



HAL
open science

Human voice pitch measures are robust across a variety of speech recordings: methodological and theoretical implications

Katarzyna Kasia Pisanski, Agata Groyecka-Bernard, Piotr Sorokowski

► To cite this version:

Katarzyna Kasia Pisanski, Agata Groyecka-Bernard, Piotr Sorokowski. Human voice pitch measures are robust across a variety of speech recordings: methodological and theoretical implications. *Biology Letters*, 2021, 17 (9), 10.1098/rsbl.2021.0356 . hal-03501104

HAL Id: hal-03501104

<https://hal.science/hal-03501104>

Submitted on 8 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Research

Cite this article: Pisanski K, Groyecka-Bernard A, Sorokowski P. 2021 Human voice pitch measures are robust across a variety of speech recordings: methodological and theoretical implications. *Biol. Lett.* **17**: 20210356. <https://doi.org/10.1098/rsbl.2021.0356>

Received: 25 June 2021

Accepted: 6 September 2021

Subject Areas:

behaviour, biomechanics, evolution

Keywords:

fundamental frequency, speech, nonverbal communication, sexual selection, source-filter theory

Author for correspondence:

Katarzyna Pisanski

e-mail: katarzyna.pisanski@cnrs.fr

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5623613>.

Animal behaviour**Human voice pitch measures are robust across a variety of speech recordings: methodological and theoretical implications**Katarzyna Pisanski^{1,2,3}, Agata Groyecka-Bernard^{1,4} and Piotr Sorokowski¹¹University of Wrocław, Wrocław, Poland²CNRS/Centre National de la Recherche Scientifique, Laboratoire Dynamique du Langage, Université Lyon 2, Lyon, France³Equipe de Neuro-Ethologie Sensorielle, Centre de Recherche en Neurosciences de Lyon, Jean Monnet University of Saint-Etienne, France⁴Johannes Gutenberg-Universität Mainz, Mainz, Germany

KP, 0000-0003-0992-2477; AG-B, 0000-0002-1932-4828; PS, 0000-0001-9225-9965

Fundamental frequency (f_0), perceived as voice pitch, is the most sexually dimorphic, perceptually salient and intensively studied voice parameter in human nonverbal communication. Thousands of studies have linked human f_0 to biological and social speaker traits and life outcomes, from reproductive to economic. Critically, researchers have used myriad speech stimuli to measure f_0 and infer its functional relevance, from individual vowels to longer bouts of spontaneous speech. Here, we acoustically analysed f_0 in nearly 1000 affectively neutral speech utterances (vowels, words, counting, greetings, read paragraphs and free spontaneous speech) produced by the same 154 men and women, aged 18–67, with two aims: first, to test the methodological validity of comparing f_0 measures from diverse speech stimuli, and second, to test the prediction that the vast inter-individual differences in habitual f_0 found between same-sex adults are preserved across speech types. Indeed, despite differences in linguistic content, duration, scripted or spontaneous production and within-individual variability, we show that 42–81% of inter-individual differences in f_0 can be explained between any two speech types. Beyond methodological implications, together with recent evidence that inter-individual differences in f_0 are remarkably stable across the lifespan and generalize to emotional speech and nonverbal vocalizations, our results further substantiate voice pitch as a robust and reliable biomarker in human communication.

1. Introduction

Largely inspired by acoustic communication research in other animals ([1], review), fundamental frequency (f_0) is arguably the most intensively studied voice parameter in human nonverbal communication. Produced by vibration of the vocal folds in the larynx (the *source* of vocal output in most terrestrial mammals [2]), f_0 and its harmonics are perceived as voice pitch and are highly perceptually salient. While f_0 signals static individual differences such as sex, size and identity in many vertebrate species [1,3], it is more sexually dimorphic in human adults than in any other extant great ape [4] and has been repeatedly linked to testosterone levels, masculinity, dominance and social power ([5,6], reviews), as well as to mate preferences across diverse human cultures ([7], review). Thousands of studies have uncovered the communicative relevance of this source signal in humans, from predicting the outcomes of competitive

contests (e.g. fights [8], sports [9] and elections [10]), to predictive reproductive success ([6,7], reviews), thus providing strong converging evidence that voice pitch, particularly in men, has been intensely shaped by sexual selection to index biologically and socially relevant information [4,5,11].

