ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Belyk M, Johnson JF, Kotz SA. 2018 Poor neuro-motor tuning of the human larynx: a comparison of sung and whistled pitch imitation. *R. Soc. open sci.* **5**: 171544.
http://dx.doi.org/10.1098/rsos.171544

Received: 4 October 2017
Accepted: 13 March 2018

Subject Category:
Psychology and cognitive neuroscience

Subject Areas:
evolution/neuroscience/behaviour

Keywords:
imitation, voice, larynx, articulation, motor control, evolution

Author for correspondence:
Michel Belyk
e-mail: belykm@gmail.com

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4050698.

THE ROYAL SOCIETY PUBLISHING

Poor neuro-motor tuning of the human larynx: a comparison of sung and whistled pitch imitation

Michel Belyk[1,2], Joseph F. Johnson[2] and Sonja A. Kotz[2,3]

[1]Bloorview Research Institute, 150 Kilgour Road, Toronto, Canada M4G 1R8
[2]Faculty of Psychology and Neuroscience, University of Maastricht, Maastricht, The Netherlands
[3]Department of Neuropsychology, Max Planck Institute for Human and Cognitive Sciences, Leipzig, Germany

MB, 0000-0002-3270-8666

Vocal imitation is a hallmark of human communication that underlies the capacity to learn to speak and sing. Even so, poor vocal imitation abilities are surprisingly common in the general population and even expert vocalists cannot match the precision of a musical instrument. Although humans have evolved a greater degree of control over the laryngeal muscles that govern voice production, this ability may be underdeveloped compared with control over the articulatory muscles, such as the tongue and lips, volitional control of which emerged earlier in primate evolution. Human participants imitated simple melodies by either singing (i.e. producing pitch with the larynx) or whistling (i.e. producing pitch with the lips and tongue). Sung notes were systematically biased towards each individual's habitual pitch, which we hypothesize may act to conserve muscular effort. Furthermore, while participants who sung more precisely also whistled more precisely, sung imitations were less precise than whistled imitations. The laryngeal muscles that control voice production are under less precise control than the oral muscles that are involved in whistling. This imprecision may be due to the relatively recent evolution of volitional laryngeal-motor control in humans, which may be tuned just well enough for the coarse modulation of vocal-pitch in speech.

1. Introduction

Vocal imitation is a hallmark of human communication that underlies the capacity to learn to speak and sing. It is the

ability to reproduce previously experienced auditory events by computing an inverse model that maps target sounds onto motor commands that reproduce them [1]. Whereas the ability to flexibly produce calls from an existing repertoire (vocal usage learning) is relatively common, the ability to add new vocalizations to an existing repertoire (vocal production learning) is rare in mammals [2,3]. Among the principal exceptions are humans, cetaceans [4–6], pinnipeds [7–11], bats [12,13] and possibly elephants [14,15]. How the hominid vocal phenotype evolved from vocal usage to vocal production learning remains a matter of speculation. Several plausible hypotheses have been advanced to suggest, for example, that emotional expression may have provided a scaffold for the evolution of speech and song [16], or that the exaggeration of social dominance cues provided selective pressure for more versatile voices [17], which may have been exploited by early hominins for the purpose of communication [18].

Vocalizations are composed of a periodic signal produced by the larynx that is filtered depending on the configuration of the vocal tract, such as the articulation of the lips and tongue [19–21]. This system of sound source and filter is a useful bioacoustic description and provides a framework for understanding the muscles of communication and the evolution of volitional control over them. For example, research has increasingly come to suggest that non-human great apes may have more flexible call repertoires than previously supposed. This includes a limited degree of flexibility at the laryngeal sound source [22–24], but much more extensive control over the shape of the vocal tract via movement of the lips and tongue. Indeed, great apes have been observed to learn to produce a variety of non-species typical oral sounds such as raspberries and whistles that involve the lips and tongue instead of the larynx [25–27]. Similarities between human speech and great ape lip-smacking behaviours [28–30], as well as in the range and flexibility of tongue movements in these species [31,32], suggest that the orofacial articulatory muscles were speech-ready in ancestral primates. By contrast, humans have a clear advantage over other primates in controlling the laryngeal sound source as we modulate vocal-pitch not only to sing, but also to encode voiced compared to voiceless phonemes [33,34], the tones of tonal languages [35], stress on particular syllables [36], emphasis on certain words [37], the intonation of sentences to contrast declarative and interrogative modes [38], and to express a broad range of genuine or feigned emotions [39–43].

Despite the vocal virtuosity of humans relative to other apes, there is a population of individuals—colloquially referred to as 'tone deaf'—who are notable in their poor abilities as singers. However, tone deafness is a misnomer, as these individuals do not necessarily have a deficit in hearing musical sounds, but rather in singing them [44–46]. These individuals are more accurately described as poor-pitch singers, as they appear to have a selective deficit in translating perceived pitches into the sequence of laryngeal-motor commands that reproduce them [47]. In some cases this results in inaccurate singing—that is consistently flat or consistently sharp—but more often it manifests as imprecise singing—that is highly variable [48]. Rather than a discrete population, poor-pitch singers appear to be the low proficiency tail of a continuous range of singing abilities [45,49,50].

Even professional opera singers, who presumably occupy the high proficiency tail of the singing proficiency continuum, are less precise with vocal-pitch when singing than violinists are with the pitch of their instruments [51,52] and may be unreliable judges of whether they themselves have just produced an error [53]. A lifetime of experience with the imprecision of the voice may explain why listeners are more generous in judging whether a vocalist is in tune than when judging an instrumentalist [54]. Across levels of training, singers match pitches more accurately with an instrument than with their voices, despite unfamiliarity with the instrument [55,56]. This pattern holds even with digital instruments that produce a vocal timbre [57,58], suggesting that poor pitch matching is rooted in vocal motor-control rather than deficient perception of vocal-pitch.

This lack of vocal proficiency is striking in humans, who are the most vocally proficient species of ape. Though there are neuro-comparative differences between humans and other primates in several brain areas related to the control of the vocal tract [59–62], one of the more striking comparative differences is specific to the laryngeal muscles that control the voice. Humans possess a direct pathway projecting from the larynx-motor cortex to the nucleus ambiguus, which is the brainstem-motor nucleus that controls the laryngeal muscles [63,64]. This direct pathway is less abundant in other great apes [65] and absent in monkeys [66,67]. However, even in humans this pathway remains sparse compared to the analogous pathways that descend to the brainstem-motor nuclei that control the muscles of the lip and tongue [63–65].

These observations lead us to hypothesize that the human vocal-motor system is not tuned as precisely as other orofacial neuro-motor systems. To test whether humans are imprecise singers, we had participants listen to and then imitate simple melodies by either singing or whistling. These tasks were highly matched in auditory and cognitive demands, differing only in whether pitches were imitated by

vocalizing the neutral vowel schwa or by producing a bilabial whistle. We hypothesized that imitation errors would tend to be larger for singing than for whistling.

2. Methods

2.1. Stimulus generation

Two sets of 45 melodies were composed by random computerized composition. The first note of each melody was selected at random from a chromatic scale. Subsequent notes were determined by sampling from a flat distribution of interval sizes, ranging ±4 semitone. This process was performed iteratively until all notes fell within the range of a single octave. All melodies consisted of a sequence of five isochronous notes each lasting 750 ms separated by 50 ms of silence. Both sets of melodies were synthesized in a vocal timbre using a neutral vowel (Vocaloid Miriam, Zero-G Limited, Okehampton, UK) as well as in a timbre that approximates a human bilabial whistle (see electronic supplementary material, S1 and S2). All stimuli had a sampling rate of 44 100 Hz with 16-bit digitization of amplitudes. Two versions of the vocal-timbre stimuli were synthesized to accommodate the disparate vocal ranges of males and females. These spanned the range A2–A3 (110–220 Hz) and A3–A4 (220–440 Hz) for males and females, respectively. The whistled-timbre stimuli were synthesized in the range A5–A6 (880–1760 Hz). Whistled-timbre stimuli were synthesized as a sine wave convolved with an empirical estimate of the onset amplitude envelope observed in pilot experiments. The pitch ranges and the onset amplitude envelope associated with the timbre of a bilabial whistle were estimated from 30 recordings taken from five males and five females. These recordings were collected during pilot experiments in which participants performed the individual assessment of producible range of pitches described below. All stimuli were synthesized at equal sound pressure levels.

2.2. Procedures

2.2.1. Participants

Thirty-four participants were recruited through two separate listings in the undergraduate testing pool of the Faculty of Psychology and Neuroscience at Maastricht University. The two listings were worded to attract either strong or poor singers in order to draw from both ends of the spectrum of singing ability, but made no reference to whistling to avoid sampling bias. Six participants were unable to produce any pitched sound by whistling. Only data from the remaining 28 participants were analysed. These participants had a median age of 21 years (range 18–29), 20 were female, nine self-identified as a good singer, 15 self-identified as a good whistler, 26 had some degree of formal musical training (2–15 years) but only two of these had any vocal training. All participants reported normal hearing and no vocal pathology. All participants provided informed consent and were compensated with either course credit or a 10€ voucher.

