



## Research

**Cite this article:** Filippi P *et al.* 2017 Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proc. R. Soc. B* **284**: 20170990.

<http://dx.doi.org/10.1098/rsob.2017.0990>

Received: 11 May 2017

Accepted: 21 June 2017

**Subject Category:**

Evolution

**Subject Areas:**

evolution, cognition

**Keywords:**

emotional arousal, language evolution, vocal communication, cross-species communication, acoustic universals, emotional prosody

**Author for correspondence:**

Piera Filippi

e-mail: [pie.filippi@gmail.com](mailto:pie.filippi@gmail.com)

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.3825532.v4>.

# Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals

Piera Filippi<sup>1,2,3,4</sup>, Jenna V. Congdon<sup>5</sup>, John Hoang<sup>5</sup>, Daniel L. Bowling<sup>6</sup>, Stephan A. Reber<sup>6</sup>, Andrius Pašukonis<sup>6</sup>, Marisa Hoeschele<sup>6</sup>, Sebastian Ocklenburg<sup>7</sup>, Bart de Boer<sup>1</sup>, Christopher B. Sturdy<sup>5,9</sup>, Albert Newen<sup>2,8</sup> and Onur Güntürkün<sup>2,7</sup>

<sup>1</sup>Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

<sup>2</sup>Center for Mind, Brain and Cognitive Evolution, Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany

<sup>3</sup>Brain and Language Research Institute, Aix-Marseille University, Avenue Pasteur 5, 13604 Aix-en-Provence, France

<sup>4</sup>Department of Language and Cognition, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands

<sup>5</sup>Department of Psychology, University of Alberta, P217 Biological Sciences Building, Edmonton, Alberta, Canada T6G 2E9

<sup>6</sup>Department of Cognitive Biology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria

<sup>7</sup>Department of Biopsychology, and <sup>8</sup>Department of Philosophy II, Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany

<sup>9</sup>Neuroscience and Mental Health Institute, University of Alberta, 4-120 Katz Group Center, Edmonton, Alberta, Canada T6G 2E1

PF, 0000-0002-2990-9344; DLB, 0000-0002-5303-5472; SAR, 0000-0002-8819-7009; AP, 0000-0002-5742-8222

Writing over a century ago, Darwin hypothesized that vocal expression of emotion dates back to our earliest terrestrial ancestors. If this hypothesis is true, we should expect to find cross-species acoustic universals in emotional vocalizations. Studies suggest that acoustic attributes of aroused vocalizations are shared across many mammalian species, and that humans can use these attributes to infer emotional content. But do these acoustic attributes extend to non-mammalian vertebrates? In this study, we asked human participants to judge the emotional content of vocalizations of nine vertebrate species representing three different biological classes—Amphibia, Reptilia (non-aves and aves) and Mammalia. We found that humans are able to identify higher levels of arousal in vocalizations across all species. This result was consistent across different language groups (English, German and Mandarin native speakers), suggesting that this ability is biologically rooted in humans. Our findings indicate that humans use multiple acoustic parameters to infer relative arousal in vocalizations for each species, but mainly rely on fundamental frequency and spectral centre of gravity to identify higher arousal vocalizations across species. These results suggest that fundamental mechanisms of vocal emotional expression are shared among vertebrates and could represent a homologous signalling system.

## 1. Introduction

Emotions are triggered by specific events and are based in physiological states that increase an animal's ability to respond appropriately to threats or danger in the surrounding environment [1]. Emotional states can be classified according to their valence (positive or negative) and their arousal level (i.e. activation

or responsiveness levels, classified as high or low) [2]. Our study focuses on arousal, which, following [3], we define as a state of the brain or the body reflecting responsiveness to sensory stimulation, ranging from sleep to frenetic excitement. Accordingly, increases in arousal are correlated with increases in behavioural, hormonal and/or neurological activity [2]. Critically, in vocalizing animals, a heightened state of arousal may be reflected in acoustic modulation of the voice. The connection between emotion and the voice can be understood in terms of the effects of emotional physiology on the physical mechanisms of voice production. According to the source–filter theory of voice production [4,5], vocalizations are generated by tissue vibrations stimulated by the passage of air in the sound ‘source’: the larynx in mammals, amphibians and non-avian reptiles, and the syrinx in birds. The signal produced by the source is subsequently filtered by the resonances of the supralaryngeal vocal tract (the ‘filter’) with certain frequencies being enhanced or attenuated. Source vibration determines the fundamental frequency of the vocalization ( $F_0$ ), and filter resonances shape its spectral content, producing concentrations of acoustic energy in particular frequency bands (called ‘formants’) [6–11]. For instance, when humans vocalize, air passes from the lungs through an opening between the vocal folds, causing them to vibrate. These vibrations are transmitted through the air in the vocal tract to the openings of the mouth and nose, where they are broadcast into the environment. Although physiological changes associated with emotional arousal affect this process in numerous ways, their effects on the muscular actions required for vocal production (e.g. of the diaphragm, intercostals and vocalis muscles) are likely to be critical, because they alter the way air flows through the system and thus the quality of the sounds produced [5].

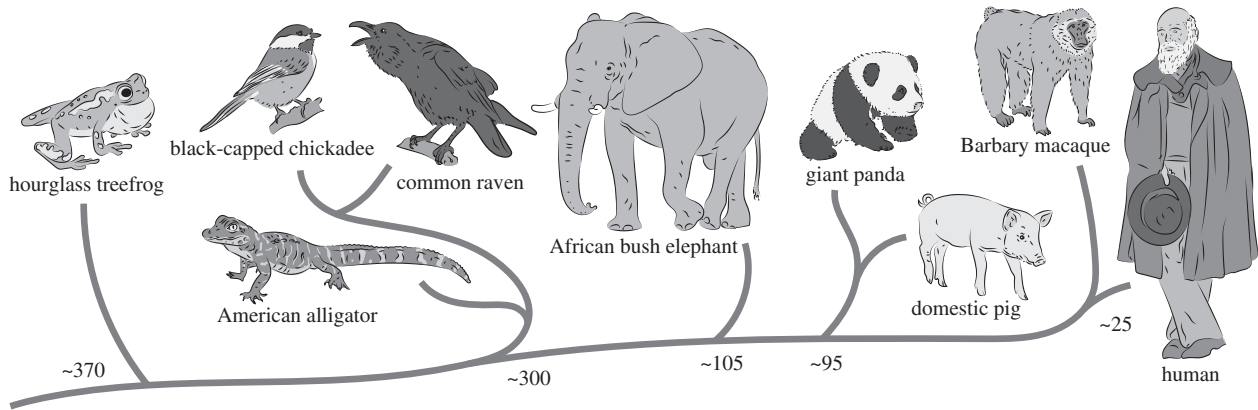
In *The descent of man*, Darwin [12] hypothesized that vocal emotional expression has ancient roots, painting a picture of how the first steps in the evolution of laryngeal vocalization may have proceeded: ‘All the air-breathing Vertebrata necessarily possess an apparatus for inhaling and expelling air, with a pipe capable of being closed at one end. Hence when the primaevial members of this class were strongly excited and their muscles violently contracted, purposeless sounds would almost certainly have been produced; and these, if they proved in any way serviceable, might readily have been modified or intensified by the preservation of properly adapted variations’ (p. 631). If this hypothesis is correct, we should expect that fundamental aspects of vocal emotional expression are shared across all extant species that trace their ancestry to early terrestrial tetrapods. Consequently, it should be possible to (i) identify acoustic universals that convey the same emotional information across a broad range of vocalizing species and (ii) use these universals to correctly infer emotional status at an interspecific level.

