

Research



Cite this article: Pisanski K, Raine J, Reby D. 2020 Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries. *R. Soc. open sci.* **7**: 191642. <http://dx.doi.org/10.1098/rsos.191642>

Received: 18 September 2019

Accepted: 21 January 2020

Subject Category:

Psychology and cognitive neuroscience

Subject Areas:

behaviour/evolution/acoustics

Keywords:

nonverbal vocalization, communication, speech, vocal, fundamental frequency, emotion

Author for correspondence:

Katarzyna Pisanski

e-mail: kasiapisanski@gmail.com

[†]These authors contributed equally to this study.

[‡]Current address: Equipe Neuro-Ethologie Sensorielle, University of Lyon/Saint-Etienne, France.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4853439>.

Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries

Katarzyna Pisanski^{1,†}, Jordan Raine^{2,†}
and David Reby^{1,2,‡}

¹Equipe de Neuro-Ethologie Sensorielle ENES/CRNL, University of Lyon/Saint-Etienne, CNRS UMR5292, INSERM UMR_S 1028, Saint-Etienne, France

²Mammal Vocal Communication and Cognition Research Group, School of Psychology, University of Sussex, Brighton, UK

KP, 0000-0003-0992-2477; JR, 0000-0003-0504-6019; DR, 0000-0001-9261-1711

Fundamental frequency (F_0 , perceived as voice pitch) predicts sex and age, hormonal status, mating success and a range of social traits, and thus functions as an important biosocial marker in modal speech. Yet, the role of F_0 in human nonverbal vocalizations remains unclear, and given considerable variability in F_0 across call types, it is not known whether F_0 cues to vocalizer attributes are shared across speech and nonverbal vocalizations. Here, using a corpus of vocal sounds from 51 men and women, we examined whether individual differences in F_0 are retained across neutral speech, valenced speech and nonverbal vocalizations (screams, roars and pain cries). Acoustic analyses revealed substantial variability in F_0 across vocal types, with mean F_0 increasing as much as 10-fold in screams compared to speech in the same individual. Despite these extreme pitch differences, sexual dimorphism was preserved within call types and, critically, inter-individual differences in F_0 correlated across vocal types ($r=0.36-0.80$) with stronger relationships between vocal types of the same valence (e.g. 38% of the variance in roar F_0 was predicted by aggressive speech F_0). Our results indicate that biologically and socially relevant indexical cues in the human voice are preserved in simulated valenced speech and vocalizations, including vocalizations characterized by extreme F_0 modulation, suggesting that voice pitch may function as a reliable individual and biosocial marker across disparate communication contexts.

1. Introduction

In addition to being the carrier of language, the human voice has been shaped by selection to communicate biologically relevant traits of the vocalizer ([1] for review). Indexical cues to a vocalizer's identity, sex and age are readily transmitted by the non-linguistic properties of modal speech, particularly fundamental frequency (F_0), perceived as voice pitch ([2] for review). Fundamental frequency in modal speech corresponds to the rate at which the vocal folds vibrate, determined by their effective mass, length and tension. In addition to fluctuating dynamically during speech production within a single vocalizer ([3,4] for reviews), F_0 varies widely across individuals, both between the sexes (men's voices are lower-pitched than women's) and within adults of the same sex, due to a combination of endocrinological, physiological, anatomical and social factors [5].

Voice pitch plays an important role in human social interactions. In addition to reliably indicating sex and age, individual differences in the mean pitch of speech utterances facilitate speaker recognition, predict mating success and social status, and influence listeners' biosocial judgements of vocalizers ([2,6,7] for reviews). However, past work in this area has focused almost exclusively on the role of pitch in speech, largely ignoring the variable nature of F_0 across a much wider range of human vocal sounds, and thus overlooking the complexity of the human vocal apparatus and its evolved functions.

Humans can readily modulate their voice F_0 to express a range of emotions and motivations during speech production [3,4], and studies conducted in real-life contexts have shown that men and women alter their speech F_0 depending on social context or to whom they are speaking (see e.g. [8,9]). Yet beyond verbal communication, humans also frequently produce non-linguistic vocalizations ranging from laughter to screams that also contain salient F_0 cues of potential biological and social relevance [10]. Such vocalizations almost certainly emerged in the human vocal repertoire before words [3], and can exploit a broader, less constrained acoustic space than modal speech. Indeed, intelligible speech production necessitates a relatively steady F_0 ensured by an ability to maintain constant subglottal pressure [2], as well as a dense harmonic structure (thus relatively low F_0) to effectively encode phonetic information, including discriminable formant frequencies [11]. By contrast, human nonverbal vocalizations, which can resemble those of non-human mammals (e.g. [12]), are often characterized by extreme fluctuations in voice pitch and, like those of other mammals, their acoustic structure typically maps onto their purported biological or social function [13]. This form-function mapping across vocal types appears to function to exploit listeners' perceptual sensitivities and biases. For example, distress vocalizations (cries, screams) are typically several orders of magnitude louder and higher-pitched than modal speech, attracting attention and eliciting arousal [14]. Such distress vocalizations also exhibit different spectral profiles than those of agonistic or aggressive vocalizations (roars, grunts), which are typically characterized by a relatively low F_0 and a high proportion of nonlinear phenomena (e.g. [15,16]).

