CrossMark

Low is large: spatial location and pitch interact in voice-based body size estimation

Katarzyna Pisanski[1,2,3] · Sari G. E. Isenstein[1] · Kelyn J. Montano[1] ·
Jillian J. M. O'Connor[1,4] · David R. Feinberg[1]

Abstract The binding of incongruent cues poses a challenge for multimodal perception. Indeed, although taller objects emit sounds from higher elevations, low-pitched sounds are perceptually mapped both to large size and to low elevation. In the present study, we examined how these incongruent vertical spatial cues (up is more) and pitch cues (low is large) to size interact, and whether similar biases influence size perception along the horizontal axis. In Experiment 1, we measured listeners' voice-based judgments of human body size using pitch-manipulated voices projected from a high versus a low, and a right versus a left, spatial location. Listeners associated low spatial locations with largeness for lowered-pitch but not for raised-pitch voices, demonstrating that pitch overrode vertical-elevation cues. Listeners associated rightward spatial locations with largeness, regardless of voice pitch. In Experiment 2, listeners performed the task while sitting or standing, allowing us to examine self-referential cues to elevation in size estimation. Listeners associated vertically low and rightward spatial cues with largeness more for lowered-

than for raised-pitch voices. These correspondences were robust to sex (of both the voice and the listener) and head elevation (standing or sitting); however, horizontal correspondences were amplified when participants stood. Moreover, when participants were standing, their judgments of how much larger men's voices sounded than women's increased when the voices were projected from the low speaker. Our results provide novel evidence for a multidimensional spatial mapping of pitch that is generalizable to human voices and that affects performance in an indirect, ecologically relevant spatial task (body size estimation). These findings suggest that crossmodal pitch correspondences evoke both low-level and higher-level cognitive processes.

Crossmodal sensory perception is fundamental for developing and calibrating normal multisensory integration (Seilheimer, Rosenberg, & Angelaki, 2014, for a review). Although typically studied as a low-level perceptual phenomenon, multisensory integration is likely to involve both low-level and high-level neurocognitive mechanisms (Campanella & Belin, 2007; Seilheimer et al., 2014). Parise (2016) defines crossmodal correspondences as mappings between pairs of cues that are neither fully redundant nor completely unrelated, such as the association between auditory pitch and visual size. In this framework, systematic mappings across sensory cues within a single modality (i.e., audition) constitute a form of cue integration that is qualitatively similar to that of cue integration across modalities (Parise, 2016).

✉ David R. Feinberg
feinberg@mcmaster.ca

[1] Department of Psychology, Neuroscience, & Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada

[2] Mammal Vocal Communication & Cognition Research Group, School of Psychology, University of Sussex, Brighton, UK

[3] Institute of Psychology, Wroclaw, Poland

[4] Department of Psychology, University of Guelph, Guelph, Ontario, Canada

Crossmodal pitch correspondences

Although auditory pitch is mapped to a wide range of dimensions (see Eitan & Timmers, 2010, for a review), two of the key crossmodal correspondences that characterize the perception of pitch are most relevant for the present study. The first involves an association between pitch and size, in which low pitch is mapped to large size, and the second involves an association between pitch and vertical spatial mapping, in which low pitch is mapped vertically low in space (Eitan & Timmers, 2010). Both crossmodal correspondences are general and robust.

Low pitch is associated with large physical size and high pitch with small size regardless whether the sounds are pure or complex tones, or even human and animal vocalizations (Feinberg, Jones, Little, Burt, & Perrett, 2005; Morton, 1977; Ohala, 1984; Parise & Spence, 2012; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; see also Bien, ten Oever, Goebel, & Sack, 2012, and Gallace & Spence, 2006, for pitch–size correspondences in synesthetes). Moreover, pitch–size correspondences have been observed across distinct cultures (Ohala, 1984; Ultan, 1978) and in infants as young as four months of age (Peña, Mehler, & Nespor, 2011; Roffler & Butler, 1968; Walker et al., 2010), and are known to affect performance in a variety of perceptual, cognitive, and attention tasks (see Marks, 2000; Spence & Deroy, 2013, for reviews). Pitch–size associations even manifest themselves semantically in many languages and in musical discourse, wherein pitch is regularly described relative to size using words such as *heavy* and *light* or *thick* and *thin* (Ashley, 2004; Dolscheid, Shayan, Majid, & Casasanto, 2013).

In Western, English-speaking culture, low-pitched sounds are perceived as originating vertically lower in physical space, whereas relatively high-pitched sounds are mapped to higher elevations, commonly known as frequency-elevation mapping (Eitan & Timmers, 2010; Evans & Treisman, 2010; Mudd, 1963; Parise, Knorre, & Ernst, 2014). In fact, the perceived spatial location of pure tones is determined more by their pitch than by their spatial elevation, such that low-pitched tones projected from high elevations are perceived as originating low to the ground (i.e., the Pratt effect; Pratt, 1930). Pitch–space associations also manifest themselves in language and musical discourse, wherein pitch is described as *high* and *low*, or as *rising* and *falling*, and have been proposed to originate from a general "up is more" perceptual bias (Cox, 1999; Lakoff & Johnson, 1980). However, low-pitched sounds appear to affect spatial perception more than do relatively high-pitched sounds (Eitan & Granot, 2006; Eitan & Timmers, 2010). This indicates an asymmetry in frequency–elevation mapping, wherein "low is less" may prevail over "up is more" for pitch-based spatial elevation cues.

As compared to the well-known vertical mapping of low pitch to spatial lowness, few studies have examined whether

pitch is mapped along the horizontal (left–right) axis. In many regions of the world, including North America, numbers increase in magnitude from left to right and text is read from left to right (Shaki, Fischer, & Petrusic, 2009). Indeed, many studies have shown that people respond faster with the left hand when presented with small numbers, and faster with the right hand when presented with larger numbers (i.e., the SNARC effect; Campbell & Scheepers, 2015; Dehaene, Bossini, & Giraux, 1993; Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015; see Wood, Willmes, Nuerk, & Fischer, 2008, for a meta-analysis). This correspondence generalizes to other ordinal sequences, such as months and letters (Gevers, Reynvoet, & Fias, 2003), suggesting that in many contexts, "right is more." Moreover, people tend to demonstrate an orthogonal association in various spatial tasks, simultaneously mapping left to low and right to high (i.e., the SRC effect; see Cho & Proctor, 2003, for a review).