In an influential study, Morton [13] proposed that harsh, low-frequency vocalizations are used in agonistic contexts, whereas more tonal, high-frequency vocalizations are used in fearful or appeasing contexts in mammals and birds. Recent studies align with this hypothesis, suggesting that increases in frequency-related parameters of the voice (e.g. fundamental frequency, frequency range and spectra shape), amplitude contours and vocalization rate, as well as decreases in the temporal interval between vocal bouts, are reliable acoustic correlates of high arousal in numerous mammalian species (see [14] for a review) and birds (black-capped chickadee, *Poecile atricapillus*

[15]; common raven, *Corvus corax* [16]). A few studies have addressed vocal correlates of arousal also in anurans and non-avian reptiles. Indeed, timing and frequency-related parameters seem to vary in response to escalating male-male competition, hence correlating with different arousal states, in some species of frogs (grey treefrog, *Hyla versicolor* [17,18]; hourglass treefrog, *Dendropsophus ebraccatus* [19,20]; African reed frog, *Hyperolius marmoratus* [21]; golden rocket frog, *Anomaloglossus beebei* [22]; neotropical treefrog, *Hyla ebraccata* [23]). In addition, frequency-related parameters and intensity parameters have been shown to correlate with arousal in the Australian freshwater crocodile (*Crocodylus johnstoni*) [24]. Notably, although the acoustic correlates of arousal have been extensively investigated in mammals, very few studies have addressed this issue in non-mammalian species.

As to the acoustic encoding of valence, based on her review of acoustic correlates of emotional valence in mammalian vocalizations, Briefer [14] identified duration as the only acoustic parameter that consistently changes as a function of valence, with positively valenced vocalizations being shorter than negatively valenced vocalizations. However, a systematic empirical investigation of emotional valence acoustic encoding in non-mammalian species is still lacking.

In parallel with comparative studies on the productive aspects of vocal emotional communication, several studies have also examined perceptual aspects. Studies have examined humans’ perception of arousal in vocalizations of a number of species including humans [25,26], piglets [27], dogs [28,29] and cats [30]. Taken together, these studies suggest that humans rely mainly on increases in fundamental frequency ( $F_0$ ) to rate both human and heterospecific vocalizations as expressing heightened levels of arousal. Sauter *et al.* [25] found that in addition to higher average  $F_0$ , shorter duration, more amplitude onsets, lower minimum  $F_0$  and less  $F_0$  variation predict humans’ higher ratings for arousal in human vocalizations. Within this framework, research shows that mule deer (*Odocoileus hemionus*) and white-tailed deer (*Odocoileus virginianus*) mothers respond to infant distress vocalizations of a number of mammalian species if the  $F_0$  falls within the deer’s frequency response range [31] and that the high-pitched, quickly-pulsating whistles of human shepherds have an activating effect on dogs [32]. Research on valence perception suggests that humans rate domestic piglets’ vocalizations with increased  $F_0$  and duration as more negative [27]. Faragó [28] showed that humans rate human and dog vocalizations with shorter duration, and human vocalizations with lower spectral centre of gravity (hereafter SCG), as more positive (but see [29]). Moreover, humans can correctly classify the emotional content of vocalizations produced by human infants, chimpanzees (*Pan troglodytes*) [33], domestic pigs (*Sus scrofa domestica*) [27,34], dogs (*Canis familiaris*) [33,35] and cats (*Felis catus*) [30,36], based on vocal production contexts varying in emotional dimensions (e.g. agonistic or food-related contexts [37]). Crucially, it has been found that dogs can identify emotional valence in both conspecific and human vocalizations [38]. Notably, these perceptual studies have focused exclusively on mammalian vocalizations. Hence, the question of whether the ability to identify emotional information in vocalizations is preserved across phylogenetically more distant species remains open. Empirical evidence in this domain is highly relevant to Darwin’s hypothesis on the shared mechanisms for vocal emotional expression, which extends across all air-breathing tetrapods [12,39].



**Figure 1.** Phylogeny of the animal species included in our experiment. Divergence times (millions of years) based on molecular data are presented at the nodes [46–48].

Research on arousal and valence perception across diverse animal classes may provide crucial insights on the adaptive value of vocal emotional expression in animals. As a first step into this research direction, the present study addresses arousal perception. The ability to correctly identify heightened arousal in vocalizations expressed through modulation of specific acoustic parameters allows animals to perceive heightened levels of threat or danger, and is thus critically important for reacting adaptively [40]. From an evolutionary standpoint, this perceptual ability provides a critical complement to the encoding of emotion in vocal production. Importantly, research suggests that animals actually use this information at a heterospecific level, integrating information gained from heterospecific vocalizations with information gained from conspecifics to determine appropriate behavioural reactions in response to potential environmental dangers [41–45]. Therefore, the investigation of vocal emotional communication in animals across all classes is key to enhance our understanding of the link between vocal signals and their adaptive nature, shedding light on the evolution of acoustic communication.

To summarize, much research has identified the acoustic correlates of vocal emotional expression in mammalian species mainly. In parallel, research has investigated the perception of emotional content across a few mammalian species. However, no studies have examined the acoustic parameters that predict the perception of arousal in non-mammals' vocalizations, and, critically, research on the acoustic universals that are responsible for arousal recognition across all classes of terrestrial tetrapods is still lacking. To address these gaps, and to shed light on Darwin's account of the ancient origins of vocal emotional communication, we examined the ability of humans to discriminate levels of arousal in the vocalizations of nine phylogenetically diverse species (figure 1): the hourglass treefrog (*Dendropsophus ebraccatus*), American alligator (*Alligator mississippiensis*), black-capped chickadee (*Poecile atricapillus*), common raven (*Corvus corax*), domestic pig (*Sus scrofa domestica*), giant panda (*Ailuropoda melanoleuca*), African bush elephant (*Loxodonta africana*), Barbary macaque (*Macaca sylvanus*) and human (*Homo sapiens*). Notably, these species span all classes of terrestrial tetrapods, including amphibians, reptiles (non-aves and aves) and mammals. We predicted that humans would be able to identify different levels of arousal across all classes of terrestrial tetrapods, and that, if this were a biologically rooted ability, it would be observed across multiple language groups. Furthermore, we performed

acoustic analyses to identify the acoustic parameters this ability is based on.