Given this considerable variability, an important question is whether acoustic cues to vocalizer attributes are shared across call types (e.g. as observed in red deer [17]) or are call-specific (e.g. as observed in zebra finches [18]). Indeed, in humans, no previous study has examined whether individual differences in F_0 are preserved across the broad range of human vocal sounds, from modal and emotionally valenced speech to nonverbal vocalizations. While it is possible that idiosyncratic differences in F_0 between different vocalizers might be overridden by the more extreme F_0 modulations that characterize agonistic and distress vocalizations, existing data on the F_0 profiles of human grunts [15], roars [16], laughs and cries [19] indicate that such vocalizations retain a degree of sexual dimorphism, wherein men produce relatively lower-pitched vocalizations than do women. There is also preliminary evidence that within each sex, F_0 in modal speech correlates with F_0 in sung speech [20], and that cues to individual identity are retained in valenced human speech [21], laughter [22], cries [23], and in the screams of both humans ([24], cf. [25]) and non-human primates [26] (with the caveat that speaker recognition is substantially reduced from these vocalizations compared to modal speech among human listeners [21,22,25]). Finally, longitudinal studies have recently revealed that individual differences in F_0 remain relatively stable across the lifespan, from infancy to childhood and throughout adulthood [27,28]. Taken together, this body of literature suggests that F_0 may represent a reliable and surprisingly stable individual marker despite its extreme dynamicity at the within-vocalizer level (e.g. across vocal types). From an evolutionary and ethological perspective, such F_0 stability could function to provide honest information about a vocalizer's identity and biosocial profile (e.g. dominance, masculinity) regardless of the mode of vocal production, or social context.

Here, we investigate the stability of individual differences in F_0 across speech and nonverbal vocalizations produced by the same men and women in neutral, aggressive, fearful and pain contexts, by contrasting within-individual variation in F_0 between and across these diverse call types.

