



**A University of Sussex PhD thesis**

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

**The Nature and Function of Human Nonverbal Vocalisations**

**Dissertation submitted to the University of Sussex for the degree of Doctor of  
Philosophy**

**Jordan Raine**

**January 2018**

**UNIVERSITY OF SUSSEX****Jordan Raine****Doctor of Philosophy****The nature and function of human nonverbal vocalisations****Summary**

Though human nonverbal vocalisations are widespread, scientific consideration of their mechanisms and communicative functions has been largely overlooked. This is despite their close alignment with the vocal communicative systems of primates and other mammals, whose primary function is to signal indexical information relevant to sexual and natural selection processes. In this thesis, I examine human nonverbal vocalisations from an evolutionary perspective, with the central hypothesis that they are functionally and structurally homologous to nonhuman mammal calls, communicating evolutionarily relevant indexical information that is perceived and utilised by listeners. In Chapter 1, I introduce the methodological framework (source-filter theory) necessary to understand the production of vocal signals in mammals, before summarising the information contained within the acoustic structure of nonhuman mammals and human speech, and the effects these cues have on both vocaliser and listener. I then examine the current evidence for functional and structural homology between human and nonhuman nonverbal vocalisations. In Chapters 2 to 5, I quantitatively analyse the acoustic structure of a number of nonverbal vocalisations, and perform playback experiments to examine their functional effects on listeners. In Chapters 2 and 3, I investigate whether aggressive roars and distress screams communicate acoustic cues to absolute and relative strength and height. In Chapter 4, I analyse the acoustic structure of pain cries of varying intensity, and conduct playback experiments to explore the acoustic and perceptual correlates of pain. In Chapter 5, I examine whether the fundamental frequency of tennis grunts produced during professional tennis matches is dependent on the sex and body posture of the vocaliser, as well as the progress and outcome of the contest, and whether listeners can infer these cues. In Chapter 6, I tie these findings together, arguing that the acoustic structure of human nonverbal vocalisations, in continuity with nonhuman mammal vocalisations, has been selected to support the functional communication of indexical and motivational information.

## Declaration

This thesis conforms to an ‘article format’ in which the middle chapters consist of discrete articles written in a style that is appropriate for publication in peer-reviewed journals in the field. The first and final chapters present synthetic overviews and discussions of the field and the research undertaken.

Chapter 2 is about to be submitted in *Journal of the Acoustical Society of America* as:

Raine, J., Pisanski, K., Bond, R., Feinberg, D. R., Simner, J., & Reby, D. (2018).

Human roars communicate upper-body strength, *Journal of the Acoustical Society of America*, submitted.

The author contributions are as follows: I was responsible for all aspects of the study design, stimulus and data collection, data analysis, and the writing of the manuscript. Prof. Reby was responsible for the initial conception of the research. Prof. Reby and Dr. Pisanski also contributed to the study design, data analysis, and provided feedback on the manuscript. Prof. Feinberg assisted with data collection, and Dr. Bond assisted with data analysis. Prof. Simner provided access to a dedicated online platform to run playback experiments.

Chapter 3 has been submitted as a short report to *Biology Letters* as:

Raine, J., Pisanski, K., Oleszkiewicz, A., Simner, J., & Reby, D. (2018). Human listeners accurately judge relative formidability from aggressive roars and speech, *Biology Letters*, submitted.

The author contributions are as follows: I was responsible for all aspects of the study design, stimulus and data collection, data analysis and the writing of the manuscript; Prof. Reby, Dr. Pisanski, and I were responsible for the initial conception of the study. Prof. Reby and Dr. Pisanski contributed to the study design, data analysis, and provided feedback on the manuscript. Dr. Oleszkiewicz assisted with data collection. Prof. Simner provided access to a dedicated online platform to run playback experiments.

Chapter 4 is in the process of being submitted to *Bioacoustics* as:

Raine, J., Pisanski, K., Simner, J., & Reby, D. (2018). Vocal communication of simulated pain, *Bioacoustics*, submitted.

The author contributions are as follows: I was responsible for all aspects of the study design, stimulus and data collection, data analysis, and the writing of the manuscript. Prof. Reby was responsible for the initial conception of the research. Prof. Reby and Dr. Pisanski also contributed to the study design, data analysis, and provided feedback on the manuscript. Prof. Simner provided access to a dedicated online platform to run playback experiments.

Chapter 5 is published in *Animal Behaviour* as:

Raine, J., Pisanski, K., & Reby, D. (2017). Tennis grunts communicate acoustic cues to sex and contest outcome, *Animal Behaviour*, 130, 47-55.

The author contributions are as follows: I was responsible for all aspects of the study design, stimulus and data collection, data analysis and the writing of the manuscript; Prof. Reby and I share responsibility for the initial conception of the study. Prof. Reby

and Dr. Pisanski contributed to the study design, data analysis, and provided feedback on the manuscript.

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:     Jordan Raine

## Acknowledgements

Where do I start... It's been a long old journey. David, I don't think it's possible to adequately put onto paper how lucky I feel to have had you to guide me along the way. A massive thank you for the wisdom you have imparted over the years, but more importantly for the little things that have made working with you such a positive experience. When I cross the final hurdle, I'll have lost a supervisor, but most definitely gained a mate (in strictly non-evolutionary terms!!). Kasia, I'm forever indebted to you also for the expertise and support you've been so kind to share since you arrived. I don't know where I'd be without your ever-present smile and proverbial shoulder to lean on. To my other office mates, in particular Val and Vicky, you've been big parts of the lab family too – thank you for helping me through the lows and sharing the highs.

Zoltan, your end-of-year review advice has been a great help. David F, thanks for having me at Mac and making me feel so welcome there, I absolutely loved it. Rod, your statistical skill was hugely appreciated, as was sharing some great moments on the football pitch... if only I'd managed to save that penalty. Jools and James, SynToolkit was an absolute godsend, thanks for letting me be your guinea pig. Big thanks also to Jane Boston, Sara Jane Bailes, and Jason Price for connecting me with an abundance of budding actors and actresses – you certainly have some good screamers in your midst!

To my parents, I'm eternally grateful for the support and security you've given me not only throughout my PhD, but for the 21 years before that too. You've always been there when I've needed you, and I love you both. And finally, to my live-in family over the years - Tom, Hebba, Babu, Paul, Gui, David, Mazz, and my PhD partner-in-crime Claire. God only knows how you endured the weird noises coming out of my bedroom (and mouth), and kept me on the right side of sanity (mostly). I owe you one.

## Table of Contents

List of Tables .....	xiv
List of Figures .....	xvii
List of Acoustic Parameters and Perceptual Correlates .....	xix
<b>Chapter 1.</b>	
<b>Introduction</b> .....	1
Overview.....	1
The source-filter model of vocal production .....	2
Acoustic measures characterising the mammal voice .....	8
Nonverbal Acoustic Communication.....	12
Functional indexical cues in nonhuman mammal vocalisations.....	14
F0 as a static cue to age, sex, size, and hormone levels .....	15
F0 as a dynamic cue to emotional or motivational state .....	17
F0 as a functional cue to male quality .....	20
Formants as a functional static cue to body size .....	22
Anatomical adaptations for formant exaggeration .....	24
Behavioural adaptations for functional formant modulation .....	26
Other vocal characteristics .....	27
Functional indexical cues in human speech .....	28
F0 and formants: static cues to age and sex .....	30
F0 as a cue to hormonal quality and fitness .....	31
Formants, but not F0, communicate body size .....	34
Does speech contain acoustic cues to strength? .....	39
F0 as a dynamic indicator of arousal, stress, emotion, and motivation ..	41

Functional indexical cues in human nonverbal vocalisations .....	43
Laughter .....	45
Infant distress cries .....	48
Arousal in nonverbal vocalisations .....	50
Other vocalisations: screams, roars and grunts .....	52
Outstanding questions .....	55
Are human nonverbal vocalisations similar in form and function to nonhuman mammal calls? .....	55
Are nonverbal vocalisations more effective carriers of certain indexical cues than speech? .....	56
How do volitionally produced vocalisations influence listeners? .....	57
The Present Thesis .....	59
<b>Chapter 2.</b>	
<b>Human roars communicate upper-body strength .....</b>	<b>62</b>
Chapter summary .....	62
Abstract .....	65
Introduction .....	66
The present study .....	68
Experiment 1: Do aggressive roars and distress screams follow motivational- structural rules, and encode strength and height? .....	69
Method .....	69
Participants .....	69
Procedure .....	70
Voice recording .....	70
Strength assessment .....	71

Acoustic analysis .....	72
Statistical analysis .....	76
Results & Discussion .....	77
Do roars, screams, and agonistic speech stimuli differ in acoustic structure? .....	77
Do roars, screams and agonistic speech stimuli contain acoustic cues to strength and height? .....	83
Experiments 2 and 3: Can listeners estimate strength and height from the human voice? .....	86
Method .....	86
Participants .....	86
Stimuli .....	87
Procedure .....	87
Statistical analysis .....	88
Results & Discussion .....	89
Are there sex differences in actual or rated strength and height? .....	89
Effects of vocaliser sex .....	89
Effects of listener sex .....	90
Did stimulus context (aggression/distress) and stimulus type affect ratings of strength and height? .....	93
Strength attributions .....	93
Height attributions .....	96
Could listeners detect strength and height from the voice? .....	97
Strength estimation .....	97
Height estimation .....	98

Were ratings of physical characteristics related to acoustic characteristics? .....	105
General Discussion .....	106
Conclusion .....	109
<b>Chapter 3.</b>	
<b>Human listeners can accurately judge relative formidability from aggressive roars and speech .....</b>	<b>111</b>
Chapter summary .....	111
Summary .....	113
Experimental Procedures .....	114
Participants .....	114
Materials .....	115
Vocal stimuli .....	115
Physical formidability measures .....	115
Procedure .....	116
Coding and statistical analysis .....	118
Results & Discussion .....	119
Judgments of relative strength .....	119
Judgments of relative height .....	122
<b>Chapter 4.</b>	
<b>Vocal communication of simulated pain .....</b>	<b>129</b>
Chapter summary .....	129
Abstract .....	131
Introduction .....	132
Method .....	136

1. Acoustic Analysis .....	136
Participants .....	136
Procedure .....	136
Voice recording .....	136
Acoustic analysis .....	137
Principal component analysis .....	139
Statistical analysis .....	140
2. Playback Experiment .....	140
Participants .....	140
Procedure .....	143
Statistical analysis .....	144
Results .....	144
Does the acoustic structure of pain vocalisations differ with pain intensity? .....	144
Principal component analysis .....	144
Discriminant function analysis .....	145
Do pain intensity level and acoustic characteristics affect ratings of pain? .....	148
Discussion .....	154
Acoustic cues to levels of pain intensity .....	154
Effect of acoustic cues on listeners' assessments of pain intensity .....	157
Are simulated vocalisations functional? .....	159
<b>Chapter 5.</b>	
<b>Tennis grunts communicate acoustic cues to sex and contest outcome .....</b>	<b>161</b>
Chapter summary .....	161

Abstract .....	163
Introduction .....	164
Method .....	167
1. Analysis of Tennis Grunts .....	167
Within-subject variables .....	167
Between-subject variables .....	168
Acoustic analysis .....	169
Statistical analysis .....	169
2. Playback Experiments .....	170
Participants .....	170
Stimuli and procedure .....	170
Statistical analysis .....	172
Ethical note .....	172
Results .....	172
Do tennis grunts contain static and dynamic cues? .....	172
Effects of sex, age, height, weight, and speaking F0 on grunt F0 .....	172
Effects of shot type, match progress, and match outcome on grunt F0 .....	173
Can listeners gauge sex and match outcome from tennis grunts? .....	176
Attribution of sex by listeners .....	176
Attribution of match outcome by listeners .....	177
Reanalysis using ERBs .....	177
Discussion .....	179
Cues to sex, but not height, weight, or age in tennis grunts .....	179

Acoustic cues to shot type .....	181
Acoustic cues to tennis match progress and outcome .....	182
Conclusion .....	184
Data availability .....	185
<b>Chapter 6. General Discussion</b> .....	186
Human and nonhuman vocalisations: structural and functional homology.....	186
Volitional production of vocalisations: the origins of speech capabilities? ....	189
References .....	193
Appendices .....	249
Appendix 1. Supplementary tables for Chapter 2 .....	249

## List of Tables

### **Chapter 2. Human roars communicate upper-body strength**

<i>Table 1.</i> Mean acoustic characteristics of male vocal stimuli .....	79
<i>Table 2.</i> Mean acoustic characteristics of female vocal stimuli .....	80
<i>Table 3.</i> Linear mixed model testing the effects of vocaliser sex, listener sex, stimulus context, and stimulus type on rated strength.....	91
<i>Table 4.</i> Linear mixed model testing the effects of vocaliser sex, listener sex, stimulus context, and stimulus type on rated height .....	92
<i>Table 5.</i> Linear mixed models testing the effects of strength, stimulus context, stimulus type, and listener sex on rated strength .....	99
<i>Table 6.</i> Standardised linear mixed model coefficients representing the sensitivity of listeners to variation in vocaliser strength and height .....	100
<i>Table 7.</i> Linear mixed models testing the effects of height, stimulus context, stimulus type, and listener sex on rated height .....	101

### **Chapter 3. Human listeners can accurately judge relative formidability from aggressive roars and speech**

<i>Table 1.</i> Mixed multinomial logistic regression testing the effects of the categorised actual difference in strength between listener and vocaliser, vocaliser sex, listener sex, and stimulus type on whether the vocaliser was rated to be stronger, weaker, or of similar strength .....	121
<i>Table 2.</i> Mixed multinomial logistic regression testing the effects of the categorised actual difference in height between listener and vocaliser, vocaliser sex, listener sex,	

and stimulus type on whether the vocaliser was rated to be taller, shorter, or of similar height .....	125
--	-----

#### **Chapter 4. Vocal communication of simulated pain**

<i>Table 1.</i> Rotated factor loadings for each principal component calculated from the acoustic variables characterising simulated pain vocalisations .....	142
<i>Table 2.</i> Principal component loadings on the discriminant functions .....	149
<i>Table 3.</i> Mean acoustic characteristics of female vocal stimuli .....	149
<i>Table 4.</i> Mean acoustic characteristics of male vocal stimuli .....	150
<i>Table 5.</i> Linear mixed model testing the effects of the intensity of pain simulated by the vocaliser and the four acoustic principal components on listeners' attributions of the level of pain conveyed by the vocaliser .....	151

#### **Chapter 5. Tennis grunts communicate acoustic cues to sex and contest outcome**

<i>Table 1.</i> Correlates of grunt F0: LMMs testing the effects of vocaliser traits and match factors on mean grunt F0 .....	175
---	-----

#### **Chapter 6. Appendix**

<i>Table A1.</i> Acoustic characteristics' loadings on the discriminant functions .....	249
<i>Table A2.</i> Acoustic characteristics' loadings on the discriminant functions, for male vocalisers only .....	250
<i>Table A3.</i> Acoustic characteristics' loadings on the discriminant functions, for female vocalisers only .....	251
<i>Table A4.</i> Standardised regression coefficients for acoustic predictors of men and women's physical strength .....	252

<i>Table A5.</i> Standardised regression coefficients for acoustic predictors of men and women's height .....	253
<i>Table A6.</i> Standardised regression coefficients for acoustic predictors of listeners' ratings of men and women's physical strength .....	254
<i>Table A7.</i> Standardised regression coefficients for acoustic predictors of listeners' ratings of men and women's physical height .....	255

## List of Figures

### Chapter 1. Introduction

*Figure 1. Top:* a diagram of the human vocal apparatus. *Bottom:* Spectrograms of the vowels /ɑ/ /i/ /ε/ /o/ and /u/ (international phonetic symbols) spoken by an adult male (top row) and an adult female (bottom row) ..... 4

### Chapter 2. Human roars communicate upper-body strength

*Figure 1.* Spectrograms illustrating the acoustic structure of a typical (a) male roar, (b) male scream, (c) female roar, and (d) female scream ..... 75

*Figure 2.* Discriminant function analysis illustrating acoustic separation of stimulus conditions, (a) for all vocalisers, (b) for male vocalisers only, and (c) for female vocalisers only ..... 78

*Figure 3.* Attributed strength as a function of actual strength, when listeners rated (a) male speech stimuli, (b) male vocalisations, (c) female speech stimuli, and (d) female vocalisations ..... 94

*Figure 4.* Attributed height as a function of actual height, when listeners rated (a) male speech stimuli, (b) male vocalisations, (c) female speech stimuli, and (d) female vocalisations ..... 94

### Chapter 3. Human listeners can accurately judge relative formidability from aggressive roars and speech

<i>Figure 1.</i> Percentage of listeners judging vocalisers as relatively weaker, of similar strength, or relatively stronger than them, as a function of the actual difference in strength between listener and vocaliser .....	122
--	-----

<i>Figure 2.</i> Percentage of listeners judging vocalisers as relatively shorter, of similar height, or relatively taller than them, as a function of the actual difference in height between listener and vocaliser .....	124
---	-----

#### **Chapter 4. Vocal communication of simulated pain**

<i>Figure 1.</i> Principal component analysis illustrating the acoustic variability of pain vocalisations across pain intensity levels.....	141
---	-----

<i>Figure 2.</i> Discriminant function analysis illustrating acoustic separation of pain vocalisations at different levels of pain intensity.....	147
---	-----

<i>Figure 3.</i> Pain ratings as a function of variation in the four principal components ....	152
--	-----

<i>Figure 4.</i> Pain ratings as a function of the average of the first three principal components .....	153
--	-----

#### **Chapter 5. Tennis grunts communicate acoustic cues to sex and contest outcome**

<i>Figure 1.</i> Frequency distribution of mean F0 for grunts accompanying serves .....	174
---	-----

<i>Figure 2.</i> Attribution of match outcome by adult listeners as a function of natural and manipulated F0 differences .....	178
--	-----

## List of Acoustic Parameters and Perceptual Correlates<sup>1</sup>

$\Delta F$ (Hz)	A reliable measure of the average dispersion of formants, related to vocal tract length (p7)
Amplitude modulation (%)	The co-occurrence of a low-frequency source of vibration with a high-frequency source, which interact to produce a signal with audible periodic variation in overall intensity
Amplitude mean (dB)	The mean level of energy in the acoustic signal (p10)
Biphonation (%)	The presence of two independent sources of vibration
Centre of gravity (Hz)	The frequency at which there is an equal amount of spectral energy above and below (p11)
Deterministic chaos (dB)	Irregular/aperiodic vocal fold vibration, occurring at the upper limits of subglottal pressure
Dominant frequency (Hz)	The frequency in the spectrum with the highest amplitude across a given time frame
Dominant frequency F4 (Hz)	Within the frequency range in which the fourth formant is expected to fall, the frequency with the highest amplitude across the duration of the signal
Duration (s)	The duration of the signal
F0 (Hz)	Fundamental frequency
F0 CV (Hz)	Variability in the rate of vocal fold vibration
F0 max (Hz)	The maximum rate of vibration of the vocal folds
F0 mean (Hz)	The average rate of vibration of the vocal folds (p3)

---

<sup>1</sup> This list is not exhaustive. It contains only those acoustic parameters mentioned or used multiple times throughout the thesis and/or those that may be unfamiliar to some readers. Page numbers are given at the end of definitions for which more information is available.

F0 min (Hz)	The minimum rate of vibration of the vocal folds
F0 modulation, major	The number of inflections in an F0 contour smoothed to suppress short-term F0 modulation, but preserve major intonation events, divided by the total duration of the voiced proportion of the signal (p72)
F0 modulation, minor	The number of inflections in an F0 contour smoothed to suppress period-to-period frequency fluctuations (jitter), but preserve short-term modulation, divided by the total duration of the voiced proportion of the signal (p72)
F0 range (Hz)	The difference between the smallest and largest values of F0 across the signal
F0 start-end (Hz)	A measure of the overall F0 contour. The F0 at the start of the signal subtracted by the F0 at the end of the signal
F1 (Hz)	First formant
F2 (Hz)	Second formant
F3 (Hz)	Third formant
F4 (Hz)	Fourth formant
Formant (Hz)	Resonance frequency of the vocal tract (p5)
Harmonics-to-noise ratio (dB)	The ratio of harmonic spectral energy to chaotic spectral energy (HNR) (p11)
Inflection	A change in the sign of the slope of F0
Intensity CV (dB)	Variability in the level of energy in the acoustic signal
Intensity range (dB)	The difference between the smallest and largest amplitude values across the signal

Jitter (Hz)	Period-to-period variability in the rate of vocal fold vibration
Loudness	Perceptual correlate of amplitude
Nonlinear phenomena (%)	The proportion of the signal for which either amplitude modulation, deterministic chaos, or subharmonics are present
Pitch	Perceptual correlate of F0 (p5)
Roughness	Perceptual correlate of non-periodic vocal fold vibration, affected by HNR, jitter, and NLP (p10)
Shimmer (dB)	Period-to-period variability in the level of energy in the acoustic signal
Subharmonics (%)	The presence of vocal fold vibration at a frequency equal to an integer fraction of the F0, an addition to the F0
Timbre	Perceptual correlate of formant frequencies, and their spacing (p7)

## Chapter 1: Introduction

### Overview

Though speech enables humans to communicate complex semantic information, nonverbal vocalisations remain in our vocal repertoire throughout life. Yet despite their evident presence in a wide variety of human interactions (e.g. sportspeople's grunts: Callison, Berg, & Slivka, 2014; Welch & Tschampl, 2012; crowd cheers: Myers, Nevill, & Al-Nakeeb, 2012; Nevill, Balmer, & Mark Williams, 2002; laughter: Scott, Lavan, Chen, & McGettigan, 2014; infant distress cries: Koutseff et al., 2017; infant play vocalisations: Fry, 1987; P. K. Smith & Lewis, 1985; coital vocalisations: Levin, 2006), scientific consideration of the mechanisms and communicative functions of human nonverbal vocalisations has been largely overlooked, with some even claiming that vocalisations such as grunting in sports are not primarily communicative (e.g. Trouvain & Truong, 2012). This is despite the close alignment between human vocalisations and those of other primates (Burling, 1993) and mammals, whose primary function is to signal indexical information relevant to sexual and natural selection processes. In this thesis, I examine the acoustic structure and functions of human nonverbal vocalisations from an evolutionary perspective, with the central hypothesis that they are functionally and structurally homologous to nonhuman mammal calls, communicating evolutionarily relevant indexical and motivational information that is perceived and utilised by listeners.

The field of mammal vocal communication has greatly benefited from the application of the source-filter theory of speech production to non-human mammals (Taylor & Reby, 2010). The source-filter model offers a framework for quantitative acoustic analysis and voice resynthesis, enabling researchers to examine how acoustic

variation is linked to anatomical and physiological attributes of the caller, and how it functionally influences perception (Taylor, Charlton, & Reby, 2016). Crucially, as the mechanisms by which vocal signals are produced are highly conserved across mammals, the source-filter model allows for comparative interpretation of human and nonhuman vocalisations from an evolutionary perspective. It is within this framework that the present thesis will investigate its general and specific hypotheses.

In Chapter 1, I introduce the source-filter framework and summarise the information contained within the acoustic structure of nonhuman mammal vocalisations, focusing on the functional effects of these acoustic cues on listeners. I then demonstrate that many of these functional cues are similarly present in human speech, and examine the current evidence for structural and functional homology between human and nonhuman nonverbal vocalisations. I conclude by raising outstanding questions that address whether human nonverbal vocalisations, like their nonhuman equivalents, primarily function to communicate evolutionarily relevant information. In Chapters 2, 3, 4 and 5, I empirically test this hypothesis by investigating the acoustic structure and function of tennis grunts, aggressive roars, distress screams, and pain vocalisations. Finally, I discuss the implications, limitations, and future directions of this work in Chapter 6.

### **The source-filter model of vocal production**

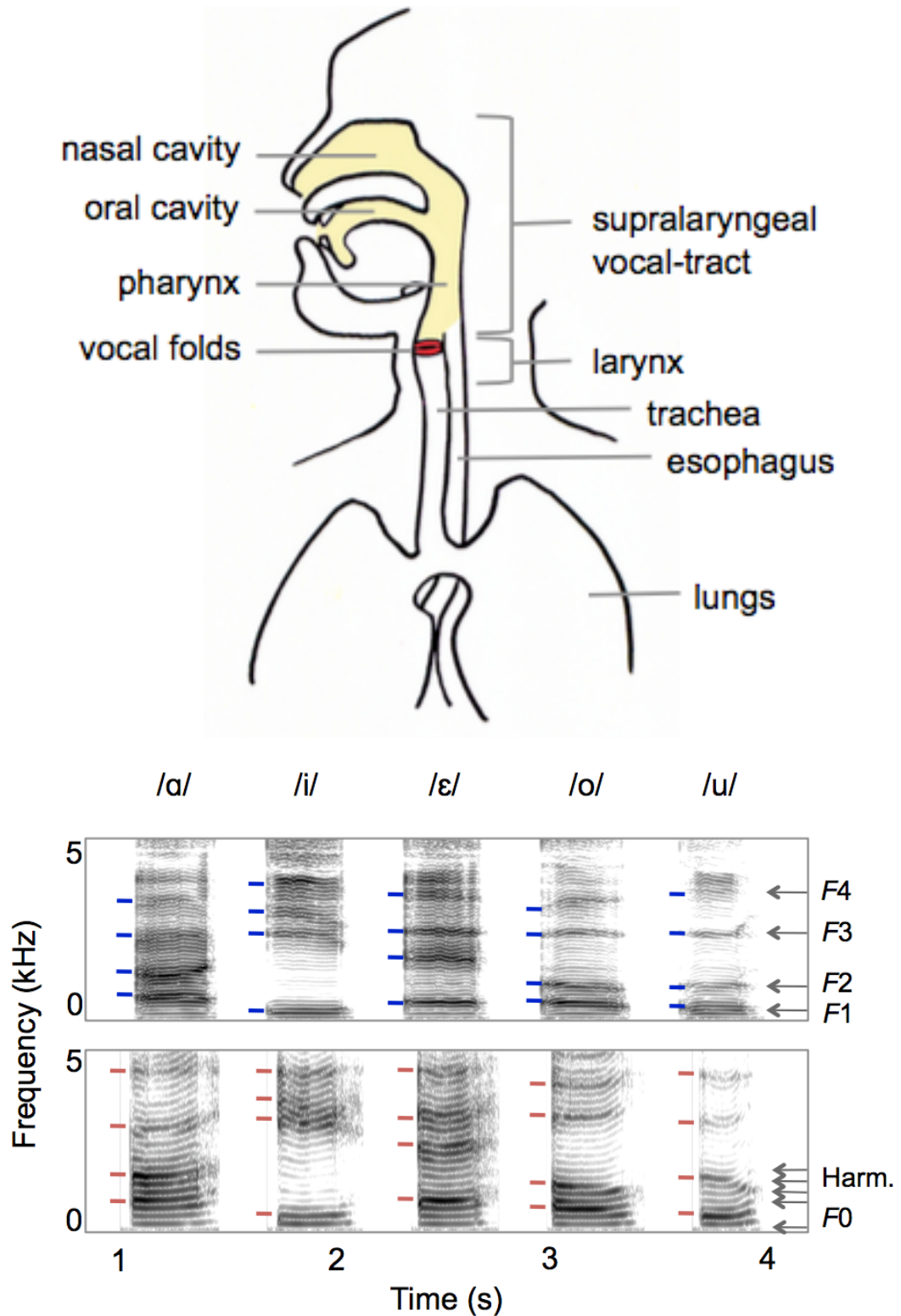
The “source-filter” framework (Chiba & Kajiyama, 1958; Fant, 1960) states that vocal production follows a two-stage process: a signal is generated by vibration of the vocal folds or turbulent airflow (source), creating a complex sound wave whose spectrum contains energy at fundamental frequency (F0) and harmonic overtones at

integer multiples of  $F_0$  if the signal is periodic, or at a wide range of random frequencies if the signal is chaotic. As this glottal wave propagates through the vocal tract (filter), the energy of selected frequencies (corresponding to the resonances of the vocal tract) is amplified, producing spectral peaks called formants ( $F_i$ ), before the acoustic signal radiates out to be perceived by listeners. This process determines the acoustic properties of mammalian vocalisations.

At the level of the source, continuous energy provided by expulsion of air from the lungs causes quasi-periodic oscillation of the vocal folds as the glottis is closed in the larynx (see Figure 1). The rate of vibration of the vocal folds determines the  $F_0$  of the voice (see Figure 1). This oscillation can be predicted by the following formula, where  $L$  is the length,  $\sigma$  is the longitudinal stress (determined by laryngeal tension and subglottal pressure), and  $\rho$  is the tissue density of the vocal folds ( $1.02 \text{ g/cm}^3$  in humans) (Titze, 1994):

$$F_0 = \frac{1}{2L} \sqrt{\frac{\sigma}{\rho}}$$

All else equal, longer, denser, and looser vocal folds vibrate at a slower rate, resulting in a lower fundamental frequency and more closely spaced harmonics. As air expulsion from the lungs forces the vocal folds into cyclical abduction and adduction, the glottis (the space between the vocal folds) alternates between open and closed states, causing air to emanate in bursts. The fundamental frequency is therefore also sometimes referred to as the glottal-pulse rate. Harmonic overtones occur at integer multiples of  $F_0$ , successively decreasing in energy.



*Figure 1. Top: a diagram of the human vocal apparatus. Bottom: Spectrograms of the vowels /a/ /i/ /ε/ /o/ and /u/ (international phonetic symbols) spoken by an adult male (top row) and an adult female (bottom row). The first to fourth formants (F1-F4), fundamental frequency (F0) and harmonics are labeled for the vowel /u/. Formant positions are labeled for each individual vowel with bars positioned to the left of each spectrogram. Note the variation in formant spacing between the male and female voices and among vowels. Reproduced with permission from Pisanski (2014).*

The perceptual correlate of fundamental frequency is the pitch of the voice (Kreiman & Sidtis, 2011; Titze, 1994). “Fundamental frequency” is generally used to describe acoustic measurement of the voice parameter, whereas pitch refers to the perception of the parameter. Though their relationship is not linear, pitch and fundamental frequency scale proportionally in the human vocal frequency range and can thus be considered equivalent. However, the relationship between F0 and pitch perception varies among listeners (Krishnan, Xu, Gandour, & Cariani, 2005), and also depends on voice resonances (Melara & Marks, 1990; Pitt, 1994) and amplitude (Arnoldner, Kaider, & Hamzavi, 2006; Stevens, 1935). In the case of whispered speech, which is characterised by turbulent, chaotic airflow through an open glottis rather than periodic vocal fold vibration, voice resonances can result in the perception of pitch in the absence of F0 (Higashikawa, Nakai, Sakakura, & Takahashi, 1996; Thomas, 1969).

The source sound produced by vibration of the vocal folds then travels up the vocal tract, which is comprised of the pharynx and oral and nasal cavities (see Figure 1), where the glottal waveform is filtered. The energy of harmonics that coincide with the resonant frequencies of the vocal tract is amplified, while the energy of harmonics between resonances is dampened, thus shaping the spectral characteristics of the glottal wave. The amplified frequency bands, corresponding to vocal tract resonances, are commonly referred to as formants (Fant, 1960; Titze, 1994), and are determined by the length, size, and shape of the vocal tract, with 4 or 5 prominent amplitude peaks in the 0 to 5 kHz frequency range in humans.

Manipulations of articulators such as the tongue, lips, and soft palate, alter the relative location of formants, giving rise to the different vowel and consonant sounds used in human speech (Fitch & Hauser, 2003; P. Lieberman & Blumstein, 1988), as well as the transitions between these sounds. For example, each of the vowel sounds /a/

/i/ /ε/ /o/ and /u/ (International Phonetic Alphabet) can be sufficiently differentiated by values of the first (F1) and second (F2) formant (See Figure 1), while the formant transition between a bilabial plosive and a high front vowel is distinct from that between a lateral fricative and a high front vowel. Although the capacity for such advanced and precise articulatory control is unique to humans, and has been argued as a critical prerequisite in the emergence of articulated speech, recent evidence suggests that some mammals possess greater control over articulatory movements than previously assumed (e.g. Lameira et al., 2015; Perlman & Clark, 2015; Schel, Townsend, Machanda, Zuberbühler, & Slocombe, 2013; see Pisanski, Cartei, McGettigan, Raine, & Reby, 2016 for a review).

Whereas the shape of the vocal tract influences the relative positions of the first and second formants, the overall spacing of formants is influenced by its length, with longer vocal tracts producing lower and more closely spaced formants. Thus, above and beyond the key role of formants in vowel production, absolute formant scaling is also an indicator of vocal tract length, and therefore body size. Modelling the vocal tract as a quarter-wave length resonator with uniform cross-sectional area, closed at the glottis and open at the lips, the relationship between successive formants (F1, F2, ... Fi) and vocal tract length can be calculated by the following formula, where  $c$  is the speed of sound in a mammal vocal tract (approximately 350 m/s in the warm, humid air of a mammalian vocal tract), and  $VTL$  is the length of the vocal tract:

$$F_i = \frac{(2i - 1)c}{4VTL}$$

Modelling the vocal tract as a quarter wave resonator is more accurate for the unconstricted schwa sound (the most common vowel sound, e.g. ‘a’ in ‘about’, or ‘e’ in ‘taken’, or ‘u’ in ‘supply’) than for other vowels produced with oral tract constrictions, which affect the relative position of primarily the first two formants (Fitch & Hauser, 2003). However, measures of the average dispersion of formants are much less sensitive to deviations from the assumed model, especially when averaging dispersion across a vocal signal of long duration (Titze, 1994). These measures also do not necessitate assumptions that the vocal tract is closed at one end, as changing conditions at glottal and lip boundaries shift the absolute frequencies of formants, but not their overall spacing (Riede & Fitch, 1999). One particularly reliable measure (Pisanski, Fraccaro, Tigue, O’Connor, Röder, et al., 2014) of formant dispersion is formant spacing ( $\Delta F$ ), which is related to vocal tract length (VTL) by the following formula (Reby & McComb, 2003):

$$\Delta F = \frac{c}{2VTL}$$

The spacing and overall frequency of these vocal tract resonances is the primary (but not the only, Houtsma, 1997) determinant of the perceptual attribute of voice “timbre”, which refers to the quality that differentiates two voices of equal pitch, loudness, and duration (Yost, 2000). The subjective nature of timbre has led to variation across disciplines in descriptive nomenclature. Literature investigating voice variation from an evolutionary perspective tends to characterise voices with lower and more closely spaced resonances as having a “deeper”, “darker”, “richer”, or more masculine

timbre (Baumann & Belin, 2010; Cartei, Bond, & Reby, 2014; Fitch & Giedd, 1999; Hollien, Green, & Massey, 1994).

According to source-filter theory, F0 and formants are assumed to be largely independent of one another (Fant, 1960; Fitch & Hauser, 1995). This means that information can be communicated via either the source (vocal folds) or the filter (vocal tract), or a combination of both, and that static inter-individual variation in anatomy, as well as dynamic intra-individual changes in physiology, can affect source- and filter-related acoustic parameters independently. Nevertheless, the perceptibility of vocal timbre can be affected by the characteristics of the source. A higher density of source energy (achieved by lowering F0, which results in more closely spaced harmonics) highlights resonances of the vocal tract, and therefore increases the perceptual salience of formants (Kewley-Port, Li, Zheng, & Neel, 1996; Ryalls & Lieberman, 1982), giving the sound of a voice a richer or fuller timbre. There is also some degree of complex interaction between the source and the filter at the level of production, whereby under certain conditions the lower vocal tract can shape the flow of air through the larynx and influence the vibratory regime of the vocal folds (Titze, 2008; Titze & Story, 1997). In addition, high vowels are produced with vocal tract configurations that increase the longitudinal tension of the vocal folds, raising F0 (Honda, 1983; Whalen & Levitt, 1995).

#### *Acoustic measures characterising the mammal voice*

The acoustic variation of the source-related components of vocal signals is typically characterised by measures of fundamental frequency (e.g. F0 mean, maximum, minimum) and its variability, including long term variability in the rate of vocal fold vibration (i.e. across the duration of the signal, e.g. F0 contour, standard deviation,

coefficient of variation, major F0 modulation) and short term perturbation (i.e. period-to-period variability in the rate of vocal fold vibration (jitter), and minor F0 modulation). In contrast, the resonance properties of the vocal tract are most commonly quantified by formant centre frequencies and measures of formant scaling or dispersion (e.g. formant spacing, Reby & McComb, 2003; formant position, Puts, Apicella, & Cárdenas, 2012). F0 and formants are highly perceptually salient and have wide-ranging functional relevance, with natural variation in both fundamental and formant frequencies reflecting maturational and sex-related differences in vocal anatomy (Briefer & McElligott, 2011; Evans, Neave, & Wakelin, 2006; Fitch & Giedd, 1999; D. E. Lieberman, McCarthy, Hiiemae, & Palmer, 2001; P. Lieberman & Blumstein, 1988; Rendall, Owren, Weerts, & Hienz, 2004; Titze, 1994; Volodin, Sibiryakova, & Volodina, 2016; Vorperian et al., 2011), as well as communicating a number of other fitness-indexing traits (e.g. Dabbs Jr. & Mallinger, 1999; Liu et al., 2015; Pisanski & Rendall, 2011; Puts, Hodges, Cárdenas, & Gaulin, 2007; Reby et al., 2005). Dedicated resynthesis techniques (e.g. Pitch Synchronous Overlap and Add (PSOLA) resynthesis, Moulines & Charpentier, 1990) allow researchers to manipulate source and filter components of the voice independently to experimentally investigate the relative functional relevance of fundamental and formant frequencies in playback experiments (Charlton, Reby, & McComb, 2008; Feinberg, Jones, Little, Burt, & Perrett, 2005; Lingle & Riede, 2014; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Puts, Gaulin, & Verdolini, 2006).

Another important aspect of the source is the level of effort or force with which the vocal folds are set into vibration. Amplitude or intensity (amplitude per unit area), and its longer-term variability (e.g. amplitude contour, standard deviation, coefficient of variation, time of max) and shorter-term variability (shimmer), characterise this force by

measuring the overall level of energy (and variability in that level) in the acoustic signal. The perceptual correlate of amplitude is loudness (Fletcher & Munson, 1933). While highly perceptually salient, amplitude is often not measured because it requires strict standardisation of distance between vocal source (e.g., the person speaking) and microphone, as well as standardised headphone volume in playback experiments. Amplitude also covaries with subglottal pressure (Gramming, Sundberg, Ternström, Leanderson, & Perkins, 1988), as raising subglottal pressure increases both the frequency of vocal fold vibration and the energy imparted to displaced air upon vocal fold opening (Behrman, 2007).

While vocal fold vibration is generally periodic (i.e. predictable and regular), non-periodic elements are sometimes present in the source signal. The modal voice register is characterised by periodic vibration and is therefore harmonic and perceived as tonal, but when vocal folds vibrate in a more chaotic (i.e. unpredictable and irregular) manner, the voice is perceived as ‘rougher’ (Bhuta, Patrick, & Garnett, 2004; Fastl & Zwicker, 2001). Roughness can be attributed to a number of acoustic regimes, collectively termed nonlinear phenomena (NLP).

Subharmonics are additional spectral components caused by vocal fold vibration at a frequency equal to an integer fraction of the  $F_0$  (in addition to the  $F_0$  itself) (Fitch, Neubauer, & Herzel, 2002; Riede, Owren, & Arcadi, 2004). Biphonation describes the presence of two independent sources of vibration (Reby et al., 2016; Riede et al., 2004; Wilden, Herzel, Peters, & Tembrock, 1998). Amplitude modulation is a subcategory of biphonation, whereby the air displacement of a slower-vibrating, low-frequency source interacts with that of a faster-vibrating, high-frequency source to produce a signal with audible periodic variation in overall intensity (Beeman, 1998; Fitch et al., 2002). Finally, at the upper limits of subglottal pressure and vocal fold oscillatory regimes, the

vocal folds can vibrate irregularly, partially or fully masking a signal's periodicity. This regime of vocal fold vibration is termed deterministic chaos, and has a particularly harsh quality akin to white noise (Fitch et al., 2002; Riede et al., 2004; Wilden et al., 1998).

Other measures of nonlinearities in vocal fold vibration (e.g. harmonics-to-noise ratio (HNR), Wiener's entropy) generally characterise the spectrum's energy distribution on a continuum from periodic (energy in a vocal signal is located only at harmonically-related frequencies) to chaotic or noisy (energy is located equally at all frequencies). Such measures have generally been overlooked in human vocal communication research because of the relative scarcity of nonlinearities in articulated human speech (Owren, Amoss, & Rendall, 2011). Nevertheless, nonlinearities are an important part of mammal vocal communication systems (Fitch et al., 2002; Reby et al., 2016; Riede et al., 2004; Tokuda, Riede, Neubauer, Owren, & Herzel, 2002; Wilden et al., 1998), and may be functionally relevant in human nonverbal vocal behaviour (particularly crying or shouting).

Finally, it can also be useful to characterise the energy contained within the voice as a whole, taking into account both the creation of energy at the source and the filtering of that energy in the vocal tract. Spectral centre of gravity measures quantify the distribution of energy across the sound spectrum by measuring the frequency at which the energy in the signal divides in half, while dominant frequency (i.e., peak frequency) is the frequency in the spectrum with the highest amplitude across a given time frame (usually but not always F0, Bowling et al., 2017). As the energy of vocal signals is generated by the vocal folds, these measures are most strongly influenced by the source, but the length and shape of the vocal tract further shape a spectrum's energy

distribution, and can increase the amplitude of a harmonic above that of F0 (Bowling et al., 2017; Paliwal, 1998).

Crucially, while humans possess unique specialisations that appear to support speech production, (e.g. highly mobile articulators and increased control of subglottal pressure, Owren et al., 2011), all mammals possess a sound-producing larynx and sound-filtering vocal tract. Moreover, the design of this vocal anatomy is highly conserved, with correspondingly conserved operating principles (Owren, 2011; Taylor et al., 2016), meaning that the various vocal parameters described above are largely comparable across mammal species. As an illustrative example, the chacma baboon grunt call strongly resembles an unarticulated human vowel sound due to similar F0 and formant frequencies (Owren, Seyfarth, & Cheney, 1997; Rendall, Kollias, Ney, & Lloyd, 2005). It is on this basis that the foundation for understanding human nonverbal vocalisations as homologous to animal vocal behaviour can be laid, and meaningful acoustic and functional comparisons between the two can be made.

### **Nonverbal Acoustic Communication**

Across the animal kingdom, vocal communication plays a crucial role in intra- and inter-species social communication (Owren, 2011). Regardless of species, acoustic characteristics of the vocal apparatus are influenced both by static body morphology and by faster-varying dynamic physiological processes (e.g., hormonal fluctuations), and thus encode indexical, physically constrained information regarding the vocaliser's age, sex, reproductive status, body size, and internal state. In mammals, as the outcome of social interactions often depends on the physical attributes of individuals (Schmidt-Nielsen, 1975), receivers who are able to perceive and assess the condition of potential

mates, group members, and competitors, are provided an evolutionary advantage in reproductive, affiliative, and survival contexts. As such, vocal signals play a key role in natural, intrasexual and intersexual selection processes (Taylor et al., 2016).

Although indexical voice cues are functional due to physically constrained relationships between acoustic parameters and anatomy or physiology (Charlton & Reby, 2016; Fitch, 1997; Fitch & Hauser, 1995; Reby & McComb, 2003), selection pressures may favour mechanisms that adaptively minimise or maximise the impression of vocally expressed physical dimensions (Maynard-Smith & Harper, 2003; Morton, 1977). In support of this contention, recent research has documented the use of behavioural strategies by nonhuman mammals to manipulate voice components (orang-utans: de Boer & Perlman, 2014; red deer: Fitch & Reby, 2001; fallow deer: McElligott, Birrer, & Vannoni, 2006), which in turn adaptively influence listener responses (red deer: Charlton, Reby, & McComb, 2007b; Charlton et al., 2008; fallow deer: Pitcher, Briefer, & McElligott, 2015). Some species have developed anatomical adaptations that permanently exaggerate traits (red deer, fallow deer, humans: Fitch & Reby, 2001; elephant seals: Sanvito, Galimberti, & Miller, 2007), but remain subject to further constraints, making the information content of signals relatively honest (Charlton, Ellis, et al., 2011; Fitch & Hauser, 2003; Reby & McComb, 2003). Humans too have been shown to volitionally modulate their voices ‘on demand’ (e.g. when imitating a different gender, Cartei, Cowles, & Reby, 2012; or exaggerating their own body size, Pisanski, Mora, et al., 2016) and are hypothesised to utilise such voice modulation across everyday social contexts to elicit favourable judgments and behaviours from others (Pisanski, Cartei, et al., 2016).

In the following section I provide a summary of the cues that are encoded and perceived in mammal vocalisations. I begin with nonhuman mammals, before

demonstrating that the nonverbal characteristics of speech, despite primarily encoding complex semantic information through mobile articulation of the vocal tract, similarly communicate many of the same cues. I then outline evidence in support of the hypothesis that human nonverbal vocalisations, whose functional relevance remains largely under-investigated, are closely aligned with the vocal communication systems of other mammals in signalling indexical and motivational information relevant to reproductive success and survival.

### **Functional indexical cues in nonhuman mammal vocalisations**

Research on nonhuman mammals has shown that as well as serving as indexical cues to identity (e.g. Briseño-Jaramillo, Estrada, & Lemasson, 2015; Cheney & Seyfarth, 1980; Levréro & Mathevon, 2013; Rendall, 2003; see Taylor et al., 2016 for review) and facilitating affiliative social bonding (Cheney, Seyfarth, & Silk, 1995; Clay & Zuberbühler, 2012; Fedurek, Machanda, Schel, & Slocombe, 2013), vocal cues to indexical characteristics mediate mate choice (Charlton, Ellis, Brumm, Nilsson, & Fitch, 2012; Charlton et al., 2007b), intra-male agonistic interaction decisions (Reby et al., 2005; Taylor, Reby, & McComb, 2010), dominance hierarchies, and mating success (Vannoni & McElligott, 2008; Wyman et al., 2012). As the majority of research in this area has concentrated on F0 and formants, I too focus primarily on these vocal parameters. I also draw a distinction between static vocal cues to slow-varying physical characteristics, and dynamic cues to faster-varying internal states. Cues to static information may be behaviourally manipulated to some extent but, due to anatomical constraints, remain relatively constant. Dynamic cues, primarily underpinned by dynamic physiological changes originating in the somatic and autonomic nervous

system that in turn influence the tension and action of vocal muscles (Scherer, 2003), are much more variable.

*F0 as a static cue to age, sex, size, and hormone levels*

As larger species usually have longer, heavier vocal folds, they tend to produce lower F0 calls than smaller animals. For example, African elephants (*Loxodonta* sp.) can produce F0s as low as 16.8 Hz, while Yinpterochiroptera bats, among the lightest mammals, vocalise with an F0 as high as 63.8 kHz (Taylor et al., 2016). This interspecific relationship between F0 and body size has been comprehensively verified in two recent reviews comparing across 67 and 91 species respectively (Bowling et al., 2017; Charlton & Reby, 2016). Furthermore, because of the allometric relationship between body size and larynx size within most mammalian species, F0 decreases with laryngeal growth as individuals age and develop (e.g. goats: Briefer & McElligott, 2011; hamadryas baboons: Pfefferle & Fischer, 2006; vervet monkeys: Seyfarth & Cheney, 1986; elephants: Stoeger & Baotic, 2016; goitred gazelles: Volodin, Efremova, Frey, Soldatova, & Volodina, 2016; Siberian wapiti: Volodin, Sibiryakova, et al., 2016; c.f. piebald shrews: Volodin, Zaytseva, Ilchenko, & Volodina, 2015). In mammalian species with sexual dimorphism in body size (or sexually dimorphic larynges), F0 can also reliably differentiate males (who tend to be the larger sex) from females (koalas: Charlton, 2015; goitred gazelles: Efremova et al., 2016; chimpanzees and bonobos: Mitani & Gros-Louis, 1995; baboons: Rendall, Owren, Weerts, & Hienz, 2004; cheetahs: Smirnova, Volodin, Demina, & Volodina, 2016).

However, because laryngeal growth is largely unconstrained by surrounding skeletal structures (Fitch, 2006), the relationship between F0 and body size does not usually hold within sex and age classes (Ey, Pfefferle, & Fischer, 2007; Garcia, Herbst,

Bowling, Dunn, & Fitch, 2017), although there are a few exceptions (giant pandas: Charlton, Zhihe, & Snyder, 2009; hamadryas baboons: Pfefferle & Fischer, 2006). In addition, the magnitude of size dimorphism and F0 dimorphism between sexes does not appear to be correlated across mammals, suggesting that F0 does not function to honestly communicate size (Charlton & Reby, 2016).

In species where sexual dimorphism in F0 is greater than size dimorphism (i.e. males produce lower than expected F0 after controlling for size differences, Corsican deer: Kidjo, Cargnelutti, Charlton, Wilson, & Reby, 2008; red deer: Reby & McComb, 2003; humans: Rendall et al., 2005), it has been suggested that F0 may be related to sex hormone levels. Indeed, as testosterone is an important determinant of male aggression, competitiveness and sperm quality in mammals (Minter & DeLiberto, 2008; Wingfield, Hegner, Dufty, & Ball, 1990; Zielinski & Vandenberg, 1993), acoustic expression of male androgen levels is likely to have functional relevance in mate choice and male-male competition contexts (see subsection on *F0 as a functional cue to male quality*, pp. 20-22). Testosterone may affect vocal fold morphology in some nonhuman mammals (as it does in humans, Saez & Sakai, 1976), as castrated lambs develop larger larynges when exposed to higher levels of testosterone and dihydrotestosterone (Beckford, Schaid, Rood, & Schanbacher, 1985). Limited empirical support shows that F0 variation in nonhuman mammal vocalisations is related to male hormonal quality (mean F0: Barelli, Mundry, Heistermann, & Hammerschmidt, 2013; dominant frequency: Fedurek et al., 2016; F0 modulation: Charlton, Keating, et al., 2011; Charlton, Swaisgood, Zhihe, & Snyder, 2012), but further studies are required to investigate inconsistencies in the nature of these relationships, and to examine whether reliable cues to testosterone levels exist in other mammal species.

In the absence of constraints on vocal fold morphology, many species have developed adaptations that further decouple F0 and body size at interspecific and intraspecific levels. For example, some bats and primates produce higher than expected vocal frequencies due to thin vocal membranes lining the edge of their vocal folds, subserving echolocation and vocal efficiency (Mergell, Fitch, & Herzel, 1999). Meanwhile, other species possess specialisations including hypertrophied larynges (hammer-headed bats: Bradbury, 1977; howler monkeys: Kelemen & Sade, 1960), fleshy padded vocal folds (Mongolian gazelles: R. Frey & Gebler, 2003; lions: Klemuk, Riede, Walsh, & Titze, 2011; tigers: Titze et al., 2010), and even additional vocal folds located outside the larynx (koalas: Charlton, Frey, et al., 2013), that enable the production of abnormally low F0. Such adaptations may serve to facilitate the communication of information encoded in formants by increasing their salience (as discussed later, see '*Formants, but not F0, communicate body size*', pp. 34-39) (Taylor et al., 2016). These adaptations may also capitalise on between-species and between-sex sound-size relationships, and where F0 signals male androgen levels, between-individual relationships between F0 and testosterone.

#### *F0 as a dynamic cue to emotional or motivational state*

In addition to the influence of static anatomical attributes on vocal fold properties, the source can be dynamically modulated to produce calls that vary in F0, dependent on physiological and environmental factors. In particular, the mammal voice expresses vocal affect, or emotion, defined in general terms as 'an intense but short-living affective reaction to a specific event or stimulus' (Briefer, 2012, p. 1). While it is debatable whether nonhuman mammals subjectively experience emotion (de Waal, 2011), the subjective awareness of a change in physiological state is likely crucial for all

mammals, as it facilitates adaptive responses to significant events (Briefer, 2012). Affective reactions are triggered by dynamic physiological changes originating in the somatic and autonomic nervous system, which in turn cause changes in respiration and salivation rates, as well as in the tension and action of vocal muscles such as the cricothyroid (Scherer, 1986b; Titze, 1994). Such changes can influence subglottal pressure (Titze, 1994) and vocal fold length/tension (Fitch, 2006) respectively, and therefore the frequency at which the vocal folds vibrate<sup>1</sup> and subsequent source-related properties of the voice.

Vocal affect in nonhuman mammals is typically characterised along two key dimensions – arousal (bodily activation) and valence (positive/negative) (Posner, Russell, & Peterson, 2005; Russell, 1980). Arousal is reliably reflected in mammalian calls. As arousal increases, so typically does the F0 of vocalisations, as well as F0 variability, energy distribution (towards higher frequencies), amplitude, calling rate, and in some cases, nonlinear phenomena (Jovanovic & Gouzoules, 2001; Riede et al., 2004; Stoeger, Baotic, Li, & Charlton, 2012; Zaytseva, Volodin, Ilchenko, & Volodina, 2017; see Briefer, 2012 for review). Such arousal-related acoustic changes may be witnessed in neutral or affiliative contexts (e.g. Rendall, 2003), but have predominantly been studied in the context of stress and distress (e.g. Jovanovic & Gouzoules, 2001; Sèbe et al., 2012). In particular, call duration, amplitude, and nonlinear phenomena have been found to increase with intensity of painful electrical stimulation in mice (Eschalièr, Marty, Trolese, Moncharmont, & Fialip, 1988; Jourdan, Ardid, Chapuy, Eschalièr, & Le Bars, 1995; Levine, Feldmesser, Tecott, Gordon, & Izdebski, 1984), and in pigs, more painful castration procedures induce vocalisations with higher peak frequencies (White et al., 1995). Research further shows that adult females of many nonhuman mammal

---

<sup>1</sup> Nervous system changes can also increase the muscle tension of supralaryngeal cavities, resulting in an upward shift in formant frequencies (and energy distribution).

species adjust their response urgency in accordance with nonlinear phenomena- (rhesus macaques: Jovanovic & Gouzoules, 2001) and F0-related arousal cues (cats: Konerding, Zimmermann, Bleich, Hedrich, & Scheumann, 2016; pigs: Weary, Lawson, & Thompson, 1996) in the distress cries of conspecific infants. Moreover, receivers are more responsive to relatively rougher adult alarm calls, indicative of higher urgency or arousal (Blumstein & Récapet, 2009; Manser, Bell, & Fletcher, 2001).

Vocal correlates of emotional valence in nonhuman mammals have received comparatively little investigation and have produced mixed findings (cats: Konerding, Zimmermann, Bleich, Hedrich, & Scheumann, 2016; pigs: Weary et al., 1996), perhaps because it remains unclear what physiological mechanism might lead to systematic acoustic differences between positive and negative affective states. An alternative but overlapping framework for differentiating between calls of similar arousal considers the effect of motivation on vocalisations. Motivational state differs from emotional state in that it refers to the external effect of context on the acoustic characteristics of vocal behaviours, rather than directly to the effect of an animal's internal state (Zahavi, 1982). As such, vocalisations may derive from ritualised display rules specific to individual contexts as much as from indexical cues to emotional state. Indeed, characterisations of the vocal repertoires of individual species often show that most calls are highly context-specific and vary with behavioural cues, rather than being produced in numerous contexts sharing an underlying emotional state (Cao et al., 2016; Dunlop, 2017; Nadhurou, Gamba, Andriaholinirina, Ouledi, & Giacoma, 2016; Smirnova et al., 2016).

Morton's (1977) theory of motivational-structural rules asserts that the acoustic structure of calls is related to the context in which they are produced. In particular, aggressive vocalisations (usually roars, barks or growls) produced in agonistic contexts are typically noisy and relatively low in F0 and formant spacing, capitalising on

perceptual associations between low frequency sounds and large size or dominance. (Morton, 1977; Ohala, 1984; Owren & Rendall, 2001). Meanwhile, distress or fearful vocalisations in hostile contexts, or vocalisations produced in appeasing contexts, are characterised by relatively higher F0 and formants (Morton, 1977; Owings & Morton, 1998), utilising perceptual associations between high frequencies, and small size or submission. Such cross-modal perceptual biases are argued to effectively convey dominance/threat and submission/apeasement, respectively, and are commonly referred to as the ‘frequency code’ (Ohala, 1984). Comparison of calls produced in aggressive and affiliative settings by multiple species (see Briefer, 2012) support the acoustic distinction in F0 predicted by Morton (1977).

In addition, calls produced in contexts of aggression tend to be of longer duration, wider frequency range, and contain fewer frequency modulations, whereas calls produced during non-aggressive behaviours are often shorter, more tonal (but can be noisy in distress contexts), and often characterised by frequency modulation (see Briefer, 2012). However, it is important to note that fearful and friendly contexts engender very different motivational states and as such are likely to display some distinct acoustic characteristics (August & Anderson, 1987). Arousal differences may explain some (but not necessarily all) of these acoustic distinctions.

#### *F0 as a functional cue to male quality*

Playback experiments have demonstrated that F0 encodes information of potential importance in mate choice and competition contexts, particularly in males. For example, male fallow deer producing lower F0 groans are in better physical condition and are more dominant in intrasexual competitive encounters, and as a result have greater mating success than males producing higher F0 groans (Pitcher, Briefer,

Vannoni, & McElligott, 2014; Vannoni & McElligott, 2008). F0 and dominance rank also share a negative relationship in Père David's deer stags (Liu et al., 2015), while mares prefer the voices of more fertile stallions who vocalise at a lower F0 (Lemasson, Remeuf, Trabalon, Cuir, & Hausberger, 2015). Male geladas of higher status produce loud calls with lower F0, and within individuals, exhaustion results in higher F0 calls (Benítez, Roux, Fischer, Beehner, & Bergman, 2016). As previously discussed, F0 expression and associated inferences regarding mate quality in these species may be related to sex hormones, not only because of the relevance of testosterone to male competitiveness and sperm quality (Minter & DeLiberto, 2008; Wingfield et al., 1990; Zielinski & Vandenberg, 1993), but also because endurance against the immunosuppressant costs of testosterone may indicate natural disease resistance (Folstad & Karter, 1992; Zahavi & Zahavi, 1997).

In some species, however, F0 and mate quality are positively rather than negatively related. Dominance status is signalled by higher F0 calls in male chacma baboons (Fischer, Kitchen, Seyfarth, & Cheney, 2004), and red deer hinds prefer males that produce higher F0 roars (Reby, Charlton, Locatelli, & McComb, 2010; Reby & McComb, 2003). Therefore, functional mate quality inferences may depend on different F0-related indexical vocaliser attributes across species. Androgen levels may dictate mating success in species where lower F0 indicates high mate quality. In contrast, where high F0 is preferred (e.g. Reby et al., 2010), receivers may derive mate quality from vocalisers' ability to sustain high arousal levels and/or muscular effort (Taylor et al., 2016). In red deer, while females prefer high F0 in mate choice contexts, males are not more attentive to high F0 roars in competitive interactions (Garcia, Charlton, Wyman, Fitch, & Reby, 2013), suggesting that the functional relevance of F0 can vary

even within species between mate choice and mate competition contexts (Taylor et al., 2016).

While the role of F0 in indexing mate quality may vary across species and contexts, F0 is overall a strongly sexually selected component of mammal vocalisations. This point is emphasised both by the increase in F0 dimorphism during evolutionary transitions towards polygyny and decrease during transitions towards monogamy (Puts et al., 2016), and by the tendency for species with relatively less sperm competition to produce sexual calls with lower F0 than expected for their body size (Charlton & Reby, 2016).

#### *Formants as a functional static cue to body size*

Body size is of great importance to social and reproductive success in mammals (Andersson, 1994), particularly in males, who tend to be larger than females (Weckerly, 1998). Indeed, relatively larger males are more likely to win resource contests with competitors, and tend to be preferred as mates by females (Bisazza & Marconato, 1988; Cooper & Vitt, 1993; Lindenfors, Gittleman, & Jones, 2007). Thus, the vocal communication of body size is likely to be functionally relevant in mammal interactions.

Body size and vocal tract length share an intraspecific (e.g. Ravignani, Gross, Garcia, Rubio-Garcia, & de Boer, 2017) and interspecific allometric relationship (Fitch, 2000), meaning that taller and larger individuals tend to have longer vocal tracts. In turn, vocal tract length is causally related to formant frequencies, such that individuals with longer vocal tracts produce lower and more closely spaced formants (e.g. Fitch, 1997; Riede & Fitch, 1999). Formants provide reliable cues to body size when comparing across mammalian species; recent work comparing across 72 species

demonstrates the robustness of this relationship at an interspecific level (Charlton & Reby, 2016).

Moreover, empirical studies show that the inverse relationship between formants and body size also operates within many species (e.g. goats: Briefer & McElligott, 2011; rhesus macaques: Fitch, 1997; domestic piglets: Garcia, Wondrak, Huber, & Fitch, 2016; giant otters: Leuchtenberger, Sousa-Lima, Ribas, Magnusson, & Mourao, 2016; red deer: Reby & McComb, 2003; domestic dogs: Riede & Fitch, 1999; elephants: Stoeger & Baotic, 2016; bison: Wyman et al., 2012). In fact, recent research showing that crocodile bellows contain reliable formant-based cues to body size (Reber et al., 2017), along with the songs of birds (e.g. Budka & Osiejuk, 2013; M. R. Jones & Witt, 2014; c.f. Favaro, Gamba, Gili, & Pessani, 2017), suggests that the principle of honest signalling via vocal tract resonances may be an evolutionarily ancient, broadly shared trait among amniotes. The negative correlation between formants and body size also often holds within same-sex and age classes (giant pandas: Charlton et al., 2009; elephant seals: Sanvito et al., 2007), though the relationship is weaker when comparing within these classes than between, and is sometimes weaker in females than males (e.g. giant pandas: Charlton et al., 2009; baboons: Rendall et al., 2005).