Importantly, despite the fact that a person's voice pitch can vary considerably as they speak (i.e. *intra*-individual differences [12]), and that people can readily modify their pitch by tensing their vocal folds or modulating airflow from their lungs [13], for instance for prosodic emphasis [14], to communicate emotion and motivation in nonverbal vocalizations [15], or even to exaggerate biological traits like body size [16], there remain sizeable *inter*-individual (between individual) differences in baseline or habitual voice pitch. These individual differences are largely imposed by anatomical and physiological constraints on vocal production. Most notably, at puberty, a surge of testosterone permanently enlarges the male vocal folds and causes male f_0 to drop to nearly half the frequency of female f_0 , producing significant pitch differences between adults and children and between men and women [17].

Yet, habitual speech f_0 can vary substantially even among adults of the same sex. In this study, for example, men's average f_0 ranged from 78 to 182 Hz and women's from 126 to 307 Hz. Thus, the magnitude of pitch differences within sexes parallels that observed between sexes and greatly exceeds just-noticeable differences in pitch perception from speech [18]. It is the vast *inter*-individual, within-sex differences in voice pitch that have been repeatedly linked to individual differences in various biosocial traits. Recently, studies have further revealed that between-individual differences in mean f_0 emerge early in life and remain remarkably stable thereafter, as the cries of 4-month-old infants predict their speech f_0 in childhood [19], and pre-pubertal f_0 predicts post-pubertal f_0 in males across the lifespan [20].

Given the broad ecological relevance of voice pitch and its popularity in the human behavioural sciences, it is important to ascertain whether the vastly different types of voice stimuli used to measure and study f_0 are valid and comparable. Researchers interested in the functions of human nonverbal vocal parameters have long relied on monophthong vowels (a e i o u), consonant–vowel–consonant (CvC) words or counting [18,21,22]. These standard speech types enjoy linguistic neutrality, high cross-cultural comparability, steady f_0 and standardized formant contrasts. However, with the desire to increase ecological validity in the voice sciences came an increase in the complexity of speech stimuli used for acoustic analysis and playback experiments, from short greetings (e.g. [23,24]) to longer scripted paragraphs (e.g. [25]) and, finally, bouts of spontaneous or conversational free speech (e.g. [26,27]), sometimes extracted from online videos of real-life vocal exchanges [10,20]. Are f_0 measures comparable across this wide variety of speech utterances? While one recent study showed that listeners' judgements of dominance, trustworthiness and competence were similar for scripted words versus sentences produced by the same vocalizers [28], the authors did not objectively measure f_0 nor compare short utterances with longer speech. Here, we use a validated supervised extraction method to measure f_0 from nearly 1000 affectively neutral speech utterances in over 150 adult men and women, each producing single words, vowels, counting, greetings, read paragraphs and free spontaneous speech.

2. Methods

(a) Participants

We audio recorded 154 adults ($n = 83$ women, mean age 35.2 ± 1.3 , range 19–67; $n = 71$ men, mean age 29.9 ± 1.4 , range 18–65). Participants were recruited from the local community in a large European city (Wroclaw, Poland) using online and public adverts. While, based on previous work, strong effect sizes may be obtained with 25 vocalizers per sex and of a similar age [29], sample sizes here were increased to accommodate a much broader age range (18–67 years). No participant reported acute conditions that could affect their voice (e.g. cold, sore throat) and all provided informed consent.

(b) Voice recording

Participants were audio recorded in private sessions in a quiet room using a Zoom H4n microphone positioned 10 cm from the mouth. Voice recordings were saved as WAV files at 96 kHz sampling frequency and 16-bit resolution. Participants first familiarized themselves with a script containing six items, presented in a randomized order between participants, and were then asked to speak each item aloud. The six randomized items included: a series of five monophthong vowels (/a/ as in 'bra', /ɜ/ as in 'bird', /i/ as in 'bee', /ɔ/ as in 'bot', /u/ as in 'boot'); a series of five words (containing the same five vowels as above, but in LvT context, 'lot, lat, lej, lit, lud'); counting from 1 to 10; a greeting ('Hello, I am from ...'); a read paragraph (5 neutral sentences regarding the weather); and free speech, in which participants were instructed to say several spontaneous sentences about the weather. Weather is a relatively affectively neutral topic and it standardized content between the read paragraph and free speech. Original and translated recording scripts are available as electronic supplementary material. In addition to voice recording, participants completed a short demographic questionnaire and their height was measured with a metric tape.

