  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  $F_1$ – $F_3$  values for synthesized modal and whispered SWS were equivalent to those of the corresponding natural speech. Table A2 provides summary statistics of formant  $F_1$ – $F_4$  measures.

In addition to  $F_1$ – $F_4$ , 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  $i$ th formant in Hz. Average formant frequency,  $F_n$  (Pisanski & Rendall, 2011) is given by:

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

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

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

where  $i$  refers to the formant number and  $c$  is the speed of sound in a uniform tube with one end closed,  $c = 35,000$  cm/s. Formant position ( $P_f$ ; Puts et al., 2012) is given by:

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

where  $F'_i$  is the standardized  $i$ th formant. Formant dispersion,  $D_f$  (Fitch, 1997; Fitch & Giedd, 1999) is given by:

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

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

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

Like previous work (Jovičić, 1998), we found that the absolute values of  $F_1$ – $F_4$  were in some cases higher for whispered than for modal vowels (Table A2). 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  $F_1$ – $F_4$  values relative to the shorter male, and relative differences in  $F_1$ ,  $F_2$ , and  $F_4$  were not statistically different between modal and whispered speech (Table A4).

**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  $3.46 \pm 0.41$  and  $3.59 \pm 0.37$  s in duration, respectively. A copy of each modal and whispered voice stimulus was additionally resynthesized into three-formant SWS in Praat (Boersma & Weenink, 2013; 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 ( $F_1$ – $F_3$ ) 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 Supplemental Materials, Audio S3 and 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 pseudorandomly 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 A1 summarizes the means and ranges of height differences between men. Height differences did not differ significantly across the four groups, one-way analysis of variance (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 /a/, /i/, /ε/, /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.,  $\leq 0.5$  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 which 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:  $W = 0.971$ ,  $p = .074$ ; 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,  $t(76) = 3.24$ ,  $p = .002$  (Figure 2a). After controlling for multiple comparisons, size assessment accuracy remained significantly above chance for modal speech, whispered speech, and whispered SWS (Bonferroni correction,  $\alpha/n = 0.0125$ , 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.

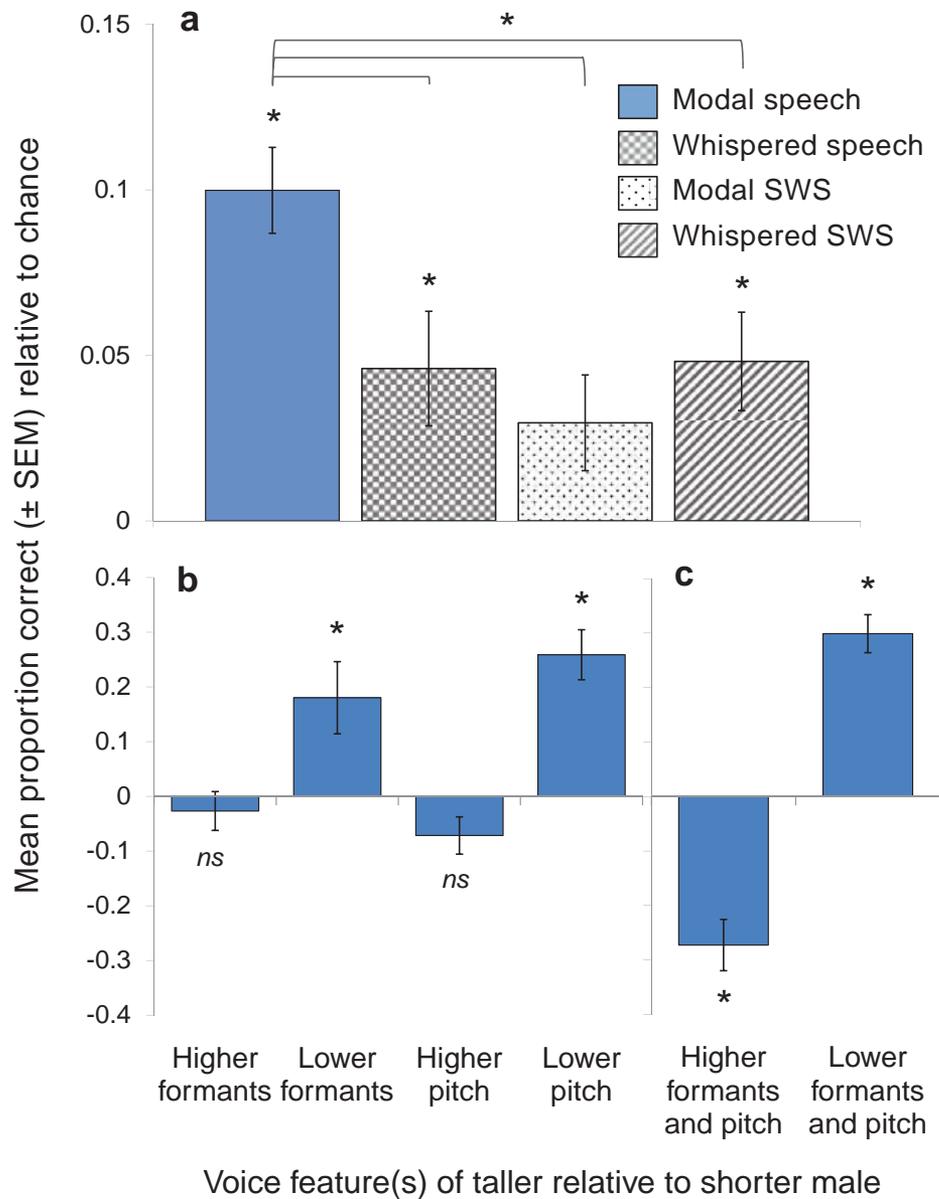
If it were true that voice pitch confounds accurate size assessment, listeners' accuracy would be expected to be higher for synthesized SWS and natural whispered speech, both of which are

