UCLA
UCLA Previously Published Works

Title
Three-month-old human infants use vocal cues of body size.

Permalink
https://escholarship.org/uc/item/9j74p8nh

Journal
Proceedings. Biological sciences, 284(1856)

ISSN
0962-8452

Authors
Pietraszewski, David
Wertz, Annie E
Bryant, Gregory A
et al.

Publication Date
2017-06-01

DOI
10.1098/rspb.2017.0656

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Downloaded from http://rspb.royalsocietypublishing.org/ on June 8, 2017

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Pietraszewski D, Wertz AE, Bryant GA, Wynn K. 2017 Three-month-old human infants use vocal cues of body size. *Proc. R. Soc. B* **284**: 20170656.
http://dx.doi.org/10.1098/rspb.2017.0656

Received: 28 March 2017
Accepted: 5 May 2017

Subject Category:
Behaviour

Subject Areas:
behaviour, cognition, developmental biology

Keywords:
acoustic properties, body size, pitch, formants, development, human infants

Author for correspondence:
David Pietraszewski
e-mail: davidpietraszewski@gmail.com

†Present address: Max Planck Institute for Human Development, Center for Adaptive Rationality, Lentzeallee 94, 14195 Berlin, Germany.
‡Present address: Max Planck Institute for Human Development, Max Planck Research Group Naturalistic Social Cognition, Lentzeallee 94, 14195 Berlin, Germany.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3780926.

THE ROYAL SOCIETY PUBLISHING

Three-month-old human infants use vocal cues of body size

David Pietraszewski[1,3,†], Annie E. Wertz[2,3,‡], Gregory A. Bryant[4] and Karen Wynn[3]

[1]Center for Adaptive Rationality, and [2]Max Planck Research Group Naturalistic Social Cognition, Lentzeallee 94, 14195 Berlin, Germany
[3]Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520-8205, USA
[4]Department of Communication, Center for Behavior, Evolution, and Culture, University of California, Los Angeles, CA 90095, USA

DP, 0000-0002-8091-0674

Differences in vocal fundamental (F_0) and average formant (F_n) frequencies covary with body size in most terrestrial mammals, such that larger organisms tend to produce lower frequency sounds than smaller organisms, both between species and also across different sex and life-stage morphs within species. Here we examined whether three-month-old human infants are sensitive to the relationship between body size and sound frequencies. Using a violation-of-expectation paradigm, we found that infants looked longer at stimuli inconsistent with the relationship—that is, a smaller organism producing lower frequency sounds, and a larger organism producing higher frequency sounds—than at stimuli that were consistent with it. This effect was stronger for fundamental frequency than it was for average formant frequency. These results suggest that by three months of age, human infants are already sensitive to the biologically relevant covariation between vocalization frequencies and visual cues to body size. This ability may be a consequence of developmental adaptations for building a phenotype capable of identifying and representing an organism's size, sex and life-stage.

1. Introduction

Size matters. In most mammals, humans included, overall body size differences correspond to differences in age, sex and reproductive stage. Body size also tracks important fitness-relevant individual differences in somatic development and maintenance, including differences in health and formidability [1–5].

Size tends to covary with two important dimensions of mammalian vocalizations: average fundamental frequency (F_0) and average formant frequency (F_n). Source-filter theory [6] describes how these vocalizations are produced in mammals. Airflow from the lungs vibrates vocal folds housed in the larynx (a pair of soft-tissue folds stretched across the opening of the glottis) and that acoustic energy is then filtered by the vocal tract. Vocal fold vibrations produce the fundamental frequency (F_0) of a vocalization (perceived as pitch), whereas formants (F_n) are the resonating frequencies of the vocal tract that correspond to perceptions of voice quality.

Fundamental and formant frequencies are decoupled in most mammals, including humans, and each independently predicts body size to varying degrees, both within and between species and sexes [7,8]. This is because differences in fundamental and formant frequencies reliably differ across sex and age morphs—male from female, prepubescent from postpubescent, and pre- from post-menopausal [9,10].[1] In particular, formant values and spacing negatively correlate with the length of the vocal tract, which in turn scales with body size, particularly with height [13,14]. Pitch negatively correlates with the thickness and length of the vocal folds, which is due primarily to exposure to male secondary sexual hormones at puberty, particularly testosterone [15]. Consequently, much of the variation in pitch and formants in mammalian

© 2017 The Author(s) Published by the Royal Society. All rights reserved.

Downloaded from http://rspb.royalsocietypublishing.org/ on June 8, 2017

vocalizations, including in humans, is due to size differences between sex and age body morphs, and to hormone-induced secondary sexual maturation divergence at puberty [10,11].

Previous experimental evidence demonstrates that human adults are sensitive to these relationships, even within age and sex morphs [11,16–18]. For instance, experimentally lowering F_0 and formant dispersion in young men's voices increases the speaker's perceived masculinity, size and age [19]. Similar effects have been found with non-human vocalizations. For example, lowering of F_n and/or F_0 in dog barks increases attributions of the dog's aggressiveness in human listeners [20]. Playback studies with male deer vocalizations also show that male deer perceive the experimental lowering of F_n and F_0 as more threatening ([21]; see also [9,11,16,22,23] for more extensive reviews). Adult domestic dogs also correctly match the growls of larger versus smaller dogs onto visual cues of dog body size [24].

Here we examined if human infants are sensitive to this association between an organism's size and the sounds it produces—in particular, the relationship between its size and the average fundamental and formant frequencies of its vocalizations. Because this relationship is robust and recurrent across taxa, and presumably phylogenetic time, natural selection may have shaped developmental processes to capture this invariance relatively early in ontogeny, as a means of building a reliably developing phenotype capable of identifying and representing an organism's size, sex, and/or life-stage via acoustic as well as visual channels. Further, because transient fitness-relevant states and behavioural intentions—such as attack versus friendly approach—are often individual-specific and signalled through vocalization content [16,25,26], this is potentially also a mechanism for reducing uncertainty in matching vocalizations to the entities producing them [27].

Further, there is some evidence that three- to five-month-old infants can detect changes in some invariant relationships between acoustic and visual modalities, particularly when the co-relation is ecologically robust and recurrent. For example, infants are sensitive to the sight and sound of an object's impact [28–30] or an object's distance [31]. Previous research also shows that three-month-old infants are able to discriminate sounds based on pitch [28] and four-month-old infants map non-rigid transformations of object thickness and rigid transformations of height to differences in pitch [32]. Further, studies using preferential looking (giving infants an option to look at two different screens) demonstrate that infants prefer to look at matched visual/acoustic events between the height and pitch of a moving object consistent with the Doppler effect [32,33], at pairings between object size and vowel type (high frontal vowels with higher F_0 and formants versus low posterior vowels with lowered F_0 and formants [34]) and at matches between pitch and the expansion and contraction of objects [35].

