

Article

Humans identify negative (but not positive) arousal in silver fox vocalizations: implications for the adaptive value of interspecific eavesdropping

Piera FILIPPI^{a,b,c,*}, Svetlana S. GOGOLEVA^d, Elena V. VOLODINA^e, Ilya A. VOLODIN^{d,e}, and Bart de BOER^a

^aArtificial Intelligence Laboratory, Department of Computer Science, Faculty of Science, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium, ^bBrain and Language Research Institute, Aix-Marseille University, Avenue Pasteur 5, 13604 Aix-en-Provence, France, ^cMax Planck Institute for Psycholinguistics, Department of Language and Cognition, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands, ^dDepartment of Vertebrate Zoology, Faculty of Biology, Lomonosov Moscow State University, Vorobiev Gory 1/12, 119991 Moscow, Russia and ^eScientific Research Department, Moscow Zoo, B. Gruzinskaya 1, 123242 Moscow, Russia

*Address correspondence to Piera Filippi. E-mail: pie.filippi@gmail.com

Received on 12 December 2016; accepted on 12 May 2017

Abstract

The ability to identify emotional arousal in heterospecific vocalizations may facilitate behaviors that increase survival opportunities. Crucially, this ability may orient inter-species interactions, particularly between humans and other species. Research shows that humans identify emotional arousal in vocalizations across multiple species, such as cats, dogs, and piglets. However, no previous study has addressed humans' ability to identify emotional arousal in silver foxes. Here, we adopted low- and high-arousal calls emitted by three strains of silver fox—Tame, Aggressive, and Unselected—in response to human approach. Tame and Aggressive foxes are genetically selected for friendly and attacking behaviors toward humans, respectively. Unselected foxes show aggressive and fearful behaviors toward humans. These three strains show similar levels of emotional arousal, but different levels of emotional valence in relation to humans. This emotional information is reflected in the acoustic features of the calls. Our data suggest that humans can identify high-arousal calls of Aggressive and Unselected foxes, but not of Tame foxes. Further analyses revealed that, although within each strain different acoustic parameters affect human accuracy in identifying high-arousal calls, spectral center of gravity, harmonic-to-noise ratio, and F0 best predict humans' ability to discriminate high-arousal calls across all strains. Furthermore, we identified in spectral center of gravity and F0 the best predictors for humans' absolute ratings of arousal in each call. Implications for research on the adaptive value of inter-specific eavesdropping are discussed.

Key words: eavesdropping, emotional arousal, emotional valence, inter-species communication, silver foxes, vocal communication.

Emotions are linked to internal brain and physiological states that may be caused by external stimuli (Mendl et al. 2010; Anderson and Adolphs 2014). At least two dimensions define emotional states:

arousal, namely a state of the brain or the body reflecting responsiveness to sensory stimulation ranging from sleep (low arousal) to frenetic excitement (high arousal) (Russell 1980), and valence, the

intrinsic attractiveness (positive valence) or averseness (negative valence) of an event (Russell 1980; Frijda 1987; Mendl et al. 2010). Some emotions can have a similar arousal level and differ in valence (e.g., happiness vs. anger) or *vice versa*, have similar valence and differ in arousal (e.g., annoyance vs. rage). Valence and arousal are described as “building blocks” of emotions (Anderson and Adolphs 2014).

Changes in emotional states may be reflected in vocalizations and express the emotional state of the signaler (Darwin 1872; Gogoleva et al. 2010a, 2010c; Briefer 2012; Volodin et al. 2017). Indeed, a key aspect of the acoustic features of the voice is that they reflect changes in the configuration and action of muscles involved in vocal production (e.g., the diaphragm and laryngeal muscles). Therefore, the way air flows through the system and ultimately the quality of the sounds produced is critically affected (Davis et al. 1996). In an extensive review on emotional vocal communication in mammals, Briefer (2012) suggested that increases in frequency-related parameters of the voice (e.g., fundamental frequency, frequency range, and spectral shape) and in amplitude contour, increase in call rate and decrease in inter-call interval predict high arousal in a number of mammalian species. In addition, Morton (1977) observed that both mammals and birds use of harsh, low-frequency vocalizations in hostile agonistic contexts, and of more tone-like, high-frequency sounds in fearful or appeasing contexts. As to research on emotional valence expression, studies comparing acoustic features underlying vocalizations produced in positive and negative situations (controlling for arousal) are sparse since animals are more likely to call in negative contexts. By reviewing findings in a number of studies conducted in a number of mammalian species, Briefer (2012) suggested that the only acoustic parameter that consistently changes as a function of valence is duration. Specifically, animal vocalizations reflecting positive valence are shorter than vocalizations expressing negative valence.