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, 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 A3).

Regression analyses showed that accuracy generally improved as the difference in height between men increased (Figures 3 and 4). Linear regression indicated that size assessment accuracy increased with the difference in height between male vocalizers for modal speech,  $F(1,51) = 4.486$ ,  $p = .039$ ,  $\beta = 0.284$ ,  $R^2 = 0.081$ ; whispered speech,  $F(1,50) = 5.714$ ,  $p = .021$ ,  $\beta = 0.320$ ,  $R^2 = 0.10$ ; and modal SWS conditions,  $F(1,51) = 6.323$ ,  $p = .015$ ,  $\beta = 0.332$ ,  $R^2 = 0.11$ . The linear regression slope for whispered SWS was not significant,  $F(1,50) = 1.471$ ,  $p = .231$ ,  $\beta = -0.169$ ,  $R^2 = 0.03$  (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 (Figure 3) would be steeper for modal speech than for whispered speech or SWS. This is because, 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 ( $\beta = 0.284$ ), whispered ( $\beta = 0.320$ ), and modal SWS ( $\beta = 0.332$ ) conditions.

We also 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  $\pm SE = 0.20 \pm 0.007$ ,  $p = .005$ ), whispered speech ( $0.25 \pm 0.007$ ,  $p = .001$ ), and modal SWS ( $0.016 \pm 0.007$ ,  $p = .028$ ), but not for whispered SWS ( $0.003 \pm 0.008$ ,  $p = .73$ ; Figure 4).

Noting the linear shape of these psychometric functions (Figure 4), we can see that the effect of men's relative height on listeners' accuracy was effectively constant across height differences. In addition, 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