These findings reveal infants' sensitivity to pitch, formant structure and certain co-relationships between visual and acoustic properties. However, it is not yet known whether infants are sensitive to the relationship between vocalization and body size, specifically that larger organisms tend to produce lower frequency sounds and smaller organisms tend to produce higher frequency sounds. Here, this question was examined using a violation-of-expectation (VOE) looking time paradigm [4,36,37] with three-month-old infants. In a VOE paradigm, infants are presented with different event types—some that are predicted to be consistent with infants' underlying expectations, and others that are predicted to be inconsistent—while their looking times are measured. If infants are indeed sensitive to the predicted relationship, this will be reflected in looking time differences, such that infants will exhibit longer looking to the inconsistent events.[2]

In the current experiment, infants were presented with two different sized organisms, each of which emitted either high or low frequency vocalizations. In half of the trials, the relationship between size and sound was consistent, such that the smaller organism produced higher frequency vocalizations and the larger organism produced lower frequency vocalizations. In the other half of the trials, the relationship between size and sound was inconsistent, such that the smaller organism produced lower frequency vocalizations and the larger organism produced higher frequency vocalizations. If infants are sensitive to the relationship between size and sound, then all else equal the inconsistent trials should evoke relatively longer looking times than the consistent trials.

2. Material and methods

(a) Experimental design

The vocalization frequency manipulated across event types varied between subjects: one group of infants heard differences in fundamental frequency (pitch; F_0), and the other group heard differences in average formant frequency (F_n). Consistency or inconsistency with the typical size/sound relationship was manipulated within subjects using a block design. Infants first saw a three trial block of one size/sound pairing (inconsistent or consistent) followed by a three trial block of the other size/sound pairing (consistent or inconsistent). The order of presentation was counterbalanced across participants.

(b) Participants

Thirty-two healthy, full term three-month-old infants participated (16 female; mean age three months, 21 days; range: three months, 0 days–four months, 18 days). Infants were recruited from the greater New Haven area and tested in the Infant Cognition Center at Yale University. Fourteen additional infants were excluded due to fussiness, eight due to procedural error, and one due to difficultly in determining looking direction (making online or offline data collection impossible). To ensure infants included in the analysis actually saw the experimental stimuli, an exclusion criterion was established *a priori* such that infants were only included if they watched at least two of each trial type (consistent and inconsistent), as determined offline by a blind coder (see *Coding* §2e(v) below). An additional five infants were excluded for failing to meet this criterion.

(c) Acoustic stimuli

An uncompressed wav file (16 bit, 44.1 kHz) of a baby goat bleat (downloaded from freesound.org) was digitally manipulated using the VTChange script (C. Darwin) in Praat (v. 5.3.1) [40]. A non-human vocal sample was used in order to maintain novelty in both the visual and auditory domains and avoid possible confusion associated with using a human voice for non-human puppets. F_0 values were altered using PSOLA (Pitch Synchronous Overlap Add) resynthesis, which maintains apparent vocal tract length (VTL). VTL is inversely correlated with the averaged distance between adjacent formants as well as absolute formant values. In other words, longer vocal tracts result in lower formant frequencies that are closer together, and shorter vocal tracts result in greater spacing between higher formants. Formant frequency alterations adjusting apparent VTL also change F_0 and duration values, which are then resampled

Downloaded from http://rspb.royalsocietypublishing.org/ on June 8, 2017

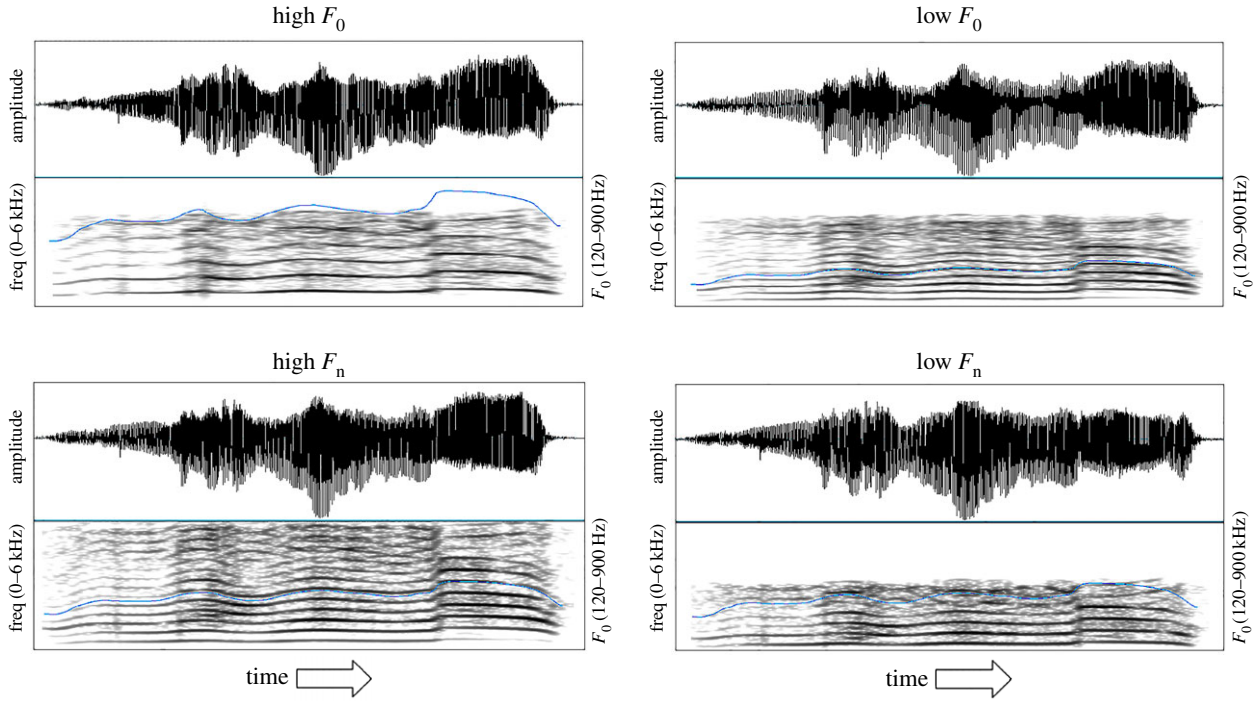

Figure 1. Four sample waveforms and wideband fast Fourier transform (FFT) spectrograms (5 ms window length, Gaussian analysis window shape, 44.1 kHz sampling rate) of all four manipulated baby goat bleats. For each image, the top panel represents the overall waveform amplitude and the bottom panel represents the spectrogram (0–6 kHz). Blue lines show F_0 values (120–900 Hz). All stimuli were 750 ms in length.