Studies testing whether pitch is mapped horizontally indicate that low pitch is most often associated with the left (Stewart, Walsh, & Frith, 2004; Timmers & Shen, 2016; Weis, Estner, & Lachmann, 2016; Weis, Estner, Van Leeuwen, & Lachmann, 2016), sometimes with the right (e.g., in nonmusicians, Stewart et al., 2004), and sometimes with neither side (or any apparent effect fails to reach statistical significance; Eitan & Granot, 2006; Eitan & Timmers, 2010; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006; Timmers & Shen, 2016 in nonmusicians). In addition to small effect sizes and variable methodologies, these mixed results for the horizontal axis might also be due in part to differences in the musical experience of participants, and pianistic experience in particular (Rusconi et al., 2006; Stewart et al., 2004; Timmers & Shen, 2016), although here too the findings of past studies are mixed. For instance, in Eitan and Timmers's word-matching task, musical experience did not affect participants' horizontal pitch correspondences. In contrast, Stewart et al. (2004) found that pianists responded faster than nonpianists in a musical Stroop task that required a leftward response to stimuli presented vertically low, whereas nonmusicians showed the reverse response. Timmers and Shen (2016) found horizontal mappings only in participants with musical training (pianists, in particular), whereas Weis and colleagues (Weis, Estner, & Lachmann, 2016; Weis, Estner, van Leeuwen, & Lachmann, 2016) found that participants mapped low tones with left and high tones with right, regardless of musical experience. Thus, the influence of musical training on horizontal pitch biases remains unclear.

It is important to note that in the present study, we tested whether listeners' judgments of body size differ for voices with high versus low pitch when originating from the right versus left. This task, although likely tapping into both spatial and frequency pitch mappings, is unique from this earlier corpus of work in which participants' responses were unrelated to size perception, making direct comparison difficult.

Crossmodal pitch correspondences in voice perception

The majority of previous work examining crossmodal pitch correspondences has used pure tones. In the present study, we examined correspondences among pitch, size and space using human voices with manipulated vocal pitch. As compared to tonal pitch, voice pitch is a complex broadband signal (Titze, 1994) and is selectively processed in higher-level regions of the auditory cortex near the superior temporal sulcus (Belin, Fecteau, & Bédard, 2004; Belin, Zatorre, & Ahad, 2002; Pernet et al., 2015). In addition, as a key indicator of various physical characteristics of the speaker such as sex and age, voice pitch plays an important role in human social interactions (Kreiman & Sidtis, 2011).

Previous studies have shown that pitch–size correspondences also manifest in voice perception tasks. Listeners consistently rate men and women with relatively lower-pitched (and/or more resonant) voices as physically larger than individuals with higher-pitched voices (Pisanski & Bryant, 2017, for review).[1] However, no previous study has examined perceptual correspondences between voice pitch and spatial cues. This is despite the suggestion that the "up is more" bias may originate in part from our experience with vocal production, wherein a rise in voice pitch is often accompanied by increases in vocal tension and effort (Cox, 1999).

Crossmodal incongruences in the mapping of pitch to size versus space?

The question of whether "up is more" applies to voice pitch is of particular interest in the context of magnitude or body size estimation, as it poses a potential paradox. In the context of true spatial cues to physical body size, voices of taller individuals will project from spatially higher elevations than those of shorter individuals. However, at the same time, listeners associate low voice pitch with tallness and largeness. This presents a perceptual incongruence between associations of *low* and *large* in the auditory domain (low-pitched objects or persons are perceived as large) and associations of *high* and *large* in the spatial domain (spatially high objects are perceived as large). The present study was designed to test how acoustic pitch and space cues interact to affect size perception, and specifically, which crossmodal correspondence prevails when incongruent pitch cues (low is large) and spatial cues (high is large) to physical body size are placed in direct conflict.

Although no previous study has examined whether the spatial location from which a sound (including a voice) is projected affects judgments of the size of the source, Eitan and Timmers (2010) used a conceptual task to test whether people associate high spatial locations with perceived largeness. The authors asked participants to pair the antonyms small and large with their concept of either high or low elevation, and found that 77% of participants associated spatially high objects with large size. This finding is in line with the "up is more" bias. At the same time, however, participants in their study also associated low pitch with concepts of large size, replicating the common "low is large" pitch bias. Eitan and Timmers (2010) concluded that people associate pitch with spatial verticality and with size in different, contrasting ways, suggesting that high pitch corresponds to "more" for spatial elevation, but to "less" for mass and size. The authors suggest that pitch perception therefore involves two contrasting magnitude representations, wherein low pitch can represent both more, and less.

Here we tested for the first time how these conflicting pitch correspondences interact in size perception. To do this we used an experimental paradigm that required listeners ($N = 112$) to estimate the body size of unseen men and women from their voices. Vocal stimuli were pitch-manipulated (raised or lowered in pitch) and projected from a high versus low, and right versus left, spatial location. On the basis of the previous literature, we predicted that lowered-pitch voices would generally elicit judgments of larger body size than would raised-pitch voices. However, we further hypothesized that, when placed in conflict with spatial location cues, pitch height cues might override spatial height (i.e., elevation) cues in the perception of body size. Thus, we predicted that listeners would associate vertically high spatial cues with largeness for voices with raised pitch, but would associate vertically low spatial cues with largeness for those same voices with lowered pitch. This prediction was based largely on studies replicating the Pratt effect (Pratt, 1930), wherein auditory pitch cues override auditory spatial cues in a variety of sound localization tasks (see, e.g., Bregman & Steiger, 1980; Cabrera, Ferguson, Tilley, & Morimoto, 2005; Morimoto & Aokata, 1984; Roffler & Butler, 1968; Rusconi et al., 2006; Trimble, 1934). Although our task was not a direct task of sound localization, we predicted that a similar bias would characterize listeners' performance in an indirect vertical spatial task involving the estimation of size.

---

[1] Interestingly, although taller individuals have lower-pitched voices than do shorter individuals when comparing adults with children or men with women (see González, 2006, for review), voice pitch explains less than 2% of the variance in height or weight when sex and age are controlled (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). As a result, the perceptual correspondence between low pitch and large body size has perplexed researchers for many years. Morton (1977) and Ohala (1984) suggested that low voice pitch evolved as a display of threat and dominance, and thus that humans and other animals are tuned to associate voice pitch with traits such as physical size. Alternatively, Rendall, Vokey, and Nemeth (2007) proposed that the perceptual association between low voice pitch and large body size may arise because listeners generalize broader pitch–size correspondences (i.e., large inanimate objects typically produce lower-frequency sounds) to within-sex judgments of body size, for which the associations are comparatively weak.

Ours was also the first study to test whether the left or the right is more strongly associated with size as assessed from the voice, and whether pitch affects these potential horizontal spatial cues to size. Along the horizontal axis, we predicted that listeners would generally associate rightward spatial locations with largeness more than leftward locations, due to common ordinal magnitude effects (i.e., the SNARC effect). However, unlike the vertical axis, we did not expect pitch and spatial cues to interact in size estimations for voices projected on the left versus the right, there is no clear incongruence or conflict between pitch–size and pitch–space correspondences along this horizontal axis.