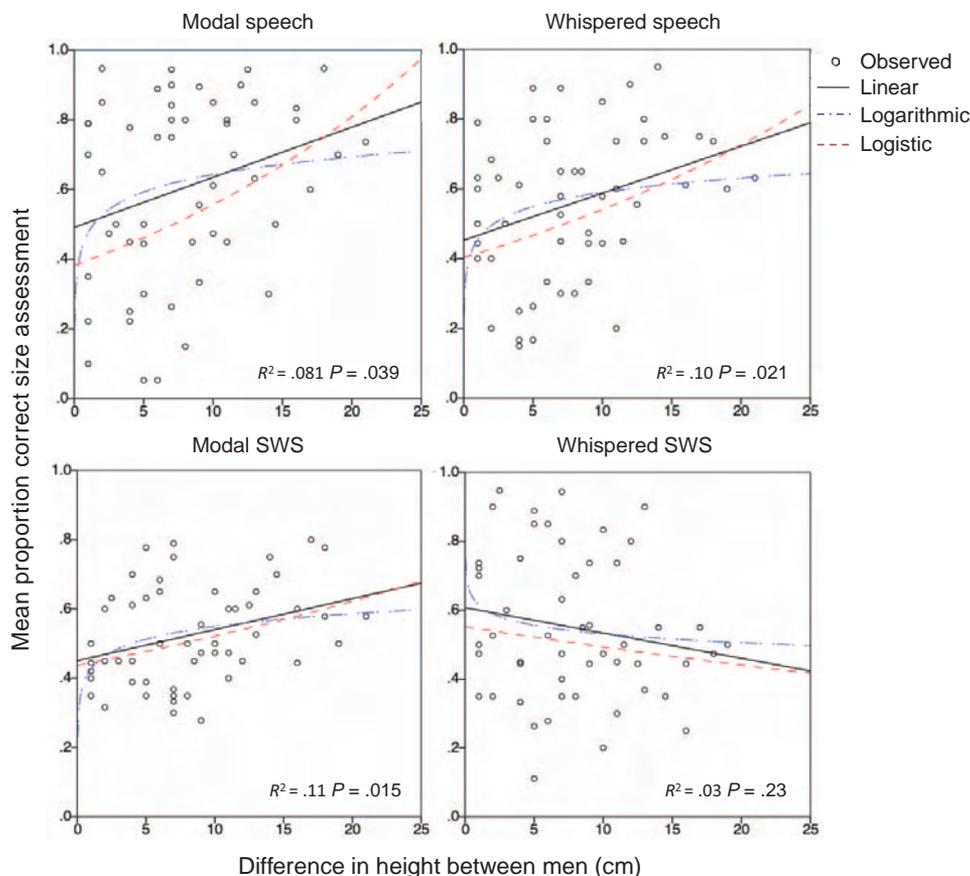


**Figure 2.** Accuracy in listeners' body size assessments (Experiment 1). Mean proportion correct size assessment ( $\pm$  SEM) as a function of (a) speech type ( $^* p < .0125$  Bonferroni correction, one-sample  $t$  tests comparing accuracy for each speech type against chance, and paired-sample  $t$  tests comparing accuracy among speech types; all tests two-tailed,  $n = 77$  listeners) 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 ( $^* p < .05$ ;  $ns$ ,  $p > .05$ , two-tailed one-sample  $t$  tests; see Table 1 for additional details). 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. The color version of this figure appears in the online article only.

(Pearson goodness-of-fit tests for probit: modal:  $\chi^2 = 71.02$ ,  $p < .001$ ; whispered:  $\chi^2 = 58.94$ ,  $p < .001$ ; modal SWS:  $\chi^2 = 22.35$ ,  $p = .27$ ; whispered SWS:  $\chi^2 = 30.2$ ,  $p = .036$ ).

**The effects of men's relative formants and voice 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 ( $F_n$ , as given by Equation 1); apparent VTL (Equation 2); formant position ( $P_f$ , Equation 3); formant dispersion ( $D_f$ , Equation 4); geometric MFF (Equation 5); fundamental frequency ( $F_0$ ); and perceived voice pitch in semitone, mel, and ERB scales (see Voice Measurement for additional details). Formant



**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. 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  $0.075$ ) 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 for modal speech, whispered speech, and modal sine-wave speech (SWS) conditions (best fitting model: linear regression,  $p < .05$ ). The slope for whispered SWS was not statistically significant ( $p = .23$ ). The color version of this figure appears in the online article only.

measures were significantly correlated with one another (two-tailed bivariate regressions:  $|r|$  modal =  $0.56$ – $0.999$ , all  $p < .01$ , Table A5), as were voice pitch measures ( $|r| = 0.993$ – $0.999$ , all  $p < .01$ , Table A6). Regardless of the measure used, formants predicted relative height better than did voice pitch (Table A3).

Nevertheless, listener's size assessment accuracy with modal speech was predicted both by differences in men's formants, ANOVA,  $F_n$ :  $F(1,53) = 24.95$ ,  $p < .001$ , 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,  $F_n + F_0$ :  $F(1,53) = 41.64$ ,  $p < .001$ . 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,  $F_n$ :  $t(52) = 2.76$ ,  $p = .01$ ;  $F_0$ :  $t(52) = 5.85$ ,  $p < .001$ . 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).