(c) Acoustic analysis

Acoustic editing and analyses were performed in Praat v. 6.1.08 [30]. Voice recordings were segregated by speech type, resulting in six utterances per vocalizer (figure 1a) for a total of 924 utterances. To retain only a single item for the 'word' category, we analysed the central, steady-state word from the series. The duration was measured from the beginning to end of voicing and ranged from an average of 0.3 s (word) to 21.5 s (read paragraph; see electronic supplementary material, table S1).

Fundamental frequency parameters were measured from the full voiced duration of each utterance using a validated custom script and Praat's pitch-extraction algorithm and *path finder* function, with the recommended search range of 60 to 300 Hz for men and 100 to 600 Hz for women, and a 0.01 time step. The f_0 contour (*path*) was systematically extracted and manually inspected (figure 1a), and any erroneous frequency candidates in the selected path (e.g. arising from octave jumps) were deselected or corrected in the Pitch editor window before computing mean f_0 (average pitch across the utterance) and f_0 CV (coefficient of variation, a measure of pitch variability that controls for baseline f_0 , computed as f_0 s.d./ f_0 mean). These established protocols have been successfully applied in numerous studies (e.g. [29,31,32]). All acoustic parameters are summarized in the electronic supplementary material, table S1.

3. Results

Corroborating previous meta-analyses [33], we first confirmed that mean f_0 (averaged across speech types within individuals)

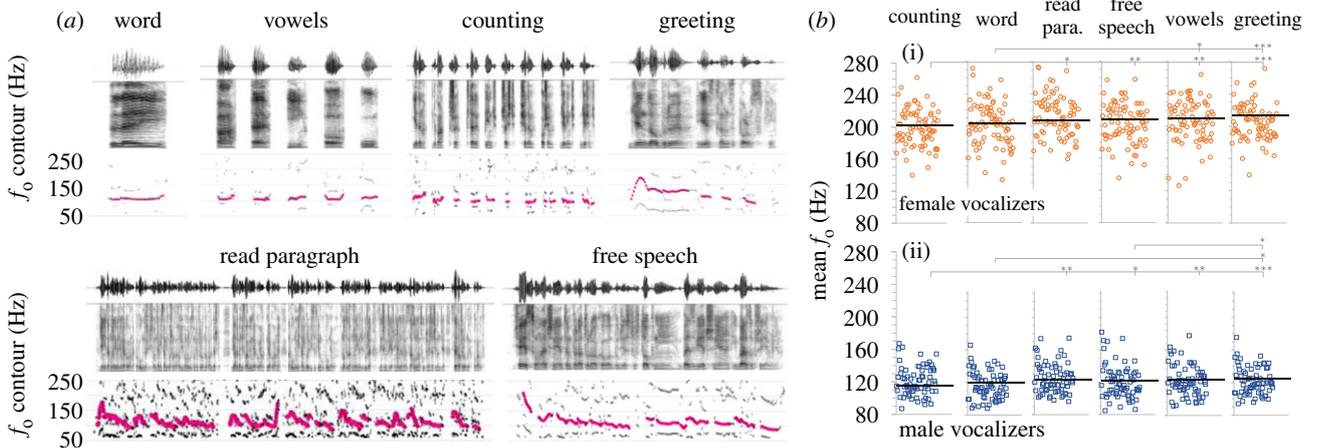


Figure 1. Voice pitch differences across vocalizers far outweigh differences across speech types. (a) Examples of the six speech types illustrated with waveforms and spectrograms (y -axis 0–5 kHz) from a single adult male, whose mean f_0 ranged from 105 Hz to 125 Hz across speech types. The f_0 contour (path) obtained using Praat's pitch tracking and path finder functions is shown in pink below each spectrogram (range 50–250 Hz). (b) Horizontal black bars indicate mean voice pitch (f_0 , Hz) averaged across all vocalizers for a given sex and speech type. Estimated marginal means and pairwise comparisons derive from a linear mixed model (LMM), where *** $p < 0.001$, ** $p < 0.01$ and * $p < 0.05$ following Šidák correction. Overlaid dot plots (jittered along with the x -axis for improved visualization) show the mean f_0 of each vocalizer plotted along the y -axis, females in (i) ($n = 83$, orange circles), males in (ii) ($n = 71$, blue squares).