Critically, from the listeners' side, correct identification of the emotional state of a signaler through accurate perception of acoustic modulation of the voice may drive adaptive survival (Nesse 1990; Anderson and Adolphs 2014) in the context of territory disputes, predators avoidance (Nesse 1990; Owings and Morton 1998; Desrochers et al. 2002; Cross and Rogers 2005; Kitchen et al. 2010), or social interactions (Gogoleva et al. 2010a, 2010c; Altenmüller et al. 2013; Bryant 2013). Furthermore, the ability to recognize emotional arousal in vocal expressions may be decisive for the survival of newborns, who require caregivers to perceive and react to their needs (Marmoset monkey *Callithrix jacchus*: Tchernichovski and Oller 2016; Zhang and Ghazanfar 2016; human *Homo Sapiens*: Fernald 1992).

Survival may be facilitated by the ability to identify emotions not only in vocalizations emitted by conspecifics, but also by members of other species (Nesse 1990). This ability may provide information that is key to responding appropriately. Indeed, it has been shown that nonhuman animals' “eavesdropping” on another species alarm calls increases opportunities for survival (Owings and Morton 1998; Kitchen et al. 2010; Fallow et al. 2011; de Boer et al. 2015; Magrath et al. 2015). Advantageous responses to inter-specific calls may occur as a result of acoustic similarity in the signals (Aubin 1991; De Kort and Carel ten 2001; Johnson et al. 2003). In other cases, listeners respond appropriately to calls that are acoustically different from their own (Templeton et al. 2005; Lea et al. 2008; Fallow and Magrath 2010), suggesting that responses are biologically rooted, or, in the case of species living in close territories, learned. For example, juvenile vervet monkeys' *Cercopithecus aethiops pygerythrus* appropriate responses to playback of alarm calls given by superb

starlings *Spreo superbus* vary depending on the rates of exposure to these alarm calls (Hauser 1988). Generally, the ability to respond appropriately to heterospecific calls, which may presuppose the ability to recognize their level of emotional arousal and valence (Mendl et al. 2010), is the result of a signaling system that affords inter-specific beneficial outcomes in dangerous contexts.

In parallel to research on the acoustic correlates of emotional dimensions in vocal production, multiple studies examined perception of valence and arousal in mammals' vocalizations. Research on arousal perception suggests that humans rate human, piglet, and dog vocalizations with higher fundamental frequency (F0) as expressing higher emotional arousal (Laukka et al. 2005; Faragó et al. 2014; Maruščáková et al. 2015). Moreover, McComb et al. (2009) suggest that humans perceive as more urgent and less pleasant cat purrs recorded while cats were actively seeking food than purrs recorded in non-solicitation contexts. The authors identified in voiced frequency peaks the acoustic predictors of humans' accuracy in classifying cat vocalizations. Sauter et al. (2010) found that humans perceive human nonverbal vocalizations with higher F0 means, shorter duration, more amplitude onsets, lower minimum F0, and less F0 variation as expressing higher arousal. In a recent study, Filippi et al. (2016) suggest that humans are able to discriminate high versus low levels of arousal in negative-valenced vocalizations of terrestrial tetrapods spanning all classes of animals. In addition, they identified in F0 and spectral center of gravity the acoustic predictors of this ability, pointing to biologically rooted acoustic universals of arousal perception. Furthermore, studies on arousal perception across species suggests that shepherds' high-pitched, quickly pulsating whistles have an activating effect on dogs (McConnell and Baylis, 2010) and that 2 species of deers *Odocoileus hemionus* and *Odocoileus virginianus* respond to infant distress vocalizations of human and nonhuman animals (infant marmots *Marmota flaviventris*, seals *Neophoca cinerea*, and *Arctocephalus tropicalis*, domestic cats *Felis catus*, bats *Lasionycteris noctivagans*, humans *H. sapiens*, and other ungulates: eland *Taurotragus oryx*, red deer *Cervus elaphus*, fallow deer *Dama dama*, sika deer *Cers nippon*, pronghorn *Antilocapridae americana*, and bighorn sheep *Ovis canadensis*) if the F0 falls within the deer's frequency response range (Lingle and Riede 2014). Moreover, research on valence perception suggests that humans recognize the negative and positive contexts in which vocalizations of human infants *H. sapiens*, chimpanzees *Pan troglodytes* (Scheumann et al. 2014), domestic pigs *Sus scrofa domesticus* (Tallet et al. 2010; Maruščáková et al. 2015), dogs *Canis familiaris* (Pongrácz et al. 2006; Scheumann et al. 2014), and cats *F. catus* (Nicastro and Owren 2003) were recorded (but see Belin et al. 2008 for contrasting results on human perception of valence in cats *F. catus*—and monkeys *Macaca mulatta* vocalizations). Albuquerque et al. (2016) found that dogs can identify emotional valence in both conspecific and human vocalizations. Further studies addressing the acoustic predictors of valence perception suggest that humans' rate human and dog vocalizations with shorter duration, and human vocalizations with lower SCOG as more positive (Faragó et al. 2014; but see Pongrácz et al. 2005). Maruščáková et al. (2015) found that humans rate domestic piglets' vocalizations with increased F0 and duration as more negative.