Multiple species not only perceive size-related formant variation in the calls of conspecifics (e.g. koalas: Charlton, Ellis, Larkin, & Fitch, 2012; red deer: Charlton, Reby, & McComb, 2007a; giant pandas: Charlton, Zhihe, & Snyder, 2010; rhesus macaques: Fitch & Fritz, 2006; domestic dogs: Taylor et al., 2010), but utilise such information as cues to identity and resource holding potential, thus mediating social interactions, mating preferences and competitive contests. For instance, formants are used by red deer and koalas as acoustic cues to assess the body size of potential mates and rivals during the breeding season (koalas: Charlton, Ellis, Brumm, et al., 2012;

Charlton, Whisson, & Reby, 2013; red deer: Charlton et al., 2007b; Reby et al., 2005). In both of these species, males and females respond to agonistic calls with lower formants as more threatening and attractive, respectively. Additional work on fallow deer has shown that bucks attend to size-related formant variation in male groans, and treat groans with lower formants mimicking larger males as more threatening (Pitcher et al 2015). Similar results have been found for male giant pandas (*Ailuropoda melanoleuca*; Charlton et al., 2010) and domestic dogs (*Canis familiaris*; Taylor, Reby, & McComb, 2010). Conversely, male Australian sea lions respond more strongly to barks with higher formants, which should represent smaller potential rivals (Australian sea lions: Charrier, Ahonen, & Harcourt, 2011). It must be noted, however, that an inverse relationship between male formant spacing and body size has not been demonstrated for this species. Finally, formants are negatively associated with mating success in male bison (Wyman et al., 2012). Taken together, these studies demonstrate that formant frequencies influence both inter- and intrasexual selection processes in a diverse range of mammals.

#### *Anatomical adaptations for formant exaggeration*

Many animals show departures in expected allometric relationships between formants and vocal tract length. Males of many polygynous species produce disproportionately lower formants than predicted by sexual dimorphism in body size alone (e.g. fallow deer: McElligott et al., 2006; red deer: Reby & McComb, 2003). Furthermore, the greater the size dimorphism, the greater the degree to which males produce vocal signals with lower than expected formant spacing ( $\Delta F$ ) but not lower than expected  $F_0$  for their size (Charlton & Reby, 2016). Such findings suggest that in species where size communication is functionally relevant, selection pressures also lead

to the evolution of adaptations enabling male callers to exaggerate their apparent body size.

Indeed, males of several mammalian species possess permanently descended or temporarily retractable larynges that extend their vocal tracts and therefore exaggerate the vocal expression of their size (koalas: Charlton, Ellis, et al., 2011; red deer: Fitch & Reby, 2001; Mongolian gazelles: R. Frey & Gebler, 2003; goitred gazelles: Roland Frey, Volodin, Volodina, Soldatova, & Juldashev, 2011; fallow deer: McElligott et al., 2006). Other species lower formant spacing using hyoid or subhyoid air sacs (howler monkeys: Dunn et al., 2015; colobus monkeys: Harris, Fitch, Goldstein, & Fashing, 2006; gorillas: Perlman & Salmi, 2017) or extended proboscises (elephant seals: Sanvito et al., 2007). Interestingly, despite humans also possessing anatomical adaptations for size exaggeration (see *Formants, but not F0, communicate body size*, pp. 34-39), they produce higher than expected  $\Delta F$  for their body size (Charlton & Reby, 2016). Competing selection pressures to decrease facial size (thereby shortening the oral cavity), serving speech production, thermoregulation, and locomotion, may be responsible for this phenomenon (Charlton & Reby, 2016).

While these anatomical specialisations exaggerate the acoustic impression of size, when vocalisers utilise such adaptations to their anatomical limits (e.g. maximum laryngeal retraction), vocalisations remain approximately honest signals (Charlton, Ellis, et al., 2011; Fitch & Hauser, 2003; Reby & McComb, 2003), and thus meaningful to listeners (e.g. Sanvito et al., 2007). In other words, while all individuals may be able to sound larger than they actually are, between-individual differences in body size appear to remain, such that larger individuals continue to sound larger than smaller individuals.

*Behavioural adaptations for functional formant modulation*

As well as anatomical adaptations, many species behaviourally manipulate the location or salience of their formant frequencies. Species with high F0, including giant pandas, sheep, and fur seals, produce calls with vibrato-like fundamental frequency modulation that may improve formant perception by increasing variability in F0 and harmonics (Charlton, Taylor, & Reby, 2017). Meanwhile, orangutans position leaves (Hardus, Lameira, Schaik, & Wich, 2009) or their hands (de Boer, Wich, Hardus, & Lameira, 2015) in front of their mouths to artificially extend their vocal tracts and exaggerate their physical body size. Such tool use is socially learned (Krützen, Willems, & van Schaik, 2011), and has to date only been documented in orang-utans.

Some species capitalise on the static and predictable relationship between formants and body size to dynamically communicate motivation or intent. Lip protrusions, which elongate the vocal tract and reduce formant spacing, tend to occur in aggressively motivated encounters (canids: Fox, 1970; colobus monkeys: Harris et al., 2006), while lip retractions often accompany affiliative or appeasement contexts (canids: Fox, 1970; cf. Faragó, Pongrácz, Range, Virányi, & Miklósi, 2010). Behavioural modulation of formants may in some cases be audience-dependent – for example, dogs with female owners growl with lower fundamental and formant frequencies to more threatening men (Balint, Farago, Miklosi, & Pongracz, 2016).

There is also evidence that some highly social mammals can modulate formants to communicate rudimentary referential information. Diana monkeys (Riede, Bronson, Hatzikirou, & Zuberbühler, 2005; Riede & Zuberbühler, 2003) and meerkats (Townsend, Charlton, & Manser, 2014) are capable of shifting one or both of the first two formants of their alarm calls to quickly distinguish different types of predators to the benefit of their social group. These same two formants, dependent on mouth and

tongue configurations, determine different vowel sounds in human speech (P. Lieberman & Blumstein, 1988; Maddieson, 2009). Such referential nonhuman mammal formant modulations could therefore be primitive precursors to complex articulated human speech (Boë et al., 2017; Pisanski, Cartei, et al., 2016).

### *Other vocal characteristics*

Other acoustic source characteristics have received less attention than F0 and formants, but may also communicate important physiological and anatomical information. Calling rate, as previously mentioned, increases with arousal (see Briefer, 2012 for review), and in species that engage in sustained calling bouts, can be an important indicator of stamina and motivation (red deer: Clutton-Brock & Albon, 1979; baboons: Fischer et al., 2004; fallow deer: Pitcher et al., 2014; Vannoni, Torriani, & McElligott, 2005). Amplitude is rarely investigated due to methodological impracticalities, but also increases with arousal (Zaytseva et al., 2017; see Briefer, 2012 for review), and in bison, with physical condition and motivation (Wyman, Mooring, McCowan, Penedo, & Hart, 2008). High-amplitude calls occur mainly in agonistic and alarm-related contexts, whereas low-amplitude calls occur more often in affiliative contexts (Gustison & Townsend, 2015).

Nonlinear phenomena in vocalisations also remain poorly understood, despite their common occurrence in mammal vocal repertoires (Cazau, Adam, Aubin, Laitman, & Reidenberg, 2016; Fitch et al., 2002; Reby et al., 2016; Riede et al., 2004; Tokuda et al., 2002; Volodin, Volodina, & Frey, 2017; Wilden et al., 1998), particularly in aggressive vocalisations and screams (Morton, 1977; Owings & Morton, 1998).

Empirical studies suggest that nonlinear phenomena signal high levels of motivation (red deer: Garcia, Wyman, Charlton, Fitch, & Reby, 2014; African wild dogs: Wilden et al., 1998) and physical condition (chimpanzees: Riede, Arcadi, & Owren, 2007; Riede et al., 2004). Their mechanistic unpredictability may also prevent habituation to important vocal signals like alarm calls (meerkats: Karp, Manser, Wiley, & Townsend, 2014), and act as attention grabbing signals (koalas: Charlton, Watchorn, & Whisson, 2017; red deer: Reby & Charlton, 2012). Finally, nonlinear phenomena have the potential to highlight formants in vocal signals by exciting a greater range of frequencies (Fitch & Hauser, 1995), and subharmonics have been hypothesised to exaggerate the perception of a given caller's body size (Fitch, 2000; Wilden et al., 1998). Nonlinear phenomena in mammal vocal signals are understudied mainly because traditional voice analysis methods make linear assumptions, and are thus poorly suited for the investigation of acoustic nonlinearities (Tokuda et al., 2002).

### **Functional indexical cues in human speech**

As the preceding section shows, nonverbal acoustic characteristics are key predictors of mate quality and social behaviour in nonhuman mammals, encoding highly perceptible and functional cues to evolutionarily relevant indexical attributes of vocalisers. In humans, however, the predominant form of vocal communication is speech, a system involving voluntary and precise coordination of vocal anatomical structures to articulate arbitrary sounds and combine them into culturally agreed-upon, meaningful combinations. Nevertheless, as I hope this thesis will underscore, humans too produce a wide range of nonverbal vocalisations in everyday communication that are likely to be structurally and functionally homologous to those of other mammals.

A key prerequisite of articulated speech production in humans is vocal control – the capacity to control the larynx and supralaryngeal vocal tract in a flexible and voluntary manner. In particular, our ability to manipulate articulators (e.g. tongue, lips, jaw, soft palate) allows us to alter the shape (rather than the length) of the vocal tract, which affects the relative positions (rather than absolute scaling) of formant frequencies (primarily F1 and F2, Peterson & Barney, 1952; Titze, 1994). Modulating the relative positions of formants enables humans to produce the wide variety of phonemes required for meaningful speech. While recent evidence suggests that baboons utilise a similar vocalic space as do humans (Boë et al., 2017), and that primates possess far greater capacity for vocal control than previously assumed (Pisanski, Cartei, et al., 2016), humans are unparalleled in our ability to volitionally and independently control the source and filter properties of our vocalisations. Furthermore, we can perform these modulations in the complete absence of an associated inducing experience or state (Fitch, 2006).

As the phonemes comprising language are culturally agreed-upon and their production is voluntary, the link between the acoustic structure and semantic meaning of spoken words is almost exclusively **biologically** arbitrary – for example, the different meanings and acoustic structures of the words ‘two’ and ‘three’ do not derive from differences in anatomical attributes or internal state. Moreover, those meanings are not fixed; in Spanish, for example, ‘tu’ (pronounced ‘two’) means ‘you/your’. In contrast, the meaning of nonhuman mammal vocalisations derives directly from the spontaneous and unbridled influence of anatomical and physiological state on vocal apparatus characteristics. Vocal production in nonhuman mammals is primarily controlled by a more evolutionarily ancient neural system, more closely linked to affective circuitry than the neural pathways predominant in human speech (Ackermann, Hage, & Ziegler,

2014). However, a plethora of studies demonstrate that despite these differences, between- and within-individual variation in indexical attributes is still encoded and perceived in the prosodic component of speech.

*F0 and formants: static cues to age and sex*

The human vocal apparatus undergoes hormone-related changes at puberty and menopause/andropause, causing changes in fundamental and formant frequencies (Abitbol, Abitbol, & Abitbol, 1999), but until puberty, sex differences in the growth rate of the vocal folds (Titze, 1994) and vocal tract (Vorperian et al., 2011) are minimal. Pre-pubertal boys speak with the same F0 as girls (250-300 Hz, Lee, Potamianos, & Narayanan, 1999), and while boys speak with lower formant frequencies, this difference is likely behavioural in origin (Cartei, Cowles, Banerjee, & Reby, 2014; Sachs, Lieberman, & Erickson, 1973) as there is little evidence that vocal tract length differs between pre-pubertal boys and girls (but see Vorperian et al., 2005).

During puberty, an increase in circulating testosterone levels among males causes the male larynx to enlarge to a much greater degree (Kahane, 1982) than in females. Post-pubertal males speak with a 50-80% lower F0 ( $M = 120$  Hz, Pisanski et al., 2014) (Hollien et al., 1994; Lee et al., 1999) than do females ( $M = 210$  Hz, see Figure 1), a difference that is clearly perceptible by listeners (Hillenbrand & Clark, 2009). These sex and age differences in F0 are closely aligned with those observed in nonhuman mammals. Additionally, in humans, sex differences in F0 plateau until roughly the age of 60 (Fouquet, Pisanski, Mathevon, & Reby, 2016), when male F0 appears to increase and that of females decreases (Titze, 1994). This is most likely because, during female menopause, a diminishing ratio of estrogens to androgens drives

F0 down, while diminishing testosterone exerts the opposite effect during male andropause (Kreiman & Sidtis, 2011; Titze, 1994).

Testosterone also causes the male larynx to descend further than in females during puberty, paralleling male-specific laryngeal descents in nonhuman mammal species (e.g. Fitch & Reby, 2001; Roland Frey et al., 2011; McElligott et al., 2006). This, in combination with males' increased growth in height (Gaulin & Boster, 1985), elongates the male vocal tract and lowers formants by 15-20% more in men than in women (see Figure 1; Fant, 1960; Fitch & Giedd, 1999; Pisanski & Rendall, 2011). Sex differences in formant frequencies are perceptible, and while F0 is more important than formants in influencing sex judgments (Markova et al., 2016), resynthesis of both F0 and formants is required to effectively change perceived sex (Hillenbrand & Clark, 2009). The voices of both adult men and women are considerably lower in F0 and formants than those of children, and thus easily distinguishable from children's voices (Hillenbrand, Getty, Clark, & Wheeler, 1995).

#### *F0 as a cue to hormonal quality and fitness*

Sexual dimorphism in human F0 is greater than expected based on sex differences in size (Titze, 1989), as with many nonhuman mammal species (Charlton & Reby, 2016). Given that testosterone directly interacts with androgen receptors in the vocal folds (Saez & Sakai, 1976) in addition to stimulating overall body growth (Verdonck, Gaethofs, Carels, & de Zegher, 1999), it has been suggested that this increased sexual dimorphism in F0 derives from sex hormones. Indeed, high circulating testosterone levels are also predictive of lower F0 within adult males (Cartei, Bond, et al., 2014; Dabbs Jr. & Mallinger, 1999; Puts et al., 2012). This relationship may be strongest in men with low cortisol levels (Puts et al., 2016), wherein concurrently high

testosterone and low cortisol levels indicate men's immune function, attractiveness (Rantala et al., 2012), and dominance (Mehta & Josephs, 2010). Human F0 may thus communicate male quality to competitors and potential mates, as in many other mammal species.

Indeed, F0 is directly related to fitness-indexing attributes. Adult men with either lower or less variable F0 self-report more sexual partners (Apicella, Feinberg, & Marlowe, 2007; c.f. Hodges-Simeon, Gaulin, & Puts, 2011), rate themselves as more dominant (Leongómez, Mileva, Little, & Roberts, 2017) and are judged as more dominant<sup>2</sup> and less cooperative (Hodges-Simeon, Gaulin, & Puts, 2010; Knowles & Little, 2016; Puts et al., 2016, 2006, 2007; Schröder, Cowie, Douglas-Cowie, Westerdijk, & Gielen, 2001) than men with relatively higher and more variable F0. Lower mean F0 is also rated as more attractive (Apicella & Feinberg, 2009; Feinberg et al., 2005; Puts et al., 2016; c.f. Sebesta et al., 2017), masculine (Cartei, Bond, et al., 2014), and indicative of better resource acquisition capability (Apicella & Feinberg, 2009), but a higher likelihood of infidelity (Hughes & Harrison, 2017; O'Connor, Re, & Feinberg, 2011) and lower likelihood of investing resources into relationships and potential offspring (O'Connor, Fraccaro, & Feinberg, 2012). Importantly, perceptual biases linked to low F0 in humans have been shown to influence social behaviour including voting preferences (Banai, Banai, & Bovan, 2017; Klostad, 2015; Tigue, Borak, O'Connor, Schandl, & Feinberg, 2012), access to emergency medical care (Boidron, Boudenia, Avena, Boucheix, & Aucouturier, 2016), men's aggressive intent (Zhang & Reid, 2017), and even women's trust (Montano, Tigue, Isenstein, Barclay, &

---

<sup>2</sup> Males and females are also judged as more dominant by both sexes when they speak with higher amplitude, amplitude variability, (Harrigan, Gramata, Lucic, & Margolis, 1989; Pereira, 2000; Tusing & Dillard, 2000), and F1, but a lower F3 (Laukka, Juslin, & Bresin, 2005).

Feinberg, 2017) and long-term memory (D. S. Smith, Jones, Feinberg, & Allan, 2012), highlighting the broad implications of F0 perception.

The importance of male F0 as an evolutionarily relevant cue is underlined by recent work showing that men dynamically shift their F0 in competitive and mating contexts. Men who perceive themselves to be dominant lower their mean F0 when addressing competing males in mate choice scenarios (Hodges-Simeon et al., 2010; Puts et al., 2006), while a decreasing F0 trajectory has been shown to predict high emergent social rank and higher dominance ratings in a group decision-making task (Cheng, Tracy, Ho, & Henrich, 2016). Opposite shifts in F0 occurred in subordinate and low ranking men. In addition, men speak with lower F0 variability when describing themselves to a potential date (Hodges-Simeon et al., 2010). Such pitch modulations appear simultaneously dependent on both perceptions of self and other – men who perceive themselves as dominant lower their mean F0 and its variability when speaking to a more dominant potential employer, while those who rate themselves as low in dominance raise these F0 characteristics (Leongómez et al., 2017). Such F0-related behavioural shifts in listeners are consistent with those observed in nonhuman mammal species, for whom low F0 signals male quality and mediates competitive interactions (see *F0 as a functional cue to male quality*, pp. 20-22).

While in nonhuman mammals, little research has been conducted into acoustic cues to female mate quality, in humans, female speaking F0 appears to serve as a signal of hormonal quality, fitness, and fertility. As women's F0 is linked to dynamic changes in oestrogen and androgen levels, women's F0 also appears to communicate reproductive status and age (e.g., women's voice F0 decreases following menopause, Abitol et al., 1999). Women with lower F0 are perceived as less attractive and less cooperative, but more dominant and better at acquiring resources (Apicella & Feinberg,

2009; Borkowska & Pawlowski, 2011; B. C. Jones, Feinberg, DeBruine, Little, & Vukovic, 2010; Knowles & Little, 2016; c.f. Tsantani, Belin, Paterson, & McAleer, 2016), rate themselves as more dominant (Leongómez et al., 2017), and are more likely to be voted for than political candidates with higher F0 (Klofstad, 2015). Within individuals, female voice F0 varies across the menstrual cycle (i.e. increases during (Bryant & Haselton, 2009) or immediately prior to ovulation (Fischer et al., 2011)), and both males and females rate female voices as more attractive when the female speaker is relatively more fertile (Karthikeyan & Locke, 2015; Pipitone & Gallup Jr, 2008; Puts et al., 2013). Accordingly, women display similar dominance-related F0 shifts to those reported in males (Cheng et al., 2016; Leongómez et al., 2017), although evidence that women modulate their F0 in response to the attractiveness of potential mates is mixed (Anolli & Ciceri, 2002; Fraccaro et al., 2011; Hughes, Farley, & Rhodes, 2010; Leongómez et al., 2014).

*Formants, but not F0, communicate body size*

Body size (i.e. height and weight) predicts fighting ability (Katić, Blažević, Krstulović, & Mulić, 2005) and physical performance (Folland, Cauley, & Williams, 2008; Samson et al., 2000), and plays an important role in predicting dominance, social success, mate choice, and the outcome of resource contests in men (Boidron et al., 2016; Courtiol, Raymond, Godelle, & Ferdy, 2010; Judge & Cable, 2004; Monden & Smits, 2009; Pisanski, Mishra, & Rendall, 2012; Pisanski & Feinberg, 2013; Puts, 2010; Yancey & Emerson, 2016; c.f. Sear & Marlowe, 2009). Weight is also relevant to female fecundity, wherein overweight and underweight women can experience amenorrhoea and other reproductive difficulties (Bolúmar, Olsen, Rebagliato, Sáez-Lloret, & Bisanti, 2000; Kirchengast & Huber, 2004; Lake, Power, & Cole, 1997; Pirke,

Schweiger, Lemmel, Krieg, & Berger, 1985; Rogers & Mitchell, 1952; Schweiger, 1991; Wise et al., 2010). Thus, the vocal communication of body size in humans is likely to have been functionally adaptive in our evolutionary past, as it is in many nonhuman mammal interactions (e.g. Charlton, Ellis, Brumm, et al., 2012; Charlton et al., 2007b; Reby et al., 2005; Vannoni & McElligott, 2008).

Laryngeal growth is not tightly constrained by surrounding anatomical structures (Fitch, 2006; D. E. Lieberman et al., 2001). Moreover, humans' capacity for advanced vocal control allows us to easily volitionally manipulate the vocal folds to dynamically raise or lower F0 (Pisanski, Cartei, et al., 2016). Consequently, the within-sex correlation between body size and F0 is weak, with a recent meta-analysis demonstrating that F0 explains less than 2% of variance in men's and 0.5% of women's heights and weights (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). This comports with abundant evidence that F0 does not function to honestly communicate size in nonhuman mammals (Charlton & Reby, 2016).

In contrast, vocal tract length, more anatomically constrained by laryngeal cartilage and the skull, is more likely to scale allometrically with body size (Fitch & Giedd, 1999). Vocal tract length is causally related to formant frequency spacing, though surprisingly, formant spacing still only explains up to 10% of the variance in men's and women's heights and weights (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). This relatively weak relationship between formants and body size may be due to the lack of a strict relationship between vocal tract length and body size, and/or behavioural modulation of the size and shape of the vocal tract, and therefore formants (see Pisanski, Mora, et al., 2016 for empirical evidence that men and women can modulate their formant spacing to exaggerate or minimize their physical body size).

Amalgamated measures representing multiple formants, such as  $\Delta F$  (formant spacing, Reby & McComb, 2003) and  $P_f$  (formant position, Puts et al., 2012), perform best at predicting body size compared to individual formants (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014), and these measures also effectively predict body size within numerous nonhuman mammal species (e.g. rhesus macaques: Fitch, 1997; giant otters: Leuchtenberger et al., 2016; red deer: Reby & McComb, 2003; domestic dogs: Riede & Fitch, 1999; bison: Wyman et al., 2012). In humans, formant-weight correlations are generally weaker than formant-height correlations (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014), presumably because vocal tract length is vertically constrained, but also possibly due in part to humans' increased susceptibility to sustained weight gain and obesity in modern environments (Bellisari, 2008; Lev-Ran, 2001; Power & Schulkin, 2008).

Studies show that human listeners can perceive formant cues to body size (Greisbach, 1999; Pisanski & Rendall, 2011; Rendall, Vokey, & Nemeth, 2007), again, in continuity with nonhuman mammals (e.g. koalas: Charlton, Ellis, Larkin, et al., 2012; red deer: Charlton et al., 2007a, giant pandas: 2010; rhesus macaques: Fitch & Fritz, 2006; domestic dogs: Taylor et al., 2010). Listeners are able to assess relative differences in men's and women's heights from their voices only, even in the absence of prior audiovisual experience (i.e. blind listeners perform as well as sighted listeners, Pisanski, Feinberg, Oleszkiewicz, & Sorokowska, 2017; Pisanski, Oleszkiewicz, & Sorokowska, 2016), and from as early as 3 months of age (Pietraszewski, Wertz, Bryant, & Wynn, 2017). However, when the voices of speakers are randomly paired such that relative differences in speaker heights represent a natural range of variation, listeners can correctly identify the taller of two men only 60% of the time on average (González, 2006; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Rendall et

al., 2007), with comparable or lower accuracy for assessment of women's relative size (González & Oliver, 2004; Pisanski et al., 2017; Rendall et al., 2007). Listeners can only reliably discriminate the relative heights of men above chance accuracy when height differences exceed 10 cm, and with high accuracy only when height differences exceed 15-20 cm (Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Pisanski, Oleszkiewicz, et al., 2016; Rendall et al., 2007). This is consistent with work showing that just noticeable differences in vocal tract length are roughly 4-7% in synthesised speech (Ives, Smith, & Patterson, 2005; D. R. Smith, Patterson, Turner, Kawahara, & Irino, 2005).

Listeners' poor accuracy in discriminating the taller of two men (or women) may be partially attributed to individual differences between listeners in their use of spectral and F0 information in height estimation, which can be misleading. Indeed, accuracy of size estimation is highly dependent on F0. Many studies report a consistent perceptual bias in human listeners to associate low-F0 voices with larger body size at the within-sex level (Pisanski & Rendall, 2011; Rendall et al., 2007; D. R. Smith & Patterson, 2005), despite F0 being a poor predictor of body size when controlling for sex and age (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). This erroneous bias may represent an overgeneralisation of between-sex and between-species sound-size relationships (Rendall et al., 2007). Thus, while listeners prioritise formant information over F0 cues when judging body size (Pisanski & Rendall, 2011), listeners' assessments are more accurate when F0 is matched between exemplars in two-alternative forced-choice trials (Rendall et al., 2007), or when the taller of two individuals has the lower F0 (men: Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; women: Pisanski et al., 2017).

Notably, although F0 can confound size assessment, it also acts as a carrier signal for vocal tract resonances. Low pitch provides a dense harmonic spectrum for better resolution of formants (Ryalls & Lieberman, 1982), and yields particularly high accuracy of size estimation (Charlton, Taylor, & Reby, 2013; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014). Finally, as vowels with lower formants and F0 are associated with taller talkers (Barreda, 2016, 2017a, 2017b), phonemically-determined spectral variability may also interfere with accurate formant-based perception of height.

There is some evidence that male listeners are more sensitive than female listeners to acoustic size cues (Charlton, Taylor, et al., 2013; but see Rendall et al., 2007), suggesting that body size perception may have been sexually selected primarily for male-male competition (rather than via female mate choice, see Puts, 2010 for discussion). Anatomical, physiological, and behavioural differences between males and females indicative of a greater role for physical competition in men than women support this assertion – men have larger bodies (Plavcan & van Schaik, 1997), larger hearts, more haemoglobin, less fat, denser bones, and more muscle (Lassek & Gaulin, 2009), and are more aggressive than women (Archer, 2009).

Given the strong influence of body size on social and reproductive outcomes, and the perceptibility of formant-based cues to size, it logically follows that, as in nonhuman mammals (e.g. koalas: Charlton, Ellis, Brumm, et al., 2012; Charlton, Whisson, et al., 2013; red deer: Charlton et al., 2007b; Reby et al., 2005; Australian sea lions: Charrier et al., 2011), formants should influence listeners' fitness-related attributions. Indeed, men with lowered formant frequencies are perceived as more socially and physically dominant by other men (Puts et al., 2012, 2007; Wolff & Puts, 2010), and as more masculine by women (Feinberg et al., 2005). The effect of formants on male voice attractiveness is less clear; relatively low formants are rated as more

attractive only when formant manipulations exceed one just-noticeable difference (Feinberg et al., 2005; Pisanski & Rendall, 2011), and preferences for low formants may vary across cultures (Šebesta et al., 2017), and/or across fertility phases within women (Hodges-Simeon et al., 2010). The relationship may also be curvilinear in some populations, with values closer to the mean considered more attractive (Bundy & Puts, 2013). In females, relatively high formants (Pisanski & Rendall, 2011) or formant characteristics conforming to community accent norms may be considered more attractive by males (Babel, McGuire, & King, 2014).

It follows too that, if formants are functionally relevant to listeners, then humans, like nonhuman mammals (e.g. orangutans: de Boer et al., 2015; fallow deer: McElligott et al., 2006), should be capable of manipulating acoustic characteristics to convey large or small size. Recent research shows that adults across multiple cultures spontaneously increase not only apparent vocal tract length but also decrease F0 to convey larger size, and do the opposite to sound smaller (Pisanski, Mora, et al., 2016), indicating that vocalisers may exploit listeners' perceptual biases. Men generally modulate their voices (formants in particular) more than women to sound physically larger (Pisanski, Mora, et al., 2016), again suggesting that the adaptive value of body size communication is rooted in male-male competition. To what extent such volitional modulation of conveyed body size (and indeed other volitional modulation of voice, but see Hughes, Mogilski, & Harrison, 2014) influences listeners' perceptions remains to be investigated (Pisanski, Cartei, et al., 2016).

#### *Does speech contain acoustic cues to strength?*

While greater physical size can afford humans various advantages (Folland et al., 2008; Katić et al., 2005; Samson et al., 2000), fighting ability, which dictates access

to resources across cultures (see Puts et al., 2012; Sell et al., 2010), is more strongly related to physical strength than to body size when visually assessed (Sell et al., 2009). Given the inconsistency in the visual relationship between strength and muscle cross-sectional area in humans (D. A. Jones, Rutherford, & Parker, 1989; E. J. Jones, Bishop, Woods, & Green, 2008; Maughan, Watson, & Weir, 1983), the ability to acoustically perceive strength is likely to have been adaptive in competitive contexts. Male strength is also a desirable trait: despite their greater aggressive tendencies, stronger males are rated as more physically attractive, have more sexual partners, and lose their virginity at earlier ages than relatively weaker men (Frederick & Haselton, 2007; Gallup, White, & Gallup, 2007; Lassek & Gaulin, 2009; Sell, Lukazsweski, & Townsley, 2017). Thus, indices of strength may be better signals of formidability and mate quality than indicators of size.

Currently, evidence that human speech contains cues to strength is limited and inconsistent. Researchers have tended to focus on upper-body strength, which correlates with history of self-reported success in conflicts (Sell et al., 2009), is more sexually dimorphic than lower-body strength (Lassek & Gaulin, 2009), and would have been the key determinant of the force produced by weaponry available to our human ancestors (Brues, 1959). In a forager population whose environment is likely to be closely aligned with the conditions in which humans evolved, stronger males speak with lower mean F0, lower formant position, and higher F0 variability (Hodges-Simeon, Gurven, Puts, & Gaulin, 2014) – although the latter finding is inconsistent with previous work on the value of F0 variability in mating contexts (Hodges-Simeon et al., 2010). These vocal cues to strength appear additive to and independent from cues to body size (Hodges-Simeon et al., 2014; Puts et al., 2012), but their existence may be culturally dependent (Puts et al., 2012).

Only one study to date has shown evidence that human listeners can assess physical strength from the voice (Sell et al., 2010). Listeners tracked strength more effectively from male than female voices, and strength-tracking capabilities were comparable across Tsimane tribes people, US students, and Romanian students. However, in Sell et al.'s study, neither F0 nor formants predicted actual physical strength, despite both acoustic variables influencing listeners' strength attributions. Furthermore, female strength did not explain variance in strength attributions when controlling for height and weight, suggesting that strength cues may only be additive in males. Clearly, further work is needed to clarify the mechanisms governing the acoustic communication of human strength.

*F0 as a dynamic indicator of arousal, stress, emotion and motivation*

As with other mammals, changes in the activity of humans' somatic and autonomic nervous systems dynamically influence vocal characteristics (Briefer, 2012; Scherer, 1986b; Titze, 1994). While humans may (de Waal, 2011) be unique in subjectively experiencing conscious 'feelings' associated with affective emotional states (Davidson, Scherer, & Goldsmith, 2003), the basis by which environmentally induced physiological changes influence vocal characteristics (termed 'affective prosody') remains the same in humans as in nonhuman mammals (Briefer, 2012). This is illustrated by the fact that empirical attempts to acoustically characterise emotions according to anthropomorphic 'felt' emotion labels produces mixed results (Banse & Scherer, 1996; Juslin & Laukka, 2003; Scherer, Johnstone, & Klasmeyer, 2003; albeit with some cultural similarities, Pell et al., 2015).

These discrepancies in previous work point towards a central role for physiological arousal levels in the communication of affective state. An attempt to

classify emotions according to physiological variation in phonation produced three components that corresponded to subglottal pressure, tension of glottal adduction, and rate of vocal fold vibration, respectively (Patel, Scherer, Björkner, & Sundberg, 2011), all of which are known to be influenced by somatic and autonomic nervous system arousal (see Briefer, 2012). Briefer (2012) conducted a comprehensive review of the literature on both human and nonhuman emotion, and found that high arousal in humans is represented by increases in F0, F0 variability, amplitude, jitter, shimmer, and speech rate – demonstrating clear parallels with nonhuman mammals. Arousal of the autonomic nervous system is a key pathway in the body's response to both psychological and physical stressors (i.e. aversive phenomena), and as such, stress and cortisol are also associated with increases in F0 (Giddens, Barron, Byrd-Craven, Clark, & Winter, 2013; Pisanski, Nowak, & Sorokowski, 2016).

Listeners can identify emotions from human speech, but this ability varies widely both within and across cultures (Juslin & Laukka, 2003; Scherer, Clark-Polner, & Mortillaro, 2011). Acoustic and perceived voice characteristics explain only a third of the variance in listeners' discrete attributions of emotion (Bänziger, Hosoya, & Scherer, 2015); psychological and cultural determinants of voice production (e.g. socio-cultural and linguistic conventions, individual differences, voluntary voice control) likely complicate the expression and perception of emotion from voice (Briefer, 2012). In contrast, acoustic variables mediate an impressive 84% of the direct effect of arousal expression on arousal perception from voice, with high mean F0, amplitude, and amplitude range constituting the most perceptually relevant cues to arousal (Bänziger et al., 2015). Thus, acoustic cues communicate arousal to a greater degree than they differentiate discrete emotion categories.

While dimensional approaches consistently demonstrate arousal as a central aspect of vocal affect communication, studies examining valence are noteworthy for their complexity and widespread inconsistencies in their results (Belyk & Brown, 2014). Positive valence has been associated with increased formants (Briefer, 2012), F0, and amplitude (Belyk & Brown, 2014), but the physiological mechanism mediating the expression of valence remains unclear. Somatic and autonomic nervous system activity can induce changes in the vocal tract; for example, a decrease in salivation raises the resonant frequencies of the vocal tract (Scherer, 1986; Zei Pollermann & Archinard, 2002). Furthermore, contraction of the sternothyroid and sternohyoid muscles pulls the larynx downward, elongating the vocal tract and therefore lowering formant frequencies (Titze, 1994). However, at present it is unclear to what extent valence is acoustically communicated, and what role the filter serves in affective communication.

In addition to the unbridled effects of emotional arousal, the prosodic quality of speech can be volitionally modulated to signal affect. Such controlled manipulation utilises relationships between sound and size (sound-size symbolism). Relatively high F0 and formant spacing, associated with smallness and submission, are predicted to be employed in social interactions to signal politeness, deference and affiliation, whereas relatively low F0 and formant spacing, associated with largeness and dominance, may be used to convey anger, aggression, and threat (Ohala, 1984).

### **Functional indexical cues in human nonverbal vocalisations**

The preceding sections show that human speech and nonhuman mammal vocalisations share striking similarities, suggesting that nonverbal communicative mechanisms have a shared evolutionary origin across mammalian species, with sexual

selection processes playing a key role (Andersson, 1994; Darwin, 1871; Owren, 2011; Taylor & Reby, 2010).

However, human vocal communication is not limited to verbal signals. Indeed, from babies' cries (Lingle, Wyman, Kotrba, Teichroeb, & Romanow, 2012) and playground squeals (Fry, 1987; P. K. Smith & Lewis, 1985), to laughter (Bryant & Aktipis, 2014) and sexual vocalisations (Levin, 2006), to sportspeople's grunts (Callison et al., 2014; Welch & Tschampl, 2012) and supporters' cheers (Myers et al., 2012; Nevill et al., 2002), nonverbal vocalisations (NVVs) constitute a meaningful and important dimension of the human vocal repertoire. They likely predate language and thus speech (e.g. laughter, Niemitz, 1990) and are closely aligned with the vocal communicative systems of other primates (Burling, 1993) and other mammals, whose sole function is to signal indexical and motivational information relevant to natural and sexual selection processes. As such, it is reasonable to hypothesise that the acoustic structure of human nonverbal vocalisations has been selected to support the functional communication of indexical information. Investigating the structure and function of these vocalisations may provide valuable insight into the nature of mammal and early human vocal communication.

Despite their widespread presence throughout life, and their apparent homology with animal vocalisations, nonverbal vocalisations remain severely under-represented in the scientific literature on human vocal communication. The majority of research to date has focused on acoustic and perceptual classification of NVVs according to emotion-label typology (e.g. Anikin & Persson, 2017; Belin, Fillion-Bilodeau, & Gosselin, 2008; Hawk, van Kleef, Fischer, & van der Schalk, 2009; Lima, Castro, & Scott, 2013; Parsons, Young, Craske, Stein, & Kringelbach, 2014; Sauter, Eisner, Calder, & Scott, 2010; Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009; Szameitat et

al., 2009). Aside from laughter (e.g. Bachorowski, Smoski, & Owren, 2001; Bryant et al., 2016; Bryant & Aktipis, 2014; Lavan, Scott, & McGettigan, 2015) and infant distress cries (e.g. Kelly et al., 2017; Koutseff et al., 2017; Lingle & Riede, 2014), the acoustic structure and communicative functions of human NVVs from an evolutionary perspective have received very little scholarly attention.

Are human nonverbal vocalisations merely cultural expressions, or do they function to mediate selection processes, similarly to nonhuman mammal vocalisations? If human and nonhuman mammal vocalisations share a common underlying motivation, do they also share a similar acoustic structure that influences receivers in a comparable way, as hypothesised by Morton (1977) and others (Owings & Morton, 1998; Owren & Rendall, 2001)? These questions have not yet been answered in a comprehensive or coherent fashion, but what little we do know points toward structural and functional similarity in the nonverbal vocalisations of humans and nonhuman mammals.

### *Laughter*

Laughter has received substantial attention from the scientific community, most likely because of its ubiquity in interpersonal interactions across human cultures, and the many emotional social signalling functions it serves (Scott et al., 2014). Humans begin to produce spontaneous laughter within the first few months of life (Sroufe & Wunsch, 1972), even in the absence of visual or acoustic experience (Eibl-Eibesfeldt, 1970; Makagon et al., 2008). Furthermore, laughter is thought to predate speech (Niemitz, 1990) and is produced and recognised across cultures (Sauter, Eisner, Ekman, & Scott, 2010), suggesting that it is a behaviour with deep evolutionary roots.

The capacity to produce volitional laughter develops later in infancy as an intentional communicative act, serving various social functions (Gervais & Wilson,

2005; Nwokah et al., 1994; Szameitat et al., 2009). These two laughter types (spontaneous and volitional) are developmentally, acoustically, perceptually, and neurally distinct (Bryant & Aktipis, 2014; Lavan & McGettigan, 2017; Lavan et al., 2015; Pisanski, Cartei, et al., 2016), although neural pathways for both spontaneous and volitional laughter interact and are partially interdependent (Wattendorf et al., 2013; Wild, Rodden, Grodd, & Ruch, 2003). As such, humans often produce laughter with both affective and volitional characteristics, such as conversational laughter (Gervais & Wilson, 2005).

Some researchers have analysed the acoustic structure of spontaneous laughter to examine its indexical content and communicative function. This research has found sex differences in F0 and formant frequencies in spontaneous laughter, analogous to those existing in human speech and nonhuman mammal vocalisations (Bachorowski et al., 2001; Rothgänger, Hauser, Cappellini, & Guidotti, 1998; Szameitat, Darwin, Szameitat, Wildgruber, & Alter, 2011). Moreover, indirect comparisons between these studies and quantitative analyses of child laughter (Hudenko, Stone, & Bachorowski, 2009; Nwokah, Davies, Islam, Hsu, & Fogel, 1993) indicate that decreases in speech F0 as a result of male and female puberty extend to laughter, while differences in the F0 of laughs between 20- and 70-year old males (La Pointe, Mowrer, & Case, 1990) are also congruent with later age-related changes in speech.

In further continuity, many of the cues that communicate arousal in speech and nonhuman mammal vocalisations also predict arousal ratings produced in response to laughter. Compared to laughter perceived as low in arousal, laughter rated as more aroused has a higher and more variable F0, an energy distribution with a higher centre of gravity, and is faster paced (higher laugh rate, lower inter-bout interval) and noisier. However, jitter, shimmer and amplitude appear not to covary with arousal in laughter

(Lavan et al., 2015; Szameitat, Darwin, Wildgruber, Alter, & Szameitat, 2011; Wood, Martin, & Niedenthal, 2017).

As well as sharing similarities with human speech, a growing body of comparative evidence points toward a close homology in form and function between human laughter and innate tickling-induced play vocalisations produced by a number of primate species. Laughter serves to extend social play across species, promoting social affiliation and the development of cooperative and competitive behaviours (see Davila-Ross, Allcock, Thomas, & Bard, 2011). Both human and nonhuman primate laughter are often produced without voicing (Bachorowski & Owren, 2001; Davila-Ross, Owren, & Zimmermann, 2009), and are characterised by similar interval duration, serial organisation, and high intra-bout variability in acoustic parameters (Vettin & Todt, 2005). In addition, some ape laughter vocalisations are produced with regular vocal fold vibration and consistently egressive airflow – call characteristics previously described as markers of human laughter and speech (Davila-Ross et al., 2009). Quantitative phylogenetic trees constructed based on laughter acoustics of humans and other great apes produce a pattern highly similar in interspecific distance to well-established trees based on genetic similarity (Davila-Ross et al., 2009). Interestingly, when spontaneous (but not volitional) human laughs are slowed down and their pitch is proportionally adjusted, they are largely indiscriminable from nonhuman primate vocalisations (Bryant & Aktipis, 2014).

Yet even human volitional laughter appears to be paralleled to some extent in other primates. In addition to spontaneous laughter, chimpanzees, our closest relatives (along with bonobos, Goodman et al., 1998), also produce acoustically distinct ‘conversational’ laughter replications in response to the laughter of conspecifics during social play (Davila-Ross et al., 2011). These laughter replications appear similar to

human conversational laughter in their developmental trajectory and social cohesive function (Davila-Ross et al., 2011; Nwokah et al., 1994; Vettin & Todt, 2005). The response latency of chimpanzee laugh replications is similar to that of human volitional replications of positive emotional expressions, and suggests that humans are not alone in possessing some degree of non-automatic vocal control of laughter (Davila-Ross et al., 2011).

### *Infant distress cries*

Both human and nonhuman mammal infants cry primarily in response to hunger, pain, and isolation (Newman, 2007; Puppe, Schön, Tuchscherer, & Manteuffel, 2005; Weary, Ross, & Fraser, 1997; Zeifman, 2001), and across species these cries are highly effective at engaging caregivers to alleviate emitters' distress (Newman, 2007; Rödel, Landmann, Starkloff, Kunc, & Hudson, 2013; Zeifman, 2001). The neural mechanisms underlying the production of infant cries are also conserved across species – located in the evolutionarily ancient brainstem (Newman, 2007; Panksepp, 2005; Zeifman, 2001).

The acoustic structure of infant distress vocalisations is remarkably similar across the mammal kingdom, taking the form of a tonal sound with a flat or descending frequency contour (Lingle et al., 2012). Nonlinearities (e.g. deterministic chaos) are often present in the infant distress vocalisations of humans (Facchini, Bellieni, Marchettini, Pulselli, & Tiezzi, 2005), nonhuman primates (Jovanovic & Gouzoules, 2001; Jovanovic, Megna, & Maestripieri, 2000; Rendall, Notman, & Owren, 2009), and many other mammal species (e.g., pandas (Stoeger et al., 2012), elephants (Stoeger, Charlton, Kratochvil, & Fitch, 2011), koalas (Charlton, Watchorn, et al., 2017)) and become more common at times of greater distress (i.e. higher arousal). In addition, increased distress in cries is interspecifically associated with higher F0 and amplitude,

and an upward shift in energy distribution (Lingle et al., 2012; Linhart, Ratcliffe, Reby, & Špinka, 2015; Vergne, Pritz, & Mathevon, 2009; c.f. Scheumann et al., 2012) – all characteristics associated with increased arousal in human speech and nonhuman mammal vocalisations (Briefer, 2012).

In distress vocalisations associated with pain, human infants experiencing greater pain produce cries with higher levels of roughness (Facchini et al., 2005; Koutseff et al., 2017; Tiezzi, Pulselli, & Facchini, 2004), higher amplitude (Fuller & Conner, 1995; Lehr et al., 2007; c.f. Maitre et al., 2017), lower variation in amplitude (Bellieni, Sisto, Cordelli, & Buonocore, 2004), longer bout duration (Johnston & O'Shaugnessy, 1987; Porter, Miller, & Marshall, 1986), and a more variable fundamental frequency, F0 (Koutseff et al., 2017; Porter et al., 1986). Mean F0 (perceived as pitch) appears not to correlate linearly with pain levels in infant cries (Johnston & O'Shaugnessy, 1987; Koutseff et al., 2017; Silva et al., 2010; c.f. Porter et al., 1986), but increases abruptly after a certain threshold of high pain is reached ('alarm threshold', Bellieni et al., 2004).

Higher-pitched (Craig, Grunau, & Aquan-Assee, 1988; Porter et al., 1986), louder, and noisier (Porter et al., 1986) human infant cries tend to be judged as more painful or urgent. Both increased F0 and nonlinear phenomena also influence human mothers' perceptions of the distress levels experienced by human infants (Baeck & Souza, 2001; Esposito, Nakazawa, Venuti, & Bornstein, 2015; Gustafson & Green, 1989; Leger, Thompson, Merritt, & Benz, 1996) and infants of other primates (F0 only, Kelly et al., 2017). Similarly, noisier cries (rhesus macaques: Jovanovic & Gouzoules, 2001) and higher frequency cries (pigs: Weary et al., 1996) provoke more urgent responses in nonhuman mammals.

The hearing of many mammalian adults is most sensitive to the peak frequency of their relative conspecific infants' distress calls (Lingle et al., 2012), suggesting that cries have anatomically shaped adaptive caregiver responses. Indeed, the F0 value of an infant's distress call is vital in determining whether a caregiver responds at all. Calls outside species-specific frequency ranges do not elicit responses, yet a caregiver will respond to calls within that range that are produced not just by conspecifics, but by a wide range of other species (Lingle & Riede, 2014; Lingle et al., 2012).

This finding provides insight into the shared nature of infant distress vocalisations. There is little evolutionary pressure to differentiate vocal distress signatures in early life, as this would only reduce the chances of an individual's suffering being alleviated. Thus, the underlying motivation and acoustic characteristics of mammal distress cries are highly conserved, leading caregivers to respond to the cries of a diverse range of mammals, whose cries are in turn also oriented towards a broad sensitivity (as the reproductive consequences of false alarms are less severe than ignoring genuinely endangered kin). Collectively, the evidence concerning infant distress cries strongly points towards homology in the form and function of mammal nonverbal vocalisations.

#### *Arousal in nonverbal vocalisations*

Assessing nonverbal vocal expressions produced while imagining scenarios representing each of the “basic” emotions (Ekman, 1992; anger, fear, disgust, sadness, surprise, Ekman & Friesen, 1971) and a further set of positive affective states, Sauter and colleagues (2010) reported that perceptions of arousal correlated positively with mean F0 and spectral centre of gravity (a measure of energy distribution), as found in nonhuman mammal vocalisations (see Briefer, 2012 for a review). However, arousal

also correlated negatively with F0 variability, contradicting nonhuman mammal research. Later work utilising a similar methodology corroborated the positive correlation between arousal ratings and spectral centre of gravity, as well as finding that arousal positively correlated with mean amplitude and amplitude variability; but found no correlation with mean F0 or F0 variability (Lima et al., 2013). In both studies, acoustic variables explained little variance in valence ratings, though listeners can distinguish positive from negative vocalisations with relative ease (Simon-Thomas et al., 2009).

While listeners recognise negative affective vocalisations more accurately than positive vocalisations and display cultural variation in recognition rates (Koeda et al., 2013; Sauter, 2010; Sauter, Eisner, Ekman, et al., 2010), many vocalisations produced by both adults (Laukka et al., 2013; Sauter, Eisner, Ekman, et al., 2010) and children (Kersken, Zuberbühler, & Gomez, 2017) are reliably recognised above chance across cultures – even those with little access to Western society (Cordaro, Keltner, Tshering, Wangchuk, & Flynn, 2016). No study has investigated whether there is also cross-cultural agreement in the voice-based assessment of arousal. However, recent research has demonstrated striking similarities in the perception of arousal across species<sup>3</sup>. Humans rate human, piglet, fox, and dog vocalisations with higher F0 as expressing higher arousal (Fragó et al., 2014; Filippi, Gogoleva, Volodina, Volodin, & Boer, 2017; Maruščáková et al., 2015), and distinguish urgent cat purrs containing a high frequency component from non-urgent cat purrs (McComb, Taylor, Wilson, & Charlton, 2009). In silver fox vocalisations, spectral centre of gravity and harmonics-to-noise ratio also contribute to the discrimination of high- from low-arousal calls. Finally, English, German, and Chinese natives have been found to use F0 and spectral centre of

---

<sup>3</sup> Evidence that humans perceive valence in other mammal species is however mixed (Fragó et al., 2014; c.f. Scheumann, Hasting, Kotz, & Zimmermann, 2014).

gravity to identify high arousal vocalisations across multiple species from various clades (amphibians, reptiles, and mammals, Filippi et al., 2017). In sum, it appears that across a wide spectrum of nonverbal vocalisations, both human and nonhuman, arousal is both expressed and perceived in a homologous manner.

*Other vocalisations: screams, roars and grunts*

Studies investigating laughter and infant distress cries indicate phylogenetic continuity in both the form and function of human and nonhuman mammal vocalisations. However, these types of nonverbal vocalisations constitute just two of a plethora that permeate our lives. Does this continuity extend to other types of vocalisations? While only one study to date has acoustically analysed a nonverbal vocalisation other than laughter or infant distress cries (adult screams: Arnal, Flinker, Kleinschmidt, Giraud, & Poeppel, 2015), a further body of more indirect evidence indicates that many of these understudied vocalisations are worth investigating. We later examine some of these less-studied vocalisations in this thesis, to expand our understanding of the structure and function of nonverbal vocalisations in social interactions.

Infant cries are not the only type of human nonverbal vocalisation expressing distress. Human adults produce vocalisations in response to pain (e.g. during childbirth, Fuller, Roberts, & McKay, 1993), and also produce screams in response to threat scenarios (Bernat, Calhoun, & Adams, 1999; Blanchard, Hynd, Minke, Minemoto, & Blanchard, 2001). These screams are characterised by rapid amplitude modulation (between 30 and 150 Hz modulation rates), a nonlinear phenomenon corresponding to the perceptual attribute of roughness (Arnal et al., 2015). The rate of amplitude modulation utilised by screams is distinct from that utilised by other speech or natural

signals, and selectively engages subcortical structures critical for receivers to rapidly appraise danger (Arnal et al., 2015).

Distress screams are also produced by adults of numerous nonhuman primate species (e.g. Bernstein & Ehardt, 1985; Cheney, 1977; H. Gouzoules & Gouzoules, 2000) and other animals (e.g. bats: Eckenweber & Knörnschild, 2016; Jiang, Huang, Wu, & Feng, 2017; rodents: Emmons, 1978) in agonistic contexts, functioning to enhance survival probability (Caro, 2005) by warning or attracting the aid of conspecifics (Bernstein & Ehardt, 1985; Cheney, 1977; Fedurek, Slocombe, & Zuberbühler, 2015; S. Gouzoules, Gouzoules, & Marler, 1984; Slocombe & Zuberbühler, 2007), and/or attracting secondary predators (Hogstedt, 1983). For example, the acoustic structure of chimpanzee screams varies with the severity of received aggression (Slocombe & Zuberbühler, 2007), and this acoustic variation is perceived and utilised by listeners (Slocombe, Kaller, Call, & Zuberbühler, 2010; Slocombe, Townsend, & Zuberbühler, 2009). The warning and attention-grabbing functions of these nonhuman mammal screams are consistent with the selective activation of neural systems mediating threat processing in response to human screams (Arnal et al., 2015). Moreover, the propensity of humans and rodents to utilise scream-like vocalisations in defensive situations varies comparably with the nature of received threat (e.g. escapability, severity) (Blanchard et al., 2001).

Chimpanzees produce acoustically distinct vocalisations within agonistic interactions depending on whether they are victims or aggressors (Slocombe & Zuberbühler, 2005), and listeners attend to these acoustic differences (Slocombe et al., 2010). Many other species also produce aggressive vocalisations (e.g. pygmy marmosets: Pola & Snowdon, 1975; red deer: Reby et al., 2005; southern elephant seals: Sanvito & Galimberti, 2000; canids: Tembrock, 1976; bison: Wyman et al., 2012),

which often function to mediate agonistic male-male interactions through the formant-based communication of body size and formidability (sea lions: Charrier et al., 2011; fallow deer: Pitcher et al., 2015; red deer: Reby et al., 2005; dogs: Taylor et al., 2010). These vocalisations appear perceptually comparable to roar vocalisations produced by humans in aggressive contexts (e.g. battle: Conlan, 1999; Merridale, 2006; Rance, 2015; predator deterrence: United States National Park Service, n.d.), and to more ritualised vocalisations produced in competitive agonistic interactions, such as tennis grunts (Sinnott & Kingstone, 2010) or martial arts kiaps (Welch & Tschampi, 2012). While the acoustic structure of these vocalisations has never been analysed, it is possible that they serve a homologous function to competitive vocalisations produced by other mammals in conveying functional indexical information relevant to formidability.

Sexual vocalisations may also point towards continuity in nonverbal vocal communication between human and nonhuman mammals. Research suggests that in primate species that advertise female fertility (e.g. through sexual skin swelling), female copulation calls, among other signalling functions (see Pradhan, Engelhardt, Schaik, & Maestriperi, 2005), serve to advertise ovulation (barbary macaques: Semple & McComb, 2000; yellow baboons: Semple, McComb, Alberts, & Altmann, 2002). The production of coital vocalisations is also an important component of sexual intercourse in humans (Levin, 2006). While such vocalisations have not been systematically investigated, the acoustic structure of women's speech changes with menstrual cycle (Banai, 2017; see Puts, Doll, & Hill, 2014 for review), and listeners perceive cues to fertility in women's modal speech (Bryant & Haselton, 2009; Pipitone & Gallup, 2012; Pipitone & Gallup Jr, 2008), leaving open the possibility that women's coital vocalisations also convey cues to cyclical variation in fertility status.

## Outstanding questions

What little we know about human nonverbal vocalisations suggests that they function to communicate evolutionarily relevant traits of the vocaliser or socially relevant contextual cues, pointing towards a shared evolutionary nature with nonhuman mammal vocal communication. However, given the scarcity with which nonverbal vocalisations have been investigated, many questions regarding their origins, structure, and function remain unanswered.

*Are human nonverbal vocalisations similar in form and function to nonhuman mammal calls?*

Existing research strongly indicates that human laughter and infant distress cries are homologous to nonhuman mammal equivalents, and suggests that other vocalisations may be too. However, many vocalisations have not yet been subjected to quantitative acoustic analysis, nor have their effects on listeners been examined. For example, we do not know if formidability is communicated in human aggressive roars, despite the fact that many nonhuman mammal roars serve this function, and that such information is also likely to be functionally relevant within the contexts in which human roars tend to be produced. Nor do we know if human vocalisations produced in competitive contexts serve similar signalling functions to those produced by nonhuman mammals, such as the roar produced by red deer during male-male competition (Reby et al., 2005); or if the acoustic mediators of pain communication in adult pain vocalisations are consistent with those observed in human infants and nonhuman mammals.

*Are nonverbal vocalisations more effective carriers of certain indexical cues than speech?*

Human speech is a highly sophisticated signal, through which precise coordination of articulatory structures enables the communication of complex referential information. However, in order to transmit linguistic information, acoustic constraints are placed on the speech channel that obfuscate and constrain the communication of indexical cues. For example, the linguistic importance of the position of lower formants (F1 and F2) may interfere with any role that these formants play in the expression of paralinguistic information, such as body size. Indeed, F1 predicts men's and women's heights much less effectively than do the higher formants (i.e. F3 and F4, Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). Speech is also usually produced with a modal, harmonic voice, restricting the expression of nonlinear phenomena such as deterministic chaos and biphonation, which are part of many mammals' normal vocal communication systems (Fitch et al., 2002; Reby et al., 2016; Riede et al., 2004; Tokuda et al., 2002; Wilden et al., 1998). Moreover, as spectral density (dictated by F0 and the spacing of harmonics at integer multiples of F0) must be sufficiently high to excite the lower formants responsible for phoneme encoding, the larynx as a sound source in speech is limited in its range and capability (Titze, 2017).

In contrast, the communicative function of nonverbal vocalisations is not constrained by language (Scott, Sauter, & McGettigan, 2010), but rather, by the anatomical limits to which the shape and tension of the vocal apparatus can be modulated. Thus, being able to utilise the vocal musculature to its full range, nonverbal vocalisations possess a much wider acoustic space within which to communicate information. Titze (2017) argues that the morphological design of the human vocal folds, like that of other mammals, is optimised for vocal communication over distances

for which higher F0, higher amplitude, and fewer unvoiced segments are used. Accordingly, evidence suggests that laughter exhibits larger F0 ranges and higher F0 (Bachorowski et al., 2001) and F1 (Szameitat, Darwin, Szameitat, et al., 2011) values than speech. Indeed, nonverbal vocalisations likely exploit larger ranges for many acoustic characteristics than afforded by prosodic variation in speech, both within and between speakers.

The production of nonverbal vocalisations (along with nonhuman mammal calls) is primarily controlled by an evolutionarily ancient neural system that is more closely linked to affective circuitry than the neural pathways predominant in speech (Ackermann et al., 2014). As unbridled, ‘pure’ (Scott et al., 1997) forms of vocal expression, they may more directly reflect physiological changes in vocal musculature in response to an affect-inducing situation (Scherer, 1986a). At the level of perception, affective information is preferentially decoded from nonverbal vocal expressions than emotionally inflected speech, being better recognised (Scott et al., 1997) and eliciting stronger, earlier, and more differentiated neural responses (Pell et al., 2015). Thus, in addition to freedom from phonological constraints, neural differences between human nonverbal vocalisations and speech may also make the former a more effective medium for the communication of indexical information. To my knowledge, however, nobody has systematically investigated whether this is the case.

#### *How do volitionally produced vocalisations influence listeners?*

While initial forays into human vocal communication focused on static cues to mate quality, recent research highlights that humans are also capable of flexibly manipulating our voices according to social context (see Pisanski, Cartei, et al., 2016). This capacity has almost exclusively been studied in the context of speech, a volitional

communication system (Pisanski et al. 2016). Yet, although nonverbal vocalisations are naturally spontaneous and affect-driven (Ackermann et al., 2014; Burling, 1993), our uniquely advanced vocal control capabilities also allow us to produce nonverbal vocalisations volitionally – that is to say, on demand and independently of immediate context and inducing physiological state. Studying the acoustic structure and functional effects of volitionally produced vocalisations on listeners may provide valuable insight into the evolution of human vocal communication and the origins of speech, particularly the adaptive value of vocal control.

Despite our known capacity to volitionally produce nonverbal vocalisations, research on laughter suggests that volitional vocalisations may not always mirror their spontaneous counterparts in form and function. In other words, it may be difficult to effectively ‘fake’ a laugh. Spontaneous and volitional laughter are produced with distinct acoustic structures (Lavan et al., 2015), and more importantly, the absence or presence of volitional vocal control in laughter has a noticeable impact on listeners’ perceptions (Bryant et al., 2016; Bryant & Aktipis, 2014; Lavan et al., 2015; McGettigan et al., 2015). Recent research suggests that such acoustic and perceptual differences may extend to a range of nonverbal emotional vocalisations (Anikin & Lima, 2017).

At the same time, in the study of both speech and nonverbal vocalisations, very little attention has been given to the notion of volitional voice modulation as an adaptive tool for influencing listeners’ perceptions. Speech studies have focused almost exclusively on the producer rather than the receiver (see Pisanski, Cartei, et al., 2016). The few studies that have examined the effectiveness of voice modulation on listener judgments have produced mixed results (Fraccaro et al., 2013; Hughes et al., 2014; Leongómez et al., 2014); thus it remains unclear whether people can volitionally

modulate the sound of their voice to effectively manipulate listeners' attributions of evolutionarily relevant traits, such as their perceived masculinity or attractiveness.

Nonverbal vocalisation research frequently utilises acted expressions, but rarely examines multiple vocalisations from the same vocaliser, and thus has not addressed whether speakers can produce vocalisations that deceptively manipulate listeners' perceptions. Indeed, the only existing research that has measured listener ratings of multiple vocalisations primarily takes a discrete emotion approach (i.e. characterising vocalisations according to discrete, anthropomorphic 'felt' emotion labels, rather than physiological dimensions such as arousal), but does show that volitional production of nonverbal vocalisations expressing various emotions influences listeners' arousal and valence ratings (Lima et al., 2013).

Pisanski, Cartei et al. (2016) discuss how formant modulation in other mammals for size exaggeration may have become increasingly complex over evolutionary time, and ultimately led to the sophisticated volitional modulation of articulators that characterises human speech. Adaptive manipulation of receiver perceptions through volitional modulation of human nonverbal vocalisations, probable speech precursors (Niemitz, 1990), would have been an important intermediary step in such evolution of vocal control. Work is now required to assess the extent to which voluntary vocal control of the nonverbal characteristics of our voices functionally and adaptively impacts listener's perceptions and behaviour.

### **The Present Thesis**

The central hypothesis of this thesis is that, in continuity with nonhuman mammal vocalisations, the acoustic structure of human nonverbal vocalisations has been selected to support the functional communication of indexical information. To

investigate this hypothesis, and address the above outstanding questions, this thesis uses source-filter analysis to quantify the acoustic structure of human vocalisations (speech and nonverbal vocalisations: roars, screams, and grunts), correlates acoustic characteristics with key speaker indices and listener ratings, and uses resynthesis techniques to experimentally test the functional role of acoustic characteristics in influencing listener's perceptions.

In Chapter 2, I assess whether volitionally produced aggressive roars and distress screams contain within their acoustic structure cues to height and upper-body strength. I then conduct playback experiments to establish relationships between actual strength/height, acoustic characteristics, and absolute ratings of strength and height. I also conduct the same analyses on aggressive and distressed speech, in order to address whether nonverbal vocalisations are more effective carriers of indexical cues than is speech. In Chapter 3, I address a crucial gap in research demonstrating a role for the voice in the communication of formidability. I employ similar playback methodology to that utilised in Chapter 2, to assess for the first time whether listeners can detect variation in the strength and height of vocalisers relative to their own.

In Chapter 4, I examine how pain is vocally communicated by investigating the acoustic characteristics of nonverbal vocalisations simulating different levels of pain intensity. I also perform playback experiments to examine whether vocalisers successfully communicate pain intensity to listeners, and with which acoustic characteristics.

In Chapter 5, I examine whether spontaneously produced tennis grunts (i.e. nonverbal vocalisations produced in a competitive context) convey static cues to sex, height, weight, and age, and dynamic cues to the progress and outcome of male and female professional tennis contests. I also perform playback experiments (using natural

and resynthesised stimuli) to assess the perceptual relevance of sex and contest outcome cues.