## 2. Material and methods

### (a) Experimental design

Participants were informed that the aim of the study was to understand whether humans are able to identify different levels of arousal expressed in animal vocalizations. We provided the following definition of arousal: 'Arousal is a state of the brain or the body reflecting responsiveness to sensory stimulation. Arousal level typically ranges from low (very subdued) to high (very excited). Examples of low arousal states (i.e. of low responsiveness to sensory stimulation) are calmness or boredom. Examples of high arousal states (i.e. of high responsiveness to sensory stimulation) are anger or excitement.' Because we included human language stimuli, we also made sure that none of the participants could speak or understand the language of these stimuli (which was Tamil), in order to exclude any influence of the semantic content on the perception of arousal.

For familiarization with the experimental procedure, each participant completed five practice trials, each consisting of a pair of human baby cries (obtained from [www.freesound.org](http://www.freesound.org)) varying in arousal. During this practice phase, explicit instructions on the experimental procedure were displayed on the monitor. In the subsequent experimental phase, ninety pairs of vocalizations (ten for each species) were played in a randomized order across participants. Each trial in both phases was divided into two phases. During the sound playback phase, one low and one high arousal vocalization produced by the same individual were played with an inter-stimulus interval of 1 s. Stimulus order within pairs was randomized across participants. The letters 'A' and 'B' appeared on the screen in correspondence with the first and second vocalization, respectively. During the subsequent relative rating phase, participants were asked to indicate which sound expressed a higher level of arousal by clicking on the corresponding letter with the mouse. Participants could replay each sound *ad libitum* by pressing either letter (A or B) on the keyboard. No feedback was provided.

### (b) Participants

25 English speakers (mean age = 19.4 years; s.d. = 1.87 years; 12 female) and 25 native Mandarin speakers (mean age = 19.96 years; s.d. = 1.45 years, 12 female) recruited at the University of Alberta (Canada), and 25 native German speakers (mean age = 22.8 years; s.d. = 5.67 years; 22 female) recruited at the

Ruhr University of Bochum (Germany), participated in this experiment in exchange for course credit. The experimental design adopted for this study was approved by both universities' ethical review panels in accordance with the Helsinki Declaration. All participants gave written informed consent.

## (c) Acoustic stimuli

### (i) Recordings and arousal classification

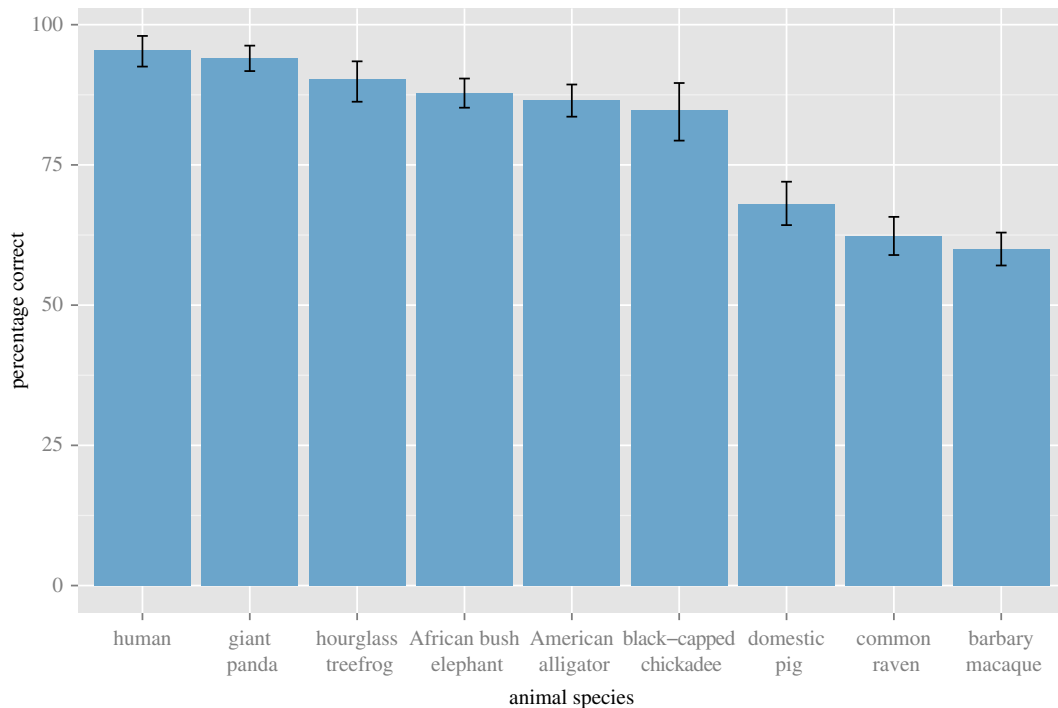
We gathered 180 recordings of vocalizations from nine different vertebrate species. These recordings were obtained from published studies for the hourglass treefrog [19,20], African bush elephant [49], giant panda [50], domestic pig [51], Barbary macaque [52] and human [53] species, and from unpublished work for the American alligator, common raven and black-capped chickadee. Except for recordings of the American alligator (made by S.A.R.), the common raven (made by A.P.) and the black-capped chickadee (made by J.V.C. and J.H.), classifications of arousal level as either high/low arousal for the purposes of this study were made in accordance with criteria presented in the original studies from which they were taken (electronic supplementary material, table S1). For the American alligator, arousal was assessed based on the status of the palatal valve (open or closed), which has been shown to correlate with different arousal states [24]. For the common raven, arousal was assessed based on the type of physical confrontation with a dominant individual [16,54]. For the chickadees, arousal was assessed based on research showing increase of neural activity in response to high-threat predator models [55]. For human recordings, speakers were instructed to express emotions (sadness and anger) that vary in arousal intensity [53]. Classification of vocalizations as high or low arousal (i.e. of high and low activation or responsiveness levels [3]) for the remaining species was based on the following indicators: escalating level of competition during sexual advertisement for the hourglass treefrog [19,20], occurrences of physiological responses (namely secretions from the temporal glands [56]) and of ears, head and tail movements for the African bush elephant [57,58] (as reported in [49]), increased motor activity for giant pandas [50] and domestic pigs [51], and increased temporal distance from the moment of disturbance originating the vocalization to the point where lack of danger is assessed for Barbary macaques [52]. Although these indicators vary across species, they provide clear correlates of relatively low or high arousal within each species, which is appropriate given that the perceptual decisions our participants were asked to make were always within a species. Importantly, we make no claim that absolute arousal levels are comparable between species. As detailed in electronic supplementary material, table S1, these indicators generally reflect the degree of threat, competition or disturbance (low or high) posed by external stimuli within the behavioural context of vocal production. Hence, all vocalizations expressing high arousal, as well as all vocalizations expressing low arousal, are considered negatively valenced. For each species, 10 pairs of low/high arousal vocalizations recordings were used. Within each pair, the vocalizations were always produced by the same individual. Vocalizations were produced by 6–10 different individuals in each species. Stimuli from the following species were produced by juveniles: American alligator, domestic pig, giant panda, African bush elephant. Hourglass treefrog vocalizations consisted of a sequence of pulses, sometimes followed by clicks [19,20]. Black-capped chickadee vocalizations consisted of a sequence of notes. Human vocalizations consisted of a Tamil sentence spoken with emotional intonation. For all the other species, each vocalization consisted of one unit. Since our stimuli were recorded in different experimental settings (using different recording equipment), and at different distances from the vocalizing animal, all vocalization recordings were equalized to the same root-mean-square amplitude. Fade in/out transitions of 5 ms were applied to all files to remove any transients.