Notably, much research has examined humans' perception of arousal or valence in vocalizations of multiple species, adopting a continuous rating scale, and their ability to infer whether the vocalizations were produced in a positive or negative context. However, humans' ability to recognize different arousal levels in nonverbal vocalizations from the same species systematically varying in valence

content remains largely unexplored. Furthermore, to the best of our knowledge, no previous work has investigated this issue focusing on the perception of heterospecific calls, linking the perception of voice modulation to the arousal state of the caller—as identified based on independent nonvocal indicators. Here, for the first time, we analyzed humans' ability to recognize emotional arousal in silver fox vocalizations produced in a positive or negative context. Furthermore, we examined the acoustic predictors of both this ability and of the level of perceived arousal in silver fox calls, as reported using a rating scale spanning from 1 (very subdued) to 7 (very excited). To this aim, we adopted vocalizations produced by the following three strains of silver fox: Tame, Aggressive, and Unselected.

These strains are the result of a genetic selection program in silver foxes. Belyaev (1979) hypothesized that selection of farm foxes for less fearful and less aggressive behavior would yield a strain of domesticated fox. To address this hypothesis, the authors started a program designed to recapitulate canine domestication in the silver fox at the Institute of Cytology and Genetics of the Russian Academy of Sciences, Novosibirsk, Russia (Trut 1999; Trut et al. 2004, 2009a). Thus, different behavioral phenotypes across these three strains of silver fox have been experimentally established by intense selective breeding (Kukekova et al. 2012). Specifically, Tame silver foxes have been experimentally selected in the course of more than 50 years of selection for positive behavior to people and show friendly response to humans, approaching any unfamiliar experimenter (Belyaev 1979; Trut 1999, 2001; Trut et al. 2009) and even kept as pets (Ratliff 2011). Aggressive foxes were selected for aggressive behaviour and can attack humans (Trut 1980, 2001; Kukekova et al. 2008a, 2008b). Unselected foxes were not deliberately selected for behavior and demonstrate aggressively fearful behavior to humans (Pedersen and Jeppesen 1990; Pedersen 1991, 1993, 1994; Trut 1999; Nimon and Broom 2001; Kukekova et al. 2008a, 2008b; Gogoleva et al. 2010c). In the presence of an unfamiliar human, the Unselected fox with its wild type attitudes toward people enlarges the animal–human distance and shows escape responses (Supplementary movies 1–3). Cross fostering, cross breeding, and embryo transplantation experiments have shown that behavioral differences between Tame and Aggressive foxes are genetically determined (Trut 1980, 2001; Kukekova et al. 2012). Critically, strict selection for tame behavior included a substantial decrease in the levels of adrenocorticotrophic hormone (ACTH) and basal levels of corticosteroids in blood plasma of Tame foxes (Oskina 1996; Trut 1999; Trut et al. 2004, 2009; Oskina et al. 2008). Both basal and 'post-stress' (after 10-min restraint in human hands) levels of cortisol and ACTH in Aggressive foxes did not differ from those of Unselected foxes, whereas in Tame foxes they were much lower (Oskina et al. 2008). These findings suggest that, in contrast to tame silver foxes, Unselected and Aggressive foxes experience negative emotional arousal in response to human approach. For this reason, Belyaev's silver foxes (Belyaev 1979) provide a unique model for studying the human ability to distinguish between high- and low-arousal levels in heterospecific vocalizations of both negative and positive emotional valence.