did not explain a significant amount of variance in the heights of men (5%) nor women (3%; figure 2b), and thus we did not control for vocalizer height in further analyses. We did observe a known small and gradual decrease in women's but not men's mean f_0 with age (figure 2b), which may be attributed to a number of biological factors (see e.g. [34,35]). Nevertheless, controlling for vocalizer age in our models, the results of which are given below, did not significantly affect the strength of inter-individual f_0 relationships across speech types (see electronic supplementary material, table S2).

An omnibus linear mixed model (LMM) fitted by restricted maximum-likelihood estimation was then used to test for differences in mean f_0 across speech types. Speech type and sex of vocalizer were entered as fixed variables, and vocalizer identity and age as random variables with random intercept. The omnibus model showed significant effects of speech type ($F_{5,760} = 15.7$, $p < 0.001$) and sex ($F_{1,148.1} = 689.5$, $p < 0.001$) on mean f_0 , the latter owing to strong sexual dimorphism in voice pitch between adult men (average f_0 207 Hz) and women (121 Hz), but no interaction ($F_{5,760} = 1.5$, $p = 0.194$). We thus conducted analogous LMMs separately for each sex. These models confirmed that men's mean f_0 ($F_{5,350} = 8.1$, $p < 0.001$) and women's mean f_0 ($F_{5,410} = 9.7$, $p < 0.001$) both varied systematically across speech types.

Pairwise tests with Šidák correction for multiple comparisons revealed that, in both sexes, counting was characterized by the lowest pitch (male f_0 116.4, female 200.2 Hz) and greetings by the highest pitch (male f_0 124.5 Hz, female 212.6 Hz), with intermediate differences among other speech types (figure 1b). However, as further illustrated in figure 1b, although f_0 differed significantly between several speech types by 4 to 12 Hz (about one or two times the just-noticeable difference in pitch perception from modal speech [18]), these differences were nevertheless small relative to the much larger inter-individual variability in f_0 observed across vocalizers within each speech type. For instance, whereas women on average spoke with a voice pitch 12.4 Hz higher in greetings than while counting, there was a 132 Hz difference between the lowest-pitched woman (139.6 Hz) and highest-pitched woman (271.4 Hz) within the greeting category itself. Notably, there

were no differences in mean f_0 between reading a paragraph and producing free speech, nor between single words and a series of words (counting from 1 to 10; figure 1b).

Analogous LMMs showed that voice pitch variability (f_0 CV) also varied across speech types in men ($F_{5,348.2} = 8.7$, $p < 0.001$) and women ($F_{5,411.8} = 45.5$, $p < 0.001$); however, pairwise tests showed that this effect was largely driven by low pitch variability in single word utterances (see electronic supplementary material, table S2; figure 1a). While women spoke with a more dynamic pitch during greetings, pitch variability did not differ substantially among all other speech types, particularly among men, for whom vowels, counting, greeting, read paragraph and free speech were all characterized by similar pitch variability (electronic supplementary material, table S3).

To test our key hypothesis that the large inter-individual differences in mean f_0 observed within each speech type are preserved across speech types, such that individuals who produce the lowest/highest-pitched speech in one category likewise produce the lowest/highest-pitched speech in all other categories, we conducted a series of simple two-tailed regressions (Pearson's correlations, r). As illustrated in figure 2a, inter-individual differences in f_0 were indeed strongly preserved across all speech types. In both sexes, the strength of bivariate relationships between values of f_0 measured from two different speech types exceeded $r = 0.65$ in all cases and reached $r = 0.90$ (all $p < 3.17 \times 10^{-11}$; 95% bootstrapping confidence interval (CI) values $r = 0.45$ and 0.94 ; figure 2a). All correlations remained highly significant following Benjamini–Hochberg correction for multiple comparisons ([36], where $m = 15$ comparisons per sex, $q = 0.05$), and when controlling for vocalizer age (electronic supplementary material, table S2). These robust relationships indicate that typically more than half and as much as 80% of the variance in inter-individual f_0 measured from a given speech utterance could be explained by the f_0 measured from any other utterance, within the same sample of vocalizers (figure 2a). While effect sizes were unanimously strong, the read paragraph explained the most variance in the f_0 of other speech types for both sexes (63%–81%), whereas free spontaneous speech explained the least variance,