back to original values using PSOLA. From the original sound file, four versions were created: high pitch (150% of baseline F_0), low pitch (73.5% of baseline F_0), high formants/shorter apparent VTL (70% of baseline VTL) and low formants/longer apparent VTL (130% of baseline VTL). Manipulations on both dimensions were well above established adult just-noticeable differences (JNDs) in human voices, which are changes of approximately 6% in both F_0 and F_n in male and female speakers [18,41], but still within a range of realistic mammalian vocalizations [42]. Changes in VTL of 150% (which would be equivalent to the F_0 change) resulted in subsequent changes in F_0 beyond F_0 JNDs. For this reason, we reduced the degree to which we changed F_n in order to maintain independence in the manipulations. Importantly, this meant that somewhat greater F_0 variation relative to F_n was generated, which will result in a slightly larger effect of pitch to the extent that the responses to our task are proportional to the objective stimulus properties (an open question). See electronic supplementary material for acoustic properties of all stimuli as well as audio files. Figure 1 shows spectrograms of all four sound stimuli.

(d) Visual stimuli

Visual stimuli were two stuffed animal creatures identical in all features but their size (small animal: 10.2 (H), 7.3 (W), 30 (circumference, C) cm; large animal: 18 (H), 15.2 (W), 45.7 (C) cm; see electronic supplementary material) and two different-coloured fabric boxes with one open panel (27 (H), 26 (W), 26 (D) cm). The stimuli were presented approximately 90 cm from the infant (see figure 2).

(e) Procedure

Infants sat in an infant seat in front of a stage with a curtain at the far end (107 cm from the infant). A second, closer curtain (64 cm from the infant) was raised and lowered to show or occlude the stage (105 cm wide).

(i) Curtain and sound familiarization

After their child was seated facing the stage, parents stood to the side or behind their infant so the infant could not see them without turning its head. A sample of each of the two sound files was then played prior to the show.[3] This allowed the parent to experience what their infant would be hearing, and exposed the infant to the two different sounds before any additional visual information was paired with the sounds. The parent was then instructed to put on a pair of electronic noise-cancelling headphones (in which classical music was playing) and to leave these on for the duration of the show so that they could not hear the size/sound pairings played for their infant during testing. The stage curtain was then raised and lowered twice to familiarize the infant to the sight and sound of the curtain being raised and lowered. Each time the curtain was raised a rattle-sound oriented the infant's attention toward the stage.

(ii) Stimulus presentation

After the familiarization procedure, the show began. A puppet creature appeared at the centre of the stage moving toward the infant at the same time a sound was played (see top row of figure 2). Upon reaching the front of the display (in front of the boxes), the creature moved laterally while wiggling contingently with the intonation of the sound to suggest that the creature was vocalizing (see second row of figure 2). After four lateral movements (two towards each side of the stage) the creature moved to the front of the upturned box, which was immediately flipped over to occlude the creature.

The same sound (i.e. 'voice' of the creature) was then played while the box under which the creature had disappeared began to move contingently, suggesting that the creature was producing the sounds and moving the box (see third panel of figure 2). When this stopped, the second box began to move contingently with a new sound that was either higher or lower in frequency (either in F_0 or F_n, between-subjects), suggesting that a second—but as yet unseen—creature was under this second box and was also vocalizing, but with a different 'voice' (see fourth row of figure 2). Each vocalization/box movement event happened three times for each box, for a total of six events. Both 'under-the-box' sounds had different manipulated frequencies, but were otherwise identical. At the termination of the sixth