Experiment 1

Method

Participants

Ten participants provided voice recordings for use as stimuli in Experiments 1 and 2 (mean age: 18 ± 0.3 years; five males, five females). Forty-six different participants took part in Experiment 1 as raters (mean age: 19.5 ± 1.6 years, all female). All participants were recruited from the psychology undergraduate research pool at McMaster University, provided informed consent, and received partial course credit for their participation.

Auditory stimuli

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

The voice stimuli averaged 3.44 s in duration (range 2.05 to 3.87 s). The pitch of each stimulus was raised or lowered by 10% from baseline using the Pitch-Synchronous Overlap Add (PSOLA) algorithm in Praat version 5.2.15 (Boersma & Weenink, 2013). This resulted in two versions (raised-pitch and lowered-pitch) of each original voice. The PSOLA method alters one voice feature (e.g., voice pitch) while leaving other features unaltered (Moulines & Charpentier, 1990). Our pitch manipulation corresponded to approximately two times the just-noticeable difference in voice pitch perception from a series of vowel sounds (Pisanski & Rendall, 2011; Re, O'Connor, Bennett, & Feinberg, 2012), as well as body size perception from voice pitch (Smith & Patterson, 2005). Pitch-manipulated men's voices ($M \pm SD$, raised-pitch: 122 ± 4 Hz;

lowered-pitch: 99 ± 3 Hz) and women's voices (raised-pitch: 238 ± 1 Hz; lowered-pitch: 194 ± 1 Hz) spanned the natural ranges in voice pitch for each sex (Titze, 1989). The sound pressure level (SPL) of each of the 20 stereo identically channeled voice stimuli was amplitude-normalized to 70 dB using the root-mean-square method, and each voice stimulus was panned 100% left or right in Praat.

Sound-speaker array and playback Four sound speakers (19-cm Bose, Companion 2 series II multimedia speaker system, Canada) were positioned parallel on the wall in a symmetrical array (see Fig. 1), within a dedicated, empty room with university-issued in-ceiling fluorescent lighting and no windows, located in the Voice Research Lab (room dimensions approximately 6 × 5 m). The sound speakers were not covered or occluded, although participants always faced away from them when listening to the voices. The room was quiet, but there was no sound treatment. When participants were seated, the center of the array was positioned directly adjacent the center of the participant's head at a distance of 213 cm (7 feet) and 0° elevation and azimuth, such that voices projected along the vertical axis would be perceived as coming from above or below (16° elevation, 0° azimuth) and voices projected along the horizontal axis would be perceived as coming from the left or the right (16° azimuth, 0° elevation) of the participant's head. Head position and elevation was standardized using a mounted chinrest. The four-sound-speaker symmetrical array allowed us to manipulate the spatial location of voices along one axis while holding the other constant. This design limited localization cues to either the azimuth or elevation planes, reducing the potential for sound source confusion (Middlebrooks & Green, 1991), and ensured that the distance from the participant and the sound source was the same for all four sound speakers.

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

Sound-pressure-level check Sound localization errors are minimized (3–5°) when the SPL of auditory stimuli is 70 dB or higher (Davis & Stephens, 1974). Thus, a sound-level meter (Brüel & Kjær, Type 2239, Denmark) was used to test the free-field SPL of the voice stimuli projected from each of the four sound speakers at the location of the chinrest. The average free-field SPL was 71.02 dB. Because louder sounds may be perceived as lower in pitch (Davis & Stephens, 1974; Wier, Jesteadt, & Green, 1977) and as originating from a larger source (Walker, 1987), we confirmed that no differences in free-field SPL were perceivable (i.e., differences were
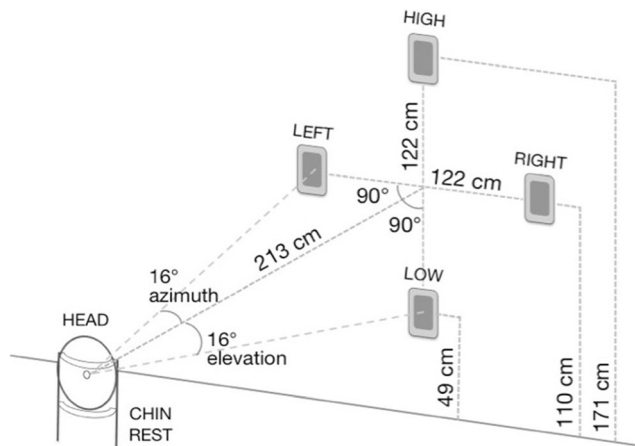
Fig. 1 Four-sound-speaker array.

<1 dB) between the lowered-pitch (70.94 dB) and raised-pitch (71.1 dB) voices.

Procedure

The McMaster Research Ethics Board approved the experiment. Each participant completed the same protocol. Before beginning the experiment, participants' heads were positioned in a chinrest and their stool height was adjusted. Once comfortable, participants were instructed to leave their head in the chinrest for the entirety of the experiment. They were then instructed that they would hear a series of voices and that their task would be to rate the body size of each person speaking on a scale from 1 (*very small*) to 7 (*very large*), and to input this response into the computer using the numeric keypad of the keyboard. The experimenter then initiated the experimental protocol on the computer and left the room.

Auditory stimuli were projected directly from the speakers mounted on the wall (free-field). Participants were presented with a single voice on each trial, and a blank computer screen on which the question "How large is this person?" was positioned above a textbox in which the response (1–7) could be inputted. Each participant completed 40 trials in each block, and voices were blocked by axis. In the *vertical-axis* block, each voice stimulus (ten raised-pitch, ten lowered-pitch) was projected once from the high and once from the low sound speaker, and in the *horizontal-axis* block, each voice stimulus was projected once from the left and once from the right sound speaker. The experimenter implemented one channel change manually between blocks, which participants did not observe. The experimenters were not visible or audible during testing. Block order was counterbalanced between participants, and the presentation of voice stimuli within each block was randomized, including sound-speaker position. Participants inputted their size judgments and their sex and age using a custom computer interface.

Following previous studies that had examined body size estimation using manipulated voice stimuli (e.g., Charlton, Taylor, & Reby, 2013; Pisanski, Oleszkiewicz, & Sorokowska, 2016; Rendall et al., 2007), participants rated the apparent "largeness" of the person whose voice they heard on each trial, and they were not instructed as to which aspect of body size (height, weight) they should focus on during this task. This method allowed for our results to be more directly comparable to those of previous work examining body size estimation, but also to studies examining pitch–size associations more broadly (see, e.g., Ohala, 1984, and Parise, 2016, for reviews).