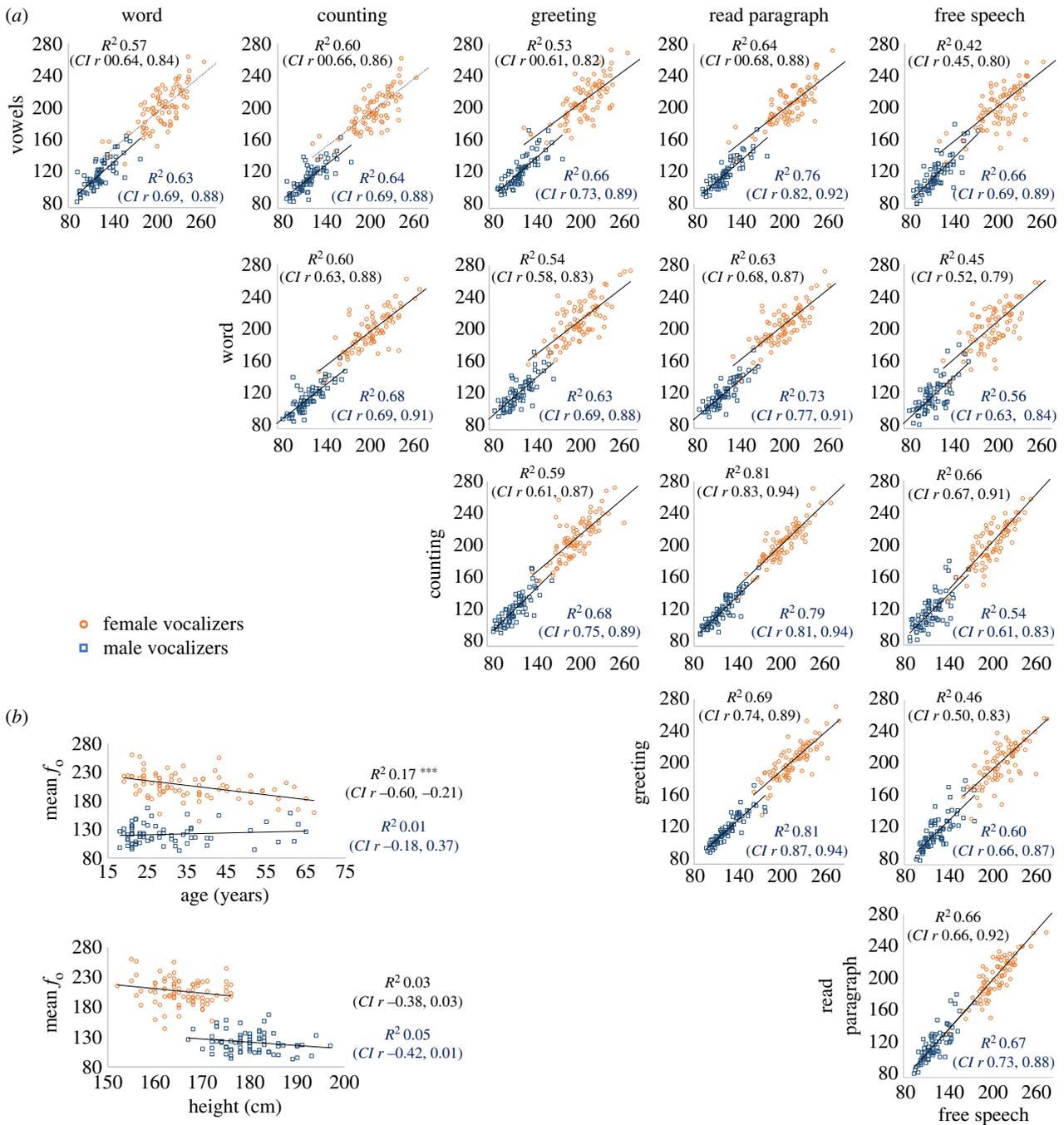


Figure 2. Inter-individual differences in voice pitch are strongly preserved across different speech utterances. (a) Correlation matrix comparing mean f_0 between all speech types for female ($n = 83$, orange circles) and male ($n = 71$, blue squares) vocalizers (Pearson's r , two-tailed, alpha 0.05). Mean f_0 plotted in Hz on both axes. All $p < 0.0000000000317$ following Benjamini–Hochberg correction. (b) Correlations between mean f_0 and vocalizer age or height for female ($n = 83$, orange circles) and male ($n = 71$, blue squares) vocalizers (Pearson's r , two-tailed, alpha 0.05), where only age predicted f_0 , and only in women ($***p < 0.01$). For all regressions, variance explained (R^2) between two speech types is given beside the regression line, followed by 95% CI derived from bootstrapping based on 1000 samples.

particularly in word, vowel and greeting f_0 (42–60%, figure 2a). Our LMM results suggest this latter result is not likely due to differences in pitch variability, as f_0CV did not differ between free speech and greetings or vowels in either sex (electronic supplementary material, table S1).

4. Discussion

We show that inter-individual differences in mean voice pitch (f_0) can be reliably and robustly measured from a variety of affectively neutral speech utterances including a single word, a series of vowels, counting, a short greeting, a longer scripted

paragraph or several sentences of spontaneous free speech. Despite differences in linguistic content and duration (500 ms to 20 s) and minor differences in speech variability (f_0CV) across these speech types, mean voice pitch measured from any of these utterances strongly predicted the pitch of every other speech utterance produced by the same individuals, explaining upwards of 80% of the variance. These results suggest that studies on human voice pitch are likely to produce comparable results whether f_0 measures are obtained from short, long, scripted or spontaneous speech. The results also support the validity of longitudinal analyses of mean f_0 measured from the same individuals at different time points, often from different speech utterances (e.g. [19,20,26]). It is

important to underscore that while this finding may allow a certain flexibility in the kinds of voice stimuli researchers can use to study between-individual differences in voice pitch, many other biologically relevant acoustic parameters cannot be compared between different kinds of voice stimuli, most notably formant frequencies.