Downloaded from http://rspb.royalsocietypublishing.org/ on June 8, 2017

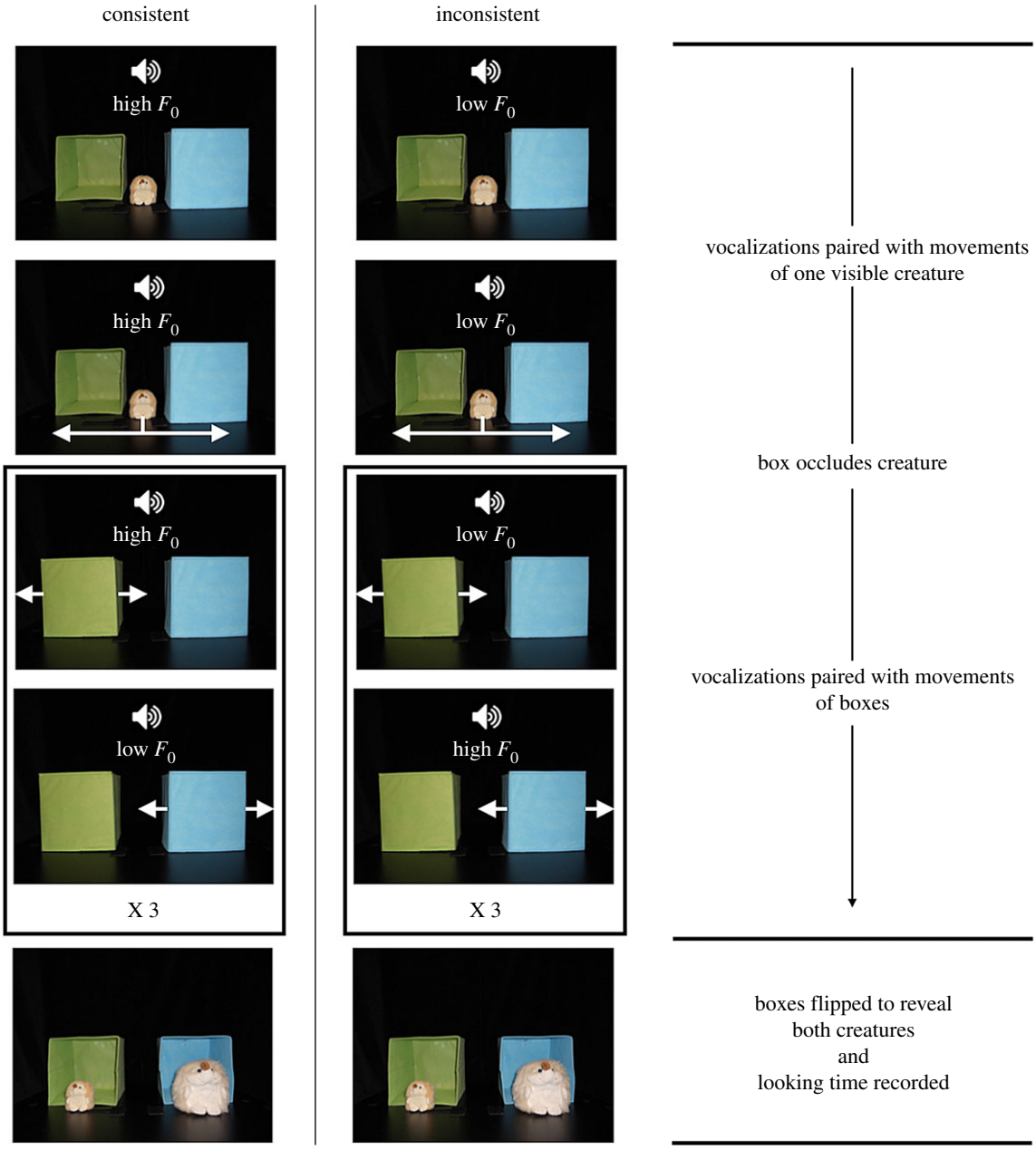

Figure 2. Schematic of the stimulus presentation trials. The two columns depict an example of a *consistent* and an *inconsistent* trial, respectively, for pitch. Within each trial, the third and fourth events (in which vocalizations were paired with box movements) were shown three times in a row. Each infant saw six trials (three consistent, three inconsistent).

and last vocalization/box movement within a single trial, both boxes were flipped back to reveal what was under each (see bottom row of figure 2). In addition to the creature seen previously, the second box contained a new, second creature that had never been seen before. This new creature was either much larger or smaller than the first creature (manipulated between-subjects; see *Visual stimuli* §2d for details).

(iii) Trial order

Infants were shown six trials in total, and a blocked trial procedure was used. For half of the infants, the first three trials were consistent sound/size pairings and the last three were inconsistent sound/size pairings. For the other half of infants, the first three trials were inconsistent and the last three were consistent.

(iv) Counterbalancing

The following were counterbalanced across participants in each experimental group (F_0 and F_n): (i) the type of trial block presented first (consistent versus inconsistent), (ii) the size of the creature presented first (large versus small), (iii) type of sound file presented first (low F_0 or F_n versus high F_0 or F_n), (iv) the colour of the box presented on the left side of the stage (blue versus green) and (v) the method of switching between consistent and inconsistent trials within a single study session, either changing the order of the sound file for the second trial block (i.e. switch low–high to high–low) and leaving the presentation of creature size the same (i.e. large–small or small–large), or changing the order of the creature size for the second trial block and leaving the presentation of the sound file the same; preliminary analyses showed that the method of switch had no impact on infants' performance. The box into which the creature moved during the show was always on the left to accommodate puppeteering.

(v) Coding

Infants' looking times for each trial were recorded during the live presentation by a trained coder using jHab [43]. Looking time recording began when the flipped boxes reached their resting

Downloaded from http://rspb.royalsocietypublishing.org/ on June 8, 2017

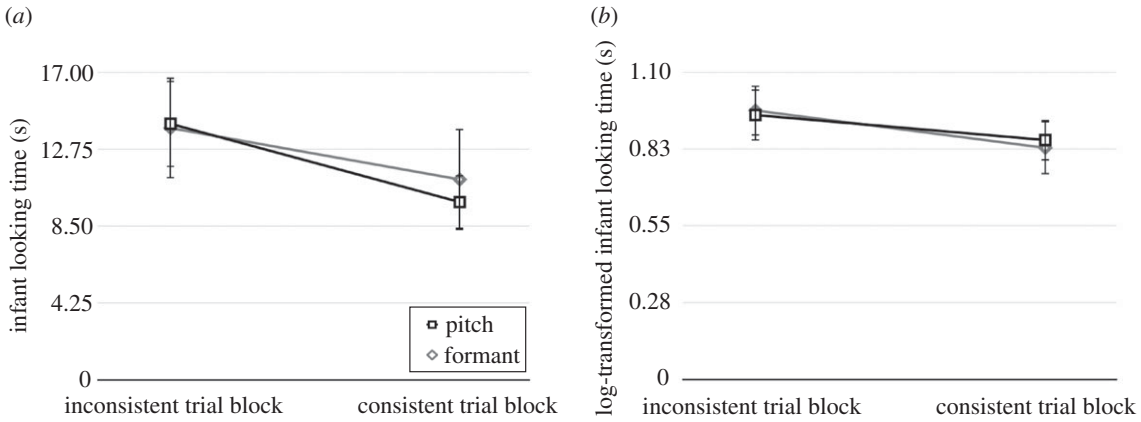

Figure 3. Mean raw looking times (*a*) and log-transformed looking times (*b*) for the inconsistent and consistent trial blocks in the pitch and formant conditions. (Error bars: ± 1 s.e.).

point on the stage, revealing both creatures, and continued until (i) the infant looked away from the display for two seconds, or (ii) the total trial length reached 45 s. A second independent coder subsequently evaluated infants' looking times from the video of the session.[4] All coders (live and video) were blind to condition during looking time coding. The live and video coders' looking times were highly correlated ($r = 0.99$); trials with coder disagreements of greater than two seconds were evaluated by a third independent coder. Looking times for five trials were excluded because (i) the trial ended too early during live coding and consequently the infant's looking time for that trial could not be accurately assessed (three trials), (ii) the live coder made an error and the relevant part of the video was lost (one trial) and (iii) the curtain was accidentally dropped midway through looking time recording (one trial).

A blind coder also determined whether infants saw critical parts of each event. An event was counted as 'seen' if infants saw (i) each creature dancing on the stage while the corresponding sound played, (ii) each box shaking with the creature underneath it while the sound played and (iii) the boxes flipping up to reveal both creatures. A second blind coder evaluated eight randomly chosen videos (25% of the final N). Coder agreement was high for whether the infants saw the creature dancing (98%), box shakes (100%), and box flip (96%). Eight looking time trials were excluded for failure to watch events; however including these trials in the analysis does not alter the pattern of results (see electronic supplementary material).

3. Results

Looking time data were log-transformed, as infant looking time data are typically not normally distributed [44–47] and the current data contained right-skew. All analyses reported below use the log-transformed looking time data; similar results are found using the raw data (see electronic supplementary material for statistical analyses) and all figures depict both the log-transformed and raw data. Infants' looking times for the inconsistent and consistent size/sound pairings are depicted in figure 3.