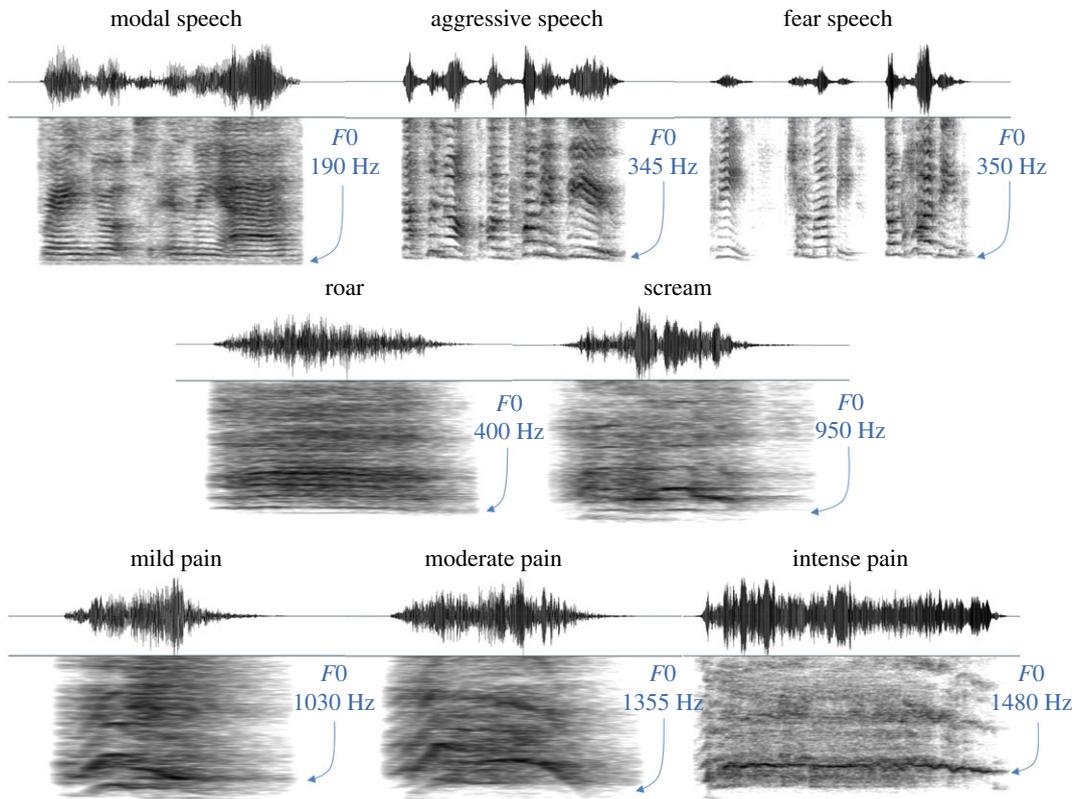


Figure 1. Waveforms and spectrograms representing each of eight vocal types produced by a single representative individual (female, aged 22), demonstrating the high degree of intra-individual variability in F_0 across vocal types.

2. Material and methods

2.1. Participants

We recorded the voices of 51 drama or acting students from the Royal Central School of Speech and Drama (London, UK) and the University of Sussex (Brighton, UK). The sample included 28 men (mean age 23.6, range 18–71) and 23 women (mean age 22.8, range 18–66) who provided informed consent and received monetary compensation. No participant reported conditions that could affect their voice (e.g. cold, sore throat, chronic smoking).

2.2. Voice recording

Participants were audio recorded in a quiet room using a Zoom H4n microphone placed at a distance of 150 cm to prevent audio clipping. Each vocalizer produced eight simulated ‘vocal types’: neutral speech, aggressive speech, fearful speech, an aggressive vocalization (i.e. ‘roar’), a fearful vocalization (i.e. scream), and pain vocalizations corresponding to mild, moderate and intense levels of pain (figure 1). Voice recordings were saved as WAV files at 44.1 kHz sampling frequency and 16-bit resolution, and later transferred to a laptop for acoustic analysis.

To obtain neutral modal speech, participants produced the first sentence of the Rainbow Passage [29] in a modal voice. They were then instructed to imagine themselves in various scenarios and to produce speech sentences and/or vocalizations to express their motivation and emotion in each given context. A description of the context or speech sentence was dictated by the experimenter and also displayed on a computer screen. In the aggression and fear contexts, participants imagined themselves on a battlefield, either attacking or being attacked, and produced the sentences, ‘That’s enough, I’m coming for you!’ and, ‘Please, show mercy, don’t hurt me!’, respectively. They were further instructed to produce a nonverbal vocalization to express that same motivation. In the pain context, participants produced vocalizations in three imagined scenarios of increasing pain intensity: mild (e.g. stubbing their toe), moderate (e.g. breaking their arm) and intense (e.g. childbirth). Full context descriptions are given in Raine *et al.* [16,30].

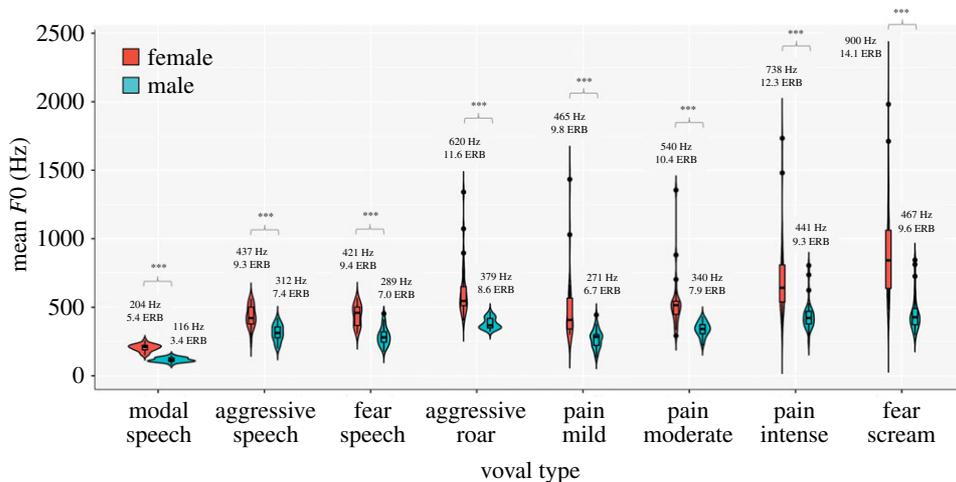


Figure 2. Violin plots representing the full distribution in mean F_0 for each sex and each vocal type. Mean F_0 values are given in both Hertz and ERBs above the violin plot for each sex and each vocal type. Significant sex differences were observed for each vocal type (** $p < 0.001$ following Šidák correction). Plots were produced in *R* ggplot 2 package.