Beyond these methodological implications, our results corroborate a growing number of studies showing that individual differences in voice pitch emerge early in life and are remarkably stable across an individual's lifetime [19,20], across diverse neutral speech utterances (this study), and even when comparing neutral speech with singing [37], with emotional speech [29,38] or with volitional nonverbal vocalizations such as screams and aggressive roars [29]. Thus, while the present study focuses on affectively neutral speech, past studies provide further evidence that between-person differences in voice pitch also generalize to emotional voice stimuli and remain stable as people age. This has theoretical implications for our understanding of the functions of voice pitch, a remarkably information-rich social and biological signal with clear evolutionary underpinnings [4,5] and real-life predictive power [9,10,26].

Stability in individual differences in voice pitch may also help to explain how human listeners can recognize vocalizers, even from extremely high-pitched volitional screams [39]. However, although our earlier work has shown that inter-individual differences in mean f_0 are preserved in emotional speech and vocalizations [29], the relationships are considerably less robust than those observed among modal speech types in the present study. Indeed, playback experiments have found that identity recognition is likewise degraded for emotional vocalizations [38,40], particularly authentic

vocalizations such as spontaneous laughs, compared with volitional (acted) laughs [40], the two being characterized by different pitch profiles [41]. Of course, recognizing speaker identity from the voice relies on much more than source characteristics (e.g. formant frequencies and temporal patterns also play a role [42]), or any single acoustic parameter [12]; however, that research does raise the possibility that inter-individual differences in f_0 may be less preserved in spontaneous than volitional vocal signals, a key prediction to test in future work. Moreover, while we show here that f_0 is preserved between scripted and spontaneous speech, researchers may also test the extent to which this stability generalizes to longer more naturalistic bouts of conversational speech produced in real-life contexts.