(a) Combined analysis

A mixed-model ANOVA was first conducted on looking time duration for both the pitch and formant conditions, with trial type as a within-subjects factor (inconsistent, consistent) and presentation order (inconsistent first, consistent first) and condition (pitch, formant) as between-subjects factors. These analyses revealed a significant main effect of trial type such that, as predicted, infants looked longer at the inconsistent sound/size pairing trials ($M = 0.96$, s.d. $= 0.36$) than the consistent sound/size pairing trials ($M = 0.84$, s.d. $= 0.33$; $F_{1,28} = 5.52$, $p = 0.026$, $\eta_p^2 = 0.17$). There was no difference found between the pitch and formant conditions $F_{1,28} = 0.003$, $p = 0.954$, $\eta_p^2 < 0.001$), and likewise no main effect of presentation order ($F_{1,28} = 1.12$, $p = 0.300$, $\eta^2 = 0.038$). However, there was also a trial type × presentation order interaction ($F_{1,28} = 9.93$, $p = 0.004$, $\eta^2 = 0.26$), meaning that the effect of trial type differed depending on which trial order infants saw (consistent first, or inconsistent first; presentation order effects are common in infant looking time studies [48]). Exploratory analyses confirmed no other significant main effects or interactions with trial type (see electronic supplementary material for analyses).

Because of the significant interaction, it was necessary to examine the effects of trial type broken up by each of the presentation orders [49] (see electronic supplementary material, figure S1). Infants who saw the inconsistent trial block first looked substantially longer at the inconsistent size/sound pairings than at the consistent size/sound pairings (inconsistent trials $M = 1.09$, s.d. $= 0.28$, consistent trials $M = 0.83$, s.d. $= 0.29$; $t_{15} = 3.75$, $p = 0.002$, $r = 0.70$). In contrast, infants who saw the consistent trial block first did not show a looking time difference across the inconsistent versus consistent pairings (inconsistent trials $M = 0.82$, s.d. $= 0.39$, consistent trials $M = 0.86$, s.d. $= 0.38$; $t_{15} = -0.61$, $p = 0.550$, $r = 0.16$). Thus, the interaction reflects that infants who saw the inconsistent events first drove the difference between inconsistent and consistent event types.

What does this mean? First and most important, the looking time difference cannot be explained by infants merely looking longer at the beginning trials and less long at later trials, reflecting a general waning of attention. If attentional waning were solely responsible for this looking time difference, the same pattern would have also been found among the infants who saw the consistent events first—they would have likewise looked longer at their first block (the consistent events) than at their second block (the inconsistent events)—yet they did not. Instead, these infants maintained their previous level of attentiveness during their second block rather than simply decreasing their looking time further. (This may have occurred because these infants had first seen expected events and thus were less attentive to stimuli changes occurring at the midway point of the trials, or because the effect of the inconsistent events was cancelling

Downloaded from http://rspb.royalsocietypublishing.org/ on June 8, 2017

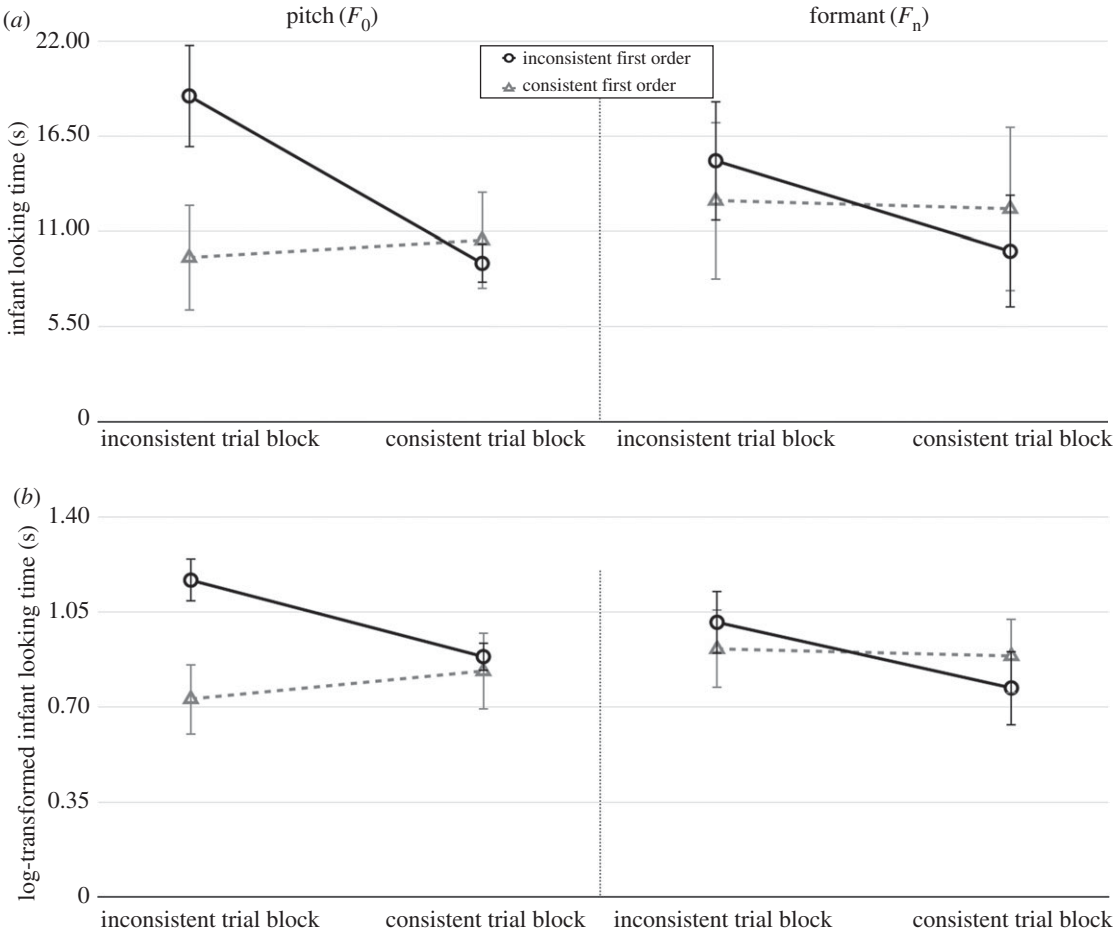

Figure 4. Mean raw looking times (*a*) and log-transformed looking times (*b*) for the inconsistent and consistent trial blocks, broken up by the order in which the trial block was presented (inconsistent first or consistent first). The pitch condition is on the left and the formant condition is on the right. (Error bars: ± 1 s.e.).

out any waning of attention occurring during the second block. Either possibility reflects an effect of experimental condition consistent with the hypothesis.) Therefore, the interaction indicates that infants' attention waned differently across the two presentation orders in a way that reflects greater attention to the inconsistent events than the consistent events.

In sum, the relative difference in the looking times found among infants can only be due to the differential response to the consistent versus inconsistent events, and, as predicted, infants looked longer at the inconsistent size/sound pairings than at the consistent size/sound pairings, even when collapsing across all participants.