Statistical analysis

We first ran an omnibus mixed-design repeated measures analysis of variance (ANOVA) in which the dependent variable was the body size estimate (coded 1–7), and the within-subjects factors included Pitch Manipulation (raised, lowered), Axis (vertical, horizontal), Sound-Speaker Position (high, low; left, right) and Sex of Voice (male, female). Average body size estimates were calculated separately for each participant. We then created contrast variables and used planned *t* tests to examine all significant effects revealed by the omnibus model. All tests were two-tailed with an alpha of .05.

Results

The significant relationships revealed by the omnibus ANOVA are summarized in Table 1. No other main or interaction effects were significant or marginal. The main effects of the omnibus model indicated that, all else being equal, lowered-pitch voices were judged as larger ($M \pm SEM$ = 4.42 ± 0.08) than raised-pitch voices (3.5 ± 0.08), men's voices were judged as larger (4.39 ± 0.09) than women's voices (3.52 ± 0.09), and voices projected from the low and right sound speakers were judged as larger (4.00 ± 0.08) than voices projected from the high and left sound speakers (3.91 ± 0.08). However, the omnibus model revealed several interactions that qualified these main effects (Table 1) and that we examined using post-hoc tests.

We examined the three-way interaction among pitch manipulation, axis, and sound-speaker position by calculating the difference in body size assessments between lowered- and raised-pitch voices at each sound-speaker location, averaging across sexes of voice, and compared these differences between sound speakers located along the vertical versus the horizontal axes. As is illustrated in Fig. 2, pitch manipulations differentially affected body size assessments for voices projected from the low versus the high sound speaker [$t(45)$ = 2.88, p = .006]. This pattern was observed for both men's

Table 1 Omnibus ANOVA for Experiment 1

Source	$F(1, 45)$	η_p^2	p
Pitch manipulation	414.9	.90	<.001
Sex of voice	76.5	.63	<.001
Sound-speaker position	9.8	.18	.003
Pitch Manipulation × Sex of Voice	26.5	.37	<.001
Pitch Manipulation × Axis × Sound-Speaker Position	8.5	.16	.005
Sex of Voice × Axis × Sound-Speaker Position	6.2	.12	.017

and women's voices [no effect of sex of voice: $t(45) = 0.89$, $p = .40$]. Thus, regardless of the sex of the voice, listeners associated vertically low spatial cues with largeness when assessing the size of lowered-pitch voices [$t(45) = 2.94$, $p = .005$] but not when assessing raised-pitch voices [$t(45) = −1.03$, $p = .31$]. In contrast, body size assessments between the left and right sound speakers did not vary as a function of voice pitch [$t(45) = –1.22$, $p = .23$]. Again, this pattern was observed for both men's and women's voices [no effect of sex of voice: $t(45) = 1.1$, $p = .28$]. Thus, regardless of the sex of the voice, listeners associated horizontally rightward spatial locations with largeness for both lowered- and raised-pitch voices [$t(45) = 3.94$, $p <.001$]. These effects were robust to Bonferroni correction for multiple comparisons.
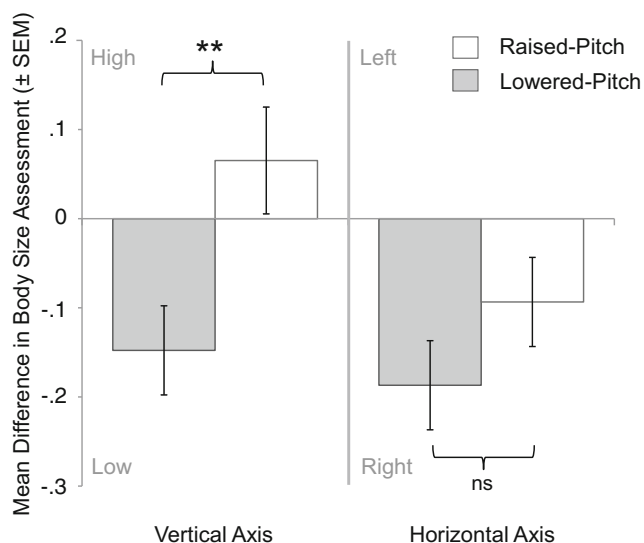

Fig. 2 Interaction between voice pitch and spatial location on body size assessments (Exp. 1). Listeners associated vertically low spatial cues with largeness (values below 0) for lowered-pitch but not for raised-pitch voices, and associated horizontally rightward spatial cues with largeness (values below 0) regardless of the voice pitch manipulation. Bars represent the mean differences ± standard errors of the means in body size assessments between sound speakers located along the vertical (high minus low) or the horizontal (left minus right) axes. Bar coloration represents the size estimates for lowered-pitch voices (gray bars) and raised-pitch voices (white bars), averaging across sexes of the voices. $^{**}p < .01$; ns = $p > .05$

We also observed a significant three-way interaction among sex of voice, axis, and sound-speaker position. To examine this interaction, we calculated the difference in body size assessments between men and women's voices at each sound-speaker location, averaging across pitch manipulations, and compared this difference between sound speakers located along the vertical versus the horizontal axis. There was a greater difference in how much larger men's voices sounded than women's when the voices were projected from the low versus the high sound-speaker location [$t(45) = 1.84$, $p = .073$], but this difference did not reach statistical significance.

Experiment 2

Parise et al. (2014) suggest that the frequency–elevation mapping of high pitch to high spatial locations stems in part from distal (environmental) information, as well as from proximal (self-referential) cues, because the spectra of sound reaching the ears depends in part on the elevation and orientation of the sound source relative to the head of the observer. Indeed, several previous studies reported that the relative position of a listener's head and torso to the sound stimulus can affect how pitch is mapped to space (e.g., Algazi, Avendano, & Duda, 2001; Middlebrooks & Green, 1991; Roffler & Butler, 1968). In their recent study, Carnevale and Harris (2016) demonstrated that people used ascending- and descending-pitch scales as spatial orientation cues, indicating which way was spatially "up" and which way was "down," respectively, when lying on their sides versus sitting upright. However, static high- or low-pitched tones (1200 vs. 200 Hz) had no effect on participants' perceptions of their own orientation.

In Experiment 2, we examined whether size assessments varied as a function of the head elevation of the listener (i.e., the elevation of the sound source relative to the head of the participant when the participant was standing vs. sitting). The spatial locations of the four sound speakers and the manipulated vocal pitch stimuli were identical in both the sitting and standing conditions; the only difference between these conditions was the position of the participant's head relative to the sound speakers. Thus, comparing between conditions allowed

us to examine the relative contributions of proximal (self-referential) and distal (environmental) cues to elevation on listeners' assessments of body size.