Converging empirical evidence on physiological and vocal correlates of arousal (Oskina 1996; Trut et al. 2009; Gogoleva 2010a, 2010c) indicate that, in response to humans approaching their cage, Tame foxes experience heightened emotional arousal with positive valence, while Aggressive and Unselected silver foxes experience high arousal with negative valence. Gogoleva et al. (2010a, 2010c) found that, across strains of silver fox, higher levels of emotional

arousal due to human approach are reflected in an increased calling rate, and the proportion of time spent vocalizing. This applies to both high-arousal calls with positive valence—in the case of tame foxes, who experience comfort in relation to human physical approach—and high-arousal calls with negative valence, in the case of Aggressive and Unselected silver foxes, who experience discomfort in relation to human physical approach. Thus, vocal responses toward humans differ according to fox strain (Gogoleva et al. 2008, 2009, 2010b, 2010c, 2011), as a consequence of genetic differences between these 3 strains (Trut 1980, 2001; Kukekova et al. 2012). Call types that silver foxes in captivity produce toward humans, include whine, moo, cackle, growl, bark, pant, snort, and cough (Gogoleva et al. 2008, 2010a, 2010b, 2010c). Selection by behavior did not affect the vocal repertoire of silver fox; all strains (Tame, Aggressive, and Unselected) retain all call types toward conspecifics (Gogoleva et al. 2010b). However, toward humans, Tame foxes selectively produce cackles and pants but never cough or snort, while Aggressive and Unselected foxes selectively produce coughs and snorts but never cackles or pants (Gogoleva et al. 2008, 2009, 2010c, 2013). Importantly, of the total of 8 call types that silver foxes direct to humans, only the whine often occurs in all 3 strains (Gogoleva et al. 2008, 2010a, 2010c, 2013). In addition, Newton-Fisher et al. (1993) have reported on the whine call-type in wild red fox, suggesting that it is used in both agonistic and affiliative contexts. Therefore, whines constitute the most appropriate call type for investigating human perception of arousal across the 3 fox strains.

The analysis of human perception of vocalizations of these 3 strains of silver foxes provides an ideal context for research on the adaptive value of emotion perception across species area for the following reasons: 1) calls emitted in comparable behavioral contexts across all 3 strains can be used; 2) our stimuli are all instances of one call type, the whine, which is produced in all 3 strains. This excludes variation in call types as a confound in the analyses; 3) both the arousal and valence states of silver foxes in response to human approach have been attested in terms of physiological measurements, namely hormonal responses, in previous studies (Oskina 1996; Trut 1999; Trut et al. 2004, 2009; Oskina et al. 2008). These 3 conditions enabled us to disentangle human sensitivity to high-arousal calls with negative valence from human sensitivity to high-arousal calls with positive valence.

Although research shows that humans identify emotional arousal in vocalizations across multiple species, such as cats, dogs, and piglets, humans' ability to identify emotional arousal in silver foxes has never been investigated. Here, we adopted low- and high-arousal calls emitted by 3 strains of silver fox—Tame, Aggressive, and Unselected—in response to human approach. Specifically, within this research framework, the present study aimed to address the following questions: 1) Are humans able to identify high-arousal calls in silver fox vocalizations? 2) If so, does this ability vary as a function of strain? 3) What are the acoustic features that predict both the ability to recognize higher levels of arousal in silver fox vocalizations and the level of perceived arousal in silver fox vocalizations?

In line with previous research on arousal perception across animal species (Pongrácz et al. 2006; Lingle et al. 2012; Teichroeb et al. 2013; Lingle and Riede 2014; Scheumann et al. 2014; Filippi 2016; Filippi et al. 2016), we hypothesized that humans would be able to discriminate between low and high levels of arousal expressed in all 3 strains of silver fox using frequency-related parameters, which are identified as acoustic correlates of arousal in mammals and birds (Morton 1977; Briefer 2012). This investigation

might provide key insights into the adaptive effects of the ability to identify different levels of arousal varying in valence in a nonhuman species.