Together, these experiments are intended to make a substantial contribution to the argument that human nonverbal vocalisations share continuity in nature and function with nonhuman mammal vocalisations, and show that individuals can volitionally manipulate the production of human nonverbal vocalisations to functionally influence listener perceptions.

## **Chapter 2: Human roars communicate and exaggerate upper-body strength**

### **Chapter summary**

While there is widespread evidence that nonverbal components of human speech signals and nonhuman mammal aggressive vocalisations communicate information about physical attributes of vocalisers, whether human nonverbal vocalisations also communicate formidability (i.e., strength and height) remains unknown. The aim of this chapter is to investigate whether the acoustic structure of aggressive roars, distress screams, and their speech equivalents reflect their function in accordance with the motivational-structural rules that govern nonhuman mammal vocalisations, and whether aggressive roars serve to communicate and exaggerate formidability, in continuity with other mammals. More specifically, the following questions will be explored:

**Question 1.** Does the acoustic structure of aggressive roars and distress screams (and their speech equivalents) align with Morton's (1977) motivational-structural rules?

**Question 2.** Does the acoustic structure of roars and screams (and their speech equivalents) encode strength and height?

**Question 3.** Can listeners detect variation in absolute strength and height from the voice, and what acoustic characteristics inform their judgments?

**Question 4.** Relative to distress stimuli, do aggressive stimuli, and in particular aggressive roars, optimally communicate and/or exaggerate formidability?

**Question 5.** Are there sex differences (in vocalisers and listeners) in the communication of formidability?

Experiment 1 investigates Questions 1 and 2 by analysing the acoustic structure of aggressive roars, aggressive speech, distress screams and distressed speech produced by male and female trained actors, and relating acoustic variation within these stimuli to anatomical variation in upper-body strength and height.

Summary of findings:

- Aggressive roars and distress screams were characterised by distinct acoustic structures aligning with motivational-structural rules: roars were relatively rougher, lower in fundamental frequency ( $F_0$ , perceived as pitch), and had a lower dominant frequency within the F4 range (DFF4). Compared to these nonverbal vocalisations, aggressive speech and distressed speech were less differentiated, and only in roughness.
- Nonverbal vocalisations displayed more variability in acoustic characteristics, and were relatively louder, higher-pitched, and exhibited more nonlinearities than their speech equivalents. Thus, nonverbal vocalisations utilised a greater acoustic space than speech, allowing for more effective distinction between aggressive and distress motivations.
- Reliable cues to strength and height were not consistently encoded in the acoustic structure of our vocal stimuli.

Experiments 2 and 3 employ playback experiments to investigate Questions 3, 4, and 5, exploring whether male and female listeners can estimate the absolute strength and height of vocalisers, and how listeners' attributions relate to acoustic characteristics of the stimuli.

### Summary of findings:

- Listeners were able to consistently estimate strength from aggressive stimuli but not distress stimuli, and most reliably from aggressive roars, consistent with research demonstrating that listeners of many nonhuman mammal species attend to formidability cues in aggressive calls. This result lends support to the emerging hypothesis that deceptive voice modulation is at the origins of selection for advanced vocal control.
- Listeners were able to estimate height from speech stimuli, but much less reliably than they could estimate strength.
- Vocalisers' voices conveyed exaggerated formidability in aggressive contexts relative to distress contexts, consistent with functional exaggeration of perceived body size by nonhuman mammals in aggressive contexts. Roars conveyed exaggerated formidability more effectively than did aggressive speech.
- There were no sex differences in formidability estimation: strength and height were estimated similarly from male and female voices, and by male and female listeners, suggesting that both mate competition and mate choice mechanisms played an important role in selection for the communication of formidability.
- Acoustic characteristics (pitch, loudness, roughness) predicted strength and height ratings, but did not consistently predict *actual* strength or height. This result suggests either that motivational signalling (i.e. variation in individuals' capacity to exaggerate/minimise perceived formidability) obfuscates the true relationship between indexical attributes and acoustic cues, or, that more complex acoustic mechanisms communicate inter-individual variation in formidability than those currently measured.

### Abstract

Although there is widespread evidence that nonverbal components of human speech (e.g., pitch) communicate information about physical attributes of vocalisers, no previous study has examined whether human nonverbal vocalisations also communicate formidability. Here, we investigated whether roars, screams, and speech sentences produced in two contrasting agonistic contexts (aggression and distress) contain acoustic cues to vocaliser strength and height. We then used playback experiments to investigate if listeners could reliably infer vocaliser strength and height from these vocalisations, and measured the acoustic correlates of listeners' judgments. While there were no consistent acoustic cues to strength, listeners accurately judged inter-individual differences in strength. They did so predominantly from aggressive stimuli, and much more reliably from aggressive roars than aggressive speech. Vocaliser height predicted listeners' judgments of height only for speech stimuli (and male distress screams). Our results show that vocalisers are able to maximise the impression of formidability in aggressive contexts relative to distress contexts, and that inter-individual variation in strength may only be honestly communicated in stimuli that function to communicate threat, particularly roars. Thus, in continuity with nonhuman mammals, the acoustic structure of human aggressive roars appears to be selected to communicate, and to some extent exaggerate, functional cues to formidability.

## Introduction

In competitive contests, evolutionary selection processes favour vocal communication of resource holding potential to settle disputes without engaging in potentially costly combat (Andersson, 1994). For example, many terrestrial mammalian species use acoustic cues to body size or dominance in aggressive calls to mediate agonistic male-male interactions (giant pandas: Charlton, Zhihe, & Snyder, 2010; sea lions: Charrier, Ahonen, & Harcourt, 2011; fallow deer: Pitcher, Briefer, & McElligott, 2015; red deer: Reby et al., 2005; domestic dogs: Taylor, Reby, & McComb, 2010).

In humans, the nonverbal components of speech also allow listeners to assess body size, including height and weight (Charlton, Taylor, & Reby, 2013; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Rendall, Vokey, & Nemeth, 2007). However, in humans, fighting ability is more closely related to physical strength than to body size (Brues, 1959; Sell et al., 2009), and thus indices of strength are likely to be better signals of formidability than are indices of size. In addition, male upper body strength explains substantially more variance in women's judgments of men's attractiveness than does height (Sell, Lukazsweski, & Townsley, 2017). Research suggests that human speech may contain acoustic cues to strength (Hodges-Simeon, Gurven, Puts, & Gaulin, 2014; Puts, Apicella, & Cárdenas, 2012; c.f. Sell et al., 2010), though evidence is limited and inconsistent. Moreover, to date only one study has shown evidence that human listeners can assess physical strength from speech stimuli (Sell et al., 2010), with actual physical strength explaining 18% and 7% of the variance in listeners' strength attributions of male and female speakers, respectively. To our knowledge, no previous study has investigated whether human nonverbal vocalisations, such as aggressive roars or distress screams, also communicate formidability (i.e.,

strength, body size). This is despite the similarity and presumed evolutionary continuity between human nonverbal vocalisations and the vocalisations of other mammals (Burling, 1993; laughter: Davila-Ross, Owren, & Zimmermann, 2009, 2010; Pisanski, Cartei, McGettigan, Raine, & Reby, 2016; infant distress screams: Lingle & Riede, 2014; Lingle, Wyman, Kotrba, Teichroeb, & Romanow, 2012; Zeifman, 2001).

To bridge this gap, we compared the ability of listeners to estimate physical strength from human speech and nonverbal vocalisations produced in two hypothetical contexts: aggression and distress. In these two distinct agonistic contexts, nonhuman mammals produce acoustically and perceptually distinct vocalisations whose acoustic features generally follow motivational-structural rules (Morton, 1977). Capitalising on perceptual associations between low frequency sounds and large size or dominance (Ohala, 1984), aggressive vocalisations (usually roars, barks or growls) are typically structurally noisy and low in fundamental frequency (i.e., low-pitched) (Morton, 1977; Ohala, 1984; Owren & Rendall, 2001). In contrast, distress vocalisations are higher-pitched and usually (but not always) tonal, exploiting perceptual associations between high frequencies and small size or submission (Morton, 1977; Ohala, 1984; Owings & Morton, 1998). While aggressive vocalisations often function to display formidability, distress vocalisations typically function to solicit aid (Bernstein & Ehardt, 1985; Hogstedt, 1983; Slocombe & Zuberbühler, 2007).

Humans produce roar-like vocalisations in aggressive contexts (e.g. battle: Conlan, 1999; Merridale, 2006; Rance, 2015; predator deterrence: United States National Park Service, n.d.), and scream-like vocalisations in distress contexts (Blanchard, Hynd, Minke, Minemoto, & Blanchard, 2001). Furthermore, women, who are on average physically weaker than men (Kim et al., 2004; Shen et al., 2004; Stoll, Huber, Seifert, Michel, & Stucki, 2000), are more likely to scream in response to threat

scenarios than are men, whose responses are biased towards aggression (Blanchard et al., 2001).

Following the hypothesis that human roars and screams are homologous to mammalian vocalisations produced in aggressive and distress contexts, respectively, and are likewise affected by anatomical and physiological constraints, we may expect that their acoustic structure encodes honest information about the physical characteristics of the vocaliser (Charlton et al., 2011; Fitch & Hauser, 2003; Fitch & Reby, 2001; Reby & McComb, 2003; Titze, 1994; Wagner Jr, 1992; Zahavi & Zahavi, 1997). However, we may also expect vocalisations produced in an aggressive context (hereafter roars) to function to maximise the expression of threat relative to those produced in a distress or submissive context (screams), which may minimise perceived threat.

### **The present study**

Here, we investigate the acoustic structure of roars and screams (and their speech equivalents), as well as their functional relevance in communicating strength and height. We measured the upper-body strength and height of men and women and audio recorded them producing aggressive roars and distress screams as well as aggressive and distressed speech sentences. We then examined differences in the acoustic structure of these four types of vocalisations, and investigated the effects of vocaliser height and strength on a range of acoustic parameters. Finally, in order to contrast the functional relevance of roars, screams, and speech equivalents in communicating formidability, we asked separate samples of participants to estimate the strength or height of the vocalisers based on their aggressive roars, distress screams, and speech equivalents. Our key hypotheses were that the acoustic structure of the vocal stimuli would reflect their

function in accordance with motivational-structural rules, and thus, that the propensity to encode and communicate cues to formidability would be maximised in the aggressive and nonverbal variants of our stimuli.

### **Experiment 1: Do Aggressive Roars and Distress Screams Follow Motivational-Structural Rules, and Encode Strength and Height?**

In Experiment 1, we analysed the acoustic structure of aggressive roars, distress screams, aggressive speech, and distressed speech, to test whether aggressive and distress stimuli follow similar motivational-structural rules to those observed in nonhuman mammal vocal behaviour. We also examined whether the acoustic structure of roars and screams, like speech, encodes cues to strength and height.

We predicted that the acoustic structure of aggressive and distress stimuli would align with Morton's (1977) motivational-structural rules, with aggressive stimuli characterised by a lower pitch than distress stimuli. We also predicted that the acoustic structure of the recorded vocal stimuli would contain cues to height, but made no predictions regarding the encoding of strength, as previous work has failed to find consistent vocal indices of strength in humans (Hodges-Simeon et al., 2014; Puts et al., 2012; Sell et al., 2010).

## **Method**

### **Participants**

We audio recorded 30 male and 31 female ( $M$  age =  $22.79 \pm 1.12$ ) drama or acting students from the Royal Central School of Speech and Drama and the University of Sussex, who received monetary compensation in exchange for their participation. All participants provided informed consent. None were currently suffering from conditions that might affect their voice (e.g. colds, sore throats).

## **Procedure**

All experiments were reviewed and approved by the University of Sussex's Life Sciences & Psychology Cluster-based Research Ethics Committee (Sci-Tec C-REC) (Certificates of approval: ER/JR307/2, ER/JR307/4, ER/JR307/8).

### *Voice recording*

Vocalisations and speech sentences ( $n = 244$ ) were recorded in a quiet room, with vocalisers standing 150 cm from a Zoom H4n microphone. A chair was placed at this distance to restrict participants from moving closer to the microphone. In the aggressive context, vocalisers were instructed to imagine themselves in a battle or war scenario, about to charge and attack. Vocalisers were instructed first to produce a given speech sentence imagining themselves in this context, and then a nonverbal vocalisation expressing the same motivation. In the distress context, vocalisers were asked to imagine that 'the tables have turned', and that they were now in a position of weakness, with an attacker charging at them, and again to produce a given speech sentence before producing an analogous nonverbal vocalisation. Each speech sentence was dictated by the experimenter and also displayed on a computer screen, and were as follows:

Aggression context: ‘That’s enough, I’m coming for you!’

Distress context: ‘Please, show mercy, don’t hurt me!’

In order to obtain realistic vocal stimuli, participants were encouraged to take as much time as they needed to immerse themselves in each imagined context, and to ‘let go of their inhibitions’. Participants were also given the option not to vocalise if they felt that they could not naturally produce the sentence or nonverbal vocalisation, and to repeat any sentence or vocalisation until they were satisfied with their portrayal.

### *Strength assessment*

After vocalising, participants’ height was measured using metric tape. The average height of our sample (male  $M = 182.03 \pm 0.97$  cm; female  $M = 167.10 \pm 1.19$  cm) compares well with that of the general UK population (male  $M = 175.3$ cm, female  $M = 161.9$  cm, Moody, 2013). Participants’ strength was assessed by measuring flexed bicep circumference, handgrip strength, and chest strength (following Sell et al. (2009), Puts et al. (2012), and others). These measures can explain approximately 55%, 24% and 35% of the variance in strength as measured by weight-lifting machines in male college students, respectively (Sell et al., 2009).

To measure flexed bicep circumference (male  $M = 32.09 \pm 0.60$  cm; female  $M = 28.96 \pm 0.70$  cm), participants were instructed to rest the elbow of their dominant arm on a table while seated, clench their fist, and curl their forearm perpendicular to the table. The experimenter measured the circumference of the bicep at its highest point. A Baseline hydraulic hand dynamometer was used to measure handgrip strength (male  $M = 41.57 \pm 1.36$  kg; female  $M = 26.98 \pm 1.06$  kg) and chest strength (male  $M = 32.70 \pm 1.55$  kg; female  $M = 19.12 \pm 0.90$  kg). We measured the handgrip strength of

participants' dominant arm with the instrument in its standard use (i.e. handle not inverted). To measure chest strength, the removable handle of the dynamometer was inverted. Subjects grasped the handles, held the device to their chest with elbows extended and perpendicular to the body, and pressed the bars together as hard as possible with both hands (Sell et al., 2009).

Each strength measure was recorded twice per subject and the highest achievable score, representing greatest strength, was used in analyses. Strength measures were z-scored and then averaged to create a single strength score for each subject that weighted each strength measure equally, as per Sell et al. (2009).

#### *Acoustic analysis*

Vocal stimuli were analysed using PRAAT 5.3.62 DSP package (Boersma & Weenink, 2017). Recordings were saved as WAV files at 44.1 kHz sampling frequency and 16 bits amplitude resolution.

We performed acoustic analyses using a dedicated batch-processing script in PRAAT containing four distinct procedures. The first procedure of the script characterised fundamental frequency (F0) and intonation (F0 contour variation). The F0 contour was extracted using the *To Pitch (cc)...* command. We systematically inspected each extracted pitch contour and verified it using a narrow band spectrogram displaying the first 2000 Hz of the signal. Erroneous pitch values (e.g. octave jumps) were manually corrected by selecting the appropriate F0 candidate values in the edited pitch object. In segments displaying subharmonics (where, in addition to F0, vocal fold vibration equal to an integer fraction of the fundamental frequency is present, Fitch, Neubauer, & Herzel, 2002), the F0 was systematically preferred over the subharmonic. Where amplitude modulation (a subcategory of biphonation, whereby the air

displacements of two independent sources of vocal energy, one of low frequency and one of higher frequency, interact to produce a signal with audible periodic variation in overall intensity, Fitch et al., 2002) was present, F0 values were selected only if clearly visible and audible. For segments where deterministic chaos (aperiodic, irregular vocal fold vibration, Fitch et al., 2002) was present, the automatically extracted pitch contour generally did not select F0 values; where it did, we manually deselected these values. Each extracted F0 contour (pitch object) was saved as a text file for future reference.

The F0 contour was used to derive the following parameters: mean F0, max F0, min F0, start-end F0 (a measure of the F0 contour), and F0CV (coefficient of variation of F0 over the duration of the signal). During inspection of each spectrogram, we also measured the proportion of the signal for which amplitude modulation was present, and created a measure representing this proportion as a percentage (%AM).

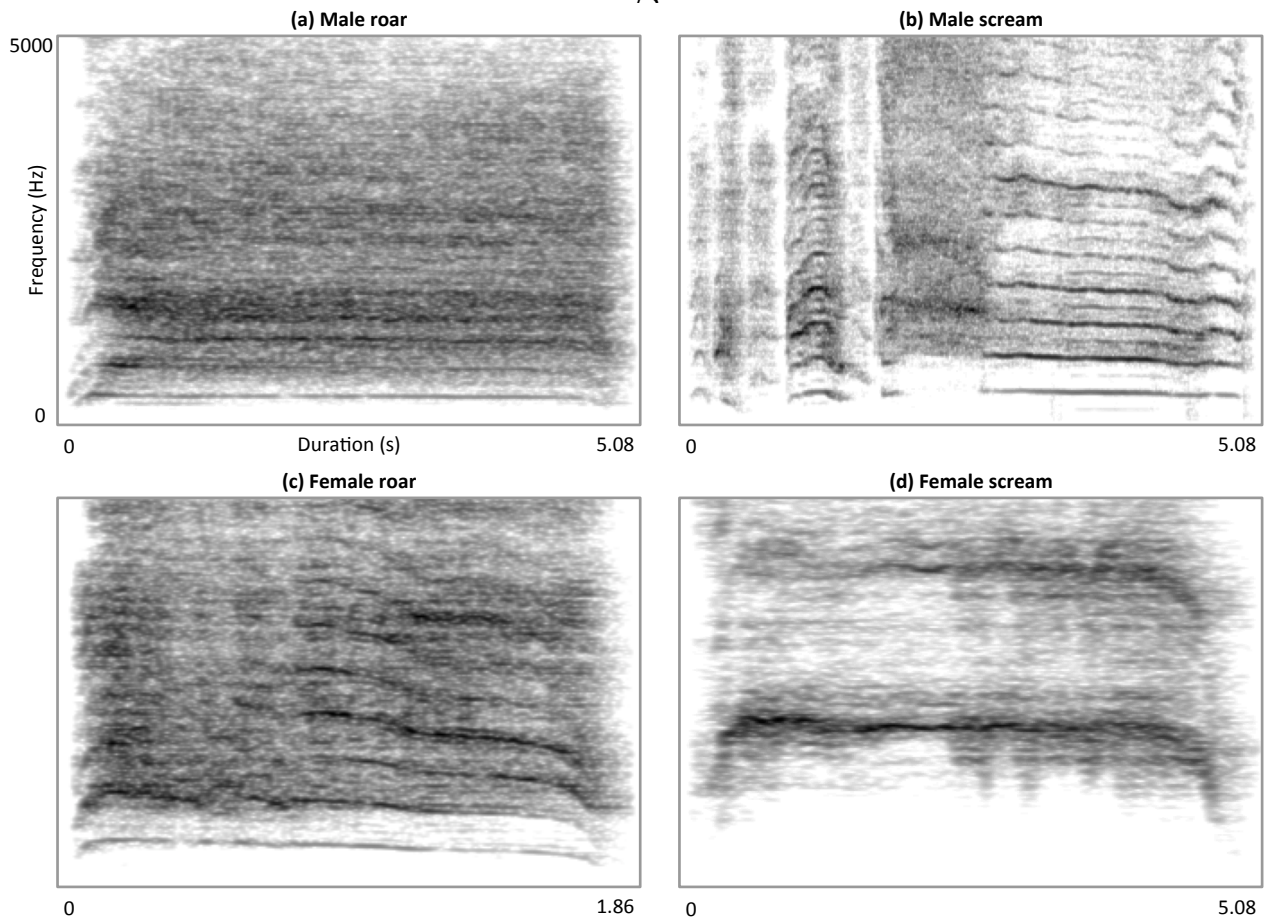
Next, two distinct smoothing algorithms (*Smooth...* command in Praat) were performed on the pitch contour: the first (*Smooth...* command parameter = 25), suppressed very short-term frequency fluctuation while preserving minor modulation events (such as frequency modulation), and the second (*Smooth...* command parameter = 2) only characterised major F0 modulation. Inflection points were counted (as each change in the sign of the contour's derivative) after each smoothing procedure, and divided by the total duration of the voiced segments in each recording, resulting in two distinct indexes of F0 modulation (inflex25 - minor inflections, and inflex2 - major inflections).

A second procedure focused on the intensity contour and characterised the mean amplitude of the stimuli, the point at which the signal's amplitude was highest (time of max intensity, expressed as a percentage of the signal's duration), as well as amplitude variability by calculating intCV, the coefficient of variation of the intensity contour

estimated using the *To intensity ...* command in PRAAT. A third procedure focused on the periodic quality of the signal and measured harmonics-to-noise ratio (HNR, a measure of the ratio of harmonic spectral energy to chaotic spectral energy), an index of jitter (small fluctuations in periodicity measured as the average of ‘local’, ‘rap’ and ‘ppq5’ measures in PRAAT) and an index of shimmer (small variation in amplitude between consecutive periods, measured as the average of ‘local’, ‘apq5’ and ‘apq11’ parameters in PRAAT). Acoustic analysis procedures similar to these have been applied successfully in previous studies of a human nonverbal vocalisation (e.g. babies’ cries, Koutseff et al., 2017; Reby, Levréro, Gustafsson, & Mathevon, 2016).

A final procedure characterised the spectral envelope of each vocal stimulus. Because many of the stimuli were relatively high-pitched (see) and therefore characterised by a low spectral density, and because amplitude modulation (present in many stimuli, see Figure 1a, c) produces sidebands in frequency spectra that can be miscategorised as formants, formant frequencies were poorly defined and difficult to both perceive and measure via cepstrum or linear predictive coding analysis (Pisanski, Fraccaro, Tigue, O’Connor, & Feinberg, 2014; Ryalls & Lieberman, 1982). Instead, this procedure measured spectral centre of gravity (indicating where the ‘centre of mass’ of the spectrum is, calculated as the amplitude-weighted mean of the frequencies present in the signal), which also carries filter-related information (Paliwal, 1998).

Finally, as there is very little or no overlap in the distributions of the third and fourth formants (F3 and F4) across vowels (Abari, Rácz, & Olaszy, 2011; Rendall, Kollias, Ney, & Lloyd, 2005), we attempted to characterise the dominant frequency within sex-specific expected frequency ranges for F4: 3108 and 4250 Hz for males, and 3524-4887 Hz for females, calculated based on published data for male and female formants (Rendall et al., 2005). These data have been used to establish formant



*Figure 1.* Spectrograms illustrating the acoustic structure of a typical (a) male roar, (b) male scream, (c) female roar, and (d) female scream. Note the higher F0 and more chaotic spectral structure of roars than screams.

thresholds in a previous investigation of the vocal communication of upper-body strength (Puts et al., 2012). Minimum values were calculated based on the mean + 0.5 SDs F3 value for (/e/), the vowel with the highest F3 mean and characterised by F3 values 300-700 Hz higher than other vowels. Maximum values were calculated based on the mean + 3 SDs F4 value for (/e/), also the vowel with the highest F4 mean.

We chose a liberal maximum as little is known about the resonance properties of nonverbal vocalisations. We consider this measure (hereafter referred to as DFF4) to be a potential proxy of vocal tract length, as articulatory manipulations of vocal tract shape minimally affect F4 (Rendall et al., 2005), and as the measurement of dominant frequency within an expected F4 range is less likely to capture strong harmonics than

for expected ranges of lower formants (as the amplitude of harmonics declines exponentially with increasing frequency, Titze, 1994). Importantly, F4 is among the strongest formant-based predictors of height in both men and women, explaining a similar amount of variance in height within-sexes as composite formant measures (e.g., formant spacing) and significantly more variance than F1, F2 or F3 (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014).

### **Statistical analysis**

To examine acoustic differences between distress and aggressive stimuli, we conducted a conventional leave-one-out DFA with forced entry (which is less vulnerable to collinear variables, random effects, and type I errors than stepwise entry, Mundry & Sommer, 2007). We entered all the aforementioned acoustic variables except duration, but computed and entered within-sex z-scores in place of raw measures for sexually dimorphic acoustic characteristics (mean F0, max F0, min F0, start-end F0, spectral COG, DFF4). We also conducted a further DFA split by sex to investigate whether there were differences in the discriminability of vocal stimuli between sexes.

To investigate whether strength and height were encoded in the acoustic structure of vocal stimuli, we computed stepwise linear regressions with the aforementioned acoustic variables as predictors, and either actual strength or actual height as outcome variables. These regression models were split by sex, stimulus type (speech/vocalisation), and stimulus context (aggression/distress), to assess whether the encoding of strength and height is dependent on these factors.

## Results & Discussion

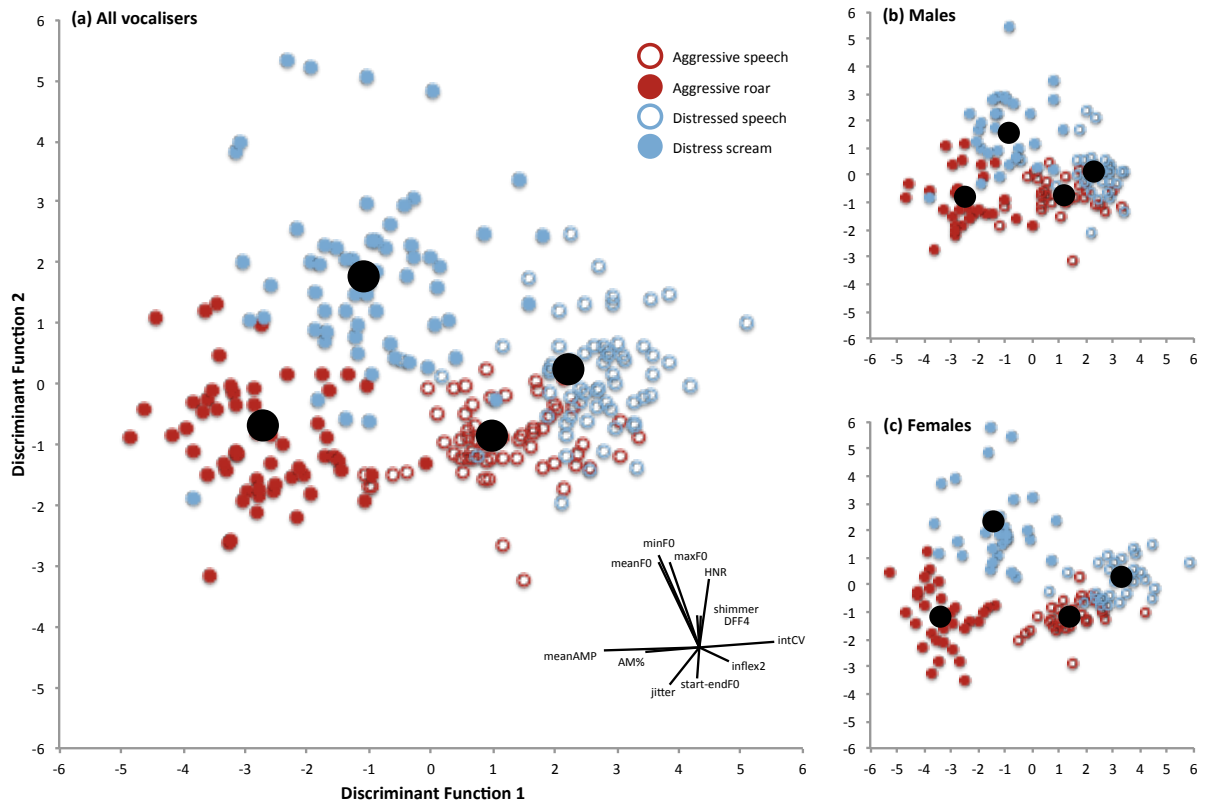
### Do roars, screams, and agonistic speech stimuli differ in acoustic structure?

Discriminant function analyses indicated that all four stimulus conditions (aggressive roars, distress screams, aggressive speech, distress speech) were acoustically distinct (Figure 2): the classification success rate was significantly greater than chance (correct classification percentage = 79.9%, against chance = 25%,  $p < 0.0005$ ). Tables A1-A3 report the factor loadings of acoustic parameters on the first three discriminant functions, collapsing across sexes (Table A1) and for male (Table A2) and female vocalisers (Table A3) separately. The first discriminant function (eigenvalue = 6.43, variance explained = 74.1%) differentiated each of the four stimulus conditions relatively equally (see Figure 2), but separated nonverbal vocalisations from speech sentences. Distressed speech stimuli tended to be the quietest of the four stimulus conditions, with the greatest amplitude variability, the least amplitude modulation, and the most major F0 inflections, followed by aggressive speech, and then distress screams. Aggressive roars were characterised by the highest amplitude, the least amplitude variability, the most amplitude modulation, and the fewest major F0 inflections.

The second discriminant function was less important in discriminating stimulus groups (eigenvalue = 1.93, variance explained = 22.2%), showing primarily that distress screams and, to a lesser degree, distressed speech sentences were more harmonic (HNR) than were aggressive roars and aggressive speech (Figures 1 and 2). F0 variables (mean, max, min) loaded primarily on this function, but also on the first function. Mean values of measured acoustic variables (reported in Tables 1 and 2) showed that distress screams were characterised by the highest F0, followed by aggressive roars, with both

speech stimulus conditions characterised by the lowest F0.

Finally, aggressive roars displayed higher jitter than all other stimuli, whereas distress screams (but not speech) were characterised by higher shimmer and a higher dominant frequency within the expected F4 range than aggressive stimuli. We excluded duration from our discriminant analyses as multiple-word speech sentences were inherently longer than single vocalisations, but we report duration means for each stimulus condition (see Tables 1 and 2). The acoustic characteristics separating vocal stimuli were similar across sexes (Figure 2, Tables A2 and A3).



*Figure 2.* Discriminant function analysis illustrating acoustic separation of stimulus conditions, (a) for all vocalisers, (b) for male vocalisers only, and (c) for female vocalisers only. Each data point represents the centroid of a vocal stimulus as a function of the first two discriminant variables that maximise individual separation. Larger black circles represent mean group centroids for each stimulus condition. The radar plot on the bottom right of panel (a) represents the loadings of the acoustic variables on the first two discriminant functions. Mean amplitude, amplitude variability, and amplitude modulation were the main factors separating stimulus conditions on the first function (DF1, Table A1). The second function (DF2, Table A1) relied mostly on F0 and harmonics-to-noise ratio. The pattern of separation was similar in male (b) and female (c) vocalisers.

*Table 1.* Mean acoustic characteristics of male vocal stimuli. Figures in square brackets represent standard errors.

<b>Acoustic variable</b>	<b>Aggressive speech</b>	<b>Aggressive roar</b>	<b>Distressed speech</b>	<b>Distress scream</b>
Duration (s)	1.92 [0.07]	1.27 [0.12]	2.66 [0.14]	1.35 [0.17]
Mean F0 (Hz)	311.6 [10.96]	378.7 [7.53]	288.5 [11.96]	466.9 [25.50]
Max F0 (Hz)	383.0 [9.04]	428.7 [7.55]	381.4 [21.80]	586.3 [33.39]
Min F0 (Hz)	213.3 [9.17]	273.2 [11.12]	204.8 [9.89]	333.8 [15.06]
Start – end F0 (Hz)	-1.62 [12.85]	31.76 [12.21]	-4.01 [16.92]	-21.64 [23.99]
F0 CV (Hz)	0.15 [0.01]	0.10 [0.01]	0.14 [0.01]	0.13 [0.01]
Minor F0 inflections	6.45 [0.36]	6.58 [0.65]	6.99 [0.41]	5.83 [0.58]
Major F0 inflections	0.88 [0.06]	0.62 [0.09]	0.94 [0.08]	0.60 [0.07]
Mean amplitude (dB)	62.57 [0.94]	71.94 [0.70]	56.39 [1.02]	67.40 [0.84]
Time of max intensity (%)	48.52 [4.66]	41.15 [3.99]	58.83 [4.21]	44.86 [3.95]
Intensity CV (dB)	1.43 [0.05]	0.81 [0.05]	1.53 [0.05]	1.05 [0.06]
Shimmer (dB)	0.14 [0.003]	0.68 [0.35]	0.66 [0.36]	1.47 [0.51]
Jitter (Hz)	0.018 [0.001]	0.029 [0.002]	0.017 [0.001]	0.019 [0.002]
HNR (dB)	7.36 [0.42]	5.51 [0.73]	9.26 [0.48]	10.13 [0.81]
Amplitude modulation (%)	24.02 [3.05]	60.99 [3.76]	11.50 [2.64]	33.81 [4.35]
Centre of gravity (Hz)	1000.3 [37.28]	1143.4 [30.68]	842.2 [41.21]	1085.2 [51.54]
Dominant frequency F4 (Hz)	3381.8 [43.53]	3314.5 [40.14]	3438.3 [45.71]	3508.3 [57.68]

Table 2. Mean acoustic characteristics of female vocal stimuli. Figures in square brackets represent standard errors.

Acoustic variable	Aggressive speech	Aggressive roar	Distressed speech	Distress scream
Duration (s)	1.98 [0.08]	1.21 [0.12]	2.54 [0.12]	1.16 [0.09]
Mean F0 (Hz)	437.1 [14.05]	620.2 [33.93]	420.8 [14.06]	898.6 [65.27]
Max F0 (Hz)	568.7 [16.57]	767.4 [59.56]	557.4 [21.50]	1087.7 [70.06]
Min F0 (Hz)	259.3 [12.11]	398.4 [21.96]	314.0 [12.22]	614.4 [43.07]
Start – end F0 (Hz)	107.1 [20.47]	62.14 [54.04]	5.56 [19.58]	-42.36 [36.19]
F0 CV (Hz)	0.17 [0.01]	0.14 [0.02]	0.13 [0.01]	0.14 [0.01]
Minor F0 inflections	6.37 [0.33]	5.41 [0.80]	8.09 [0.42]	6.41 [0.49]
Major F0 inflections	0.81 [0.07]	0.56 [0.08]	1.02 [0.08]	0.57 [0.06]
Mean amplitude (dB)	61.11 [0.91]	73.97 [0.69]	53.35 [1.21]	68.24 [0.99]
Time of max intensity (%)	38.58 [4.19]	39.39 [3.88]	59.36 [4.41]	43.60 [4.16]
Intensity CV (dB)	1.42 [0.04]	0.76 [0.03]	1.43 [0.05]	0.94 [0.05]
Shimmer (dB)	0.44 [0.30]	1.58 [0.56]	2.10 [0.67]	2.86 [0.67]
Jitter (Hz)	0.018 [0.001]	0.026 [0.003]	0.014 [0.001]	0.015 [0.002]
HNR (dB)	8.36 [0.43]	7.85 [1.14]	10.56 [0.44]	14.02 [0.97]
Amplitude modulation (%)	28.42 [3.19]	48.04 [5.17]	14.52 [1.89]	46.48 [3.79]
Centre of gravity (Hz)	1321.2 [44.60]	1411.8 [43.43]	1156.5 [63.72]	1413.5 [55.91]
Dominant frequency F4 (Hz)	3763.3 [52.94]	3789.6 [57.27]	3881.5 [59.02]	3947.1 [81.69]

The high classification accuracy of our discriminant analysis shows that our stimuli were characterised by distinct acoustic structures that varied according to both stimulus type (speech/vocalisation) and stimulus context (aggression/distress). Nonverbal emotional expressions of anger and have been commonly confused in a previous discriminant analysis (albeit among eight other expressions, Sauter, Eisner, Calder, & Scott, 2010), offering an explanation for our departure from perfect accuracy.

Nonverbal vocalisations displayed more variability in acoustic characteristics, were louder, higher-pitched, and exhibited more nonlinearities (amplitude modulation) than did their speech equivalents. This is consistent with evidence that laughter exhibits higher F0, F0 range (Bachorowski et al., 2001), and F1 (Szameitat, Darwin, Szameitat, et al., 2011) values than speech. Thus, the lack of linguistic constraints in nonverbal vocalisations (Scott et al., 2010) enables the utilisation of a wider acoustic space compared to speech, which necessitates a relatively low pitch (providing sufficiently low spectral density to excite formants responsible for phoneme encoding, Titze, 2017) and places constraints on intonation (contributing to semantic encoding, Brown, 2017; and maintaining phoneme recognition, Miller, Schlauch, & Watson, 2010).

The co-occurrence of high F0, amplitude, and nonlinear phenomena in nonverbal vocalisations indicates that they were produced with high vocal effort (Traunmüller & Eriksson, 2000). F0 and amplitude are both known to increase with subglottal pressure (Behrman, 2007; Herbst, 2016), and nonlinear phenomena (indicating a transition to unstable regimes of vocal fold vibration) arise at the upper limits of subglottal pressure (Berry, Herzel, Titze, & Story, 1996; Fitch et al., 2002; Herbst, 2016; Jiang, Zhang, & Stern, 2001; Zhang & Jiang, 2005). By operating at or near these limits, nonverbal vocalisations may be more subject to anatomical constraints (known to play a major role in ensuring the honesty of acoustic indexical cues: Charlton

et al., 2011; Fitch & Hauser, 2003; Reby & McComb, 2003), and thus signal formidability more effectively than speech. This may be particularly true of aggressive roars, which exhibited the most nonlinearities of all stimuli.

In accordance with motivational-structural rules (Morton, 1977; Ohala, 1984; Owings & Morton, 1998), distress stimuli were more tonal (higher HNR and, within stimulus type, less amplitude modulation) than aggressive stimuli. In nonhuman mammals, distress vocalisations are usually tonal, but may be noisy if fear and aggression are conflicting, or if functioning to solicit support from distant allies (Owings and Morton, 1998, Gouzoules & Gouzoules, 2000). Aggressive roars and distress screams occupied opposite extremes of HNR, again suggesting that vocalisations exploit wider ranges of acoustic space than does speech.

Distress screams were characterised by a higher F0 (see Figure 1), lower jitter, and a higher dominant frequency within the F4 range than aggressive roars (as predicted by motivational-structural rules, Morton, 1977; Ohala, 1984; Owings & Morton, 1998), however these differences were not observed between aggressive and distressed speech. Our results therefore suggest that the acoustic constraints necessary to intelligibly communicate speech (i.e. low F0, harmonic voice, Fitch et al., 2002; Titze, 2017) may limit the expression of motivational-structural rules in speech, while the greater acoustic space afforded by nonverbal vocalisations allows for more effective distinction between aggression and distress.

## **Do roars, screams and agonistic speech stimuli contain acoustic cues to strength and height?**

Strength did not correlate with height among either male ( $r = -.04$ ,  $p = .833$ ) or female ( $r = .083$ ,  $p = .655$ ) vocalisers. Therefore, at least in our sample, these two physical measurements appear to characterise distinct aspects of physical formidability.

We observed very few significant, systematic relationships between acoustic variables and vocaliser height and strength (Tables A4 and A5). The only notable exception was that the dominant frequency within the expected frequency range for the fourth formant (DFF4) was negatively associated with strength for female vocalisers in all calls except distress screams (Table A4).

Our results indicate that cues to height were not consistently encoded in the acoustic structure of our vocal stimuli. In modal speech, formant frequencies (and their spacing) are the only acoustic characteristic known to correlate reliably with vocal tract length, and thus height within sexes (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). However, the prevalence of high pitch (resulting in low spectral density) and/or amplitude modulation (producing sidebands that can resemble formants) (producing spectral prominences that can resemble formants (termed pseudoformants), see Fitch & Fritz, 2006) in our nonverbal stimuli resulted in poor representation of vocal tract resonances. This was also observed to some extent in speech sentences that were produced with high vocal effort. Interestingly, our indices of filter-related (formant) information (COG, DFF4) did not predict height, suggesting that they either failed to capture variation in vocal tract length, or that our stimuli did not contain reliable filter-related cues to speaker size (in part due to the poor acoustic representation of vocal tract

resonances, but possibly also as a consequence of variation in vocalisers' propensity to exaggerate size in aggressive stimuli or minimise size in distress stimuli).

While formant spacing is well-established to correlate with height (Pisanski, Fraccaro, Tighe, O'Connor, Röder, et al., 2014), previous research has produced inconsistent and mixed findings regarding the acoustic encoding of physical strength in speech (Hodges-Simeon et al., 2014; Puts et al., 2012; Sell et al., 2010). Formant dispersion has been reported to correlate with strength in males (Hodges-Simeon et al., 2014; Puts et al., 2012; c.f. Sell et al., 2010), but only in cases where correlations between height and strength were strong (Hodges-Simeon et al., 2014; Puts et al., 2012), indicating that these relationships between strength and formant spacing were mediated by height. Because formant frequencies were not systematically measurable in our stimuli, and because strength was not correlated with height, we suggest that formant spacing is unlikely to constitute a functional cue to strength in these vocal signals. However, the unexpected but consistent association between DFF4 and strength in females suggests that spectral characteristics reflecting complex contributions of both source and filter may still play a role in encoding strength in female vocalisations.

Despite indications that our aggressive roars and distress screams utilised a wider acoustic space than did speech sentences, and despite measuring a much wider set of acoustic characteristics than previous studies examining neutral speech (Hodges-Simeon et al., 2014; Puts et al., 2012; Sell et al., 2010), our investigations failed to reveal consistent acoustic cues to strength. Thus, while one study has reported an association between F0 and strength (Hodges-Simeon et al., 2014) in speech signals, **our study corroborates the more commonly observed lack of significant relationship between F0 and physical strength** (Puts et al., 2012; Sell et al., 2010). In sum, while the acoustic structure of vocal stimuli varied between contexts (aggression vs. distress) in a

way that is consistent with the hypothesis that aggressive roars and distress screams have evolved to maximise or minimise the impression of strength, acoustic structure did not vary systematically with either vocaliser strength or height within call types.

### **Experiments 2 and 3: Can Listeners Estimate Strength and Height from the Human Voice?**

In Experiments 2 and 3, we assessed the functional relevance of aggressive roars, aggressive speech, distress screams, and distressed speech in communicating formidability. Separate samples of listeners judged either the physical strength or height of the vocalisers recorded in Experiment 1.

Given the predominance of male-male competition in shaping men's vocal signals (Hill et al., 2013; Hill, Bailey, & Puts, 2017), we predicted that strength and height would be more reliably estimated from male than female voices. However, as formidability is relevant in both mate competition and mate choice contexts, we did not predict sex differences in listeners' judgments of strength. Moreover, due to the greater relevance of strength than body size to fighting ability (Puts et al., 2012; Sell et al., 2009), and the poor accuracy with which listeners estimate men's relative height (Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Rendall et al., 2007), we predicted that our vocal stimuli would communicate strength more reliably than they communicate height.

Within individuals, one would expect the perception of strength cues to be optimal in aggressive stimuli, as such vocalisations index quantitative information regarding the severity of potential threat (i.e. the formidability of the aggressor),

adaptively influencing decision-making in competitive interactions. In contrast, in distress stimuli, attention to the level of distress experienced by the emitter (indirectly indicating the severity of potential threat) may be selected for, rather than the signaller's formidability. Indeed, in nonhuman mammals, vocalisations produced in aggressive contexts function specifically to signal formidability (giant pandas: Charlton et al., 2010; sea lions: Charrier et al., 2011; fallow deer: Pitcher et al., 2015; red deer: Reby et al., 2005; domestic dogs: Taylor et al., 2010); and in these contexts many species functionally exaggerate acoustic cues to body size (de Boer, Wich, Hardus, & Lameira, 2015; Fox, 1970; Hardus, Lameira, Schaik, & Wich, 2009; Harris, Fitch, Goldstein, & Fashing, 2006; Reby & McComb, 2003). Thus, we predicted that aggressive stimuli would communicate strength more reliably than would distress stimuli, and that aggressive stimuli would be perceived as expressing greater strength than distress stimuli.

## **Method**

### **Participants**

Participants from the USA were recruited from Amazon Mechanical Turk to provide voice-based assessments of strength and height. All participants provided informed consent, and completed the experiments online using a custom computer interface. Participants were compensated with \$3.50 USD. In Experiment 1, 48 females and 42 males (age =  $33.82 \pm 9.60$ ) took part, while in Experiment 2, 30 females and 30 males (age =  $33.80 \pm 8.98$ ) participated. Data from four participants in Experiment 1 and from six participants in Experiment 2 who did not complete the experiment but rated more than half of the stimuli were included in our analysis.

## Stimuli

Participants rated all 244 voice stimuli acquired in Experiment 1 (61 vocalisers x 4 stimulus types), on one dimension (either strength or height). In order to reliably assess the effect of amplitude on listeners' attributions, it was necessary for listeners to maintain the same volume for the duration of the playback experiment. As the difference in amplitude between the quietest (mean = 40.40 dB) and loudest (mean = 81.66 dB) stimulus was large, we partially normalised the amplitude of the stimuli to minimise auditory discomfort while ensuring that listeners could clearly hear all stimuli. Speech stimuli (mean = 58.31 dB) were consistently quieter than vocalisations across sexes (mean = 70.27 dB), therefore we increased the amplitude of speech stimuli and decreased the amplitude of vocalisations by 4 dB each.

## Procedure

This study was hosted by *Syntoolkit*, a dedicated online testing platform for designing and running psychology studies (e.g., Hughes, Gruffydd, Simner & Ward, in press; see Simner & Alvarez, forthcoming). Participants were directed to the URL testing site and provided consent before beginning the study. Listeners were instructed to use headphones and complete the experiment in a quiet place. To allow listeners to complete the experiment at a comfortable but suitable volume to hear all stimuli clearly, they were instructed to first set their volume to its lowest level. Listeners were then told to play a demo sound file (amalgamating the loudest stimulus and the fifth quietest stimulus), and to raise their volume until they could clearly hear the quiet vocalisation, while the loudest vocalisation did not cause discomfort. Following this, listeners were asked not to adjust the volume during the experiment unless it became too uncomfortable. Listeners were asked at the end of the experiment if they adjusted their

volume at any point. Due to the agonistic nature of the stimuli, listeners were made aware that if they felt uncomfortable or distressed listening to the sounds, they could stop the experiment.

Voice stimuli were blocked by sex (male/female), stimulus type (speech/vocalisation), and stimulus context (aggression/distress). The order of blocks and stimuli within blocks was randomised. Before each block, participants were reminded to listen to each stimulus in full before rating it, and informed that they could take a break at any time. Listeners rated the physical strength (Experiment 2) or height (Experiment 3) of each voice stimulus (“Rate how strong/tall this vocaliser is”) on a 101-point scale from 0 (extremely weak/short) to 100 (extremely strong/tall).

Listeners were debriefed upon completion that the roars and screams were acted, and that the vocalisers were not really experiencing aggression or distress. We inspected listeners’ ratings and compared their reaction times against stimulus duration to ensure that they completed the experiments properly. Data from two participants who did not do so were removed (these participants were not reported in the participant statistics given above).

### **Statistical analysis**

We computed a series of linear mixed models. To examine whether males were actually stronger/taller than females, we first tested the effect of vocaliser sex on actual strength/height. Next, we tested the effects of vocaliser sex, listener sex, stimulus context, and stimulus type on attributed strength/height ratings to ascertain whether males were rated as stronger/taller than females, and to investigate how volitional production of multiple vocalisations affects ratings of physical characteristics within vocalisers. The third set of models added actual strength/height into the previous

models, to assess the capacity of listeners to accurately estimate physical characteristics from the voice. As the strength and height distributions for males and females displayed little overlap, we split these models by vocaliser sex rather than including sex as a factor.

In all models, we included listener identity as a subject variable, and vocaliser identity as a random factor, thus allowing the intercepts and slopes of the relationships between predictors and outcomes to vary between both vocalisers and listeners and testing null hypotheses based on the average of these intercepts and slopes. Effect sizes were estimated using  $R^2$  coefficients derived from simple linear regressions among relevant variables, and using  $\gamma$  coefficients derived from the linear mixed models.

Subsequently, we computed stepwise linear multiple regressions to assess the relationships between acoustic characteristics and strength/height ratings. The previously measured acoustic variables were used as predictors, and either mean strength or mean height ratings as outcome variables. Participants who modified their volume during the experiment (Experiment 1  $n = 4$ , Experiment 2  $n = 15$ ) were excluded from the calculation of mean ratings, enabling valid analysis of the effect of amplitude on ratings. The regression models were split by sex, stimulus type (speech/vocalisation), and stimulus context (aggression/distress), to assess whether the encoding of strength and height is dependent on these factors.

## **Results & Discussion**

### **Are there sex differences in actual or rated strength and height?**

#### *Effects of vocaliser sex*

Linear mixed model analysis revealed that males ( $M = 0.81 \pm 0.11$ ) were physically stronger than females ( $M = -0.46 \pm 0.11$ ,  $F_{(1, 61)} = 64.83$ ,  $p < .0005$ ). Males ( $M = 182.03 \pm 1.09$  cm) were also taller than females ( $M = 166.94 \pm 1.04$  cm,  $F_{(1, 61)} = 101.02$ ,  $p < .0005$ ).

However, males ( $M = 60.64 \pm 1.17$ ) were only rated as stronger than females ( $M = 55.53 \pm 1.22$ ) by male listeners judging aggressive roars (Table 3xiv,  $p = .032$ ). For all other conditions (listener sex x stimulus context x stimulus type), females ( $M = 44.63 \pm 1.20$ ) were either rated as equally strong or slightly stronger than males ( $M = 43.91 \pm 1.20$ , Figure 3), indicating that listeners' strength attributions were generally not consistent with sexual dimorphism in actual strength.

Height ratings, on the other hand, were consistent with sexual dimorphism in height. Listeners rated males ( $M = 53.59 \pm 0.50$ ) as taller than females ( $M = 47.81 \pm 0.50$ ) across all stimulus types and contexts (Figure 4, Table 4ii,  $p < .0005$ ). This sex difference in height ratings was larger for aggressive ( $M$  difference = 7.04) than distress stimuli ( $M$  difference = 4.51, Table 6vii,  $p < .0005$ ), and for nonverbal vocalisations ( $M$  difference = 6.50) than for speech sentences ( $M$  difference = 5.06, Table 4viii,  $p = .009$ ).

#### *Effects of listener sex*

Female listeners rated aggressive roars produced by female vocalisers as stronger than did male listeners ( $M$  difference = 2.58, Table 3xvi,  $p = .032$ ), but otherwise produced comparable strength ratings ( $M$  difference for other stimulus conditions = 0.37). Female listeners ( $M = 52.04 \pm 0.66$ ) generally judged vocalisers as taller than did male listeners ( $M = 49.36 \pm 0.66$ , Table 4iii,  $p = .005$ ), particularly when listening to aggressive roars ( $M$  difference = 4.9,  $M$  difference other stimulus conditions = 1.94, Table 6xv,  $p = .046$ ).

*Table 3.* Strength attributions: linear mixed model testing the effects of vocaliser sex, listener sex, stimulus context, and stimulus type on rated strength.

<b>Source</b>		<b><i>df</i><sub>1</sub>, <i>df</i><sub>2</sub></b>	<b><i>F</i></b>	<b><i>p</i></b>
i.	Intercept	1, 88.01	3892.10	< <b>.001</b>
ii.	Vocaliser sex	1, 5398.65	0.00	.970
iii.	Listener sex	1, 88.01	0.06	.813
iv.	Stimulus context	1, 16376.86	2940.38	< <b>.001</b>
v.	Stimulus type	1, 16376.86	285.87	< <b>.001</b>
vi.	Vocaliser sex * listener sex	1, 5398.65	0.02	.876
vii.	Vocaliser sex * stimulus context	1, 16390.45	9.33	<b>.002</b>
viii.	Vocaliser sex * stimulus type	1, 16390.45	13.96	< <b>.001</b>
ix.	Listener sex * stimulus context	1, 16376.86	1.20	.273
x.	Listener sex * stimulus type	1, 16376.86	0.21	.648
xi.	Stimulus context * stimulus type	1, 16376.86	176.99	< <b>.001</b>
xii.	Voc sex * list sex * stimulus context	1, 16390.45	3.38	.066
xiii.	Voc sex * list sex * stimulus type	1, 16390.45	0.01	.921
xiv.	Voc sex * stimulus context * stimulus type	1, 16390.45	33.17	< <b>.001</b>
xv.	List sex * stimulus context * stimulus type	1, 16376.86	7.22	<b>.007</b>
xvi.	Voc sex * list sex * stim context * stim type	1, 16390.45	4.58	<b>.032</b>

*Table 4.* Height attributions: linear mixed model testing the effects of vocaliser sex, listener sex, stimulus context, and stimulus type on rated height.

<b>Source</b>		<b><i>df</i><sub>1</sub>, <i>df</i><sub>2</sub></b>	<b><i>F</i></b>	<b><i>p</i></b>
i.	Intercept	1, 58.16	11922.30	< . <b>.001</b>
ii.	Vocaliser sex	1, 3618.53	279.44	< . <b>.001</b>
iii.	Listener sex	1, 58.16	8.34	<b>.005</b>
iv.	Stimulus context	1, 10577.56	234.15	< . <b>.001</b>
v.	Stimulus type	1, 10476.98	19.87	< . <b>.001</b>
vi.	Vocaliser sex * listener sex	1, 3618.53	1.82	.177
vii.	Vocaliser sex * stimulus context	1, 10578.54	21.54	< . <b>.001</b>
viii.	Vocaliser sex * stimulus type	1, 10421.61	6.91	<b>.009</b>
ix.	Listener sex * stimulus context	1, 10577.56	5.60	<b>.018</b>
x.	Listener sex * stimulus type	1, 10476.98	14.38	< . <b>.001</b>
xi.	Stimulus context * stimulus type	1, 10432.64	5.20	<b>.023</b>
xii.	Voc sex * list sex * stimulus context	1, 10578.54	0.17	.684
xiii.	Voc sex * list sex * stimulus type	1, 10421.61	0.92	.339
xiv.	Voc sex * stimulus context * stimulus type	1, 10406.88	3.81	.051
xv.	List sex * stimulus context * stimulus type	1, 10432.64	3.97	<b>.046</b>
xvi.	Voc sex * list sex * stim context * stim type	1, 10406.88	3.81	.051

Thus, we observed vocaliser and listener sex differences in height attributions, but not strength attributions. Males and females were generally rated as similarly strong, despite sexual dimorphism in actual strength, and female and male listeners generally produced comparable strength ratings. In contrast, listeners' height attributions were consistent with sexual dimorphism in height, and female listeners generally judged vocalisers as taller than did male listeners. We suggest that because strength is a relatively abstract construct (i.e. consisting of multiple dimensions and not commonly quantified or well-defined), listeners may have judged it according to within-sex

expectations; whereas height, as a solitary, commonly used, and well-defined measure, was judged in absolute terms.

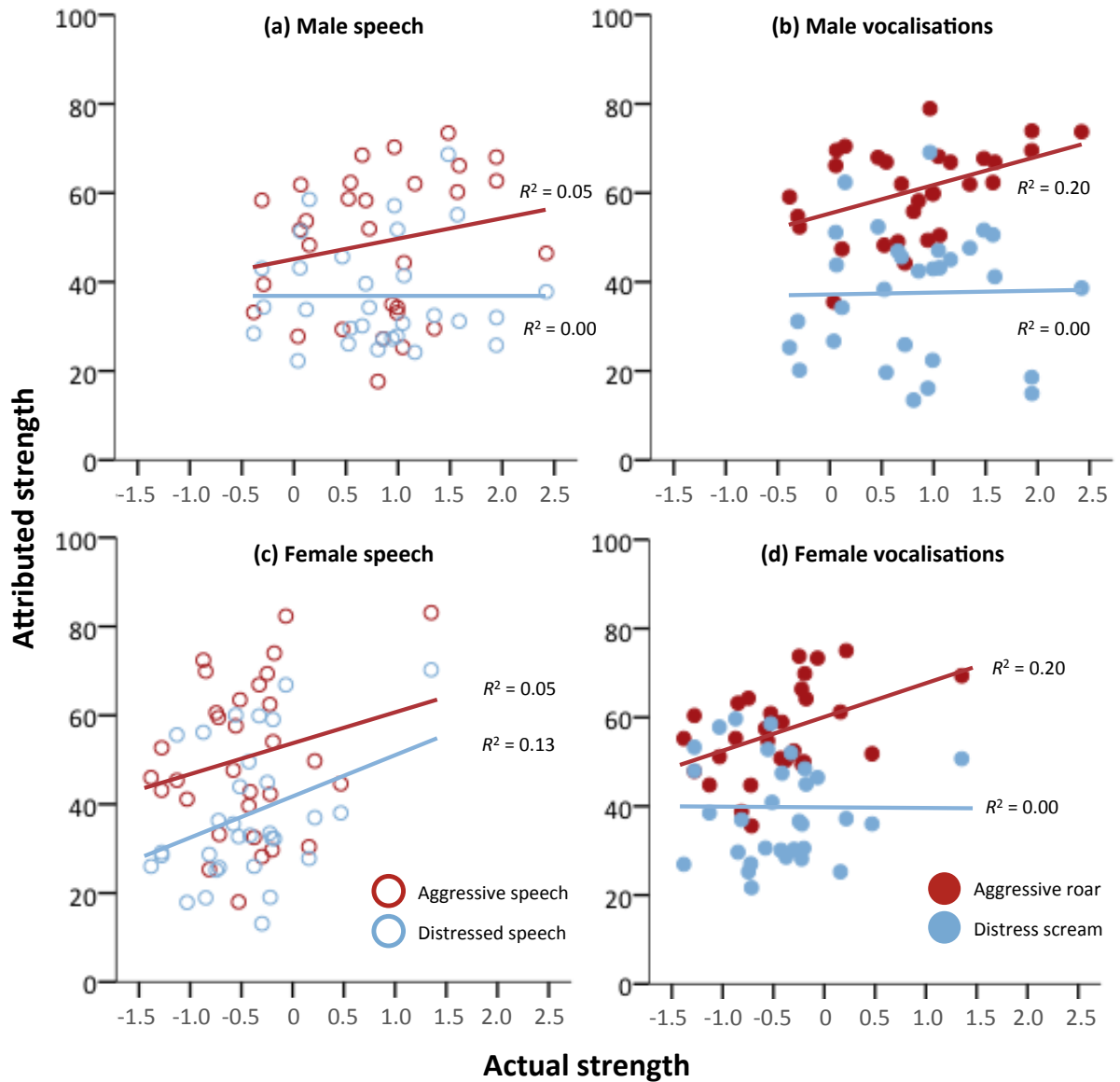
Because men have larger bodies (Plavcan & van Schaik, 1997), larger hearts, less fat, more muscle (Bishop, Cureton, & Collins, 1987; see Lassek & Gaulin, 2009), and are more aggressive than women (Archer, 2009), the communication of formidability is generally considered to have been sexually selected for primarily in male-male competition contexts (Hill et al., 2013, 2017). However, for ecologically relevant stimuli, reliable communication of strength and body size may be equally relevant in male and female vocalisers. Furthermore, consistent with the substantial influence of strength (and to a lesser degree, height) on women's judgments of males' bodily attractiveness (Sell et al., 2017), our results suggest that mate choice mechanisms played an important role in selection for the communication of formidability.

### **Did stimulus context and type affect ratings of strength and height?**

#### *Strength attributions*

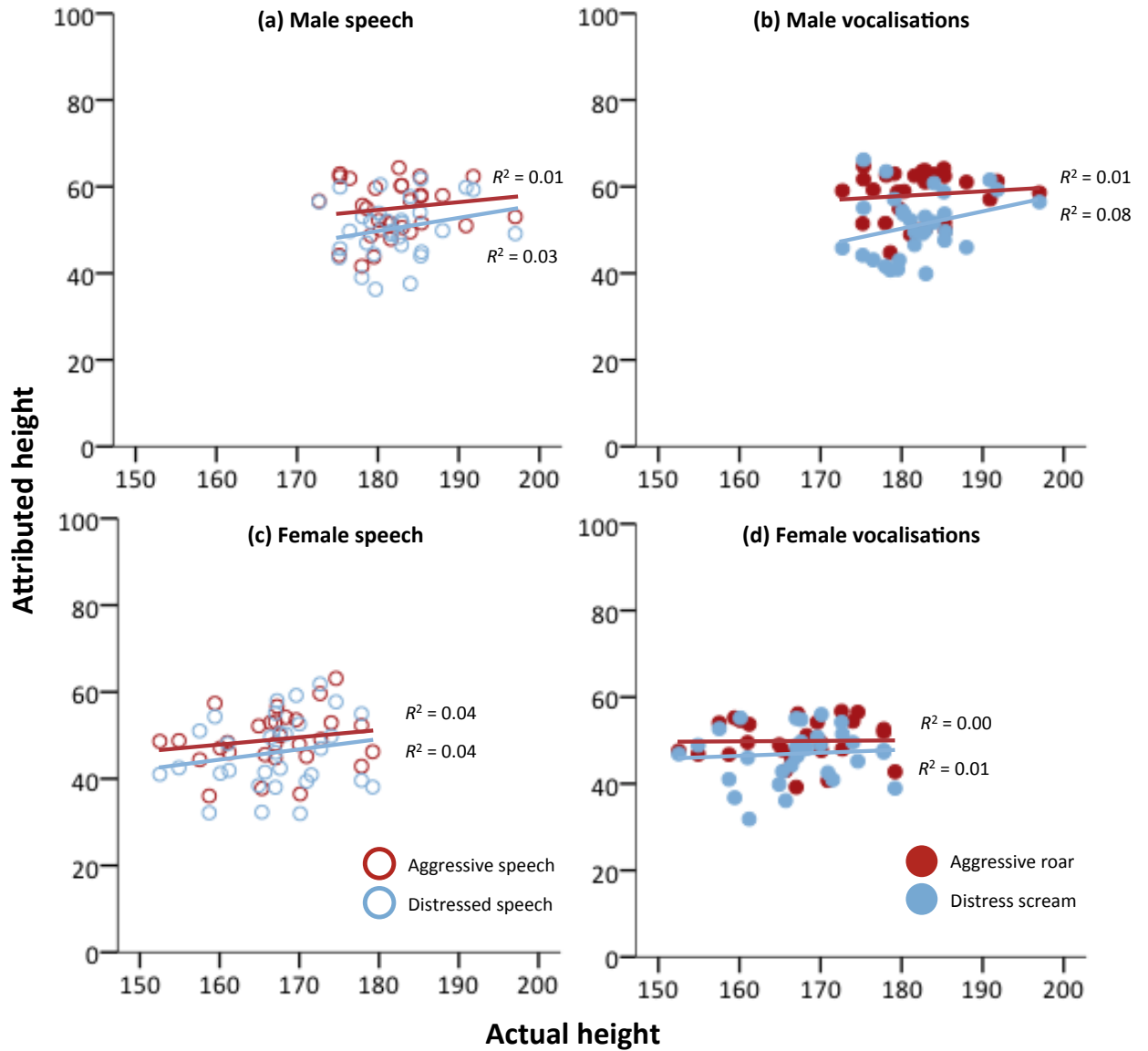
Aggressive stimuli ( $M = 54.15 \pm 0.75$ ) were rated as stronger than distress stimuli ( $M = 37.84 \pm 0.75$ , Figure 3, Table 3iv,  $p < .0005$ ). This difference was significantly larger when listeners rated nonverbal vocalisations ( $M$  difference = 20.31) than when they rated speech sentences ( $M$  difference = 12.31, Figure 3, Table 3xi,  $p < .0005$ ; except when male listeners rated female vocalisers, Table 3xvi,  $p < .001$ ).