In addition, by asking participants to assess the body sizes of voices while standing, we introduced a spatial asymmetry between the listener and sound speakers along each axis (see Fig. 1). This allowed us to test for interactions between listeners' horizontal and vertical biases (i.e., an orthogonal stimulus–response compatibility effect; Cho & Proctor, 2003). Moreover, if the effects observed in Experiment 1 could be attributed largely to low-level perceptual mechanisms, the effects should be reduced when the listener was standing versus sitting. Finally, in this type of task, standing could introduce a difference in size estimations based on perceived social factors relevant to judging the body sizes of other people (e.g., potential rivals). For instance, Fessler and Holbrook (2013) found that men who were tied to a chair underestimated their own height and overestimated the heights of other men. Indeed, although this is controversial, it has been posited that bodily "power poses" affect people's perceptions and experiences of dominance and control (Carney, Cuddy, & Yap, 2010). Thus, standing could further alter size judgments via higher-level social constructs.

In Experiment 2, we additionally tested for an interaction between sex of voice and sound location on size assessments, the results of Experiment 1 indicated a marginally stronger association between low elevation and large size for men's than for women's voices. We also included both male and female listeners in Experiment 2 to rule out effects of the sex of the listener on body size assessments.

Method

Participants Sixty-six participants (48 females, mean age 18.6 ± 2 years; 18 males, mean age 18.7 ± 1.5 years) took part in Experiment 2 as raters, none of whom had participated in Experiment 1. All participants were recruited from the psychology undergraduate research pool at McMaster University, provided informed consent, and received partial course credit for their participation.

Procedure We used the same voice stimuli and experimental setup as in Experiment 1, with the following modifications. There were two between-subjects conditions: standing and sitting (i.e., head elevation). In neither condition did participants use the chinrest. When standing, participants placed their feet on designated markers on the floor to minimize positional differences across participants. When participants were seated, the chair was also placed on the same markers on the floor. We measured standing height (171.09 ± 12.97 cm) and sitting height (148.13 ± 19.15 cm), from both the floor and the seat of the chair to the top of the participant's head, using metric tape. Participants were randomly assigned to

either the sitting or the standing condition. As in Experiment 1, in both conditions here participants faced away from the speakers and could not see them during the experiment.

Also as in Experiment 1, participants were instructed that they would hear a series of voices and that their task would be to rate the body size of each person speaking on a scale from 1 (*very small*) to 7 (*very large*), using the numeric pad of a keyboard. The experimenter then initiated the experimental protocol on the computer and left the room.

Statistical analysis As in Experiment 1, we ran an omnibus repeated measures ANOVA in which the dependent variable was body size estimates (coded 1–7) and the within-subjects factors included Pitch Manipulation (raised, lowered), Axis (vertical, horizontal), Sound-Speaker Position (high, low; left, right), and Sex of Voice (male, female). The model for Experiment 2 additionally included the between-subjects factors Head Elevation (sitting, standing) and Sex of Listener (male, female). We examined any significant effects revealed by the omnibus model using contrast variables and post-hoc t tests.

Results

The significant relationships revealed by the omnibus ANOVA are summarized in Table 2. No other main or interaction effects were significant or marginal, including any effects of the sex of the listener. Thus, all subsequent analyses were collapsed across male and female raters.

The main effects of the omnibus model indicated that voices lowered in pitch were rated as larger ($M \pm SEM$ = 4.25 ± 0.07) than voices raised in pitch (3.80 ± 0.07), and men's voices were rated as larger (4.07 ± 0.11) than women's voices (3.97 ± 0.06), replicating the results of Experiment 1. Here we also found a main effect of head elevation on size assessments, wherein voices were generally rated as larger when participants were standing (4.17 ± 0.11) than when they were sitting (3.88 ± 0.07). The omnibus model revealed several interactions that qualified these main effects (Table 2) and that we examined using post-hoc tests. For brevity and clarity, here we provide the results of planned post-hoc tests examining only those effects for which we had specific predictions based on the results of Experiment 1. All other post-hoc analyses are provided in the supplementary online material.

We examined the four-way interaction among pitch manipulation, axis, sound-speaker position, and head elevation by calculating the difference in body size assessments between lowered- and raised-pitch voices at each sound-speaker location and axis, averaging across sexes of voice. We then calculated the difference in how much larger lowered-pitch voices were rated as compared to raised-pitch voices when they were projected from high versus low and from left versus right spatial locations, and

Table 2 Omnibus ANOVA for Experiment 2

Source	$F(1, 52)$	η_p^2	p
Pitch manipulation	178.49	.77	<.001
Sex of voice	145.38	.74	<.001
Head elevation	4.9	.09	.031
Pitch Manipulation × Sound-Speaker Position	58.97	.53	<.001
Pitch Manipulation × Sex of Voice	17.69	.25	<.001
Pitch Manipulation × Head Elevation	7.61	.13	.008
Sound-Speaker Position × Sex of Voice	5.21	.09	.027
Axis × Sound-Speaker Position	24.8	.32	<.001
Pitch Manipulation × Sound-Speaker Position × Sex of Voice	46.13	.47	<.001
Pitch Manipulation × Axis × Sex of Voice	8.05	.13	.007
Axis × Sound-Speaker Position × Sex of Voice	12.6	.20	.001
Pitch Manipulation × Axis × Sound-Speaker Position × Head Elevation	4.48	.08	.039
Sex of Voice × Axis × Sound-Speaker Position × Head Elevation	5.92	.10	.018

compared these differences between the sitting and standing conditions.

As is illustrated in Fig. 3, in both the sitting [$t(29) = −6.76$, $p < .001$; Fig. 3a] and standing [$t(29) = −4.58$, $p < .001$; Fig. 3b] conditions, we observed a greater difference in how much lowered-pitch voices sounded larger than raised-pitch voices when they were projected from the low than from the high sound speaker. There was no difference between the sitting or standing conditions [independent-samples t test, using Brown–Forsyth degrees of freedom with equal variances not assumed: $t(51.534) = −1.67$, $p = .10$; Levine's test for equality of the variances: $F = 5.64$, $p = .021$]. Moreover, both when sitting [$t(29) = −4.66$, $p < .001$; Fig. 3a] and when standing [$t(29) = −5.91$, $p < .001$; Fig. 3b], a significantly greater difference emerged in how much larger lowered-pitch voices sounded than raised-pitch voices when they were projected from the right than from the left sound speaker. This effect was larger in the standing than in the sitting condition [independent-sample t test: $t(61) = 2.159$, $p = .035$]. The difference between the sitting and standing conditions, when comparing the aforementioned differences across the vertical versus horizontal axes, was also significant [$t(61) = −3.35$, $p = .001$].

To examine the four-way interaction among sex of voice, axis, sound-speaker position, and head elevation, we calculated the difference in body size assessments between men's and women's voices at each sound-speaker location and axis, averaging across pitch manipulations. We then calculated the differences in size assessments between men and women when voices were projected from high versus low and from left versus right spatial locations, and compared these differences between the sitting and standing conditions.