Materials and Methods

Acoustic recordings and emotion classification

The stimuli adopted in our study are selected from acoustic recordings collected at the experimental farm of the Institute of Cytology and Genetics, Novosibirsk, Russia under a framework used in previous studies (Gogoleva et al. 2010a, 2010c). Three study groups included Tame (selected for tameness toward humans, 45–47 generations since the start of selection), Aggressive (selected for aggressiveness toward humans; 34–36 generations since the start of selection), and Unselected (unselected for any behavioral trait) adult female silver foxes. The foxes were kept and tested in individual outdoor cages (for keeping details see Gogoleva et al. 2010a, 2010c). Human-approach tests were made when foxes were in their home cages, out of breeding or pup-raising seasons.

The same researcher (S. S. G.), unfamiliar to the foxes, performed all human-approach tests (1 per fox), while acoustic recordings were collected. Each test lasted 10 min and included 5 successive steps, each lasting 2 min. A test started at the moment of the researcher's approach to a focal fox cage at a distance of 50 cm. At Step 1, the researcher was motionless; at Step 2, the researcher performed smooth body and hand movements left to right, maintaining a distance of 50 cm; at Step 3, the researcher shortened the human–fox distance with 1-step forward, and performed body and hand movements forward and back, touching the cage door with her fingers; at Step 4, the researcher enhanced the human–fox distance with 1-step back; and repeated Step 2. Finally, at Step 5, the research was motionless as in Step 1. Thus, the human impact on an animal increased between the Steps 1 and 3, and decreased between Steps 3 and 5. The shifts in the levels of emotional arousal and valence were estimated by nonvocal indicators, that is, by increased degree of striving to approach the front door (i.e., by striving to contact with a human either friendly or aggressively, see Supplementary movies 1–3).

Hence, the same recording procedure applied to different fox strains made it possible to obtain high- and low-arousal stimuli of negative valence for the Aggressive and Unselected foxes and high- and low-arousal stimuli of positive valence for the Tame foxes (Gogoleva et al. 2010a, 2010c). The unfamiliar human represented an external stimulus for the foxes. The level of arousal in the focal fox depended on the distance between the human and the focal fox. Specifically, the arousal level in the focal fox increased by decreasing the distance between the human experimenter and the focal fox. In silver foxes, the changes of emotional arousal states in response to human approach, have been established in previous studies (Pedersen and Jeppesen 1990; Pedersen 1993; Bakken 1998; Bakken et al. 1999; Trut 1999; Kukekova et al. 2008b). Critically, research show that, in response to humans, Unselected silver foxes show fearful behaviors, Aggressive silver foxes show aggressive behaviors, and Tame silver foxes show friendly behaviors (Trut 1999; Trut et al. 2009). The valence content of vocalizations produced by the focal foxes (negative for Aggressive and Unselected silver foxes, and positive for Tame silver foxes) was inferred on the basis of these studies.

For audio recordings (distance between vocalizing fox and the microphone: 0.25–1 m), we used a Marantz PMD-222 (D&M Professional, Kanagawa, Japan) cassette recorder with an AKG-

C1000S (AKG-Acoustics GmbH, Vienna, Austria) cardioid electret condenser microphone, and Type II chrome audiocassettes EMTEC-CS II (EMTEC Consumer Media, Ludwigshafen, Germany). The system had a frequency response of 0.04–14 kHz at a tape speed of 4.75 cm/s.

For the purposes of the present study, recordings were digitized (with each test step taken as a separate file) at a 22.05 kHz sampling rate with 16 bit precision and then high-pass filtered at 0.1 kHz with Avisoft-SASLab Pro (Avisoft Bioacoustics, Berlin, Germany, Specht 2002). SSG classified each call visually to 1 of 8 types (whine, moo, cackle, growl, bark, pant, snort, and cough) by spectrogram (Hamming window, FFT-length 1024 points, frame 50%, and overlap 87.5%), blindly to the fox strain and to the number of the test step (based on Gogoleva et al. 2008),

Only whines were adopted as playback stimuli, as only the whine is shared by Tame, Aggressive, and Unselected foxes in the human-approach context (Gogoleva et al. 2008, 2010a, 2010c, 2013). For creating the playback stimuli, we selected 27 individual foxes (9 foxes per strain). From each individual, we selected 1 low-arousal whine (from Step 1) and 1 high-arousal whine (from Step 3), which provided 27 paired low-/high-arousal stimuli in total (Figure 1). Our choice was based on the quality of the recordings, background noise, and vocalizations of other animals. We equalized all experimental stimuli to the same root-mean square amplitude (70 dB). Fade in/out transitions of 5 ms were applied to all stimuli to remove any transients.