2.2.2. Individual assessment of producible frequency ranges

Recordings were performed in a sound-attenuated chamber using a desk-mounted Sennheiser microphone and Adobe Audition software (v. 1.5). Participants were instructed to sing (i) a stable and comfortable note, (ii) a descending sweep as a smoothly varying pitch contour from a comfortable note to their lowest producible note, and (iii) an ascending sweep from a comfortable note to their highest producible note. Each production task was repeated three times. The mean frequency of the comfortable note was measured using Praat (v. 6.0.17; www.fon.hum.uva.nl/praat/) and taken as each participant's habitual pitch. The highest and lowest frequencies produced during vocal sweeps were used to estimate each participant's producible range. The same procedure was repeated for whistling.

2.2.3. Imitation task

In the same recording environment, participants performed two audio-motor imitation tasks: once imitating one set of melodies presented in a vocal timbre and sex-appropriate vocal range by singing, and once imitating a second set of melodies presented in a whistled timbre by whistling. Participants were instructed to sing using only a neutral vowel that was also the carrier vowel of the stimulus. Each task consisted of listening to and then repeating 45 melodies consisting of five notes each. Each melody was presented one at a time and separated by 7 s silent gaps during which participants'

imitations were recorded. Stimulus onset times were jittered by 250, 500 or 750 ms. Participants were given the opportunity to rest for a duration of their choosing after every 15th trial. Both the order of imitation conditions and the sets of target melodies were counter-balanced across participants. Melodies were presented in random order within conditions. Stimulus presentation and sound recordings were managed through Python (v. 2.7; python.org). Target stimuli were played over free field speakers at a comfortable volume.

2.2.4. Melodic discrimination

We assessed participants' ability to perceive pitches within a melodic context and retain them in working memory using a computerized version of the Montreal Battery of Amusia Evaluation (MBEA) [68] programmed in Python. Stimuli were presented over free field speakers while participants were seated alone in a sound attenuated booth. Only the subscales of the MBEA that assess pitch perception (1a–1c) were completed. Participants listened to three sets of 30 pairs of melodies that were either identical or had one note transposed, and indicated whether the melodies were the same or different. Each set contained transpositions that were increasingly difficult to detect and each set was preceded by two practice trials.

2.3. Acoustic analysis

An in-house Praat script was used to semi-automate the extraction of fundamental frequency (F_0) from the centre 250 ms of each imitated note. This script is available in the online data supplement to this article (http://dx.doi.org/10.5061/dryad.504t7 [69]). Melodies that were produced with too few or too many notes were excluded from further analysis because the positions of omitted or duplicated notes were not possible to determine (totalling 3.6% of trials). Responses to stimuli that were outside of each participant's producible range were excluded as they may reflect limitations of the producible range rather than imitation ability.

F_0 values were converted from hertz to cents relative to lowest scale degree of the stimulus set, where 100 cents is equal to one semitone and 1200 cents is equal to one octave of the equal temperament scale (equation (2.1)). Note error was calculated as the differences between the pitches of the target melody and participants' imitations. Intervals are the difference between adjacent notes in a melody. Interval error was calculated as the differences between intervals in the target melody and intervals in participants' imitations. All produced notes associated with errors greater than 1000 cents were verified for measurement errors, including octave errors.

$$\log_2\left(\frac{A}{B}\right) \times 1200. \tag{2.1}$$

We applied the approach of Pfordresher *et al*. [48] in separately calculating the accuracy and precision of imitated melodies. Inaccuracy reflects a consistent bias to produce responses that err in the same direction, for example, by consistently singing flat. Imprecision reflects the variability across repeated attempts to produce the same pitch, for example, by intermittently singing responses that are flat and sharp by varying degrees. Inaccuracy scores were calculated for each participant as the mean signed difference between target notes or intervals and imitated notes or intervals. Imprecision scores were calculated for each participant by finding the standard deviation of differences between the target and imitated notes or intervals within each pitch class, and taking the average across pitch classes.

3. Results

MBEA scores varied widely along a continuous range from 71% to 100% correct responses (mean 86.4%, s.d. 9.6%). The scores of three participants were below the conventional cut-off suggested to identify individuals with amusia [68]. The continuous range of scores observed in this sample and reported by Peretz *et al*. lead us to retain the data from all participants but include MBEA scores as a continuous predictor in subsequent analyses.

Figure 1 shows violin plots of imitation errors across 11 433 notes produced by participants in this experiment and demonstrates a clear violation of heterogeneity of variance because singing appears to be more variable than whistling. We chose not to model these data using nonlinear regression techniques that are robust to heteroscedasticity because the systematic difference in variability between conditions is of theoretical interest. As an alternative, we computed (in)accuracy and (im)precision scores for each participant [48]. These scores passed all tests of assumptions for linear mixed models (LMMs).

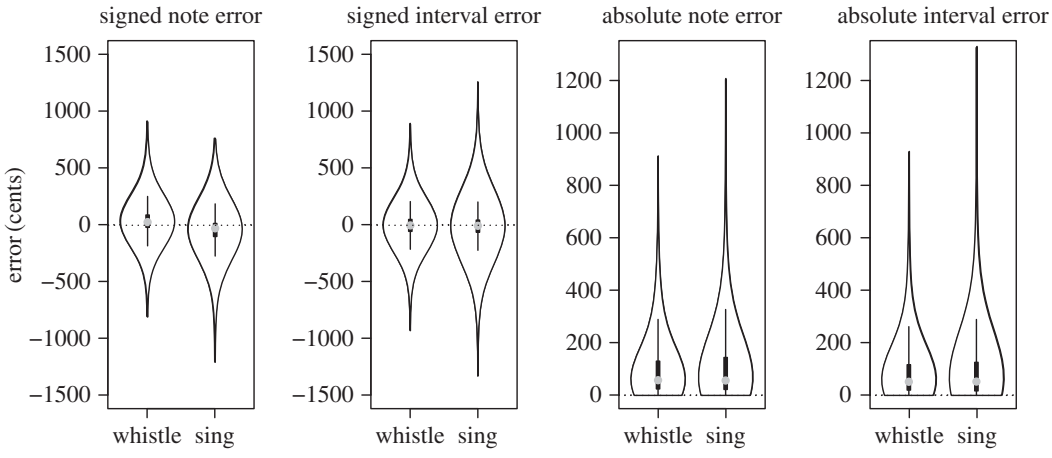

Figure 1. Violin plots. The four panels show box plots surrounded by density distributions of pitch errors for all notes produced by all participants. The two leftmost panels show signed errors and demonstrate that there is a small tendency for imitated notes to be sung flat or whistled sharp. The rightmost panels demonstrate that imitation errors while singing were much more variable than imitation errors while whistling.

We constructed four sets of nested LMMs to predict note inaccuracy, interval inaccuracy, note imprecision and interval imprecision from the two modalities of imitation and from MBEA scores, with participant modelled as a random intercept [70,71]. The effects of pitch-production modality and perceptual ability were tested by comparing nested LMMs; degrees of freedom and *p*-values were calculated using the Kenward–Roger approximation [72]. We report standardized estimates calculated by refitting each statistical model with input variables centred and scaled by 2 s.d. [73,74]. These effect sizes should be interpreted as the expected differences in outcome for levels of the predictor variables that are 1 s.d. below the mean compared to 1 s.d. above the mean [75]. For categorical predictors (such as whistling versus singing), this is equivalent to the estimated difference between conditions. For continuous predictors (such as MBEA score), this is equivalent to the difference in outcomes for participants with scores 1 s.d. below the mean (a score of 76%) to 1 s.d. above the mean (a score of 96%). Confidence intervals (CIs) for these estimates were determined by bootstrapping with 1000 iterations.

3.1. Imprecision

Note imprecisions ($F_{1,25.3} = 12.02$, $p < 0.05$, standardized estimate = −25.5, 95% CI = −38.4 to −11.3) and interval imprecisions ($F_{1,25.2} = 6.14$, $p < 0.05$, standardized estimate = −17.0, 95% CI = −30.4 to −3.5) were both significantly higher for singing than whistling (figure 2). Both note imprecision ($F_{1,26.2} = 36.92$, $p < 0.05$, standardized estimate = −83.6, 95% CI = −112.3 to −54.9) and interval imprecision ($F_{1,25.8} = 6.14$, $p < 0.05$, standardized estimate = −92.4, 95% CI = −127.6 to −61.1) were significantly predicted by perceptual ability.

Figure 2 also highlights a strong relationship between singing and whistling precision scores. However, in the light of the common influence of perceptual ability on both of these scores we conducted partial correlations between singing and whistling scores, controlling for perceptual ability as estimated by the pitch subscales of the MBEA. There were significant partial correlations between singing and whistling scores for both note imprecision ($r^2 = 0.29$, $p < 0.05$) and interval imprecision ($r^2 = 0.50$, $p < 0.05$), demonstrating that the relationship between singing and whistling imprecision scores is not solely due to the common influence of perceptual ability.