The difference in strength ratings between aggressive and distress stimuli was larger when listeners rated vocalisations than when they rated speech because for aggressive stimuli, but not distress stimuli, vocalisations elicited higher strength ratings than did speech sentences ( $M$  difference aggressive = 9.09,  $M$  difference distress = 1.08, Table 3xi,  $p < .0005$ ).



*Figure 3.* Attributed strength as a function of actual strength, when listeners rated (a) male speech stimuli, (b) male vocalisations, (c) female speech stimuli, and (d) female vocalisations. Each data point represents the mean strength rating averaged across listeners attributed to each vocalisation. Blue circles represent distress stimuli, red circles represent aggressive stimuli. Open circles represent speech stimuli, closed circles represent vocalisations.  $R^2$  values for each regression line are reported in the graphs.

Removing the strongest female vocaliser from our analyses did not affect the general pattern or significance of our results.



*Figure 4.* Attributed height as a function of actual height, when listeners rated (a) male speech stimuli, (b) male vocalisations, (c) female speech stimuli, and (d) female vocalisations. Each data point represents the mean height rating averaged across listeners attributed to each vocalisation. Blue circles represent distress stimuli, red circles represent aggressive stimuli. Open circles represent speech stimuli, closed circles represent vocalisations.  $R^2$  values for each regression line are reported in the graphs.

### *Height attributions*

Vocalisers were rated as taller when producing aggressive stimuli than when producing distress stimuli. This was particularly true for male vocalisers ( $M$  difference male vocalisers = 5.44,  $M$  difference female vocalisers = 2.91, Figure 4, Table 4vii,  $p < .001$ ), and by female raters judging vocalisations ( $M$  difference = 5.98,  $M$  difference other stimulus conditions = 3.61, Table 4xv,  $p = .046$ ).

Speech sentences and nonverbal vocalisations generally elicited similar height ratings, except when female listeners rated aggressive stimuli, in which case they rated vocalisers as taller when producing roars ( $M = 56.16 \pm 0.74$ ) than when producing aggressive speech ( $M = 52.75 \pm 0.73$ ,  $M$  difference = 3.41,  $M$  difference other stimulus conditions = 0.48, Table 4xv,  $p = .046$ ).

Our results indicate that roars maximise impressions of strength relative to other vocal stimuli. Listeners attributed higher strength and height ratings to aggressive stimuli (aggressive speech and roars) than to distress stimuli (distress speech and screams), consistent with functional exaggeration of acoustic cues to body size by nonhuman mammals in aggressive contexts (de Boer et al., 2015; Fox, 1970; Hardus et al., 2009; Harris et al., 2006; Reby & McComb, 2003). This difference may be due to acoustic differences between stimuli: aggressive roars were characterised by higher roughness and amplitude than distress screams, as well as a lower F0 and DFF4. This suggests that aggressive roars capitalised on perceptual associations between low frequency sounds and large size, exaggerating perceived formidability relative to distress screams, which likely exploited perceptual associations between high frequencies and small size or submission (as predicted by motivational-structural rules, Ohala, 1984; Owings & Morton, 1998; Rendall et al., 2007; Sell et al., 2010).

In the absence of differences in F0 and DFF4 between aggressive speech and distressed speech, the smaller difference in strength ratings between these speech stimuli (compared to roars and screams) may be attributed to differences in roughness and amplitude, consistent with the observation that both roughness and amplitude consistently predicted listeners' formidability ratings within stimulus conditions (see below).

While distressed speech and distress screams were rated comparably, aggressive vocalisations were consistently rated as expressing greater formidability than aggressive speech (except when male listeners rated height). Thus, in addition to communicating strength more reliably than aggressive speech, roars also conveyed exaggerated formidability more effectively, in accordance with motivational-structural rules.

### **Could listeners estimate strength and height from the voice?**

#### *Strength estimation*

For male vocalisers, actual strength predicted attributed strength only when listeners rated aggressive stimuli (Table 5vi,  $p < .001$ ). For female vocalisers, listeners could estimate strength from aggressive roars, aggressive speech, and distressed speech, but not distress screams (Table 5xii,  $p < .001$ ). Differences in slope gradient between conditions indicate linear differences in listeners sensitivity' to variation in vocaliser strength (as represented by the gamma statistics reported in Table 6, denoting the standardised increase in rated strength per one unit increase in actual strength).  $R^2$  values denote the percentage of variance in mean strength ratings explained by variance in actual strength, and can be interpreted as representing the overall reliability of listeners' strength estimations, adjusted to the linear sensitivity of listeners to variation

in actual strength within each condition. For both male and female vocalisers, the reliability of strength estimation was higher for aggressive roars than for aggressive speech or female distressed speech (Figure 3). Thus, listeners consistently estimated strength from aggressive but not distress stimuli, and estimated strength most reliably from aggressive roars.

Moreover, there was little evidence for listener sex or vocaliser sex differences in the capacity to estimate strength. The only exception was for distressed speech, whereby listeners were more sensitive to variation in actual strength when rating female than male vocalisers.

#### *Height estimation*

For male vocalisers, actual height predicted rated height when listeners rated distress stimuli but not aggressive stimuli (Figure 4, Table 7vi,  $p = .008$ ). For female vocalisers, actual height predicted attributed height when listeners rated speech stimuli but not vocalisations (Figure 4, Table 7vii,  $p = .007$ ). Effect sizes for the relationship between actual and attributed height were much smaller than those for the relationship between actual and attributed strength (Figures 2 and 3).

There were few sex differences in height estimation, except that when rating distress screams, listeners were more sensitive to variation in actual strength when rating male than female vocalisers.

*Table 5.* Strength estimation: Linear mixed models testing the effects of actual strength, stimulus context, stimulus type, and listener sex on rated strength. Separate models are reported for female and male vocalisers.

Source		Females			Males		
		<i>df</i> <sub>1</sub> , <i>df</i> <sub>2</sub>	<i>F</i>	<i>p</i>	<i>df</i> <sub>1</sub> , <i>df</i> <sub>2</sub>	<i>F</i>	<i>p</i>
i.	Intercept	1, 110.71	3159.40	< .001	1, 106.86	2814.58	< .001
ii.	Actual strength	1, 2697.52	162.96	< .001	1, 606.93	55.03	< .001
iii.	Stimulus context	1, 8309.89	706.95	< .001	1, 8063.01	598.41	< .001
iv.	Stimulus type	1, 8309.89	2.70	.100	1, 8063.01	99.14	< .001
v.	Listener sex	1, 110.71	0.21	.651	1, 106.86	0.06	.810
vi.	Strength * stimulus context	1, 8317.01	9.80	.002	1, 8066.40	80.17	< .001
vii.	Strength * stimulus type	1, 8317.01	38.67	< .001	1, 8066.40	2.35	.126
viii.	Strength * listener sex	1, 2697.52	0.42	.515	1, 2606.93	0.05	.826
ix.	Stimulus context * stimulus type	1, 8309.89	77.82	< .001	1, 8063.01	88.97	< .001
x.	Stimulus context * listener sex	1, 8309.89	2.12	.145	1, 8063.01	0.45	.502
xi.	Stimulus type * listener sex	1, 8309.89	1.47	.226	1, 8063.01	0.10	.749
xii.	Strength * stimulus context * stimulus type	1, 8317.01	50.25	< .001	1, 8066.40	1.15	.284
xiii.	Strength * stimulus context * listener sex	1, 8317.01	0.01	.910	1, 8066.40	0.16	.686
xiv.	Strength * stimulus type * listener sex	1, 8317.01	1.72	.190	1, 8066.40	0.04	.851
xv.	Stimulus context * stimulus type * listener sex	1, 8309.89	11.32	.001	1, 8063.01	1.80	.180
xvi.	Strength * stimulus context * stimulus type * listener sex	1, 8317.01	2.20	.138	1, 8066.40	2.41	.120

*Table 6.* Standardised linear mixed model coefficients representing the sensitivity of listeners to variation in vocaliser strength and height. Each coefficient represents the average of listeners' individual slopes for the relationship between actual strength/height and attributed strength/height. Significances represent whether each average slope was significantly different from zero. Separate models are reported for male and female vocalisers.

Source	Females		Males	
	$\gamma$	$p$	$\gamma$	$p$
<b>Strength</b>				
Aggressive speech	.18	< .001	.15	< .001
Distressed speech	.24	< .001	.01	.283
Aggressive roar	.20	< .001	.20	< .001
Distress scream	-.03	.198	.02	.379
<b>Height</b>				
Aggressive speech	.07	.003	.03	.171
Distressed speech	.09	< .001	.05	.021
Aggressive roar	.01	.749	.02	.270
Distress scream	.03	.140	.11	< .001

Table 7. Height estimation: linear mixed models testing the effects of actual height, stimulus context, stimulus type, and listener sex on rated height. Separate models are reported for female and male vocalisers.

	Source	Females			Males		
		<i>df</i> <sub>1</sub> , <i>df</i> <sub>2</sub>	<i>F</i>	<i>p</i>	<i>df</i> <sub>1</sub> , <i>df</i> <sub>2</sub>	<i>F</i>	<i>p</i>
i.	Intercept	1, 1782.63	18.64	< .001	1, 1727.83	6.30	.012
ii.	Actual height	1, 1751.63	13.45	< .001	1, 1713.07	16.08	< .001
iii.	Stimulus context	1, 5286.69	2.15	.143	1, 5154.25	9.29	.002
iv.	Stimulus type	1, 5294.85	7.66	.006	1, 5155.61	0.95	.331
v.	Listener sex	1, 1782.63	.32	.571	1, 1727.83	0.03	.855
vi.	Height * stimulus context	1, 5291.60	1.36	.244	1, 5154.25	6.95	.008
vii.	Height * stimulus type	1, 5294.95	7.38	.007	1, 5155.62	1.24	.265
viii.	Height * listener sex	1, 1751.63	0.09	.761	1, 1713.07	0.00	.956
ix.	Stimulus context * stimulus type	1, 5251.09	0.02	.888	1, 5155.61	2.73	.099
x.	Stimulus context * listener sex	1, 5286.69	.73	.391	1, 5154.25	0.03	.858
xi.	Stimulus type * listener sex	1, 5294.85	1.11	.293	1, 5155.61	0.37	.542
xii.	Height * stimulus context * stimulus type	1, 5251.18	0.02	.897	1, 5155.62	2.44	.118
xiii.	Height * stimulus context * listener sex	1, 5291.60	0.83	.362	1, 5154.25	0.02	.901
xiv.	Height * stimulus type * listener sex	1, 5294.95	0.85	.357	1, 5155.62	0.30	.582
xv.	Stimulus context * stimulus type * listener sex	1, 5251.09	.11	.743	1, 5155.61	0.29	.593
xvi.	Height * stimulus context * stimulus type * listener sex	1, 5251.18	.11	.742	1, 5155.62	0.38	.540

Our results indicate that listeners estimate strength reliably from aggressive voice stimuli, but estimate height poorly and inconsistently. As has been previously reported in neutral speech stimuli (Sell et al., 2010), we observed that listeners were able to detect strength from the voice. However, with the noticeable exception of female distressed speech, this ability was limited only to aggressive stimuli. Thus, aggressively motivated vocal behaviour, whether in the form of speech or nonverbal vocalisations, appears to be optimised to communicate threat potential. These results are consistent with an extensive body of research demonstrating that listeners attend to formidability cues in aggressive calls across a wide range of mammals (giant pandas: Charlton et al., 2010; sea lions: Charrier et al., 2011; fallow deer: Pitcher et al., 2015; red deer: Reby et al., 2005; domestic dogs: Taylor et al., 2010). Moreover, the fact that variation in strength was generally not detected in distress stimuli (with the exception of female distressed speech) indicates that the availability of formidability cues varies with the putative function of the signal, possibly reflecting differential selection on vocalisers to encode formidability cues in aggressive rather than submissive voice signal.

Examination of  $\gamma$  (sensitivity) and  $R^2$  (reliability) values in our analyses indicates that listeners were more sensitive to variation in strength, and estimated strength more reliably, from aggressive roars than aggressive speech. These results accord with evidence that affective information is preferentially decoded from nonverbal vocalisations over emotionally inflected speech (Pell et al., 2015; Scott et al., 1997), suggesting that nonverbal vocalisations may be more effective carriers of indexical cues than speech. While previously reported effect sizes for strength estimation in neutral speech produced by males (Sell et al., 2010) were of similar magnitude to those we observed in our aggressive roars, within-study comparisons are more informative than comparisons between studies, which may be confounded by methodological differences (i.e. differences

in the blocking of stimuli in playback experiments, affecting how listeners use rating scales).

Consistent with the poor accuracy with which listeners estimate men's relative height (Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Rendall et al., 2007), sensitivity to height variation and reliability of height estimation was poor. Given that listeners only discriminate between the taller of two men with above chance accuracy when men's height differences exceed 10 cm, and only reach 90% accuracy when the difference is at least 20 cm, it is unsurprising that in a sample characterised by within-sex height ranges of less than 30 cm, listeners estimated the absolute height of individual speakers unreliably.

Listeners were able to detect a small but significant proportion of variation in height from male and female distressed speech, female aggressive speech, and male distress screams, which were on average characterised by relatively lower F0 (facilitating formant perception through increased spectral density, Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Ryalls & Lieberman, 1982) and less amplitude modulation (minimising the interference of sidebands with formant perception). Thus, when available, listeners may have utilised formant cues to estimate height in these stimuli. However, while in neutral speech formants explain 6 to 10% of the variance in men's and women's heights (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014), in our study, variance in actual height only explained greater than 4% of variance in rated height in one condition, indicating that formant cues to height may not have been systematically available to listeners.

The finding that F0 predicted listeners' height ratings but not actual height suggests that F0 confounded height assessment. Many studies report a consistent perceptual bias in listeners to associate low-F0 speech with larger body size at the within-

sex level (Pisanski & Rendall, 2011; Rendall et al., 2007; D. R. Smith & Patterson, 2005), despite F0 being a very poor predictor of body size when controlling for sex and age (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). We show that this bias, presumably driven by overgeneralisation of sound-size relationships (González, 2006; Rendall et al., 2007) and long thought to interfere with accurate body size estimation (Bruckert, Liénard, Lacroix, Kreutzer, & Leboucher, 2006; Greisbach, 1999; Rendall et al., 2007; but see Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014), extends beyond speech to judgments of nonverbal vocalisations. While it has also been reported that low F0 may elicit higher strength attributions in neutral speech (Sell et al., 2010), we did not corroborate this finding.

Listeners were less sensitive to variation in actual height than strength, and estimated height less reliably. This has previously been reported in male but not female vocalisers, for whom the ability to estimate female strength from speech sentences disappeared when controlling for height and weight (Sell et al., 2010). We also found that, when rating aggressive stimuli and roars in particular, listeners consistently and reliably detected inter-individual variation in strength, but not height. As strength and height were not correlated in the present study, our results provide strong evidence that the human voice contains independent cues to strength and height and that strength cues are more perceptually salient. This finding complements the greater relevance of physical strength than body size to men's fighting behaviour and to perceptions of men's fighting ability (Sell et al., 2009), as well as previous work showing a greater influence of strength than height on female listeners' judgments of male bodily attractiveness (Sell et al., 2017).

Contrary to previous speech research showing that strength and height are more reliably estimated from male than female voices (Rendall et al., 2007; Sell et al., 2010), we found that strength and height were estimated comparably well from male and female

voices. Furthermore, while male listeners have previously been reported to be more sensitive than female listeners to acoustic cues to body size (Charlton, Taylor, et al., 2013; but see Rendall et al., 2007), we did not observe listener sex differences in strength or height estimation.

### **Were ratings of physical characteristics related to acoustic characteristics?**

Mean amplitude was the only acoustic characteristic that consistently predicted physical strength across stimulus categories and sexes (Tables S6 and S7). In addition, vocalisers who were rated as stronger generally produced rougher voice stimuli. Decreases in F0 variability, and increases in amplitude modulation and duration with rated strength were also inconsistently observed (Table S6).

The influence of acoustic characteristics on height ratings was in general much less consistent than for strength ratings (Table S7). In males, louder and lower-pitched stimuli were consistently judged as produced by taller vocalisers. Male roars and screams characterised by higher jitter were also rated as produced by taller vocalisers. No acoustic characteristic consistently predicted height ratings of female vocalisers, but louder aggressive roars and distressed speech were rated as produced by taller vocalisers.

Thus, the acoustic mediators of formidability communication remain unknown. Greater loudness and roughness were consistently associated with higher strength ratings, and greater loudness and lower F0 were often associated with greater attributed height, but these acoustic characteristics did not predict actual strength or height, and thus cannot account for the ability of listeners to reliably estimate strength, and to a lesser degree, height, solely from the acoustic structure of our vocal stimuli. Similarly, while listeners detected strength variation in stimulus conditions for which the dominant frequency within

the expected F4 range negatively correlated with actual strength, DFF4 did not predict listeners' strength ratings. Moreover, listeners also detected strength variation from male aggressive speech and roars despite the absence of acoustic predictors of actual strength for these stimuli. Despite measuring a wide set of acoustic characteristics, our analyses failed to determine what acoustic pathways mediate strength communication, confirming previous observations based on a much smaller set of variables (F0 and formants, Sell et al., 2010).

To summarise, we found that vocalisers conveyed exaggerated formidability in aggressive contexts relative to distress contexts. Moreover, while we did not identify acoustic mechanisms mediating its communication, listeners were able to consistently estimate strength from aggressive but not distress stimuli, and most reliably from aggressive roars. Listeners were also able to estimate height from speech stimuli, but much less reliably than they could estimate strength.

## General Discussion

We compared the acoustic structure of aggressive roars, distress screams, and their speech equivalents, and examined the effectiveness of these various speech stimuli in communicating formidability. We found that the acoustic structure of human aggressive and distress signals, particularly nonverbal vocalisations (roars and screams), varied according to Morton's motivational-structural rules (Morton, 1977). Accordingly, aggressive stimuli exaggerated impressions of strength and body size, relative to distress stimuli and aggressive speech. In addition, while our acoustic analyses failed to identify vocal features mediating the communication of strength and size, listeners could nevertheless accurately estimate strength from male and female aggressive vocal stimuli,

and to a lesser degree, could estimate height from speech stimuli. Aggressive roars conveyed honest inter-individual variation in strength more reliably than did distress screams or speech sentences, and also exaggerated impressions of formidability most effectively.

Our study provides the first evidence that the volitional production of vocal stimuli with either aggressive or submissive motivation effectively maximises or minimises listeners' impressions of the vocaliser's strength and body size. Differences in acoustic structure between these stimuli point towards an exploitation of perceptual biases linking low and harsh voice frequencies to large body size and dominance (Morton, 1977; Ohala, 1984; Owings & Morton, 1998; Pisanski, Fraccaro, Tighe, O'Connor, & Feinberg, 2014; Puts, Hodges, Cárdenas, & Gaulin, 2007; Rendall et al., 2007; Smith & Patterson, 2005).

The ability to exaggerate one's size or strength through vocal production is likely to have conferred an evolutionary advantage, as both larger body size and greater strength are associated with various socioeconomic, competitive, and mating benefits (Brues, 1959; Frederick & Haselton, 2007; Gallup, White, & Gallup, 2007; Judge & Cable, 2004; Monden & Smits, 2009; Pisanski & Feinberg, 2013; Sell et al., 2017). Indeed, in line with our findings, other recent evidence indicates that the capacity to volitionally exaggerate or minimise body size via simulated nonverbal emotional expressions is not limited to actors (R. Jürgens, Grass, Drolet, & Fischer, 2015; Pisanski, Mora, et al., 2016).

Our finding that screams and roars had a particularly large effect on listeners' ratings of strength and height is furthermore consistent with the emerging hypothesis that deceptive voice modulation may be at the origins of selection for humans' uniquely advanced vocal control (Brown, 2017; Oesch, 2016; Pisanski, Cartei, et al., 2016). Indeed, some nonhuman mammals already demonstrate limited capacity for functional vocal deception (see Oesch, 2016) and body size exaggeration (de Boer et al., 2015; Hardus et

al., 2009; Reby & McComb, 2003; see Pisanski, Cartei, et al., 2016) in agonistic contexts, as well as more voluntary vocal flexibility in nonhuman primates (e.g. Lameira et al., 2015; Perlman & Clark, 2015; Schel et al., 2013; see Pisanski, Cartei, et al., 2016 for a review). Survival benefits conferred to those able to modulate the expression of primary indexical cues may have given rise to increasingly greater vocal control, paving the way for the evolution of complex speech capabilities (Oesch, 2016; Pisanski, Cartei, et al., 2016).

However, while the co-optation of primary relationships between acoustic cues and physical attributes may more effectively serve motivational signalling, variation in individuals' capacity to modulate these cues may result in a decoupling between the cues and attributes - as observed in the present study, where we could not identify consistent acoustic correlates of actual height or strength. Despite this, listeners were able to detect variation in these physical traits, suggesting that listeners could detect vocal deception and adjust their judgments accordingly. Evolutionary accounts of vocal signalling contend that in competitive contexts vocalisers should evolve strategies to better manipulate receivers (thus obfuscating indexical information in favour of motivational signalling), while receivers should evolve to detect and resist such manipulation (thus reliably estimating indexical characteristics in spite of deceptive voice modulation) (Knight, 1998; Krebs & Dawkins, 1984; Oesch, 2016). To empirically test this account, acoustic analyses could investigate whether cues to deception are encoded in nonverbal vocalisations, and playback experiments could be employed to assess whether listeners can detect volitional exaggeration or minimisation of body size and strength. Future research may also examine not only whether other nonverbal vocalisations relevant to the signalling of formidability (e.g. martial arts kiaps) communicate indexical cues, but also whether these vocalisations

more reliably communicate motivational state than does speech (e.g. aggression, submission, distress, experienced pain (see Chapter 4)).

It is also possible that cues to formidability in our study may have been communicated by acoustic characteristics that were not captured by our acoustic analyses. For example, information may be contained in the dynamic temporal variation of these vocal parameters; indeed such information is commonly utilised in the construction of model-based emotion recognition from speech (Le & Provost, 2013; Li et al., 2013; Nwe, Foo, & De Silva, 2003). Listeners may also rely on complex linear or nonlinear combinations of acoustic parameters. While analysis of the individual contribution of acoustic characteristics has revealed numerous indexical cues in human and nonhuman mammal vocal behaviour (Taylor, Charlton, & Reby, 2016), future research should utilise alternative acoustic analytical approaches (e.g. linear interactions between acoustic characteristics, deep neural networks, hidden Markov models) to elucidate more complex acoustic mechanisms potentially communicating not only inter-individual variation in strength, but also other functional cues for which linear acoustic analysis has been unable to account (e.g. above-chance discrimination of vocaliser sex from babies' cries, Reby et al., 2016).

## **Conclusion**

We show that listeners can detect variation in vocaliser strength and size from simulated nonverbal and verbal vocal stimuli produced in agonistic contexts (aggression and distress, i.e., contexts in which the communication of formidability is most ecologically relevant). Roars were particularly effective in communicating strength; the lack of linguistic constraints on aggressive roars appears to afford a greater acoustic space

with which to both honestly communicate variation in strength between individuals, and convey exaggerated strength relative to other vocal signals within individuals. These results complement studies examining the vocal communication and exaggeration of formidability and threat in nonhuman mammal species (Charlton et al., 2011; Fitch & Hauser, 2003; Harris et al., 2006; Reby et al., 2005; Reby & McComb, 2003) and add to a growing body of evidence indicating structural and functional homology between human and nonhuman mammal vocalisations (e.g. laughter: Davila-Ross et al., 2010; Pisanski, Cartei, et al., 2016; infant distress vocalisations: Lingle et al., 2012, pain vocalisations: Chapter 4). Nonverbal vocalisations, and the ability to voluntarily produce and modulate them, may constitute a direct intermediary link between involuntary control of stereotyped calls in nonhuman mammals, and full-blown volitional speech in humans (Brown, 2017; Oesch, 2016; Pisanski, Cartei, et al., 2016). As such, further investigation into the structure and function of nonverbal vocalisations may be essential to understanding the origins and evolution of human vocal communication, and its relationship to animal vocal signals.

### **Chapter 3: Human listeners can assess relative formidability from aggressive roars and speech**

#### **Chapter Summary**

The previous chapter demonstrated that aggressive roars, and to a lesser degree, aggressive speech, communicate formidability cues, and that roars serve to exaggerate perceived formidability relative to speech (both aggressive and distressed speech). However, this and all other studies to date focus on listeners' ability to judge individual vocalisers in absolute terms, rather than on the more ecologically relevant capacity to assess the formidability of opponents relative to their own. The main aim of this chapter is to address this crucial shortcoming. More specifically, the following questions will be explored:

**Question 6.** Can listeners estimate the relative strength and height of vocalisers from aggressive roars and aggressive speech?

**Question 7.** Do roars maximise the expression of formidability relative to aggressive speech?

**Question 8.** Are there sex differences (vocaliser and listener) in the communication of relative formidability?

Chapter 3 investigates these questions by exposing listeners of known strength and height to aggressive roars and aggressive speech produced by vocalisers (from Chapter 2), whose strength and height were also measured. In playback experiments, I asked listeners to rate to what degree each vocaliser was stronger/weaker or taller/shorter than them.

Summary of findings:

- Our results provide the first evidence that listeners are able to estimate the formidability of vocalisers relative to their own.
- Male vocalisers were more likely to be perceived as relatively stronger than listeners when producing roars than when producing aggressive speech, consistent with the hypothesis that human roars, like many of their nonhuman analogues, are sexually selected to exaggerate formidability.
- Male and female listeners estimated relative strength with high accuracy, but only male listeners estimated relative height with high accuracy, suggesting that the capacity to assess strength may derive from mate competition and mate choice selection pressures, while size-related information may be of greater importance to males than females.
- Female listeners tended to overestimate the strength of male vocalisers relative to their own, suggesting that women are better physically equipped to defend themselves against male violence than they perceive themselves to be.

## Summary

In nonhuman mammals, vocal cues to body size (a proxy of threat potential) mediate behaviour in agonistic male-male interactions (koalas: Charlton, Whisson, & Reby, 2013; sea lions: Charrier, Ahonen, & Harcourt, 2011; fallow deer: Pitcher, Briefer, & McElligott, 2015; red deer: Reby et al., 2005; domestic dogs: Taylor, Reby, & McComb, 2010). The nonverbal components of human speech also signal physical formidability, but actual strength and height typically explain only a small proportion of variance in listeners' voice-based judgments of absolute height (Charlton, Taylor, & Reby, 2013; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Rendall, Vokey, & Nemeth, 2007) or strength (Chapter 2, Sell et al., 2010), or in their judgments of the relative height of two same-sex vocalisers (e.g. Charlton, Taylor, et al., 2013; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Rendall et al., 2007). No previous study has examined the more ecologically relevant capacity of listeners to assess the formidability of a vocaliser relative to their own. To address this crucial shortcoming, we used playback experiments to investigate whether listeners could estimate the strength and height of vocalisers relative to their own from two ecologically relevant vocal signals: aggressive roars and aggressive speech. The strength of listeners and vocalisers was quantified using a standardised amalgamated measure of flexed bicep circumference and handgrip strength. Our results show that listeners can estimate the relative formidability (strength and height) of vocalisers with high accuracy. In male vocalisers only, roars functioned to exaggerate the expression of threat compared to aggressive speech, as men were more likely to be rated as relatively stronger when producing aggressive roars than aggressive speech. When assessing roars, male listeners accurately identified substantially relatively stronger vocalisers in 88% of trials, and never as weaker. These results indicate

that, as in other mammals, the acoustic structure of human aggressive vocal signals (and in particular roars) may have been selected to communicate functional information relevant to **vocalisers'** survival.

### **Experimental Procedures**

All experiments were approved by the University of Sussex's Life Sciences & Psychology Cluster-based Research Ethics Committee (C-REC) (Certificates of approval: ER/JR307/8, ER/JR307/9) and comply with the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct.

### **Participants**

Vocal stimuli were recorded from 30 male and 31 female ( $M$  age =  $22.79 \pm 1.12$ ) drama or acting students from the Royal Central School of Speech and Drama and the University of Sussex, United Kingdom, who received monetary compensation in exchange for their participation.

We recruited separate samples of participants to provide voice-based assessments of the relative strength and height of vocalisers. The sample rating strength (hereafter Experiment 1) consisted of 19 females and 26 males (age =  $31.44 \pm 8.33$ ) recruited from Tromsø and surrounding rural towns in Norway ( $N = 11$ , all fluent English speakers), and from the University of Sussex, United Kingdom ( $N = 34$ ), in return for prize draw monetary compensations (5 x £20). The sample rating height (hereafter Experiment 2) consisted of 31 females and 25 males (age =  $34.27 \pm 10.39$ ), recruited from the USA using

Amazon Mechanical Turk, and compensated with \$1.75 USD. All participants provided informed consent and completed the experiment online using a custom computer interface. Data from one female and male participant in Experiment 1, and from two female and two male participants in Experiment 2, who did not complete the experiment but rated more than half of the stimuli were included in our analysis.

## **Materials**

### *Vocal stimuli*

Vocalisers were audio recorded producing an aggressive roar and aggressive speech in a quiet, anechoic room, standing 150 cm from a Zoom H4n microphone. A chair was placed at this distance to restrict participants from moving closer to the microphone. Vocalisers were instructed to produce the speech sentence, ‘That’s enough, I’m coming for you!’, followed by a nonverbal vocalisation expressing the same motivation, while imagining themselves in a battle or war scenario, about to charge and attack. This resulted in a total of 122 vocal stimuli.

To obtain realistic vocal stimuli, participants were encouraged to take as much time as they needed to immerse themselves in each imagined context, and to ‘let go of their inhibitions’. Participants were also given the option not to vocalise if they felt that they could not naturally produce the sentence or nonverbal vocalisation, and to repeat any sentence or vocalisation until they were satisfied with their portrayal.

Recordings were saved as WAV files at 44.1 kHz sampling frequency and 16 bits amplitude resolution.

### *Physical formidability measures*

We measured participants' height using metric tape. The average height of our sample (male  $M = 182.03 \pm 0.97$  cm; female  $M = 167.10 \pm 1.19$  cm) compares well with that of the general UK population (male  $M = 175.3$ cm, female  $M = 161.9$  cm, Moody, 2013). Flexed bicep circumference and handgrip strength measurements were aggregated to produce a single, equally weighted, z-scored strength value for each subject (following Sell et al. (2009), Puts et al. (2012), and others). These measures explain approximately 55% and 24% of the variance in strength as measured by weight-lifting machines in male college students, respectively (Sell et al., 2009).

To measure flexed bicep circumference (male  $M = 32.09 \pm 0.60$  cm; female  $M = 28.96 \pm 0.70$  cm), participants were instructed to rest the elbow of their dominant arm on a table while seated, clench their fist, and curl their forearm perpendicular to the table. The experimenter measured the circumference of the bicep at its highest point. A Baseline hydraulic hand dynamometer in its standard use was used to measure the handgrip strength of participants' dominant arm (male  $M = 41.57 \pm 1.36$  kg; female  $M = 26.98 \pm 1.06$  kg). Each strength measure was recorded twice per subject and the highest achievable score, representing greatest strength, was used in analyses.

## Procedure

All playback experiments were completed online on Syntoolkit, a dedicated online testing platform for psychology studies (e.g., Hughes, Gruffydd, Simner & Ward, in press; see Simner & Alvarez, forthcoming) that is particularly suited to running studies with sensory or multisensory stimuli. Listeners were instructed to use headphones and complete the experiment in a quiet place. To allow listeners to complete the experiment at a comfortable but audible volume, they were instructed to first set their volume to its lowest level. Listeners then heard a demo sound file (amalgamating a loud and quiet stimulus), and were instructed to raise their volume until they could clearly hear the quiet stimulus,

while the louder stimulus did not cause discomfort. Following this, listeners were asked not to adjust the volume settings during the experiment unless it became too uncomfortable, and were asked at the end of the experiment if they had done so. Due to the agonistic nature of the stimuli, listeners were made aware that if they felt uncomfortable or distressed listening to the sounds, they could stop the experiment.

In playback experiments, vocal stimuli ( $n = 122$ ) were blocked by sex and stimulus type (speech/roar). The order of blocks and stimuli within blocks was randomised. Before each block, participants were reminded to listen to each stimulus in full, and informed that they could take a break at any time. Listeners rated the physical strength (Experiment 1) or height (Experiment 2) of each voice stimulus (“Rate by how much this person is stronger/taller or weaker/shorter than you”) on a 101-point scale from -50 (much weaker/shorter) to 50 (much stronger/taller). We set the slider’s default position to 0 (described as ‘same as you’) and did not compel listeners to move the slider so as not to artificially force directional judgments.

Listeners were debriefed upon completion that the roars and screams were acted, and that the vocalisers were not really experiencing aggression or distress. We examined reaction times against stimulus durations to ensure that participants completed the experiments properly. No participants were removed as a result of this process.

To assess whether listeners could accurately judge the physical characteristics of vocalisers relative to their own, we measured listeners’ own physical characteristics. In Experiment 1, we used a tailor’s tape measure to measure bicep circumference (male  $M = 33.89 \pm 0.46$  cm; female  $M = 28.12 \pm 0.57$  cm), and a Takei hand dynamometer to measure handgrip strength (male  $M = 46.11 \pm 1.67$  kg; female  $M = 33.03 \pm 1.10$  kg), in identical fashion to measurements taken from vocalisers. Both vocaliser and listener strength z-scores were calculated based on a pooled sample of the listeners and the

vocalisers' measurements. Experiment 2 relied on a self-report measure of height given at the start of the playback experiment (male  $M = 176.38 \pm 1.30$  cm; female  $M = 169.36 \pm 1.48$  cm). The validity of self-report measures of height has been extensively studied, and despite slight overestimations, self-reported height closely reflects measured height within the age range of our sample of listeners (Krul, Daanen, & Choi, 2011; Lim, Seubsman, & Sleigh, 2009; Parker, Dillard, & Phillips, 1994; Wada et al., 2005).

### **Coding and statistical analysis**

To examine strength/height estimation in functionally relevant terms, we divided the actual difference in strength/height into five categories. In Experiment 1, percentage differences between -10% and 10% were coded as 'similar strength', differences between  $\pm 10\%$  and  $\pm 30\%$  were coded as 'vocaliser is stronger (weaker) than listener', and differences greater than  $\pm 30\%$  were coded as 'vocaliser is much stronger (weaker) than listener'. In Experiment 2, we calculated by how many centimetres the vocaliser was taller than the listener. Values were coded into identical categories of 11 cm intervals. This interval was chosen as it produced a similar distribution to that observed for our actual strength difference categories.

In both experiments, we coded the rated difference in strength/height between listener and vocaliser into three categories. Ratings between 45 and 55 were categorised as 'rated as similar strength', and ratings above (below) this range were coded as 'vocaliser rated as stronger (weaker)'. We computed a linear mixed multinomial logistic regression, testing the effects of the actual strength difference between listener and vocaliser, vocaliser sex, listener sex, and stimulus type on the rated difference between listener and vocaliser, excluding actual difference categories with sample sizes less than 15. In all models, we included listener identity as a subject variable, and vocaliser identity as a

random factor, thus allowing the intercepts and slopes of the relationships between predictors and outcomes to vary between both vocalisers and listeners and testing null hypotheses based on the average of these intercepts and slopes.

## Results and Discussion

Strength did not correlate with height among either male ( $r = -.04, p = .833$ ) or female ( $r = .083, p = .655$ ) vocalisers. Therefore, at least in our sample, these two physical measurements appear to characterise distinct aspects of physical formidability.

### Judgments of relative strength

A linear mixed multinomial logistic regression, with the actual strength difference between vocaliser and listener, vocaliser sex, listener sex, and stimulus type as predictors, and rated relative strength difference as a categorical outcome variable, showed that overall, the actual strength difference was a significant predictor of the perceived strength difference (Table 1ii). Relatively stronger vocalisers were rated as relatively stronger, and vice versa (Figure 1). This demonstrates that listeners of both sexes are capable of making accurate functional judgments of the strength of other men and women, relative to their own, from both verbal and nonverbal vocal stimuli.

The model showed a significant main effect of vocaliser sex (Table 1iii), with male vocalisers overall more likely to be judged as relatively stronger than females, and vice versa, independently of the actual strength difference between vocaliser and listener (Figure 1). The main effects of vocaliser sex and actual strength difference interacted significantly (Table 1vi), with listeners more likely to judge relatively weaker males, but relatively stronger females, as of similar strength to themselves than relatively stronger males or weaker females (Figure 1). We also observed a significant interaction between listener sex and actual strength difference (Table 1vii). Female listeners were more likely

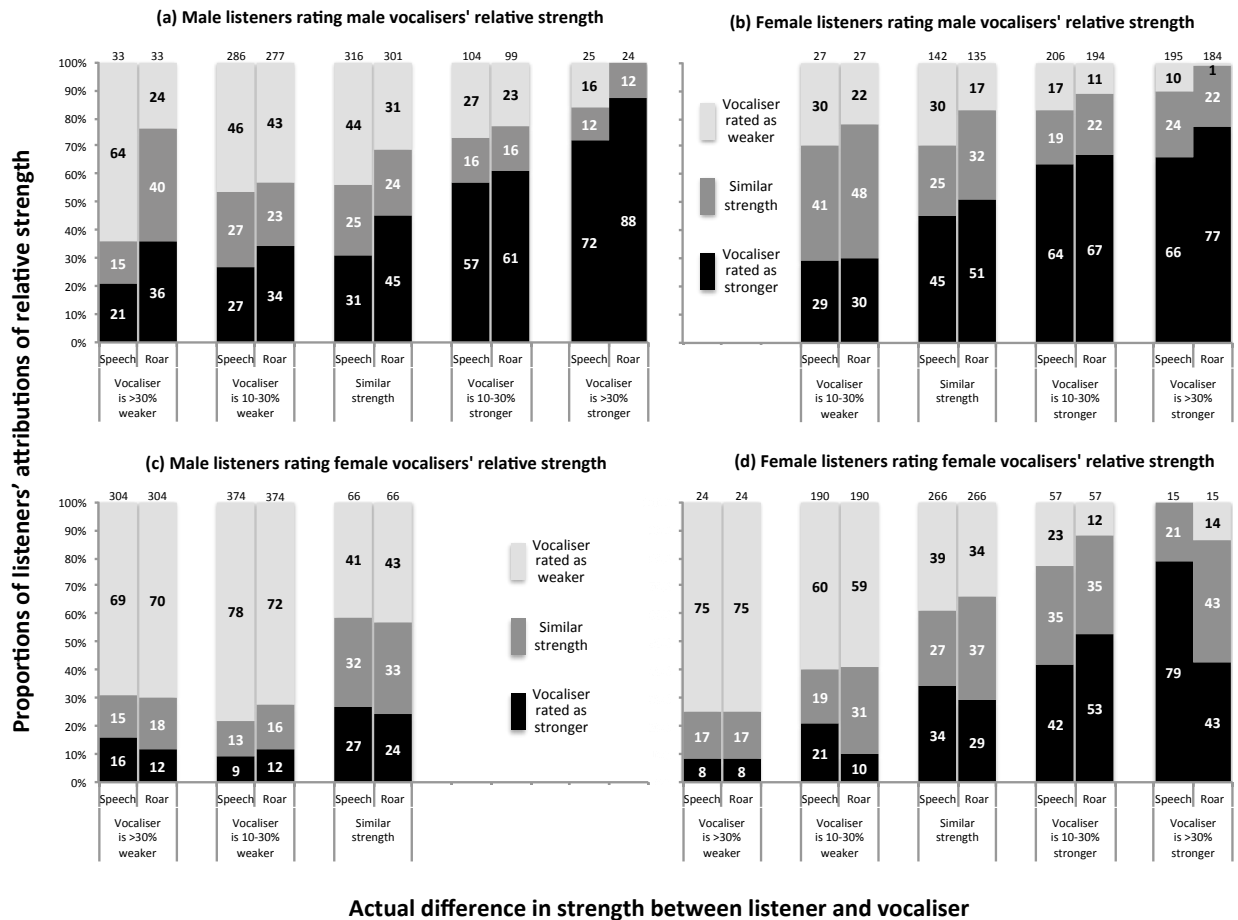
to judge vocalisers as stronger or of similar strength to themselves than were male listeners, except when the vocaliser was much weaker or much stronger (Figure 3).

The combined effects of vocaliser sex and listener sex resulted in a tendency for male listeners to underestimate the relative strength of female vocalisers (Figure 1a & c), and for female listeners to overestimate the relative strength of male vocalisers (Figure 1b & d). The significant interaction between listener sex and vocaliser sex (Table 1ix) indicated that female listeners overestimated male vocalisers more than expected from the combined main effects (Figure 1). Together, these results suggest that listeners, particularly females, may overgeneralise population-level sex differences in strength (Chapter 2; Bishop, Cureton, & Collins, 1987; see Lassek & Gaulin, 2009 for a review). Such overgeneralisations are common in human perception of nonverbal vocal cues (Reby, Levréro, Gustafsson, & Mathevon, 2016; Rendall et al., 2007), and are likely to reflect stereotypical biases. The stronger bias among female than male listeners is consistent with previous indications that women perceive gender differences to be larger than do men, across a wide range of psychological traits (Zell et al., 2016).

Finally, the model revealed a significant main effect of stimulus type, showing that overall, listeners were more likely to rate vocalisers as stronger or of similar strength when judging roars compared to speech. A significant interaction with vocaliser sex (Table 1x) indicated that this was only the case when listeners rated male vocalisers (Figure 1). Furthermore, a three-way interaction between stimulus type, vocaliser sex, and actual strength difference indicated that this effect was strongest when male vocalisers were much weaker than male listeners (Table 1xiii, Figure 1). This suggests that while male roars increase the perceived difference in strength between listeners and vocalisers, compared to aggressive speech, this difference is particularly functional in the weakest male vocalisers.

*Table 1.* Mixed multinomial logistic regression testing the effects of the categorised actual difference in strength between listener and vocaliser, vocaliser sex, listener sex, and stimulus type on whether the vocaliser was rated as stronger, weaker, or of similar strength to the listener.

<b>Source</b>		<b><i>df</i><sub>1</sub>, <i>df</i><sub>2</sub></b>	<b><i>F</i></b>	<b><i>p</i></b>
i.	Intercept	33, 5135	23.37	<b>&lt; .001</b>
ii.	Actual strength difference	4, 5135	19.03	<b>&lt; .001</b>
iii.	Vocaliser sex	1, 5135	78.59	<b>&lt; .001</b>
iv.	Listener sex	1, 5135	3.73	.054
v.	Stimulus type	1, 5135	4.91	<b>.027</b>
vi.	Actual strength difference * vocaliser sex	4, 5135	3.25	<b>.011</b>
vii.	Actual strength difference * listener sex	4, 5135	2.97	<b>.018</b>
viii.	Actual strength difference * stimulus type	4, 5135	0.52	.720
ix.	Vocaliser sex * listener sex	1, 5135	4.21	<b>.040</b>
x.	Vocaliser sex * stimulus type	1, 5135	14.91	<b>&lt; .001</b>
xi.	Listener sex * stimulus type	1, 5135	0.56	.453
xii.	Strength difference * vocaliser sex * listener sex	1, 5135	0.67	.412
xiii.	Strength difference * vocaliser sex * stimulus type	4, 5135	3.60	<b>.006</b>
xiv.	Strength difference * listener sex * stimulus type	4, 5135	0.37	.832
xv.	Vocaliser sex * listener sex * stimulus type	1, 5135	0.01	.932
xvi.	Strength diff * vocaliser sex * listener sex * stimulus type	1, 5135	1.30	.255



*Figure 1.* Percentage of listeners judging vocalisers as relatively weaker (black), of similar strength (dark grey), or as relatively stronger (light grey) than themselves, as a function of the actual difference in strength between listener and vocaliser. Separate graphs are reported for (a) male listeners rating male vocalisers, (b) female listeners rating male vocalisers, (c) male listeners rating female vocalisers, and (d) female listeners rating female vocalisers. Within each panel, for each actual strength difference category, separate bars are reported for listeners rating aggressive speech (left) and aggressive roars (right).

### Judgments of relative height

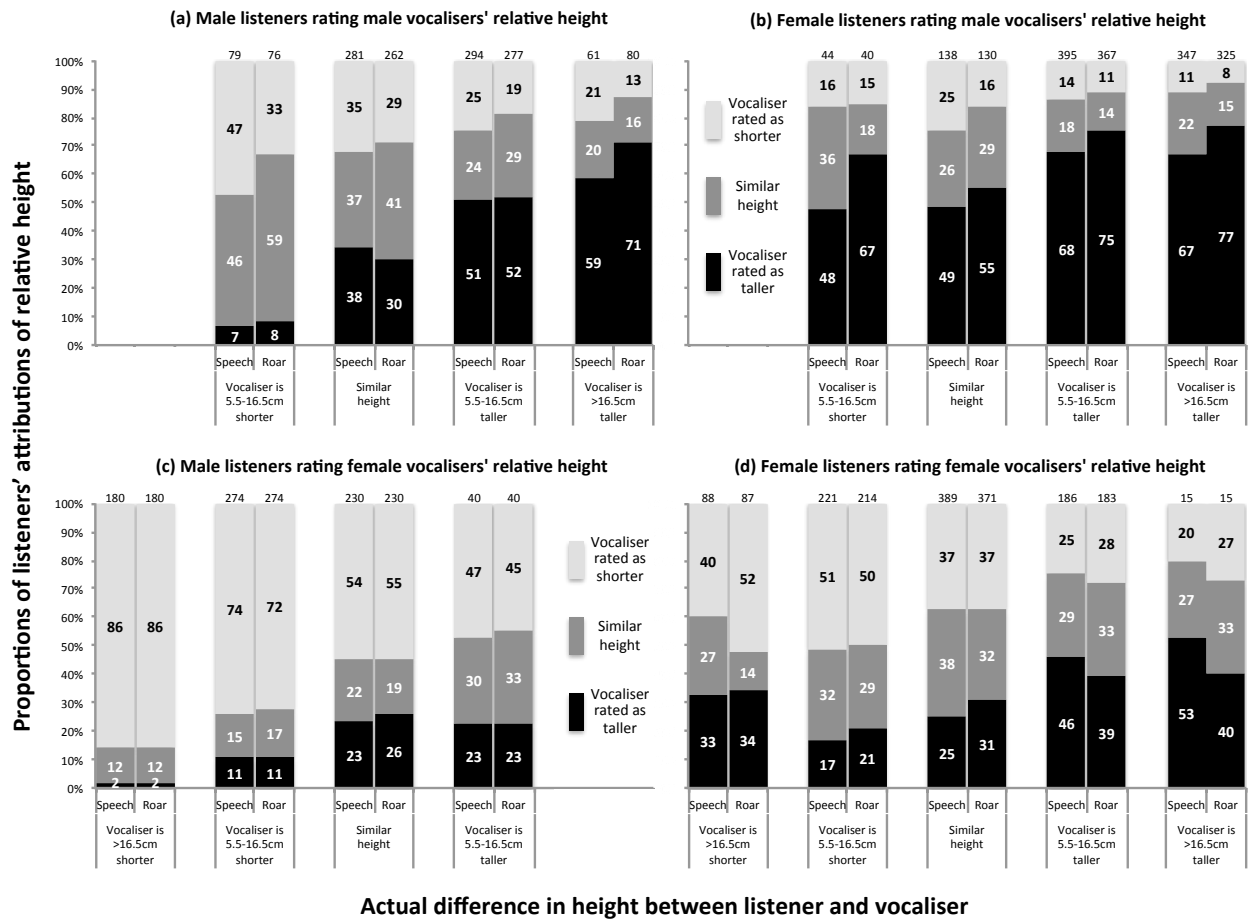
A second linear mixed multinomial logistic regression, with the actual height difference between vocaliser and listener, vocaliser sex, listener sex, and stimulus type as predictors, and rated height difference as a categorical outcome variable, showed that overall, the actual height difference was a significant predictor of the perceived strength

difference (Table 2ii). Relatively taller vocalisers were rated as relatively taller, and vice versa (Figure 2). This demonstrates that listeners of both sexes can judge the body size of other men and women, relative to their own, from both verbal and nonverbal stimuli.

This effect was qualified by an interaction with listener sex, whereby male listeners were more sensitive to relative size variation than were female listeners: as actual size differences increased, male listeners were increasingly more likely to rate the vocaliser as relatively taller than were female listeners. These findings support the hypothesis that size assessment abilities may have arisen primarily through male-male competition, and are consistent with previous observations that men are better than women at estimating body size from synthesised vocal stimuli (Charlton, Taylor, et al., 2013). A significant three-way interaction between actual height difference, listener sex, and vocaliser sex indicated that the effect of actual height difference was minimal when female listeners rated female vocalisers (Figure 2d). This is consistent with evidence that male body size plays a role in female mate choice (Bruckert, Liénard, Lacroix, Kreutzer, & Leboucher, 2006; Sell, Lukazsweski, & Townsley, 2017).

The model showed a significant main effect of vocaliser sex (Table 2iii), with male vocalisers more likely to be judged as taller relative to the listener than females, and vice versa, independently of the actual height difference between vocaliser and listener (Figure 2). The main effect of listener sex was also significant (Table 2iv), showing that female listeners were generally more likely to judge vocalisers as relatively taller or of similar height to themselves than were male listeners (Figure 2). Thus, as with strength, male listeners tended to underestimate the relative height of female vocalisers (Figure 2a & c), and female listeners tended to overestimate the relative height of male vocalisers (Figure 2b & d). This suggests that sexual dimorphism in actual height in adult humans (i.e. men are approximately 7% to 10% taller than women, Pisanski, Fraccaro, Tigue, O'Connor,

Röder, et al., 2014) may induce disproportionate sex-dependent biases in listeners' relative height judgments.



*Figure 2.* Percentage of listeners judging vocalisers as relatively shorter (black), of similar height (dark grey), or as relatively taller (light grey) than themselves, as a function of the actual difference in height between listener and vocaliser. Separate graphs are reported for (a) male listeners rating male vocalisers, (b) female listeners rating male vocalisers, (c) male listeners rating female vocalisers, and (d) female listeners rating female vocalisers. Within each panel, for each actual height difference category, separate bars are reported for listeners rating aggressive speech (left) and aggressive roars (right).

*Table 2.* Mixed multinomial logistic regression testing the effects of the categorised actual difference in height between listener and vocaliser, vocaliser sex, listener sex, and stimulus type on whether the vocaliser was rated as taller, shorter, or of similar height to the listener.

<b>Source</b>		<b><i>df</i><sub>1</sub>, <i>df</i><sub>2</sub></b>	<b><i>F</i></b>	<b><i>p</i></b>
i.	Intercept	33, 6738	31.51	< . <b>.001</b>
ii.	Actual height difference	4, 6738	5.26	< . <b>.001</b>
iii.	Vocaliser sex	1, 6738	193.37	< . <b>.001</b>
iv.	Listener sex	1, 6738	25.43	< . <b>.001</b>
v.	Stimulus type	1, 6738	3.62	.057
vi.	Actual height difference * vocaliser sex	3, 6738	0.60	.616
vii.	Actual height difference * listener sex	4, 6738	3.47	<b>.008</b>
viii.	Actual height difference * stimulus type	4, 6738	0.50	.735
ix.	Vocaliser sex * listener sex	1, 6738	0.60	.438
x.	Vocaliser sex * stimulus type	1, 6738	6.01	<b>.014</b>
xi.	Listener sex * stimulus type	1, 6738	0.01	.951
xii.	Height difference * vocaliser sex * listener sex	2, 6738	4.24	<b>.014</b>
xiii.	Height difference * vocaliser sex * stimulus type	3, 6738	0.34	.794
xiv.	Height difference * listener sex * stimulus type	4, 6738	0.32	.865
xv.	Vocaliser sex * listener sex * stimulus type	1, 6738	1.21	.272
xvi.	Height diff * vocaliser sex * listener sex * stimulus type	2, 6738	0.33	.722

Lastly, the interaction between stimulus type and vocaliser sex was significant (Table 2x), with listeners more likely to rate male vocalisers (but not female vocalisers) as stronger or of similar strength to themselves when judging roars than speech (Figure 2). This is consistent with the hypothesis that roars serve to exaggerate physical formidability, as already observed in nonhuman mammals (Charlton et al., 2011; Harris, Fitch, Goldstein, & Fashing, 2006; Reby & McComb, 2003) and humans (Chapter 2).

Investigations of humans' capacity to estimate formidability from the voice have exclusively focused on absolute judgments (e.g. Bruckert et al., 2006; Sell et al., 2010; Smith & Patterson, 2005) or comparison of pairs of vocalisers (e.g. Charlton, Taylor, et al., 2013; Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Rendall et al., 2007). Our results provide the first evidence that listeners are able to estimate the formidability of vocalisers relative to their own, a judgment more closely aligned with the hypothesised central role of mate competition in selecting for the communication of formidability (Hill et al., 2013; Hill, Bailey, & Puts, 2017).

Indeed, while previous studies typically report that strength and height explain relatively modest proportions of variance in listeners' formidability judgements, we show that both male and female listeners can use available formidability cues to make ecologically relevant judgments ecologically relevant to competitive interactions with high accuracy. For example, listeners erroneously judged relatively stronger vocalisers as weaker in only 18% of cases, and substantially stronger vocalisers as weaker in only 6% of cases. The finding that female listeners estimated strength (but not height) with high accuracy adds to a small but growing body of evidence suggesting that the capacity to assess strength may not only derive from sexual selection for mate competition, but also

from female mate choice, with body size being of less importance than strength to females' perceptions of males' attractiveness (Sell et al., 2017).

Given that in many nonhuman mammals acoustic cues to formidability mediate dyadic agonistic interactions between competing males (koalas: Charlton, Whisson, et al., 2013; sea lions: Charrier et al., 2011; fallow deer: Pitcher et al., 2015; red deer: Reby et al., 2005; domestic dogs: Taylor et al., 2010), it is assumed that nonhuman mammals are also able to assess opponents' formidability relative to their own. To empirically verify this assumption, future research should now examine how between-individual variation in the formidability of nonhuman receivers mediates vocal behaviour (e.g. call response latency, calling rate, Charlton, Whisson, et al., 2013; Reby et al., 2005).

Male vocalisers were more likely to be perceived as stronger relative to listeners when producing roars than aggressive speech. This effect was more pronounced when strength differences were extreme, with listeners almost never (less than 1% of cases) rating substantially stronger male vocalisers as weaker when judging roars. In turn, male listeners correctly identified substantially weaker vocalisers as weaker on only 24% of trials when judging roars. This exaggerative function of roars is consistent with the conformity of their acoustic structure to motivational-structural rules (Morton, 1977, see Chapter 2). More generally, our results support the hypothesis that human roars, like many of their nonhuman analogues, are sexually selected to exaggerate formidability in male-male competitive interactions (Charlton et al., 2011; Harris et al., 2006; Reby & McComb, 2003), but may also afford advantages to males in mate choice contexts (Charlton, Ellis, Brumm, Nilsson, & Fitch, 2012; Charlton, Reby, & McComb, 2007b, 2007a), likely as a result of resource holding potential benefits conferred by greater formidability (Brues, 1959; Frederick & Haselton, 2007; Gallup et al., 2007; Judge & Cable, 2004; Monden & Smits, 2009; Pisanski & Feinberg, 2013; Sell et al., 2017).

The observation that women were more likely to rate vocalisers as relatively stronger than were men at the same actual difference in strength is consistent with a general tendency for women to underestimate, and for men to overestimate, their skills and abilities (Bleidorn et al., 2016; Ehrlinger & Dunning, 2003; Erkut, 1983; Freund & Kasten, 2012; Gold, Brush, & Sprotzer, 1980; Kosakowska-Berezecka, Jurek, Besta, & Badowska, 2017; Syzmanowicz & Furnham, 2011). Of particular interest is that women correctly identified relatively weaker male vocalisers on only 25% of trials, and tended to judge similar strength male vocalisers as stronger than themselves. Our findings suggest that women may be better physically equipped to defend themselves against violence than they perceive themselves to be. As such, confidence-based interventions (already shown to ameliorate the 'confidence gap' in cognitive tasks, Bench, Lench, Liew, Miner, & Flores, 2015; Ehrlinger & Dunning, 2003; Estes & Felker, 2012) specifically targeting this negative bias are likely to improve the efficacy of, for instance, sexual assault resistance programmes (Jordan & Mossman, 2017; Senn et al., 2015, 2017; Wong & Balemba, 2016).

## Chapter 4: Vocal communication of simulated pain

### Chapter Summary

The previous chapters established that volitionally produced roars honestly communicate indexical information to listeners at a between-individual level, but also that vocalisers can modulate (exaggerate or minimise) the expression of such cues in accordance with motivation (aggression or distress). The present chapter explores the ability of trained actors to communicate dynamic motivational cues to pain in the complete absence of an associated genuine experience of actual pain. More specifically, the following question will be explored:

**Question 9.** Does the acoustic encoding and perception of pain levels in adult simulated pain vocalisations follow similar patterns to those observed in human infant and nonhuman mammal pain cries?

Chapter 4 investigates Question 9 by asking trained actors to produce pain vocalisations in three simulated contexts of increasing pain intensity, and examining how pain levels are encoded in their acoustic structure. I also perform playback experiments to examine whether these vocalisations successfully communicate pain intensity to listeners, and which acoustic characteristics are responsible for variation in pain ratings.

Summary of findings:

- The acoustic structure of pain cries changes systematically with simulated level of intensity. The mean and range of F0, the amplitude of the vocalisation, the harmonicity of the voiced proportion of the vocalisations,

and the proportion of the signal displaying nonlinear phenomena all increased with the level of pain. These parameters also predicted increases in listeners' ratings of pain intensity.

- This pattern is largely consistent with acoustic mediators of pain communication observed in authentic human infant and nonhuman mammal pain vocalisations. This suggests that vocal pain exaggeration or simulation may be an adaptive survival-enhancing strategy, and supports the hypothesis that volitional modulation of nonverbal vocalisations may have been at the origins of selection for increased vocal control.
- While different voice features contributed to increases in pain ratings within each level of expressed pain, a combination of these features most strongly and reliably predicted listeners' pain ratings both across and within levels. These features also highly reliably classified vocalisations according to pain intensity level.
- Our analyses highlight potential for the development of a practical quantitative tool to improve pain assessment in populations unable to self-report their subjective pain experience.

**Abstract**

While evidence suggests that pain cries produced by human babies and other mammal infants communicate acoustic cues to pain intensity, whether the pain vocalisations of human adults also encode pain intensity, and which acoustic characteristics influence listeners' perceptions remain unexplored. Here, we investigated how trained actors communicated pain by contrasting the acoustic characteristics of nonverbal vocalisations expressing different levels of pain intensity (mild, moderate, and severe). We then performed playback experiments to examine whether vocalisers successfully communicated pain intensity to listeners, and which acoustic characteristics were responsible for variation in pain ratings. We found that the mean and range of voice fundamental frequency (F0, perceived as pitch), the amplitude of the vocalisation, the harmonicity of the voiced proportion of the vocalisations, and the proportion of the signal displaying nonlinear phenomena all increased with the level of simulated pain intensity. In turn, these parameters predicted increases in listeners' ratings of pain intensity. We also found that while different voice features contributed to increases in pain ratings within each level of expressed pain, a combination of these features explained an impressive amount of the variance in listeners' pain ratings across (76%) and within (31-54%) pain levels. Our acoustic analyses highlight potential for the development of a practical quantitative tool to improve pain assessment in populations unable to self-report their subjective pain experience.

## Introduction

Mammal vocal signals communicate key indexical information that is relevant in social and competitive contexts (Briefer, 2012; A. M. Taylor, Charlton, & Reby, 2016), and highly conserved across species, (Owren, 2011; A. M. Taylor et al., 2016) including humans (e.g. Feinberg, Jones, Little, Burt, & Perrett, 2005; Koutseff et al., 2017; Puts, Hodges, Cárdenas, & Gaulin, 2007; Rendall, Vokey, & Nemeth, 2007; Sell et al., 2010). When experiencing pain, human infants (Bellieni, 2012; Levine & Gordon, 1982), human adults (Baker & Kenner, 1993; Fuller, Roberts, & McKay, 1993) and many nonhuman mammals (Bars, Gozariu, & Cadden, 2001; Bufalari, Adami, Angeli, & Short, 2007; Dubner, 1994; Mogil, 2009) produce pain vocalisations in response to noxious stimuli, that are in turn detected and processed via similar mechanisms in humans and nonhuman mammals (Schnitzler & Ploner, 2000; Tracey & Mantyh, 2007; X. J. Zhang, Zhang, Hu, & Xu, 2011). Vocal communication of pain is likely to provide survival advantages to signallers by attracting attention, aid, or protection (Craig, 2009; Levine & Gordon, 1982; Sullivan, 2008; Williams, 2002), and may also be advantageous to friendly receivers (warning of threat and danger, Craig, 2009; Sullivan, 2008; Williams, 2002). As such, pain vocalisations are likely to have been selected to communicate honest cues to pain levels in their acoustic structure (Hadjistavropoulos & Craig, 2002).