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

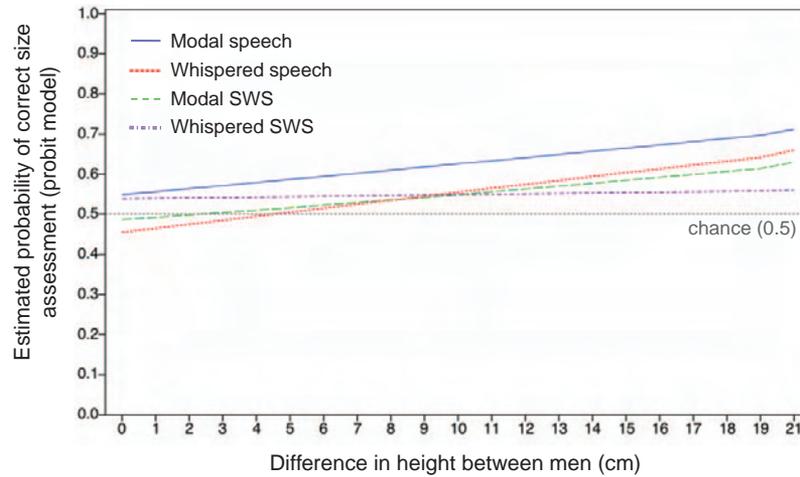


Figure 4. Probit inverse cumulative distribution functions fitted to listeners' size assessment accuracy scores (Experiment 1) as a function of men's relative height (taller–shorter male) for each speech type. Estimated probability of correct size assessment increased with the difference in height between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

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.

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 <sup>b</sup> ( $M \pm SD$ )	$df^c$	$t^c$	$p^c$
Independent effects <sup>a</sup>				
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

<sup>a</sup> Effect of one voice feature while controlling for the other. Formant measure =  $F_n$ , mean formant frequency (see Equation 1); Pitch measure =  $F_0$ , mean fundamental frequency. <sup>b</sup> Proportion correct size assessment ( $M \pm SD$ s) based on a total of 54 vocalizer pairs whose difference in height exceeded 0.5 cm. <sup>c</sup> Two-tailed one-sample  $t$  tests against chance (0.50).

## Participants

For Experiment 2, a new group of 40 women (age =  $19.38 \pm 2.62$  years) and 18 men (age =  $21.17 \pm 5.15$  years) 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 men) 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 (Lee et al., 1999; Peterson & Barney, 1952; Pisanski & Rendall, 2011). 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 vs. 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 counterbalanced 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 /a/, /i/, /e/, /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 ( $M \pm SD = 75 \pm 29\%$ ) and whispered speech ( $71 \pm 29\%$ ), 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's 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.

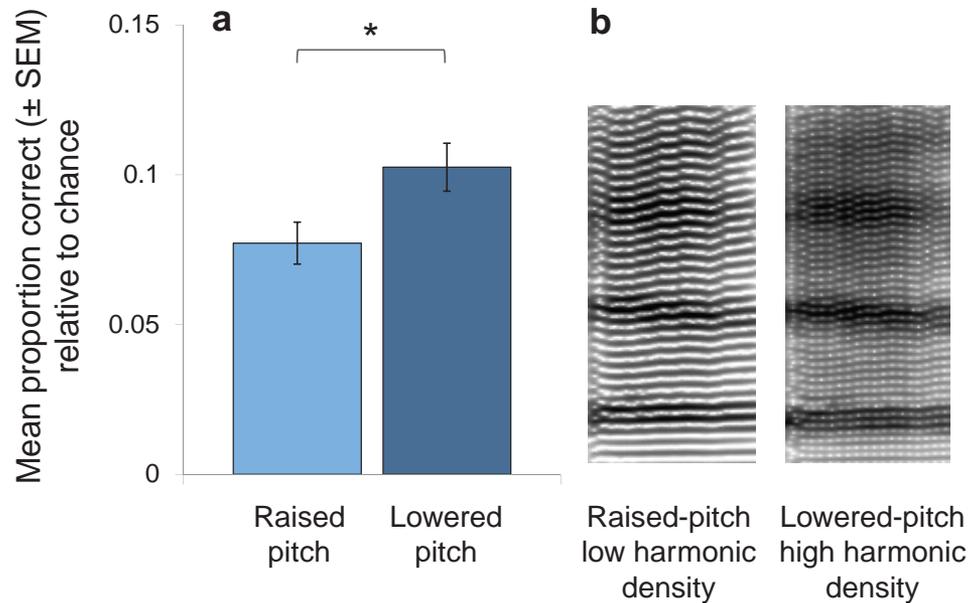
Consistent with the psychoacoustic work of Irino et al. (2012), the absence of voice pitch did not hinder listeners' ability to extract size information from the formant frequencies of the voice. Thus, the absence of pitch cannot explain the results of Experiment 1.