3.2. Inaccuracy

Figure 3 highlights an overall tendency for sung notes to be flat and whistled notes to be sharp. Note inaccuracy ($F_{1,25.7} = 27.03$ $p < 0.05$, standardized estimate = 100.8, 95% CI = 6.5 to 138.8) and interval inaccuracy ($F_{1,25.5} = 8.86$, $p < 0.05$, standardized estimate = 9.2, 95% CI = 2.5 to 15.5) scores were significantly lower for singing compared to whistling. Neither note inaccuracy ($F_{1,26.8} = 2.47$, $p = 0.13$, standardized estimate = 41.9, 95% CI = −9.5 to 71.5) or interval inaccuracy ($F_{1,26.6} = 0.64.0$, $p = 0.43$, standardized estimate = 3.6, 95% CI = −4.9 to 11.9) were significantly predicted by perceptual

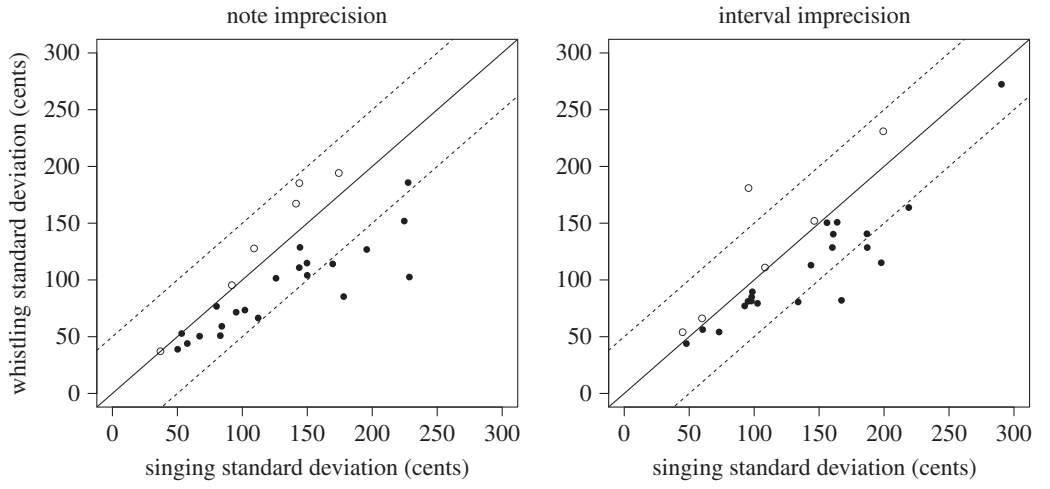

Figure 2. Imprecision. Singing imprecision is plotted on the *x*-axis and whistling imprecision is plotted on the *y*-axis for each participant. Larger scores indicate less precise imitation. The solid line indicates a hypothetical one-to-one correspondence between singing and whistling imprecision scores. Of 28 participants, 22 were below this line (filled circles) indicating that they were less precise when singing than whistling. Dashed lines indicate ±50 cents from the equal performance line, which is a conventional threshold for poor imitation.

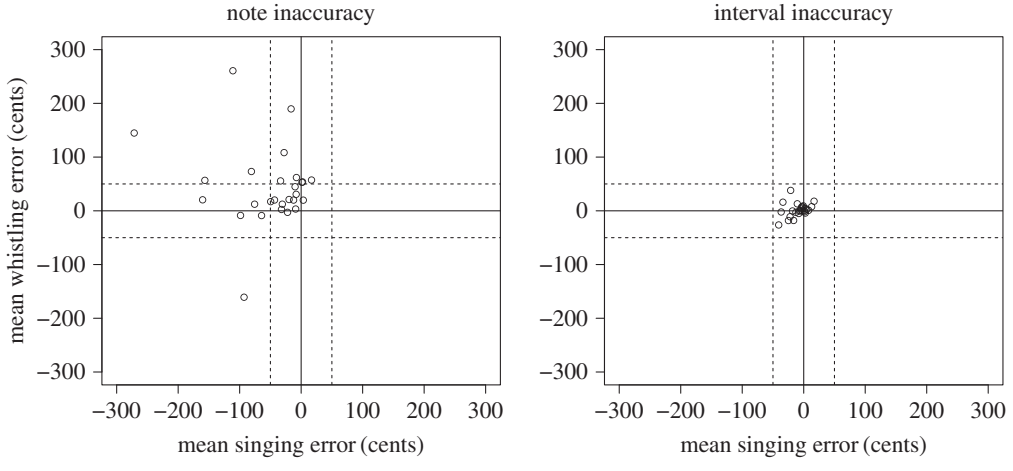

Figure 3. Inaccuracy. Singing inaccuracy is plotted on the *x*-axis and whistling inaccuracy is plotted on the *y*-axis for each participant. More extreme scores indicate less accurate imitation. Negative values indicate an average imitation that is flat and positive values indicate an average imitation that is sharp. Solid horizontal and vertical lines indicate inaccuracy scores of 0, or perfect performance. Dashed lines indicate ±50 cents as a conventional threshold for poor performance. Participants with scores beyond the dashed lines produced imitations that were closer to an out of tune note than to the target note. The notable disparity between participants note inaccuracy and interval inaccuracy scores may be explained by transposition. Participants appear to have consistently sung entire melodies up to 300 cents lower, or whistled melodies up to 200 cents higher, than the target melodies while retaining the correct relationship between notes within melodies.

ability. There was no significant partial correlation between singing and whistling scores for either note inaccuracy ($r^2 = 0.08$, $p = 0.15$) or interval inaccuracy ($r^2 = 0.12$, $p = 0.07$).

In order to explore the possible causes of the flatness of singing and sharpness of whistling, we conducted a *post hoc* test of imitation inaccuracy as a function of the target pitch. Figure 4 plots mean imitation errors of whistling and singing for each target note. As before we observed consistently sharper scores for whistling than singing ($F_{1,643.4} = 6.53$, $p < 0.05$, standardized effect = 85.8, 95% CI = 72.8 to 98.5). We also observed a strong tendency for imitations to become more flat as the pitch height of target notes increased ($F_{1,642.8} = 104.0$, $p < 0.05$, standardized effect = −69.1, 95% CI = −83.1 to −55.2), and an interaction indicating that this effect was stronger for singing than for whistling ($F_{1,643.4} = 14.4$, $p < 0.05$, standardized effect = 51.5, 95% CI = 23.7 to 78.9). Figure 4 makes plain that high notes were sung flat,

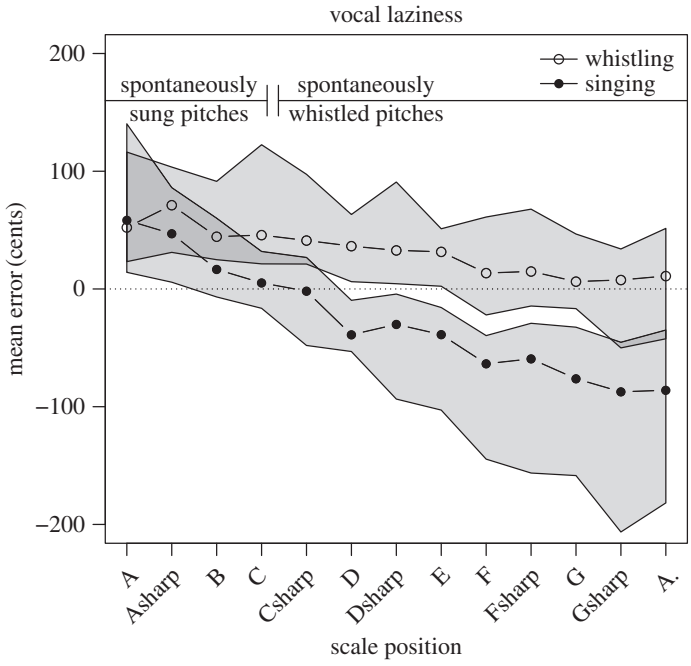

Figure 4. Vocal laziness. Median imitation inaccuracy is plotted for each stimulus note following the chromatic scale. Singing for males, singing for females, and whistling were performed in task appropriate octaves (starting from A2, A3 and A5, respectively), and are displayed on a common scale on this plot to facilitate comparison. The grey area encloses the interquartile range. The dotted line indicates a maximally accurate imitation score. Positive scores indicate imitations that were sharp and negative scores indicate imitations that were flat. Solid lines in the top portion of the figure indicate the ranges of participants' spontaneous singing and whistling pitches. Participant's spontaneous pitches tended to occupy the lower end of the stimulus range for singing and the upper end of the stimulus range for whistling. For both modes of production, accuracy was best for notes within the range of spontaneous pitches. Hence, while participants sung flat on average, they did not consistently transpose melodies downwards. Rather, participants shifted sung notes towards a point within the range of habitually produced pitches. This point may reflect a configuration of least vocal effort, distance from which may make pitch production more effortful. Vocal laziness, or a conservation of vocal effort, may lead singers to compress sung notes towards the pitch associated with a default or preferred configuration of the larynx.

whereas low notes were sung sharp. This suggests that sung imitations may be compressed towards a particular point within a singer's vocal range.

3.3. Musical training

Each LMM was refitted with additional linear predictors for musical experience measured in years of formal training and the sex of the participant, excluding two participants who had prior training as vocalists. Neither musical experience nor sex significantly predicted any outcome measure. Parameter estimates for task and MBEA predictors were similar to those reported above, although with broader CIs that presumably reflect lost residual degrees of freedom from modelling a smaller sample with more predictors (see electronic supplementary material, file 3).

4. Discussion

The current experiment aimed to test the relative proficiency of human singing and whistling in the light of previous indications that although humans are the most proficient vocal learners among primates they may none-the-less have relatively coarse control over the laryngeal muscles that regulate vocal-pitch. We observed that participants tended to sing flat but whistle sharp. Singing and whistling imprecision were highly correlated, suggesting that some common mechanisms may contribute to errorfulness in both domains. Singing was also systematically less precise than whistling, suggesting a differential error proneness for the laryngeal and oral muscles, respectively. Since singing and whistling probe muscles that control the laryngeal sound source and the vocal-tract filter, respectively, they provide an avenue to study the motor control of both sound source and filter using a common metric.