While multiple studies have reported differences in the occurrence, acoustic characteristics and perceptual characteristics of human infant and nonhuman mammal cries produced in response to pain versus other stressors, such as hunger or isolation (e.g. Boero, Bianchi, Volpe, Marcello, & Lenti, 1998; Calvino, Besson, Boehrer, & Depaulis, 1996; Fuller, 1991; Lindová, Špinka, & Nováková, 2015; Marx, Horn, Thielebein, Knubel, & von Borell, 2003; Watts & Stookey, 1999; Weary, Braithwaite, & Fraser,

1998), ethical considerations limit the degree to which the graded acoustic communication of pain intensity can be investigated. Thus, most research in this area takes advantage of painful procedures already performed for purposes other than scientific investigation (e.g. medical: Facchini, Bellieni, Marchettini, Pulselli, & Tiezzi, 2005; Koutseff et al., 2017; agricultural: Puppe, Schön, Tuchscherer, & Manteuffel, 2005; White et al., 1995).

As human infants experience increases in pain, they produce cries with higher levels of roughness (irregular/chaotic vocal fold vibration) (Facchini et al., 2005; Koutseff et al., 2017; Tiezzi et al., 2004), higher amplitude (Fuller & Conner, 1995; Lehr et al., 2007; c.f. Maitre et al., 2017), lower variation in amplitude (Bellieni et al., 2004), longer bout duration (Johnston & O'Shaughnessy, 1987; Porter et al., 1986), and a more variable fundamental frequency, F0 (Koutseff et al., 2017; Porter et al., 1986). Mean F0 (perceived as pitch) appears not to correlate linearly with pain levels in infant cries (Johnston & O'Shaughnessy, 1987; Koutseff et al., 2017; Silva et al., 2010; c.f. Porter et al., 1986), but rather increases abruptly after a certain threshold of high pain is reached ('alarm threshold', Bellieni et al., 2004).

Acoustic cues to pain in nonhuman mammals have received relatively little consideration, and research has tended to focus disproportionately on calling rate (e.g. Kurejova et al., 2010; A. A. Taylor & Weary, 2000) rather than on variation in the spectral characteristics of calls. However, several studies have shown that call duration, amplitude and acoustic nonlinearities (irregular vocal fold vibration regimes, Fitch, Neubauer, & Herzel, 2002) increase with the intensity of electrical stimulation in mice (Eschali r et al., 1988; Jourdan et al., 1995; Levine et al., 1984). In pigs, more painful castration procedures also induce vocalisations with higher peak frequencies (White et al., 1995), indicating cross-specific commonalities in the acoustic gradation of pain intensity.

Acoustic correlates of pain also co-vary with arousal in human speech (see Briefer, 2012), nonhuman mammal vocalisations (see Blumstein & Chi, 2012; Briefer, 2012), and human nonverbal vocalisations (Lima, Castro, & Scott, 2013; Nwokah, Davies, Islam, Hsu, & Fogel, 1993; Sauter, Eisner, Calder, & Scott, 2010; Szameitat, Darwin, Wildgruber, Alter, & Szameitat, 2011). This is because activation of the autonomic nervous system – which occurs when experiencing either pain (Benarroch, 2006) or arousal (Briefer, 2012) – affects respiratory and phonatory aspects of voice production (Briefer, 2012). Indeed, pain cries are assumed to exhibit higher F0 compared to distress cries caused by other stressors because they reflect a more highly aroused state (Boero et al., 1998; Fuller & Horii, 1986, 1988; Grunau, Johnston, & Craig, 1990; Gustafson & Harris, 1990; Johnston & O'Shaughnessy, 1987; Lingle, Wyman, Kotrba, Teichroeb, & Romanow, 2012).

Human listeners are able to distinguish infant pain cries from distress cries produced in response to other stressors (Gustafson & Harris, 1990; Koutseff et al., 2017; Porter et al., 1986; but see Lindová et al., 2015), and can discriminate more invasive from less invasive surgical circumcision procedures (Porter et al., 1986). However, listeners cannot reliably distinguish between pain levels elicited by different vaccines, even though acoustic analyses reveal that more painful vaccines elicit cries with greater roughness (Koutseff et al., 2017). The few studies investigating perception of pain intensity suggest that higher-pitched (Craig et al., 1988; Porter et al., 1986), louder, and noisier (Porter et al., 1986) cries tend to be judged as more painful or urgent. In distress cries associated with other stressors (e.g. hunger, isolation), increased F0, F0 variability, duration, and roughness predict humans' perceptions of the urgency or level of distress experienced by human infants (Dessureau, Kurowski, & Thompson, 1998; Esposito, Nakazawa, Venuti, & Bornstein, 2012, 2015; Out, Pieper, Bakermans-Kranenburg, Zeskind, & van IJzendoorn,

2010; Reby, Levréro, Gustafsson, & Mathevon, 2016; Wood, 2009; Zanchi et al., 2016; Zeifman, 2004; see LaGasse, Neal, & Lester, 2005 for review) and infants of other primate species (F0 only, Kelly et al., 2017). Similarly, noisier (rhesus macaques: Jovanovic & Gouzoules, 2001) and higher frequency (pigs: Weary, Lawson, & Thompson, 1996) cries provoke more urgent responses in other mammals.

While in adult humans, the experience of pain can be reported verbally (X. J. Zhang et al., 2011), pain is also frequently expressed with nonverbal cries or screams, for example as a consequence of high-intensity pain (e.g. during childbirth, Fuller et al., 1993). Vocalisations are also considered valuable indicators of pain in groups unable to submit reliable self-reports regarding their subjective pain experience, such as older adults with advanced dementia, persons with intellectual disabilities, and patients at the end of life (Carter, McArthur, & Cunliffe, 2002; Herr, Coyne, McCaffery, Manworren, & Merkel, 2011; McGrath, Rosmus, Canfield, Campbell, & Hennigar, 1998; van Iersel, Timmerman, & Mullie, 2006). However, the acoustic structure of adult nonverbal pain vocalisations, and their effects on listeners' perceptions, have not yet been systematically investigated.

Here, to investigate the communication of pain in adult human pain vocalisations, trained actors were asked to produce pain vocalisations in three simulated contexts of increasing pain intensity. Using acoustic analysis, we examined how simulated pain levels were encoded in the acoustic structure of these vocalisations. We then asked listeners to rate the pain levels experienced by the vocalisers, to test whether listeners correctly judged higher-intensity pain vocalisations as expressing more pain, and which acoustic characteristics affected their judgments. Given the apparent evolutionary continuity between other kinds of vocalisations produced by adult humans, infants and other mammals (Burling, 1993; laughter: Davila-Ross, Owren, & Zimmermann, 2009, 2010;

Pisanski, Cartei, McGettigan, Raine, & Reby, 2016; infant distress cries: Lingle & Riede, 2014; Lingle et al., 2012; Zeifman, 2001), we predicted that acoustic encoding and perception of pain levels in adult simulated pain vocalisations would follow similar patterns to those observed in human infant and nonhuman mammal pain cries.

## **Method**

### ***1. Acoustic Analysis***

#### **Participants**

We audio recorded 30 male and 30 female students of drama or acting from the Royal Central School of Speech and Drama and the University of Sussex, who received monetary compensation in exchange for their participation. All participants provided informed consent. None were currently suffering from any conditions that might affect their voice (e.g. cold, sore throat). This experiment was approved by the University of Sussex's Life Sciences & Psychology Cluster-based Research Ethics Committee (C-REC) (Certificate of approval: ER/JR307/4) and complies with the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct.

#### **Procedure**

##### *Voice recording*

Vocalisations were recorded in a quiet room, with vocalisers standing 150 cm from a Zoom H4n microphone. A chair was positioned between the vocaliser and the microphone to restrict forward movement. Vocalisers were asked to imagine themselves in three painful situations of increasing intensity, and to produce a vocalisation in response to each imagined scenario. A description of each context was dictated by the experimenter

and also displayed on a computer screen. The descriptions for each pain context were as follows:

Mild: Imagine you are experiencing a mild pain, one that is noticeable but manageable. Scalding your finger with boiling water or stubbing your toe are examples of this level of pain.

Moderate: Imagine you are experiencing a strong pain, one that is serious but not life-threatening. Examples of this level of pain are breaking your arm or dislocating your shoulder. Produce a vocalisation to express your pain.

Severe: Imagine you are experiencing the most intense pain you can think of. Examples are childbirth, or a life-threatening injury. Produce a vocalisation to express your pain.

In order to obtain realistic vocal stimuli, participants were encouraged to take as much time as they needed to immerse themselves in each imagined context, and to ‘let go of their inhibitions’. Participants were also given the option not to vocalise if they felt that they could not naturally produce the sentence or nonverbal vocalisation, and to repeat any sentence or vocalisation until they were satisfied with their portrayal.

#### *Acoustic analysis*

Vocal stimuli were analysed using PRAAT 5.3.62 DSP package (Boersma & Weenink, 2017). Recordings were saved as WAV files at 44.1 kHz sampling frequency and 16 bits amplitude resolution.

We performed acoustic analyses using a dedicated batch-processing script in PRAAT containing three distinct procedures. The first procedure characterized

fundamental frequency (F0) and modulation (F0 contour variation). The F0 contour was extracted using the *To Pitch (cc)...* command. We systematically inspected each extracted pitch contour and verified it using a narrow band spectrogram displaying the first 2000 Hz of the signal. Erroneous pitch values (e.g. octave jumps) were manually corrected by selecting the appropriate F0 candidate values in the edited pitch object. In segments displaying subharmonics (the presence of vocal fold vibration at a frequency equal to an integer multiple of the F0 in addition to the F0 itself, Fitch et al., 2002), the F0 was systematically preferred over the subharmonic. Where amplitude modulation (a subcategory of biphonation, whereby the air displacements of two independent sources of vocal energy, one of low frequency and one of higher frequency, interact to produce a signal with audible periodic variation in overall intensity, Fitch et al., 2002) was present, F0 values were selected only if clearly visible and audible. For segments where deterministic chaos (aperiodic, irregular vocal fold vibration, Fitch et al., 2002) was present, the automatically extracted pitch contour generally did not select F0 values; where it did, we manually deselected these values.

The F0 contour was used to derive the following parameters: mean F0, max F0, min F0, range F0, and F0CV (coefficient of variation in F0 across the entire duration of the signal). During inspection of each spectrogram, we also measured the proportion of the signal for which nonlinear phenomena (amplitude modulation, subharmonics, or deterministic chaos) were present, and created a measure representing this proportion as a percentage (%NLP).

Next, two distinct smoothing algorithms (*Smooth...* command in Praat) were performed on the pitch contour: the first (*Smooth...* command parameter = 25), suppressed period-to-period frequency fluctuations while preserving short-term, minor modulation events (such as vibrato-like frequency modulation, Charlton, Taylor, & Reby, 2017). The

second (*Smooth...* command parameter = 2) suppressed short-term modulation, characterising only major F0 modulation events. After each smoothing procedure, inflection points were counted as each change in the sign of the contour's derivative, and divided by the total duration of the voiced segments in each recording. This resulted in two distinct indexes of F0 modulation (inflex25 - minor inflections, and inflex2 - major inflections).

A second procedure characterised the mean amplitude of the stimuli, as well as amplitude range (intRange) and variability (intCV, the coefficient of variation of the intensity contour estimated using the *To intensity ...* command in PRAAT). A third procedure focused on the periodic quality of the voiced proportion of the signal and measured harmonics-to-noise ratio (HNR, a measure of the ratio of harmonic spectral energy to chaotic spectral energy), jitter (small fluctuations in periodicity measured as the average of 'local', 'rap' and 'ppq5' measures in PRAAT) and shimmer (small variation in amplitude between consecutive periods, measured as the average of 'local', 'apq5' and 'apq11' parameters in PRAAT). Together, HNR, jitter and shimmer represent the overall 'harmonicity' of the voiced proportion of vocal signals. Acoustic analysis procedures similar to these have been applied successfully in previous studies of human nonverbal vocalisations (e.g. babies' cries, Koutseff et al., 2017; Reby, Levréro, et al., 2016).

### *Principal component analysis*

To reduce our set of correlated acoustic variables to a smaller number of uncorrelated factors, we performed a principal component analysis (PCA) with varimax rotation on all aforementioned acoustic variables extracted from the full dataset of 180 vocalisations (Abdi & Williams, 2010) (see Table 3 for mean  $\pm$  SDs of these variables for

each pain intensity level). We entered within-sex z-scores in place of raw measures for sexually dimorphic acoustic characteristics (mean F0, max F0, min F0, range F0).

### **Statistical analysis**

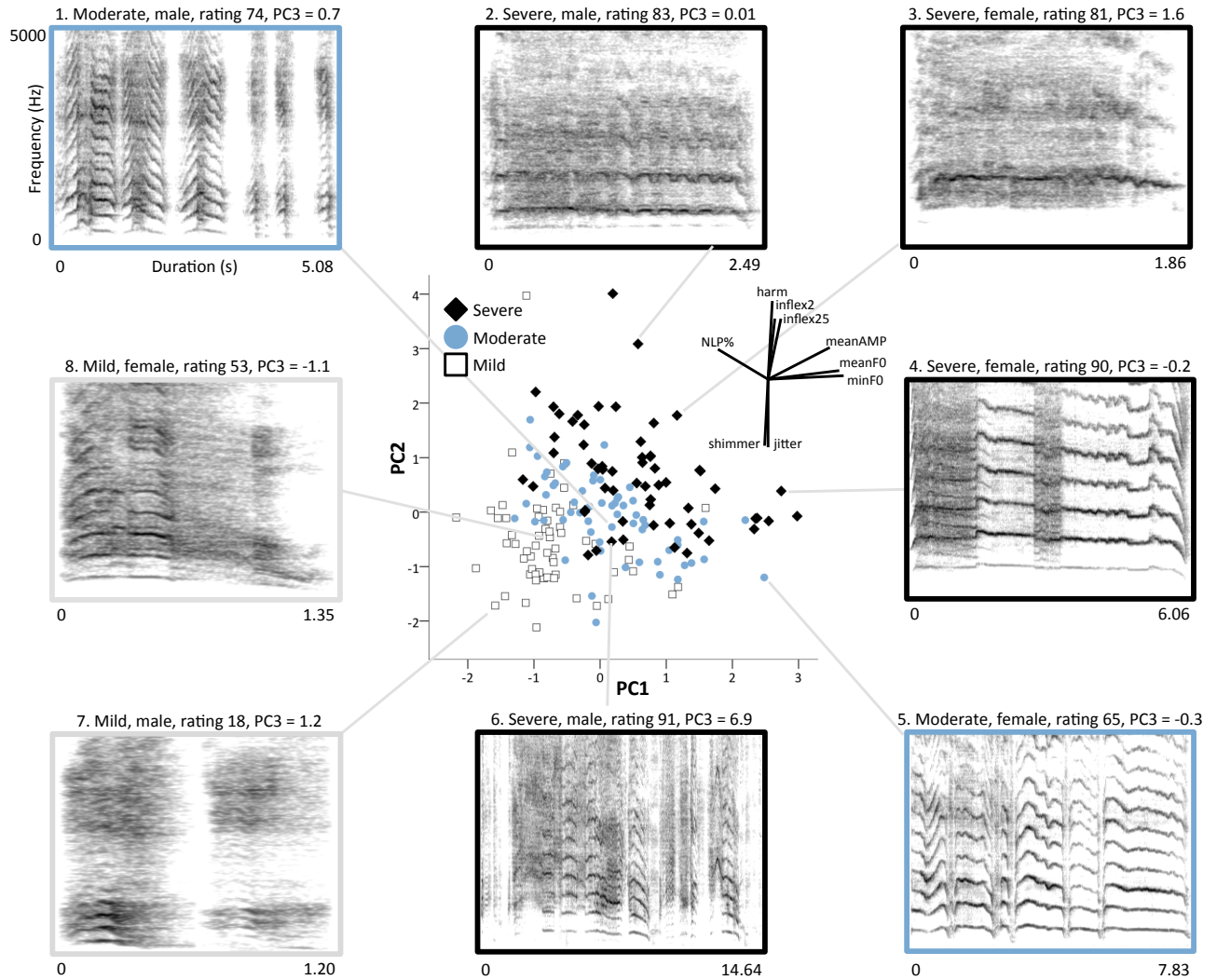
To examine acoustic differences between pain intensities, we conducted a conventional leave-one-out DFA with forced entry (which is less vulnerable to collinear variables, random effects, and type I errors than stepwise entry, Mundry & Sommer, 2007) of the four principal components produced from the acoustic variables. We also conducted a MANOVA to establish whether there were significant differences in each raw acoustic variable between groups.

## ***2. Playback Experiment***

### **Participants**

Thirty females and 34 males ( $M$  age =  $35.65 \pm 9.53$ ) from the USA were recruited using Amazon Mechanical Turk to provide voice-based assessments of the 180 previously acquired pain vocalisations (60 vocalisers x 3 vocalisations). Participants completed the experiment using a custom computer interface. All participants provided informed consent, and were compensated with \$4.

In order to reliably assess the effect of amplitude on listeners' attributions, it was necessary for listeners to maintain the same volume for the duration of the playback experiment. Eight participants who reported that they adjusted their volume settings were excluded from analyses.



*Figure 1.* Principal component analysis (PCA) illustrating the acoustic variability of pain vocalisations across pain intensity levels. Each pain vocalisation is plotted against its score along the first two principal components. The radar plot in the top right corner of the scatterplot represents PC factor loadings of the acoustic variables. Spectrograms illustrate how the vocalisations vary along the principal components. The text directly above each spectrogram describes: the name of the corresponding audio file accessible in the Electronic Supplementary Materials, vocaliser sex, pain intensity, the mean pain rating attributed to the vocalisation, and score on the third principal component.

*Table 1.* Rotated factor loadings for each principal component calculated from the acoustic variables characterising simulated pain vocalisations. Percentage of explained variance in acoustic characteristics and eigenvalues for each factor are given below. The highest factor loading for each acoustic variable across PCs is highlighted in bold.

	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>
	Variance = 33%	Variance = 21%	Variance = 14%	Variance = 10%
<b>Acoustic variable</b>	Eigenvalue = 4.57	Eigenvalue = 2.96	Eigenvalue = 1.97	Eigenvalue = 1.39
HNR (dB)	<b>.92</b>	.04	-.02	.04
Jitter (Hz)	<b>-.79</b>	.00	.08	-.15
Shimmer (dB)	<b>-.76</b>	-.04	.09	.00
Minor F0 inflections	<b>.72</b>	.17	.18	-.11
Major F0 inflections	<b>.72</b>	.03	.23	-.01
Minimum F0 (Hz)	.05	<b>.91</b>	-.11	.05
Mean F0 (Hz)	.10	<b>.87</b>	.31	.02
Mean amplitude (dB)	.37	<b>.75</b>	.22	-.02
Nonlinear phenomena (%)	-.34	<b>.59</b>	.11	-.30
F0 CV (Hz)	-.07	-.09	<b>.95</b>	.05
F0 range (Hz)	.11	.28	<b>.93</b>	.06
Max F0 (Hz)	.12	.56	<b>.77</b>	.08
Intensity CV (dB)	-.22	-.24	-.03	<b>.92</b>
Intensity range (dB)	.31	.25	.25	<b>.81</b>

This experiment was approved by the University of Sussex's Life Sciences & Psychology Cluster-based Research Ethics Committee (C-REC) (Certificate of approval: ER/JR307/8) and complies with the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct.

## **Procedure**

All playback experiments were completed online on Syntoolkit, a dedicated online testing platform for psychology studies (e.g., Hughes, Gruffydd, Simner & Ward, in press; see Simner & Alvarez, forthcoming) that is particularly suited to running studies with sensory or multisensory stimuli. Listeners were instructed to use headphones and complete the experiment in a quiet place. To allow listeners to complete the experiment at a comfortable but audible volume, they were instructed to first set their volume to its lowest level. Listeners then heard a demo sound file (amalgamating a loud and a quiet stimulus), and were instructed to raise their volume until they could clearly hear the quiet vocalisation, while the loud vocalisation did not cause discomfort. Following this, listeners were asked not to adjust their volume settings during the experiment unless it became too uncomfortable. Listeners were asked at the end of the experiment if they adjusted their volume at any point. Due to the agonistic nature of the stimuli, listeners were made aware that if they felt uncomfortable or distressed listening to the sounds, they could stop the experiment.

Voice stimuli were blocked by sex. The order of blocks and stimuli within blocks was randomised. Before each block, participants were reminded to listen to each stimulus in full, and informed that they could take a break at any time. Listeners were instructed to, "Rate how much pain this vocalisation is conveying" on a 101-point Likert scale from 0 (no pain) to 100 (extreme pain).

Listeners were debriefed upon completing the study. They were told that the pain vocalisations were simulated, and that the vocalisers were not really experiencing pain. We examined reaction times against stimulus durations to ensure that participants completed the experiments properly. No participants were removed as a result of this process.

### **Statistical analysis**

We conducted a linear model testing the effects of intensity level and the four acoustic principal components on listeners' pain ratings. The model included main effects and 2-way interactions between pain intensity level and each of the four principal components. We allowed the slopes of the relationship between pain ratings and the predictors to vary between both vocalisers and listeners, and tested null hypotheses based on the average of these slopes. The model included listener ID as a random subject variable, and vocaliser ID as a random factor. Effect sizes (provided in the Figures) were estimated using  $R^2$  coefficients derived from simple linear regressions among relevant variables.

## **Results**

### **Does the acoustic structure of pain vocalisations differ with pain intensity?**

#### *Principal component analysis*

This unsupervised analysis produced four components with eigenvalues greater than 1 (Kaiser's criterion). These components explained 33%, 21%, 14%, and 10% of the

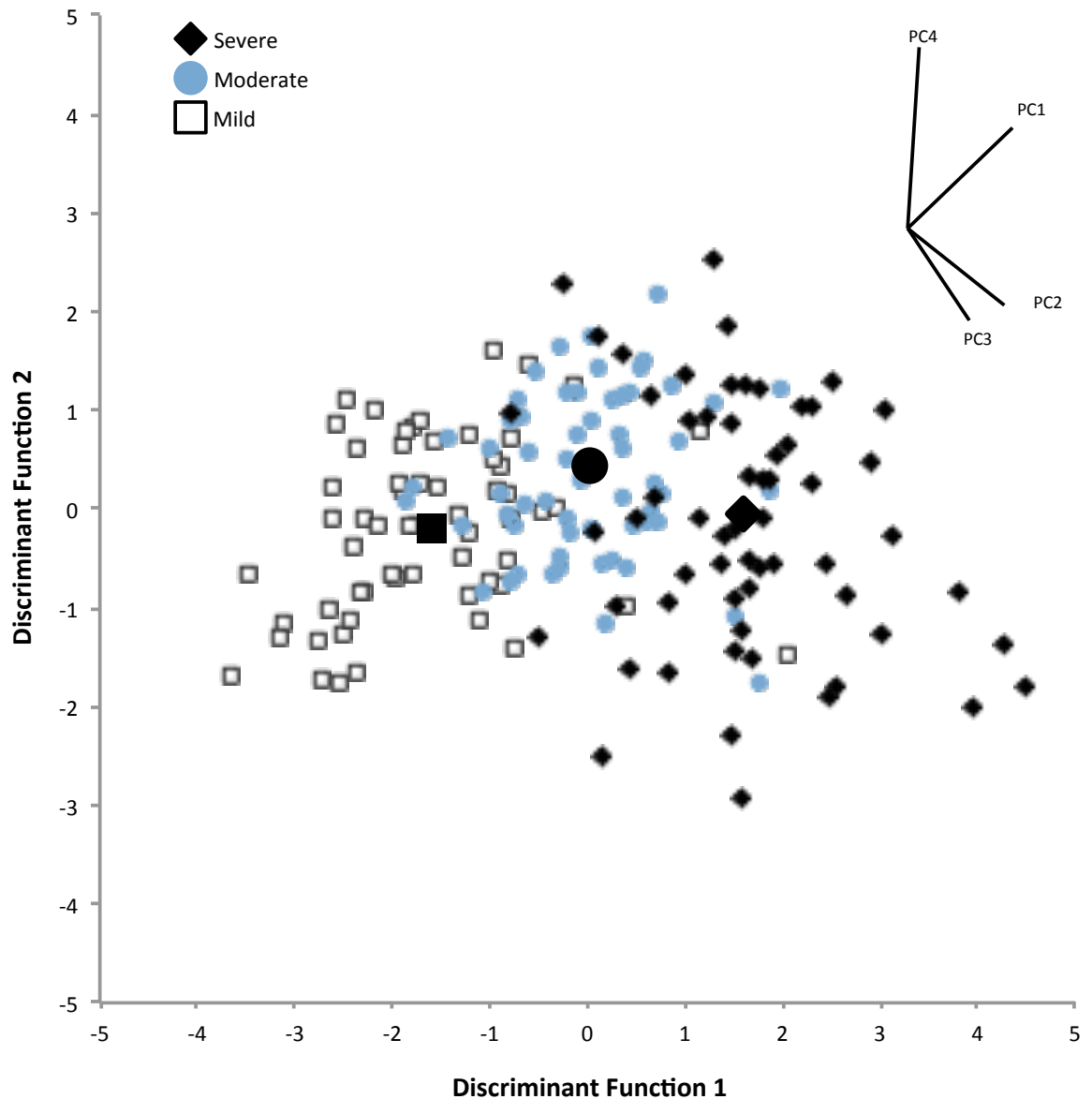
variance in acoustic characteristics, respectively. Acoustic variable loadings on the components are reported in Table 1.

Variable loadings indicated that the first principal component (PC1) indexed the harmonicity and F0 modulation of the voiced proportion of the vocalisations: vocalisations with higher PC1 scores were more harmonic, had a lower level of jitter and shimmer, and had more minor (short-term) and major (longer-term) F0 inflections. Vocalisations with higher PC2 values had a higher mean amplitude, a higher minimum, mean, and maximum F0, and displayed more nonlinear phenomena. PC2 can reasonably be interpreted as an index of subglottal pressure. Indeed, amplitude and F0 both increase with subglottal pressure, as increasing pressure below the glottis raises both the speed at which the vocal folds vibrate and the energy imparted to displaced air upon vocal fold opening (Behrman, 2007; Herbst, 2016); nonlinearities are also observed at the upper limits of subglottal pressure (Berry et al., 1996; Fitch et al., 2002; Herbst, 2016; J. J. Jiang et al., 2001; Y. Zhang & Jiang, 2005). PC3 characterised the range of F0, primarily driven by high maximum F0 values (resulting in higher F0 range, and higher F0CV). The final component (PC4) indexed amplitude variability: vocalisations with higher PC4 scores had higher intCV and intRange values.

#### *Discriminant function analysis*

Discriminant function analysis indicated that the three pain intensities were acoustically distinct (Figure 2): the classification success rate was significantly greater than chance (correct classification percentage = 75.6%, chance = 33.33%,  $p < .0005$ ). Table 2 reports the loadings of the acoustic principal components on the first three discriminant functions. The first discriminant function (eigenvalue = 1.82, variance explained = 96.8%) was the key differentiator of intensity categories (Figure 2),

demonstrating that the harmonicity and F0 modulation of the voiced proportion of the signal (PC1), F0, amplitude, nonlinear phenomena (PC2), and F0 variation (PC3) all increased with pain intensity. The second discriminant function (eigenvalue = 0.06, variance explained = 3.2%) was not important in discriminating groups.



*Figure 2.* Discriminant function analysis illustrating acoustic separation of pain vocalisations at different levels of pain intensity. Each data point represents the centroid of a vocal stimulus as a function of the first two discriminant variables that maximise individual separation. Larger black data points represent mean group centroids for each stimulus condition. The radar plot on the top right represents the loadings of the principal components on the first two discriminant functions. Pain intensity categories were mainly separated on the first three principal components (see Table 1).

Using Pillai's trace, a MANOVA revealed that there was a significant effect of pain intensity on the raw acoustic variables ( $V = 8.75$ ,  $F(28, 330) = 8.75$ ,  $p < .0005$ ). Separate univariate ANOVAs revealed that the effect of pain intensity was significant for each acoustic variable (all  $ps < .012$ ). Tables 3 and 4 report the mean values of the raw acoustic variables, as well as the principal components, for each vocaliser sex and pain intensity. Patterns of acoustic variation were comparable across sexes.

### **Do pain intensity level and acoustic characteristics affect ratings of pain?**

Linear mixed model analysis revealed a significant effect of pain intensity level on pain ratings (Table 5): mild intensity pain vocalisations were rated as conveying the least pain ( $M = 16.61 \pm 1.31$ ), followed by moderate intensity vocalisations ( $M = 44.21 \pm 1.19$ ), with severe intensity vocalisations rated as conveying the most pain ( $M = 75.25 \pm 1.20$ ).

All four principal components significantly predicted pain ratings (Table 5). Higher pain ratings were associated with greater harmonicity in the voiced proportion of the signal (PC1), higher F0, amplitude, and nonlinear phenomena (PC2), greater F0 variation (PC3), and greater intensity variation (PC4), although the effect size for PC4 was minimal (Figure 3).

As illustrated in Figure 3, pain ratings increased as principal component values increased. However, the relative contribution of individual principal components in predicting listeners' ratings of pain intensity differed across intensity levels. Variation in PC1 had the greatest effect on pain ratings in moderate pain vocalisations, a smaller effect in mild pain vocalisations, and no effect in severe pain vocalisations. The effect of PC1 on pain ratings was also more reliable in moderate than in mild pain vocalisations. PC2 only reliably affected pain ratings within the mild intensity category. Listeners were sensitive to

*Table 2.* Principal component (PC) loadings on the discriminant functions (DF).

Principal components are defined in Table 1.

Acoustic variable	DF1	DF2
PC1 – Harmonicity (voiced proportion of signal)	.48	.44
PC2 – F0, amplitude, nonlinear phenomena	.43	-.32
PC3 – F0 variation	.24	-.41
PC4 – Intensity variation	.06	.78

*Table 3.* Mean acoustic characteristics of female vocal stimuli. Figures in square brackets represent standard errors of the mean.

Acoustic variable	Pain intensity		
	Mild	Moderate	Severe
Mean F0 (Hz)	465.5 [43.31]	539.9 [35.57]	737.8 [55.28]
Max F0 (Hz)	537.4 [50.86]	697.4 [53.58]	983.6 [78.92]
Min F0 (Hz)	367.2 [37.07]	383.2 [17.41]	464.6 [34.41]
Range F0 (Hz)	170.2 [28.48]	314.2 [41.68]	519.0 [67.06]
F0 CV (Hz)	0.11 [0.02]	0.15 [0.02]	0.17 [0.03]
Minor F0 inflections	1.84 [0.25]	3.92 [0.42]	6.77 [0.65]
Major F0 inflections	0.19 [0.04]	0.41 [0.06]	0.64 [0.08]
Mean amplitude (dB)	55.10 [1.38]	62.70 [1.32]	71.38 [0.93]
Intensity range (dB)	22.74 [1.07]	29.54 [0.90]	31.26 [1.41]
Intensity CV (dB)	1.19 [0.06]	1.16 [0.06]	0.90 [0.07]
Shimmer (dB)	0.15 [0.01]	0.11 [0.01]	0.10 [0.01]
Jitter (Hz)	0.029 [0.002]	0.017 [0.002]	0.018 [0.002]
HNR (dB)	5.73 [0.99]	10.57 [0.92]	12.91 [1.09]
Nonlinear phenomena (%)	44.69 [6.34]	39.70 [4.55]	55.07 [5.05]
Principal component	Mild	Moderate	Severe
PC1 – Harmonicity (voiced proportion)	-0.49 [0.15]	0.44 [0.17]	0.92 [0.19]
PC2 – F0, amplitude, %NLP	-0.40 [0.19]	-0.21 [0.12]	0.49 [0.18]
PC3 – F0 variation	-0.42 [0.12]	-0.05 [0.14]	0.45 [0.23]
PC4 – Intensity variation	-0.26 [0.15]	0.01 [0.12]	-0.42 [0.18]

*Table 4.* Mean acoustic characteristics of male vocal stimuli. Figures in square brackets represent standard errors of the mean.

Acoustic variable	Pain intensity		
	Mild	Moderate	Severe
Mean F0 (Hz)	270.7 [12.54]	340.4 [9.38]	440.8 [20.84]
Max F0 (Hz)	312.3 [14.98]	420.4 [15.79]	654.9 [59.72]
Min F0 (Hz)	209.0 [11.19]	251.6 [10.07]	296.1 [12.85]
Range F0 (Hz)	103.3 [9.08]	168.8 [17.42]	358.8 [61.96]
F0 CV (Hz)	0.13 [0.01]	0.13 [0.01]	0.18 [0.03]
Minor F0 inflections	0.80 [0.15]	2.79 [0.37]	5.22 [0.52]
Major F0 inflections	0.12 [0.03]	0.27 [0.05]	0.49 [0.06]
Mean amplitude (dB)	52.33 [1.30]	62.62 [1.37]	69.99 [1.15]
Intensity range (dB)	22.33 [1.40]	32.42 [1.21]	37.19 [1.89]
Intensity CV (dB)	1.27 [0.07]	1.39 [0.07]	1.17 [0.10]
Shimmer (dB)	0.17 [0.01]	0.14 [0.01]	0.13 [0.01]
Jitter (Hz)	0.038 [0.003]	0.022 [0.001]	0.020 [0.002]
HNR (dB)	2.51 [0.44]	6.82 [0.49]	8.44 [0.76]
Nonlinear phenomena (%)	40.70 [3.94]	48.58 [5.35]	64.84 [4.58]
Principal component	Mild	Moderate	Severe
PC1 – Harmonicity (voiced proportion)	-0.96 [0.10]	-0.17 [0.10]	0.26 [0.14]
PC2 – F0, amplitude, %NLP	-0.74 [0.13]	0.03 [0.14]	0.83 [0.17]
PC3 – F0 variation	-0.22 [0.08]	-0.21 [0.11]	0.44 [0.28]
PC4 – Intensity variation	-0.20 [0.18]	0.52 [0.15]	0.35 [0.24]

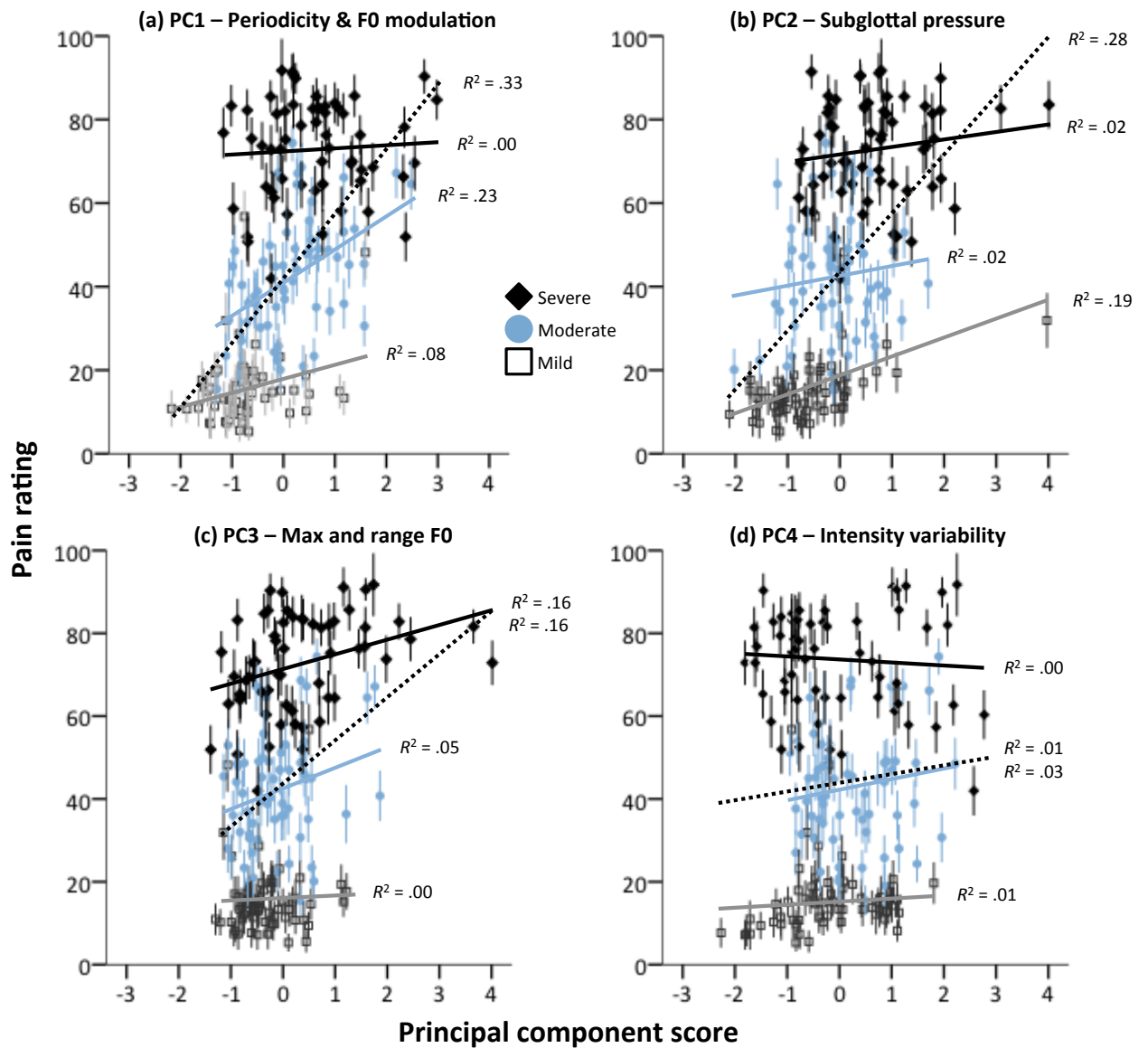
variation in PC3 only in severe and moderate pain vocalisations, but the effect of PC3 on pain ratings was much more reliable in severe pain vocalisations. Finally, PC4 increased marginally with pain ratings within moderate pain vocalisations, but effect sizes were minimal. These results demonstrate that acoustic variables contribute differently to listeners' perceptions of pain at different pain intensity levels.

*Table 5.* Linear mixed model testing the effects of the intensity of pain simulated by the vocaliser and the four acoustic principal components on listeners' attributions of the level of pain conveyed by the vocaliser.

<b>Source</b>		<b><i>df</i><sub>1</sub>, <i>df</i><sub>2</sub></b>	<b><i>F</i></b>	<b><i>p</i></b>
i.	Intercept	59.02	1070.73	< .001
ii.	Pain intensity	9102.63	1117.40	< .001
iii.	PC1	8873.25	1038.61	< .001
iv.	PC2	9728.38	905.87	< .001
v.	PC3	9695.95	525.95	< .001
vi.	PC4	9012.57	30.52	< .001
vii.	Pain intensity * PC1	9173.89	168.87	< .001
viii.	Pain intensity * PC2	9547.43	67.28	< .001
ix.	Pain intensity * PC3	9412.55	37.59	< .001
x.	Pain intensity * PC4	8972.28	31.96	< .001

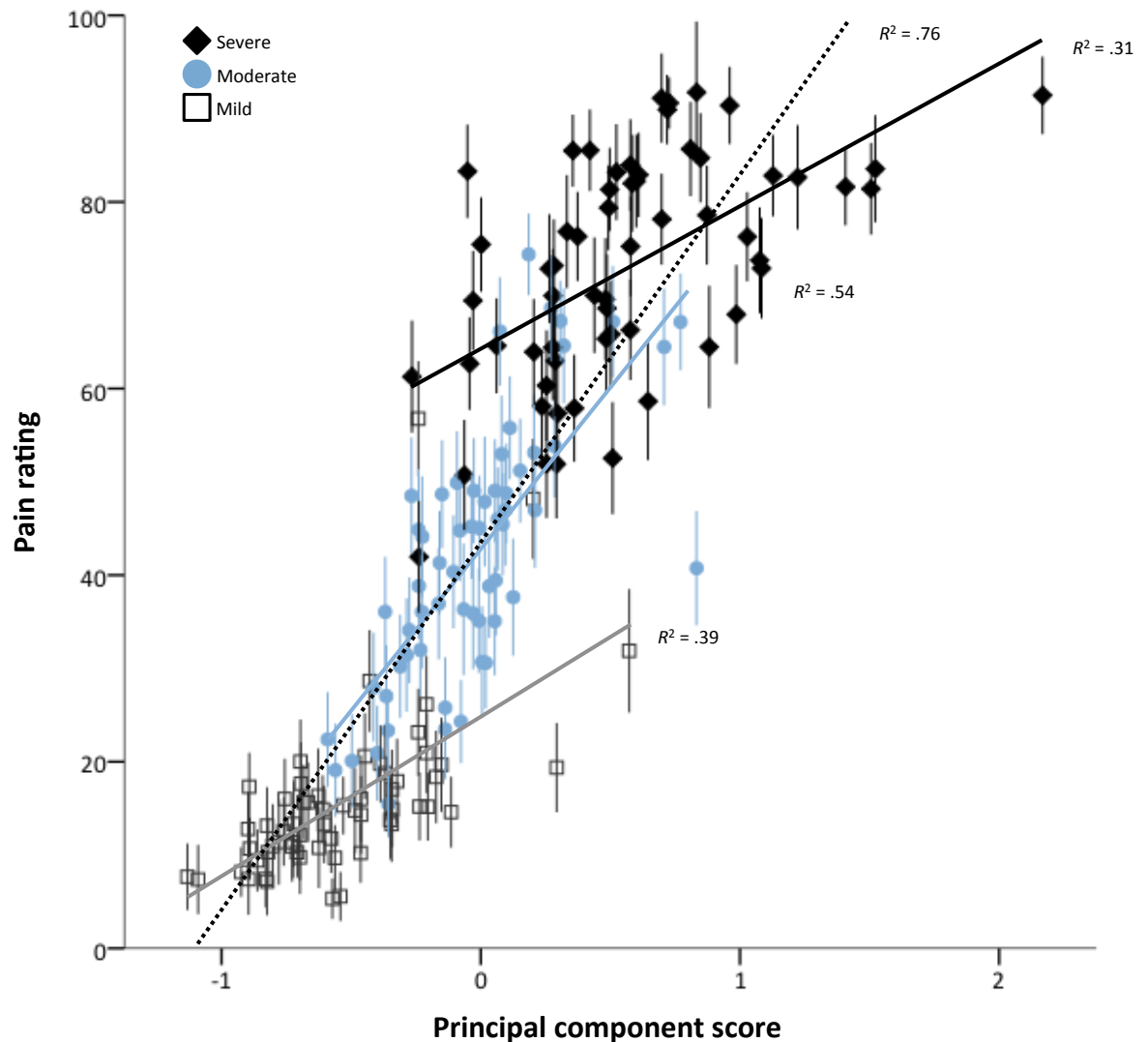
Inspection of spectrograms (see examples in Figure 1) suggested that vocalisations often transitioned between highly periodic (PC1) and highly chaotic (PC2) regimes of vocal fold vibration (e.g. Figure 1 spectrograms 4 and 6). Vocalisations with such bifurcations would not score highly on individual components, despite exhibiting multiple characteristics associated with high pain ratings. In addition, some vocalisations exhibited octave jumps or other forms of F0 variation (producing high PC3 scores) concurrently with periodic or chaotic vibratory regimes (e.g. Figure 1 spectrogram 6), of which the possible additive effect on pain ratings cannot be assessed by testing each PC individually.

Therefore, for each vocalisation, we computed an average of values for the first three principal components (excluding PC4 due to the observed minimal effect sizes), and conducted a fully factorial linear mixed model with only pain intensity and the average of PCs 1-3 (PC123) as predictors. Both main effects and interaction terms were highly significant (all *F*s > 132.36, all *p*s < .001). Higher pain ratings were associated with higher



*Figure 3.* Pain rating as a function of variation in the four principal components (a)-(d). Each data point represents the mean pain rating averaged across listeners for each pain vocalisation. Error bars represent 95% confidence intervals. White squares represent mild pain simulations; blue circles represent moderate pain simulations; black diamonds represent severe pain simulations.  $R^2$  values for each regression line (calculated based on mean pain ratings) are reported in the graphs. Dotted regression lines represent the relationship between each principal component and pain ratings across pain intensity levels. (a) PC1 represents the harmonicity and F0 modulation of the voiced proportion of the signal. (b) PC2 represents indicators of subglottal pressure (mean amplitude, F0, and the proportion of the signal displaying nonlinear phenomena). (c) PC3 represents max and range F0. One value (6.87, 91) is not represented in the graph but is included in the regression lines. (d) PC4 represents intensity variability.

PC123 scores, with PC123 explaining 75.5% of the variance in listeners' pain ratings (Figure 4). Within pain intensity levels, variation in PC123 had the greatest effect on pain ratings in moderate pain vocalisations.



*Figure 4.* Pain ratings as a function of the average of the first three principal components. Each data point represents the mean pain rating averaged across listeners for each pain vocalisation. Error bars represent 95% confidence intervals. White squares represent mild pain simulations; blue circles represent moderate pain simulations; black diamonds represent severe pain simulations.  $R^2$  values for each regression line (calculated based on mean pain ratings) are reported in the graphs. The dotted regression line represents the relationship between each principal component and pain ratings across pain intensity levels. This regression line explains the most (76%) variance in pain ratings.

## Discussion

Our results show that acoustic variation in simulated pain vocalisations produced by adult men and women is organised along three main groups of acoustic characteristics. Together, these acoustic components are sufficient to reliably separate vocalisations by their simulated level of pain intensity (mild, moderate and severe) and in turn reliably predict the assessments of pain intensity by adult listeners. Moreover, while the relative contribution of these acoustic characteristics to listeners' pain ratings varied within each level of simulated pain intensity, their combination (by averaging) was the strongest and most reliable predictor of listeners' pain ratings across and within pain levels.

### *Acoustic cues to levels of pain intensity*

The results of the principal component analysis revealed that the acoustic variability of simulated pain vocalisations could be described by three uncorrelated groups of acoustic variables. A first group of variables (all loading on PC1) characterised the harmonicity (HNR, jitter and shimmer) and F0 modulation (rate of short term and long-term inflections) of the voiced proportion of the signal. A second group of variables (loading on PC2) characterised the pitch (min and mean F0), amplitude (mean amplitude) and occurrence of nonlinear phenomena (percentage presence – %NLP), which are all known to increase with subglottal pressure. Finally, a third group of variables (max and range F0, F0 CV, all loading on PC3) represented pitch range and variability. The fact that the vocalisations were clearly organised according to the three increasing levels of simulated pain intensity in the three-dimensional space created by these components strongly indicates that a large proportion of the acoustic variation in our vocalisations served to express pain intensity (see Figures 1 & 3).

These results support our predictions, which stem from previous work on human infants and other mammals. Indeed, indicators of subglottal pressure and pitch range have previously been shown to encode pain intensity in pain vocalisations produced by human infants (F0: Bellieni et al., 2004; roughness: Facchini et al., 2005; Koutseff et al., 2017; Tiezzi et al., 2004; amplitude: Fuller & Conner, 1995; Lehr et al., 2007; F0 variability/range: Koutseff et al., 2017; Porter et al., 1986) and nonhuman mammals (F0: White et al., 1995; roughness: Levine et al., 1984; amplitude: Eschali r et al., 1988; Jourdan et al., 1995). These acoustic features also influence perceived urgency of caregivers in nonhuman mammals (Jovanovic & Gouzoules, 2001; Weary et al., 1996), as well as assessments of pain (Craig et al., 1988; Porter et al., 1986) or distress (Esposito et al., 2015; Kelly et al., 2017; LaGasse et al., 2005; Wood, 2009) in adult humans listening to infant distress cries. Thus, the observed increases in acoustic indicators of subglottal pressure and pitch range with simulated pain intensity level (and perceived pain intensity) are consistent with acoustic mediators of pain communication observed in authentic pain vocalisations produced by human infants and infant or adult nonhuman mammals.

We also found that, as vocalisers simulated higher pain levels, they produced vocalisations with more modulated F0 (short- and long-term inflections, contributing to PC1). To our knowledge, this is the first time that frequency modulation has been identified as communicating pain intensity, although high frequency modulation is associated with calls produced in fearful contexts in nonhuman mammals (Briefer, 2012). Inspection of spectrograms suggested that vocalisations with high short-term F0 modulation were either characterised by vibrato-like frequency modulation (Figure 1, spectrograms 2, 3, 4 and 5), and/or numerous glottal stops (Figure 1, spectrograms, 1 4, 5, and 6), both giving the vocalisations a cry/sob-like quality. Similar shifts in vocal quality

have been observed in infant cries, where individual cries within bouts become shorter and more frequent as pain increases (Porter et al., 1986).

We also observed that, while the occurrence of nonlinear phenomena (contributing to PC2) increased as levels of simulated pain intensity increased, the harmonicity of the voiced proportion of the signal (i.e. the proportion with a detectable pitch, loading on PC1) increased, thus contrasting previous research on human infant pain cries (Koutseff et al., 2017). We argue that this is primarily driven by the breathy voice quality that characterised the majority of mild intensity vocalisations (see Figure 1, spectrograms 7 & 8), but is not observed in infant cries (Facchini et al., 2005; Koutseff et al., 2017). Breathy voice is produced with minimal glottal closure (Gobl & Chasaide, 1992), resulting in turbulent airflow accompanying vocal fold vibration and therefore producing a much less periodic acoustic signal than modal speech (de Krom, 1995; Gobl & Chasaide, 1992; Herbst, 2016; Hillenbrand, 1988; Hillenbrand, Cleveland, & Erickson, 1994; Hillenbrand & Houde, 1996; Scherer, 1986) or shouted speech (C. Zhang & Hansen, 2007). In contrast, the higher amplitude of moderate and severe intensity vocalisations is associated with greater and more abrupt glottal closure (Backstrom, Alku, & Vilkmann, 2002; Södersten, Hertegård, & Hammarberg, 1995), achieved through high vocal fold tension and resulting in relatively less turbulent air leakage (associated with “pressed” voice quality, Gobl & Chasaide, 1992; Herbst, 2016; Södersten et al., 1995), and therefore a more periodic acoustic signal.

However, as subglottal pressure reaches the upper limits at which the vocal folds vibrate stably, the vocal folds transition to chaotic regimes of vibration (Fitch et al., 2002; Herbst, 2016; see Figure 1), which can overlay or replace periodic spectral components (as observed in infant cries, Facchini et al., 2005; Koutseff et al., 2017). Highly irregular, unvoiced portions in acoustic recordings (during which pitch is absent or undetectable) are

not considered in harmonicity measures, but are represented by the percentage of the signal for which nonlinear phenomena are present (contributing to PC2). Thus, vocalisations may be characterised by either high PC1 values (highly periodic), high PC2 values (highly chaotic), or, where vocalisations transition between periodic and highly chaotic vocal regimes (bifurcations, Fitch et al., 2002; Herbst, 2016; e.g. Figure 1, spectrograms 4 & 6), a combination of the two. The prevalence of nonlinear phenomena in vocalisations associated with higher pain corroborates a growing body of evidence that this acoustic characteristic typically serves to attract attention (Arnal, Flinker, Kleinschmidt, Giraud, & Poeppel, 2015; Blumstein & Récapet, 2009; Charlton, Watchorn, & Whisson, 2017; Mitani & Stuht, 1998; Reby & Charlton, 2012).

A follow-up discriminant analysis based on the first three acoustic principal components reliably classified vocalisations according to the three levels of simulated pain intensity (76% correct classification). Should this high reliability extend to authentic pain vocalisations, our multivariate acoustic analyses may form the basis for the development of a practical quantitative tool to improve pain assessment in populations unable to self-report their subjective pain experience (Docking, Lane, & Schofield, 2017; Herr et al., 2011), especially as pain levels appear to be discriminated more sensitively by acoustic analysis than by perceptual judgments (Koutseff et al. 2017). Future research could apply this methodology to real pain vocalisations such as childbirth vocalisations (Fuller et al., 1993), wherein acoustic indicators of pain may offer a viable alternative (Baker & Kenner, 1993) to obtrusive and much-criticised vaginal examination (Dahlen, Downe, Duff, & Gyte, 2013; Shepherd & Cheyne, 2013) in monitoring labour stage.

#### *Effect of acoustic cues on listeners' assessments of pain intensity*

The values of each principal component varied within pain intensity levels, and predicted pain ratings both within and across these levels, supporting the contention that

acoustic communication of pain is graded (likely as a function of distress-related arousal), rather than discrete (Bellieni, 2012; Bellieni et al., 2004; Briefer, 2012; Kelly et al., 2017; Out et al., 2010; Porter et al., 1986; Sauter et al., 2010). Moreover, acoustic variation in our simulated pain vocalisations influenced listeners' perceptions of pain in a manner consistent with reported effects of F0 and roughness on the urgency of nonhuman mammals' responses to distress cries (Jovanovic & Gouzoules, 2001; Weary, Lawson, & Thompson, 1996). Interestingly, the relative contribution of each component to listeners' pain ratings varied within pain intensity levels. Mild intensity vocalisations tended to be characterised by indicators of low subglottal pressure, indicators of breathy voice quality, low F0 range, and elicited low pain ratings. Within this category, pain ratings were mainly driven by indicators of subglottal pressure (PC2), and to a lesser extent, breathiness (PC1). Moderate intensity vocalisations elicited higher pain ratings, but ratings were primarily influenced by harmonicity (i.e. breathiness of voice quality) and pitch range (e.g. due to frequency jumps, high max F0). Finally, severe pain vocalisations tended to either be highly periodic, highly chaotic, or transitioned between the two vocal fold vibration regimes, and elicited the highest pain ratings. Yet regardless of vibratory regime, pitch range largely determined whether severe pain vocalisations were rated relatively low or high.

The increase in both harmonicity and nonlinear phenomena (characterised by PC1 and PC2 scores) with pain intensity and pain ratings suggests that pain can be communicated via distinct, seemingly opposing acoustic regimes (periodic vs. chaotic). Opposite relationships between roughness and distress-related arousal have also been documented in different species (Facchini et al., 2005; Levine & Gordon, 1982; Stoeger, Charlton, Kratochvil, & Fitch, 2011; c.f. Blumstein & Chi, 2012; Linhart, Ratcliffe, Reby, & Špinka, 2015; Puppe et al., 2005), and, in piglets, between call types in the same

distress-inducing context (Linhart et al., 2015). Importantly, the combined influence of the first three acoustic principal components contributed substantially to the high accuracy of our discriminant analysis (76%), and more strongly and reliably predicted pain ratings both across ( $R^2 = 76\%$ ) and within ( $R^2$ s = 31-54%) intensity levels than did any individual acoustic component. Therefore, while pain can be conveyed via multiple acoustic routes, and the relative influence of each individual acoustic component on pain ratings varies across pain levels, it is their additive presence that most effectively communicates pain intensity.

*Are simulated vocalisations functional?*

The fact that we focused on simulated pain vocalisations may be seen as a limitation affecting the ecological relevance of our results. However, while there are acoustic, perceptual, and neural differences between simulated and authentic nonverbal vocalisations (Anikin & Lima, 2017; Bryant & Aktipis, 2014; Lavan, Scott, & McGettigan, 2015; McGettigan et al., 2015), acted portrayals are generally considered acceptably similar to spontaneous nonverbal vocalisations (Sauter et al., 2010; Sauter & Fischer, 2017). In particular, simulated pain vocalisations are among the most likely to be classified as authentic, and there is a smaller difference in listeners' judgments of authenticity between spontaneous and simulated pain vocalisations than for most other vocalisations (Anikin & Lima, 2017). Consistent with this, we found that the expression and perception of pain in these vocalisations appeared to follow similar rules to those reported in the vocalisations of preverbal human infants and nonhuman mammals (as discussed in the previous sections). In particular, the substantially larger increase in F0 between moderate and severe intensities than between mild and moderate intensities that we report suggests that actors produced vocalisations mirroring the previously observed

‘alarm threshold’ in human infant pain cries (Bellieni et al., 2004), rather than capitalising on more linear associations between F0 and perceived pain (Craig et al., 1988; Porter et al., 1986) to influence listeners’ attributions.

Moreover, simulation is likely to be an integral component of the spontaneous communication of vocal pain in adult humans. Recent evidence that non-actors may provide vocal expressions as realistic as do actors (R. Jürgens et al., 2015) suggests that the capability to accurately simulate spontaneous vocalisations and elicit appropriate listener responses may not be limited to actors. Indeed, humans can even modulate (exaggerate or minimise) responses to genuine pain depending on context, mood, and cognition (see Tracey & Mantyh, 2007), indicating that spontaneous expression of pain is dependent not just on nociceptive input, but also on communicative intentions. Future work could investigate whether listeners can detect exaggeration in partially or fully simulated pain vocalisations.

Humans’ ability to modulate or simulate pain expression is also consistent with functional vocal deception in other social mammals, which is commonly observed in survival contexts despite the potential costs associated with ‘crying wolf’ (Oesch, 2016; Schmid, Karg, Perner, & Tomasello, 2017). For example, in capuchin monkeys, deceptive alarm calls are acoustically indistinguishable from predator-elicited alarm calls, and evoke comparable responses from conspecifics (Wheeler & Hammerschmidt, 2013). Vocal pain exaggeration or simulation may thus be an adaptive survival-enhancing strategy, for example eliciting urgent aid. Such volitional modulation of nonverbal vocalisations may have been at the origins of selection for increased vocal control, eventually culminating in the emergence of articulated speech in humans (Oesch, 2016; Pisanski et al., 2016).

## Chapter 5: Tennis grunts communicate cues to sex and contest outcome

### Summary

The previous chapters examine the communication of motivational and indexical information in volitionally produced vocalisations (aggressive roars, distress screams and pain cries). The final empirical chapter sets out to investigate the communication of such information in spontaneously produced tennis grunts – a nonverbal vocalisation produced during a competitive context. More specifically, the following questions will be explored:

**Question 10.** Does the F0 of tennis grunts communicate information about the vocaliser and context, similar to nonhuman mammal vocal displays?

Chapter 5 investigated Question 10 by testing whether the F0 of tennis grunts encodes static cues to vocalisers' indexical characteristics (sex, height, weight, age), and covaries dynamically with tennis shot type (a proxy of body posture), and the progress and outcome of male and female professional tennis contests. I then investigated whether tennis grunts have perceptual and functional relevance in playback experiments using a separate sample of natural and F0-resynthesised grunts.

Summary of findings:

- The F0 of tennis grunts predicted player sex, consistent with sexual dimorphism in human vocal folds and speaking F0, but not age or body size, consistent with the weak relationship between F0 and body size in human speech and other mammal vocalisations, and with the relative stability of F0 in human speech after puberty.

- The F0 of grunts accompanying serves was higher than that of grunts accompanying groundstrokes, consistent with the influence of dynamic biomechanical constraints on vocal production mechanisms
- Grunts produced later in contests had higher F0 than those produced earlier, consistent with the effects of fatigue, arousal and stress on F0 in speech.
- The F0 of grunts occurring during contests that players lost had a higher F0 than those produced during contests they won, consistent with the negative relationship between F0 and dominance, and positive relationships between F0 and stress and distress observed in human speech and nonhuman mammal vocalisations.
- This difference emerged early in matches, and did not change in magnitude as matches progressed, suggesting a possible role of physiological and/or psychological factors manifesting early or before matches.
- Listeners with tennis playing experience used grunt F0 to infer sex and contest outcome, indicating that tennis grunts, like nonhuman mammal calls, contain perceptible static and dynamic cues to vocalisers' anatomical and motivational state.

**Abstract**

Despite their ubiquity in human behaviour, the communicative functions of nonverbal vocalisations remain poorly understood. Here, we analysed the acoustic structure of tennis grunts, nonverbal vocalisations produced in a competitive context. We predicted that tennis grunts convey information about vocalizer and context, similar to nonhuman vocal displays. Specifically, we tested whether the fundamental frequency (F0) of tennis grunts conveys static cues to a player's sex, height, weight, and age, and covaries dynamically with tennis shot type (a proxy of body posture) and the progress and outcome of male and female professional tennis contests. We also performed playback experiments (using natural and resynthesised stimuli) to assess the perceptual relevance of tennis grunts. The F0 of tennis grunts predicted player sex, but not age or body size. Serve grunts had higher F0 than forehand and backhand grunts, grunts produced later in contests had higher F0 than those produced earlier, and grunts produced during contests that players won had a lower F0 than those produced during lost contests. This difference in F0 between losses and wins emerged early in matches, and did not change in magnitude as the match progressed, suggesting a possible role of physiological and/or psychological factors manifesting early or even before matches. Playbacks revealed that listeners use grunt F0 to infer sex and contest outcome. These findings indicate that tennis grunts communicate information about both vocalizer and contest, consistent with nonhuman mammal vocalisations.

## Introduction

Despite their ubiquitous use across ages, sexes, contexts and cultures, human nonverbal vocalisations remain underinvestigated. In fact, aside from laughter (see Bryant et al., 2016; Scott, Lavan, Chen, & McGettigan, 2014) and infant cries (Lingle, Wyman, Kotrba, Teichroeb, & Romanow, 2012 for review), human nonverbal vocalisations (such as moans, sighs, roars, screams, and grunts) have received little attention, especially from a functional and evolutionary perspective.

Indeed, the majority of research on human nonverbal vocalisations has focused on their classification according to emotional content (e.g. Belin, Fillion-Bilodeau, & Gosselin, 2008; Lima, Castro, & Scott, 2013; Sauter, Eisner, Calder, & Scott, 2010), while overlooking their potential to convey indexical cues about the vocalizer such as age, sex, body size, and social dominance. Such cues are typically present in the vocal signals of nonhuman mammals, and function to mediate interactions in social and sexual contexts (Briefer, 2012; Taylor, Charlton, & Reby, 2016). Human nonverbal vocalisations likely predate language, and appear homologous in structure and function with nonhuman vocalisations (e.g. laughter Davila-Ross, Owren, & Zimmermann, 2010; Pisanski, Cartei, McGettigan, Raine, & Reby, 2016; infant distress vocalisations Lingle et al., 2012). As such, they may constitute a relatively direct link between animal and human vocal systems. Investigating their production, control and perception may therefore provide a unique window into the origins and evolution of human vocal behaviour (Pisanski, Cartei, et al., 2016).

Here, we examined whether the acoustic structure of tennis grunts – a nonverbal vocalisation produced during a competitive contest – contains functionally relevant and perceptible cues. While there are clear qualitative differences between tennis matches and

nonhuman mammal contests (tennis players do not voluntarily yield to dominant competitors, and there are more clearly defined rules and endpoints), animal contests often follow ritualised rules and structures, during which competitors produce signals that contain static and dynamic information about their respective physical condition and motivation (e.g. ungulates: Jennings & Gammell, 2013). Thus tennis matches provide a potentially useful model to examine whether similar information is communicated in human competitive interactions.