### 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 2, b 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 =  $20 \pm 2.6$  years) or lowered-pitch condition ( $n = 60$ , age =  $19 \pm 2.19$  years).



**Figure 5.** Harmonic density hypothesis. (a) Mean proportion correct ( $\pm$  SEM) 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. 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 analysis of variance [ANOVA],  $n = 120$ ). (b) Narrowband spectrograms depicting a stimulus male's voice (vowel / $\epsilon$ /) with raised pitch ( $F_0 = 122$  Hz) and lowered pitch ( $F_0 = 83$  Hz). Lower voice pitch (i.e., denser harmonics) provided a better carrier signal from which to resolve formants and assess body size. The color version of this figure appears in the online article only.

## 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 (e.g., 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 ( $M \pm SD = 60.25 \pm 6.59\%$  correct) than in the raised-pitch condition where harmonics were sparser ( $57.72 \pm 5.93\%$ ); one-way ANOVA,  $F(1,118) = 4.893$ ,  $p = .029$ . Thus, as predicted, we found that harmonic density facilitated accurate voice-based size perception. 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  $F_1$ - $F_4$ ) 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 a periodic 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 ( $\sim 5\%$ ) for synthesized modal and synthesized whispered speech (Iriño et al., 2012). Whereas Iriño 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 that our work provides the first evidence to our knowledge 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  $F_1$  and  $F_2$  shift constantly in continuous speech within individuals, facilitating vowel perception, whereas  $F_3$  and  $F_4$  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,  $F_1$ - $F_3$  (Monahan & Idsardi, 2010; Sjerps, Mitterer & McQueen, 2011). Thus,  $F_4$  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  $F_1$ - $F_3$ . 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  $F_4$  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; Künnel, 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. It is interesting that 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 for low-pitched than for 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) would 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 nonlinear 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 voice pitch nevertheless aids listeners in 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*, 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*, 1054–1063. doi:10.1121/1.427115
- Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* (Version 5.2.15). [Software]. Retrieved from <http://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*, 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*, 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*, 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*, 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. doi:10.1007/s10329-006-0033-y
- Fant, F. (1960). *Acoustic theory of speech production*. The Hague, the Netherlands: Mouton.
- 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*, 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*, 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*, 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*, 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*, 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*, 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*, 277–287. doi:10.1016/S0095-4470(03)00049-4
- 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*, 215–250. doi:10.2466/pms.105.1.215-220
- Graddol, D., & Swann, J. (1983). Speaking fundamental frequency: Some physical and social correlates. *Language and Speech*, *26*, 351–366. doi:10.1177/002383098302600403
- Greisbach, R. (1999). Estimation of speaker height from formant frequencies. *International Journal of Speech Language and Law*, *6*, 265–277. doi:10.1558/sll.1999.6.2.265
- 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. 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 Otolaryngology*, *112*, 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*, *142*, 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*, 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*, 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, United Kingdom: Wiley-Blackwell. doi:10.1002/9781444395068.ch4
- Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, *46*, 117–125. doi:10.1159/000261832
- 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. doi:10.1121/1.381808
- 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*, 195–204.
- 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*, 117–128. doi:10.1016/S0003-9969(00)00108-4
- Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge, MA: Cambridge University Press. doi:10.1017/CBO9781139165952
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables* (Vol. 7). Thousand Oaks, CA: Sage.
- Majewski, W., Hollien, H., & Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males. *Phonetica*, *25*, 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*, 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*, *111*, 855–869. doi:10.1086/283219
- 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, A. N. Popper, 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. 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*, 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*, 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*, 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*, 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*, 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*, 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*, 1631–1634. doi:10.1121/1.388499
- Ryan, M. J. (1988). Constraints and patterns in the evolution of anuran acoustic communication. In B. Fritzsche (Ed.), *The evolution of the amphibian auditory system* (pp. 637–677). Hoboken, NJ: Wiley
- 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. doi:10.1098/rspb.2010.0769
- Shields, M., Gorber, S. C., Janssen, I., & Tremblay, M. S. (2011). Bias in self-reported estimates of obesity in Canadian health surveys: An update on correction equations for adults. *Health Reports*, *22*, 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*, 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*, 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*, 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*, 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*, 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*, 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*, 429–439. doi:10.1121/1.3596462