Previous research has observed that trained singers are less precise at matching pitches with their voices than instrumentalists are at matching pitches with their instruments [51,55,56]. Although the violin produces a continuous range of pitches and is, therefore, capable of erring to the same degree as the voice, design features of the violin may give instrumentalists an advantage. For instance, the strings of the violin are tuned to ensure that a given placement of the bow will produce a reliable pitch. By contrast, pitch production by the larynx depends on complex and nonlinear interactions between multiple laryngeal muscles [76–82], the configuration of the articulatory muscles [19,83] and the action of the lungs [84–86], within a biological frame whose tuning may change as it matures and senesces.

We observed that pitch production by singing is also imprecise compared to pitch production by whistling. Although less is known about the bioacoustics of whistling relative to singing, both rely on the careful configuration of complex and interacting muscle groups within the vocal tract. Moreover, even people who seldom sing have more extensive experience controlling vocal-pitch than a whistled pitch through daily experience with speech. Speech involves a constant regulation of voice onsets and vocal-pitch height. Speakers use these cues to encode voiced compared to voiceless phonemes [33,34], the tones of a tonal language [35], stress on particular syllables [36], emphasis on certain words [37], the intonation of sentences to contrast declarative and interrogative modes [38] and to express a broad range of genuine or feigned emotions [39–43]. Sung pitches remain imprecise despite a lifetime of daily voice experience, suggesting that there are fundamental limitations on the precision of human vocal-pitch control at the level of the laryngeal muscles or the neuro-motor system that controls them [87–89].

4.1. Music and language share a voice in song and speech

Music and language are two frameworks that humans use to interact and communicate with one another. They are not exclusive to any one mode of production; for example, music can be expressed by blowing in some instruments or by banging on others and language can be expressed by writing or by making signs. Music and language share the use of the voice when they are expressed as singing and speaking, respectively.

The use of vocal-pitch as both the carrier for melody in song and for providing prosodic and phonetic cues in speech reflects part of a broader framework linking the evolution of musical and linguistic abilities in humans [90–92]. Many features are shared between language and music, such as processing sequences of sound over time [93–97], interpreting their meaning within the broader context of a musical or linguistic phrase [98–102], syntactic ordering of events [103–105], pacing of rhythmic movements [106–108] and vocal production learning [109–112]. The evolution of any of these abilities, including vocal production learning, may have been driven by selective pressures that predate singing or speaking, though they support both of these behaviours [3,96,113–115].

This shared history of selective pressures not specific to music is consistent with an existing view that the development of human musical scales may have been constrained to accommodate the imprecision of the voice [51]. The music of most cultures is built on scales containing a small number of degrees [116], leading most note categories to be separated by a full tone (200 cents). Scales of this construction may have allowed even novice singers to sing notes that were closer to being in tune than out of tune, most of the time.

4.2. The relatively recent evolution of the vocal-motor system

The greater precision of human orofacial-pitch control in whistling, relative to laryngeal-pitch control in singing, is consistent with a relatively recent evolution of the neuro-motor system that controls the laryngeal muscles [87]. Although many species can volitionally produce their species typical calls, few species have the capacity to add new calls to their repertoire through imitation. This ability is found in three lineages of songbird [3,117] and several lineages of mammal, including cetaceans [4–6], pinnipeds [7–11], bats [12,13] and possibly elephants [14,15]. Humans are notable as the only primate with a strong capacity for vocal imitation.

Non-human apes have been observed to produce a variety of novel sounds, but these are most often in the form of oral sounds, such as a ‘raspberry’ or a whistle that use the lips or tongue as a sound source [25–27,118], although these species may also have a limited degree of flexibility at the laryngeal sound source [22–24]. The most well-documented case is that of Koko the encultured Gorilla. Koko learned an extensive repertoire of novel sounds that she used primarily during play [23]. These sounds demonstrated a considerable degree of control over the muscles of articulation and respiration, but little

of Koko's vocabulary involved voice production from the laryngeal sound source, suggesting a more limited degree of control over the laryngeal sound source than the rest of the vocal tract.

Sound imitation is a complex behaviour that engages a broad network of brain areas. Neuroimaging research in which human participants imitated articulatory patterns [119–122] or pitch patterns [123,124] have observed activation in brain areas related to motor-planning and execution, including the inferior frontal gyrus, anterior cingulate cortex, supplementary motor area, basal ganglia, cerebellum, primary somatosensory cortex and primary-motor cortex. Much of this network is conserved across primates [63,65,125–129] and contains somatotopic maps with separate representations of the various muscles of the body [128,130–135], inviting a neuro-comparative analysis at the level of muscle groups.

While the larynx-motor area of monkeys is found in premotor cortex and has limited involvement in vocal behaviour [136–139], the human larynx-motor area is found in primary-motor cortex and has a clear involvement in regulating vocal behaviour [124,133,140–147]. Non-human great apes have an intermediate phenotype [128,148,149]. These brain areas have distinct cytoarchitectural profiles; the primary-motor cortex has a greater abundance of descending motor fibres than the premotor cortex [150,151]. Likewise, monkeys lack a direct connection between the larynx-motor cortex and the nucleus ambiguus, which is the brainstem-motor nucleus that controls the laryngeal muscles [66,67]. Apes have a sparse, but extant, direct pathway between these areas, that is slightly more abundant in humans [63,64]. Vocal behaviour driven by primary-motor cortex began to evolve before the divergence of humans from other apes, but was elaborated over human evolution. Several theorists have speculated that the emergence of this pathway may have been a prerequisite to the evolution of speech [105,152–155].

By contrast, the motor areas controlling the lips and tongue are found in similar cytoarchitectural zones across primates [65,128]. Likewise, motor fibres descending to the facial nucleus, which is the brainstem nucleus that controls the lips and tongue, are more abundant than the equivalent pathway for the larynx [63–65,149]. This abundance of orofacial-motor fibres is common to monkeys, non-human apes and humans, suggesting an evolutionary history that predates the divergence of these clades.

This comparative analysis of both vocal-learning ability and its underlying neurophysiology suggests that volitional control over the orofacial muscles evolved earlier in the primate lineage than volitional control over the laryngeal muscles; whereas orofacial-motor control is evident in all primates, volitional control over the laryngeal muscles is lacking in monkeys, incipient in non-human apes and most evident humans.

Although human vocal-motor abilities are elaborated beyond the poorer vocal-motor abilities of other primates, the relatively imprecise control of pitch by the larynx may have been sufficient to satisfy the selective pressures for which it evolved. Singing may impose demands on vocal-pitch control beyond the scope for which this ability evolved.

4.3. Imprecise pitch imitation as the accumulation of neuro-motor noise

We observed a strong relationship between pitch perception abilities and imprecision in audio-motor imitation for both singing and whistling. Audio-motor imitation requires singers to listen to a target melody, compute an inverse model that maps the target melody onto a sequence of movements that would reproduce it, and finally, execute those movements. We propose that computational noise at each stage of this process may be propagated to subsequent stages, such that imitation errors are the accumulation of errors at each stage of processing (figure 5).

Perceiving target melodies probably engages similar processes within the auditory system for both singing and whistling, and computational noise in perceiving pitch targets and retaining them in memory may explain the high degree of correlation between singing and whistling imprecision. Executing movements for sound production probably engages muscle-specific domains within the motor system from primary-motor cortex through descending corticobulbar pathways. The somatotopic organization of motor cortex by muscle effector [128,133,144,156] may cause neuro-motor noise for laryngeal movements during singing to be independent from the neuro-motor noise for tongue movements during whistling.

From analogy with songbirds, which are the most extensively studied animal model of vocal learning and imitation, inverse models that map target pitches onto motor commands appear to be computed by a thalamo-cortico-striatal loop [157,158]. Non-invasive brain imaging studies in humans have begun to support this analogy [121,122,124]. The parts of the thalamus, striatum and cortex that are relevant to movement are somatotopically organized into populations that control different groups of muscles [130,132,133,159]. Hence, it seems possible that separate but parallel neural networks compute inverse models for the larynx and the tongue, although further research is needed to assess the separation

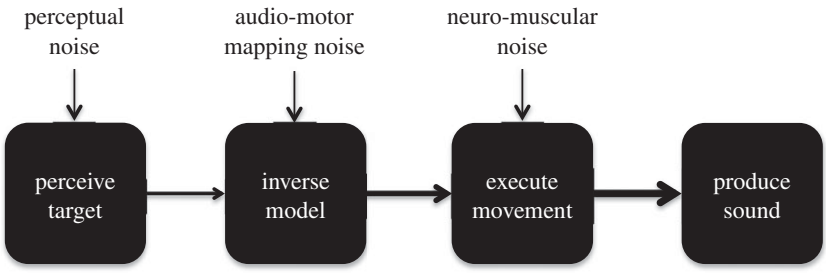

Figure 5. Propagation of error model. Audio-motor imitation requires singers to listen to a target melody, compute an inverse model that maps the target melody onto a sequence of movements that will reproduce it, then execute those movements. We posit that computational noise at each stage of this process may be propagated to subsequent stages. Perceiving target melodies probably engages similar processes within the auditory system for both singing and whistling, and computational noise in this system may explain the high degree of correlation between singing and whistling imprecision. Executing movements probably engages muscle-specific domains within the motor system and may explain why levels of imprecision differed for melodies sung with the larynx versus whistled with the tongue.

of inverse-model processes in the striatum for different muscle groups. Indeed, the sensorimotor computations of the inverse model have been proposed to be a key deficit in poor-pitch singing [47].