Investigations of the function of tennis grunts have so far focused on their distracting quality to opponents (Farhead & Punt, 2015; Sinnett & Kingstone, 2010), and their enhancement of ball velocity without increasing oxygen cost (e.g. O’Connell, Hinman, Hearne, Michael, & Nixon, 2014). No previous study has examined their possible communicative value. We hypothesized that tennis grunts are competitive nonverbal vocalisations homologous to those produced by nonhuman animals in agonistic contexts. Accordingly, we predicted that grunts would contain static and dynamic cues to anatomical and physiological traits of the vocaliser. As vocalisations produced in a competitive context, tennis grunts may be affected by changes in players’ physiological and psychological state (e.g. arousal and stress, which correlate positively with F0, Briefer, 2012; Pisanski, Nowak, & Sorokowski, 2016). Thus, we predicted that grunt F0 would correlate with the outcome of competitive contests (i.e., tennis matches). Finally, we predicted that listeners would be able to use these static and dynamic cues to make functionally relevant inferences about both the tennis player and the match.

Recent research generalising the source-filter model of speech production (Fant, 1960) to vertebrate vocal signals has highlighted the function of fundamental frequency (F0, affecting perceived pitch) and formant frequencies (resonances of the supralaryngeal vocal tract, affecting perceived timbre) in communicating various static and dynamic cues

in nonhuman mammal vocalisations (see Briefer, 2012; Taylor et al., 2016). For example, sexually-selected calls communicate F0-based cues to dominance (e.g. male deer groans Liu et al., 2015; Vannoni & McElligott, 2008) and formant cues to body size (red deer Reby et al., 2005; Australian sea lions Charrier, Ahonen, & Harcourt, 2011; dogs Taylor, Reby, & McComb, 2010). However, whether fundamental and formant frequencies also communicate similar information in human nonverbal vocalisations remains to be determined.

To address this, we investigated the acoustic structure of tennis grunts produced by male and female tennis players during professional matches. Because tennis grunts are relatively high-pitched vocalisations characterised by a low spectral density, formant frequencies were poorly defined and difficult to both perceive and measure (Pisanski, Fraccaro, Tigue, O'Connor, & Feinberg, 2014; Ryalls & Lieberman, 1982). We therefore focused our analyses on the mean F0 of grunts. We recorded the sex, height, weight, and age of the vocalisers (static cues), the type of tennis shot accompanying the grunt (forehand/backhand/serve), and the outcome of the given match (vocaliser won/lost) (dynamic cues). We then investigated whether tennis grunts have perceptual and functional relevance in playback experiments using a separate sample of natural and F0-resynthesised tennis grunts.

We predicted that (i) females would produce tennis grunts with higher F0 than males due to sexual dimorphism in human vocal folds and F0 (Taylor et al., 2016; Titze, 1994); (ii) F0 would not indicate height, weight, or age in our sample of young adult players due to the weak relationship between speaking F0 and body size within-sexes in human speech (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014 for meta-analysis) and other mammal vocalisations (Taylor et al., 2016 for review), and based on the relative stability of F0 in human speech after puberty (Fouquet, Pisanski, Mathevon, &

Reby, 2016); (iii) postural differences between shot types would affect F0 due to the influence of dynamic biomechanical constraints on vocal production mechanisms (Fitch & Hauser, 1995; Titze, 1994); and (iv) the F0 of vocalisations occurring during match losses would be higher than during match wins. This final prediction stems from the aforementioned negative relationship between F0 and dominance, and evidence that F0 increases under stress, distress, and arousal in both humans (Pisanski, Nowak, et al., 2016) and nonhuman mammals (Briefer, 2012 for review). Finally, we predicted that F0 cues in tennis grunts would influence listeners' attributions of vocalizer sex and match outcome.

## Method

### *1. Analysis of Tennis Grunts*

#### **Within-subject variables**

From the top 30 professional tennis players in the world at the time of data collection, we identified seven males and seven females who consistently grunt when hitting both serves and groundstrokes within and between matches (see Appendix for additional details).

Using PRAAT 5.3.62 DSP package (Boersma & Weenink, 2017) and Boom 2 software (Global Delight Technologies, 2014), we extracted 367 tennis grunts from direct audio output of television footage of 50 matches provided by the International Tennis Federation and the IMG Sport Video Archive. Recordings were saved as WAV files at 44.1 kHz sampling frequency and 16 bits amplitude resolution. For each grunt, we coded shot type (forehand/backhand/serve – the most common shot types, Johnson & McHugh, 2006), the duration of the match at the point of grunt production (hereafter *match*

*progress*, expressed as a percentage of total match duration), and match outcome (win/loss). For each match, two grunts per set were recorded for males, and three for females. This resulted in an equal number of grunts per match for each sex, as women play best-of-three sets while men play best-of-five. Grunts were sampled at equally spaced time intervals across the duration of each set. Within each match, an equal number of forehands, backhands, and serves were recorded. Within each vocaliser, we recorded a roughly equal number of grunts from match wins and losses. Within the constraints of the limited number of televised matches available to us, we matched wins and losses as closely as possible in terms of tournament stage, so as to control for the potential effect of match importance on physiological and psychological state.

### **Between-subject variables**

To test whether player sex, height, weight, and age predicted grunt F0, we conducted an additional between-subject analysis. Because F0 variation between serve vocalisations in the within-subjects dataset was relatively small (see Figure 1a), we randomly selected two serve vocalisations from each of the 14 players described above, and extracted 92 serve vocalisations from an additional 23 male and 23 female professional tennis players (mean age  $\pm$  SD = 25.09  $\pm$  0.42), to achieve an adequate sample size of 30 players per sex and 120 serve vocalisations. We chose serve vocalisations because they are always produced at the start of rallies, from an initially stationary position where posture is most standardised. Player sex, height, weight, and age data were acquired from [www.atpworldtour.com](http://www.atpworldtour.com) and [www.wtatennis.com](http://www.wtatennis.com). We also extracted 10-second pre-match interview clips to measure each player's mean speaking F0.

### **Acoustic analysis**

We extracted mean F0 using a dedicated processing script in PRAAT (Boersma & Weenink, 2014). We systematically inspected each extracted pitch contour and verified it using a narrow band spectrogram displaying the first 2,000 Hz of the signal. Erroneous pitch values (e.g. octave jumps) were manually corrected. Fifty recordings were omitted from the within-subject dataset due to insufficient quality for pitch extraction ( $N = 41$ ) or aperiodicity exceeding 50% of clip duration ( $N = 9$ ). Thus, 317 grunts were used in the final within-subject analysis (mean  $\pm$  SD per player =  $23.29 \pm 0.19$ ). No recordings were omitted from the between-subject analysis. Each extracted F0 contour (pitch object) was used to derive mean F0. This acoustic analysis procedure has been applied successfully in previous studies of a human nonverbal vocalisation (e.g. babies' cries Reby, Levréro, Gustafsson, & Mathevon, 2016).

### **Statistical analysis**

We ran all statistical analyses with IBM SPSS Statistics 22 (IBM Corp., Armonk, NY, U.S.A.). We tested for effects of predictor variables on mean grunt F0 using Linear Mixed Models (LMMs, covariance structure: variance components, restricted maximum likelihood). We expected the distributions of F0 between sexes to show little to no overlap (due to strong sexual dimorphism in F0), and therefore conducted separate analyses for male and female players. For models examining within-subject variables, we calculated AICc values with every possible combination of variables and interactions, and employed model selection to arrive at a best-fitting final model (Symonds & Moussalli, 2011). Further detail on model selection parameters is provided in the Appendix. Model structures detailing fixed and random effects are given in the footnotes of Table 1.

## **2. *Playback Experiments***

We tested whether listeners can accurately gauge sex and match outcome from tennis grunts using a newly recorded set taken from television and YouTube footage of professional matches. We also examined whether natural or experimentally manipulated variation in grunt F0 influenced how listeners attributed sex and match outcome.

### **Participants**

Thirty University of Sussex students (16 male, 14 female, mean  $\pm$  SD age =  $24.97 \pm 9.46$ ) participated in the sex attribution task, 16 of whom were competitive tennis team players. Eighteen competitive tennis players (including the 16 who completed the sex attribution task) participated in the match outcome attribution task (11 males, 7 females, mean  $\pm$  SD age =  $20.89 \pm 2.61$ ).

### **Stimuli and procedure**

Participants completed the playback experiments in a quiet room. Stimuli were presented on a laptop computer via the Experiment Multiple Forced Choice tool in PRAAT and through Sony MDR XB500 headphones at a comfortable pre-set volume, standardised within participants.

For the sex attribution task, participants were presented with 1 grunt per trial from 10 male and 10 female players and asked to identify the sex of the player. Participants were only presented with natural grunts and completed a total of 20 trials.

For the match attribution task, participants were presented with pairs of 6-grunt sequences from 6 male and 6 female players and asked to identify which sequence in the pair came from the match the player lost. Each pair consisted of one sequence of grunts from the end of a match the player won without losing a set (win sequence), and a

sequence of grunts from a match that same player lost without winning a set (loss sequence). Participants completed 12 trials for natural grunt-sequence pairs (though only 11 pairs were included in statistical analysis, see Appendix).

In addition, we resynthesised the F0 of grunt-sequence pairs to create four F0-resynthesis conditions: two in which the loss sequence was higher in F0 than the win sequence, and two in which the win sequence was higher in F0 than the loss sequence. We resynthesised the F0 of entire loss and win grunt sequences by equal amounts, such that the mean F0 difference between grunt-sequences within each pair was equal to the mean F0 difference between loss and win grunts measured from our samples of professional tennis players ( $\pm 27$  Hz and  $\pm 58$  Hz for males, and  $\pm 39$  Hz and  $\pm 100$  Hz for females, to fit the mean difference  $\pm N$  SD, with  $N = 0$  and 1). Two of the 12 grunt-sequences could not be resynthesised because F0 could not be detected. Thus, participants completed 40 trials for resynthesised grunt-sequence pairs (10 pairs  $\times$  4 resynthesis conditions). For additional details see Appendix.

Participants entered ratings by clicking on the chosen button on the screen. They could either confirm their choice (“OK” button), replay the sound (“replay” button), or change their rating (“back” button), before moving on to the next stimulus.

Non-tennis playing participants completed only the sex attribution task. Competitive tennis playing participants completed both the sex attribution task and the two match outcome attribution tasks in separate blocks. As the resynthesised match outcome task stimuli consisted of four repetitions of previously heard stimuli, this task was split into two blocks and pseudo-randomised so that two resyntheses of the same stimulus pair were presented at least six trials apart. These two ‘match outcome’ blocks were separated by the sex attribution task (for all but two listeners who did not complete the sex attribution task), so as to minimise recognition of exemplar repetition. Block order

was counterbalanced, and within blocks, stimulus presentation was randomised and counterbalanced.

### **Statistical analysis**

We used Generalised Linear Mixed Models (GLMMs) with binary logistic regression link in SPSS to test for accuracy in listeners' attributions. We also examined the effects of natural and artificial variation in mean F0, and for the sex attribution task, the effect of level of tennis experience, on these attributions. All GLMMs included listener identity as a subject variable, and the model testing the effect of F0 on match outcome attributions in resynthesised grunts also included actual match outcome as a random factor. Effect sizes were estimated using R coefficients derived from simple linear regressions between the mean F0 (or mean F0 difference) and the ratings averaged by exemplar and listener (sex, match outcome). Listener sex did not significantly predict how listeners attributed sex or match outcome in any model, and therefore was not included in reported models.

### **Ethical note**

The Sciences and Technology Cross-Schools Research Ethics Committee of the University of Sussex reviewed and approved this research (ER/JR307/1).

## **Results**

### **Do tennis grunts contain static and dynamic cues?**

#### *Effects of sex, age, height, weight, and speaking F0 on grunt F0*

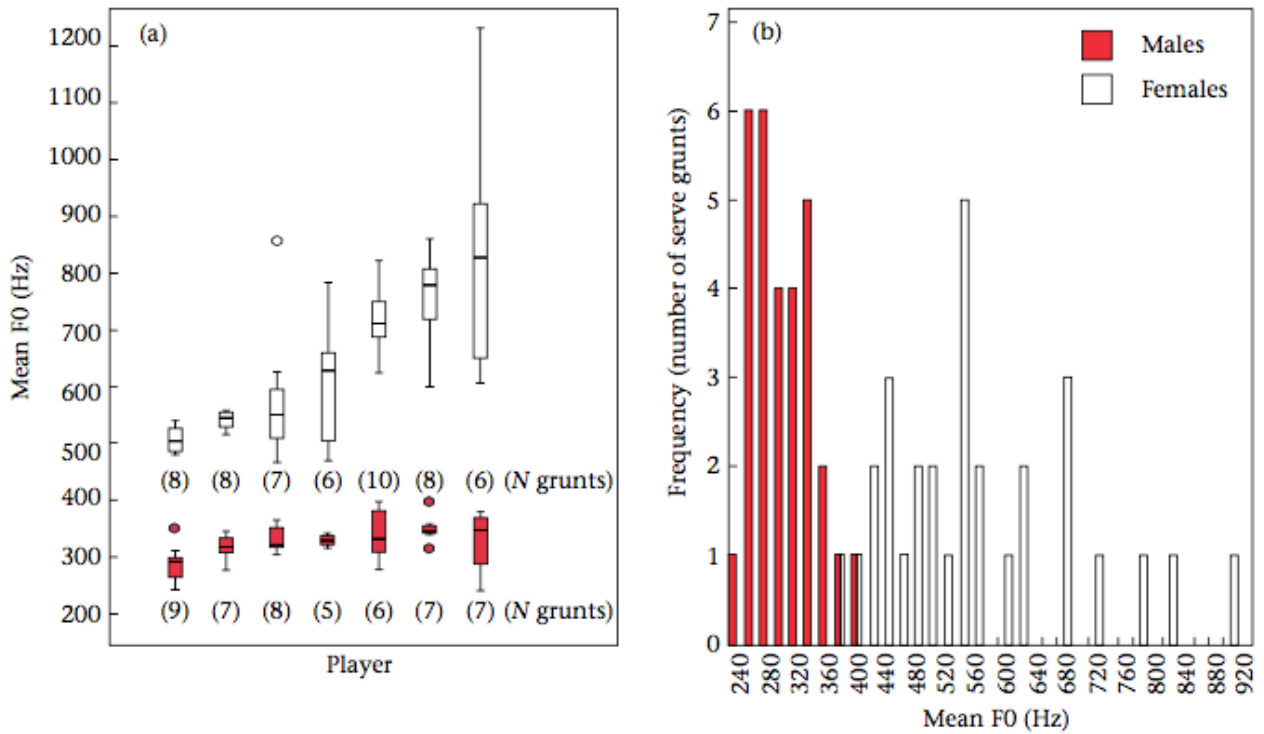
Player sex significantly predicted mean grunt F0 ( $F_{1,58} = 104.73$ ,  $P < 0.001$ ): the mean F0 of female grunts ( $573.9 \pm 24.36$  Hz) was higher than that of male grunts ( $296.9 \pm$

24.46 Hz). F0 variation was greater among females than males (Figure 1a, Levene's test:  $F_{1,100} = 47.87$ ,  $P < 0.001$ ). Figure 1b illustrates that there was very little overlap in the frequency distributions of male and female grunts. Age, height, and weight (examined separately for each sex) did not significantly predict mean grunt F0 (all  $P$ s  $> 0.198$ ), nor did speaking F0 ( $P$ s  $> 0.161$ ), although there was a non-significant trend for female grunt F0 to increase with speaking F0 (Table 1).

*Effects of shot type, match progress, and match outcome on grunt F0*

The effects of shot type, match progress, and match outcome on mean grunt F0 were assessed separately for males and females. For both sexes, shot type predicted mean grunt F0 (males:  $F_{2,145} = 8.12$ ,  $P < 0.001$ , females:  $F_{2,138} = 3.64$ ,  $P = 0.029$ , Table 1). Planned pairwise comparisons indicated that the F0 of grunts accompanying serves (males: mean  $\pm$  SD =  $325.0 \pm 7.3$  Hz, females: mean  $\pm$  SD =  $650.0 \pm 44.1$  Hz) was significantly higher than for forehands (males: mean  $\pm$  SD =  $311.8 \pm 7.0$  Hz,  $p = 0.052$ ; females: mean  $\pm$  SD =  $616.0 \pm 44.0$  Hz,  $P = 0.045$ ) and backhands (males: mean  $\pm$  SD =  $297.5 \pm 7.1$  Hz,  $P < 0.001$ ; females: mean  $\pm$  SD =  $606.1 \pm 44.1$  Hz,  $P = 0.011$ ).

For both sexes, grunt F0 increased as the match progressed (males:  $F_{1,146} = 6.70$ ,  $P = 0.011$ , females:  $F_{1,140} = 4.19$ ,  $P = 0.043$ , Table 1). To examine this effect further we categorised grunts according to which third of the match they were produced in. Male grunt F0 was significantly higher in the last third of matches (mean  $\pm$  SD =  $320.87 \pm 7.14$  Hz) than in the first third (mean  $\pm$  SD =  $305.89 \pm 7.32$  Hz,  $P = 0.027$ ). There was a similar but non-significant trend among female players (last third:  $633.70 \pm 44.06$  Hz; first third:  $602.36 \pm 44.26$  Hz,  $P = 0.059$ ).



*Figure 1.* (a) Frequency distribution of mean F0 for grunts accompanying serves: male  $N = 7$ , female  $N = 7$  players.  $N$  of grunts per player is displayed beneath each box plot. Box plots represent the distribution of serve grunt F0 for each recorded player, showing that in general, variation in serve F0 was small, and that variation was greater within and across females than males. Central lines within boxes represent the median; box edges represent the interquartile range. For boxes without outliers (represented by circles), whiskers represent minima and maxima. For boxes with outliers, whiskers represent the highest and lowest cases within 1.5 times the interquartile range, and circles represent cases lying outside this range. (b) Frequency distribution of mean F0 for grunts accompanying serves. Male  $N = 30$ , female  $N = 30$  players.  $N$  grunts per player = 2. The distributions of grunt F0 differ significantly between sexes.

*Table 1.* Correlates of grunt F0: linear mixed models (LMMs) testing the main effects of height, weight, age, speaking F0, shot type, match outcome, and match progress on professional tennis players' mean grunt F0. Model 1 (30 males, 30 females; 2 grunts per player): LMM included player identity as subject variable. Height and weight were strongly correlated in males ( $r = 0.59$ ) and females ( $r = 0.53$ ). Model 2 (7 females, mean grunts per player: 23.29, 7 males, mean grunts per player: 23.29): LMM included player identity as subject variable, and the match identity as a random factor.

Model	Males			Females		
	$df_1, df_2$	$F$	$P$	$df_1, df_2$	$F$	$P$
<b><i>Model 1: Effect of between-subjects predictors on mean grunt F0 for each sex</i></b>						
Intercept	1, 26.7	0.01	0.921	1, 25.4	0.01	0.928
Height	1, 26.3	0.54	0.468	1, 25.1	0.94	0.341
Weight	1, 26.3	0.18	0.674	1, 25.2	1.75	0.198
Age	1, 29.0	0.89	0.352	1, 28.0	0.17	0.688
Speaking F0	1, 24.9	1.15	0.295	1, 25.3	2.09	0.161
<b><i>Model 2: Effect of within-subjects predictors on mean grunt F0 for each sex</i></b>						
Intercept	1, 17.0	1435.57	<b>&lt; 0.001</b>	1, 6.9	180.15	<b>&lt; 0.001</b>
Shot type	2, 145.2	8.12	<b>&lt; 0.001</b>	2, 138.4	3.64	<b>0.029</b>
Match outcome	1, 16.5	7.53	<b>0.014</b>	1, 12.3	3.07	0.105
Match progress	1, 145.8	6.70	<b>0.011</b>	1, 139.6	4.19	<b>0.043</b>

Among males, grunt F0 was higher during losses (mean  $\pm$  SD = 319.97  $\pm$  6.97 Hz) than wins (mean  $\pm$  SD = 302.97  $\pm$  6.65 Hz,  $F_{1,17} = 7.53$ ,  $P = 0.014$ ). Among female players, grunt F0 was not significantly higher during losses (mean  $\pm$  SD = 640.53  $\pm$  44.18 Hz) than wins (mean  $\pm$  SD = 607.65  $\pm$  43.87 Hz,  $F_{1,12} = 3.07$ ,  $P = 0.105$ , Table 1). Following AICc comparison for model selection (Symonds & Moussalli, 2011), our final reported models were 2.94 times (males) and 2.31 times (females) more likely to be the best approximating model than the models with the next lowest AICc values, which included the non-significant interaction between match outcome and match progress for

both sexes ( $P_s > 0.451$ ). The final model showed that the difference in F0 between losses and wins in male players emerged early in matches, and the magnitude of this difference did not change as the match progressed; F0 increased with match progress for both winners and losers by an equal degree.

### **Can listeners gauge sex and match outcome from tennis grunts?**

#### *Attribution of sex by listeners*

Acoustic analysis of the playback stimuli confirmed that player sex significantly predicted mean grunt F0 ( $F_{1,18} = 475.88$ ,  $P < 0.001$ ). Males produced grunts with lower F0 (mean  $\pm$  SD =  $329 \pm 27.70$  Hz) than did females (mean  $\pm$  SD =  $525.59 \pm 27.70$  Hz).

Playback experiments showed that listeners correctly identified the sex of players from their grunts ( $F_{1,598} = 256.53$ ,  $P < 0.001$ ). Listeners correctly attributed the sex of the player for 95% of male and 91% of female grunts. All female exemplars were identified with at least 93% accuracy except for two females with relatively low grunt F0, who were identified with 57% and 67% accuracy. All male exemplars were identified with 87% accuracy or above. Competitive tennis players attributed sex with a higher degree of accuracy (mean  $\pm$  SD =  $96 \pm 1\%$ ) than did non-playing participants (mean  $\pm$  SD =  $90 \pm 2\%$ ,  $F_{1,598} = 7.32$ ,  $P = 0.007$ ).

Grunts with higher F0 were significantly more likely to be identified as female, and those with lower F0 as male ( $F_{1,598} = 162.94$ ,  $P < 0.001$ ), even within sexes for females ( $F_{2,597} = 119.78$ ,  $P < 0.001$ ). This effect was large collapsing across sexes ( $R = 0.81$ ), and within females ( $R = 0.50$ ).

*Attribution of match outcome by listeners*

Acoustic analysis of the playback stimuli showed that match outcome significantly predicted mean grunt F0 in both males ( $F_{1,15} = 10.85$ ,  $P = 0.005$ ) and females ( $F_{1,15} = 5.65$ ,  $P = 0.031$ ). Grunt F0 was higher during lost than won matches for both males (mean  $\pm$  SD diff =  $26.68 \pm 8.10$  Hz) and females (mean  $\pm$  SD diff =  $39.08 \pm 16.44$  Hz).

In playback experiments involving natural grunt sequences, actual match outcome significantly predicted how often listeners accurately choose the match that the player lost ( $F_{1,196} = 8.63$ ,  $P = 0.004$ ). The mean percentage of correct identification ( $61\% \pm 3.5\%$ ) was marginally higher than chance (50%). Furthermore, the relatively higher F0 grunt sequence was more likely to be attributed as coming from a loss ( $F_{1,196} = 25.41$ ,  $P < 0.001$ ), and the higher the F0 of the loss sequence relative to the win sequence, the more accurately listeners attributed match outcome ( $F_{1,196} = 17.06$ ,  $P < 0.001$ , Figure 2a). These effects of F0 were large (both  $R_s = 0.71$ ). In playbacks involving resynthesised grunt sequences, relatively higher F0 grunt sequences were significantly more likely to be attributed as coming from losses than were relatively lower F0 sequences ( $F_{1,718} = 16.29$ ,  $P < 0.001$ , Figure 2b). This effect was also large ( $R = 0.51$ ).

**Reanalysis using ERBs**

We re-ran our acoustic analyses using equivalent rectangular bandwidths (ERBs), a semilogarithmic scale that controls for the nonlinear relationship between F0 and perceived pitch. The significance of all statistical tests remained unchanged, and therefore these statistics are not reported.

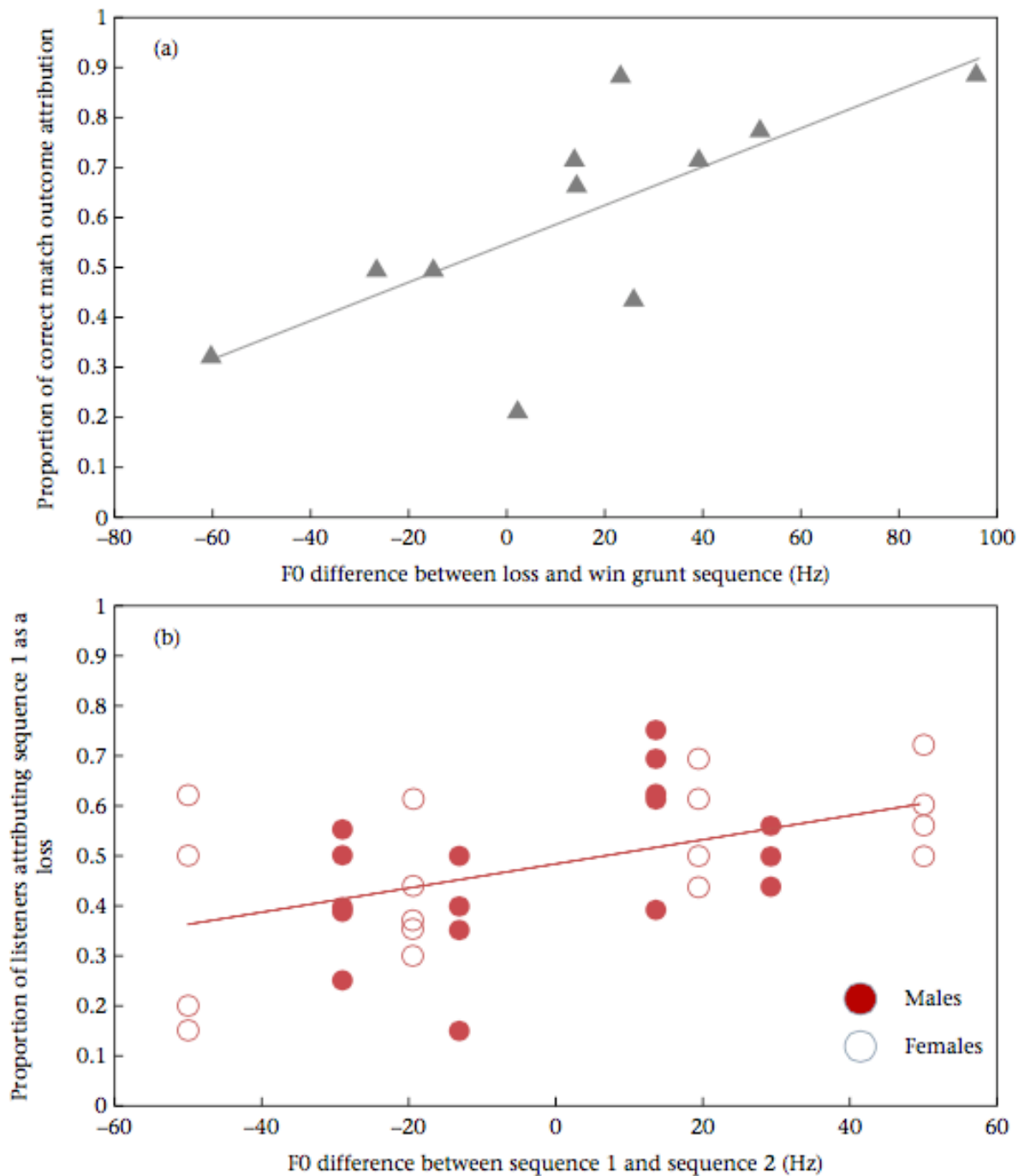


Figure 2. Attribution of match outcome by adult listeners as a function of natural and manipulated F0 differences. (a) Triangles represent the average probability that listeners (male  $N = 11$ , female  $N = 7$  listeners) correctly identified which of two natural six-grunt sequences with different F0 profiles came from a match the player lost (11 pairs (win/loss), male  $N = 6$ , female  $N = 5$  players). Positive x-axis values indicate that the F0 of the loss sequence was higher. (b) Dots represent the average probability that listeners (male  $N = 11$ , female  $N = 7$  listeners) presented with pairs of six-grunt sequences attributed the sequence presented first as coming from a match loss (10 pairs, male  $N = 5$ , female  $N = 5$  players), for each of four resynthesis variants of the mean F0 difference between sequences. Positive x-axis values indicate that sequence 1's F0 was higher.

## Discussion

Our results show that tennis grunts contain static cues to speaker sex, and dynamic cues to tennis shot type, match progress and contest outcome. Female grunts, serve grunts, and loss grunts were characterised by a higher F0 (voice pitch) than were male, groundstroke, and win grunts, respectively, supporting our predictions. We also found that listeners could accurately judge player sex and match outcome, with F0 cues strongly predicting listeners' categorisations. These findings are consistent with existing literature describing the influence and functional relevance of sex, dominance, muscular control, and arousal on F0 production in humans and other mammals (Taylor et al., 2016).

### *Cues to sex, but not height, weight, or age in tennis grunts*

The mean F0 of tennis serve grunts (females: 574 Hz, males: 297 Hz) was three times higher than in modal human speech (i.e. speaking F0) for both males (120 Hz) and females (210 Hz) (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). Nevertheless, we found that sexual dimorphism in tennis grunts (a ratio of 1:1.9) is roughly the same as that previously observed for human speech (1:1.8, Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). Sexual dimorphism in adult laughter also varies around this ratio (Bachorowski, Smoski, & Owren, 2001; Szameitat, Darwin, Szameitat, Wildgruber, & Alter, 2011), indicating that the constraints imposed by sex differences in vocal fold dimensions on the F0 of human speech extend to adult nonverbal vocalisations.

We found that listeners accurately attributed the sex of tennis players. Although competitive tennis players attributed sex more accurately than did non-playing listeners, possibly because they may have recognised the identity of some players, the difference in accuracy was small (6%) and both groups correctly identified sex with very few errors.

Higher F0 grunts were more likely to be perceived as female, as is the case in adult human speech (Hillenbrand & Clark, 2009), and babies' cries (despite no sex differences in F0) (Reby et al., 2016). However, other acoustic characteristics known to be sexually dimorphic in speech signals, such as vocal tract resonances (Titze, 1994) and the variability of the pitch contour (Puts, Apicella, & Cárdenas, 2012), may also have helped listeners to accurately attribute the sex of players.

In concordance with widespread evidence that F0 is a poor predictor of body size when controlling for sex and age in animal vocalisations (Taylor et al., 2016) and human speech (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014), neither height nor weight predicted mean grunt F0 in our study. This lack of covariation has been attributed to the fact that vocal fold growth is not tightly constrained by skull and body dimensions (Lieberman, McCarthy, Hiimae, & Palmer, 2001), thus the relationship between body size and vocal fold size – and therefore F0 – does not generate reliable information on inter-individual variation in body size (Pisanski, Fraccaro, Tigue, O'Connor, Röder, et al., 2014). Between-individual differences in grunt F0 may however covary with between-individual differences in androgen levels. Indeed, adult men with higher circulating levels of testosterone speak with a lower modal F0 than do men with lower levels of testosterone (Cartei, Bond, & Reby, 2014; Dabbs Jr. & Mallinger, 1999). While we did not have access to hormonal data, future studies should investigate the influence of androgens on F0 in nonverbal vocalisations.

Age did not predict grunt F0 in our sample of adult players, aged 16 to 35. Due to overall body growth (and differential vocal fold growth in males during puberty), the F0 of children's voices is much higher than that of adults in both modal speech (e.g. Peterson & Barney, 1952) and laughter (Bachorowski et al., 2001; Hudenko, Stone, & Bachorowski, 2009). However, despite sizeable between-individual differences in F0

among men, longitudinal observations show that within individuals, F0 remains largely stable after puberty and throughout men's adulthood, at least until about age 60 (Fouquet et al., 2016), consistent with our findings.

Speaking F0 did not reliably predict grunt F0. Thus, while F0 reliably indicates sex in both speech and grunts, there is a decoupling of inter-individual F0 variation between modal speech and this vocalisation. This decoupling suggests that the biomechanical constraints affecting inter-individual differences in F0 differ between these two modes of vocal production.

#### *Acoustic cues to shot type*

Grunt F0 differed across shot types. The observation that serves were characterised by a significantly higher F0 than forehands and backhands may be attributed to biomechanics: groundstrokes involve more pelvic rotation and lower limb drive, whereas serves involve powerful contractions of the abdominal muscles to facilitate the shoulder-over-shoulder trunk flexion that accelerates the body before ball impact (Elliott, 2006). As such, abdominal muscle activity is higher during serves than groundstrokes (Chow, Park, & Tillman, 2009; Knudson & Blackwell, 2000). Contraction of the abdominal muscles is known to produce an increase in subglottic air pressure – for example, singers actively manipulate their abdominal muscles to precisely control subglottic pressure (e.g. Sundberg, Leanderson, von Euler, & Knutsson, 1991). Given that F0 increases with subglottic pressure during speech (Plant & Younger, 2000) and singing (Sundberg, Titze, & Scherer, 1993), the increased abdominal muscle contraction required for serving may lead to the production of grunts characterised by a higher F0.

*Acoustic cues to tennis match progress and outcome*

Our acoustic analyses showed that both male and female players produced grunts with a lower F0 at the beginning than at the end of matches, consistent with evidence that F0 increases under moderate physical load, and again prior to exhaustion (Johannes et al., 2007). Professional players tire as matches progress (Reid & Duffield, 2014), but maintain constant serve speed and accuracy over the duration of long matches (Maquirriain, Baglione, & Cardey, 2016). Thus, the degradation of shot technique (e.g. lower height of ball toss for the serve) may be compensated for by an increase in muscle activation to achieve the same kinetic force (Kibler, 2014). This increased physical effort may be accompanied by increased vocal effort when grunting, which would raise F0 (Lagier et al., 2010). Players are also likely to experience greater physiological arousal and stress in later stages of matches as, being more proximal to the outcome of the match, individual points become more important in determining the winner. As F0 positively correlates with arousal and stress (Briefer, 2012; Pisanski, Nowak, et al., 2016), including psychological stress elicited during short-term physical tasks (Wittels, Johannes, Enne, Kirsch, & Gunga, 2002), this may contribute to the rise in F0 with match progress.

We also found that, independent of match progress, male players produced grunts with overall higher F0 in losses than wins, by an average of 17 Hz or roughly one semitone. This is consistent with observations that F0 increases in response to stressors in both humans and nonhuman mammals (e.g. Briefer, 2012; Pisanski, Nowak, et al., 2016; Wittels, Johannes, Enne, Kirsch, & Gunga, 2002). Men's testosterone levels, which correlate negatively with F0 (Cartei et al., 2014; Dabbs Jr. & Mallinger, 1999), are also higher in winners than losers of competitive encounters (Booth, Shelley, Mazur, Tharp, & Kittok, 1989; Campbell, Orourke, & Rabow, 1988; Elias, 1981; Mazur & Lamb, 1980), and dynamic F0 shifts depend on perceived dominance in dyadic competitive scenarios

(Puts, Gaulin, & Verdolini, 2006). In addition, losing in competitive sport is stressful (Scanlan & Passer, 1977) and may result in increased physical and vocal effort during the match in attempts to avoid defeat, which, as discussed previously, may also contribute to the F0 difference between grunts occurring in wins and losses.

The effect of match outcome on mean F0 was marginally non-significant in female players. This may be partially attributed to the larger intra-individual variance in female than male grunt F0 (Figure 1a), and to our modest sample size. This result may also reflect humans' polygynous evolutionary history (Puts, 2016), whereby the emphasis on male competition and female choice results in greater sexual dimorphism through greater pressures for acoustic communication of dominance and formidability in males than in females (Puts et al., 2016). Nevertheless, we observed a similar trend in females, with grunts during match losses on average 33 Hz higher in F0 than those occurring during wins, a difference perceptually equivalent to that found in males (roughly one semitone).

Crucially, the average intra-individual difference in F0 between grunts produced in lost and won matches did not vary significantly as matches progressed. Thus, while grunt F0 dynamically varies within matches irrespective of match outcome, likely as a function of fatigue and/or arousal, players consistently grunt at a relatively higher F0 in lost than won matches, even at the start of matches. This suggests that rather than dynamic shifts due to short-term scoreboard dominance, the difference in F0 between male winners and losers may reflect longer-term physiological and/or psychological factors that may manifest even before the commencement of a match. For example, head-to-head record (i.e. the outcome of all previous matches between the two players), the outcome of the previous encounter, current form, world ranking, and physical condition may influence self-perceived dominance and/or stress, and affect grunt F0 from the beginning of matches. Future research could examine the influence of these factors.

Playback experiments revealed that listeners with experience of competitive tennis could identify which of two grunt sequences produced by the same player came from a match that the player lost. The F0 difference between these sequences predicted the accuracy with which listeners attributed match outcome, and systematic F0 resynthesis of grunt sequences further confirmed that F0 influenced match outcome judgments. These results are consistent with the sensitivity of male fallow deer to F0 resynthesis of competitive calls of other males (indicating dominance)(Pitcher, Briefer, & McElligott, 2015), and suggest that tennis grunts, like visual nonverbal cues (basketball, table tennis Furley & Schweizer, 2014b), provide functional cues that allow human receivers to infer contest outcome.

## **Conclusion**

Our results indicate that the acoustic structure of tennis grunts, like nonhuman mammal calls, contains perceptible static and dynamic cues to anatomical and physiological attributes of the speaker. In future work researchers may examine the functional relevance of these cues by testing how they affect tennis players' behaviour and internal state. For example, acoustic cues to contest outcome in tennis grunts may influence opposing players' confidence, as is the case with visual nonverbal cues (Furley & Schweizer, 2014a). Furthermore, although our sample included players who grunt sufficiently frequently to allow for behavioural observation (see Appendix), future work may investigate predictors of variation in the occurrence of grunting both within- and between-players to more fully understand the mechanisms and functions of tennis grunts. Finally, researchers may also examine whether, in addition to grunts, other types of human vocalisations such as aggressive roars, fear screams, and sexual vocalisations convey evolutionarily important information (Chapters 2, 3 & 4). Such vocalisations constitute an

intermediary link between nonhuman mammal vocalisations and human speech, and investigating their production and perception may provide additional insight into the evolution of human vocal communication and vocal control, a necessary prerequisite for speech (Pisanski, Cartei, et al., 2016).

**Data availability**

The dataset supporting this article has been uploaded to the Sussex Research Online (SRO) repository (Item #68818, <http://sro.sussex.ac.uk/68818/>).

## **Chapter 6: General discussion**

While it is well understood that the nonverbal components of human speech communicate a wide variety of indexical information, in continuity with nonhuman mammal vocalisations, the nature and function of human nonverbal vocalisations from an evolutionary perspective has until recently been largely overlooked. Previous investigations of laughter (e.g. Bachorowski et al., 2001; Bryant et al., 2016; Bryant & Aktipis, 2014; Lavan et al., 2015) and infant distress cries (e.g. Kelly et al., 2017; Koutseff et al., 2017; Lingle & Riede, 2014) indicated that the nonverbal vocalisations of humans and nonhuman mammals are likely mediated by shared evolutionary mechanisms. However, examination of the acoustic structure and perceptual relevance of a wider range of vocalisations is necessary to more comprehensively understand their potential origins and functions.

Throughout this thesis, I provide evidence that human nonverbal vocalisations are functionally and structurally homologous to nonhuman mammals calls, communicating honest anatomical and physiological cues relevant to natural and sexual selection processes, but also functioning as an adaptive social tool with which to motivationally influence listeners' perceptions. Furthermore, I argue that volitional modulation and simulation of these vocalisations may have been at the origins of selection for humans' uniquely advanced vocal control, paving the way for the emergence of full-blown articulated speech.

### **Human and nonhuman vocalisations: structural and functional homology**

The preceding chapters provide strong evidence for similarities between human and nonhuman nonverbal vocalisations in both the acoustic encoding and perception of

indexical and motivational information. Over the course of the thesis, I demonstrate commonalities between human and nonhuman mammal vocalisations in the acoustic expression of indexical traits (sex, body size and strength) and motivational or affective states (arousal, pain, aggression, and distress). Critically, the vocalisations studied in this thesis are not characterised by arbitrary acoustic structures, but instead, largely align with the evolutionarily grounded motivational-structural rules that govern the spectrotemporal structure and communication of vocal signals in nonhuman mammals (Morton, 1977; Ohala, 1984; Owings & Morton, 1998). Indeed, Chapter 2 demonstrates that humans are capable of conveying aggression or distress in at least two distinct types of vocalisations, characterised by distinct acoustic structures that also reflect the typical characteristics of aggressive and distress vocalisations in other mammals. In Chapters 4 and 5, I show that even within a single call type, acoustic variation corresponds with these rules – for example, F0 was higher in pain vocalisations simulating higher levels of pain, and in spontaneous tennis grunts produced during losses relative to wins.

While we could not corroborate the role of formants in expressing formidability in aggressive roars (a point discussed further in the following section), I show that these vocalisations not only communicate inter-individual variation in both absolute and relative strength, but also convey exaggerated formidability to listeners relative to distress screams and speech (Chapters 2 and 3). The combined indexical and motivational signalling functions of roars strongly parallels the communication and exaggeration of formidability cues in aggressive vocalisations produced by other mammals (Charlton, Ellis, et al., 2011; Charlton, Whisson, et al., 2013; Fitch & Hauser, 2003; Reby et al., 2005; Reby & McComb, 2003). Moreover, listeners are able to infer contest outcome from competitive tennis grunts (Chapter 5), in continuity with the key role that many nonhuman mammal agonistic vocalisations play in mediating agonistic interactions in male competition

contexts (giant pandas: Charlton et al., 2010; sea lions: Charrier et al., 2011; fallow deer: Pitcher et al., 2015; red deer: Reby et al., 2005; domestic dogs: A. M. Taylor et al., 2010). Further research is now required to assess whether contest outcome can also be inferred from formidability cues in competitive vocalisations produced in contexts involving physical combat, such as martial arts kiaps (Welch & Tschampl, 2012). Finally, I show that acoustic variation in simulated pain vocalisations influences human listeners' perceptions of pain, in a manner consistent with the effect of acoustic characteristics on the urgency of nonhuman mammals' responses to distress cries (Chapter 4, Jovanovic & Gouzoules, 2001; Weary et al., 1996). Overall, the research presented in this thesis strongly indicates that human nonverbal vocalisations are homologous to nonhuman mammal calls in both structure and function.

Furthermore, my findings suggest that human nonverbal vocalisations are better suited to communicate evolutionarily relevant information than is articulated speech. In Chapters 2 and 3, I directly compared the signalling of formidability in nonverbal vocalisations and speech equivalents, finding that roars and screams are distinguished by a wider acoustic space than are aggressive and distressed speech, and that relative to aggressive speech, roars communicate absolute formidability more reliably, while also exaggerating absolute and relative cues to formidability. Thus, these results support the hypothesis that compared to nonverbal vocalisations, the acoustic constraints placed on the speech channel to communicate arbitrarily determined semantic information obfuscate the effective signalling of indexical and motivational cues. In contrast, the communication of strength cues in roars - but not screams (Chapter 2) - and the considerable explanatory power of acoustic variation in pain vocalisations (both in discriminating pain levels and predicting pain ratings, Chapter 4) indicate that the production and perception of nonverbal vocalisations are not arbitrary, but instead grounded in basic, conserved

relationships between acoustic characteristics and anatomical or physiological attributes of the vocaliser.

To further investigate this conclusion, future research should build on recent work demonstrating cross-cultural commonalities in the perception of laughter and nonverbal vocalisations expressing the ‘basic emotions’ (Bryant et al., 2016; Sauter, 2010; Sauter, Eisner, Ekman, et al., 2010), to examine whether the observed functional relevance of roars, screams, and grunts generalises to non-Western cultures. Researchers should also examine whether functional homology between humans and other mammals extends to other nonverbal vocalisations. Nevertheless, my findings clearly indicate that nonverbal vocalisations, like nonhuman mammal calls, are a highly effective medium for the communication of evolutionarily relevant information.

### **Volitional production of vocalisations: the origins of speech capabilities?**

While this thesis demonstrates that human and nonhuman mammal nonverbal vocalisations share continuity in form and function, they differ in one key respect: humans can produce and modulate vocalisations in the complete absence of an associated inducing experience or state (Pisanski, Cartei, et al., 2016). This uniquely advanced capacity for vocal control highlights a fundamental trade-off between the honest communication of indexical cues, and motivational signalling to manipulate the expression of those cues. Primary conserved relationships between acoustic characteristics and static physical attributes may be co-opted to imitate aggressive or submissive motivation, thus exaggerating (roars), minimising, or masking (screams) the expression of anatomical indexical cues (Chapters 2 and 3; Morton, 1977; Ohala, 1984; Owings & Morton, 1998; Pisanski, Fraccaro, Tigue, O’Connor, & Feinberg, 2014; Puts et al., 2007; Rendall et al.,

2007; D. R. Smith & Patterson, 2005). Moreover, relationships between dynamic physiological state and acoustic parameters may be exploited to communicate motivational states such as pain, in the absence of authentic nociceptive input (Chapter 4).

Importantly, the observed lack of consistent acoustic cues to strength or body size (Chapter 2) suggests that variation in individuals' capacity to modulate their voices may decouple relationships between physical attributes and acoustic characteristics at the level of production. Yet, the fact that listeners were still able to reliably estimate vocalisers' anatomical characteristics when listening to aggressive stimuli suggests that listeners may be able to detect volitional voice modulation, and at least partially correct their attributions accordingly. This interpretation is consistent with previous evidence that listeners can perceive differences in authenticity between spontaneous and volitional nonverbal vocalisations (Anikin & Lima, 2017; Bryant & Aktipis, 2014; Lavan et al., 2015), and with evolutionary accounts of vocal signalling contending that emitters should evolve strategies to manipulate receivers, while receivers should evolve to detect and resist manipulation (Knight, 1998; Krebs & Dawkins, 1984; Oesch, 2016). Additional work is now needed to establish to what extent listeners can disentangle the relative contributions of honest cues to indexical and motivational state, and volitional modulation of these cues. In particular, future research may benefit from simultaneously assessing listeners' attributions of vocalisers' anatomical and physiological state (e.g. strength/pain), motivational intentions (e.g. aggression/submission/urgency), and authenticity, to examine how these attributions correspond with acoustic variation in vocalisations, and the accuracy of listeners' estimations of indexical characteristics.

While I recruited acting students to produce the volitional vocalisations examined in this thesis, evidence suggests that the capacity to volitionally simulate or modulate vocalisations to exaggerate or minimise the expression of pain, formidability, and other

cues is not limited to those with acting training (R. Jürgens et al., 2015; Pisanski, Mora, et al., 2016). Indeed, government organisations recommend the volitional production of roars as a predator-deterrent strategy (United States National Park Service, n.d.), and context-, mood-, and cognition-dependent modulations are likely an important component of the vocal communication of spontaneous pain (see Tracey & Mantyh, 2007). Functional vocal deception (see Oesch, 2016) and body size exaggeration (de Boer et al., 2015; Hardus et al., 2009; Reby & McComb, 2003; see Pisanski, Cartei, et al., 2016) have even been observed in primitive but not volitional form in a number of nonhuman mammal species, while recent evidence indicates that nonhuman primates are capable of more voluntary vocal flexibility (e.g. Lameira et al., 2015; Perlman & Clark, 2015; Schel et al., 2013; see Pisanski, Cartei, et al., 2016 for a review) than previously assumed.

The volitional modulation and simulation of nonverbal vocalisations to adaptively influence listeners' perceptions may represent a key intermediary step between limited voiced modulation in nonhuman mammals, and articulated human speech (Brown, 2017; Oesch, 2016; Pisanski, Cartei, et al., 2016). This step may involve the emergence of monosynaptic innervation pathways between the motor cortex and vocal motor neurons, now believed to be a crucial early prerequisite of language capabilities (the Küypers/Jürgens hypothesis of speech motor control, Ackermann et al., 2014; Feierman, 2017; Fitch, 2010; U. Jürgens, 1994; Küypers, 1958). Indeed, the species-atypical vocal modulation capabilities of enculturated apes raises the possibility that such neural connections are at least partly experience-dependent (Adkins, Boychuk, Remple, & Kleim, 2006; Holtmaat & Svoboda, 2009; see Pisanski, Cartei, et al., 2016); investigating the neural substrates underpinning these ape's enhanced vocal control may provide crucial insight into the evolutionary pathway leading to articulated speech. Early volitional modulation of nonverbal vocalisations could have marked the first intentional departure

from honest, biologically grounded communication of indexical attributes, paving the way for an increasingly flexible and arbitrary relationship between form and function, ultimately culminating in the emergence of complex speech.

## References

- Abari, K., Rácz, Z. Z., & Olaszy, G. (2011). Formant maps in Hungarian vowels - online data inventory for research, and education. In *Proceedings of Interspeech 2011* (pp. 1609–1612). Florence, Italy. Retrieved from [https://www.researchgate.net/profile/Gabor\\_Olaszy/publication/221484320\\_Formant\\_Maps\\_in\\_Hungarian\\_Vowels\\_Online\\_Data\\_Inventory\\_for\\_Research\\_and\\_Education/links/0046351b842a46eb95000000.pdf](https://www.researchgate.net/profile/Gabor_Olaszy/publication/221484320_Formant_Maps_in_Hungarian_Vowels_Online_Data_Inventory_for_Research_and_Education/links/0046351b842a46eb95000000.pdf)
- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice*, 13(3), 424–446. [https://doi.org/10.1016/S0892-1997\(99\)80048-4](https://doi.org/10.1016/S0892-1997(99)80048-4)
- Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *The Behavioral and Brain Sciences*, 37(6), 529–546. <https://doi.org/10.1017/S0140525X13003099>
- Adkins, D. L., Boychuk, J., Remple, M. S., & Kleim, J. A. (2006). Motor training induces experience-specific patterns of plasticity across motor cortex and spinal cord. *Journal of Applied Physiology*, 101(6), 1776–1782. <https://doi.org/10.1152/jappphysiol.00515.2006>
- Andersson, M. B. (1994). *Sexual selection*. Princeton, NJ: Princeton University Press.
- Anikin, A., & Lima, C. F. (2017). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *Quarterly Journal of Experimental Psychology*, 1–21. <https://doi.org/10.1080/17470218.2016.1270976>
- Anikin, A., & Persson, T. (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods*, 49(2), 758–771. <https://doi.org/10.3758/s13428-016-0736-y>
- Anolli, L., & Ciceri, R. (2002). Analysis of the vocal profiles of male seduction: From exhibition to self-disclosure. *The Journal of General Psychology*, 129(2), 149–169. <https://doi.org/10.1080/00221300209603135>
- Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter–gatherers. *Proceedings of the Royal Society of London B*:

- Biological Sciences*, 276(1659), 1077–1082.  
<https://doi.org/10.1098/rspb.2008.1542>
- Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters*, 3(6), 682–684.  
<https://doi.org/10.1098/rsbl.2007.0410>
- Archer, J. (2009). Does sexual selection explain human sex differences in aggression? *Behavioral and Brain Sciences*, 32(3–4), 249–266.  
<https://doi.org/10.1017/S0140525X09990951>
- Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, 25(15), 2051–2056. <https://doi.org/10.1016/j.cub.2015.06.043>
- Arnoldner, C., Kaider, A., & Hamzavi, J. (2006). The role of intensity upon pitch perception in cochlear implant recipients. *The Laryngoscope*, 116(10), 1760–1765.  
<https://doi.org/10.1097/01.mlg.0000228214.02606.42>
- August, P. V., & Anderson, J. G. T. (1987). Mammal sounds and motivation-structural rules: A test of the hypothesis. *Journal of Mammalogy*, 68(1), 1–9.  
<https://doi.org/10.2307/1381039>
- Babel, M., McGuire, G., & King, J. (2014). Towards a more nuanced view of vocal attractiveness. *PLOS ONE*, 9(2), e88616.  
<https://doi.org/10.1371/journal.pone.0088616>
- Bachorowski, J. A., & Owren, M. J. (2001). Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science*, 12(3), 252–257. <https://doi.org/10.1111/1467-9280.00346>
- Bachorowski, J. A., Smoski, M. J., & Owren, M. J. (2001). The acoustic features of human laughter. *The Journal of the Acoustical Society of America*, 110(3), 1581–1597. <https://doi.org/10.1121/1.1391244>
- Backstrom, T., Alku, P., & Vilkman, E. (2002). Time-domain parameterization of the closing phase of glottal airflow waveform from voices over a large intensity range. *IEEE Transactions on Speech and Audio Processing*, 10(3), 186–192.  
<https://doi.org/10.1109/TSA.2002.1001983>
- Baeck, H. E., & Souza, M. N. (2001). Study of acoustic features of newborn cries that correlate with the context. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Vol. 3, pp. 2174–2177). <https://doi.org/10.1109/IEMBS.2001.1017201>

- Baker, A., & Kenner, A. N. (1993). Communication of pain: Vocalization as an indicator of the stage of labour. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 33(4), 384–385. <https://doi.org/10.1111/j.1479-828X.1993.tb02115.x>
- Balint, A., Farago, T., Miklosi, A., & Pongracz, P. (2016). Threat-level-dependent manipulation of signaled body size: dog growls' indexical cues depend on the different levels of potential danger. *Animal Cognition*, 19(6), 1115–1131. <https://doi.org/10.1007/s10071-016-1019-9>
- Banai, I. P. (2017). Voice in different phases of menstrual cycle among naturally cycling women and users of hormonal contraceptives. *Plos One*, 12(8), e0183462. <https://doi.org/10.1371/journal.pone.0183462>
- Banai, I. P., Banai, B., & Bovan, K. (2017). Vocal characteristics of presidential candidates can predict the outcome of actual elections. *Evolution and Human Behavior*, 38(3), 309–314. <https://doi.org/10.1016/j.evolhumbehav.2016.10.012>
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Bänziger, T., Hosoya, G., & Scherer, K. R. (2015). Path models of vocal emotion communication. *PLOS ONE*, 10(9), e0136675. <https://doi.org/10.1371/journal.pone.0136675>
- Barelli, C., Mundry, R., Heistermann, M., & Hammerschmidt, K. (2013). Cues to androgens and quality in male gibbon songs. *PLOS ONE*, 8(12), e82748. <https://doi.org/10.1371/journal.pone.0082748>
- Barreda, S. (2016). Investigating the use of formant frequencies in listener judgments of speaker size. *Journal of Phonetics*, 55(Supplement C), 1–18. <https://doi.org/10.1016/j.wocn.2015.11.004>
- Barreda, S. (2017a). An investigation of the systematic use of spectral information in the determination of apparent-talker height. *Journal of the Acoustical Society of America*, 141(6), 4781–4792. <https://doi.org/10.1121/1.4985192>
- Barreda, S. (2017b). Listeners respond to phoneme-specific spectral information when assessing speaker size from speech. *Journal of Phonetics*, 63, 1–18. <https://doi.org/10.1016/j.wocn.2017.03.002>
- Bars, D. L., Gozariu, M., & Cadden, S. W. (2001). Animal models of nociception. *Pharmacological Reviews*, 53(4), 597–652.

- Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research PRPF*, 74(1), 110. <https://doi.org/10.1007/s00426-008-0185-z>
- Beckford, N. S., Schaid, D., Rood, S. R., & Schanbacher, B. (1985). Androgen stimulation and laryngeal development. *Annals of Otolaryngology, Rhinology and Laryngology*, 94(6 I), 634–640.
- Beeman, K. (1998). Digital signal analysis, editing, and synthesis. In *Animal Acoustic Communication* (pp. 59–103). Berlin: Springer. [https://doi.org/10.1007/978-3-642-76220-8\\_3](https://doi.org/10.1007/978-3-642-76220-8_3)
- Behrman, A. (2007). *Speech and voice science*. San Diego, California: Plural Pub.
- Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, 40(2), 531–539.
- Bellieni, C. V. (2012). Pain assessment in human fetus and infants. *The AAPS Journal*, 14(3), 456–461. <https://doi.org/10.1208/s12248-012-9354-5>
- Bellieni, C. V., Sisto, R., Cordelli, D. M., & Buonocore, G. (2004). Cry features reflect pain intensity in term newborns: An alarm threshold. *Pediatric Research*, 55(1), 142–146. <https://doi.org/10.1203/01.PDR.0000099793.99608.CB>
- Bellisari, A. (2008). Evolutionary origins of obesity. *Obesity Reviews*, 9(2), 165–180. <https://doi.org/10.1111/j.1467-789X.2007.00392.x>
- Belyk, M., & Brown, S. (2014). The acoustic correlates of valence depend on emotion family. *Journal of Voice*, 28(4), 523.e9-523.e18. <https://doi.org/10.1016/j.jvoice.2013.12.007>
- Benarroch, E. E. (2006). Pain-autonomic interactions. *Neurological Sciences*, 27(2), s130–s133. <https://doi.org/10.1007/s10072-006-0587-x>
- Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender gaps in overestimation of math performance. *Sex Roles*, 72(11–12), 536–546. <https://doi.org/10.1007/s11199-015-0486-9>
- Benítez, M. E., Roux, A. le, Fischer, J., Beehner, J. C., & Bergman, T. J. (2016). Acoustic and temporal variation in gelada (*Theropithecus gelada*) loud calls advertise male quality. *International Journal of Primatology*, 37(4–5), 568–585. <https://doi.org/10.1007/s10764-016-9922-0>

- Bernat, J. A., Calhoun, K. S., & Adams, H. E. (1999). Sexually aggressive and nonaggressive men: sexual arousal and judgments in response to acquaintance rape and consensual analogues. *Journal of Abnormal Psychology, 108*(4), 662–673.
- Bernstein, I. S., & Ehardt, C. L. (1985). Agonistic aiding: Kinship, rank, age, and sex influences. *American Journal of Primatology, 8*(1), 37–52. <https://doi.org/10.1002/ajp.1350080105>
- Berry, D. A., Herzel, H., Titze, I. R., & Story, B. H. (1996). Bifurcations in excised larynx experiments. *Journal of Voice, 10*(2), 129–138. [https://doi.org/10.1016/S0892-1997\(96\)80039-7](https://doi.org/10.1016/S0892-1997(96)80039-7)
- Bhuta, T., Patrick, L., & Garnett, J. D. (2004). Perceptual evaluation of voice quality and its correlation with acoustic measurements. *Journal of Voice, 18*(3), 299–304. <https://doi.org/10.1016/j.jvoice.2003.12.004>
- Bisazza, A., & Marconato, A. (1988). Female mate choice, male-male competition and parental care in the river bullhead, *Cottus gobio* L. (Pisces, Cottidae). *Animal Behaviour, 36*(5), 1352–1360. [https://doi.org/10.1016/S0003-3472\(88\)80204-5](https://doi.org/10.1016/S0003-3472(88)80204-5)
- Bishop, P., Cureton, K., & Collins, M. (1987). Sex difference in muscular strength in equally-trained men and women. *Ergonomics, 30*(4), 675–687. <https://doi.org/10.1080/00140138708969760>
- Blanchard, C. D., Hynd, A. L., Minke, K. A., Minemoto, T., & Blanchard, R. J. (2001). Human defensive behaviors to threat scenarios show parallels to fear- and anxiety-related defense patterns of non-human mammals. *Neuroscience & Biobehavioral Reviews, 25*(7–8), 761–770. [https://doi.org/10.1016/S0149-7634\(01\)00056-2](https://doi.org/10.1016/S0149-7634(01)00056-2)
- Bleidorn, W., Arslan, R. C., Denissen, J. J. A., Rentfrow, P. J., Gebauer, J. E., Potter, J., & Gosling, S. D. (2016). Age and gender differences in self-esteem - A cross-cultural window. *Journal of Personality and Social Psychology, 111*(3), 396–410. <https://doi.org/10.1037/pspp0000078>
- Blumstein, D. T., & Chi, Y. Y. (2012). Scared and less noisy: glucocorticoids are associated with alarm call entropy. *Biology Letters, 8*(2), 189–192. <https://doi.org/10.1098/rsbl.2011.0832>
- Blumstein, D. T., & Récapet, C. (2009). The sound of arousal: The addition of novel nonlinearities increases responsiveness in marmot alarm calls. *Ethology, 115*(11), 1074–1081. <https://doi.org/10.1111/j.1439-0310.2009.01691.x>
- Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., ... Fagot, J. (2017). Evidence of a vocalic proto-system in the baboon (*Papio papio*) suggests

- pre-hominin speech precursors. *PLOS ONE*, 12(1), e0169321. <https://doi.org/10.1371/journal.pone.0169321>
- Boero, D. L., Bianchi, C., Volpe, C., Marcello, A., & Lenti, C. (1998). Newborns crying in different contexts: Discrete or graded signals? *Perceptual and Motor Skills*, 86(3 suppl), 1123–1140. <https://doi.org/10.2466/pms.1998.86.3c.1123>
- Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer (Version 5.4). Retrieved from <http://www.praat.org/>
- Boidron, L., Boudenia, K., Avena, C., Boucheix, J.-M., & Aucouturier, J.-J. (2016). Emergency medical triage decisions are swayed by computer-manipulated cues of physical dominance in caller's voice. *Scientific Reports*, 6, 30219. <https://doi.org/10.1038/srep30219>
- Bolúmar, F., Olsen, J., Rebagliato, M., Sáez-Lloret, I., & Bisanti, L. (2000). Body mass index and delayed conception: A European multicenter study on infertility and subfecundity. *American Journal of Epidemiology*, 151(11), 1072–1079. <https://doi.org/10.1093/oxfordjournals.aje.a010150>
- Booth, A., Shelley, G., Mazur, A., Tharp, G., & Kittok, R. (1989). Testosterone, and winning and losing in human competition. *Hormones and Behavior*, 23(4), 556–571. [https://doi.org/10.1016/0018-506X\(89\)90042-1](https://doi.org/10.1016/0018-506X(89)90042-1)
- Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82(1), 55–59. <https://doi.org/10.1016/j.anbehav.2011.03.024>
- Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., Frommolt, K.-H., & Fitch, W. T. (2017). Body size and vocalization in primates and carnivores. *Scientific Reports*, 7, 41070. <https://doi.org/10.1038/srep41070>
- Bradbury, J. W. (1977). Lek mating behavior in the hammer-headed bat. *Zeitschrift Für Tierpsychologie*, 45(3), 225–255. <https://doi.org/10.1111/j.1439-0310.1977.tb02120.x>
- Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology*, 288(1), 1–20. <https://doi.org/10.1111/j.1469-7998.2012.00920.x>
- Briefer, E. F., Mandel, R., Maigrot, A.-L., Freymond, S. B., Bachmann, I., & Hillmann, E. (2017). Perception of emotional valence in horse whinnies. *Frontiers in Zoology*, 14(1), 8. <https://doi.org/10.1186/s12983-017-0193-1>

- Briefer, E. F., & McElligott, A. G. (2011). Indicators of age, body size and sex in goat kid calls revealed using the source–filter theory. *Applied Animal Behaviour Science*, 133(3–4), 175–185. <https://doi.org/10.1016/j.applanim.2011.05.012>
- Briseño-Jaramillo, M., Estrada, A., & Lemasson, A. (2015). Individual voice recognition and an auditory map of neighbours in free-ranging black howler monkeys (*Alouatta pigra*). *Behavioral Ecology and Sociobiology*, 69(1), 13–25. <https://doi.org/10.1007/s00265-014-1813-9>
- Brown, S. (2017). A joint prosodic origin of language and music. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01894>
- Bruckert, L., Liénard, J.-S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men’s characteristics. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1582), 83–89. <https://doi.org/10.1098/rspb.2005.3265>
- Brues, A. (1959). The spearman and the archer - an essay on selection in body build. *American Anthropologist*, 61(3), 457–469. <https://doi.org/10.1525/aa.1959.61.3.02a00080>
- Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evolution and Human Behavior*, 35(4), 327–335. <https://doi.org/10.1016/j.evolhumbehav.2014.03.003>
- Bryant, G. A., Fessler, D. M. T., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., ... Zhou, Y. (2016). Detecting affiliation in laughter across 24 societies. *Proceedings of the National Academy of Sciences*, 113(17), 4682–4687. <https://doi.org/10.1073/pnas.1524993113>
- Bryant, G. A., & Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biology Letters*, 5(1), 12–15. <https://doi.org/10.1098/rsbl.2008.0507>
- Budka, M., & Osiejuk, T. S. (2013). Formant frequencies are acoustic cues to caller discrimination and are a weak indicator of the body size of corncrake males. *Ethology*, 119(11), 960–969. <https://doi.org/10.1111/eth.12141>
- Bufalari, A., Adami, C., Angeli, G., & Short, C. E. (2007). Pain assessment in animals. *Veterinary Research Communications*, 31(1), 55–58. <https://doi.org/10.1007/s11259-007-0084-6>
- Bundy, J. N., & Puts, D. A. (2013). *Total sexual selection on men’s voices*. Doctoral dissertation, Pennsylvania State University. Retrieved from

[http://forms.gradsch.psu.edu/diversity/mcnair/mcnair\\_jrnl2012\\_14/files/2012-Bundy.pdf](http://forms.gradsch.psu.edu/diversity/mcnair/mcnair_jrnl2012_14/files/2012-Bundy.pdf)

- Burling, R. (1993). Primate calls, human language, and nonverbal communication. *Current Anthropology*, 34(1), 25–53. <https://doi.org/10.1086/204132>
- Callison, E. R., Berg, K. E., & Slivka, D. R. (2014). Grunting in tennis increases ball velocity but not oxygen cost. *Journal of Strength and Conditioning Research*, 28(7), 1915–1919. <https://doi.org/10.1519/JSC.0000000000000333>
- Calvino, B., Besson, J. M., Boehrer, A., & Depaulis, A. (1996). Ultrasonic vocalization (22–28 kHz) in a model of chronic pain, the arthritic rat: effects of analgesic drugs. *NeuroReport*, 7(2), 581.
- Campbell, B. C., Orourke, M. T., & Rabow, M. W. (1988). Pulsatile response of salivary testosterone and cortisol to aggressive competition in young males. In *American Journal of Physical Anthropology* (Vol. 75, pp. 193–194). New York, NY, US: John Wiley & Sons.
- Cao, D., Zhou, H., Wei, W., Lei, M., Yuan, S., Qi, D., & Zhang, Z. (2016). Vocal repertoire of adult captive red pandas (*Ailurus fulgens*). *Animal Biology*, 66(2), 145–155. <https://doi.org/10.1163/15707563-00002493>
- Caro, T. (2005). *Antipredator defenses in birds and mammals*. Chicago, US: University of Chicago Press.
- 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. *Hormones and Behavior*, 66(4), 569–576. <https://doi.org/10.1016/j.yhbeh.2014.08.006>
- Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PLOS ONE*, 7(2), e31353. <https://doi.org/10.1371/journal.pone.0031353>
- Cartei, V., Cowles, W., Banerjee, R., & Reby, D. (2014). Control of voice gender in pre-pubertal children. *British Journal of Developmental Psychology*, 32(1), 100–106. <https://doi.org/10.1111/bjdp.12027>
- Carter, B., McArthur, E., & Cunliffe, M. (2002). Dealing with uncertainty: parental assessment of pain in their children with profound special needs. *Journal of Advanced Nursing*, 38(5), 449–457. <https://doi.org/10.1046/j.1365-2648.2002.02206.x>

- Cazau, D., Adam, O., Aubin, T., Laitman, J. T., & Reidenberg, J. S. (2016). A study of vocal nonlinearities in humpback whale songs: from production mechanisms to acoustic analysis. *Scientific Reports*, 6, 31660. <https://doi.org/10.1038/srep31660>
- Charlton, B. D. (2015). The acoustic structure and information content of female koala vocal signals. *PLOS ONE*, 10(10), e0138670. <https://doi.org/10.1371/journal.pone.0138670>
- Charlton, B. D., Ellis, W. A. H., Brumm, J., Nilsson, K., & Fitch, W. T. (2012). Female koalas prefer bellows in which lower formants indicate larger males. *Animal Behaviour*, 84(6), 1565–1571. <https://doi.org/10.1016/j.anbehav.2012.09.034>
- Charlton, B. D., Ellis, W. A. H., Larkin, R., & Fitch, W. T. (2012). Perception of size-related formant information in male koalas (*Phascolarctos cinereus*). *Animal Cognition*, 15(5), 999–1006. <https://doi.org/10.1007/s10071-012-0527-5>
- Charlton, B. D., Ellis, W. A. H., McKinnon, A. J., Cowin, G. J., Brumm, J., Nilsson, K., & Fitch, W. T. (2011). Cues to body size in the formant spacing of male koala (*Phascolarctos cinereus*) bellows: honesty in an exaggerated trait. *Journal of Experimental Biology*, 214(20), 3414–3422. <https://doi.org/10.1242/jeb.061358>
- Charlton, B. D., Frey, R., McKinnon, A. J., Fritsch, G., Fitch, W. T., & Reby, D. (2013). Koalas use a novel vocal organ to produce unusually low-pitched mating calls. *Current Biology*, 23(23), R1035–R1036. <https://doi.org/10.1016/j.cub.2013.10.069>
- Charlton, B. D., Keating, J. L., Kersey, D., Rengui, L., Huang, Y., & Swaisgood, R. R. (2011). Vocal cues to male androgen levels in giant pandas. *Biology Letters*, 7(1), 71–74. <https://doi.org/10.1098/rsbl.2010.0582>
- Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications*, 7, 12739. <https://doi.org/10.1038/ncomms12739>
- Charlton, B. D., Reby, D., & McComb, K. (2007a). Female perception of size-related formant shifts in red deer, *Cervus elaphus*. *Animal Behaviour*, 74(4), 707–714. <https://doi.org/10.1016/j.anbehav.2006.09.021>
- Charlton, B. D., Reby, D., & McComb, K. (2007b). Female red deer prefer the roars of larger males. *Biology Letters*, 3(4), 382–385. <https://doi.org/10.1098/rsbl.2007.0244>
- Charlton, B. D., Reby, D., & McComb, K. (2008). Effect of combined source (F0) and filter (formant) variation on red deer hind responses to male roars. *The Journal of*

- the Acoustical Society of America*, 123(5), 2936–2943.  
<https://doi.org/10.1121/1.2896758>
- Charlton, B. D., Swaisgood, R. R., Zhihe, Z., & Snyder, R. J. (2012). Giant pandas attend to androgen-related variation in male bleats. *Behavioral Ecology and Sociobiology*, 66(6), 969–974. <https://doi.org/10.1007/s00265-012-1345-0>
- Charlton, B. D., Taylor, A. M., & Reby, D. (2013). Are men better than women at acoustic size judgements? *Biology Letters*, 9(4), 20130270. <https://doi.org/10.1098/rsbl.2013.0270>
- Charlton, B. D., Taylor, A. M., & Reby, D. (2017). Function and evolution of vibrato-like frequency modulation in mammals. *Current Biology*, 27(17), 2692–2697. <https://doi.org/10.1016/j.cub.2017.07.046>
- Charlton, B. D., Watchorn, D. J., & Whisson, D. A. (2017). Subharmonics increase the auditory impact of female koala rejection calls. *Ethology*, 123(8), 571–579. <https://doi.org/10.1111/eth.12628>
- Charlton, B. D., Whisson, D. A., & Reby, D. (2013). Free-ranging male koalas use size-related variation in formant frequencies to assess rival males. *PLOS ONE*, 8(7), e70279. <https://doi.org/10.1371/journal.pone.0070279>
- Charlton, B. D., Zhihe, Z., & Snyder, R. J. (2009). The information content of giant panda, *Ailuropoda melanoleuca*, bleats: acoustic cues to sex, age and size. *Animal Behaviour*, 78(4), 893–898. <https://doi.org/10.1016/j.anbehav.2009.06.029>
- Charlton, B. D., Zhihe, Z., & Snyder, R. J. (2010). Giant pandas perceive and attend to formant frequency variation in male bleats. *Animal Behaviour*, 79(6), 1221–1227. <https://doi.org/10.1016/j.anbehav.2010.02.018>
- Charrier, I., Ahonen, H., & Harcourt, R. G. (2011). What makes an Australian sea lion (*Neophoca cinerea*) male's bark threatening? *Journal of Comparative Psychology*, 125(4), 385–392. <https://doi.org/10.1037/a0024513>
- Cheney, D. L. (1977). The acquisition of rank and the development of reciprocal alliances among free-ranging immature baboons. *Behavioral Ecology and Sociobiology*, 2(3), 303–318. <https://doi.org/10.1007/BF00299742>
- Cheney, D. L., & Seyfarth, R. M. (1980). Vocal recognition in free-ranging vervet monkeys. *Animal Behaviour*, 28(2), 362–367. [https://doi.org/10.1016/S0003-3472\(80\)80044-3](https://doi.org/10.1016/S0003-3472(80)80044-3)

- Cheney, D. L., Seyfarth, R. M., & Silk, J. B. (1995). The role of grunts in reconciling opponents and facilitating interactions among adult female baboons. *Animal Behaviour*, 50(1), 249–257. <https://doi.org/10.1006/anbe.1995.0237>
- Cheng, J. T., Tracy, J. L., Ho, S., & Henrich, J. (2016). Listen, follow me: Dynamic vocal signals of dominance predict emergent social rank in humans. *Journal of Experimental Psychology: General*, 145(5), 536–547. <https://doi.org/10.1037/xge0000166>
- Chiba, T., & Kajiyama, M. (1958). *The vowel, its nature and structure*. Tokyo: Phonetic Society of Japan.
- Chow, J. W., Park, S.-A., & Tillman, M. D. (2009). Lower trunk kinematics and muscle activity during different types of tennis serves. *BMC Sports Science, Medicine and Rehabilitation*, 1(1), 24. <https://doi.org/10.1186/1758-2555-1-24>
- Clay, Z., & Zuberbühler, K. (2012). Communication during sex among female bonobos: effects of dominance, solicitation and audience. *Scientific Reports*, 2. <https://doi.org/10.1038/srep00291>
- Clutton-Brock, T. H., & Albon, S. D. (1979). The roaring of red deer and the evolution of honest advertisement. *Behaviour*, 69(3–4), 145–170. <https://doi.org/10.1163/156853979X00449>
- Conlan, T. (1999). The nature of warfare in fourteenth-century Japan: The record of Nomoto Tomoyuki. *Journal of Japanese Studies*, 25(2), 299–330. <https://doi.org/10.2307/133314>
- Cooper, W. E., & Vitt, L. J. (1993). Female mate choice of large male broad-headed skinks. *Animal Behaviour*, 45(4), 683–693. <https://doi.org/10.1006/anbe.1993.1083>
- Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion*, 16(1), 117–128. <https://doi.org/10.1037/emo0000100>
- Courtiol, A., Raymond, M., Godelle, B., & Ferdy, J.-B. (2010). Mate choice and human stature: Homogamy as a unified framework for understanding mating preferences. *Evolution*, 64(8), 2189–2203. <https://doi.org/10.1111/j.1558-5646.2010.00985.x>
- Craig, K. D. (2009). The social communication model of pain. *Canadian Psychology*, 50(1), 22–32.

- Craig, K. D., Grunau, R. V. E., & Aquan-Assee, J. (1988). Judgment of pain in newborns: Facial activity and cry as determinants. *Canadian Journal of Behavioural Science*, 20(4), 442–451. <https://doi.org/10.1037/h0079934>
- Dabbs Jr., J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences*, 27(4), 801–804. [https://doi.org/10.1016/S0191-8869\(98\)00272-4](https://doi.org/10.1016/S0191-8869(98)00272-4)
- Dahlen, H. G., Downe, S., Duff, M., & Gyte, G. M. (2013). Vaginal examination during normal labor : routine examination or routine intervention? *International Journal of Childbirth*, 3(3), 142–152. <https://doi.org/10.1891/2156-5287.3.3.142>
- Darwin, C. (1871). *The descent of man, and selection in relation to sex*. John Murray.
- Davidson, R. J., Scherer, K. R., & Goldsmith, H. H. (2003). *Handbook of affective sciences*. Oxford: Oxford University Press.
- Davila-Ross, M., Allcock, B., Thomas, C., & Bard, K. A. (2011). Aping expressions? Chimpanzees produce distinct laugh types when responding to laughter of others. *Emotion*, 11(5), 1013–1020. <https://doi.org/10.1037/a0022594>
- Davila-Ross, M., Owren, M. J., & Zimmermann, E. (2009). Reconstructing the evolution of laughter in great apes and humans. *Current Biology*, 19(13), 1106–1111. <https://doi.org/10.1016/j.cub.2009.05.028>
- Davila-Ross, M., Owren, M. J., & Zimmermann, E. (2010). The evolution of laughter in great apes and humans. *Communicative & Integrative Biology*, 3(2), 191–194. <https://doi.org/10.4161/cib.3.2.10944>
- de Boer, B., & Perlman, M. (2014). Physical mechanisms may be as important as brain mechanisms in evolution of speech. *Behavioral and Brain Sciences*, 37(06), 552–553. <https://doi.org/10.1017/S0140525X13004007>
- de Boer, B., Wich, S. A., Hardus, M. E., & Lameira, A. R. (2015). Acoustic models of orangutan hand-assisted alarm calls. *Journal of Experimental Biology*, 218(6), 907–914. <https://doi.org/10.1242/jeb.110577>
- de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech and Hearing Research*, 38(4), 794–811.
- de Waal, F. B. M. (2011). What is an animal emotion? *Annals of the New York Academy of Sciences*, 1224(1), 191–206. <https://doi.org/10.1111/j.1749-6632.2010.05912.x>

- Dessureau, B. K., Kurowski, C. O., & Thompson, N. S. (1998). A reassessment of the role of pitch and duration in adults' responses to infant crying. *Infant Behavior and Development*, 21(2), 367–371. [https://doi.org/10.1016/S0163-6383\(98\)90013-3](https://doi.org/10.1016/S0163-6383(98)90013-3)
- Docking, R. E., Lane, M., & Schofield, P. A. (2017). Usability testing of the iPhone app to improve pain assessment for older adults with cognitive impairment (prehospital setting): A qualitative study. *Pain Medicine*. <https://doi.org/10.1093/pm/pnx028>
- Dubner, R. (1994). Methods of assessing pain in animals. In Wall, P. D., & Melzack, R. (Eds.), *Textbook of pain* (3rd ed., pp. 293–302). Edinburgh: Churchill Livingstone.
- Dunlop, R. A. (2017). Potential motivational information encoded within humpback whale non-song vocal sounds. *Journal of the Acoustical Society of America*, 141(3), 2204–2213. <https://doi.org/10.1121/1.4978615>
- Dunn, J. C., Halenar, L. B., Davies, T. G., Cristobal-Azkarate, J., Reby, D., Sykes, D., ... Knapp, L. A. (2015). Evolutionary trade-off between vocal tract and testes dimensions in howler monkeys. *Current Biology*, 25(21), 2839–2844. <https://doi.org/10.1016/j.cub.2015.09.029>
- Eckenweber, M., & Knörnschild, M. (2016). Responsiveness to conspecific distress calls is influenced by day-roost proximity in bats (*Saccopteryx bilineata*). *Open Science*, 3(5), 160151. <https://doi.org/10.1098/rsos.160151>
- Efremova, K. O., Frey, R., Volodin, I. A., Fritsch, G., Soldatova, N. V., & Volodina, E. V. (2016). The postnatal ontogeny of the sexually dimorphic vocal apparatus in goitred gazelles (*Gazella subgutturosa*). *Journal of Morphology*, 277(6), 826–844. <https://doi.org/10.1002/jmor.20538>
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, 84(1), 5–17.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200. <https://doi.org/10.1080/02699939208411068>
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129.
- Elias, M. (1981). Serum cortisol, testosterone, and testosterone-binding globulin responses to competitive fighting in human males. *Aggressive Behavior*, 7(3), 215–224. [https://doi.org/10.1002/1098-2337\(1981\)7:3<215::AID-AB2480070305>3.0.CO;2-M](https://doi.org/10.1002/1098-2337(1981)7:3<215::AID-AB2480070305>3.0.CO;2-M)

- Elliott, B. (2006). Biomechanics and tennis. *British Journal of Sports Medicine*, 40(5), 392–396. <https://doi.org/10.1136/bjism.2005.023150>
- Emmons, L. H. (1978). Sound communication among African rainforest squirrels. *Zeitschrift Für Tierpsychologie*, 47(1), 1–49. <https://doi.org/10.1111/j.1439-0310.1978.tb01821.x>
- Erkut, S. (1983). Exploring sex differences in expectancy, attribution, and academic achievement. *Sex Roles*, 9(2), 217–231. <https://doi.org/10.1007/BF00289625>
- Eschali r, A., Marty, H., Trolese, J. F., Moncharmont, L., & Fialip, J. (1988). An automated method to analyze vocalization of unrestrained rats submitted to noxious electrical stimuli. *Journal of Pharmacological Methods*, 19(2), 175–184. [https://doi.org/10.1016/0160-5402\(88\)90038-1](https://doi.org/10.1016/0160-5402(88)90038-1)
- Esposito, G., Nakazawa, J., Venuti, P., & Bornstein, M. H. (2012). Perceptions of distress in young children with autism compared to typically developing children: A cultural comparison between Japan and Italy. *Research in Developmental Disabilities*, 33(4), 1059–1067. <https://doi.org/10.1016/j.ridd.2012.01.014>
- Esposito, G., Nakazawa, J., Venuti, P., & Bornstein, M. H. (2015). Judgment of infant cry: The roles of acoustic characteristics and sociodemographic characteristics. *Japanese Psychological Research*, 57(2), 126–134. <https://doi.org/10.1111/jpr.12072>
- Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Archives of Sexual Behavior*, 41(3), 557–570. <https://doi.org/10.1007/s10508-011-9875-5>
- Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. *Biological Psychology*, 72(2), 160–163. <https://doi.org/10.1016/j.biopsycho.2005.09.003>
- Ey, E., Pfefferle, D., & Fischer, J. (2007). Do age- and sex-related variations reliably reflect body size in non-human primate vocalizations? A review. *Primates*, 48(4), 253–267. <https://doi.org/10.1007/s10329-006-0033-y>
- Facchini, A., Bellieni, C. V., Marchettini, N., Pulselli, F. M., & Tiezzi, E. B. P. (2005). Relating pain intensity of newborns to onset of nonlinear phenomena in cry recordings. *Physics Letters A*, 338(3–5), 332–337. <https://doi.org/10.1016/j.physleta.2005.02.048>

- Fant, G. (1960). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*. Walter de Gruyter.
- Faragó, T., Andics, A., Devecseri, V., Kis, A., Gácsi, M., & Miklósi, Á. (2014). Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biology Letters*, 10(1), 20130926. <https://doi.org/10.1098/rsbl.2013.0926>
- Faragó, T., Pongrácz, P., Range, F., Virányi, Z., & Miklósi, Á. (2010). ‘The bone is mine’: affective and referential aspects of dog growls. *Animal Behaviour*, 79(4), 917–925. <https://doi.org/10.1016/j.anbehav.2010.01.005>
- Farhead, N., & Punt, T. D. (2015). Silencing sharapova’s grunt improves the perception of her serve speed. *Perceptual and Motor Skills*, 120(3), 722–730. <https://doi.org/10.2466/30.PMS.120v20x2>
- Fastl, H., & Zwicker, E. (2001). *Psychoacoustics: Facts and models* (3rd ed.). Berlin: Springer.
- Favaro, L., Gamba, M., Gili, C., & Pessani, D. (2017). Acoustic correlates of body size and individual identity in banded penguins. *PLOS ONE*, 12(2), e0170001. <https://doi.org/10.1371/journal.pone.0170001>
- Fedurek, P., Machanda, Z. P., Schel, A. M., & Slocombe, K. E. (2013). Pant hoot chorusing and social bonds in male chimpanzees. *Animal Behaviour*, 86(1), 189–196. <https://doi.org/10.1016/j.anbehav.2013.05.010>
- Fedurek, P., Slocombe, K. E., Enigk, D. K., Thompson, M. E., Wrangham, R. W., & Muller, M. N. (2016). The relationship between testosterone and long-distance calling in wild male chimpanzees. *Behavioral Ecology and Sociobiology*, 70(5), 659–672. <https://doi.org/10.1007/s00265-016-2087-1>
- Fedurek, P., Slocombe, K. E., & Zuberbühler, K. (2015). Chimpanzees communicate to two different audiences during aggressive interactions. *Animal Behaviour*, 110, 21–28. <https://doi.org/10.1016/j.anbehav.2015.09.010>
- Feierman, J. R. (2017). ‘Y/Our’ vocal sounds: Toward an ethological classification of human vocalising behavior and sounds. *Human Ethology Bulletin*, 32(3), 3–33.
- Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, 69(3), 561–568. <https://doi.org/10.1016/j.anbehav.2004.06.012>

- Filippi, P., Gogoleva, S. S., Volodina, E. V., Volodin, I. A., & Boer, B. de. (2017). Humans identify negative (but not positive) arousal in silver fox vocalizations: implications for the adaptive value of interspecific eavesdropping. *Current Zoology*, 63(4), 445–456. <https://doi.org/10.1093/cz/zox035>
- Fischer, J., Kitchen, D. M., Seyfarth, R. M., & Cheney, D. L. (2004). Baboon loud calls advertise male quality: acoustic features and their relation to rank, age, and exhaustion. *Behavioral Ecology and Sociobiology*, 56(2), 140–148. <https://doi.org/10.1007/s00265-003-0739-4>
- Fischer, J., Semple, S., Fickenscher, G., Jürgens, R., Kruse, E., Heistermann, M., & Amir, O. (2011). Do women's voices provide cues of the likelihood of ovulation? The importance of sampling regime. *PLOS ONE*, 6(9), e24490. <https://doi.org/10.1371/journal.pone.0024490>
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, 102(2), 1213–1222. <https://doi.org/10.1121/1.421048>
- Fitch, W. T. (2000). Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry. *Zoology - Analysis of Complex Systems*, 103(1–2), 40–58.
- Fitch, W. T. (2006). Production of vocalizations in mammals. In Brown, K. (Ed.), *Encyclopedia of language and linguistics* (pp. 115–121). Oxford: Elsevier. Retrieved from <http://homepage.univie.ac.at/tecumseh.fitch/wp-content/uploads/2010/08/Fitch2006MammalVocalProd.pdf>
- Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.
- Fitch, W. T., & Fritz, J. B. (2006). Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *The Journal of the Acoustical Society of America*, 120(4), 2132–2141. <https://doi.org/10.1121/1.2258499>
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511–1522. <https://doi.org/10.1121/1.427148>
- Fitch, W. T., & Hauser, M. D. (1995). Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on “honest” advertisement. *American Journal of Primatology*, 37(3), 191–219. <https://doi.org/10.1002/ajp.1350370303>

- Fitch, W. T., & Hauser, M. D. (2003). Unpacking “honesty”: Vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, R. R. Fay, & A. N. Popper (Eds.), *Acoustic communication* (pp. 65–137). Springer New York. Retrieved from [http://link.springer.com/chapter/10.1007/0-387-22762-8\\_3](http://link.springer.com/chapter/10.1007/0-387-22762-8_3)
- Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, 63(3), 407–418. <https://doi.org/10.1006/anbe.2001.1912>
- Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1477), 1669–1675. <https://doi.org/10.1098/rspb.2001.1704>
- Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, 12(4), 377–430. <https://doi.org/10.1002/j.1538-7305.1933.tb00403.x>
- Folland, J. P., Cauley, T. M. M., & Williams, A. G. (2008). Allometric scaling of strength measurements to body size. *European Journal of Applied Physiology*, 102(6), 739–745. <https://doi.org/10.1007/s00421-007-0654-x>
- Folstad, I., & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *The American Naturalist*, 139(3), 603–622.
- 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. *Open Science*, 3(10), 160395. <https://doi.org/10.1098/rsos.160395>
- Fox, M. W. (1970). A comparative study of the development of facial expressions in canids; wolf, coyote and foxes. *Behaviour*, 36(1), 49–73.
- Fraccaro, P. J., Jones, B. C., Vukovic, J., Smith, F. G., Watkins, C. D., Feinberg, D. R., ... DeBruine, L. M. (2011). Experimental evidence that women speak in a higher voice pitch to men they find attractive. *Journal of Evolutionary Psychology*, 9(1), 57–67. <https://doi.org/10.1556/JEP.9.2011.33.1>
- Fraccaro, P. J., O'Connor, J. J. M., Re, D. E., Jones, B. C., DeBruine, L. M., & Feinberg, D. R. (2013). Faking it: deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour*, 85(1), 127–136. <https://doi.org/10.1016/j.anbehav.2012.10.016>
- Frederick, D. A., & Haselton, M. G. (2007). Why is muscularity sexy? Tests of the fitness indicator hypothesis. *Personality and Social Psychology Bulletin*, 33(8), 1167–1183. <https://doi.org/10.1177/0146167207303022>

- Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, 138(2), 296–321. <https://doi.org/10.1037/a0026556>
- Frey, R., & Gebler, A. (2003). The highly specialized vocal tract of the male Mongolian gazelle (*Procapra gutturosa*). *Journal of Anatomy*, 203(5), 451–471. <https://doi.org/10.1046/j.1469-7580.2003.00232.x>
- Frey, R., Volodin, I., Volodina, E., Soldatova, N. V., & Juldachev, E. T. (2011). Descended and mobile larynx, vocal tract elongation and rutting roars in male goitred gazelles (*Gazella subgutturosa*). *Journal of Anatomy*, 218(5), 566–585. <https://doi.org/10.1111/j.1469-7580.2011.01361.x>
- Fry, D. P. (1987). Differences between playfighting and serious fighting among Zapotec children. *Ethology and Sociobiology*, 8(4), 285–306. [https://doi.org/10.1016/0162-3095\(87\)90029-X](https://doi.org/10.1016/0162-3095(87)90029-X)
- Fuller, B. F. (1991). Acoustic discrimination of three types of infant cries. *Nursing Research*, 40(3), 156–160.
- Fuller, B. F., & Conner, D. A. (1995). The effect of pain on infant behaviors. *Clinical Nursing Research*, 4(3), 253–273. <https://doi.org/10.1177/105477389500400304>
- Fuller, B. F., & Horii, Y. (1986). Differences in fundamental frequency, jitter, and shimmer among four types of infant vocalizations. *Journal of Communication Disorders*, 19(6), 441–447. [https://doi.org/10.1016/0021-9924\(86\)90019-5](https://doi.org/10.1016/0021-9924(86)90019-5)
- Fuller, B. F., & Horii, Y. (1988). Spectral energy distribution in four types of infant vocalizations. *Journal of Communication Disorders*, 21(3), 251–261. [https://doi.org/10.1016/0021-9924\(88\)90033-0](https://doi.org/10.1016/0021-9924(88)90033-0)
- Fuller, B. F., Roberts, J. E., & McKay, S. (1993). Acoustical analysis of maternal sounds during the second stage of labor. *Applied Nursing Research*, 6(1), 8–12. [https://doi.org/10.1016/S0897-1897\(05\)80036-9](https://doi.org/10.1016/S0897-1897(05)80036-9)
- Furley, P., & Schweizer, G. (2014a). ‘I’m pretty sure that we will win!’: The influence of score-related nonverbal behavioral changes on the confidence in winning a basketball game. *Journal of Sport and Exercise Psychology*, 36(3), 316–320. <https://doi.org/10.1123/jsep.2013-0199>
- Furley, P., & Schweizer, G. (2014b). The expression of victory and loss: Estimating who’s leading or trailing from nonverbal cues in sports. *Journal of Nonverbal Behavior*, 38(1), 13–29. <https://doi.org/10.1007/s10919-013-0168-7>

- Gallup, A. C., White, D. D., & Gallup, G. G. (2007). Handgrip strength predicts sexual behavior, body morphology, and aggression in male college students. *Evolution and Human Behavior*, 28(6), 423–429. <https://doi.org/10.1016/j.evolhumbehav.2007.07.001>
- Garcia, M., Charlton, B. D., Wyman, M. T., Fitch, W. T., & Reby, D. (2013). Do red deer stags (*Cervus elaphus*) use roar fundamental frequency (F0) to assess rivals? *PLOS ONE*, 8(12), e83946. <https://doi.org/10.1371/journal.pone.0083946>
- Garcia, M., Herbst, C. T., Bowling, D. L., Dunn, J. C., & Fitch, W. T. (2017). Acoustic allometry revisited: morphological determinants of fundamental frequency in primate vocal production. *Scientific Reports*, 7, 10450. <https://doi.org/10.1038/s41598-017-11000-x>
- Garcia, M., Wondrak, M., Huber, L., & Fitch, W. T. (2016). Honest signaling in domestic piglets (*Sus scrofa domestica*): vocal allometry and the information content of grunt calls. *Journal of Experimental Biology*, 219(12), 1913–1921. <https://doi.org/10.1242/jeb.138255>
- Garcia, M., Wyman, M. T., Charlton, B. D., Fitch, W. T., & Reby, D. (2014). Response of red deer stags (*Cervus elaphus*) to playback of harsh versus common roars. *Naturwissenschaften*, 101(10), 851–854. <https://doi.org/10.1007/s00114-014-1217-8>
- Gaulin, S., & Boster, J. (1985). Cross-cultural differences in sexual dimorphism. *Ethology and Sociobiology*, 6(4), 219–225. [https://doi.org/10.1016/0162-3095\(85\)90014-7](https://doi.org/10.1016/0162-3095(85)90014-7)
- Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *The Quarterly Review of Biology*, 80(4), 395–430. <https://doi.org/10.1086/498281>
- Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., & Winter, A. S. (2013). Vocal indices of stress: A review. *Journal of Voice*, 27(3), 390.e21–390.e29. <https://doi.org/10.1016/j.jvoice.2012.12.010>
- Global Delight Technologies. (2014). Boom 2 (Version 1.6). Global Delight Technologies. Retrieved from [www.globaldelight.com/boom](http://www.globaldelight.com/boom)
- Gobl, C., & Chasaide, A. N. (1992). Acoustic characteristics of voice quality. *Speech Communication*, 11(4), 481–490. [https://doi.org/10.1016/0167-6393\(92\)90055-C](https://doi.org/10.1016/0167-6393(92)90055-C)
- Gold, A. R., Brush, L. R., & Sprotzer, E. R. (1980). Developmental changes in self-perceptions of intelligence and self-confidence. *Psychology of Women Quarterly*, 5(2), 231–239. <https://doi.org/10.1111/j.1471-6402.1980.tb00959.x>

- González, J. (2006). Research in acoustics of human speech sounds: Correlates and perception of speaker body size. *Recent Research Developments in Applied Physics*, 37661(9), 1–15.
- González, J., & Oliver, J. C. (2004). Percepción a través de la voz de las características físicas del hablante: identificación de la estatura a partir de una frase o una vocal. *Revista de Psicología General y Aplicada*, 57, 21–33.
- Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J., ... Groves, C. P. (1998). Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Molecular Phylogenetics and Evolution*, 9(3), 585–598. <https://doi.org/10.1006/mpev.1998.0495>
- Gouzoules, H., & Gouzoules, S. (2000). Agonistic screams differ among four species of macaques: the significance of motivation-structural rules. *Animal Behaviour*, 59(3), 501–512. <https://doi.org/10.1006/anbe.1999.1318>
- Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour*, 32(1), 182–193. [https://doi.org/10.1016/S0003-3472\(84\)80336-X](https://doi.org/10.1016/S0003-3472(84)80336-X)
- Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., & Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice*, 2(2), 118–126. [https://doi.org/10.1016/S0892-1997\(88\)80067-5](https://doi.org/10.1016/S0892-1997(88)80067-5)
- Greisbach, R. (1999). Estimation of speaker height from formant frequencies. *International Journal of Speech Language and the Law*, 6(2), 265–277.
- Grunau, R. V. E., Johnston, C. C., & Craig, K. D. (1990). Neonatal facial and cry responses to invasive and non-invasive procedures. *Pain*, 42(3), 295–305. [https://doi.org/10.1016/0304-3959\(90\)91142-6](https://doi.org/10.1016/0304-3959(90)91142-6)
- Gustafson, G. E., & Green, J. A. (1989). On the importance of fundamental frequency and other acoustic features in cry perception and infant development. *Child Development*, 60(4), 772–780.
- Gustafson, G. E., & Harris, K. L. (1990). Women's responses to young infants' cries. *Developmental Psychology*, 26(1), 144–152.
- Gustison, M. L., & Townsend, S. W. (2015). A survey of the context and structure of high- and low-amplitude calls in mammals. *Animal Behaviour*, 105, 281–288. <https://doi.org/10.1016/j.anbehav.2015.04.021>
- Hadjistavropoulos, T., & Craig, K. D. (2002). A theoretical framework for understanding self-report and observational measures of pain: a communications model.

- Behaviour Research and Therapy*, 40(5), 551–570. [https://doi.org/10.1016/S0005-7967\(01\)00072-9](https://doi.org/10.1016/S0005-7967(01)00072-9)
- Hardus, M. E., Lameira, A. R., Schaik, C. P. V., & Wich, S. A. (2009). Tool use in wild orang-utans modifies sound production: a functionally deceptive innovation? *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20091027. <https://doi.org/10.1098/rspb.2009.1027>
- Harrigan, J. A., Gramata, J. F., Lucic, K. S., & Margolis, C. (1989). It's how you say it: Physicians' vocal behavior. *Social Science & Medicine*, 28(1), 87–92. [https://doi.org/10.1016/0277-9536\(89\)90310-9](https://doi.org/10.1016/0277-9536(89)90310-9)
- Harris, T. R., Fitch, W. T., Goldstein, L. M., & Fashing, P. J. (2006). Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass. *Ethology*, 112(9), 911–920.
- Hawk, S. T., van Kleef, G. A., Fischer, A. H., & van der Schalk, J. (2009). 'Worth a thousand words': Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, 9(3), 293–305. <https://doi.org/10.1037/a0015178>
- Herbst, C. T. (2016). Biophysics of vocal production in mammals. In *Vertebrate sound production and acoustic communication* (pp. 159–189). Springer, Cham. [https://doi.org/10.1007/978-3-319-27721-9\\_6](https://doi.org/10.1007/978-3-319-27721-9_6)
- Herr, K., Coyne, P. J., McCaffery, M., Manworren, R., & Merkel, S. (2011). Pain assessment in the patient unable to self-report: position statement with clinical practice recommendations. *Pain Management Nursing: Official Journal of the American Society of Pain Management Nurses*, 12(4), 230–250. <https://doi.org/10.1016/j.pmn.2011.10.002>
- Higashikawa, M., Nakai, K., Sakakura, A., & Takahashi, H. (1996). Perceived pitch of whispered vowels - relationship with formant frequencies: a preliminary study. *Journal of Voice: Official Journal of the Voice Foundation*, 10(2), 155–158.
- Hill, A. K., Bailey, D. H., & Puts, D. A. (2017). Gorillas in our midst? Human sexual dimorphism and contest competition in men. In M. Tibayrenc & F. J. Ayala (Eds.), *On human nature* (pp. 235–249). San Diego: Academic Press. <https://doi.org/10.1016/B978-0-12-420190-3.00015-6>
- Hill, A. K., Hunt, J., Welling, L. L. M., Cárdenas, R. A., Rotella, M. A., Wheatley, J. R., ... Puts, D. A. (2013). Quantifying the strength and form of sexual selection on men's traits. *Evolution and Human Behavior*, 34(5), 334–341. <https://doi.org/10.1016/j.evolhumbehav.2013.05.004>

- Hillenbrand, J. (1988). Perception of aperiodicities in synthetically generated voices. *The Journal of the Acoustical Society of America*, 83(6), 2361–2371. <https://doi.org/10.1121/1.396367>
- Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4), 769–778. <https://doi.org/10.1044/jshr.3704.769>
- Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39(2), 311–321. <https://doi.org/10.1044/jshr.3902.311>
- Hillenbrand, J. M., & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, 71(5), 1150–1166. <https://doi.org/10.3758/APP.71.5.1150>
- Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5 Pt 1), 3099–3111.
- Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, 21(4), 406–427. <https://doi.org/10.1007/s12110-010-9101-5>
- Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2011). Voice correlates of mating success in men: examining “contests” versus “mate choice” modes of sexual selection. *Archives of Sexual Behavior*, 40(3), 551–557. <https://doi.org/10.1007/s10508-010-9625-0>
- Hodges-Simeon, C. R., Gurven, M., Puts, D. A., & Gaulin, S. J. C. (2014). Vocal fundamental and formant frequencies are honest signals of threat potential in peripubertal males. *Behavioral Ecology*, aru081. <https://doi.org/10.1093/beheco/aru081>
- Hogstedt, G. (1983). Adaptation unto death: function of fear screams. *The American Naturalist*, 121(4), 562–570. <https://doi.org/10.1086/284083>
- Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *The Journal of the Acoustical Society of America*, 96(5), 2646–2654. <https://doi.org/10.1121/1.411275>
- Holtmaat, A., & Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience*, 10(9), 647. <https://doi.org/10.1038/nrn2699>

- Honda, K. (1983). Relationship between pitch control and vowel articulation. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (Vol. 73, pp. 286–297). San Diego: College Hill Press. Retrieved from <http://www.dtic.mil/get-tr-doc/pdf?AD=ADA129713#page=275>
- Houtsma, A. J. M. (1997). Pitch and timbre: Definition, meaning and use. *Journal of New Music Research*, 26(2), 104–115. <https://doi.org/10.1080/09298219708570720>
- Hudenko, W. J., Stone, W., & Bachorowski, J. A. (2009). Laughter differs in children with autism: An acoustic analysis of laughs produced by children with and without the disorder. *Journal of Autism and Developmental Disorders*, 39(10), 1392–1400. <https://doi.org/10.1007/s10803-009-0752-1>
- Hughes, S. M., Farley, S. D., & Rhodes, B. C. (2010). Vocal and physiological changes in response to the physical attractiveness of conversational partners. *Journal of Nonverbal Behavior*, 34(3), 155–167. <https://doi.org/10.1007/s10919-010-0087-9>
- Hughes, S. M., & Harrison, M. A. (2017). Your cheatin' voice will tell on you: Detection of past infidelity from voice. *Evolutionary Psychology*, 15(2). <https://doi.org/10.1177/1474704917711513>
- Hughes, S. M., Mogilski, J. K., & Harrison, M. A. (2014). The perception and parameters of intentional voice manipulation. *Journal of Nonverbal Behavior*, 38(1), 107–127. <https://doi.org/10.1007/s10919-013-0163-z>
- Ives, D. T., Smith, D. R., & Patterson, R. D. (2005). Discrimination of speaker size from syllable phrases. *The Journal of the Acoustical Society of America*, 118(6), 3816–3822.
- Jennings, D. J., & Gammell, M. P. (2013). Contest behaviour in ungulates. In I. C. W. Hardy & M. Briffa (Eds.), *Animal contests* (pp. 304–321). Cambridge: Cambridge University Press.
- Jiang, J. J., Zhang, Y., & Stern, J. (2001). Modeling of chaotic vibrations in symmetric vocal folds. *The Journal of the Acoustical Society of America*, 110(4), 2120–2128. <https://doi.org/10.1121/1.1395596>
- Jiang, T., Huang, X., Wu, H., & Feng, J. (2017). Size and quality information in acoustic signals of *Rhinolophus ferrumequinum* in distress situations. *Physiology & Behavior*, 173, 252–257. <https://doi.org/10.1016/j.physbeh.2017.02.025>
- Johannes, B., Wittels, P., Enne, R., Eisinger, G., Castro, C. A., Thomas, J. L., ... Gerzer, R. (2007). Non-linear function model of voice pitch dependency on physical and

- mental load. *European Journal of Applied Physiology*, 101(3), 267–276. <https://doi.org/10.1007/s00421-007-0496-6>
- Johnson, C. D., & McHugh, M. P. (2006). Performance demands of professional male tennis players. *British Journal of Sports Medicine*, 40(8), 696–699. <https://doi.org/10.1136/bjsm.2005.021253>
- Johnston, C. C., & O'Shaughnessy, D. (1987). Acoustical attributes of infant pain cries: Discriminating features. *PAIN*, 30, S233. [https://doi.org/10.1016/0304-3959\(87\)91526-0](https://doi.org/10.1016/0304-3959(87)91526-0)
- Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, 79(1), 57–62. <https://doi.org/10.1016/j.anbehav.2009.10.003>
- Jones, D. A., Rutherford, O. M., & Parker, D. F. (1989). Physiological changes in skeletal muscle as a result of strength training. *Quarterly Journal of Experimental Physiology*, 74(3), 233–256. <https://doi.org/10.1113/expphysiol.1989.sp003268>
- Jones, E. J., Bishop, P. A., Woods, A. K., & Green, J. M. (2008). Cross-sectional area and muscular strength. *Sports Medicine*, 38(12), 987–994. <https://doi.org/10.2165/00007256-200838120-00003>
- Jones, M. R., & Witt, C. C. (2014). Migrate small, sound big: functional constraints on body size promote tracheal elongation in cranes. *Journal of Evolutionary Biology*, 27(6), 1256–1264. <https://doi.org/10.1111/jeb.12397>
- Jordan, J., & Mossman, E. (2017). “Back off buddy, this is my body, not yours”: Empowering girls through self-defense. *Violence Against Women*, 1077801217741217. <https://doi.org/10.1177/1077801217741217>
- Jourdan, D., Ardid, D., Chapuy, E., Eschali r, A., & Le Bars, D. (1995). Audible and ultrasonic vocalization elicited by single electrical nociceptive stimuli to the tail in the rat. *PAIN*, 63(2), 237–249. [https://doi.org/10.1016/0304-3959\(95\)00049-X](https://doi.org/10.1016/0304-3959(95)00049-X)
- Jovanovic, T., & Gouzoules, H. (2001). Effects of nonmaternal restraint on the vocalizations of infant rhesus monkeys (*Macaca mulatta*). *American Journal of Primatology*, 53(1), 33–45. [https://doi.org/10.1002/1098-2345\(200101\)53:1<33::AID-AJP3>3.0.CO;2-C](https://doi.org/10.1002/1098-2345(200101)53:1<33::AID-AJP3>3.0.CO;2-C)
- Jovanovic, T., Megna, N. L., & Maestripieri, D. (2000). Early maternal recognition of offspring vocalizations in rhesus macaques (*Macaca mulatta*). *Primates*, 41(4), 421–428. <https://doi.org/10.1007/BF02557653>

- Judge, T. A., & Cable, D. M. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *Journal of Applied Psychology*, 89(3), 428–441. <https://doi.org/10.1037/0021-9010.89.3.428>
- 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. *Journal of Nonverbal Behavior*, 39(3), 195–214. <https://doi.org/10.1007/s10919-015-0209-5>
- Jürgens, U. (1994). The role of the periaqueductal grey in vocal behaviour. *Behavioural Brain Research*, 62(2), 107–117. [https://doi.org/10.1016/0166-4328\(94\)90017-5](https://doi.org/10.1016/0166-4328(94)90017-5)
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, 129(5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Kahane, J. C. (1982). Growth of the human prepubertal and pubertal larynx. *Journal of Speech, Language, and Hearing Research*, 25(3), 446–455. <https://doi.org/10.1044/jshr.2503.446>
- Karp, D., Manser, M. B., Wiley, E. M., & Townsend, S. W. (2014). Nonlinearities in meerkat alarm calls prevent receivers from habituating. *Ethology*, 120(2), 189–196. <https://doi.org/10.1111/eth.12195>
- Karthikeyan, S., & Locke, J. L. (2015). Men's evaluation of women's speech in a simulated dating context: Effects of female fertility on vocal pitch and attractiveness. *Evolutionary Behavioral Sciences*, 9(1), 55–67. <https://doi.org/10.1037/ebbs0000014>
- Katić, R., Blažević, S., Krstulović, S., & Mulić, R. (2005). Morphological structures of elite karateka and their impact on technical and fighting efficiency. *Collegium Antropologicum*, 29(1), 79–84.
- Kelemen, G., & Sade, J. (1960). The vocal organ of the Howling monkey (*Alouatta palliata*). *Journal of Morphology*, 107(2), 123–140. <https://doi.org/10.1002/jmor.1051070202>
- Kelly, T., Reby, D., Levréro, F., Keenan, S., Gustafsson, E., Koutseff, A., & Mathevon, N. (2017). Adult human perception of distress in the cries of bonobo, chimpanzee, and human infants. *Biological Journal of the Linnean Society*, 120(4), 919–930. <https://doi.org/10.1093/biolinnean/blw016>

- Kersken, V., Zuberbühler, K., & Gomez, J.-C. (2017). Listeners can extract meaning from non-linguistic infant vocalisations cross-culturally. *Scientific Reports*, 7. <https://doi.org/10.1038/srep41016>
- Kewley-Port, D., Li, X., Zheng, Y., & Neel, A. T. (1996). Fundamental frequency effects on thresholds for vowel formant discrimination. *The Journal of the Acoustical Society of America*, 100(4), 2462–2470. <https://doi.org/10.1121/1.417954>
- Kibler, W. B. (2014). Understanding the kinetic chain in tennis performance and injury. *Aspetar Sports Medicine Journal*, 2, 492–497.
- Kidjo, N., Cargnelutti, B., Charlton, B. D., Wilson, C., & Reby, D. (2008). Vocal behaviour in the endangered Corsican deer: description and phylogenetic implications. *Bioacoustics*, 18(2), 159–181.
- Kim, J., Heshka, S., Gallagher, D., Kotler, D. P., Mayer, L., Albu, J., ... Heymsfield, S. B. (2004). Intermuscular adipose tissue-free skeletal muscle mass: estimation by dual-energy X-ray absorptiometry in adults. *Journal of Applied Physiology*, 97(2), 655–660. <https://doi.org/10.1152/jappphysiol.00260.2004>
- Kirchengast, S., & Huber, J. (2004). Body composition characteristics and fat distribution patterns in young infertile women. *Fertility and Sterility*, 81(3), 539–544. <https://doi.org/10.1016/j.fertnstert.2003.08.018>
- Klemuk, S. A., Riede, T., Walsh, E. J., & Titze, I. R. (2011). Adapted to roar: Functional morphology of tiger and lion vocal folds. *PLOS ONE*, 6(11), e27029. <https://doi.org/10.1371/journal.pone.0027029>
- Klofstad, C. A. (2015). Candidate voice pitch influences election outcomes. *Political Psychology*, n/a-n/a. <https://doi.org/10.1111/pops.12280>
- Knight, C. (1998). Ritual/speech coevolution: A solution to the problem of deception. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 68–91). Cambridge: Cambridge University Press.
- Knowles, K. K., & Little, A. C. (2016). Vocal fundamental and formant frequencies affect perceptions of speaker cooperativeness. *Quarterly Journal of Experimental Psychology*, 69(9), 1657–1675. <https://doi.org/10.1080/17470218.2015.1091484>
- Knudson, D., & Blackwell, J. (2000). Trunk muscle activation in open stance and square stance tennis forehands. *International Journal of Sports Medicine*, 21(5), 321–324. <https://doi.org/10.1055/s-2000-3776>

- Koeda, M., Belin, P., Hama, T., Masuda, T., Matsuura, M., & Okubo, Y. (2013). Cross-cultural differences in the processing of non-verbal affective vocalizations by Japanese and Canadian listeners. *Frontiers in Psychology*, 4, 105. <https://doi.org/10.3389/fpsyg.2013.00105>
- Konerding, W. S., Zimmermann, E., Bleich, E., Hedrich, H.-J., & Scheumann, M. (2016). Female cats, but not males, adjust responsiveness to arousal in the voice of kittens. *BMC Evolutionary Biology*, 16. <https://doi.org/10.1186/s12862-016-0718-9>
- Kosakowska-Berezecka, N., Jurek, P., Besta, T., & Badowska, S. (2017). Self-presentation strategies, fear of success and anticipation of future success among university and high school students. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01884>
- Koutseff, A., Reby, D., Martin, O., Levvero, F., Patural, H., & Mathevon, N. (2017). The acoustic space of pain: cries as indicators of distress recovering dynamics in pre-verbal infants. *Bioacoustics*, 0(0), 1–13. <https://doi.org/10.1080/09524622.2017.1344931>
- Krebs, J. R., & Dawkins, R. (1984). Animal signals: mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (2nd ed., pp. 380–402). Oxford: Blackwell Scientific Publications.
- Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons.
- Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25(1), 161–168. <https://doi.org/10.1016/j.cogbrainres.2005.05.004>
- Krul, A. J., Daanen, H. A. M., & Choi, H. (2011). Self-reported and measured weight, height and body mass index (BMI) in Italy, the Netherlands and North America. *European Journal of Public Health*, 21(4), 414–419. <https://doi.org/10.1093/eurpub/ckp228>
- Krützen, M., Willems, E. P., & van Schaik, C. P. (2011). Culture and geographic variation in orangutan behavior. *Current Biology*, 21(21), 1808–1812. <https://doi.org/10.1016/j.cub.2011.09.017>
- Kurejova, M., Nattenmüller, U., Hildebrandt, U., Selvaraj, D., Stösser, S., & Kuner, R. (2010). An improved behavioural assay demonstrates that ultrasound vocalizations constitute a reliable indicator of chronic cancer pain and neuropathic pain. *Molecular Pain*, 6, 18. <https://doi.org/10.1186/1744-8069-6-18>

- Küypers, H. G. J. M. (1958). Corticobulbar connections to the pons and lower brain stem in man. *Brain*, *81*, 364–388.
- La Pointe, L. L., Mowrer, D. M., & Case, J. L. (1990). A comparative acoustic analysis of the laugh responses of 20- and 70-year-old males. *The International Journal of Aging and Human Development*, *31*(1), 1–9. <https://doi.org/10.2190/GU55-AF9Q-M586-4HM3>
- LaGasse, L. L., Neal, A. R., & Lester, B. M. (2005). Assessment of infant cry: Acoustic cry analysis and parental perception. *Mental Retardation and Developmental Disabilities Research Reviews*, *11*(1), 83–93. <https://doi.org/10.1002/mrdd.20050>
- Lagier, A., Vaugoyeau, M., Ghio, A., Legou, T., Giovanni, A., & Assaiante, C. (2010). Coordination between posture and phonation in vocal effort behavior. *Folia Phoniatrica et Logopaedica*, *62*(4), 195–202. <https://doi.org/10.1159/000314264>
- Lake, J. K., Power, C., & Cole, T. J. (1997). Women's reproductive health: the role of body mass index in early and adult life. *International Journal of Obesity and Related Metabolic Disorders: Journal of the International Association for the Study of Obesity*, *21*(6), 432–438.
- Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., & Menken, S. B. J. (2015). Speech-like rhythm in a voiced and voiceless orangutan call. *PLoS ONE*, *10*(1), e116136. <https://doi.org/10.1371/journal.pone.0116136>
- Lassek, W. D., & Gaulin, S. J. C. (2009). Costs and benefits of fat-free muscle mass in men: relationship to mating success, dietary requirements, and native immunity. *Evolution and Human Behavior*, *30*(5), 322–328. <https://doi.org/10.1016/j.evolhumbehav.2009.04.002>
- Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W., ... Thingujam, N. S. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology*, *4*. <https://doi.org/10.3389/fpsyg.2013.00353>
- Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, *19*(5), 633–653. <https://doi.org/10.1080/02699930441000445>
- Lavan, N., & McGettigan, C. (2017). Increased discriminability of authenticity from multimodal laughter is driven by auditory information. *Quarterly Journal of Experimental Psychology*, *70*(10), 2159–2168. <https://doi.org/10.1080/17470218.2016.1226370>

- Lavan, N., Scott, S. K., & McGettigan, C. (2015). Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter. *Journal of Nonverbal Behavior*, 1–17. <https://doi.org/10.1007/s10919-015-0222-8>
- Le, D., & Provost, E. M. (2013). Emotion recognition from spontaneous speech using Hidden Markov models with deep belief networks. In *2013 IEEE Workshop on Automatic Speech Recognition and Understanding* (pp. 216–221). <https://doi.org/10.1109/ASRU.2013.6707732>
- Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105(3), 1455–1468. <https://doi.org/10.1121/1.426686>
- Leger, D. W., Thompson, R. A., Merritt, J. A., & Benz, J. J. (1996). Adult perception of emotion intensity in human infant cries: effects of infant age and cry acoustics. *Child Development*, 67(6), 3238–3249.
- Lehr, V. T., Zeskind, P. S., Ofenstein, J. P., Cepeda, E., Warrier, I., & Aranda, J. V. (2007). Neonatal facial coding system scores and spectral characteristics of infant crying during newborn circumcision. *The Clinical Journal of Pain*, 23(5), 417–424. <https://doi.org/10.1097/AJP.0b013e31805476f2>
- Lemasson, A., Remeuf, K., Trabalon, M., Cuir, F., & Hausberger, M. (2015). Mares prefer the voices of highly fertile stallions. *PLOS ONE*, 10(2), e0118468. <https://doi.org/10.1371/journal.pone.0118468>
- Leongómez, J. D., Binter, J., Kubicová, L., Stolařová, P., Klapilová, K., Havlíček, J., & Roberts, S. C. (2014). Vocal modulation during courtship increases proceptivity even in naive listeners. *Evolution and Human Behavior*, 35(6), 489–496. <https://doi.org/10.1016/j.evolhumbehav.2014.06.008>
- Leongómez, J. D., Mileva, V. R., Little, A. C., & Roberts, S. C. (2017). Perceived differences in social status between speaker and listener affect the speaker's vocal characteristics. *PLOS ONE*, 12(6), e0179407. <https://doi.org/10.1371/journal.pone.0179407>
- Leuchtenberger, C., Sousa-Lima, R., Ribas, C., Magnusson, W. E., & Mourao, G. (2016). Giant otter alarm calls as potential mechanisms for individual discrimination and sexual selection. *Bioacoustics*, 25(3), 279–291. <https://doi.org/10.1080/09524622.2016.1157704>

- Levin, R. J. (2006). Vocalised sounds and human sex. *Sexual and Relationship Therapy*, 21(1), 99–107. <https://doi.org/10.1080/14681990500438014>
- Levine, J. D., Feldmesser, M., Tecott, L., Gordon, N. C., & Izdebski, K. (1984). Pain-induced vocalization in the rat and its modification by pharmacological agents. *Brain Research*, 296(1), 121–127. [https://doi.org/10.1016/0006-8993\(84\)90517-1](https://doi.org/10.1016/0006-8993(84)90517-1)
- Levine, J. D., & Gordon, N. C. (1982). Pain in prelingual children and its evaluation by pain-induced vocalization. *Pain*, 14(2), 85–93. [https://doi.org/10.1016/0304-3959\(82\)90090-2](https://doi.org/10.1016/0304-3959(82)90090-2)
- Lev-Ran, A. (2001). Human obesity: an evolutionary approach to understanding our bulging waistline. *Diabetes/Metabolism Research and Reviews*, 17(5), 347–362. <https://doi.org/10.1002/dmrr.230>
- Lévréro, F., & Mathevon, N. (2013). Vocal signature in wild infant chimpanzees. *American Journal of Primatology*, 75(4), 324–332. <https://doi.org/10.1002/ajp.22108>
- Li, L., Zhao, Y., Jiang, D., Zhang, Y., Wang, F., Gonzalez, I., ... Sahli, H. (2013). Hybrid deep neural network–hidden Markov model (DNN-HMM) based speech emotion recognition. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 312–317). <https://doi.org/10.1109/ACII.2013.58>
- Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology*, 46(2), 117–128. [https://doi.org/10.1016/S0003-9969\(00\)00108-4](https://doi.org/10.1016/S0003-9969(00)00108-4)
- Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.
- Lim, L. L., Seubsman, S., & Sleight, A. (2009). Validity of self-reported weight, height, and body mass index among university students in Thailand: Implications for population studies of obesity in developing countries. *Population Health Metrics*, 7, 15. <https://doi.org/10.1186/1478-7954-7-15>
- Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods*, 45(4), 1234–1245. <https://doi.org/10.3758/s13428-013-0324-3>
- Lindenfors, P., Gittleman, J. L., & Jones, K. E. (2007). Sexual size dimorphism in mammals. In D. J. Fairbairn, W. U. Blanckenhorn, & T. Székely (Eds.), *Sex, size and gender roles : evolutionary studies of sexual size dimorphism*. New York, NY, US: Oxford University Press. Retrieved from

[http://www.zoologi.su.se/research/lindenfors/publications/Lindenfors\\_et\\_al\\_2007c.pdf](http://www.zoologi.su.se/research/lindenfors/publications/Lindenfors_et_al_2007c.pdf)

- Lindová, J., Špinka, M., & Nováková, L. (2015). Decoding of baby calls: Can adult humans identify the eliciting situation from emotional vocalizations of preverbal infants? *PLOS ONE*, 10(4), e0124317. <https://doi.org/10.1371/journal.pone.0124317>
- Lingle, S., & Riede, T. (2014). Deer mothers are sensitive to infant distress vocalizations of diverse mammalian species. *The American Naturalist*, 184(4), 510–522. <https://doi.org/10.1086/677677>
- Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., & Romanow, C. A. (2012). What makes a cry a cry? A review of infant distress vocalizations. *Current Zoology*, 58(5), 698–726.
- Linhart, P., Ratcliffe, V. F., Reby, D., & Špinka, M. (2015). Expression of emotional arousal in two different piglet call types. *PLOS ONE*, 10(8), e0135414.
- Liu, N., Jiang, Z., Zhang, L., Zhong, Z., Ping, X., Xu, H., & Li, C. (2015). Bioacoustic cues and their relations to dominance rank in Père David's deer stags. *Animal Production Science*.
- Maddieson, I. (2009). *Patterns of sounds*. Cambridge: Cambridge University Pres.
- Maitre, N. L., Stark, A. R., Menser, C. C. M., Chorna, O. D., France, D. J., Key, A. F., ... Bruehl, S. (2017). Cry presence and amplitude do not reflect cortical processing of painful stimuli in newborns with distinct responses to touch or cold. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 102(5), F428–F433. <https://doi.org/10.1136/archdischild-2016-312279>
- Manser, M. B., Bell, M. B., & Fletcher, L. B. (2001). The information that receivers extract from alarm calls in suricates. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1484), 2485–2491. <https://doi.org/10.1098/rspb.2001.1772>
- Maquirriain, J., Baglione, R., & Cardey, M. (2016). Male professional tennis players maintain constant serve speed and accuracy over long matches on grass courts. *European Journal of Sport Science*, 16(7), 845–849. <https://doi.org/10.1080/17461391.2016.1156163>
- Markova, D., Richer, L., Pangelinan, M., Schwartz, D. H., Leonard, G., Perron, M., ... Paus, T. (2016). Age- and sex-related variations in vocal-tract morphology and

- voice acoustics during adolescence. *Hormones and Behavior*, 81, 84–96. <https://doi.org/10.1016/j.yhbeh.2016.03.001>
- Maruščáková, I. L., Linhart, P., Ratcliffe, V. F., Tallet, C., Reby, D., & Špinka, M. (2015). Humans (*Homo sapiens*) judge the emotional content of piglet (*Sus scrofa domestica*) calls based on simple acoustic parameters, not personality, empathy, nor attitude toward animals. *Journal of Comparative Psychology (Washington, D.C.: 1983)*, 129(2), 121–131. <https://doi.org/10.1037/a0038870>
- Marx, G., Horn, T., Thielebein, J., Knubel, B., & von Borell, E. (2003). Analysis of pain-related vocalization in young pigs. *Journal of Sound and Vibration*, 266(3), 687–698. [https://doi.org/10.1016/S0022-460X\(03\)00594-7](https://doi.org/10.1016/S0022-460X(03)00594-7)
- Maughan, R. J., Watson, J. S., & Weir, J. (1983). Strength and cross-sectional area of human skeletal muscle. *The Journal of Physiology*, 338(1), 37–49. <https://doi.org/10.1113/jphysiol.1983.sp014658>
- Maynard-Smith, J., & Harper, D. (2003). *Animal Signals*. Oxford: Oxford University Press.
- Mazur, A., & Lamb, T. A. (1980). Testosterone, status, and mood in human males. *Hormones and Behavior*, 14(3), 236–246. [https://doi.org/10.1016/0018-506X\(80\)90032-X](https://doi.org/10.1016/0018-506X(80)90032-X)
- McComb, K., Taylor, A. M., Wilson, C., & Charlton, B. D. (2009). The cry embedded within the purr. *Current Biology*, 19(13), R507–R508. <https://doi.org/10.1016/j.cub.2009.05.033>
- McElligott, A. G., Birrer, M., & Vannoni, E. (2006). Retraction of the mobile descended larynx during groaning enables fallow bucks (*Dama dama*) to lower their formant frequencies. *Journal of Zoology*, 270(2), 340–345. <https://doi.org/10.1111/j.1469-7998.2006.00144.x>
- McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2015). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral Cortex*, 25(1), 246–257. <https://doi.org/10.1093/cercor/bht227>
- McGrath, P. J., Rosmus, C., Canfield, C., Campbell, M. A., & Hennigar, A. (1998). Behaviours caregivers use to determine pain in non-verbal, cognitively impaired individuals. *Developmental Medicine and Child Neurology*, 40(5), 340–343.