## Appendix

### Additional Data Tables

Table A1

*Height Differences Between Men in Vocalizer Pairs Used in Experiments 1 and 3*

Group	Male pairs	Height difference between men in pair		
		Mean $\pm$ SD (cm) <sup>a</sup>	SE (cm)	Range (cm)
1	1–15	7.87 $\pm$ 6.13	1.58	0–19
2	16–30	7.6 $\pm$ 6.22	1.61	0.5–21
3	31–45	7.06 $\pm$ 5.05	1.30	0–16
4	46–60	7.13 $\pm$ 5.37	1.39	0–18
All (1–4)	1–60	7.42 $\pm$ 5.58	0.72	0–21

<sup>a</sup> Relative height between men did not differ significantly across the four groups of pairs (one-way analysis of variance [ANOVA], two-tailed),  $F(3,56) = 0.067$ ,  $p = .997$ .

(Appendix continues)

Table A2  
*Sample Voice Characteristics: Men's Average Formant and Pitch Measures*

Vowel $n^a$	$F_1$ (Hz)	$F_2$ (Hz)	$F_3$ (Hz)	$F_4$ (Hz)	$F_0$ (Hz) <sup>b</sup>
Modal speech					
/a/	720 ± 63	1,228 ± 88	2,496 ± 195	3,600 ± 200	115 ± 13
$n$	30	30	30	30	30
/i/	297 ± 39	2,245 ± 179	3,030 ± 179	3,638 ± 197	114 ± 13
$n$	30	30	30	29	30
/ε/	579 ± 66	1,744 ± 142	2,535 ± 160	327 ± 212	111 ± 13
$n$	30	30	30	30	30
/o/	480 ± 46	974 ± 67	2,434 ± 157	3,273 ± 175	109 ± 12
$n$	30	30	30	30	30
/u/	338 ± 34	1,070 ± 159	2,246 ± 266	3,272 ± 187	108 ± 12
$n$	30	30	30	30	30
All <sup>c</sup>	483 ± 29	1,453 ± 78	2,549 ± 118	3,461 ± 132	112 ± 12
Whispered speech					
/a/	945 ± 77	1,435 ± 154	2,622 ± 143	3,584 ± 188	
$n$	30	30	30	30	
/i/	438 ± 77	2,387 ± 199	3,018 ± 222	3,763 ± 258	
$n$	13	28	27	28	
/ε/	766 ± 86	1,892 ± 185	2,653 ± 146	3,615 ± 189	
$n$	30	30	30	29	
/o/	702 ± 89	1,121 ± 96	2,582 ± 164	3,379 ± 200	
$n$	24	28	28	28	
/u/	438 ± 73	1,236 ± 219	2,469 ± 191	3,384 ± 219	
$n$	20	29	29	29	
All <sup>c</sup>	724 ± 104	1,617 ± 135	2,666 ± 133	3,548 ± 157	

*Note.*  $F_1$ – $F_4$  = first to fourth formant;  $F_0$  = fundamental frequency; Hz = hertz. Means ± SDs of the first four formants and the fundamental frequency (voice pitch) of each of five vowels measured from men's modal or whispered speech.

<sup>a</sup>  $n$  = number of men's voices included in calculating the mean voice measure. <sup>b</sup> Whispered speech does not contain  $F_0$ . <sup>c</sup> Averaged across all five vowels.

Table A3  
*Average Measures of Formant Structure and Voice Pitch, and the Degree to Which Each Predicted Relative Height Among Vocalizer Pairs*

	$M$	$SD$	Predictive of relative height <sup>a</sup>
Formants			
$F_n$	1,986.4 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)

*Note.*  $F_n$  = 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 and SDs of voice measures taken from  $n = 30$  men.