The movement outcomes of motor commands are assessed by a cerebro-cerebellar forward model network that compares the intended motor command to feedback from the sensory periphery [160–162]. Although the forward model was conceived as a mechanism that corrects motor commands based on proprioceptive feedback, it may also use auditory feedback for movements that produce a sound [163]. Since the cerebellum also contains somatotopic representations of the body [164], neuronal noise in forward model processes may also contribute to effector-specific imprecision.

We also observed a correlation between singing and whistling precision abilities after controlling for the mutual influence of perceptual abilities. There may be additional factors that have a mutual influence on singing and whistling production. One candidate is the mutual influence of respiratory motor-control, as expiration provides the mechanical drive for both singing and whistling. For both modes of sound production, increases in sound pressure level are related to higher frequencies [165,166]. Hence, fluctuations in expiatory flow may translate into unstable singing and whistling. Computational noise from shared processes such as respiration, together with independent and effector-specific noise in motor execution processes, may explain the strong correlation between singing and whistling imprecision with a consistent shift towards lower levels of precision in singing.

4.4. Vocal laziness

We observed no correlation between singing and whistling accuracy after controlling for perceptual ability, but instead observed a consistent bias to sing flat. This replicated a previous finding that untrained singers tend to compress pitches towards a habitual range [44]. From these exploratory analyses, we hypothesize that each individual's larynx may have a preferred frequency that it produces in a default configuration. Pitches produced at a participant's most accurate note may require the least muscular effort, while pitches further from this preferred note, in either direction, may require greater muscular effort.

The cricothyroid (CT) and thyroarytenoid (TA) muscles are the primary regulators of vocal-pitch in mammals. Contraction of the CT muscle rocks the thyroid cartilage forward, thereby stretching and increasing the tension of the vocal folds, and causing them to vibrate at a higher fundamental frequency (F_0) [76,78–80,167]. The TA muscle may relax the vocal folds, and in that sense acts as an antagonist to the CT muscle to decrease F_0. However, the role of the TA muscle is complicated by strong interactions with the state of the CT muscle [86,168,169].

Electromyographic studies have not examined the muscular profiles of pitch levels above compared to below participant specific habitual levels. However, as F_0 decreases to the level where the CT muscle is no longer active, other laryngeal muscles may become engaged [80]. A different but equally active process may be engaged when singers lower F_0 from the habitual level compared to raising it above the habitual level. An active process for producing lower than habitual pitches may explain why participants tended to sing sharp within this range because erring towards a habitual pitch may be less effortful than

erring away from it. Vocal laziness, or a conservation of vocal effort, may lead singers to compress sung notes towards the pitch associated with a default or preferred configuration of the larynx.

5. Conclusion

We report the results of a study on pitch imitation in singing with the laryngeal sound source compared with whistling with an oral sound source. Sung imitations were less precise than whistled imitations, although neither were on the order of precision that have previously been reported for continuous pitch instruments, such as the violin. While biological pitch production in general may be less reliable than instrumental pitch production, the neuro-motor control of the larynx for pitch production is particularly coarse. From the relatively recent evolution of vocal production learning in great apes, we suggest that evolution has not tuned the human vocal-motor system to the same degree as other neuro-muscular systems.

Ethics. The experimental protocol was approved by the Ethical Review Committee of Maastricht University. All participants provided written informed consent.

Data accessibility. The raw data, processed data and experimental materials are archived in Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.504t7 [69].

Authors' contributions. M.B. conceived of the study, designed the study, analysed the data and drafted the manuscript; J.F.J. collected the data; S.A.K. helped design the study and provided critical revision to the manuscript. All authors contributed comments and critical revisions. All authors gave final approval for publication.

Competing interests. The authors report no competing interests.

Funding. This work was funded by grants from the Auditory Cognitive Neuroscience Society (ACN) to M.B. and S.A.K. and the Biotechnological and Biological Sciences Research Council (BBSRC) of the UK to S.A.K. (BB/M009742/1).

Acknowledgements. We thank Dr Peter Pfordresher for critical comments on the manuscript, and the 'LOUD POINTS' Data Visualization group for statistical consultation.

References

1. Mercado E, Mantell JT, Pfordresher PQ. 2014 Imitating sounds: a cognitive approach to understanding vocal imitation. *Comp. Cogn. Behav. Rev.* **9**, 17–74. (doi:10.3819/ccbr.2014.90002)
2. Janik VM, Slater PJB. 2000 The different roles of social learning in vocal communication. *Anim. Behav.* **60**, 1–11. (doi:10.1006/anbe.2000.1410)
3. Petkov CI, Jarvis ED. 2012 Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* **4**, 1–24. (doi:10.3389/fnevo.2012.00012)
4. Janik VM. 2014 Cetacean vocal learning and communication. *Curr. Opin. Neurobiol.* **28**, 60–65. (doi:10.1016/j.conb.2014.06.010)
5. King S, Sayigh L. 2013 Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proc. R. Soc. B* **280**, 1–9. (doi:10.1098/rspb.2013.0053)
6. Noad MJ, Cato DH, Bryden MM, Jenner MN, Jenner KC. 2000 Cultural revolution in whale songs. *Nature* **408**, 537. (doi:10.1038/35046199)
7. Ralls K, Fiorelli P, Gish S. 1985 Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Can. J. Zool.* **63**, 1050–1056. (doi:10.1139/z85-157)
8. Sanvito S, Galimberti F, Miller EH. 2007 Observational evidences of vocal learning in southern elephant seals: a longitudinal study. *Ethology* **113**, 137–146. (doi:10.1111/j.1439-0310.2006.01306.x)
9. Schusterman RJ, Feinstein SH. 1965 Shaping and discriminative control of underwater click vocalizations in a California sea lion. *Science* **150**, 1743–1744. (doi:10.1126/science.150.3704.1743)
10. Reichmuth C, Casey C. 2014 Vocal learning in seals, sea lions, and walruses. *Curr. Opin. Neurobiol.* **28**, 66–71. (doi:10.1016/j.conb.2014.06.011)
11. Ravignani A, Fitch WT, Hanke FD, Heinrich T, Hurgitsch B, Kotz SA, Scharff C, Stoeger AS, de Boer B. 2016 What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Front. Neurosci.* **10**, 274. (doi:10.3389/fnins.2016.00274)
12. Knörnschild M, Nagy M, Metz M, Mayer F, von Helversen O. 2010 Complex vocal imitation during ontogeny in a bat. *Biol. Lett.* **6**, 156–159. (doi:10.1098/rsbl.2009.0685)
13. Vernes SC. 2016 What bats have to say about speech and language. *Psychon. Bull. Rev.* **24**, 111–117. (doi:10.3758/s13423-016-1060-3)
14. Stoeger AS, Mietchen D, Oh S, de Silva S, Herbst CT, Kwon S, Fitch WT. 2012 An Asian elephant imitates human speech. *Curr. Biol.* **22**, 2144–2148. (doi:10.1016/j.cub.2012.09.022)
15. Poole JH, Tyack PL, Stoeger-Horwath AS, Watwood S. 2005 Elephants are capable of vocal learning. *Nature* **434**, 455–456. (doi:10.1029/2001GL014051)
16. Brown S. 2017 A joint prosodic origin of language and music. *Front. Psychol.* **8**, 1–20. (doi:10.3389/fpsyg.2017.01894)
17. Pisanski K, Cartei V, McGettigan C, Raine J, Reby D. 2016 Voice modulation: a window into the origins of human vocal control? *Trends Cogn. Sci.* **20**, 304–318. (doi:10.1016/j.tics.2016.01.002)
18. Filippi P. 2016 Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Front. Psychol.* **7**, 1–19. (doi:10.3389/fpsyg.2016.01393)
19. Titze IR. 2008 Nonlinear source–filter coupling in phonation: theory. *J. Acoust. Soc. Am.* **123**, 2733. (doi:10.1121/1.2832337)
20. 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)
21. Fant G. 1960 *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.
22. Wich SA *et al.* 2012 Call cultures in orang-utans? *PLoS ONE* **7**, 1–9. (doi:10.1371/journal.pone.0036180)
23. Perlman M, Clark N. 2015 Learned vocal and breathing behavior in an enculturated gorilla. *Anim. Cogn.* **18**, 1165–1179. (doi:10.1007/s10071-015-0889-6)
24. Lameira AR, Hardus ME, Mielke A, Wich SA, Shumaker RW. 2016 Vocal fold control beyond the species-specific repertoire in an orangutan. *Sci. Rep.* **6**, 1–10. (doi:10.1038/srep30315)
25. Bergman TJ. 2013 Speech-like vocalized lip-smacking in geladas. *Curr. Biol.* **23**, R268–R2689. (doi:10.1016/j.cub.2013.02.038)
26. Hopkins WD, Taglialatela J, Leavens DA. 2007 Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Anim. Behav.* **73**, 281–286. (doi:10.1016/j.anbehav.2006.08.004)
27. Wich SA, Swartz KB, Hardus ME, Lameira AR, Stromberg E, Shumaker RW. 2009 A case of spontaneous acquisition of a human sound by an orangutan. *Primates* **50**, 56–64. (doi:10.1007/s10329-008-0117-y)
28. Morrill RJ, Paukner A, Ferrari PF, Ghazanfar AA. 2012 Monkey lipsmacking develops like the human

speech rhythm. *Dev. Sci.* **15**, 557–568. (doi:10.1111/j.1467-7687.2012.01149.x)