### (ii) Acoustic analysis

To explore how specific acoustic cues affect humans' ratings of arousal across animal taxa, we measured four parameters for each stimulus:  $F_0$ , tonality (harmonics-to-noise-ratio, HNR), SCG and duration.  $F_0$  and HNR are related to the tension in the vocal fold and are reliable indicators of the emotional state of the vocalizing individual [4,13]. Duration is also typically measured as a parameter linked to the emotional state of the vocalizing individual [14,59]. Finally, SCG has been found to affect the perception of arousal in humans [25,28]. Acoustic analyses were performed in PRAAT v. 5.2.26 ([www.praat.org](http://www.praat.org)) [60] and SIGNAL v. 5.00.24 sound analysis software (Engineering Design, RTS, Berkeley, California, USA) (electronic supplementary material, table S2). For the measurements of duration and SCG, analysis settings were identical for all the recordings. Duration was measured in SIGNAL. SCG was measured in PRAAT using the 'To spectrum' and 'Get center of gravity' commands (power = 2.0). The analysis of  $F_0$  was performed in PRAAT, and restricted to vocalizations with clear harmonics visible in the spectrogram.  $F_0$  measurements were made using the 'Get pitch' algorithm. Typically 'Pitch floor' and 'Pitch ceiling', but sometimes also 'Silence threshold' and 'Voicing threshold' within the 'Advanced pitch settings' menu, were adjusted until the values identified by the algorithm visually matched the frequency distance between harmonics seen in the PRAAT spectrogram view window. If harmonics could not be identified (e.g. in the presence of subharmonics or bifurcations),  $F_0$  was not measured. Following these criteria,  $F_0$  was measured in 84/90 vocalizations (93.33%). Because HNR can only be measured in vocalizations with  $F_0$ , HNR measurements were also limited to this subset of our data. HNR was measured using the 'To Harmonicity (cc)' command in PRAAT (time step = 0.01 s; minimum frequency based on the settings used for  $F_0$  measurement of the same vocalization). Each vocalization of the hourglass treefrogs, black-capped chickadees and humans, which consisted of a sequence of units (pulses sometimes followed by clicks, notes and words, respectively), were analysed as a whole stimulus, averaging across the entire vocalization.

### (d) Statistical analysis

Statistical analyses were performed in R Studio v. 1.0.136 [61]. In order to assess participants' accuracy for each species, we performed a binomial test. In order to assess any effect of repetition of sound playback on participants' responses within the relative rating task, we computed a binary logistic regression model within the generalized linear model framework. Furthermore, we used a generalized linear mixed model (GLMM) to compare humans' accuracy in identifying the vocalization that expressed a higher level of arousal within each trial across language groups. Within this model, we assessed the effect of language group and acoustic parameters of the vocalizations on humans' accuracy across our species sample. Data across all participants were modelled using a binomial distribution with a logit link function. The dependent variable was a binary response (i.e. correct or incorrect response). Participant's response was correct when they identified the vocalization expressing the higher level of arousal, as was independently assessed based on arousal indicators. Participant, animal species and behavioural context of vocalization were entered as random factors. Ratios of  $F_0$ , HNR, SCG and duration between low and high arousal vocalizations within each trial and language group were entered as fixed factors (glmer function, lme4 library). We assessed the statistical significance of each factor by comparing the model with and without the factor included using likelihood-ratio tests. Pairwise comparisons within language groups were performed within the same model (glht function, multcomp library), using the Holm–Bonferroni correction procedure [62]. To assess which acoustic parameters affected human ability to identify the vocalization expressing a higher level of arousal within each species, we performed separate GLMMs for each species. These models were identical to the one described above, except that



**Figure 2.** Accuracy across animal species. Percentage of correct responses for each animal species in the relative rating of arousal. Here participants were instructed to indicate which sound expressed a higher level of arousal after one low- and one high-arousal vocalization emitted by the same individual were played. (Online version in colour.)

only *participant ID* was entered as a random factor. For all the analyses within GLMMs, we used a model selection procedure based on Akaike's information criterion adjusted for small sample size (AICc) to identify the model(s) with the highest power to explain variation in the dependent variable [63,64]. AICc was used to rank the GLMMs and to obtain model weights (model.sel function, MuMIn library). Selection of the models (i.e. of model(s) with the highest power to explain variation in the dependent variable) is based on lowest AICc. Models with  $AICc \leq 2$  compared with the best model's AICc are considered as good as the best model [64].

### 3. Results

The binomial test revealed that the proportion of correct answers was higher than expected by chance (50%) for all species (hourglass treefrog: 90%; American alligator: 87%; black-capped chickadee: 85%; common raven: 62%; domestic pig: 68%; giant panda: 94%; African bush elephant: 88%; Barbary macaque: 60%; human: 95%;  $p < 0.001$  for all species; figure 2). Our analysis did not reveal any significant effect of number of repetitions on responses (effect of number of repetitions of 'A':  $z = -1.728$ ,  $p = 0.08$ ; effect of number of repetitions of 'B':  $z = 1.866$ ,  $p = 0.06$ ).

As detailed in electronic supplementary material, table S3, our analyses revealed a significant effect of  $F_0$  and SCG ratios for identification of vocalizations expressing a higher level of arousal within each vocalization pair. Specifically, increases in  $F_0$  and SCG ratios predicted higher human accuracy in identifying vocalizations expressing a higher level of arousal ( $F_0$ :  $\chi^2_{(1)} = 60.043$ ,  $p < 0.001$ ; SCG:  $\chi^2_{(1)} = 32.301$ ,  $p < 0.001$ ). Within this model, no significant effect was reached by HNR and duration ratios. In line with this result, the model selection computed within the GLMM ranked the models where  $F_0$  or SCG ratios were excluded from the analyses as the weakest models (electronic supplementary material, table S3). In addition, the effect of language group was not significant

( $\chi^2_{(2)} = 2.048$ ,  $p = 0.36$ ). Pairwise comparisons between language groups were also not significant (English–German:  $z = -0.123$ ,  $p = 0.90$ ; German–Mandarin:  $z = -1.182$ ,  $p = 0.57$ ; English–Mandarin:  $z = -1.303$ ,  $p = 0.57$ ).