Ethics. The study was performed in accordance with the American Psychological Association's ethical standards for the treatment of human participants, including obtaining informed consent from all participants, and was approved by the Ethical Committee of the Institute of Psychology, University of Wrocław (project 2016/23/B/HS6/00771).

Data accessibility. The data are provided in the electronic supplementary material.

Authors' contributions. P.S., K.P. and A.G.-B. conceptualized, designed and conducted the study. K.P. performed statistical analyses, wrote the original manuscript and created figures and tables. All authors revised the manuscript, approved the final version, and agree to be accountable for all aspects of the work.

Competing interests. We declare we have no competing interests

Funding. This work was supported by a grant from the National Science Centre (grant no. 2016/23/B/HS6/00771). A.G.-B. was supported by the Foundation for Polish Science (FNP).

Acknowledgements. The authors thank Katarzyna Bugaj and Tomasz Frackowiak for assisting with voice recordings.

References

- Charlton BD, Pisanski K, Raine J, Reby D. 2020 Coding of static information in terrestrial mammal vocal signals. In *Animal signals and communication* (eds T Aubin, N Mathevon), pp. 115–136. Berlin, Germany: Springer Nature.
- Taylor AM, Reby D. 2010 The contribution of source-filter theory to mammal vocal communication research: advances in vocal communication research. *J. Zool.* **280**, 221–236. (doi:10.1111/j.1469-7998.2009.00661.x)
- Fitch WT, Hauser MD. 2003 Unpacking 'honesty': vertebrate vocal production and the evolution of acoustic signals. In *Acoustic communication* (eds AM Simmons, RR Fay, AN Popper), pp. 65–137. New York, NY: Springer.
- Puts D *et al.* 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)
- Aung T, Puts D. 2020 Voice pitch: a window into the communication of social power. *Curr. Opin. Psychol.* **33**, 154–161. (doi:10.1016/j.copsyc.2019.07.028)
- Pisanski K, Bryant GA. 2019 The evolution of voice perception. In *The Oxford handbook of voice studies* (eds NS Eidsheim, KL Meizel), pp. 269–300. New York, NY: Oxford University Press.
- Pisanski K, Feinberg DR. 2019 Voice attractiveness. In *The Oxford handbook of voice perception* (eds S Frühholz, P Belin), pp. 606–626. New York, NY: Oxford University Press, 606–626.
- Aung T, Goetz S, Adams J, McKenna C, Hess C, Roytman S, Cheng JT, Zilioli S, Puts D. 2021 Low fundamental and formant frequencies predict fighting ability among male mixed martial arts fighters. *Scient. Rep.* **11**, 1. (doi:10.1038/s41598-020-79139-8)
- 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)
- Klofstad CA. 2016 Candidate voice pitch influences election outcomes. *Polit. Psychol.* **37**, 725–738. (doi:10.1111/pops.12280)
- Puts DA, Doll LM, Hill AK. 2014 Sexual selection on human voices. In *Evolutionary perspectives on human sexual psychology and behavior* (eds V Weekes-Shackelford, T Shackelford), pp. 69–86. New York, NY: Springer. (doi:10.1007/978-1-4939-0314-6_3)
- Kreiman J, Park SJ, Keating PA, Alwan A. 2015 The relationship between acoustic and perceived intraspeaker variability in voice quality. *Interspeech* **2015**, 2357–2360.
- 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)
- Ponsot E, Burred JJ, Belin P, Aucouturier JJ. 2018 Cracking the social code of speech prosody using reverse correlation. *Proc. Natl Acad. Sci. USA* **115**, 3972–3977. (doi:10.1073/pnas.1716090115)
- Laukka P, Effenbein HA. 2021 Cross-cultural emotion recognition and in-group advantage in vocal expression: a meta-analysis. *Emot. Rev.* **13**, 3–11. (doi:10.1177/1754073919897295)
- Pisanski K, Reby D. 2021 Efficacy in deceptive vocal exaggeration of human body size. *Nat. Commun.* **12**, 968. (doi:10.1038/s41467-021-21008-7)