2.3. Acoustic analysis

Acoustic editing and analysis were performed in PRAAT 5.3.62 [31]. Recordings were segregated by vocal type, resulting in 408 audio clips, and edited manually to remove silences or acute background noises. Fundamental frequency (F_0) was measured using a custom script with a search range of 60–2000 Hz, 0.05 s window length and 0.01 time step. Extracted F_0 contours were systematically inspected and verified, and any measurement errors (e.g. arising from octave jumps, sub-harmonics or deterministic chaos) were de-selected or corrected. This established method has been applied successfully in a number of studies to measure F_0 both in human speech and nonverbal vocalizations characterized by extreme F_0 values (e.g. babies cries [32], tennis grunts [15]). Measured F_0 was converted from hertz (Hz) to equivalent rectangular bandwidths (ERBs), a quasi-logarithmic scale that controls for any discrepancy between measured F_0 and perceived voice pitch.

2.4. Statistical analysis

A linear mixed model (LMM) fit by restricted maximum-likelihood estimation was first used to test for differences in F_0 (ERB) across vocal types and between sexes. Vocal type and sex of vocalizer were entered into the omnibus model as fixed variables, and vocalizer ID was entered as a random variable with random intercept. This was followed by separate LMMs for each vocal type to more closely examine sex differences. Significant effects were examined using pairwise tests with Šidák correction for multiple comparisons.

Correlation matrices (Spearman's ρ , r_s , one-tailed) were then conducted to test for positive relationships in F_0 between all vocal types, separately for each sex and each F_0 scale (ERB, Hz). Shapiro–Wilk tests indicated that F_0 was not normally distributed in nonverbal vocalizations (see electronic supplementary material, table S1), hence non-parametric tests were used for regression analyses. Bootstrapping was used to compute 95% confidence intervals for each bivariate correlation. The Benjamini–Hochberg procedure ([33], where $m = 28$, $q = 0.1$) was employed to control for the inflated false discovery rate owing to multiple comparisons. Datasets and statistical scripts are provided as electronic supplementary material.

3. Results

3.1. Sexual dimorphism in F_0 across vocal types

Figure 2 illustrates distributions in mean F_0 across the eight vocal types for each sex. For illustrative purposes F_0 values are plotted in Hertz along the y -axis and means are given in both Hertz and ERBs above each violin plot.

The omnibus LMM was significant (intercept: $F_{1,49} = 4949.6$, $p < 0.001$) and showed significant effects of vocalizer sex ($F_{1,49} = 123.1$, $p < 0.001$), vocal type ($F_{7,343} = 121.4$, $p < 0.001$) and a significant interaction between sex and vocal type ($F_{7,343} = 4.1$, $p < 0.001$) on mean F_0 .

Separate LMMs conducted for each vocal type confirmed that men's mean F_0 was significantly lower than women's for each vocal type, including modal speech, all valenced speech and nonverbal vocalizations (all $F_{1,49} > 23.7$, all $p < 0.001$; see electronic supplementary material, table S2 for pairwise comparisons, where all $p < 0.001$ following Šidák correction). Sexual dimorphism was particularly pronounced for nonverbal vocalizations, wherein the F_0 s of women's fear screams and aggressive roars were on average 4.5 ERBs (433 Hz) and 3 ERBs (241 Hz) higher than men's, respectively. The F_0 s of women's pain cries ranged from an average of 2.6–3.2 ERBs (194–297 Hz) higher than men's (figure 2; electronic supplementary material, table S2).

3.2. Stability of individual differences in F_0 across vocal types

Within each sex, significant moderate to strong positive relationships between F_0 s ($r = 0.36$ – 0.80) were also observed across various vocal types (table 1; see electronic supplementary material, table S4 for comparable results in Hz). For significant correlations, the mean F_0 of a given vocal type explained between 13 and 64% of the variance in the mean F_0 of another vocal type, on average, within the same sample of same-sex vocalizers (table 1).