Acoustic analysis

To explore the effect of specific acoustic cues on humans' perception of arousal in whine calls across the 3 strains of silver fox, for each call, we measured the following 5 parameters: duration, tonality (harmonics-to-noise-ratio: HNR), SCOG, dominant frequency (DF)—that is, the frequency with the highest amplitude in the spectrum, and mean F0. We based the choice of the parameters to include in our analysis on findings from previous studies: duration, HNR, and F0 are shown to be linked to the emotional state of the caller (Morton 1977; Taylor and Reby 2010; Briefer 2012; Zimmermann et al. 2013) and SCOG affects the perception of arousal in humans (Sauter et al. 2010; Faragó et al. 2014).

We performed an automated acoustic analysis of the acoustic parameters in PRAAT (v. 5.2.26; Boersma 2002) for all parameters, except for duration, which was measured in Avisoft-SASLab Pro (Table 1). The duration was measured with the standard marker cursor in the main window of Avisoft. HNR was measured using the “To Harmonicity (cc)” command in PRAAT with standard settings.

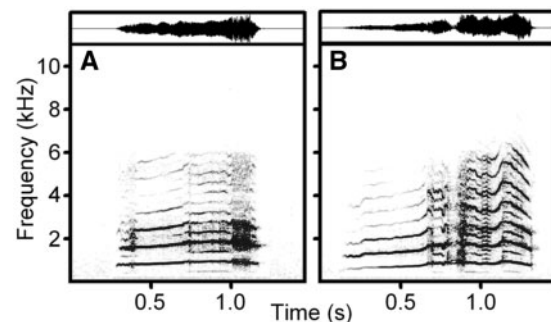


Figure 1. Spectrogram (below) and waveform (above) of: (A) low-arousal whine and (B) high-arousal whine of the same individual Aggressive silver fox.

Table 1. Descriptive table of acoustic values of the low- and high-arousal calls of each strain. For each of the acoustic parameters, namely DF (dominant frequency), duration, HNR (harmonic-to-noise-ratio), SCOG (spectral center of gravity), and F0 (mean fundamental frequency), the following values are provided: minimum and maximum, mean, and standard deviation of the mean

Strain	Arousal level	Acoustic parameter	N	Minimum	Maximum	Mean	SD
Aggressive	high	df (Hz)	9	378	2070	1024.667	602.020
		duration (s)	9	0.726	1.184	0.990	0.140
		HNR (dB)	9	3.415	14.451	9.584	4.103
		scog (Hz)	9	857.522	2072.206	1427.819	467.095
		F0 (Hz)	9	377.064	858.885	628.297	198.311
	low	df (Hz)	9	618	2022	1130.000	471.296
		duration (s)	9	0.679	1.358	1.016	0.233
		HNR (dB)	9	7.214	15.684	11.843	3.546
		scog (Hz)	9	902.781	1795.137	1225.265	311.676
		F0 (Hz)	9	379.728	851.781	573.562	171.136
Tame	high	df (Hz)	9	402	1206	714	327.866
		duration (s)	9	0.262	1.399	0.700	0.370
		HNR (dB)	9	-1.087	16.072	10.950	5.529
		scog (Hz)	9	554.540	3074.646	1274.658	719.675
		F0 (Hz)	9	401.655	1259.259	596.700	271.998
	low	df (Hz)	9	438	2130	1226	463.128
		duration (s)	9	0.232	1.074	0.573	0.255
		HNR (dB)	9	4.176	16.268	10.723	4.341
		scog (Hz)	9	874.089	2327.472	1447.183	434.674
		F0 (Hz)	9	320.834	842.181	522.081	157.799
Unselected	high	df (Hz)	9	306	1698	603.333	460.821
		duration (s)	9	0.749	1.864	1.102	0.342
		HNR (dB)	9	4.325	19.052	10.682	4.745
		scog (Hz)	9	470.349	1931.269	869.907	556.211
		F0 (Hz)	9	182.492	817.729	392.011	182.205
	low	df (Hz)	9	198	1986	591.333	580.197
		duration (s)	9	0.435	1.173	0.824	0.298
		HNR (dB)	9	6.246	14.969	10.046	2.695
		scog (Hz)	9	325.590	1304.692	607.799	333.890
		F0 (Hz)	9	213.290	505.722	347.699	102.763