When standing [$t(29) = 5.50$, $p < .001$] but not when sitting [$t(29) = 1.31$, $p = .201$], there was a greater difference in how much larger men's voices sounded than women's when they were projected from the low than from the high sound-speaker location. The difference between the sitting and standing conditions was significant [$t(61) = −2.92$, $p = .005$]. In neither the sitting [$t(29) = 0.14$, $p = .201$] nor the standing [$t(29) = 1.149$, $p = .148$] condition did a greater difference emerge in how much larger men's voices sounded than women's when they were projected from the right than from the left sound-speaker location. The difference between the sitting and standing conditions was not significant [$t(61) = 1.14$, $p = .26$]. The difference between sitting or standing, when we compared the aforementioned differences across the vertical and horizontal orientations, was significant [$t(61) = −2.84$, $p = .006$].

Discussion

In everyday perception, we are faced with the challenge of integrating multiple and often incongruent cues. One prime example is the integration of incongruent auditory pitch cues to size and to spatial location, wherein high pitch is mapped to "more" for spatial elevation, but to "less" for mass and size (Eitan & Timmers, 2010). Here we examined how voice pitch (lowered vs. raised), the spatial location of the sound source (high vs. low, left vs. right), and the spatial location of the listener (head elevation) interact to influence assessments of size. Rather than using tonal pitch, we manipulated the pitches of men and women's voices, and we utilized an ecologically relevant task of body size estimation.
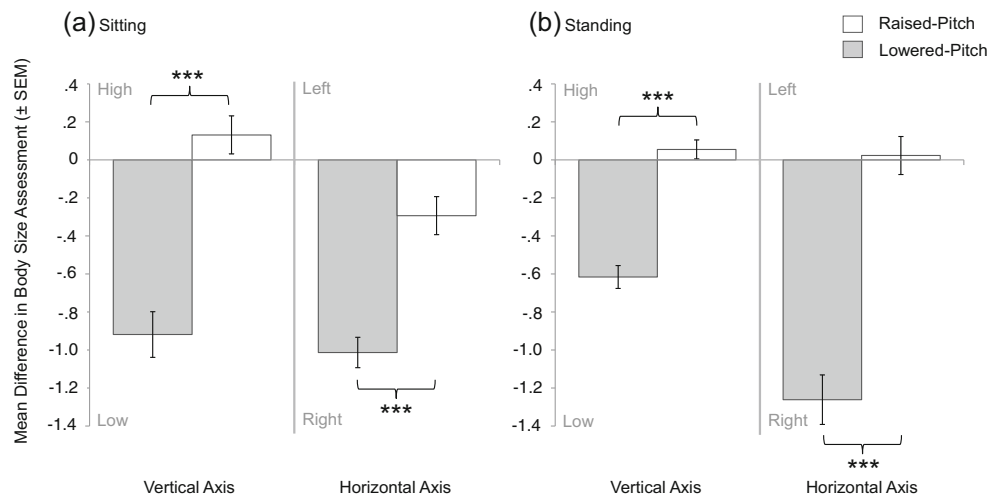
Fig. 3 Interaction between voice pitch and spatial location on body size assessments in the sitting (a) and standing (b) conditions (Exp. 2). When either sitting or standing, listeners associated vertically low and horizontally rightward spatial cues with largeness (values below 0) for lowered-pitch but not for raised-pitch voices. Comparing between the sitting and standing conditions, we found no difference between conditions for size estimates of voices presented along the vertical axis. However, along the horizontal axis, there was a stronger association between rightward spatial cues and largeness for lowered-pitch than for raised-pitch voices when listeners were standing then when they were sitting. Bars represent the mean differences ± standard errors of the means in body size assessments between sound-speaker locations on the vertical or horizontal axis for lowered-pitch (gray bars) and raised-pitch voices (white bars), averaging across sexes of both the voices and listeners. $^{***}p < .001$

Summary of results

Experiment 1 revealed that, regardless of the sex of the voice, listeners associated vertically low spatial locations with physical largeness when rating voices that were lowered in pitch, but not when rating those same voices that were raised in pitch. In contrast, listeners associated horizontally rightward spatial locations with largeness for all voices, both those lowered and those raised in pitch (Fig. 2). These results indicate that lower-pitched voices projected low and close to the Earth were perceived as belonging to larger people. This correspondence between low and large emerged despite the a priori probability for humans (as opposed to nonhuman objects; see Parise et al., 2014) that a larger individual's voice will *always* project from relatively higher in space than that of the smaller individual when two individuals are standing.[2]

This key result was replicated in Experiment 2, in which a different group of participants completed an analogous task while either sitting (as in Exp. 1) or standing. Here, regardless of the sex of a voice and regardless of the listeners' head elevation relative to the sound source, listeners once again associated vertically low spatial cues with largeness more for lowered- than for raised-pitch voices (Fig. 3). Thus, frequency–elevation mapping for human voices in particular does not appear to follow from the observation of statistical regularities linking low-frequency sounds to objects close to the earth, as is often the case for nonvocal sounds (Parise et al., 2014).

The results of Experiment 2 further revealed that listeners associated horizontally rightward spatial locations with largeness more for lowered- than for raised-pitch voices. This effect was found for both men's and women's voices, and both when listeners were sitting and standing. However, Experiment 2 revealed that head elevation can affect some aspects of size estimation. Only when standing did listeners rate men's voices as larger than women's when the voices were projected from a low spatial location. This is unlikely to have been due to the fact that men's voices are typically lower than women's, as we would have then expected to see this effect in Experiment 1 and when participants were seated. We suggest that perhaps higher-level social constructs are at play here (e.g., those described by Carney et al., 2010; Fessler & Holbrook, 2013). Experiment 2 also revealed a stronger association between rightward spatial cues and largeness for lowered-pitch than for raised-pitch voices when listeners were standing versus sitting (Fig. 3).

Interpretation and implications

Our results build upon the classic Pratt effect (Pratt, 1930), demonstrating that pitch affects vertical elevation perception not only in a direct sound localization task (e.g., Bregman & Steiger, 1980; Morimoto & Aokata, 1984; Roffler & Butler, 1968; Trimble, 1934), but also in an indirect spatial task involving size estimation. Our findings suggest that the crossmodal correspondence between *low pitch* and *largeness*

---

[2] In other words, we suggest that it is highly unlikely that in the real world people have developed Bayesian priors based on an observed statistical regularity, whereby taller people constantly bend over to speak with their heads lowered to the earth.

is relatively stronger than the incongruent correspondence between *low spatial location* and *smallness*. Although associating low pitch with low space can lead to errors in body size estimation (people with low-pitched voices are generally taller than people with high-pitched voices, and do not speak with their heads low to the earth), our study does not necessarily refute the hypothesis that frequency–elevation mapping is functionally adaptive within a broader context (Parise, Knorre, & Ernst, 2014; Stumpf, 1883; Walker et al., 2010). Indeed, the mapping of high-pitched sounds to high spatial locations may be tuned to the statistics of natural auditory scenes (see Parise et al., 2014). Thus, frequency–elevation mapping appears to be generally adaptive, even though it is not useful in assessing human body size.