In both sexes, F_0 measured from aggressive speech correlated strongly with F_0 measured from fearful speech. Moreover, both of these valenced speech types predicted the mean F_0 of corresponding nonverbal vocalizations, where the relationships were largely valence-specific: hence, aggressive speech F_0 predicted roar F_0 , whereas fearful speech F_0 predicted scream F_0 . Following the same valence-specificity, roar F_0 did not significantly predict scream F_0 in either sex (though these weak relationships approached significance, $r = 0.26$ and 0.30 , $p < 0.10$). Within pain vocalizations, F_0 correlated across at least two pain intensity levels in either sex, with significant relationships between mild and moderate or intense pain (table 1; electronic supplementary material, table S3).

While these results were generally consistent between the sexes, some sex differences emerged. The F_0 s of pain vocalizations correlated with roar F_0 in men only, and with scream F_0 in women only. Interestingly, F_0 measured from modal speech was a relatively poor predictor of F_0 in emotionally valenced speech or nonverbal vocalizations, particularly among women, with the exception that modal speech predicted men's fearful speech and pain vocalizations.

4. Discussion

We show that individual differences in human fundamental frequency (F_0 , perceived as voice pitch), previously observed in neutral speech and known to function as indexical and social signals in human conversation ([2,6,7] for reviews), are also present in valenced speech and in simulated non-linguistic vocal sounds, including those characterized by extreme F_0 values (e.g. screams reaching nearly 2000 Hz). We further show that, despite a high degree of variability in F_0 across eight different vocal types, between-individual differences in F_0 are preserved across speech and vocalizations in a largely valence-specific manner. For instance, the mean F_0 of men and women's aggressive speech sentences reliably predicted the mean F_0 of their roars, with an analogous relationship between the F_0 s of fearful speech and screams. Individual differences in F_0 were also preserved across pain vocalizations representing varying levels of pain intensity. All vocal types also retained sexual dimorphism and thus were significantly lower-pitched among men than women, with the most salient sex differences observed for screams, roars and pain cries.

Screams showed the most extreme F_0 values overall (figure 1) and in turn the greatest degree of variability among vocalizers both between and within sexes. For example, the mean F_0 s of women's screams ranged from 484 to 1981 Hz, representing a difference of almost 1500 Hz (11 ERBs) between the highest- and lowest-pitched women. The fearful speech was by contrast much more constrained, with mean F_0 s ranging between 307 and 570 Hz in women (a maximum difference of 263 Hz or 3.7 ERBs between women). It is thus quite remarkable that individual differences in the F_0 s of valenced speech predicted those of such extreme vocalizations. By contrast, modal speech showed the lowest degree of F_0 variability, with a maximum difference of only 112 Hz or 2.3 ERBs between women (F_0 range 148–260 Hz) and 72 Hz or 1.8 ERBs between men (F_0 range 81–153 Hz). Moreover, the F_0 of modal speech was a poor predictor of the F_0 of both valenced speech and nonverbal vocalizations, particularly among women for whom modal speech did not predict the F_0 of any other vocal type. These results suggest that, in addition to valence-specificity, individual differences in F_0 are also more strongly preserved within the broader category of emotional (compared to neutral) sounds.

Table 1. Correlation matrix of F0 (ERB), pairwise comparisons across all vocal types. Results for men's voices are given in the top right portion of the table (significant relationships are highlighted in blue), and results for women's voices are given in the bottom left portion of the table (significant relationships are highlighted in beige). See electronic supplementary material, table S3 for exact *p*-values and electronic supplementary material, table S4 for analogous comparisons on Hz scale. See electronic supplementary material, figure S1 for corresponding scatterplot matrices.