SCOG was measured using the “To spectrum” and “Get center of gravity” commands (Power = 2.0). DF was measured using the “To Ltas:” command (bandwidth = 10 Hz, with no interpolation). Finally, the analysis of F0 was restricted to harmonics (integer multiples of F0). F0 was measured with the “Get pitch” command (Pitch settings: View range = 30–4,000 Hz, analysis method = cross-correlation. Advanced pitch settings: very accurate: “yes”, voicing threshold = 0.3).

In addition, the first author inspected the F0 contour of each stimulus visually. When the visible contour did not overlap with the first harmonic, the parameters “Pitch floor” and “Pitch Ceiling”, but sometimes also “Silence Threshold”, “Voicing Threshold”, and “Octave jump” 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. These settings’ adjustments were made for 23 (out of 54) stimuli.

Human participants

Twenty-seven participants (mean age = 26.26 years; SD = 5.65 years; 15 female), recruited at the Vrije Universiteit Brussel (Belgium) participated in this experiment in exchange for monetary compensation. The experimental design adopted for this study was approved by the university ethical review panel in accordance with the Helsinki Declaration. All participants gave written informed consent.

Experimental design

To avoid any bias in data collection and preliminary analyses, the experimenter (PF) was blind to the fox strain each caller belonged to. The experimental interface was created in *PsychoPy* (standalone version 1.81.0orc1; Peirce 2007). Participants were individually tested in a sound-attenuated room. The entire procedure was computerized. Stimuli were played binaurally over Shure SRH440 headphones.

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. Before the start of the experiment they were instructed to read an information sheet where the definition of arousal and the experimental procedure were explained. 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 (e.g., of low responsiveness to sensory stimulation) are calmness or boredom. Examples of high-arousal states (e.g., of high responsiveness to sensory stimulation) are anger or excitement.’

For familiarization with the experimental procedure, each participant completed 5 practice trials, each consisting of a pair of baby cries (www.freesound.org) varying in arousal level. During this practice phase, explicit instructions on the experimental procedure were displayed on the monitor. In the experimental phase, 27 pairs of calls were played in a randomized order across participants. Each trial in both phases was divided into 3 parts:

1. Sound playback: One low and one high-arousal vocalization emitted by the same individual were played with an inter-stimulus interval of 1 s. Order within pairs was randomized across participants. One sound would play while the letter “A” appeared on screen. At the end of sound A playback, the letter “A” faded out and then the other sound played while the letter “B” appeared on screen.
2. Relative rating of arousal: Participants were asked to indicate which vocalization expressed a higher level of arousal by clicking on the corresponding letter with the mouse. Given the short duration of our stimuli (see Table 1), to favor accurate assessment of sound features, participants could replay each sound *ad libitum* by pressing either letter (A or B) on the keyboard. No feedback was provided.
3. Absolute rating of arousal: Participants were asked to rate the level of arousal expressed in each vocalization by using a Likert scale ranging from 1 = very subdued to 7 = very excited. Again, they could replay each vocalization separately by pressing the corresponding letter on the keyboard. No feedback was provided.

Statistical analysis

All statistical analyses were performed using R 3.1.2 (R Development Core Team 2013). A binomial test and a signal detection analysis were performed to assess participants' accuracy in identifying high-arousal calls within each call pair across the 3 strains of silver fox. The dependent variable was the proportion of correct choices in participants' responses (where chance = 0.50).