As shown in electronic supplementary material, table S4, the GLMMs computed within each species revealed significant effects for the following acoustic variables on identification of vocalizations expressing higher levels of arousal:  $F_0$  ratio for hourglass treefrog, American alligator, common raven, giant panda and domestic pig; SCG ratio for common raven, African bush elephant, giant panda, domestic pig and Barbary macaque; HNR ratio for black-capped chickadee, common raven, African bush elephant, giant panda, domestic pig and Barbary macaque; ratios of duration for African bush elephant, giant panda, domestic pig and Barbary macaque. The effect of language group did not reach significance in any of the species. Model selection computed within each of these GLMMs was consistent with these results (see electronic supplementary material, table S4). None of the acoustic parameters included in our model reached significant effects for high-arousal identification in human vocalizations.

### 4. Discussion

We show that humans are able to reliably identify higher levels of arousal in vocalizations of nine species spanning all classes of air-breathing tetrapods (figure 1). This finding held true for English, German and Mandarin native speakers, suggesting that this ability is biologically rooted in humans. In addition, although different acoustic parameters affect humans' arousal perception within each species, higher  $F_0$  and SCG ratios best predict humans' ability to identify higher levels of arousal in vocalizations produced by all classes of terrestrial tetrapods. In addition, our data suggest that duration and HNR are not among the best predictors of human accuracy in identifying

arousal across a wide range of tetrapods. We cannot exclude that, besides the acoustic parameters included in our statistical model, amplitude might also play a role in the human ability to discriminate levels of arousal.

Ever since Darwin argued for a shared set of mechanisms grounding vocal emotional expression across terrestrial vertebrates [12,39], there have been attempts to pinpoint the phylogenetic continuity of emotional communication across species, in terms of both the production [13–24,49–53] and the perception [25–38] of emotional vocalizations. However, no study had investigated the ability of humans to recognize emotional information in vocalizations of non-mammalian species. To our knowledge, our study is the first to directly address this issue, providing evidence on the acoustic parameters grounding the human ability to identify higher levels of arousal expressed in the vocalizations across all classes of terrestrial tetrapods. Indeed, the species included here not only exhibit greater phylogenetic diversity than has been previously assessed, but also considerable diversity in size, ecology and social structure (figure 1). Hence, our results are consistent with the hypothesis that fundamental mechanisms underlying emotion perception in vocalizations, a biological phenomenon key to survival, may have emerged in the early stages of their evolution and have been preserved across a broad range of animal species [28,37,59,65,66]. However, in order to provide stronger empirical support for this hypothesis, more species need to be tested on their ability to infer the arousal state of signallers from a similarly wide range of hetero-specific vocalizations. Findings on the evolutionary roots of arousal perception in animal vocalizations will complement further evidence on the mechanisms grounding the production of vocalizations with arousal-related content, supporting Darwin's hypothesis on homologous mechanisms of emotional expressions across terrestrial tetrapods [12,39].

Our findings extend research suggesting that perception mechanisms of frequency-related information, which is critical in human audition, originated early in primate evolution [65]. Crucially, our work corroborates and extends Morton's [13] observation that the use of frequency-related parameters in vocalizations serves emotional expression in mammals and birds, an ability that might have triggered appropriate behaviours in response to surrounding threats. Moreover, our data confirm outcomes from studies on duration and HNR as vocal correlates of arousal, which provide contrasting evidence in different species [14,51,67]. Here too we found that in certain species increases in these parameters predict identification of high arousal, while in other species the reverse pattern applies. This suggests that these parameters are not reliable indicators of arousal at an interspecific level. Our within-species findings on the effect of higher  $F_0$  ratios on humans' accuracy in identifying high-arousal vocalizations also extends earlier findings showing, for example, that increases in  $F_0$  predict humans' ability to identify high arousal in conspecifics [25,26], dogs [28], cats [30] and domestic pigs [27]. In the case of the hourglass treefrog and black-capped chickadee, it might be that one of the acoustic parameters that best predicts arousal identification is the repetition rate of vocalizations' units [15,19,20]. Unfortunately, we could not include this parameter in our model, since it was not measurable across all the species included in our stimuli set.

Sauter *et al.* [25], who also included SCG in their analyses, found that, in addition to increase in  $F_0$ -related measures, increase in SCG predicted humans' arousal rating of human

nonverbal emotional vocalizations such as screams and laughter. Somewhat surprisingly, and in contrast, we found that none of the acoustic correlates of arousal included in our analyses affected arousal identification in human vocalizations. One explanation might be that verbal emotional sentences, even when spoken in an unfamiliar language, are processed differently than nonverbal emotional vocalizations (e.g. screams), which are less constrained by precise articulatory movements [68,69]. In fact, nonverbal stimuli are typically employed for emotional identification in human stimuli [25,70–72]. Additional studies should compare emotional processing of both verbal and nonverbal human sounds (speech and vocal or instrumental music), possibly examining effects of multiple emotional dimensions, namely arousal, valence, approach–withdrawal or persistence [40]. Furthermore, it would be interesting to disentangle language effects from cultural effects in human perception of vocal arousal across species. To this aim, future studies should extend our work to language groups with different cultural backgrounds, or to one group with the same cultural background, but speaking two native languages.

Notably, the vocalizations of American alligator, domestic pig, giant panda and African bush elephant used in our study were produced by infants. Lingle & Riede [31] suggested that infant distress vocalizations, which evolved to elicit a response by caregivers, have a similar acoustic structure across species within the contexts of isolation from the mother and human capture. It is possible that differences in the acoustic structure of arousal vocalizations produced by infants compared with adults, as well as differences in arousal assessment in caregivers compared with non-caregivers, affect arousal perception. Further experimental investigations are needed to estimate these effects.

In our study, the classification of the recordings as high- or low-arousal vocalizations, for most species, was based on observational indicators, which reflect the underlying level of threat, competition or disturbance in the context of vocalization production (electronic supplementary material, table S1). Future research would benefit from combining recordings of behavioural observations with other types of data, such as brain activity [73,74] and physiological measures (heart rate, temperature, adrenaline or stress hormone levels) of each signaller during the production of emotional vocalizations (as in [18,75–77]). Moreover, one crucial limitation of our study is that the behavioural contexts in which our vocal stimuli were produced varied considerably across species, including laboratory settings, as in the case of domestic pig vocalizations included in our stimuli set. Hence, future studies should aim at including vocalizations recorded in qualitatively and functionally comparable and biologically relevant contexts across all species. This objectively quantified classification of vocalizations across animal species, which may assess different degrees of both arousal and valence in the vocalizations, could be used to disentangle the relative effect of arousal and valence in perception of emotional content in animal vocalizations.