| | statistic | modal speech | aggress speech | fear speech | aggress roar | fear scream | pain mild | pain mod | pain intense |
|----------------|-----------|--------------|----------------|-------------|--------------|-------------|-------------|-------------|--------------|
| modal speech | r_s | | -0.06 | 0.36* | 0.16 | 0.15 | 0.80** | 0.31 | 0.39* |
| | 95% CI | | -0.41, 0.30 | -0.07, 0.68 | -0.24, 0.50 | -0.29, 0.53 | 0.63, 0.89 | -0.12, 0.65 | 0.01, 0.69 |
| aggress speech | r_s | 0.09 | | 0.55** | 0.43* | 0.26 | 0.11 | 0.08 | 0.15 |
| | 95% CI | -0.33, 0.47 | | 0.15, 0.77 | 0.04, 0.75 | -0.13, 0.53 | -0.30, 0.48 | -0.28, 0.45 | -0.22, 0.48 |
| fear speech | r_s | 0.04 | 0.46* | | 0.41* | 0.54** | 0.50** | 0.12 | 0.26 |
| | 95% CI | -0.47, 0.51 | 0.09, 0.75 | | 0.03, 0.68 | 0.16, 0.77 | 0.10, 0.75 | -0.23, 0.45 | -0.12, 0.58 |
| aggress roar | r_s | 0.23 | 0.62** | 0.32 | | 0.26 | 0.08 | 0.22 | 0.55** |
| | 95% CI | -0.19, 0.58 | 0.30, 0.83 | -0.06, 0.64 | | -0.14, 0.58 | -0.29, 0.45 | -0.18, 0.56 | 0.23, 0.74 |
| fear scream | r_s | -0.08 | 0.30 | 0.45* | 0.30 | | 0.26 | 0.28 | 0.25 |
| | 95% CI | -0.51, 0.38 | -0.14, 0.71 | 0.10, 0.71 | -0.13, 0.63 | | -0.18, 0.61 | -0.14, 0.65 | -0.13, 0.57 |
| pain mild | r_s | 0.15 | -0.17 | -0.13 | -0.01 | 0.55** | | 0.39* | 0.28 |
| | 95% CI | -0.28, 0.48 | -0.59, 0.31 | -0.56, 0.29 | -0.44, 0.47 | 0.20, 0.80 | | -0.03, 0.67 | -0.12, 0.62 |
| pain mod | r_s | -0.07 | -0.12 | 0.07 | -0.17 | 0.25 | 0.43* | | 0.30 |
| | 95% CI | -0.49, 0.36 | -0.51, 0.30 | -0.40, 0.49 | -0.60, 0.30 | -0.13, 0.54 | -0.01, 0.74 | | -0.11, 0.67 |
| pain intense | r_s | -0.06 | 0.18 | 0.12 | 0.03 | 0.64** | 0.44* | 0.38× | |
| | 95% CI | -0.45, 0.35 | -0.25, 0.66 | -0.36, 0.57 | -0.42, 0.51 | 0.34, 0.81 | -0.03, 0.80 | -0.05, 0.70 | |

Notes: Spearman's ρ (r_s) correlation coefficients are followed by lower and upper 95% confidence intervals (bootstrapping, 1000 samples). Highlighted cells show significant relationships where ** $p < 0.01$, * $p < 0.05$. A single relationship (marked ×) did not pass Benjamini–Hochberg correction (unadjusted $p = 0.04$).

These results complement two longitudinal studies that recently showed that the $F0$ of babies' cries predicts the $F0$ of their speech later in childhood [28], and that individual differences in $F0$ stabilize after puberty [27], suggesting that voice pitch is determined early in life and remains fairly stable throughout adulthood [27,28]. Those findings, together with the results of the present study and other recent work showing that individuals can generally be recognized when producing emotional speech [21], laughter [22], cries [23] and screams ([24], cf. [25]), suggest that the human voice functions as a stable individual marker despite its extreme dynamicity. With regards to voice $F0$, this may be due to anatomical constraints (probably linked to individual differences in vocal fold length) that maintain the relative inter-individual differences in $F0$ across the typical ranges for each vocal type. While few studies have investigated the stability of $F0$ differences across the vocal repertoires of other species, there is evidence that cues to individual identity may also be preserved across call types in non-human mammals, including red deer [17], rhesus macaques [34] and cows [35].

The results of the present study warrant replication on a larger, more representative sample, as the current study is unpowered with the small sample size of 28 men and 23 women. As a consequence, our analyses could fail to detect statistical significance in weak and even moderate correlations ($r = 0.10$ – 0.30). This is further illustrated by the broad confidence intervals observed for most correlations (table 1). Replication studies should also include individuals without acting experience, as the potential influence of vocal training on the spectral and perceptual dimensions of speech and nonverbal vocalizations remains unclear (see e.g. [36,37]). Finally, given the influence of social and cultural factors on voice production and vocal expressions of emotion, additional research is also needed to examine the generalizability of these results to other cultures.

Extending the present line of inquiry, follow-up studies may include a broader range of speech and vocalizations (e.g. positively valenced), and could test whether modal speech $F0$, which proved to be a poor predictor in the present study, correlates with the $F0$ of less extreme and comparatively 'neutral' vocalizations such as yawns. As researchers have observed some differences between simulated and spontaneous vocal sounds (e.g. identity-related information is more reliably encoded in genuine laughter [19,22]), future studies may also compare $F0$ relationships in volitional versus spontaneous vocal types, including naturally occurring vocal displays produced in real-life social contexts.