Ours was the first study to test for a left–right bias in size perception and a potential interaction between horizontal spatial cues and pitch height on size judgments. Listeners consistently associated the right with large size, and in Experiment 2 this association was strongest for lowered-pitch voices. Horizontal size estimates might reflect semantic or numeric coding. For example, large numbers are associated with rightward responses in various cognitive tasks (Campbell & Scheepers, 2015; Dehaene et al., 1993; Dehaene et al., 2015; Shaki et al., 2009; Wood et al., 2008). However, the interactions between head elevation (standing or sitting) and voice pitch manipulation observed in Experiment 2 suggest that our findings are not the result of simple numeric-size mapping. Indeed, there is no a priori prediction that numerical order would affect size judgments differently when people sit versus stand, or for assessments of lowered- versus raised-pitch voices. Moreover, participants used a numeric pad to report body size, on which the arrangement of digits was neither ascending nor descending in magnitude, but rather was ordered 7–8–9 (top row), 4–5–6 (center row), and 1–2–3 (bottom row). If our experimental results were merely due to numerical mapping, we would have also expected that sounds from the low speaker would be rated smaller than sounds from the high speaker, since the smallest numbers were lowest on the keypad and the largest numbers were highest on the keypad. This did not happen. Therefore, another interesting possibility is that the right–large correspondence reflects hemispheric specialization that develops through experience. This is supported by evidence of reversed horizontal biases in people from cultures that read from right to left (Maass & Russo, 2003) and in piano players (Stewart et al., 2004).

In Experiment 2, participants completed the size estimation task while either sitting or standing. This introduced an asymmetry in the spatial paradigm that allowed us to test for interactions between listeners' horizontal and vertical biases (when standing, voices projected from the high sound-speaker location were now closer to the listener's head than were voices projected from the low sound speaker). We were also able to examine the contribution of self-referential elevation cues to listeners' assessments of body size. We found that this asymmetry had no effect on listeners' size estimates along the vertical plane. Thus, the results of Experiment 2 demonstrate that the interaction between low pitch and low elevation on size perception is robust, unaffected by the head elevation of the listener or the relative degree of spatial lowness. Carnevale and Harris (2016) similarly found no effect of pitch-based auditory spatial cues on people's perceptions of upright orientation when they were lying down versus sitting. We argue that the insensitivity of crossmodal pitch correspondences to changes in head position provides evidence that the vertical correspondence involves some degree of higher-level cognitive processing, and supports the general ubiquity of pitch–size and frequency–elevation correspondences (Parise, 2016). In contrast, the interaction of low pitch and rightward spatial location on assessments of size was stronger in the standing than sitting condition. When participants were standing rather than sitting, voices projected from the right originated from a low spatial location relative to the participant's head (i.e., right and low, rather than just right). Thus, when standing, the perceptual biases linking low pitch to both low and rightward spatial locations may have additively combined to exaggerate listeners' size estimates along the horizontal plane. The independent "low is large" and "right is large" biases observed in our study may therefore have an addictive effect on size perception that is similar to the orthogonal (stimulus–response compatibility; Cho & Proctor, 2003) effects in spatial location tasks.

The results of Experiment 2 also showed stronger associations between low elevation and large size for assessments of men's than of women's voices; however, this only occurred when raters were standing. Although listeners may associate vertically low spatial cues with physical largeness more for men's than for women's voices because men's voices are almost twice as low in pitch (Titze, 1989)—and may be more readily associated with dominance and masculinity, which often map onto perceptions of body size (Pisanski, Mishra, & Rendall, 2012)—this cannot explain why this association was only present when participants were standing. One possibility is that standing introduces an added social dimension to body size estimation—for instance, related to dominance. Indeed, Fessler and Holbrook (2013) showed that visual estimates of men's body size are also sensitive to the body position of the rater (i.e., men are visually assessed as taller by raters who are strapped to a chair).

In both Experiments 1 and 2, experimentally lowering voice pitch affected size estimates more than did raising pitch, indicating an asymmetry in pitch–size correspondences. Past studies had also reported strong perceptual associations between low pitch and low elevation, but weak or no associations between high pitch and high elevation (Eitan & Granot, 2006; Eitan & Timmers, 2010). This provides additional support in refutation of the *directional symmetry hypothesis*

(Eitan & Granot, 2006), and suggests that low frequencies may elicit stronger crossmodal correspondences than do high frequencies. This pitch asymmetry also suggests that our findings cannot be explained by low-level interactions (e.g., that low-pitched sounds transmit better from lower than from higher space; Morton, 1977).

Limitations and future directions

When using manipulated vocal stimuli, most studies, including the present study, have asked participants to assess the body size of speakers on "largeness," without biasing listeners to focus on either height or weight (e.g., Charlton et al., 2013; Rendall et al., 2007). This is of course also the case for studies examining pitch–size mapping more broadly—for instance, between tones and inanimate objects (Parise, 2016, for a review), for which a height/weight distinction is less sensible. Measuring assessments of largeness rather than of height or weight allows for comparisons between these two literatures; however, it also poses the possibility that the different participants in our study relied on different markers of size (e.g., height, weight, or a combination of the two) to gauge the largeness of speakers. Although this possibility cannot explain our findings, which were based on *within-participant* variation in size judgments across conditions, future studies may examine whether the reported effects of voice pitch and spatial location on body size perception are magnified when participants are specifically instructed to estimate a person's physical height (i.e., a direct verticality judgment).

Among humans and many other mammals, body size is more reliably communicated by vocal-tract resonances (formant frequencies) than by voice pitch (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). Previous work has shown that both low pitch and low formants are independently associated with perceptions of large size, but that they also interact in complex ways that affect size estimation (Feinberg, Jones, DeBruine, O'Connor, Tigue, & Borak, 2011; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Smith & Patterson, 2005). For instance, although pitch is only weakly related to body size among same-sex adults, low voice pitch increases the spectral density of a vocal signal and the saliency of formant frequencies, making it easier for listeners to estimate body size (Charlton et al., 2013; Pisanski et al., 2014). Future work may examine whether formant frequencies and voice pitch elicit similar frequency–elevation mappings, and whether spatial cues affect formant-based size estimation.