- Titze IR. 1989 Physiologic and acoustic differences between male and female voices. *J. Acoust. Soc. Am.* **85**, 1699–1707. (doi:10.1121/1.397959)
- Pisanski K, Rendall D. 2011 The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *J. Acoust. Soc. Am.* **129**, 2201. (doi:10.1121/1.3552866)
- Levero 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)
20. 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)
 21. Collins SA. 2000 Men's voices and women's choices. *Anim. Behav.* **60**, 773–780. (doi:10.1006/anbe.2000.1523)
 22. Bruckert L, Liénard JS, Lacroix A, Kreutzer M, Leboucher G. 2006 Women use voice parameters to assess men's characteristics. *Proc. R. Soc. B* **273**, 83–89. (doi:10.1098/rspb.2005.3265)
 23. McAleer P, Todorov A, Belin P. 2014 How do you say 'Hello'? Personality impressions from brief novel voices. *PLoS ONE* **9**, e90779. (doi:10.1371/journal.pone.0090779)
 24. Sell A, Bryant GA, Cosmides L, Tooby J, Sznycer D, von Rueden C, Krauss A, Gurven M. 2010 Adaptations in humans for assessing physical strength from the voice. *Proc. R. Soc. B* **277**, 3509–3518. (doi:10.1098/rspb.2010.0769)
 25. Puts D, Apicella CL, Cardenas RA. 2012 Masculine voices signal men's threat potential in forager and industrial societies. *Proc. R. Soc. B* **279**, 601–609. (doi:10.1098/rspb.2011.0829)
 26. Pisanski K, Oleszkiewicz A, Plachetka J, Gmiterek M, Reby D. 2018 Voice pitch modulation in human mate choice. *Proc. R. Soc. B* **285**, 20181634. (doi:10.1098/rspb.2018.1634)
 27. Puts D, Gaulin SJC, Verdolini K. 2006 Dominance and the evolution of sexual dimorphism in human voice pitch. *Evol. Hum. Behav.* **27**, 283–296. (doi:10.1016/j.evolhumbehav.2005.11.003)
 28. Mahrholz G, Belin P, McAleer P. 2018 Judgements of a speaker's personality are correlated across differing content and stimulus type. *PLoS ONE* **13**, e0204991. (doi:10.1371/journal.pone.0204991)
 29. 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. (doi:10.1098/rsos.191642)
 30. Boersma P, Weenink D. 2017 *Praat: doing phonetics by computer*. See <http://www.praat.org/>.
 31. Raine J, Pisanski K, Simmer J, Reby D. 2019 Vocal communication of simulated pain. *Bioacoustics* **28**, 404–426. (doi:10.1080/09524622.2018.1463295)
 32. Cartei V, Bond R, Reby D. 2014 What makes a voice masculine: physiological and acoustical correlates of women's ratings of men's vocal masculinity. *Horm. Behav.* **66**, 569–576. (doi:10.1016/j.yhbeh.2014.08.006)
 33. Pisanski K *et al.* 2014 Vocal indicators of body size in men and women: a meta-analysis. *Anim. Behav.* **95**, 89–99. (doi:10.1016/j.anbehav.2014.06.011)
 34. Abitbol J, Abitbol P, Abitbol B. 1999 Sex hormones and the female voice. *J. Voice* **13**, 424–446. (doi:10.1016/S0892-1997(99)80048-4)
 35. Kreiman J, Sidtis D. 2011 Physical characteristics and the voice: can we hear what a speaker looks like? In *Foundations of voice studies: an interdisciplinary approach to voice production and perception*, pp. 110–155. Chichester, UK: John Wiley & Sons.
 36. Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Methodol.* **57**, 289–300. (doi:10.1111/j.2517-6161.1995.tb02031.x)
 37. 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)
 38. 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)
 39. Engelberg JW, Schwartz JW, Gouzoules H. 2019 Do human screams permit individual recognition? *PeerJ* **7**, e7087. (doi:10.7717/peerj.7087)
 40. 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)
 41. 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)
 42. Kreiman J, Sidtis D. 2011 Recognizing speaker identity from voice: theoretical and ethological perspectives and a psychological model. In *Foundations of voice studies: an interdisciplinary approach to voice production and perception*, pp. 156–188. Chichester, UK: John Wiley & Sons.