29. Ghazanfar AA, Morrill RJ, Kayser C. 2013 Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proc. Natl Acad. Sci. USA* **110**, 1959–1963. (doi:10.1073/pnas.1214956110)
30. Ghazanfar AA, Takahashi DY, Mathur N, Fitch WT. 2012 Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Curr. Biol.* **22**, 1176–1182. (doi:10.1016/j.cub.2012.04.055)
31. Fitch WT, de Boer B, Mathur N, Ghazanfar AA. 2016 Monkey vocal tracts are speech-ready. *Sci. Adv.* **2**, e1600723. (doi:10.1126/sciadv.1600723)
32. Boë L-J, Berthommier F, Legou T, Captier G, Kemp C, Sawallis TR, Becker Y, Rey A, Fagot J. 2017 Evidence of a vocalic proto-system in the baboon (*Papio papio*) suggests pre-Hominin speech precursors. *PLoS ONE* **12**, e0169321. (doi:10.1371/journal.pone.0169321)
33. Lisker L, Abramson AS. 1966 Some effects of context on voice onset time in English stops. *Lang. Speech* **10**, 1–28. (doi:10.1177/002383096701000101)
34. Lisker L, Abramson AS. 1964 A cross-language study of voicing in initial stops: acoustical measurements. *Word* **20**, 384–422. (doi:10.1080/00437956.1964.11659830)
35. Yip M. 2002 *Tone*. New York, NY: Cambridge University Press.
36. Lieberman P. 1960 Some acoustic correlates of word stress in American English. *J. Acoust. Soc. Am.* **32**, 22–25. (doi:10.1121/1.1908095)
37. Ladd DR, Morton R. 1997 The perception of intonational emphasis: continuous or categorical? *J. Phon.* **25**, 313–342. (doi:10.1006/jpho.1997.0046)
38. Ohala JJ. 1984 An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* **41**, 1–16. (doi:10.1159/000261706)
39. Banse R, Sherer K. 1996 Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* **70**, 614–636. (doi:10.1037/0022-3514.70.3.614)
40. Belyk M, Brown S. 2014 The acoustic correlates of valence depend on emotion family. *J. Voice* **28**, 523.e9–523.e18. (doi:10.1016/j.jvoice.2013.12.007)
41. Jürgens R, Hammerschmidt K, Fischer J. 2011 Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Front. Psychol.* **2**, 180–191. (doi:10.3389/fpsyg.2011.00180)
42. Lieberman P, Michaels SB. 1962 Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *J. Acoust. Soc. Am.* **34**, 922–927. (doi:10.1121/1.1918222)
43. Sauter DA, Scott SK. 2007 More than one kind of happiness: can we recognize vocal expressions of different positive states? *Motiv. Emot.* **31**, 192–199. (doi:10.1007/s11031-007-9065-x)
44. Pfordresher PQ, Brown S. 2007 Poor-pitch singing in the absence of 'tone deafness'. *Music Percept.* **25**, 95–115. (doi:10.1525/mp.2007.25.2.95)
45. Dalla Bella S, Giguère J-F, Peretz I. 2007 Singing proficiency in the general population. *J. Acoust. Soc. Am.* **121**, 1182. (doi:10.1121/1.2427111)
46. Bradshaw E, McHenry MA. 2005 Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *J. Voice* **19**, 431–439. (doi:10.1016/j.jvoice.2004.07.010)
47. Pfordresher PQ, Mantell JT. 2014 Singing with yourself: evidence for an inverse modeling account of poor-pitch singing. *Cogn. Psychol.* **70**, 31–57. (doi:10.1016/j.cogpsych.2013.12.005)
48. Pfordresher PQ, Brown S, Meier KM, Belyk M, Liotti M. 2010 Imprecise singing is widespread. *J. Acoust. Soc. Am.* **128**, 2182–2190. (doi:10.1121/1.3478782)
49. Pfordresher PQ, Larrouy-Maestri P. 2015 On drawing a line through the spectrogram: how do we understand deficits of vocal pitch imitation? *Front. Hum. Neurosci.* **9**, 1–12. (doi:10.3389/fnhum.2015.00271)
50. Dalla Bella S, Berkowska M. 2009 Singing proficiency in the majority: normality and 'phenotypes' of poor singing. *Ann. N. Y. Acad. Sci.* **1169**, 99–107. (doi:10.1111/j.1749-6632.2009.04558.x)
51. Pfordresher PQ, Brown S. 2016 Vocal mistuning reveals the origin of musical scales. *J. Cogn. Psychol.* **5911**, 1–18. (doi:10.1080/20445911.2015.1132024)
52. Geringer JM. 1978 Intonational performance and perception of ascending scales. *J. Res. Music Educ.* **26**, 32–40. (doi:abs/10.2307/3344787)
53. Vurma A, Ross J. 2006 Production and perception of musical intervals. *Music Percept.* **23**, 331–345. (doi:10.1525/mp.2006.23.4.331)
54. Hutchins S, Roquet C, Peretz I. 2012 The vocal generosity effect: how bad can your singing be? *Music Percept.* **30**, 147–160. (doi:10.1525/mp.2012.30.2.147)
55. Demorest SM. 2001 Pitch-matching performance of junior high boys: a comparison of perception and production. *Bull. Counc. Res. Music Educ.* **151**, 63–70. (doi:jstor.org/stable/40319118)
56. Hutchins SM, Peretz I. 2012 A frog in your throat or in your ear? Searching for the causes of poor singing. *J. Exp. Psychol. Gen.* **141**, 76–97. (doi:10.1037/a0025064)
57. Hutchins S, Larrouy-Maestri P, Peretz I. 2014 Singing ability is rooted in vocal-motor control of pitch. *Atten. Percept. Psychophys.* **76**, 1–9. (doi:10.3758/s13414-014-0732-1)
58. Lévêque Y, Giovanni A, Schön D. 2012 Pitch-matching in poor singers: human model advantage. *J. Voice* **26**, 293–298. (doi:10.1016/j.jvoice.2011.04.001)
59. Sherwood CC, Holloway RL, Erwin JM, Schleicher A, Zilles K, Hof PR. 2004 Cortical orofacial motor representation in Old World monkeys, great apes, and humans: I. Quantitative analysis of cytoarchitecture. *Brain. Behav. Evol.* **63**, 61–81. (doi:10.1159/000075672)
60. Neubert F-X, Mars RB, Sallet J, Rushworth MFS. 2015 Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proc. Natl Acad. Sci. USA* **11**, E2695–E2704. (doi:10.1073/pnas.1410767112)
61. Neubert FX, Mars RB, Thomas AG, Sallet J, Rushworth MFS. 2014 Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* **81**, 700–713. (doi:10.1016/j.neuron.2013.11.012)
62. Rilling JK, Glasser MF, Preuss TM, Ma X, Zhao T, Hu X, Behrens TEJ. 2008 The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* **11**, 426–428. (doi:10.1038/nn2072)
63. Kuypers HGJM. 1958 Corticobulbar connexions to the pons and lower brain-stem in man. *Brain* **81**, 364–388. (doi:10.1093/brain/81.3.364)
64. Iwatsubo T, Kuzuhara S, Kanemitsu A. 1990 Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. *Neurology* **40**, 309–312. (doi:10.1212/WNL.40.2.309)
65. Kuypers HGJM. 1958 Some projections from the peri-central cortex to the pons and lower brain stem in monkey and chimpanzee. *J. Comp. Neurol.* **110**, 221–255. (doi:10.1002/cne.901100205)
66. Jürgens U, Ehrenreich L. 2007 The descending motorcortical pathway to the laryngeal motoneurons in the squirrel monkey. *Brain Res.* **1148**, 90–95. (doi:10.1016/j.brainres.2007.02.020)
67. Simonyan K, Jürgens U. 2003 Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Res.* **974**, 43–59. (doi:10.1016/S0006-8993(03)02548-4)
68. Peretz I, Champod AS, Hyde K. 2003 Varieties of musical disorders: the Montreal battery of evaluation of amusia. *Ann. N. Y. Acad. Sci. USA* **999**, 58–75. (doi:10.1196/annals.1284.006)
69. Belyk M, Johnson JF, Kotz SA. 2018 Data from: Poor neuro-motor tuning of the human larynx: a comparison of sung and whistled pitch imitation. Dryad Digital Repository. (doi:10.5061/dryad.504t7)
70. Bates D, Maechler M, Bolker B, Walker S. 2015 Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48. (doi:10.18637/jss.v067.i01)
71. R Core Team 2017 R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. See https://www.R-project.org/.
72. Fox J, Weisberg S. 2011 *An {R} companion to applied regression*, 2nd edn. Thousand Oaks, CA: SAGE.
73. Schielzeth H. 2010 Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* **1**, 103–113. (doi:10.1111/j.2041-210X.2010.00012.x)
74. Gelman A, Yo Y-S. 2015 arm: Data analysis using regression and multilevel/hierarchical models. See https://cran.r-project.org/web/packages/arm/index.html.
75. Gelman A. 2008 Scaling regression inputs by dividing by two standard deviations. *Stat. Med.* **27**, 2865–2873. (doi:10.1002/sim.3107)
76. Buchthal F. 1959 Electromyography of intrinsic laryngeal muscles. *Q. J. Exp. Physiol. Cogn. Med. Sci.* **44**, 137–148. (doi:10.1177/107424841037638 2)
77. Gay EC, Hirose H, Strome M, Sawashima M. 1972 Electromyography of the intrinsic laryngeal muscles during phonation. *Ann. Otol. Rhinol. Laryngol.* **81**, 401–410. (doi:10.1177/000348947208100311)
78. Hollien H, Moore P. 1960 Measurements of the vocal folds during changes in pitch. *J. Speech Lang. Hear. Res.* **3**, 157–165. (doi:10.1044/jshr.0302.157)
79. Kempster GB, Larson CR, Kistler MK. 1988 Effects of electrical stimulation of cricothyroid and thyroarytenoid muscles on voice fundamental frequency. *J. Voice* **2**, 221–229. (doi:10.1016/S0892-1997(88)80080-8)