<sup>a</sup> 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.

(Appendix continues)

Table A4  
Average Formant Differences in Modal and Whispered Speech Between Men in Vocalizer Pairs Used in Experiments 1 and 3

Formant	Speech type	Mean difference <sup>a</sup> ± SD (Hz)	<i>t</i> <sup>b</sup>	<i>p</i> <sup>b</sup>
<i>F</i> <sub>1</sub>	Modal	-3.56 ± 36.25	1.486	.143
	Whispered	-25.1 ± 143.4		
<i>F</i> <sub>2</sub>	Modal	-38.43 ± 89.32	0.069	.945
	Whispered	-40.22 ± 177.48		
<i>F</i> <sub>3</sub>	Modal	-23.39 ± 168.78	2.316	.024*
	Whispered	-80.95 ± 166.77		
<i>F</i> <sub>4</sub>	Modal	-79.76 ± 173.99	-0.436	.665
	Whispered	-68.31 ± 218.85		

Note. Hz = hertz; *F*<sub>1</sub>-*F*<sub>4</sub> = first to fourth formant. Means ± SDs 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.

<sup>a</sup> Differences in men's *F*<sub>1</sub>-*F*<sub>4</sub> within each pair were calculated by subtracting the *F*<sub>1</sub> of the shorter male in the pair from the corresponding *F*<sub>1</sub> of the taller male in the pair, such that mean differences below 0 reflect relatively lower *F*<sub>1</sub> in the taller male. <sup>b</sup> Results of paired sample *t* tests (*df* = 53, two-tailed) indicate that men's relative *F*<sub>1</sub>, *F*<sub>2</sub>, and *F*<sub>4</sub> were no different for modal and whispered speech. Differences in *F*<sub>3</sub> between men were greater for whispered than for modal speech but were in the predicted direction (the taller man had lower *F*<sub>3</sub> 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 with modal speech in Experiment 1.

\* *p* < .05, two-tailed paired sample *t* test.

Table A5  
Relationships Among Five Different Measures of Formant Structure Taken From Men's Modal or Whispered Speech

Speech type	Formant measure	Modal				Whispered				
		VTL	<i>P</i> <sub>f</sub>	<i>D</i> <sub>f</sub>	MFF	<i>F</i> <sub>n</sub>	VTL	<i>P</i> <sub>f</sub>	<i>D</i> <sub>f</sub>	MFF
Modal	<i>F</i> <sub>n</sub>	-0.909**	0.925**	0.847**	0.93**	0.467**	-0.401*	0.447*	0.555**	0.335†
	VTL		-0.998**	-0.575**	-0.999**	-0.433*	0.391*	-0.420*	-0.421*	-0.345†
	<i>P</i> <sub>f</sub>			0.601**	0.994**	0.440*	-0.394*	0.427*	0.439*	0.319†
	<i>D</i> <sub>f</sub>				0.56**	0.468**	-0.404*	0.445*	0.610**	0.350†
	MFF					0.43**	-0.392*	0.42*	0.42*	0.347†
Whispered	<i>F</i> <sub>n</sub>						-0.969**	0.998**	0.456*	0.958**
	VTL							-0.979**	-0.306†	-0.992**
	<i>P</i> <sub>f</sub>								0.410*	0.973**
	<i>D</i> <sub>f</sub>									0.237

Note. *F*<sub>n</sub> = average formant frequency; VTL = apparent vocal-tract length; *P*<sub>f</sub> = formant position; *D*<sub>f</sub> = 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* < .1. \* *p* < .05. \*\* *p* < .01, two-tailed bivariate Pearson correlations.

Table A6  
Relationships Among Four Different Measures of Voice Pitch Taken From Men's Modal Speech

Pitch measure	<i>F</i> <sub>0</sub>	Semitones	Mel	ERB
<i>F</i> <sub>0</sub>		0.988**	0.999**	0.998**
Semitones			0.993**	0.995**
Mel				0.999**
ERB				

Note. ERB = equivalent rectangular bandwidth; *F*<sub>0</sub> = fundamental frequency. Pearson correlation coefficients (*r*) are given in each cell, *n* = 30 men per cell.

\*\* *p* < .01, two-tailed bivariate Pearson correlations.

Received August 3, 2013  
Revision received January 29, 2014  
Accepted March 19, 2014 ■