(b) Analyses broken up by pitch and formant

Because pitch and formant differences correspond to different kinds of body size differences within and between organisms [7–14], it was also of theoretical interest to conduct an analysis broken up by the type of frequency manipulation infants were exposed to. This allowed us to verify that both types of manipulations produced a similar pattern to that found in the combined analysis, and to look for any differences in the magnitude of the looking time effects they produced (particularly because the pitch stimuli differences were slightly larger than those in the formant conditions).

Two separate 2×2 mixed-model ANOVA's with trial type as a within-subjects factor and order as a between-subjects factor were conducted. Again, a trial type X presentation order interaction was evident in the pitch condition, $F_{1,14} = 9.89$, $p = 0.007$, $\eta_p^2 = 0.41$, although the interaction effect was weaker for formant, $F_{1,14} = 2.18$, $p = 0.162$, $\eta_p^2 = 0.13$. As can be seen in figure 4, infants who saw the inconsistent trial block first looked longer at the inconsistent size/sound pairings than at the consistent size/sound pairings in the pitch condition (inconsistent trials $M = 1.17$, s.d. $= 0.22$, consistent trials $M = 0.88$, s.d. $= 0.15$; $t_7 = 3.29$, $p = 0.013$, $r = 0.78$) and marginally longer in the formant condition (inconsistent trials $M = 1.01$, s.d. $= 0.33$, consistent trials $M = 0.77$, s.d. $= 0.38$; $t_7 = 2.09$, $p = 0.075$, $r = 0.62$). And, as in the combined analyses above, no looking time differences were found for infants who saw the consistent trial block first (*pitch*: inconsistent trials $M = 0.73$, s.d. $= 0.36$, consistent trials $M = 0.83$, s.d. $= 0.40$; $t_7 = -1.18$, $p = 0.277$, $r = 0.41$; *formant*: inconsistent trials $M = 0.91$, s.d. $= 0.41$, consistent trials $M = 0.89$, s.d. $= 0.39$; $t_7 = 0.30$, $p = 0.775$, $r = 0.11$).

In summary, as predicted, infants looked longer at inconsistent size/sound pairings than at consistent size/sound pairings, both for changes in pitch and in formant frequencies. Both frequency changes produced a similar pattern of results, and the effect of pitch was stronger than that of formants.

4. Discussion

Three-month-old infants were exposed to size/sound pairings that were either consistent or inconsistent with the natural covariation between vocalization fundamental and formant frequencies and the physical size of an organism. Infants looked longer at events that depicted size/sound relationships

Downloaded from http://rspb.royalsocietypublishing.org/ on June 8, 2017

inconsistent with this covariation, namely when a smaller creature produced a lower-frequency vocalization than a larger creature. This occurred for both manipulations of average pitch (F_0) and formant frequencies (F_n), indicating that such inconsistent relationships were unexpected. Overall, these results suggest an early-developing sensitivity to the relationship between the size of an organism and the properties of its vocalizations. To our knowledge, this is the first study to show that human infants (or any other mammalian species) expect differences in both fundamental and formant frequencies in vocalizations to map onto differences in organisms' size.

(a) Implications

Because both pitch and formants independently predict body size across a variety of species, and in humans distinguish sex and age morphs [10,11], these results open up the interesting possibility that developmental adaptations are taking advantage of acoustic invariances, and do so in order to help infants learn about the attributes of conspecifics in the environment, including differences in their age, sex, and size [50].

This finding opens up three new directions for future research: (i) exploring the relative contributions of fundamental and formant frequencies in judgments of other differences between organisms, such as differences in dominance and formidability [4,16,22], (ii) more precisely determining if this size/sound expectation reflects expectations about different species, different age and sex morphs within a species, and/or expectations of size differences within these morphs, and (iii) assessing the ecological and taxonomic distribution of this competence by exploring if and when similar representations develop in non-human animal species, particularly in species that differ in the range of body size differences they encounter in their natural ecology.

In the current findings there was a stronger effect for pitch than for formants. Why this happened warrants some consideration, and should be pursued in future studies. The first possibility is that infants, like adults, find pitch differences easier to perceive than formant differences.[5] Furthermore, although pitch and formant manipulations were far above known perceptual thresholds for adults listening to human speech, a slightly larger objective difference in pitch was used because of the constraints of the vocalization used (see *Acoustic stimuli* §2c and figure 1). Future work should therefore explore the relative role of pitch and formant frequency information in infants' body size perception by using stimuli that can be more readily equated on pitch and formants (i.e. human voices) and by using a factorial design incorporating step wise variations in both dimensions, particularly because lower-pitched acoustic stimuli facilitate more accurate judgements of formants, and consequently body size, due to greater harmonic density in lower frequency sounds [10].

The second possibility is that the stronger effect of pitch may not only reflect arbitrary limitations of perceptual discrimination (i.e. differences in range and in JNDs), but may also reflect a more principled developmental prioritization. In particular, formants better distinguish size *within* age and sex categories [11], both in humans and a number of other large terrestrial and aquatic mammals, whereas pitch better distinguishes *between* different ages and sexes [10]. Evidence from other modalities (such as face perception [50]) suggests that early developmental processes often start out carving up the world along broader category differences, and then once established, start to build up more fine-grained within-category differences (which may either be ecologically functional or a developmental constraint). A similar developmental process may be happening here in the acoustic modality.

Finally, it is too early to tell if the discrimination and association abilities documented in the current study are produced by adaptations for tracking organism acoustics specifically, or the broader category of object acoustics. That is, motion paired with emitted sounds may satisfy the input conditions of adaptations for picking up on invariant sound/object co-relations, irrespective of whether the object is a living thing vocalizing, or a non-living object producing noise [27–34,54]. However, because the acoustics produced by the living and non-living worlds are interestingly different from one another, both in terms of their properties and also in their affordances [9,16,22,23], the psychological mechanisms for representing and reasoning about the living versus the non-living worlds will likely become increasingly distinct from one another over the course of development. Future research will therefore be needed to fully establish when and how this happens, and theoretical analyses will be needed to more fully establish the evidentiary standards of specialized design [55] for acoustic perception adaptations designed around the living versus non-living world.

Moreover, it is worth noting that existing work on visual and acoustic cross-modality correspondences often uses inanimate objects or shapes [32–35]. However, from an evolutionary perspective, it is plausible that the proper domain of some cross-modality expectations is the living world. Thus—and as this study demonstrates—it may be fruitful to also examine cross-modal expectations using more ecologically valid stimuli, including presenting cues of animate organisms when appropriate.

(b) Conclusion

Because the relationship between organisms' size and the sounds they produce is robust, phylogenetically recurrent and useful for learning, it is plausible that developmental adaptations produce a phenotype that binds together these size and sound invariances at relatively early stages of ontogeny. Using one of the youngest populations possible for our looking-time methodology, we indeed found evidence for this—three-month-old human infants expected that differences in vocalization properties would map onto size differences between organisms.