Finally, research on the acoustic parameters involved in the production and perception of arousal in emotional vocalizations across terrestrial tetrapods is relevant to understanding the evolution of human language. Indeed, responding adaptively to the emotional content of vocal expressions, which appears to be dominant over verbal content, is likely to be evolutionarily older than speech articulation, and might have paved the path for its emergence [78,79]. Critically, arousal-related acoustic universals also appear to be shared by

music [26]. Comparison between animal vocalizations, speech and music are thus likely to further our understanding of the shared evolutionary roots between music and emotional prosody in verbal language.

In conclusion, our findings provided empirical evidence for the universality of acoustic correlates of arousal among tetrapods, suggesting that important aspects of vocal expression and perception are deeply rooted in our terrestrial vertebrate ancestors. This research framework has direct implications for our understanding of emotion processing across a broad variety of land-living animal species.

**Ethics.** Ethical approval was granted by the local ethics committee of the Faculty of Psychology, Ruhr-University Bochum (Germany) and the University of Alberta (Canada) Research Ethics Board. Written informed consent was obtained by each participant prior to the study. All participants were treated in accordance with the declaration of Helsinki.

**Data accessibility.** Data are available from the Dryad Digital Repository (<http://dx.doi.org/10.5061/dryad.7k23g>) [80].

**Authors' contributions.** P.F. developed the study concept. P.F., J.V.C., J.H., D.L.B., S.A.R., A.P., M.H., S.O., B.d.B., C.B.S., A.N. and O.G. contributed to the study design. P.F., J.V.C., J.H. and D.L.B. performed the acoustic measurements of the stimuli. P.F. performed data analysis and interpretation. P.F. drafted the manuscript, and all the other authors provided critical revisions for important intellectual content. S.A.R. drew figure 1. All authors approved the final version of the manuscript for submission.

**Competing interests.** We have no competing interests.

**Funding.** P.F. was supported by the following research grants: visiting fellowship awarded by the Center of Mind, Brain and Cognitive Evolution (Ruhr-Universität Bochum, Germany), the European Research Council Starting Grant 'ABACUS' [293435] awarded to B.d.B.;

ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and ANR-11-IDEX-0001-02 (A\*MIDEX); a visiting fellowship awarded by the Max Planck Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. J.V.C. was supported by an Alexander Graham Bell Canada Graduate Scholarship-Master's (CGSM) and Walter H. Johns Graduate Fellowship. J.H. was supported by a Queen Elizabeth II Master's scholarship. D.L.B. was supported by a Lise Meitner Postdoctoral Fellowship from the Austrian Science Fund (FWF). S.A.R. was supported by a Marietta Blau Grant, the European Research Council (ERC) Advanced Grant 'SOMACCA' [230604] awarded to W.T. Fitch, and the Austrian Science Fund (FWF) [W1234-G17]. A.P. was funded by the Austrian Science Fund (FWF) [W1234-G17]. M.H. is currently funded by a Lise Meitner Postdoctoral Fellowship [M 1732-B19] from the Austrian Science Fund (FWF) and was also funded by a Banting Postdoctoral Fellowship awarded by the Natural Sciences and Engineering Research Council of Canada during this project. C.B.S. was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and Discovery Accelerator Supplement, an Alberta Ingenuity Fund (AIF) New Faculty Grant, a Canada Foundation for Innovation (CFI) New Opportunities Fund (NOF) and Infrastructure Operating Fund (IOF) grants along with start-up funding and CFI partner funding from the University of Alberta (UofA). A.N. was supported by the Fritz Thyssen Foundation Grant [Az. 10.14.2.015]. O.G. was supported by the German Research Foundation (DFG) [Gu 227/16-1].

**Acknowledgements.** We thank the following collaborators: Sima Parsey and Larissa Heege for their help in testing participants and for translating the instructions into German; Natalie Kuo-Hsuan Yang and Christian Poon for their help in testing participants in both Mandarin and English; Nigel Mantou Lou for translating the instructions into Mandarin; Kerem Eryilmaz for help building the interface; and Andrea Ravignani and Gary Lupyan for valuable comments on the statistical analyses. We are especially grateful to all authors of the studies referenced in section 'Recordings and arousal classification', who have kindly provided the stimuli adopted in our experiment.

## References

- Nesse RM. 1990 Evolutionary explanations of emotions. *Hum. Nat.* **1**, 261–289. (doi:10.1007/BF02733986)
- Mendl M, Burman OHP, Paul ES. 2010 An integrative and functional framework for the study of animal emotion and mood. *Proc. R. Soc. B* **277**, 2895–2904. (doi:10.1098/rspb.2010.0303)
- Russell JA. 1980 A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**, 1161–1178. (doi:10.1037/h0077714)
- Taylor A, Reby D. 2010 The contribution of source-filter theory to mammal vocal communication research. *J. Zool.* **280**, 221–236. (doi:10.1111/j.1469-7998.2009.00661.x)
- Titze IR. 1994 *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.
- Charlton BD, Reby D. 2016 The evolution of acoustic size exaggeration in terrestrial mammals. *Nat. Commun.* **7**, 12739. (doi:10.1038/ncomms12739)
- Reber SA, Nishimura T, Janisch J, Robertson M, Fitch WT. 2015 A Chinese alligator in heliox: formant frequencies in a crocodylian. *J. Exp. Biol.* **218**, 2442–2447. (doi:10.1242/jeb.119552)
- Bowling DL, Garcia M, Dunn JC, Ruprecht R, Stewart A, Frommolt K-H, Fitch WT. 2017 Body size and vocalization in primates and carnivores. *Sci. Rep.* **24**, 41070. (doi:10.1038/srep41070)
- Gerhardt HC, Huber F. 2002 *Acoustic communication in insects and anurans: common problems and diverse solutions*. Chicago, IL: University of Chicago Press.
- Stanley Rand A, Dudley R. 1993 Frogs in helium: the anuran vocal sac is not a cavity resonator frogs in helium: the anuran vocal sac is not a cavity resonator. *Physiol. Zool.* **66**, 793–806. (doi:10.1086/physzool.66.5.30163824)
- Fitch WT. 1999 Acoustic exaggeration of size in birds via tracheal elongation: comparative and theoretical analyses. *J. Zool.* **248**, 31–48. (doi:10.1111/j.1469-7998.1999.tb01020.x)
- Darwin C. 1871 *The descent of man and selection in relation to sex*. London, UK: John Murray.
- Morton ES. 1977 On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* **111**, 855–869. (doi:10.1086/283219)
- Briefer E. 2012 Vocal expression of emotions in mammals: mechanisms of production and evidence. *J. Zool.* **288**, 1–20. (doi:10.1111/j.1469-7998.2012.00920.x)
- Templeton CN, Greene E, Davis K. 2005 Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science* **308**, 1934–1937. (doi:10.1126/science.1108841)
- Szipl G. 2017 Social relationships and acoustic communication in common ravens. University of Vienna.
- Pierce JR, Ralin DB. 1972 Vocalization and behavior of the males of three species in the *Hyla versicolor* complex. *Herpetologica* **28**, 329–337.
- Reichert MS. 2013 Sources of variability in advertisement and aggressive calling in competitive interactions in the grey treefrog, *Hyla versicolor*. *Bioacoustics* **22**, 195–214. (doi:10.1080/09524622.2013.777942)
- Reichert MS. 2011 Aggressive calls improve leading callers' attractiveness in the treefrog *Dendropsophus ebraccatus*. *Behav. Ecol.* **22**, 951–959. (doi:10.1093/beheco/arr074)
- Reichert MS. 2013 Patterns of variability are consistent across signal types in the treefrog *Dendropsophus ebraccatus*. *Biol. J. Linn. Soc.* **109**, 131–145. (doi:10.1111/bij.12028)
- Grafe TU. 1995 Graded aggressive calls in the African painted reed frog *Hyperolius marmoratus* (Hyperoliidae). *Ethology* **101**, 67–81. (doi:10.1111/j.1439-0310.1995.tb00346.x)
- Pettitt BA, Bourne GR, Bee MA. 2012 Quantitative acoustic analysis of the vocal repertoire of the golden rocket frog (*Anomalaglossus beebei*). *J. Acoust. Soc. Am.* **131**, 4811–4820. (doi:10.1121/1.4714769)