80. Roubeau B, Chevrie-Muller C, Saint Guily J. 1997 Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Otolaryngol.* **117**, 459–464. (doi:10.3109/00016489709113421)
81. Shipp T, Izdebski K. 1975 Vocal frequency and vertical larynx positioning by singers. *J. Acoust. Soc. Am.* **58**, 1104–1106. (doi:10.1121/1.380776)
82. Shipp T. 1987 Vertical laryngeal position: research findings and application for singers. *J. Voice* **1**, 217–219. (doi:10.1016/S0892-1997(87)80002-4)
83. Vilkman E, Sonninen A, Hurme P, Körkkö P. 1996 External laryngeal frame function in voice production revisited: a review. *J. Voice* **10**, 78–92. (doi:10.1016/S0892-1997(96)80021-X)
84. Sundberg J, Leanderson R, von Euler C. 1989 Activity relationship between diaphragm and cricothyroid muscles. *J. Voice* **3**, 225–232. (doi:10.1016/S0892-1997(89)80004-9)
85. Titze IR. 1989 On the relation between subglottal pressure and fundamental frequency in phonation. *J. Acoust. Soc. Am.* **85**, 901–906. (doi:0.1121/1.397562)
86. Titze IR, Luschei ES, Hirano M. 1989 Role of the thyroarytenoid muscle in regulation of fundamental frequency. *J. Voice* **3**, 213–224. (doi:10.1016/S0892-1997(89)80003-7)
87. Belyk M, Brown S. 2017 The origins of the vocal brain in humans. *Neurosci. Biobehav. Rev.* **77**, 177–193. (doi:10.1016/j.neubiorev.2017.03.014)
88. Jürgens U. 2009 The neural control of vocalization in mammals: a review. *J. Voice* **23**, 1–10. (doi:10.1016/j.jvoice.2007.07.005)
89. Jürgens U. 2002 Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* **26**, 235–258. (doi:10.1016/S0149-7634(01)00068-9)
90. Fitch WT. 2006 The biology and evolution of music: a comparative perspective. *Cognition* **100**, 173–215. (doi:10.1016/j.cognition.2005.11.009)
91. Brown S. 2000 The 'musilanguage' model of music evolution. In *The origins of music* (eds N Wallin, B Merker, S Brown), pp. 271–300. Cambridge, MA: MIT Press.
92. Mithen S. 2005 *The singing Neanderthals: the origins of music, language, mind, and body*. Cambridge, MA: Harvard University Press.
93. Bonin TL, Trainor LJ, Belyk M, Andrews PW. 2016 The source dilemma hypothesis: perceptual uncertainty contributes to musical emotion. *Cognition* **154**, 174–181. (doi:10.1016/j.cognition.2016.05.021)
94. Zatorre RJ, Gandour JT. 2008 Neural specializations for speech and pitch: moving beyond the dichotomies. *Phil. Trans. R. Soc. B* **363**, 1087–1104. (doi:10.1098/rstb.2007.2161)
95. Loui P, Wu EH, Wessel DL, Knight RT. 2009 A generalized mechanism for perception of pitch patterns. *J. Neurosci.* **29**, 454–459. (doi:10.1523/JNEUROSCI.4503-08.2009)
96. Trainor LJ. 2015 The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Proc. R. Soc. B* **379**, 20140089.
97. Schön D, Magne C, Besson M. 2004 The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* **41**, 341–349. (doi:10.1111/1469-8986.00172.x)
98. Belyk M, Brown S, Lim J, Kotz SA. 2017 Convergence of semantics and emotional expression within the IFG pars orbitalis. *Neuroimage* **156**, 240–248. (doi:10.1016/j.neuroimage.2017.04.020)
99. Belyk M, Brown S, Kotz SA. 2017 Demonstration and validation of kernel density estimation for spatial meta-analyses in cognitive neuroscience using simulated data. *Data Br.* **13**, 346–352. (doi:10.1016/j.dib.2017.06.003)
100. Koelsch S. 2005 Neural substrates of processing syntax and semantics in music. *Curr. Opin. Neurobiol.* **15**, 207–212. (doi:10.1016/j.conb.2005.03.005)
101. Koelsch S. 2011 Towards a neural basis of processing musical semantics. *Phys. Life Rev.* **8**, 89–105. (doi:10.1016/j.plrev.2011.04.004)
102. Binder JR, Desai RH, Graves WW, Conant L. 2009 Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19**, 2767–2796. (doi:10.1093/cercor/bhp055)
103. Patel AD. 2003 Language, music, syntax and the brain. *Nat. Neurosci.* **6**, 674–681. (doi:10.1038/nn1082)
104. Slevc LR, Rosenberg JC, Patel AD. 2009 Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* **16**, 374–381. (doi:10.3758/16.2.374)
105. Fitch WT. 2011 The evolution of syntax: an exaptationist perspective. *Front. Evol. Neurosci.* **3**, 1–12. (doi:10.3389/fnevo.2011.00009)
106. Penhune VB, Zattore RJ, Evans AC. 1998 Cerebellar contributions to motor timing: a PET study of auditory and visual rhythm reproduction. *J. Cogn. Neurosci.* **10**, 752–765. (doi:10.1162/089892998563149)
107. Abrams DA, Bhatara A, Ryali S, Balaban E, Levitin DJ, Menon V. 2011 Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex* **21**, 1507–1518. (doi:10.1093/cercor/bhq198)
108. Port RF. 2003 Meter and speech. *J. Phon.* **31**, 599–611. (doi:10.1016/j.wocn.2003.08.001)
109. Racette A, Bard C, Peretz I. 2006 Making non-fluent aphasics speak: sing along! *Brain* **129**, 2571–2584. (doi:10.1093/brain/awl250)
110. Brown S, Martinez MJ, Parsons LM. 2006 Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *Eur. J. Neurosci.* **23**, 2791–2803. (doi:10.1111/j.1460-9568.2006.04785.x)
111. Özdemir E, Norton A, Schlaug G. 2006 Shared and distinct neural correlates of singing and speaking. *Neuroimage* **33**, 628–635. (doi:10.1016/j.neuroimage.2006.07.013)
112. Zarate JM. 2013 The neural control of singing. *Front. Hum. Neurosci.* **7**, 1–12. (doi:10.3389/fnhum.2013.00237)
113. Honing H, ten Cate C, Peretz I, Trehub SE. 2015 Without it no music: cognition, biology and evolution of musicality. *Phil. Trans. R. Soc. B* **370**, 20140088. (doi:10.1098/rstb.2014.0088)
114. Hoeschele M, Merchant H, Kikuchi Y, Hattori Y, ten Cate C. 2015 Searching for the origins of musicality across species. *Phil. Trans. R. Soc. B* **370**, 20140094. (doi:10.1098/rstb.2014.0094)
115. Hauser MD, Chomsky N, Fitch WT. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)
116. Savage PE, Brown S, Sakai E, Currie TE. 2015 Statistical universals reveal the structures and functions of human music. *Proc. Natl Acad. Sci. USA* **112**, 8987–8992. (doi:10.1073/pnas.1414495112)
117. Nottebohm F. 1972 The origins of vocal learning. *Am. Nat.* **106**, 116–140. (doi:10.1086/282756)
118. Hayes KJ, Hayes C. 1951 The intellectual development of a home-raised chimpanzee. *Proc. Am. Philos. Soc.* **95**, 105–109.
119. Carey D, Krishnan S, Callaghan MF, Sereno MI, Dick F. 2017 Functional and quantitative MRI mapping of somatosensory representations of human supralaryngeal vocal tract. *Cereb. Cortex* **27**, 265–278. (doi:10.1093/cercor/bhw393)
120. Carey D, Miquel ME, Evans BG, Adank P, McGettigan C. 2017 Vocal tract images reveal neural representations of sensorimotor transformation during speech imitation. *Cereb. Cortex* **27**, 3064–3079. (doi:10.1093/cercor/bhx056)
121. Simmonds AJ, Leech R, Iverson P, Wise RJS. 2014 The response of the anterior striatum during adult human vocal learning. *J. Neurophysiol.* **112**, 792–801. (doi:10.1152/jn.00901.2013)
122. Segawa JA, Tourville JA, Beal DS, Guenther FH. 2015 The neural correlates of speech motor sequence learning. *J. Cogn. Neurosci.* **27**, 819–831. (doi:10.1162/jocn_a_00737)
123. Garnier M, Lamalle L, Sato M. 2013 Neural correlates of phonetic convergence and speech imitation. *Front. Psychol.* **4**, 1–15. (doi:10.3389/fpsyg.2013.00600)
124. Belyk M, Pfordresher PQ, Liotti M, Brown S. 2016 The neural basis of vocal pitch imitation in humans. *J. Cogn. Neurosci.* **28**, 621–635. (doi:10.1162/jocn_a_00914)
125. Petrides M, Cadoret G, Mackey S. 2005 Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature* **435**, 1235–1238. (doi:10.1038/nature03628)
126. Petrides M, Pandya DN. 2002 Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur. J. Neurosci.* **16**, 291–310. (doi:10.1046/j.1460-9568.2002.02090.x)
127. Shmuelof L, Krakauer JW. 2011 Are we ready for a natural history of motor learning? *Neuron* **72**, 469–476. (doi:10.1016/j.neuron.2011.10.017)
128. Leyton S, Sherrington C. 1917 Observations on the excitable cortex of the chimpanzee, organ-utan, and gorilla. *Exp. Physiol.* **11**, 135–222. (doi:10.1113/expphysiol.1917.sp000240)
129. Petrides M, Tomaiuolo F, Yeterian EH, Pandya DN. 2012 The prefrontal cortex: comparative architectonic organization in the human and the macaque monkey brains. *Cortex* **48**, 46–57. (doi:10.1016/j.cortex.2011.07.002)
130. Nambu A. 2011 Somatotopic organization of the primate basal ganglia. *Front. Neuroanat.* **5**, 1–9. (doi:10.3389/fnana.2011.00026)
131. Amiez C, Petrides M. 2014 Neuroimaging evidence of the anatomo-functional organization of the human cingulate motor areas. *Cereb. Cortex* **24**, 563–578. (doi:10.1093/cercor/bhs329)