Ethics. This research was conducted in the Department of Psychology at Yale University, and as such was carried out in a manner consistent with all Yale University ethical requirements and review processes, including obtaining ethics approval from the Yale University IRB, and obtaining informed consent from parents for their child's participation.

Data accessibility. Supporting data can be found in the electronic supplementary material, and also by contacting the corresponding author.

Authors' contributions. D.P. and A.E.W. designed and conducted the studies and analysed the data. G.A.B. created and analysed the acoustic stimuli. D.P., A.E.W. and G.A.B. wrote the manuscript. K.W. provided feedback on data analyses and edited the manuscript.

Competing interests. The authors have no competing interests to declare.

Funding. This research was supported by NSF grant no. BCS-9201515 and NIH grant no. RO1 MH 081877 to K.W.

Downloaded from http://rspb.royalsocietypublishing.org/ on June 8, 2017

Acknowledgements. We would like to thank Shelley McKinnon for her assistance recruiting participants and running the study, and the infants and parents who kindly participated.

Endnotes

[1]Formants better distinguish size *within* age and sex morphs [11], both in humans and a number of other large terrestrial and aquatic mammals [12].

[2]Looking time paradigms are used with young infants because perceptual and oculomotor maturation occurs more quickly and uniformly than does gross and fine motor maturation, and thus looking times are more a reliable index of attentional patterns early in ontogeny [4,28–37]. Three-month-olds were tested because previous findings suggest that infants of this age have adequate perceptual and oculomotor maturity for the current task [28–39].

[3]All sounds were played through a set of centrally located speakers behind the back curtain of the stage.

[4]Two sessions could not be independently coded a second time, due to video loss. Inclusion or exclusion of these sessions has no effect on the results (both losses occurred in the *formant*, *consistent first* conditions; see §3).

[5]In adults, JNDs for pitch changes in stable vowel sounds are as low as 2% [51], whereas JNDs for formant changes that allow listeners to discriminate simple words are approximately 5% [52,53]. Although these perceptual thresholds are likely to vary as a function of the particular judgment task and stimulus set, it is possible that pitch is a slightly more noticeable dimension of vocalizations for developing infants as well.

References

1. Archer J. 1988 *The behavioural biology of aggression*. Cambridge, UK: Cambridge University Press.
2. Enquist M, Leimar O. 1983 Evolution of fighting behaviour: decision rules of relative strength. *J. Theor. Biol.* **102**, 387–410. (doi:10.1016/0022-5193(83)90376-4)
3. Lassek WD, Gaulin SJC. 2009 Costs and benefits of fat-free muscle mass in men: relationship to mating success, dietary requirements, and native immunity. *Evol. Hum. Behav.* **30**, 322–328. (doi:10.1016/j.evolhumbehav.2009.04.002)
4. Thomsen L, Frankenhuis WE, Ingold-Smith M, Carey S. 2011 Big and mighty: preverbal infants mentally represent social dominance. *Science* **331**, 477–480. (doi:10.1126/science.1199198)
5. Sell A, Cosmides L, Tooby J, Sznycer D, von Rueden C, Gurven M. 2009 Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proc. R. Soc. B* **276**, 575–584. (doi:10.1098/rspb.2008.1177)
6. Fant F. 1960 *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.
7. Fitch WT, Hauser M. 2003 Unpacking 'honesty': vertebrate vocal production and the evolution of acoustic signals. In *Acoustic communication* (eds A Simmons, AN Popper, RR Fay), pp. 65–137. New York, NY: Springer.
8. Taylor AM, Reby D. 2010 The contribution of source-filter theory to mammal vocal communication research. *J. Zool.* **280**, 221–236. (doi:10.1111/j.1469-7998.2009.00661.x)
9. Hodges-Simeon CR, Gurven M, Puts DA, Gaulin SJC. 2014 Vocal fundamental and formant frequencies are honest signals of threat potential in peripubertal males. *Behav. Ecol.* **25**, 984–988. (doi:10.1093/beheco/aru081)
10. Pisanski K, Fraccaro PJ, Tigue CC, O'Connor JJM, Feinberg DR. 2014 Return to Oz: voice pitch facilitates assessments of men's body size. *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 1316–1331. (doi:10.1037/a0036956)
11. Rendall D, Vokey JR, Nemeth C. 2007 Lifting the curtain on the Wizard of Oz: biased voice-based impressions of speaker size. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 1208–1219. (doi:10.1037/0096-1523.33.5.1208)
12. Pisanski K *et al.* 2014 Vocal indicators of body size in men and women: a meta-analysis. *Anim. Behav.* **95**, 89–99. (doi:10.1016/j.anbehav.2014.06.011)
13. Fitch WT. 1997 Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* **102**, 1213–1222. (doi:10.1121/1.421048)
14. Fitch WT, Giedd J. 1999 Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.* **106**, 1511–1522. (doi:10.1121/1.427148)
15. Titze IR. 1994 *Principles of voice production*. Englewood Cliffs, NJ: Pretence Hall.
16. Sell A, Bryant GA, Cosmides L, Tooby J, Sznycer D, von Rueden C, Krauss A, Gurven M. 2010 Adaptations in humans for assessing physical strength from the voice. *Proc. R. Soc. B* **277**, 3509–3518. (doi:10.1098/rspb.2010.0769)
17. Evans KK, Treisman A. 2010 Natural cross-modal mappings between visual and auditory features. *J. Vis.* **10**, 1–12. (doi:10.1167/10.1.6)
18. Gallace A, Spence C. 2006 Multisensory synesthetic interactions in the speeded classification of visual size. *Percept. Psychophys.* **68**, 1191–1203. (doi:10.3758/BF03193720)
19. Feinberg DR, Jones BC, Little AC, Burt DM, Perrett DI. 2005 Manipulations of fundamental and/or formant frequencies influence the attractiveness of human male voices. *Anim. Behav.* **69**, 561–568. (doi:10.1016/j.anbehav.2004.06.012)
20. Taylor AM, Reby D, McComb K. 2010 Why do large dogs sound more aggressive to human listeners: Acoustic bases of motivational misattributions. *Ethology* **116**, 1155–1162. (doi:10.1111/j.1439-0310.2010.01829.x)
21. Pitcher BJ, Briefer EF, McElligott AG. 2015 Intrasexual selection drives sensitivity to pitch, formants, and duration in the competitive calls of fallow bucks. *BMC Evol. Biol.* **15**, 221. (doi:10.1186/s12862-015-0429-7)
22. Puts DA, Hodges CR, Cárdenas RA, Gaulin SJC. 2007 Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evol. Hum. Behav.* **28**, 340–344. (doi:10.1016/j.evolhumbehav.2007.05.002)
23. Puts DA, Apicella CL, Cárdenas RA. 2012 Masculine voices signal men's threat potential in forager and industrial societies. *Proc. R. Soc. B* **279**, 601–609. (doi:10.1098/rspb.2011.0829)
24. Faragó T, Pongrácz P, Miklósi Á, Huber L, Virányi Z, Range F. 2010 Dogs' expectations about signalers' body size by virtue of their growls. *PLoS ONE* **5**, e15175. (doi:10.1371/journal.pone.0015175)
25. Briefer EF. 2012 Vocal expressions of emotions in mammals: mechanisms of production and evidence. *J. Zool.* **288**, 1–20. (doi:10.1111/j.1469-7998.2012.00920.x)
26. Maynard Smith J, Harper D. 2003 *Animal signals*. Oxford, UK: Oxford University Press.
27. Coward SW, Stevens CJ. 2004 Extracting meaning from sound: nomic mappings, everyday listening, and perceiving object size from frequency. *Psychol. Rec.* **54**, 349–364.
28. Bahrick LE. 1994 The development of infants' sensitivity to arbitrary intermodal relations. *Ecol. Psychol.* **6**, 111–123. (doi:10.1207/s15326969eco0602_2)
29. Spelke ES. 1979 Perceiving bimodally specified events in infancy. *Dev. Psychol.* **15**, 626–636. (doi:10.1037/0012-1649.15.6.626)
30. Spelke ES. 1981 The infants' acquisition of knowledge of bimodally specified events. *J. Exp. Child Psychol.* **31**, 279–299. (doi:10.1016/0022-0965(81)90018-7)
31. Pickens J. 1994 Perception of auditory-visual distance relations by 5-month-old infants. *Dev. Psychol.* **30**, 537–544. (doi:10.1037/0012-1649.30.4.537)
32. Dolscheid S, Hunnis S, Casasanto D, Majid A. 2014 Prelinguistic infants are sensitive to space–pitch associations found across cultures. *Psychol. Sci.* **25**, 1256–1261. (doi:10.1177/0956797614528521)
33. Walker P, Bremner JG, Mason U, Spring J, Mattock K, Slater A, Johnson SP. 2010 Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychol. Sci.* **21**, 21–25. (doi:10.1177/0956797609354734)