23. Wells K, Schwartz J. 1984 Vocal communication in a neotropical treefrog, *Hyla ebraccata*: aggressive calls. *Behaviour* **91**, 128–145. (doi:10.1163/156853984X00254)
24. Britton ARC. 2001 Review and classification of call types of juvenile crocodylians and factors affecting distress calls. In *Crocodylian biology and evolution* (eds GC Grigg, F Seebacher, CE Franklin), pp. 364–377. Chipping Norton, Australia: Surrey Beatty & Sons.
25. Sauter DDA, Eisner F, Calder AAJ, Scott SSK. 2010 Perceptual cues in nonverbal vocal expressions of emotion. *Q. J. Exp. Psychol.* **63**, 2251–2272. (doi:10.1080/17470211003721642)
26. Juslin PN, Laukka P. 2003 Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* **129**, 770–814. (doi:10.1037/0033-2909.129.5.770)
27. Maruščáková IL, Linhart P, Ratcliffe VF, 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. *J. Comp. Psychol.* **129**, 121–131. (doi:10.1037/a0038870)
28. Faragó T, Andics A, Devescseri V, Kis A, Gácsi M, Miklósi D. 2014 Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biol. Lett.* **10**, 20130926. (doi:10.1098/rsbl.2013.0926)
29. Pongrácz P, Molnár C, Miklósi Á. 2006 Acoustic parameters of dog barks carry emotional information for humans. *Appl. Anim. Behav. Sci.* **100**, 228–240. (doi:10.1016/j.applanim.2005.12.004)
30. McComb K, Taylor AM, Wilson C, Charlton BD. 2009 The cry embedded within the purr. *Curr. Biol.* **19**, R507–R508. (doi:10.1016/j.cub.2009.05.033)
31. Lingle S, Riede T. 2014 Deer mothers are sensitive to infant distress vocalizations of diverse mammalian species. *Am. Nat.* **184**, 510–522. (doi:10.1086/677677)
32. McConnell PB. 1990 Acoustic structure and receiver response in domestic dogs, *Canis familiaris*. *Anim. Behav.* **39**, 897–904. (doi:10.1016/S0003-3472(05)80954-6)
33. Scheumann M, Hasting AS, Kotz SA, Zimmermann E. 2014 The voice of emotion across species: how do human listeners recognize animals' affective states? *PLoS ONE* **9**, e91192. (doi:10.1371/journal.pone.0091192)
34. Tallet C, Spinka M, Maruščáková I, Simecek P. 2010 Human perception of vocalizations of domestic piglets and modulation by experience with domestic pigs (*Sus scrofa*). *J. Comp. Psychol.* **124**, 81–91. (doi:10.1037/a0017354)
35. Pongrácz P, Molnár C, Miklósi Á, Csányi V. 2005 Human listeners are able to classify dog (*Canis familiaris*) barks recorded in different situations. *J. Comp. Psychol.* **119**, 136–144. (doi:10.1037/0735-7036.119.2.136)
36. Nicastro N, Owren MJ. 2003 Classification of domestic cat (*Felis catus*) vocalizations by naive and experienced human listeners. *J. Comp. Psychol.* **117**, 44–52. (doi:10.1037/0735-7036.117.1.44)
37. Belin P, Fecteau S, Charest I, Nicastro N, Hauser MD, Armony JL. 2008 Human cerebral response to animal affective vocalizations. *Proc. R. Soc. B* **275**, 473–481. (doi:10.1098/rspb.2007.1460)
38. Albuquerque N, Guo K, Wilkinson A, Savalli C, Otta E, Mills D. 2016 Dogs recognize dog and human emotions. *Biol. Lett.* **12**, 20150883. (doi:10.1098/rsbl.2015.0883)
39. Darwin C. 1872 *The expression of the emotions in man and animals*. London, UK: John Murray.
40. Anderson DJ, Adolphs R. 2014 A framework for studying emotions across species. *Cell* **157**, 187–200. (doi:10.1016/j.cell.2014.03.003)
41. Kitchen D, Bergman T, Cheney D, Nicholson J, Seyfarth Robert M. 2010 Comparing responses of four ungulate species to playbacks of baboon alarm calls. *Anim. Cogn.* **13**, 861–870. (doi:10.1007/s10071-010-0334-9)
42. Fallow PM, Magrath RD. 2009 Eavesdropping on other species: mutual interspecific understanding of urgency information in avian alarm calls. *Anim. Behav.* **79**, 411–417. (doi:10.1016/j.anbehav.2009.11.018)
43. De Boer B, Wich SA, Hardus ME, Lameira AR. 2015 Acoustic models of orangutan hand-assisted alarm calls. *J. Exp. Biol.* **218**, 907–914. (doi:10.1242/jeb.110577)
44. Magrath RD, Pitcher BJ, Gardner JL. 2009 Recognition of other species' aerial alarm calls: speaking the same language or learning another? *Proc. R. Soc. B* **276**, 769–774. (doi:10.1098/rspb.2008.1368)
45. Marc DH. 1988 How infant vervet monkeys learn to recognize starling alarm calls: the role of experience. *Behaviour* **105**, 187–201. (doi:10.1163/156853988X00016)
46. Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. 2003 Placental mammal diversification and the Cretaceous–Tertiary boundary. *Proc. Natl Acad. Sci. USA* **100**, 1056–1061. (doi:10.1073/pnas.0334222100)
47. Hugall A, Foster R, Lee M. 2007 Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Syst. Biol.* **56**, 543–563. (doi:10.1080/10635150701477825)
48. Stevens NJ *et al.* 2013 Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. *Nature* **497**, 611–614. (doi:10.1038/nature12161)
49. Stoeger AS, Charlton BD, Kratochvil H, Fitch WT. 2011 Vocal cues indicate level of arousal in infant African elephant roars. *J. Acoust. Soc. Am.* **130**, 1700–1710. (doi:10.1121/1.3605538)
50. Stoeger AS, Baotic A, Li D, Charlton BD. 2012 Acoustic features indicate arousal in infant giant panda vocalisations. *Ethology* **118**, 896–905. (doi:10.1111/j.1439-0310.2012.02080.x)
51. Linhart P, Ratcliffe VF, Reby D, Špinka M. 2015 Expression of emotional arousal in two different piglet call types. *PLoS ONE* **10**, e0135414. (doi:10.1371/journal.pone.0135414)
52. Fischer J, Hammerschmidt K, Todt D. 1995 Factors affecting acoustic variation in Barbary-macaque (*Macaca sylvanus*) disturbance calls. *Ethology* **101**, 51–66. (doi:10.1111/j.1439-0310.1995.tb00345.x)
53. Bowling DL, Gingras B, Han S, Sundararajan J, Opitz ECL. 2013 Tone of voice in emotional expression: relevance for the affective character of musical mode. *J. Interdiscip. Music Stud.* **7**, 29–44.
54. Massen JJM, Pašukonis A, Schmidt J, Bugnyar T. 2014 Ravens notice dominance reversals among conspecifics within and outside their social group. *Nat. Commun.* **5**, 3679. (doi:10.1038/ncomms4679)
55. Avey MT, Hoeschele M, Moscicki MK, Bloomfield LL, Sturdy CB. 2011 Neural correlates of threat perception: neural equivalence of conspecific and heterospecific mobbing calls is learned. *PLoS ONE* **6**, e23844. (doi:10.1371/journal.pone.0023844)
56. Buss I, Rasmussen L, Smuts G. 1976 The role of stress and individual recognition in the function of the African elephant's temporal gland. *Mammalia* **40**, 437–451. (doi:10.1515/mamm.1976.40.3.437)
57. Buss IO, Estes JA. 1971 The functional significance of movements and positions of the Pinnae of the African. *J. Mammal.* **52**, 21–2743. (doi:10.2307/1378428)
58. Poole JH. 1999 Signals and assessment in African elephants: evidence from playback experiments. *Anim. Behav.* **58**, 185–193. (doi:10.1006/anbe.1999.1117)
59. Zimmermann E, Leliveld L, Schehka S. 2013 Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices. In *Evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man* (eds E Altenmüller, S Schmidt, E Zimmermann), pp. 116–132. Oxford, UK: Oxford University Press.
60. Boersma P. 2002 Praat, a system for doing phonetics by computer. *Glott Int.* **5**, 341–345.
61. Team R. 2015 *RStudio: integrated development for R*. Boston, MA: RStudio. See <http://www.rstudio.com>.
62. Holm S. 1979 A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70.
63. Symonds MRE, Moussalli A. 2011 A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav. Ecol. Sociobiol.* **65**, 13–21. (doi:10.1007/s00265-010-1037-6)
64. Burnham KP, Anderson DR, Huyvaert KP. 2011 AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* **65**, 23–35. (doi:10.1007/s00265-010-1029-6)
65. Song X, Osmanski MS, Guo Y, Wang X. 2016. 19 Complex pitch perception mechanisms are shared by humans and a New World monkey. *Proc. Natl Acad. Sci. USA* **113**, 781–786. (doi:10.1073/pnas.1516120113)
66. Andics A, Gácsi M, Faragó T, Kis A, Miklósi Á. 2014 Voice-sensitive regions in the dog and human brain