We did not record the musical training of participants. Although it is unclear how musical training might affect the mapping of pitch and spatial cues in the estimation of physical size, musical expertise is known to affect the spatial mapping of pitch along both the vertical and horizontal axes (see Lega, Cattaneo, Merabet, Vecchi, & Cucchi, 2014). Future studies may therefore test whether musical experience affects pitch

and/or spatial cues to body size. Studies could also test whether there are differences between musicians who play instruments for which left is low (such as the piano and guitar) versus instruments for which left is high (such as the flute and French horn). Similarly, for the vertical axis, differences may emerge between musicians who play the sitar and contrabass, for whom low frequencies are played by fingering at higher elevations, and musicians who play the clarinet or saxophone, for whom low frequencies are played by fingering at lower elevations. Replication studies may also include additional measures, such as reaction time analyses, handedness analysis, and brain imaging, that could allow for a more comprehensive understanding of the mechanisms driving pitch–size–space correspondences.

Conclusion

In summary, in our study, listeners associated vertically low elevations with largeness when assessing the body sizes of men and women with lowered- but not with raised-pitch voices. Listeners also associated horizontally rightward spatial locations with physical largeness, and this effect was stronger for lowered- than for raised-pitch voices. Voice pitch is structurally and functionally more complex than is tonal pitch. Yet our study demonstrates that pitch correspondences, typically studied with tones and music, generalize to human voice pitch. Moreover, whereas previous studies using tonal stimuli have used broad pitch ranges spanning 200–8000 Hz (see, e.g., Cabrera et al., 2005; Carnevale & Harris, 2016; Mudd, 1963; Parise et al., 2014; Pratt, 1930), we have demonstrated that frequency–elevation mapping is elicited with voice pitch manipulations of only 20–40 Hz at pitch centers of 100–250 Hz (although this degree of manipulation is still an order of magnitude larger than the just-noticeable differences for pitch detection in similar vocal stimuli; see Re et al., 2012). In addition, our results show that the influence of pitch on spatial perception goes beyond localizing sounds in space, but also affects performance in an indirect spatial task, and one that has ecological relevance. It is evident that pitch correspondences in spatial and size perception are likely to affect multisensory integration in a wide range of contexts, including when judging the body sizes of other people.

References

Algazi, V. R., Avendano, C., & Duda, R. O. (2001). Elevation localization and head-related transfer function analysis at low frequencies. *Journal of the Acoustical Society of America, 109,* 1110–1122.

Ashley, R. (2004). Musical pitch space across modalities: Spatial and other mappings through language and culture. In S. D. Lipscomb, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 64–71). Adelaide, South Australia: Causal Productions.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8,* 129–135. doi:10.1016/j.tics.2004.01.008

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13,* 17–26.

Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch–size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage, 59,* 663–672. doi:10.1016/j.neuroimage.2011.06.095

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

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

Cabrera, D., Ferguson, S., Tilley, S., & Morimoto, M. (2005, July). *Recent studies on the effect of signal frequency on auditory vertical localization.* Paper presented at the Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland.

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

Campbell, J. I. D., & Scheepers, F. (2015). Effects of pitch on auditory number comparisons. *Psychological Research, 79,* 389–400. doi:10.1007/s00426-014-0571-7

Carnevale, M. J., & Harris, L. R. (2016). Which direction is up for a high pitch? *Multisensory Research, 29,* 113–132.

Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science, 21,* 1363–1368.

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

Cho, Y. S., & Proctor, R. W. (2003). Stimulus and response representations underlying orthogonal stimulus response compatibility effects. *Psychonomic Bulletin & Review, 10,* 45–73. doi:10.3758/BF03196468

Cox, A. W. (1999). *The metaphoric logic of musical motion and space.* Unpublished Ph.D. dissertation, University of Oregon, Eugene.

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

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122,* 371–396. doi:10.1037/0096-3445.122.3.371

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron, 88,* 2–19. doi:10.1016/j.neuron.2015.09.019

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

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

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

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 6. doi:10.1167/10.1.6

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.

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

Fessler, D. M. T., & Holbrook, C. (2013). Bound to lose: Physical incapacitation increases the conceptualized size of an antagonist in men. *PLoS ONE, 8,* e71306. doi:10.1371/journal.pone.0071306

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68,* 1191–1203. doi:10.3758/BF03193720

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87,* B87–B95. doi:10.1016/S0010-0277(02)00234-2

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.

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception.* New York, NY: Wiley-Blackwell.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4,* 195–208.

Lega, C., Cattaneo, Z., Merabet, L. B., Vecchi, T., & Cucchi, S. (2014). The effect of musical expertise on the representation of space. *Frontiers in Human Neuroscience, 8,* 250. doi:10.3389/fnhum.2014.00250

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events nature or culture? *Psychological Science, 14,* 296–301. doi:10.1111/1467-9280.14421

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

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

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

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

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

Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tone. *Journal of Experimental Psychology, 66,* 347–352. doi:10.1037/h0040045

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica, 41,* 1–16.

Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research, 29,* 7–28. doi:10.1163/22134808-00002502

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

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association

test. *Experimental Brain Reseasrch, 220,* 319–333. doi:10.1007/s00221-012-3140-6

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

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., … Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, *119*, 164–174. doi:10.1016/j.neuroimage.2015.06.050

Pisanski, K., & Bryant, G. A. (2017). The evolution of voice perception. In N. S. Eidsheim & K. L. Meizel (Eds.), *The Oxford handbook of voice studies*. Oxford, UK: Oxford University Press.

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

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

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.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology, 13,* 278–285. doi:10.1037/h0072651

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

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

Roffler, S. K., & Butler, R. A. (1968). Localization of tonal stimuli in the vertical plane. *Journal of the Acoustical Society of America, 43,* 1260–1266.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition, 99,* 113–129.

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

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect.

Psychonomic Bulletin & Review, 16, 328–331. doi:10.3758/PBR.16.2.328

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.

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

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

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

Timmers, R., & Shen, L. (2016). Representation of pitch in horizontal space and its dependence on musical and instrumental experience. *Psychomusicology: Music, Mind, and Brain, 26,* 139–148. doi:10.1037/pmu0000146

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 vocal production*. Englewood Cliffs, NJ: Prentice-Hall.

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

Ultan, R. (1978). Size-sound symbolism. *Universals of Human Language, 2,* 525–568.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21,* 21–25. doi:10.1177/0956797609354734

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

Weis, T., Estner, B., & Lachmann, T. (2016). When speech enhances Spatial Musical Association of Response Codes: Joint spatial associations of pitch and timbre in nonmusicians. *Quarterly Journal of Experimental Psychology, 69,* 1687–1700. doi:10.1080/17470218.2015.1091850

Weis, T., Estner, B., van Leeuwen, C., & Lachmann, T. (2016). SNARC (spatial–numerical association of response codes) meets SPARC (spatial–pitch association of response codes): Automaticity and interdependency in compatibility effects. *Quarterly Journal of Experimental Psychology, 69,* 1366–1383. doi:10.1080/17470218.2015.1082142

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

Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly, 50,* 489–525.