Downloaded from http://rspb.royalsocietypublishing.org/ on June 8, 2017

34. Peña M, Mehler J, Nespor M. 2011 The role of audiovisual processing in early conceptual development. *Psychol. Sci.* **22**, 1419–1421. (doi:10.1177/0956797611421791)
35. Fernández-Prieto I, Navarra J, Pons F. 2015 How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behav. Dev.* **38**, 77–81. (doi:10.1016/j.infbeh.2014.12.008)
36. Baillargeon R, Spelke ES, Wasserman S. 1985 Object permanence in five-month-old infants. *Cognition* **20**, 191–208. (doi:10.1016/0010-0277(85)90008-3)
37. Woodward AL. 1998 Infants selectively encode the goal object of an actor's reach. *Cognition* **69**, 1–34. (doi:10.1016/S0010-0277(98)00058-4)
38. Kellman PJ, Arterberry ME. 2006 Infant visual perception. In *Handbook of child psychology: vol. 2. Cognition, perception, and language* (eds D Kuhn, RS Siegler) (Series Ed., Damon W), pp. 109–160, 6th edn. Hoboken, NJ: Wiley.
39. Slater A, Kirby R. 1998 Innate and learned perceptual abilities in the newborn infant. *Exp. Brain Res.* **123**, 90–94. (doi:10.1007/s002210050548)
40. Boersma P, Weenink D. 2012 Praat: doing phonetics by computer [Computer program]. Version 5.1.3, retrieved from http://www.praat.org/
41. Pisanski K, Rendall D. 2011. The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *J. Acoust. Soc. Am.* **129**, 2201–2212. (doi:10.1121/1.3552866)
42. Hauser MD. 1993 The evolution of nonhuman primate vocalizations: effects of phylogeny, body weight, and social context. *Am. Nat.* **142**, 528–542. (doi:10.1086/285553)
43. Casstevens RM. 2007 jHab: Java Habituation Software (Version 1.0.2) [Computer Software]. Chevy Chase, MD.
44. Csibra G, Hernik M, Mascaro O, Tatone D, Lengyel M. 2016 Statistical treatment of looking-time data. *Dev. Psychol.* **52**, 521–536. (doi:10.1037/dev0000083)
45. Hernik M, Csibra G. 2015 Infants learn enduring functions of novel tools from action demonstrations. *J. Exp. Child Psychol.* **130**, 176–192. (doi:10.1016/j.jecp.2014.10.004)
46. Kibbe MM, Leslie AM. 2016 The ring that does not bind: topological class in infants' working memory for objects. *Cogn. Dev.* **38**, 1–9. (doi:10.1016/j.cogdev.2015.12.001)
47. Kestenbaum R, Termine N, Spelke ES. 1987 Perception of objects and object boundaries by 3-month-old infants. *Br. J. Dev. Psychol.* **5**, 367–383. (doi:10.1111/j.2044-835X.1987.tb01073.x)
48. Luo Y, Baillargeon R, Brueckner L, Munakata Y. 2003 Reasoning about a hidden object after a delay: evidence for robust representations in 5-month-old infants. *Cognition* **88**, B23-B32. (doi:10.1016/S0010-0277(03)00045-3)
49. Rosenthal R, Rosnow RL. 2007 *Essentials of behavioral research*, 3rd edn. New York, NY: McGraw-Hill.
50. Ramsey JL, Langlois JH, Marti NC. 2005 Infant categorization of faces: ladies first. *Dev. Rev.* **25**, 212–246. (doi:10.1002/dev.20412)
51. Moore BC. 2008 Basic auditory processes involved in the analysis of speech sounds. *Phil. Trans. R. Soc. B* **363**, 947–963. (doi:10.1098/rstb.2007.2152)
52. Sinnott JM, Owren MJ, Petersen MR. 1987 Auditory frequency discrimination in primates: species differences (*Cercopithecus, Macaca, Homo*). *J. Comp. Psychol.* **101**, 126–131. (doi:10.1037/0735-7036.101.2.126)
53. Smith DR, Patterson RD, Turner R, Kawahara H, Irino T. 2005 The processing and perception of size information in speech sounds. *J. Acoust. Soc. Am.* **117**, 305–318. (doi:10.1121/1.1828637)
54. Spence C. 2011 Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* **73**, 971–995. (doi:10.3758/s13414-010-0073-7)
55. Williams G. 1966 *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.