- are revealed by comparative fMRI. *Curr. Biol.* **24**, 574–578. (doi:10.1016/j.cub.2014.01.058)
67. Blumstein DT, Chi YY. 2012 Scared and less noisy: glucocorticoids are associated with alarm call entropy. *Biol. Lett.* **8**, 189–192. (doi:10.1098/rsbl.2011.0832)
68. Pell MD, Rothermich K, Liu P, Paulmann S, Sethi S, Rigoulot S. 2015 Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biol. Psychol.* **111**, 14–25. (doi:10.1016/j.biopsycho.2015.08.008)
69. Rauschecker JP, Scott SK. 2009 Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724. (doi:10.1038/nn.2331)
70. Sauter DA, Eisner F, Ekman P, Scott SK. 2010 Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc. Natl Acad. Sci. USA* **107**, 2408–2412. (doi:10.1073/pnas.0908239106)
71. 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. *Behav. Res. Methods* **40**, 531–539. (doi:10.3758/BRM.40.2.531)
72. Simon-Thomas ER, Keltner DJ, Sauter D, Sinicropi-Yao L, Abramson A. 2009 The voice conveys specific emotions: evidence from vocal burst displays. *Emotion* **9**, 838–846. (doi:10.1037/a0017810)
73. Ocklenburg S, Ströckens F, Güntürkün O. 2013 Lateralisation of conspecific vocalisation in non-human vertebrates. *Laterality* **18**, 1–31. (doi:10.1080/1357650X.2011.626561)
74. Panksepp J. 2011 The basic emotional circuits of mammalian brains: do animals have affective lives? *Neurosci. Biobehav. Rev.* **35**, 1791–1804. (doi:10.1016/j.neubiorev.2011.08.003)
75. Briefer EF, Tettamanti F, Mcelligott AG. 2015 Animal studies repository emotions in goats: mapping physiological, behavioural and vocal profiles. *Anim. Behav.* **99**, 131–143. (doi:10.1016/j.anbehav.2014.11.002)
76. Briefer EF, Maigrot A-L, Mandel R, Freymond SB, Bachmann I, Hillmann E. 2015 Segregation of information about emotional arousal and valence in horse whinnies. *Sci. Rep.* **4**, 9989. (doi:10.1038/srep09989)
77. Stocker M, Munteanu A, Stöwe M, Schwab C, Palme R, Bugnyar T. 2016 Loner or socializer? Ravens' adrenocortical response to individual separation depends on social integration. *Horm. Behav.* **78**, 194–199. (doi:10.1016/j.yhbeh.2015.11.009)
78. Filippi P, Ocklenburg S, Bowling DL, Heege L, Güntürkün O, Newen A, De Boer B. 2017 More than words (and faces): evidence for a Stroop effect of prosody in emotion word processing. *Cogn. Emot.* **31**, 879–891. (doi:10.1080/02699931.2016.1177489)
79. Filippi P. 2016 Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Front. Psychol.* **7**, 1393. (doi:10.3389/fpsyg.2016.01393)
80. Filippi P *et al.* 2017 Data from: Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. Dryad Digital Repository. (<http://dx.doi.org/10.5061/7k23g>)