The key implication of stable inter-individual differences in voice pitch across vocal types is that honest biosocial cues about the speaker are likely to be communicated regardless of the mode of vocal production. Thus, our results also suggest that individual differences in $F0$ may be preserved across social (and emotional) contexts, though this hypothesis should be explicitly tested. This work also contributes to a small but growing body of research examining form and function in human nonverbal vocalizations, which, despite being understudied, constitute a substantial portion of the human vocal repertoire and promise to offer new insight into the evolutionary origins of vocal communication and the evolution of speech [3,10,12].

Ethics. The study was reviewed and approved by the Sciences and Technology Cross-Schools Research Ethics Committee (C-REC) of the University of Sussex (ER/JR307/4).

Data accessibility. The datasets and tables supporting this article are available as electronic supplementary material.

Authors' contributions. D.R., J.R. and K.P. designed the investigation. J.R. collected the data. J.R. and K.P. performed acoustic analyses. D.R. and K.P. performed statistical analyses. K.P. wrote the original manuscript and created figures and tables. K.P., J.R. and D.R. edited, revised and approved the final manuscript, and agree to be accountable for the work.

Competing interests. We declare we have no competing interests.

Funding. This work was partly supported by the University of Sussex through a scholarship to J.R. Author K.P. was supported by the University of Lyon IDEXLYON project as part of the 'Programme Investissements d'Avenir' (ANR-16-IDEX-0005) to D.R., and by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 655859.

References

- Owren MJ. 2011 Human voice in evolutionary perspective. *Acoust. Today* **7**, 24–33. (doi:10.1121/1.3684225)
- Kreiman J, Sidtis D. 2011 *Foundations of voice studies: an interdisciplinary approach to voice production and perception*. Chichester, UK: Wiley-Blackwell. See <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0631222979.html>.
- 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)
- Lavan N, Burton AM, Scott SK, McGettigan C. 2019 Flexible voices: identity perception from variable vocal signals. *Psychon. Bull. Rev.* **26**, 90–102. (doi:10.3758/s13423-018-1497-7)
- Titze IR. 1989 Physiologic and acoustic differences between male and female voices. *J. Acoust. Soc. Am.* **85**, 1699–1707. (doi:10.1121/1.397959)
- Mathias SR, von Kriegstein K. 2014 How do we recognise who is speaking. *Front. Biosci. (Schol. Ed.)* **6**, 92–109. (doi:10.2741/S417)