132. Penfield W, Welch K. 1951 The supplementary motor area of the cerebral cortex: a clinical and experimental study. *Arch. Neurol. Psychiatry* **66**, 289–317. (doi:10.1001/archneurpsyc.1951.02320090038004)
133. Penfield W, Boldrey E. 1937 Somatic motor and sensory representations in the cerebral cortex of man as studied by electrical stimulation. *Brain* **60**, 389–443. (doi:10.1192/bjp.84.352.868-a)
134. Koziol L *et al.* 2014 Consensus paper: the cerebellum's role in movement and cognition. *Cerebellum* **13**, 151–177. (doi:10.1007/s12311-013-0511-x)
135. Cerkevich CM, Qi HX, Kaas JH. 2014 Corticocortical projections to representations of the teeth, tongue, and face in somatosensory area 3b of macaques. *J. Comp. Neurol.* **522**, 546–572. (doi:10.1002/cne.23426)
136. Coudé G, Ferrari PF, Rodà F, Maranesi M, Borelli E, Veroni V, Monti F, Rozzi S, Fogassi L. 2011 Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS ONE* **6**, 1–10. (doi:10.1371/journal.pone.0026822)
137. Hast MH, Fischer JM, Wetzel AB, Thompson VE. 1974 Cortical motor representation of the laryngeal muscles in *Macaca mulatta*. *Brain* **73**, 229–240. (doi:10.1016/0006-8993(74)91046-4)
138. Hast MH, Milojkvic R. 1966 The response of the vocal folds to electrical stimulation of the inferior frontal cortex of the squirrel monkey. *Acta Otolaryngol.* **61**, 196–204. (doi:10.3109/00016486609127056)
139. Jürgens U. 1974 On the elicitability of vocalization from the cortical larynx area. *Brain Res.* **81**, 564–566. (doi:10.1016/0006-8993(74)90853-1)
140. Belyk M, Brown S. 2016 Pitch underlies activation of the vocal system during affective vocalization. *Soc. Cogn. Affect. Neurosci.* **11**, 1078–1088. (doi:10.1093/scan/nsv074)
141. Belyk M, Brown S. 2014 Somatotopy of the extrinsic laryngeal muscles in the human sensorimotor cortex. *Behav. Brain Res.* **270**, 364–371. (doi:10.1016/j.bbr.2014.05.048)
142. Breshears JD, Molinaro AM, Chang EF. 2015 A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. *J. Neurosurg.* **123**, 340–349. (doi:10.3171/2014.11.JNS14889)
143. Brown S, Ngan E, Liotti M. 2008 A larynx area in the human motor cortex. *Cereb. Cortex* **18**, 837–845. (doi:10.1093/cercor/bhm131)
144. Foerster O. 1931 The cerebral cortex in man. *Lancet* **218**, 309–312. (doi:10.1016/S0140-6736(00)47063-7)
145. Jürgens U, Kirzinger A, von Cramon D. 1982 The effects of deep-reaching lesions in the cortical face area on phonation: a combined case report and experimental monkey study. *Cortex* **18**, 125–139. (doi:10.1016/S0010-9452(82)80024-5)
146. Loucks TMJ, Poletto CJ, Simonyan K, Reynolds CL, Ludlow CL. 2007 Human brain activation during phonation and exhalation: common volitional control for two upper airway functions. *Neuroimage* **36**, 131–143. (doi:10.1016/j.neuroimage.2007.01.049)
147. Simonyan K, Ostuni J, Ludlow CL, Horwitz B. 2009 Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *J. Neurosci.* **29**, 14 912–14 923. (doi:10.1523/JNEUROSCI.4897-09.2009)
148. Hines M. 1940 Movements elicited from precentral gyrus of adult chimpanzees by stimulation with sine wave currents. *J. Neurophysiol.* **3**, 442–466. (doi:10.1152/jn.1940.3.5.442)
149. Walker AE, Green HD. 1938 Electrical excitability of the motor face area: a comparative study in primates. *J. Neurophysiol.* **1**, 152–165. (doi:jn.physiology.org/content/1/2/152)
150. Cambell AW. 1904 Histological studies on the localisation of cerebral function. *Br. J. Psychiatry* **50**, 651–662. (doi:10.1192/bjp.50.211.651)
151. Brodmann K. 1909 *Localisation in the cerebral cortex*. 3rd edn. New York, NY: Springer.
152. Fischer J, Hammerschmidt K. 2011 Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. *Genes Brain Behav.* **10**, 17–27. (doi:10.1111/j.1601-183X.2010.00610.x)
153. Fitch WT. 2010 *The evolution of language*. Cambridge, UK: Cambridge University Press.
154. Jarvis ED. 2004 Learned birdsong and the neurobiology of human language. *Ann. N. Y. Acad. Sci.* **1016**, 749–777. (doi:10.1196/annals.1298.038)
155. Simonyan K, Horwitz B. 2011 Laryngeal motor cortex and control of speech in humans. *Neurosci.* **17**, 197–208. (doi:10.1177/1073858410386727)
156. Takai O, Brown S, Liotti M. 2010 Representation of the speech effectors in the human motor cortex: somatotopy or overlap? *Brain Lang.* **113**, 39–44. (doi:10.1016/j.bandl.2010.01.008)
157. Jarvis E. 2007 Neural systems for vocal learning in birds and humans: a synopsis. *J. Ornithol.* **148**, 35–44. (doi:10.1007/s10336-007-0243-0)
158. Pfenning AR *et al.* 2014 Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346**, 1–13. (doi:10.1126/science.1256846)
159. Vitek JL, Ashe J, DeLong MR, Alexander GE. 1994 Physiologic properties and somatotopic organization of the primate motor thalamus. *J. Neurophysiol.* **71**, 1498–1513. (doi:10.1152/jn.1994.71.4.1498)
160. Wolpert DM, Ghahramani Z, Jordan MI. 1995 An internal model for sensorimotor integration. *Science* **269**, 1880–1882. (doi:10.1126/science.7569931)
161. Scott S. 2004 Optimal feedback control and the neural basis of volitional motor control. *Nat. Rev. Neurosci.* **5**, 532–546. (doi:10.1038/nrn1427)
162. Ishikawa T, Tomatsu S, Izawa J, Kakei S. 2016 The cerebro-cerebellum: could it be loci of forward models? *Neurosci. Res.* **104**, 72–79. (doi:10.1016/j.neures.2015.12.003)
163. Knolle F, Schröger E, Baess P, Kotz SA. 2012 The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *J. Cogn. Neurosci.* **24**, 698–706. (doi:10.1162/jocn_a_00167)
164. Glickstein M, Sultan F, Voogd J. 2011 Functional localization in the cerebellum. *Cortex* **47**, 59–80. (doi:10.1016/j.cortex.2009.09.001)
165. Meyer J. 2004 Bioacoustics of human whistled languages: an alternative approach to the cognitive processes of language. *An. Acad. Bras. Cienc.* **76**, 405–412. (doi:/S0001-37652004000200033)
166. Gramming P, Sundberg J, Ternstrom S, Leanderson R, Perkins WH. 1988 Relationship between changes in voice pitch and loudness. *J. Voice* **2**, 118–126. (doi:10.1016/S0892-1997(88)80067-5)
167. Gay T, Rendell JK, Spiro J. 1994 Oral and laryngeal muscle coordination during swallowing. *Laryngoscope* **104**, 341–349. (doi:10.1288/00005537-199403000-00017)
168. Kochis-Jennings KA, Finnegan EM, Hoffman HT, Jaiswal S, Hull D. 2014 Cricothyroid muscle and thyroarytenoid muscle dominance in vocal register control: preliminary results. *J. Voice* **28**, 652.e21–652.e29. (doi:10.1016/j.jvoice.2014.01.017)
169. Lowell SY, Story BH. 2006 Simulated effects of cricothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration. *J. Acoust. Soc. Am.* **120**, 386–397. (doi:10.1121/1.2204442)