- Mehta, P. H., & Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Hormones and Behavior*, 58(5), 898–906. <https://doi.org/10.1016/j.yhbeh.2010.08.020>
- Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, 48(2), 169–178. <https://doi.org/10.3758/BF03207084>
- Mergell, P., Fitch, W. T., & Herzel, H. (1999). Modeling the role of nonhuman vocal membranes in phonation. *The Journal of the Acoustical Society of America*, 105(3), 2020–2028. <https://doi.org/10.1121/1.426735>
- Merridale, C. (2006). Culture, ideology and combat in the Red Army, 1939-45. *Journal of Contemporary History*, 41(2), 305–324.
- Miller, S. E., Schlauch, R. S., & Watson, P. J. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *Journal of the Acoustical Society of America*, 128(1), 435–443. <https://doi.org/10.1121/1.3397384>
- Minter, L. J., & DeLiberto, T. J. (2008). Seasonal variation in serum testosterone, testicular volume, and semen characteristics in the coyote (*Canis latrans*). *Theriogenology*, 69(8), 946–952. <https://doi.org/10.1016/j.theriogenology.2008.01.010>
- Mitani, J. C., & Gros-Louis, J. (1995). Species and sex differences in the screams of chimpanzees and bonobos. *International Journal of Primatology*, 16(3), 393–411. <https://doi.org/10.1007/BF02735794>
- Mitani, J. C., & Stuht, J. (1998). The evolution of nonhuman primate loud calls: Acoustic adaptation for long-distance transmission. *Primates*, 39(2), 171–182. <https://doi.org/10.1007/BF02557729>
- Mogil, J. S. (2009). Animal models of pain: progress and challenges. *Nature Reviews Neuroscience*, 10(4), nrn2606. <https://doi.org/10.1038/nrn2606>
- Monden, C. W. S., & Smits, J. (2009). Maternal height and child mortality in 42 developing countries. *American Journal of Human Biology*, 21(3), 305–311. <https://doi.org/10.1002/ajhb.20860>
- Montano, K. J., Tigue, C. C., Isenstein, S. G. E., Barclay, P., & Feinberg, D. R. (2017). Men's voice pitch influences women's trusting behavior. *Evolution and Human Behavior*, 38(3), 293–297. <https://doi.org/10.1016/j.evolhumbehav.2016.10.010>

- Moody, A. (2013). *Adult anthropometric measures, overweight and obesity* (Health Survey for England - 2012). NHS. Retrieved from <http://digital.nhs.uk/catalogue/PUB13218>
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111(981), 855–869.
- Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9(5), 453–467. [https://doi.org/10.1016/0167-6393\(90\)90021-Z](https://doi.org/10.1016/0167-6393(90)90021-Z)
- Myers, T., Nevill, A., & Al-Nakeeb, Y. (2012). The influence of crowd noise upon judging decisions in Muay Thai. *Advances in Physical Education*, 02(04), 148–152. <https://doi.org/10.4236/ape.2012.24026>
- Nadhurou, B., Gamba, M., Andriaholinirina, N. V., Ouledi, A., & Giacoma, C. (2016). The vocal communication of the mongoose lemur (*Eulemur mongoz*): phonation mechanisms, acoustic features and quantitative analysis. *Ethology Ecology & Evolution*, 28(3), 241–260. <https://doi.org/10.1080/03949370.2015.1039069>
- Nevill, A. M., Balmer, N. J., & Mark Williams, A. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, 3(4), 261–272. [https://doi.org/10.1016/S1469-0292\(01\)00033-4](https://doi.org/10.1016/S1469-0292(01)00033-4)
- Newman, J. D. (2007). Neural circuits underlying crying and cry responding in mammals. *Behavioural Brain Research*, 182(2), 155–165. <https://doi.org/10.1016/j.bbr.2007.02.011>
- Niemitz, C. (1990). Visuelle Zeichen, Sprache und Gehirn in der Evolution des Menschen—Eine Entgegnung auf McFarland (Visual signs, language and the brain in the evolution of humans—A reply to McFarland). *Zeitschrift für Semiotik*, 12, 323–336.
- Nwe, T. L., Foo, S. W., & De Silva, L. C. (2003). Speech emotion recognition using hidden Markov models. *Speech Communication*, 41(4), 603–623. [https://doi.org/10.1016/S0167-6393\(03\)00099-2](https://doi.org/10.1016/S0167-6393(03)00099-2)
- Nwokah, E. E., Davies, P., Islam, A., Hsu, H. C., & Fogel, A. (1993). Vocal affect in three-year-olds: a quantitative acoustic analysis of child laughter. *The Journal of the Acoustical Society of America*, 94(6), 3076–3090.
- Nwokah, E. E., Hsu, H.-C., Dobrowolska, O., & Fogel, A. (1994). The development of laughter in mother-infant communication: Timing parameters and temporal

- sequences. *Infant Behavior and Development*, 17(1), 23–35.  
[https://doi.org/10.1016/0163-6383\(94\)90019-1](https://doi.org/10.1016/0163-6383(94)90019-1)
- O’Connell, D. G., Hinman, M., Hearne, K. F., Michael, Z. S., & Nixon, S. L. (2014). The effects of “grunting” on serve and forehand velocity in collegiate tennis players. *Journal of Strength and Conditioning Research*, 1.  
<https://doi.org/10.1519/JSC.0000000000000604>
- O’Connor, J. J. M., Fraccaro, P. J., & Feinberg, D. R. (2012). The influence of male voice pitch on women’s perceptions of relationship investment. *Journal of Evolutionary Psychology*, 10(1), 1–13. <https://doi.org/10.1556/JEP.10.2012.1.1>
- O’Connor, J. J. M., Re, D. E., & Feinberg, D. R. (2011). Voice pitch influences perceptions of sexual infidelity. *Evolutionary Psychology: An International Journal of Evolutionary Approaches to Psychology and Behavior*, 9(1), 64–78.
- Oesch, N. (2016). Deception as a derived function of language. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01485>
- Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, 41(1), 1–16. <https://doi.org/10.1159/000261706>
- Out, D., Pieper, S., Bakermans-Kranenburg, M. J., Zeskind, P. S., & van IJzendoorn, M. H. (2010). Intended sensitive and harsh caregiving responses to infant crying: The role of cry pitch and perceived urgency in an adult twin sample. *Child Abuse & Neglect*, 34(11), 863–873. <https://doi.org/10.1016/j.chiabu.2010.05.003>
- Owings, D. H., & Morton, E. S. (1998). *Animal vocal communication: A new approach*. Cambridge University Press.
- Owren, M. J. (2011). Human voice in evolutionary perspective. *Acoustics Today*, 7(4), 24–33.
- Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, 73(6), 530–544. <https://doi.org/10.1002/ajp.20913>
- Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews*, 10(2), 58–71.  
<https://doi.org/10.1002/evan.1014>
- Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for

- production processes and functions. *The Journal of the Acoustical Society of America*, 101(5), 2951–2963. <https://doi.org/10.1121/1.418523>
- Paliwal, K. K. (1998). Spectral subband centroid features for speech recognition. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing* (Vol. 2, pp. 617–620). <https://doi.org/10.1109/ICASSP.1998.675340>
- Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, 14(1), 30–80. <https://doi.org/10.1016/j.concog.2004.10.004>
- Parker, J. M., Dillard, T. A., & Phillips, Y. Y. (1994). Impact of using stated instead of measured height upon screening spirometry. *American Journal of Respiratory and Critical Care Medicine*, 150(6 Pt 1), 1705–1708. <https://doi.org/10.1164/ajrccm.150.6.7952637>
- Parsons, C. E., Young, K. S., Craske, M. G., Stein, A. L., & Kringelbach, M. L. (2014). Introducing the Oxford Vocal (OxVoc) Sounds database: a validated set of non-acted affective sounds from human infants, adults, and domestic animals. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00562>
- Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology*, 87(1), 93–98. <https://doi.org/10.1016/j.biopsycho.2011.02.010>
- Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, 111, 14–25. <https://doi.org/10.1016/j.biopsycho.2015.08.008>
- Pereira, C. (2000). Dimensions of emotional meaning in speech. In *SpeechEmotion-2000* (pp. 25–28). Newcastle, Northern Ireland. Retrieved from [http://www.isca-speech.org/archive\\_open/speech\\_emotion/spem\\_025.html](http://www.isca-speech.org/archive_open/speech_emotion/spem_025.html)
- Perlman, M., & Clark, N. (2015). Learned vocal and breathing behavior in an enculturated gorilla. *Animal Cognition*, 18(5), 1165–1179. <https://doi.org/10.1007/s10071-015-0889-6>
- Perlman, M., & Salmi, R. (2017). Gorillas may use their laryngeal air sacs for whinny-type vocalizations and male display. *Journal of Language Evolution*, 2(2), 126–140. <https://doi.org/10.1093/jole/lzx012>

- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184. <https://doi.org/10.1121/1.1906875>
- Pfefferle, D., & Fischer, J. (2006). Sounds and size: identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*. *Animal Behaviour*, 72(1), 43–51. <https://doi.org/10.1016/j.anbehav.2005.08.021>
- Pietraszewski, D., Wertz, A. E., Bryant, G. A., & Wynn, K. (2017). Three-month-old human infants use vocal cues of body size. *Proceedings of the Royal Society B-Biological Sciences*, 284(1856), 20170656. <https://doi.org/10.1098/rspb.2017.0656>
- Pipitone, R. N., & Gallup, G. G. (2012). The unique impact of menstruation on the female voice: implications for the evolution of menstrual cycle cues. *Ethology*, 118(3), 281–291. <https://doi.org/10.1111/j.1439-0310.2011.02010.x>
- Pipitone, R. N., & Gallup Jr, G. G. (2008). Women's voice attractiveness varies across the menstrual cycle. *Evolution and Human Behavior*, 29(4), 268–274. <https://doi.org/10.1016/j.evolhumbehav.2008.02.001>
- Pirke, K. M., Schweiger, U., Lemmel, W., Krieg, J. C., & Berger, M. (1985). The influence of dieting on the menstrual cycle of healthy young women. *The Journal of Clinical Endocrinology and Metabolism*, 60(6), 1174–1179. <https://doi.org/10.1210/jcem-60-6-1174>
- Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice modulation: A window into the origins of human vocal control? *Trends in Cognitive Sciences*, 20(4), 304–318. <https://doi.org/10.1016/j.tics.2016.01.002>
- Pisanski, K., & Feinberg, D. R. (2013). Cross-cultural variation in mate preferences for averageness, symmetry, body size, and masculinity. *Cross-Cultural Research*, 47(2), 162–197. <https://doi.org/10.1177/1069397112471806>
- Pisanski, K., Feinberg, D. R., Oleszkiewicz, A., & Sorokowska, A. (2017). Voice cues are used in a similar way by blind and sighted adults when assessing women's body size. *Scientific Reports*, 7, 10329. <https://doi.org/10.1038/s41598-017-10470-3>
- Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., & Feinberg, D. R. (2014). Return to Oz: Voice pitch facilitates assessments of men's body size. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1316–1331. <https://doi.org/10.1037/a0036956>
- Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., Röder, S., Andrews, P. W., ... Feinberg, D. R. (2014). Vocal indicators of body size in men and women: a

- meta-analysis. *Animal Behaviour*, 95, 89–99. <https://doi.org/10.1016/j.anbehav.2014.06.011>
- Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*, 33(5), 509–519. <https://doi.org/10.1016/j.evolhumbehav.2012.01.004>
- Pisanski, K., Mora, E. C., Pisanski, A., Reby, D., Sorokowski, P., Frackowiak, T., & Feinberg, D. R. (2016). Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. *Scientific Reports*, 6. <https://doi.org/10.1038/srep34389>
- Pisanski, K., Nowak, J., & Sorokowski, P. (2016). Individual differences in cortisol stress response predict increases in voice pitch during exam stress. *Physiology & Behavior*, 163, 234–238. <https://doi.org/10.1016/j.physbeh.2016.05.018>
- Pisanski, K., Oleszkiewicz, A., & Sorokowska, A. (2016). Can blind persons accurately assess body size from the voice? *Biology Letters*, 12(4), 20160063. <https://doi.org/10.1098/rsbl.2016.0063>
- Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, 129(4), 2201–2212. <https://doi.org/10.1121/1.3552866>
- Pitcher, B. J., Briefer, E. F., & McElligott, A. G. (2015). Intrasexual selection drives sensitivity to pitch, formants and duration in the competitive calls of fallow bucks. *BMC Evolutionary Biology*, 15, 149. <https://doi.org/10.1186/s12862-015-0429-7>
- Pitcher, B. J., Briefer, E. F., Vannoni, E., & McElligott, A. G. (2014). Fallow bucks attend to vocal cues of motivation and fatigue. *Behavioral Ecology*, 25(2), 392–401. <https://doi.org/10.1093/beheco/art131>
- Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 976–986. <https://doi.org/10.1037/0096-1523.20.5.976>
- Plant, R. L., & Younger, R. M. (2000). The interrelationship of subglottic air pressure, fundamental frequency, and vocal intensity during speech. *Journal of Voice*, 14(2), 170–177. [https://doi.org/10.1016/S0892-1997\(00\)80024-7](https://doi.org/10.1016/S0892-1997(00)80024-7)
- Plavcan, J. M., & van Schaik, C. P. (1997). Intrasexual competition and body weight dimorphism in anthropoid primates. *American Journal of Physical Anthropology*,

- 103(1), 37–68. [https://doi.org/10.1002/\(SICI\)1096-8644\(199705\)103:1<37::AID-AJPA4>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1096-8644(199705)103:1<37::AID-AJPA4>3.0.CO;2-A)
- Pola, Y. V., & Snowdon, C. T. (1975). The vocalizations of pygmy marmosets (*Cebuella pygmaea*). *Animal Behaviour*, 23, 826–842. [https://doi.org/10.1016/0003-3472\(75\)90108-6](https://doi.org/10.1016/0003-3472(75)90108-6)
- Porter, F. L., Miller, R. H., & Marshall, R. E. (1986). Neonatal pain cries: Effect of circumcision on acoustic features and perceived urgency. *Child Development*, 57(3), 790–802. <https://doi.org/10.2307/1130355>
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3), 715–734. <https://doi.org/10.1017/S0954579405050340>
- Power, M. L., & Schulkin, J. (2008). Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *British Journal of Nutrition*, 99(5), 931–940. <https://doi.org/10.1017/S0007114507853347>
- Pradhan, G. R., Engelhardt, A., Schaik, C. P. van, & Maestripieri, D. (2005). The evolution of female copulation calls in primates: a review and a new model. *Behavioral Ecology and Sociobiology*, 59(3), 333–343. <https://doi.org/10.1007/s00265-005-0075-y>
- Puppe, B., Schön, P. C., Tuchscherer, A., & Manteuffel, G. (2005). Castration-induced vocalisation in domestic piglets, *Sus scrofa*: Complex and specific alterations of the vocal quality. *Applied Animal Behaviour Science*, 95(1), 67–78. <https://doi.org/10.1016/j.applanim.2005.05.001>
- Puts, D. A. (2010). Beauty and the beast: mechanisms of sexual selection in humans. *Evolution and Human Behavior*, 31(3), 157–175. <https://doi.org/10.1016/j.evolhumbehav.2010.02.005>
- Puts, D. A. (2016). Human sexual selection. *Current Opinion in Psychology*, 7, 28–32. <https://doi.org/10.1016/j.copsyc.2015.07.011>
- Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society of London B: Biological Sciences*, 279(1728), 601–609. <https://doi.org/10.1098/rspb.2011.0829>
- Puts, D. A., Bailey, D. H., Cárdenas, R. A., Burriss, R. P., Welling, L. L. M., Wheatley, J. R., & Dawood, K. (2013). Women's attractiveness changes with estradiol and

- progesterone across the ovulatory cycle. *Hormones and Behavior*, 63(1), 13–19. <https://doi.org/10.1016/j.yhbeh.2012.11.007>
- Puts, D. A., Doll, L. M., & Hill, A. K. (2014). Sexual selection on human voices. In *Evolutionary perspectives on human sexual psychology and behavior* (pp. 69–86). Springer, New York, NY. [https://doi.org/10.1007/978-1-4939-0314-6\\_3](https://doi.org/10.1007/978-1-4939-0314-6_3)
- Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, 27(4), 283–296. <https://doi.org/10.1016/j.evolhumbehav.2005.11.003>
- Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., ... Ramos-Fernandez, G. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings of the Royal Society of London B: Biological Sciences*, 283(1829), 20152830. <https://doi.org/10.1098/rspb.2015.2830>
- Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, 28(5), 340–344. <https://doi.org/10.1016/j.evolhumbehav.2007.05.002>
- Rance, P. (2015). War Cry. In *The encyclopedia of the Roman army*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118318140.wbra1621>
- Rantala, M. J., Moore, F. R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., & Krams, I. (2012). Evidence for the stress-linked immunocompetence handicap hypothesis in humans. *Nature Communications*, 3, 694. <https://doi.org/10.1038/ncomms1696>
- Ravignani, A., Gross, S., Garcia, M., Rubio-Garcia, A., & de Boer, B. (2017). How small could a pup sound? The physical bases of signaling body size in harbor seals. *Current Zoology*, (zox026), 1–9. <https://doi.org/10.1093/cz/zox026>
- Reber, S. A., Janisch, J., Torregrosa, K., Darlington, J., Vliet, K. A., & Fitch, W. T. (2017). Formants provide honest acoustic cues to body size in American alligators. *Scientific Reports*, 7, 1816. <https://doi.org/10.1038/s41598-017-01948-1>
- Reby, D., & Charlton, B. D. (2012). Attention grabbing in red deer sexual calls. *Animal Cognition*, 15(2), 265–270. <https://doi.org/10.1007/s10071-011-0451-0>
- Reby, D., Charlton, B. D., Locatelli, Y., & McComb, K. (2010). Oestrous red deer hinds prefer male roars with higher fundamental frequencies. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1695), 2747–2753. <https://doi.org/10.1098/rspb.2010.0467>

- Reby, D., Levréro, F., Gustafsson, E., & Mathevon, N. (2016). Sex stereotypes influence adults' perception of babies' cries. *BMC Psychology*, 4(1). <https://doi.org/10.1186/s40359-016-0123-6>
- Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, 65(3), 519–530. <https://doi.org/10.1006/anbe.2003.2078>
- Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1566), 941–947. <https://doi.org/10.1098/rspb.2004.2954>
- Reby, D., Wyman, M. T., Frey, R., Passilongo, D., Gilbert, J., Locatelli, Y., & Charlton, B. D. (2016). Evidence of biphonation and source–filter interactions in the bugles of male North American wapiti (*Cervus canadensis*). *Journal of Experimental Biology*, 219(8), 1224–1236. <https://doi.org/10.1242/jeb.131219>
- Reid, M., & Duffield, R. (2014). The development of fatigue during match-play tennis. *British Journal of Sports Medicine*, 48(Suppl 1), i7–i11. <https://doi.org/10.1136/bjsports-2013-093196>
- Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *The Journal of the Acoustical Society of America*, 113(6), 3390–3402. <https://doi.org/10.1121/1.1568942>
- Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America*, 117(2), 944–955.
- Rendall, D., Notman, H., & Owren, M. J. (2009). Asymmetries in the individual distinctiveness and maternal recognition of infant contact calls and distress screams in baboons. *The Journal of the Acoustical Society of America*, 125(3), 1792–1805. <https://doi.org/10.1121/1.3068453>
- Rendall, D., Owren, M. J., Weerts, E., & Hienz, R. D. (2004). Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners. *The Journal of the Acoustical Society of America*, 115(1), 411–421.
- Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size. *Journal of Experimental*

- Psychology: Human Perception and Performance*, 33(5), 1208–1219.  
<https://doi.org/10.1037/0096-1523.33.5.1208>
- Riede, T., Arcadi, A. C., & Owren, M. J. (2007). Nonlinear acoustics in the pant hoots of common chimpanzees (*Pan troglodytes*): Vocalizing at the edge. *The Journal of the Acoustical Society of America*, 121(3), 1758–1767.  
<https://doi.org/10.1121/1.2427115>
- Riede, T., Bronson, E., Hatzikirou, H., & Zuberbühler, K. (2005). Vocal production mechanisms in a non-human primate: morphological data and a model. *Journal of Human Evolution*, 48(1), 85–96. <https://doi.org/10.1016/j.jhevol.2004.10.002>
- Riede, T., & Fitch, T. (1999). Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). *Journal of Experimental Biology*, 202(20), 2859–2867.
- Riede, T., Owren, M. J., & Arcadi, A. C. (2004). Nonlinear acoustics in pant hoots of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos. *American Journal of Primatology*, 64(3), 277–291. <https://doi.org/10.1002/ajp.20078>
- Riede, T., & Zuberbühler, K. (2003). The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. *The Journal of the Acoustical Society of America*, 114(2), 1132–1142.  
<https://doi.org/10.1121/1.1580812>
- Rödel, H. G., Landmann, C., Starkloff, A., Kunc, H. P., & Hudson, R. (2013). Absentee mothering – Not so absent? Responses of European rabbit (*Oryctolagus cuniculus*) mothers to pup distress calls. *Ethology*, 119(11), 1024–1033.  
<https://doi.org/10.1111/eth.12149>
- Rogers, J., & Mitchell, G. W. J. (1952). The relation of obesity to menstrual disturbances. *New England Journal of Medicine*, 247(2), 53–55.  
<https://doi.org/10.1056/NEJM195207102470204>
- Rothgänger, H., Hauser, G., Cappellini, A. C., & Guidotti, A. (1998). Analysis of laughter and speech sounds in Italian and German students. *Naturwissenschaften*, 85(8), 394–402. <https://doi.org/10.1007/s001140050522>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–78.

- Ryalls, J. H., & Lieberman, P. (1982). Fundamental frequency and vowel perception. *The Journal of the Acoustical Society of America*, 72(5), 1631–1634. <https://doi.org/10.1121/1.388499>
- Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. *Language Attitudes: Current Trends and Prospects*, 74–84.
- Saez, S., & Sakai, F. (1976). Androgen receptors in human pharyngo-laryngeal mucosa and pharyngo-laryngeal epithelioma. *Journal of Steroid Biochemistry*, 7(11), 919–921. [https://doi.org/10.1016/0022-4731\(76\)90011-X](https://doi.org/10.1016/0022-4731(76)90011-X)
- Samson, M. M., Meeuwsen, I. B., Crowe, A., Dessens, J. A., Duursma, S. A., & Verhaar, H. J. (2000). Relationships between physical performance measures, age, height and body weight in healthy adults. *Age and Ageing*, 29(3), 235–242. <https://doi.org/10.1093/ageing/29.3.235>
- Sanvito, S., & Galimberti, F. (2000). Bioacoustics of southern elephant seals. I. Acoustic structure of male aggressive vocalisations. *Bioacoustics*, 10(4), 259–285. <https://doi.org/10.1080/09524622.2000.9753438>
- Sanvito, S., Galimberti, F., & Miller, E. H. (2007). Vocal signalling of male southern elephant seals is honest but imprecise. *Animal Behaviour*, 73(2), 287–299. <https://doi.org/10.1016/j.anbehav.2006.08.005>
- Sauter, D. A. (2010). Are positive vocalizations perceived as communicating happiness across cultural boundaries? *Communicative & Integrative Biology*, 3(5), 440–442. <https://doi.org/10.4161/cib.3.5.12285>
- Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology*, 63(11), 2251–2272. <https://doi.org/10.1080/17470211003721642>
- Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107(6), 2408–2412. <https://doi.org/10.1073/pnas.0908239106>
- Sauter, D. A., & Fischer, A. H. (2017). Can perceivers recognise emotions from spontaneous expressions? *Cognition and Emotion*, 0(0), 1–12. <https://doi.org/10.1080/02699931.2017.1320978>
- Scanlan, T. K., & Passer, M. W. (1977). Factors related to competitive stress among male youth sport participants. *Medicine and Science in Sports*, 10(2), 103–108.

- Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE*, 8(10), e76674. <https://doi.org/10.1371/journal.pone.0076674>
- Scherer, K. R. (1986a). Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, 99(2), 143–165.
- Scherer, K. R. (1986b). Voice, stress, and emotion. In M. H. Appley & R. Trumbull (Eds.), *Dynamics of stress: Physiological, psychological, and social perspectives* (pp. 157–179). New York, NY, US: Plenum Press.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1–2), 227–256. [https://doi.org/10.1016/S0167-6393\(02\)00084-5](https://doi.org/10.1016/S0167-6393(02)00084-5)
- Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, 46(6), 401–435. <https://doi.org/10.1080/00207594.2011.626049>
- Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433–456). Oxford University Press.
- Scheumann, M., Hasting, A. S., Kotz, S. A., & Zimmermann, E. (2014). The voice of emotion across species: how do human listeners recognize animals' affective states? *PloS One*, 9(3), e91192. <https://doi.org/10.1371/journal.pone.0091192>
- Scheumann, M., Roser, A.-E., Konerding, W., Bleich, E., Hedrich, H.-J., & Zimmermann, E. (2012). Vocal correlates of sender-identity and arousal in the isolation calls of domestic kitten (*Felis silvestris catus*). *Frontiers in Zoology*, 9, 36. <https://doi.org/10.1186/1742-9994-9-36>
- Schmid, B., Karg, K., Perner, J., & Tomasello, M. (2017). Great apes are sensitive to prior reliability of an informant in a gaze following task. *PLOS ONE*, 12(11), e0187451. <https://doi.org/10.1371/journal.pone.0187451>
- Schmidt-Nielsen, K. (1975). Scaling in biology: The consequences of size. *Journal of Experimental Zoology*, 194(1), 287–307. <https://doi.org/10.1002/jez.1401940120>
- Schnitzler, A., & Ploner, M. (2000). Neurophysiology and functional neuroanatomy of pain perception. *Journal of Clinical Neurophysiology*, 17(6), 592–603.
- Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In

- Proceedings of Eurospeech 2001*. Aalborg, Denmark. Retrieved from [http://www.academia.edu/download/45662845/schroeder\\_etal2001.pdf](http://www.academia.edu/download/45662845/schroeder_etal2001.pdf)
- Schweiger, U. (1991). Menstrual function and luteal-phase deficiency in relation to weight changes and dieting. *Clinical Obstetrics and Gynecology*, 34(1), 191–197.
- Scott, S. K., Lavan, N., Chen, S., & McGettigan, C. (2014). The social life of laughter. *Trends in Cognitive Sciences*, 18(12), 618–620. <https://doi.org/10.1016/j.tics.2014.09.002>
- Scott, S. K., Sauter, D., & McGettigan, C. (2010). Brain mechanisms for processing perceived emotional vocalizations in humans. *Handbook of Behavioral Neuroscience*, 19, 187–197. <https://doi.org/10.1016/B978-0-12-374593-4.00019-X>
- Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnsons, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature*, 385(6613), 254–257. <https://doi.org/10.1038/385254a0>
- Sear, R., & Marlowe, F. W. (2009). How universal are human mate choices? Size does not matter when Hadza foragers are choosing a mate. *Biology Letters*, rsbl20090342. <https://doi.org/10.1098/rsbl.2009.0342>
- Sèbe, F., Poindron, P., Andanson, S., Chandeze, H., Delval, E., Despres, G., ... Boissy, A. (2012). Bio-acoustic analyses to assess emotion in animals: Acoustic patterns are linked to behavioural, cardiac and hormonal responses of ewes to the separation from their lambs. *Bioacoustics*, 21(1), 54.
- Sebesta, P., Kleisner, K., Turecek, P., Kocnar, T., Akoko, R. M., Trebicky, V., & Havlicek, J. (2017). Voices of Africa: acoustic predictors of human male vocal attractiveness. *Animal Behaviour*, 127, 205–211. <https://doi.org/10.1016/j.anbehav.2017.03.014>
- Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., ... Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences*, 277(1699), 3509–3518. <https://doi.org/10.1098/rspb.2010.0769>
- Sell, A., Cosmides, L., Tooby, J., Sznycer, D., Rueden, C. von, & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1656), 575–584. <https://doi.org/10.1098/rspb.2008.1177>
- Sell, A., Lukazsweski, A. W., & Townsley, M. (2017). Cues of upper body strength account for most of the variance in men's bodily attractiveness. *Proceedings of the*

- Royal Society of London B: Biological Sciences*, 284(1869), 20171819.  
<https://doi.org/10.1098/rspb.2017.1819>
- Semple, S., & McComb, K. (2000). Perception of female reproductive state from vocal cues in a mammal species. *Proceedings of the Royal Society B: Biological Sciences*, 267(1444), 707–712.
- Semple, S., McComb, K., Alberts, S., & Altmann, J. (2002). Information content of female copulation calls in yellow baboons. *American Journal of Primatology*, 56(1), 43–56. <https://doi.org/10.1002/ajp.1062>
- Senn, C. Y., Eliasziw, M., Barata, P. C., Thurston, W. E., Newby-Clark, I. R., Radtke, H. L., & Hobden, K. L. (2015). Efficacy of a sexual assault resistance program for university women. *The New England Journal of Medicine*, 372(24), 2326–2335. <https://doi.org/10.1056/NEJMsa1411131>
- Senn, C. Y., Eliasziw, M., Hobden, K. L., Newby-Clark, I. R., Barata, P. C., Radtke, H. L., & Thurston, W. E. (2017). Secondary and 2-year outcomes of a sexual assault resistance program for university women. *Psychology of Women Quarterly*, 41(2), 147–162. <https://doi.org/10.1177/0361684317690119>
- Seyfarth, R. M., & Cheney, D. L. (1986). Vocal development in vervet monkeys. *Animal Behaviour*, 34(6), 1640–1658. [https://doi.org/10.1016/S0003-3472\(86\)80252-4](https://doi.org/10.1016/S0003-3472(86)80252-4)
- Shen, W., Punyanitya, M., Wang, Z., Gallagher, D., St-Onge, M.-P., Albu, J., ... Heshka, S. (2004). Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. *Journal of Applied Physiology*, 97(6), 2333–2338. <https://doi.org/10.1152/jappphysiol.00744.2004>
- Shepherd, A., & Cheyne, H. (2013). The frequency and reasons for vaginal examinations in labour. *Women and Birth*, 26(1), 49–54. <https://doi.org/10.1016/j.wombi.2012.02.001>
- Silva, M., Mijovic, B., Van den Bergh, B. R. H., Allegaert, K., Aerts, J.-M., Van Huffel, S., & Berckmans, D. (2010). Decoupling between fundamental frequency and energy envelope of neonate cries. *Early Human Development*, 86(1), 35–40. <https://doi.org/10.1016/j.earlhumdev.2009.12.006>
- Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion*, 9(6), 838–846. <https://doi.org/10.1037/a0017810>
- Sinnett, S., & Kingstone, A. (2010). A preliminary investigation regarding the effect of tennis grunting: does white noise during a tennis shot have a negative impact on

- shot perception? *PLoS ONE*, 5(10), e13148.  
<https://doi.org/10.1371/journal.pone.0013148>
- Slocombe, K. E., Kaller, T., Call, J., & Zuberbühler, K. (2010). Chimpanzees extract social information from agonistic screams. *PLoS ONE*, 5(7), e11473–e11473.  
<https://doi.org/10.1371/journal.pone.0011473>, 10.1371/journal.pone.0011473
- Slocombe, K. E., Townsend, S. W., & Zuberbühler, K. (2009). Wild chimpanzees (*Pan troglodytes schweinfurthii*) distinguish between different scream types: evidence from a playback study. *Animal Cognition*, 12(3), 441–449.  
<https://doi.org/10.1007/s10071-008-0204-x>
- Slocombe, K. E., & Zuberbühler, K. (2005). Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role. *Journal of Comparative Psychology*, 119(1), 67–77.
- Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, 104(43), 17228–17233. <https://doi.org/10.1073/pnas.0706741104>
- Smirnova, D. S., Volodin, I. A., Demina, T. S., & Volodina, E. V. (2016). Acoustic structure and contextual use of calls by captive male and female cheetahs (*Acinonyx jubatus*). *PLOS ONE*, 11(6), e0158546.  
<https://doi.org/10.1371/journal.pone.0158546>
- Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, 118(5), 3177–3186.
- Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *The Journal of the Acoustical Society of America*, 117(1), 305–318.
- Smith, D. S., Jones, B. C., Feinberg, D. R., & Allan, K. (2012). A modulatory effect of male voice pitch on long-term memory in women: evidence of adaptation for mate choice? *Memory & Cognition*, 40(1), 135–144. <https://doi.org/10.3758/s13421-011-0136-6>
- Smith, P. K., & Lewis, K. (1985). Rough-and-tumble play, fighting, and chasing in nursery school children. *Ethology and Sociobiology*, 6(3), 175–181.  
[https://doi.org/10.1016/0162-3095\(85\)90029-9](https://doi.org/10.1016/0162-3095(85)90029-9)

- Södersten, M., Hertegård, S., & Hammarberg, B. (1995). Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women. *Journal of Voice*, 9(2), 182–197. [https://doi.org/10.1016/S0892-1997\(05\)80252-8](https://doi.org/10.1016/S0892-1997(05)80252-8)
- Stevens, S. S. (1935). The relation of pitch to intensity. *The Journal of the Acoustical Society of America*, 6(3), 150–154. <https://doi.org/10.1121/1.1915715>
- Stoeger, A. S., & Baotic, A. (2016). Information content and acoustic structure of male African elephant social rumbles. *Scientific Reports*, 6, 27585. <https://doi.org/10.1038/srep27585>
- Stoeger, A. S., Baotic, A., Li, D., & Charlton, B. D. (2012). Acoustic features indicate arousal in infant giant panda vocalisations. *Ethology*, 118(9), 896–905.
- Stoeger, A. S., Charlton, B. D., Kratochvil, H., & Fitch, W. T. (2011). Vocal cues indicate level of arousal in infant African elephant roars. *The Journal of the Acoustical Society of America*, 130(3), 1700–1710. <https://doi.org/10.1121/1.3605538>
- Stoll, T., Huber, E., Seifert, B., Michel, B. A., & Stucki, G. (2000). Maximal isometric muscle strength: Normative values and gender-specific relation to age. *Clinical Rheumatology*, 19(2), 105–113. <https://doi.org/10.1007/s100670050026>
- Sullivan, M. J. L. (2008). Toward a biopsychomotor conceptualization of pain: Implications for research and intervention. *The Clinical Journal of Pain*, 24(4), 281. <https://doi.org/10.1097/AJP.0b013e318164bb15>
- Sundberg, J., Leanderson, R., von Euler, C., & Knutsson, E. (1991). Influence of body posture and lung volume on subglottal pressure control during singing. *Journal of Voice*, 5(4), 283–291. [https://doi.org/10.1016/S0892-1997\(05\)80057-8](https://doi.org/10.1016/S0892-1997(05)80057-8)
- Sundberg, J., Titze, I., & Scherer, K. R. (1993). Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. *Journal of Voice*, 7(1), 15–29. [https://doi.org/10.1016/S0892-1997\(05\)80108-0](https://doi.org/10.1016/S0892-1997(05)80108-0)
- Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, 65(1), 13–21. <https://doi.org/10.1007/s00265-010-1037-6>
- Syzmanowicz, A., & Furnham, A. (2011). Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta analyses. *Learning and Individual Differences*, 21(5), 493–504. <https://doi.org/10.1016/j.lindif.2011.07.001>

- Szameitat, D. P., Alter, K., Szameitat, A. J., Wildgruber, D., Sterr, A., & Darwin, C. J. (2009). Acoustic profiles of distinct emotional expressions in laughter. *The Journal of the Acoustical Society of America*, 126(1), 354–366. <https://doi.org/10.1121/1.3139899>
- Szameitat, D. P., Darwin, C. J., Szameitat, A. J., Wildgruber, D., & Alter, K. (2011). Formant characteristics of human laughter. *Journal of Voice*, 25(1), 32–37. <https://doi.org/10.1016/j.jvoice.2009.06.010>
- Szameitat, D. P., Darwin, C. J., Wildgruber, D., Alter, K., & Szameitat, A. J. (2011). Acoustic correlates of emotional dimensions in laughter: Arousal, dominance, and valence. *Cognition and Emotion*, 25(4), 599–611. <https://doi.org/10.1080/02699931.2010.508624>
- Taylor, A. A., & Weary, D. M. (2000). Vocal responses of piglets to castration: identifying procedural sources of pain. *Applied Animal Behaviour Science*, 70(1), 17–26. [https://doi.org/10.1016/S0168-1591\(00\)00143-X](https://doi.org/10.1016/S0168-1591(00)00143-X)
- Taylor, A. M., Charlton, B. D., & Reby, D. (2016). Vocal production by terrestrial mammals: Source, filter, and function. In R. A. Suthers, W. T. Fitch, R. R. Fay, & A. N. Popper (Eds.), *Vertebrate sound production and acoustic communication* (pp. 229–259). Springer International Publishing. [https://doi.org/10.1007/978-3-319-27721-9\\_8](https://doi.org/10.1007/978-3-319-27721-9_8)
- Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology*, 280(3), 221–236. <https://doi.org/10.1111/j.1469-7998.2009.00661.x>
- Taylor, A. M., Reby, D., & McComb, K. (2010). Size communication in domestic dog, *Canis familiaris*, growls. *Animal Behaviour*, 79(1), 205–210. <https://doi.org/10.1016/j.anbehav.2009.10.030>
- Tembrock, G. (1976). Canid vocalizations. *Behavioural Processes*, 1(1), 57–75. [https://doi.org/10.1016/0376-6357\(76\)90007-3](https://doi.org/10.1016/0376-6357(76)90007-3)
- Thomas, I. B. (1969). Perceived pitch of whispered vowels. *The Journal of the Acoustical Society of America*, 46(2B), 468–470. <https://doi.org/10.1121/1.1911712>
- Tiezzi, E. B. P., Pulselli, F. M., & Facchini, A. (2004). Determination of pain intensity in newborns by time series analysis. *WIT Transactions on Ecology and the Environment*, 73, 443–450.

- Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, 33(3), 210–216. <https://doi.org/10.1016/j.evolhumbehav.2011.09.004>
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, 85(4), 1699–1707. <https://doi.org/10.1121/1.397959>
- Titze, I. R. (1994). *Principles of voice production*. National Center for Voice and Speech.
- Titze, I. R. (2008). Nonlinear source–filter coupling in phonation: Theory. *The Journal of the Acoustical Society of America*, 123(5), 2733–2749. <https://doi.org/10.1121/1.2832337>
- Titze, I. R. (2017). Human speech: A restricted use of the mammalian larynx. *Journal of Voice*, 31(2), 135–141. <https://doi.org/10.1016/j.jvoice.2016.06.003>
- Titze, I. R., Fitch, W. T., Hunter, E. J., Alipour, F., Montequin, D., Armstrong, D. L., ... Walsh, E. J. (2010). Vocal power and pressure–flow relationships in excised tiger larynges. *Journal of Experimental Biology*, 213(22), 3866–3873. <https://doi.org/10.1242/jeb.044982>
- Titze, I. R., & Story, B. H. (1997). Acoustic interactions of the voice source with the lower vocal tract. *The Journal of the Acoustical Society of America*, 101(4), 2234–2243.
- Tokuda, I., Riede, T., Neubauer, J., Owren, M. J., & Herzel, H. (2002). Nonlinear analysis of irregular animal vocalizations. *The Journal of the Acoustical Society of America*, 111(6), 2908–2919. <https://doi.org/10.1121/1.1474440>
- Townsend, S. W., Charlton, B. D., & Manser, M. B. (2014). Acoustic cues to identity and predator context in meerkat barks. *Animal Behaviour*, 94, 143–149. <https://doi.org/10.1016/j.anbehav.2014.05.021>
- Tracey, I., & Mantyh, P. W. (2007). The cerebral signature for pain perception and Its modulation. *Neuron*, 55(3), 377–391. <https://doi.org/10.1016/j.neuron.2007.07.012>
- Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *J Acoust Soc Am*, 107(6), 3438–3451.
- Trouvain, J., & Truong, K. P. (2012). Comparing non-verbal vocalisations in conversational speech corpora. In L. Devillers, B. Schuller, A. Batliner, P. Rosso, E. Douglas-Cowie, R. Cowie, & C. Pelachaud (Eds.), *Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals (ES3 2012)* (pp. 36–39). Paris, France: European Language

- Resources Association (ELRA). Retrieved from <http://www.lrec-conf.org/proceedings/lrec2012/workshops/18.Proceedings%20ES3%202012.pdf>
- Tsantani, M. S., Belin, P., Paterson, H. M., & McAleer, P. (2016). Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. *Perception*, 45(8), 946–963. <https://doi.org/10.1177/0301006616643675>
- Tusing, K. J., & Dillard, J. P. (2000). The sounds of dominance. *Human Communication Research*, 26(1), 148–171. <https://doi.org/10.1111/j.1468-2958.2000.tb00754.x>
- United States National Park Service. (n.d.). What to do if you see a bear. Retrieved 8 February 2017, from <http://www.nps.gov/yose/planyourvisit/scarebears.htm/>
- van Iersel, T., Timmerman, D., & Mullie, A. (2006). Introduction of a pain scale for palliative care patients with cognitive impairment. *International Journal of Palliative Nursing*, 12(2), 54–59. <https://doi.org/10.12968/ijpn.2006.12.2.20531>
- Vannoni, E., & McElligott, A. G. (2008). Low frequency groans indicate larger and more dominant fallow deer ( *Dama dama* ) males. *PLOS ONE*, 3(9), e3113. <https://doi.org/10.1371/journal.pone.0003113>
- Vannoni, E., Torriani, M. V. G., & McElligott, A. G. (2005). Acoustic signalling in cervids: A methodological approach for measuring vocal communication in fallow deer. *Cognition, Brain and Behaviour*, IX, 551–565.
- Verdonck, A., Gaethofs, M., Carels, C., & de Zegher, F. (1999). Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *European Journal of Orthodontics*, 21(2), 137–143. <https://doi.org/10.1093/ejo/21.2.137>
- Vergne, A. L., Pritz, M. B., & Mathevon, N. (2009). Acoustic communication in crocodilians: from behaviour to brain. *Biological Reviews*, 84(3), 391–411. <https://doi.org/10.1111/j.1469-185X.2009.00079.x>
- Vettin, J., & Todt, D. (2005). Human laughter, social play, and play vocalizations of non-human primates: an evolutionary approach. *Behaviour*, 142(2), 217–240. <https://doi.org/10.1163/1568539053627640>
- Volodin, I. A., Efremova, K. O., Frey, R., Soldatova, N. V., & Volodina, E. V. (2016). Vocal changes accompanying the descent of the larynx during ontogeny from neonates to adults in male and female goitred gazelles (*Gazella subgutturosa*). *Zoology*. <https://doi.org/10.1016/j.zool.2016.09.001>

- Volodin, I. A., Sibiryakova, O. V., & Volodina, E. V. (2016). Sex and age-class differences in calls of Siberian wapiti *Cervus elaphus sibiricus*. *Mammalian Biology - Zeitschrift Für Säugetierkunde*, 81(1), 10–20. <https://doi.org/10.1016/j.mambio.2015.09.002>
- Volodin, I. A., Volodina, E. V., & Frey, R. (2017). Bull bellows and bugles: a remarkable convergence of low and high-frequency vocalizations between male domestic cattle *Bos taurus* and the rutting calls of Siberian and North American wapiti. *Bioacoustics-the International Journal of Animal Sound and Its Recording*, 26(3), 271–284. <https://doi.org/10.1080/09524622.2016.1275805>
- Volodin, I. A., Zaytseva, A. S., Ilchenko, O. G., & Volodina, E. V. (2015). Small mammals ignore common rules: A comparison of vocal repertoires and the acoustics between pup and adult piebald shrews *Diplomesodon pulchellum*. *Ethology*, 121(2), 103–115. <https://doi.org/10.1111/eth.12321>
- Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, 117(1), 338–350. <https://doi.org/10.1121/1.1835958>
- Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research: JSLHR*, 54(4), 995–1010. [https://doi.org/10.1044/1092-4388\(2010/10-0097\)](https://doi.org/10.1044/1092-4388(2010/10-0097))
- Wada, K., Tamakoshi, K., Tsunekawa, T., Otsuka, R., Zhang, H., Murata, C., ... Toyoshima, H. (2005). Validity of self-reported height and weight in a Japanese workplace population. *International Journal of Obesity (2005)*, 29(9), 1093–1099. <https://doi.org/10.1038/sj.ijo.0803012>
- Wagner Jr, W. E. (1992). Deceptive or honest signalling of fighting ability? A test of alternative hypotheses for the function of changes in call dominant frequency by male cricket frogs. *Animal Behaviour*, 44, Part 3, 449–462. [https://doi.org/10.1016/0003-3472\(92\)90055-E](https://doi.org/10.1016/0003-3472(92)90055-E)
- Wattendorf, E., Westermann, B., Fiedler, K., Kaza, E., Lotze, M., & Celio, M. R. (2013). Exploration of the neural correlates of ticklish laughter by functional magnetic resonance imaging. *Cerebral Cortex*, 23(6), 1280–1289. <https://doi.org/10.1093/cercor/bhs094>

- Watts, J. M., & Stookey, J. M. (1999). Effects of restraint and branding on rates and acoustic parameters of vocalization in beef cattle. *Applied Animal Behaviour Science*, 62(2), 125–135. [https://doi.org/10.1016/S0168-1591\(98\)00222-6](https://doi.org/10.1016/S0168-1591(98)00222-6)
- Weary, D. M., Braithwaite, L. A., & Fraser, D. (1998). Vocal response to pain in piglets. *Applied Animal Behaviour Science*, 56(2), 161–172. [https://doi.org/10.1016/S0168-1591\(97\)00092-0](https://doi.org/10.1016/S0168-1591(97)00092-0)
- Weary, D. M., Lawson, G. L., & Thompson, B. K. (1996). Sows show stronger responses to isolation calls of piglets associated with greater levels of piglet need. *Animal Behaviour*, 52(6), 1247–1253. <https://doi.org/10.1006/anbe.1996.0272>
- Weary, D. M., Ross, S., & Fraser, D. (1997). Vocalizations by isolated piglets: a reliable indicator of piglet need directed towards the sow. *Applied Animal Behaviour Science*, 53(4), 249–257. [https://doi.org/10.1016/S0168-1591\(96\)01173-2](https://doi.org/10.1016/S0168-1591(96)01173-2)
- Weckerly, F. W. (1998). Sexual-size dimorphism: influence of mass and mating systems in the most dimorphic mammals. *Journal of Mammalogy*, 79(1), 33–52. <https://doi.org/10.2307/1382840>
- Welch, A. S., & Tschampl, M. (2012). Something to shout about: A simple, quick performance enhancement technique improved strength in both experts and novices. *Journal of Applied Sport Psychology*, 24(4), 418–428. <https://doi.org/10.1080/10413200.2012.688787>
- Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23(3), 349–366. [https://doi.org/10.1016/S0095-4470\(95\)80165-0](https://doi.org/10.1016/S0095-4470(95)80165-0)
- Wheeler, B. C., & Hammerschmidt, K. (2013). Proximate factors underpinning receiver responses to deceptive false alarm calls in wild tufted capuchin monkeys: Is It counterdeception? *American Journal of Primatology*, 75(7), 715–725. <https://doi.org/10.1002/ajp.22097>
- White, R. G., DeShazer, J. A., Tressler, C. J., Borchert, G. M., Davey, S., Waninge, A., ... Clemens, E. T. (1995). Vocalization and physiological response of pigs during castration with or without a local anesthetic. *Journal of Animal Science*, 73(2), 381–386. <https://doi.org/10.2527/1995.732381x>
- Wild, B., Rodden, F. A., Grodd, W., & Ruch, W. (2003). Neural correlates of laughter and humour. *Brain*, 126(10), 2121–2138. <https://doi.org/10.1093/brain/awg226>
- Wilden, I., Herzel, H., Peters, G., & Tembrock, G. (1998). Subharmonics, biphonation, and deterministic chaos in mammal vocalization. *Bioacoustics*, 9(3), 171–196. <https://doi.org/10.1080/09524622.1998.9753394>

- Williams, A. C. de C. (2002). Facial expression of pain: An evolutionary account. *Behavioral and Brain Sciences*, 25(4), 439–455. <https://doi.org/10.1017/S0140525X02000080>
- Wingfield, J. C., Hegner, R. E., Dufty, A. M., & Ball, G. F. (1990). The ‘challenge hypothesis’: Theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. *The American Naturalist*, 136(6), 829–846.
- Wise, L. A., Rothman, K. J., Mikkelsen, E. M., Sørensen, H. T., Riis, A., & Hatch, E. E. (2010). An internet-based prospective study of body size and time-to-pregnancy. *Human Reproduction*, 25(1), 253–264. <https://doi.org/10.1093/humrep/dep360>
- Wittels, P., Johannes, B., Enne, R., Kirsch, K., & Gunga, H.-C. (2002). Voice monitoring to measure emotional load during short-term stress. *European Journal of Applied Physiology*, 87(3), 278–282. <https://doi.org/10.1007/s00421-002-0625-1>
- Wolff, S. E., & Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. *Behavioral Ecology and Sociobiology*, 64(10), 1673–1683. <https://doi.org/10.1007/s00265-010-0981-5>
- Wong, J. S., & Balemba, S. (2016). The effect of victim resistance on rape completion: A meta-analysis. *Trauma, Violence, & Abuse*, 1524838016663934. <https://doi.org/10.1177/1524838016663934>
- Wood, A., Martin, J., & Niedenthal, P. (2017). Towards a social functional account of laughter: Acoustic features convey reward, affiliation, and dominance. *PLOS ONE*, 12(8), e0183811. <https://doi.org/10.1371/journal.pone.0183811>
- Wood, R. M. (2009). Changes in cry acoustics and distress ratings while the infant is crying. *Infant and Child Development*, 18(2), 163–177. <https://doi.org/10.1002/icd.597>
- Wyman, M. T., Mooring, M. S., McCowan, B., Penedo, M. C. T., & Hart, L. A. (2008). Amplitude of bison bellows reflects male quality, physical condition and motivation. *Animal Behaviour*, 76(5), 1625–1639. <https://doi.org/10.1016/j.anbehav.2008.05.032>
- Wyman, M. T., Mooring, M. S., McCowan, B., Penedo, M. C. T., Reby, D., & Hart, L. A. (2012). Acoustic cues to size and quality in the vocalizations of male North American bison, *Bison bison*. *Animal Behaviour*, 84(6), 1381–1391. <https://doi.org/10.1016/j.anbehav.2012.08.037>

- Yancey, G., & Emerson, M. O. (2016). Does height matter? An examination of height preferences in romantic coupling. *Journal of Family Issues*, 37(1), 53–73. <https://doi.org/10.1177/0192513X13519256>
- Yost, W. A. (2000). *Fundamentals of hearing: An introduction* (4th ed.). San Diego, California: Academic Press. Retrieved from <http://asa.scitation.org/doi/abs/10.1121/1.1398047>
- Zahavi, A. (1982). The pattern of vocal signals and the information they convey. *Behaviour*, 80(1), 1–8. <https://doi.org/10.1163/156853982X00409>
- Zahavi, A., & Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle*. Oxford: Oxford University Press.
- Zanchi, P., Fasolo, M., Spinelli, M., Zampini, L., Suttora, C., & Salerni, N. (2016). The role of acoustic and contextual features in the recognition of crying causes. *Psicologia Clinica Dello Sviluppo*, 2016(1), 103–124. <https://doi.org/10.1449/83132>
- Zaytseva, A. S., Volodin, I. A., Ilchenko, O. G., & Volodina, E. V. (2017). Discomfort-related changes in pup ultrasonic calls of fat-tailed gerbils *Pachyuromys duprasi*. *Bioacoustics-the International Journal of Animal Sound and Its Recording*, 26(1), 1–13. <https://doi.org/10.1080/09524622.2016.1164076>
- Zeifman, D. M. (2001). An ethological analysis of human infant crying: Answering Tinbergen's four questions. *Developmental Psychobiology*, 39(4), 265–285. <https://doi.org/10.1002/dev.1005>
- Zeifman, D. M. (2004). Acoustic features of infant crying related to intended caregiving intervention. *Infant and Child Development*, 13(2), 111–122. <https://doi.org/10.1002/icd.344>
- Zhang, C., & Hansen, J. H. L. (2007). Analysis and classification of speech mode: whispered through shouted. In *INTERSPEECH-2007* (pp. 2289–2292). Antwerp, Belgium.
- Zhang, J., & Reid, S. A. (2017). Aggression in young men high in threat potential increases after hearing low-pitched male voices: two tests of the retaliation-cost model. *Evolution and Human Behavior*, 38(4), 513–521. <https://doi.org/10.1016/j.evolhumbehav.2017.02.005>
- Zhang, X. J., Zhang, T. W., Hu, S. J., & Xu, H. (2011). Behavioral assessments of the aversive quality of pain in animals. *Neuroscience Bulletin*, 27(1), 61–67. <https://doi.org/10.1007/s12264-011-1035-3>

- Zhang, Y., & Jiang, J. J. (2005). Spatiotemporal chaos in excised larynx vibrations. *Physical Review E*, 72(3), 035201. <https://doi.org/10.1103/PhysRevE.72.035201>
- Zielinski, W. J., & Vandenberg, J. G. (1993). Testosterone and competitive ability in male house mice, *Mus musculus*: laboratory and field studies. *Animal Behaviour*, 45(5), 873–891. <https://doi.org/10.1006/anbe.1993.1108>

## Appendices

### Appendix 1. Supplementary tables for Chapter 2

*Table A1.* Factor loadings of acoustic variables on the discriminant functions (DF), for both sexes combined. Highest factor loadings (for each acoustic variable) are highlighted in bold.

Acoustic variable	DF1	DF2	DF3
	Variance = 75%	Variance = 21%	Variance = 3%
	Eigenvalue = 4.28	Eigenvalue = 1.22	Eigenvalue = 0.19
Mean amplitude (dB)	<b>-.63</b>	-.03	.34
Intensity CV (dB)	<b>.51</b>	-.12	.39
Amplitude modulation (%)	<b>-.37</b>	-.03	.06
Major F0 inflections	<b>.20</b>	-.11	-.14
Minimum F0 (Hz)	-.27	<b>.61</b>	.13
Max F0 (Hz)	-.21	<b>.57</b>	.30
Mean F0 (Hz)	-.28	<b>.57</b>	.33
HNR (dB)	.06	<b>.44</b>	.07
Jitter (Hz)	-.19	<b>-.24</b>	-.24
Shimmer (dB)	-.03	<b>.21</b>	-.21
Dominant frequency F4 (Hz)	.03	<b>.21</b>	-.01
Start – end F0 (Hz)	-.02	<b>-.20</b>	.02
Time of max intensity (%)	.13	.06	<b>-.38</b>
Centre of gravity (Hz)	-.20	.02	<b>.33</b>
F0 CV (Hz)	.05	-.01	<b>.28</b>
Minor F0 inflections	.09	.00	<b>-.23</b>

*Table A2.* Factor loadings of acoustic variables on the discriminant functions (DF), for male vocalisers only. Highest factor loadings (for each acoustic variable) are highlighted in bold.

Acoustic variable	DF1	DF2	DF3
	Variance = 75%	Variance = 21%	Variance = 3%
	Eigenvalue = 4.28	Eigenvalue = 1.22	Eigenvalue = 0.19
Mean amplitude (dB)	<b>-.64</b>	-.06	-.36
Intensity CV (dB)	<b>.51</b>	-.04	-.22
Amplitude modulation (%)	<b>-.49</b>	-.23	.13
Major F0 inflections	<b>.18</b>	-.13	.03
Max F0 (Hz)	-.20	<b>.62</b>	-.24
Minimum F0 (Hz)	-.33	<b>.54</b>	-.16
Mean F0 (Hz)	-.31	<b>.54</b>	-.28
HNR (dB)	.13	<b>.45</b>	.00
Dominant frequency F4 (Hz)	.05	<b>.25</b>	-.05
Shimmer (dB)	-.05	<b>.22</b>	.12
Start – end F0 (Hz)	-.05	<b>-.16</b>	.15
Jitter (Hz)	-.24	-.26	<b>.33</b>
Centre of gravity (Hz)	-.26	.00	<b>-.30</b>
F0 CV (Hz)	.13	.07	<b>-.21</b>
Time of max intensity (%)	.14	.02	<b>.20</b>
Minor F0 inflections	.04	-.09	<b>.18</b>

*Table A3.* Factor loadings of acoustic variables on the discriminant functions (DF), for female vocalisers only. Highest factor loadings (for each acoustic variable) are highlighted in bold.

Acoustic variable	DF1	DF2	DF3
	Variance = 75%	Variance = 21%	Variance = 3%
	Eigenvalue = 4.28	Eigenvalue = 1.22	Eigenvalue = 0.19
Mean amplitude (dB)	<b>.57</b>	-.03	.26
Intensity CV (dB)	<b>-.48</b>	-.14	.43
Amplitude modulation (%)	<b>.26</b>	.09	.23
Minimum F0 (Hz)	.21	<b>.58</b>	.01
Mean F0 (Hz)	.24	<b>.49</b>	.22
Max F0 (Hz)	.21	<b>.45</b>	.19
HNR (dB)	-.01	<b>.40</b>	.05
Jitter (Hz)	.14	<b>-.21</b>	-.06
Start – end F0 (Hz)	.00	<b>-.20</b>	.18
Dominant frequency F4 (Hz)	-.01	<b>.15</b>	-.09
Time of max intensity (%)	-.10	.08	<b>-.43</b>
Centre of gravity (Hz)	.13	.03	<b>.26</b>
F0 CV (Hz)	.01	-.05	<b>.24</b>
Shimmer (dB)	.02	.18	<b>-.23</b>
Minor F0 inflections	-.11	.07	<b>-.19</b>
Major F0 inflections	-.19	-.06	<b>-.19</b>

*Table A4.* Standardised regression coefficients for acoustic predictors of men and women's physical strength. Separate stepwise regressions were computed for aggressive speech, aggressive roars, distressed speech, and distress screams. \*  $p < .05$  \*\*  $p < .01$

Acoustic variable	Females				Males			
	Aggr. speech	Aggr. roar	Distr. speech	Distr. scream	Aggr. speech	Aggr. roar	Distr. speech	Distr. scream
Duration (s)								
Mean F0 (Hz)								
Max F0 (Hz)								
Min F0 (Hz)								
Start – end F0 (Hz)								
F0CV (Hz)								
Minor F0 inflections								
Major F0 inflections	.42*							
Mean intensity (dB)								
Time of max intensity (%)								
Intensity CV (dB)								
Shimmer (dB)								
Jitter (Hz)								
HNR (dB)								
Amplitude modulation (%)			.55**					
Centre of gravity (Hz)								
Dominant frequency F4 (Hz)	-.47**	-.47**	-.32*					

[illegible]

*Table A6.* Standardised regression coefficients for acoustic predictors of listeners' ratings of men and women's physical strength. Separate stepwise regressions were computed for aggressive speech, aggressive roars, distressed speech, and distress screams. \*  $p < .05$  \*\*  $p < .01$

Acoustic variable	Females				Males			
	Aggr. speech	Aggr. roar	Distr. speech	Distr. scream	Aggr. speech	Aggr. roar	Distr. speech	Distr. scream
Duration (s)		<b>.53<sup>***</sup></b>		<b>.21<sup>*</sup></b>			<b>-.19<sup>*</sup></b>	
Mean F0 (Hz)								<b>-.49<sup>***</sup></b>
Max F0 (Hz)			<b>-.29<sup>***</sup></b>					
Min F0 (Hz)								
Start – end F0 (Hz)						<b>.40<sup>**</sup></b>		
F0CV (Hz)		<b>-.38<sup>***</sup></b>					<b>-.28<sup>**</sup></b>	<b>-.18<sup>*</sup></b>
Minor F0 inflections								
Major F0 inflections								
Mean intensity (dB)	<b>.80<sup>***</sup></b>	<b>.86<sup>***</sup></b>	<b>1.11<sup>***</sup></b>	<b>.71<sup>***</sup></b>	<b>.37<sup>**</sup></b>	<b>.36<sup>*</sup></b>	<b>.95<sup>***</sup></b>	<b>.60<sup>***</sup></b>
Time of max intensity (%)								
Intensity CV (dB)						<b>-.34<sup>*</sup></b>		
Shimmer (dB)		<b>-.33<sup>**</sup></b>			<b>-.18<sup>*</sup></b>			
Jitter (Hz)						<b>.43<sup>**</sup></b>	<b>-.32<sup>*</sup></b>	
HNR (dB)	<b>-.35<sup>***</sup></b>	<b>-.72<sup>***</sup></b>	<b>-.34<sup>***</sup></b>				<b>-.76<sup>***</sup></b>	<b>-.23<sup>*</sup></b>
Amplitude modulation (%)				<b>.40<sup>***</sup></b>	<b>.58<sup>***</sup></b>			<b>.40<sup>**</sup></b>
Centre of gravity (Hz)								
Dominant frequency F4 (Hz)								

*Table A7.* Standardised regression coefficients for acoustic predictors of listeners' ratings of men and women's height. Separate stepwise regressions were computed for aggressive speech, aggressive roars, distressed speech, and distress screams. \*  $p < .05$  \*\*  $p < .01$

Acoustic variable	Females				Males			
	Aggr. speech	Aggr. roar	Distr. speech	Distr. scream	Aggr. speech	Aggr. roar	Distr. speech	Distr. scream
Duration (s)								
Mean F0 (Hz)		-.54***			-.49*	-.54**	-1.34***	-.80***
Max F0 (Hz)			-.34*					
Min F0 (Hz)							.69*	
Start – end F0 (Hz)								
F0CV (Hz)								
Minor F0 inflections								
Major F0 inflections								
Mean intensity (dB)		.70***	.78***		.76**	.32*	.90***	.46**
Time of max intensity (%)						-.28*		
Intensity CV (dB)			.32**					
Shimmer (dB)		-.38*						
Jitter (Hz)						.42**		.34*
HNR (dB)			-.45***					
Amplitude modulation (%)								
Centre of gravity (Hz)	.48**			-.52**				
Dominant frequency F4 (Hz)								