7. Puts D. 2016 Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proc. R. Soc. B* **283**, 20152830. (doi:10.1098/rspb.2015.2830)
8. Pisanski K, Oleszkiewicz A, Plachetka J, Gmiterek M, Reby D. 2019 Voice pitch modulation in human mate choice. *Proc. R. Soc. B* **285**, 20181634. (doi:10.1098/rspb.2018.1634)
9. Sorokowski P *et al.* 2019 Voice of authority: professionals lower their vocal frequencies when giving expert advice. *J. Nonverbal Behav.* **43**, 257–269. (doi:10.1007/s10919-019-00307-0)
10. Anikin A, Bääth R, Persson T. 2018 Human non-linguistic vocal repertoire: call types and their meaning. *J. Nonverbal Behav.* **42**, 53–80. (doi:10.1007/s10919-017-0267-y)
11. Ryalls JH, Lieberman P. 1982 Fundamental frequency and vowel perception. *J. Acoust. Soc. Am.* **72**, 1631–1634. (doi:10.1121/1.388499)
12. Bryant GA, Aktipis CA. 2014 The animal nature of spontaneous human laughter. *Evol. Hum. Behav.* **35**, 327–335. (doi:10.1016/j.evolhumbehav.2014.03.003)
13. Morton ES. 1977 On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* **111**, 855–869. (doi:10.1086/283219)
14. Pisanski K, Bryant GA. 2019 The evolution of voice perception. In *The Oxford handbook of voice studies* (eds NS Eidsheim, KL Meizel). New York, NY: Oxford University Press.
15. Raine J, Pisanski K, Reby D. 2017 Tennis grunts communicate acoustic cues to sex and contest outcome. *Anim. Behav.* **130**, 47–55. (doi:10.1016/j.anbehav.2017.06.022)
16. Raine J, Pisanski K, Bond R, Simmer J, Reby D. 2019 Human roars communicate upper-body strength more effectively than do screams or aggressive and distressed speech. *PLoS ONE* **14**, e0213034. (doi:10.1371/journal.pone.0213034)
17. Reby D, André-Obrecht R, Galinier A, Farinas J, Cargnelutti B. 2006 Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (*Cervus elaphus*) stags. *J. Acoust. Soc. Am.* **120**, 4080–4089. (doi:10.1121/1.2358006)
18. Elie JE, Theunissen FE. 2018 Zebra finches identify individuals using vocal signatures unique to each call type. *Nat. Commun.* **9**, 4026. (doi:10.1038/s41467-018-06394-9)
19. Lavan N, Domone A, Fisher B, Kenigzstein N, Scott SK, McGettigan C. 2019 Speaker sex perception from spontaneous and volitional nonverbal vocalizations. *J. Nonverbal Behav.* **43**, 1–22. (doi:10.1007/s10919-018-0289-0)
20. Valentova JV, Tureček P, Varella MAC, Mendes FDC, Pereira KJ, Kubicova L, Stolarova P, Havlicek J. 2019 Vocal parameters of speech and singing covary and are related to vocal attractiveness, body measures, and sociosexuality: a cross-cultural study. *Front. Psychol.* **10**, 2029. (doi:10.3389/fpsyg.2019.02029)
21. Lavan N, Burston LF, Ladwa P, Merriman SE, Knight S, McGettigan C. 2018 Breaking voice identity perception: expressive voices are more confusable for listeners. *Q. J. Exp. Psychol.* **72**, 2240–2248. (doi:10.1177/1747021819836890)
22. Lavan N, Short B, Wilding A, McGettigan C. 2018 Impoverished encoding of speaker identity in spontaneous laughter. *Evol. Hum. Behav.* **39**, 139–145. (doi:10.1016/j.evolhumbehav.2017.11.002)
23. Gustafsson E, Levrero F, Reby D, Mathevon N. 2013 Fathers are just as good as mothers at recognizing the cries of their baby. *Nat. Commun.* **4**, 1698. (doi:10.1038/ncomms2713)
24. Engelberg JW, Schwartz JW, Gouzoules H. 2019 Do human screams permit individual recognition? *PeerJ* **7**, e7087. (doi:10.7717/peerj.7087)
25. Hansen JH. 1996 Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. *Speech Commun.* **20**, 151–173. (doi:10.1016/S0167-6393(96)00050-7)
26. Seyfarth RM, Cheney DL. 2015 The evolution of concepts about agents: or, what do animals recognize when they recognize an individual? In *The conceptual mind: new directions in the study of concepts* (eds E Margolis, S Laurence), pp. 57–75. Cambridge, MA: MIT Press.
27. Fouquet M, Pisanski K, Mathevon N, Reby D. 2016 Seven and up: individual differences in male voice fundamental frequency emerge before puberty and remain stable throughout adulthood. *R. Soc. open sci.* **3**, 160395. (doi:10.1098/rsos.160395)
28. Levrero F, Mathevon N, Pisanski K, Gustafsson E, Reby D. 2018 The pitch of babies' cries predicts their voice pitch at age 5. *Biol. Lett.* **14**, 20180065. (doi:10.1098/rsbl.2018.0065)
29. Fairbanks G. 1960 The rainbow passage. In *Voice and articulation drillbook*, pp. 124–139. New York, NY: Harper & Row.
30. Raine J, Pisanski K, Simmer J, Reby D. 2018 Vocal communication of simulated pain. *Biocoustics*, **28**, 1–23. (doi:10.1080/09524622.2018.1463295)
31. Boersma P, Weenink D. 2017 Praat: doing phonetics by computer. See <http://www.praat.org/>.
32. Reby D, Levrero F, Gustafsson E, Mathevon N. 2016 Sex stereotypes influence adults' perception of babies' cries. *BMC Psychol.* **4**, 19. (doi:10.1186/s40359-016-0123-6)
33. Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300.
34. Rendall D, Rodman PS, Emond RE. 1996 Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim. Behav.* **51**, 1007–1015. (doi:10.1006/anbe.1996.0103)
35. Green A, Clark C, Favaro L, Lomax S, Reby D. 2020 Vocal individuality of Holstein-Friesian cattle is maintained across putatively positive and negative farming contexts. *Sci. Rep.* **9**, 1–9. (doi:10.1038/s41598-019-54968-4) <https://www.nature.com/articles/s41598-019-54968-4>
36. Jürgens R, Grass A, Drolet M, Fischer J. 2015 Effect of acting experience on emotion expression and recognition in voice: non-actors provide better stimuli than expected. *J. Nonverbal Behav.* **39**, 195–214. (doi:10.1007/s10919-015-0209-5)
37. Livingstone SR, Choi D, Russo FA. 2014 The influence of vocal training and acting experience on measures of voice quality and emotional genuineness. *Front. Psychol.* **5**, 156. (doi:10.3389/fpsyg.2014.00156)