We used a generalized linear mixed model (GLMM) to analyze humans' overall accuracy in identifying the high-arousal vocalization across strains. GLMMs were used because they allow fixed and random factors to be defined. Data across all participants were modeled using a binomial distribution. Participant ID was entered as a random factor, fox strain was entered as a fixed factor, and correct/incorrect response was entered as the outcome variable. False discovery rate (FDR) adjustments were applied to conduct pairwise comparisons. FDR was controlled at α level 0.05 following the procedure proposed by Benjamini and Hochberg (1995): For m tests, rank the P -values in ascending order $P(1) \leq P(2) \leq \dots \leq P(m)$, and denote by $H(i)$ the null hypothesis corresponding to $P(i)$; Let k be the largest i for which $P(i) \leq (\alpha/m) * i$ and reject all null hypotheses $H(1) \dots H(k)$. This means that, starting with the highest P -value each P is checked for this requirement; at the first P that meets the requirement its corresponding null hypothesis and all those having smaller P 's are rejected (Verhoeven et al. 2005). In addition, Cohen's (Cohen 1992) d effect sizes were calculated.

To assess which acoustic parameters affect human ability to identify the vocalization expressing a higher level of arousal within each strain, we performed separate GLMMs for each strain.

A separate GLMM was used to examine the acoustic parameters that predict participants' correct identification of high-arousal calls within each pair of arousal calls. Here, participant ID was entered as a random factor, acoustic parameters (duration, DF, SCOG, HNR, and F0 ratios) were entered as fixed factors, and the correct or incorrect response was entered as the outcome variable. Finally, a multiple linear regression analysis was computed to detect the acoustic parameters that predict absolute ratings for arousal level in each call. In this analysis, we included acoustic parameters as fixed factors and mean ratings for perceived emotional arousal in each call as outcome variable.

For all the analyses including acoustic parameters as fixed factors, we used a model selection procedure based on the 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. The AICc was used to rank the GLMMs and to obtain model weights (model.sel function, MuMIn library). Selection of the model(s), that is, of the model(s) with the highest power to explain variation in the dependent variable, is based on lowest AICc. When the difference between the AICc values of 2 models (Δ AICc) is less than 2 units, both models are considered as good as the best model (Symonds and Moussalli 2011). Models with Δ AICc up to 6 have considerably less support by the data. Models with values greater than 10 are sufficiently poorer than the best AIC model as to be considered implausible (Anderson and Burnham 2002).

Finally, we computed binary logistic regression models within the generalized linear model framework to assess any effect of order of sound playback within each call pair on participants response and of number of sound replays on participants' accuracy within the relative rating task.

Results

The binomial test and signal detection analysis revealed that participants' accuracy was above chance level for vocalizations of Aggressive silver foxes (proportion of correct responses: 0.59; $P = 0.007$; $d' = 0.455$) and of Unselected silver foxes (proportion of correct responses: 0.65; $P < 0.001$; $d' = 0.771$). Remarkably, we detected a significant effect in the opposite direction for stimuli of Tame silver foxes, where proportion of incorrect responses was higher than the expected 0.50 (proportion of incorrect responses: 0.60; $P = 0.003$; $d' = -0.507$) (Figure 2). In line with this result, the analysis performed within the GLMM revealed a significant effect of strain (Wald $\chi^2_2 = 31.681$, $P < 0.001$). Pairwise comparisons performed using the FDR correction revealed a significant difference between the effect of Tame and Aggressive strains ($Q < 0.001$; $d = -0.386$) and between Tame and Unselected strains ($Q < 0.001$; $d = -0.515$).

As shown in Table 2 (A), analyses performed within the GLMM revealed a significant effect of all the acoustic predictors included in our analysis on humans' accuracy in identifying arousal level in whine calls across all strains. However, critically, the model selection procedure ranked the models where F0, HNR, or SCOG ratios were excluded from the analyses as the weakest models, not resulting in a Δ AICc < 2 . A separate GLMM computed on each strain identified the following acoustic parameters as predictors of humans' accuracy in identifying arousal level in whine calls: DF and F0 ratios for Aggressive foxes, SCOG, F0, and HNR ratios for Unselected foxes, and duration and HNR ratios for Tame foxes. Model selection procedure applied on each of these models was in line with these results (Table 2 (B)).

The multiple linear regression analysis for the absolute ratings identified a significant effect of F0 and SCOG values of each call as the best predictors of the rated level of emotional arousal in each call—across strains (Table 3 (A)). The model selection procedure applied to this model was in line with these findings, and identified the models excluding DF as the weakest model, not resulting in a Δ AICc < 2 . However, according to this procedure, the model excluding duration or HNR, which have Δ AICc < 6 , as models that should not be discounted (Table 3 (B)). Results from a Shapiro Wilk test indicated that errors in the multiple linear regression analysis

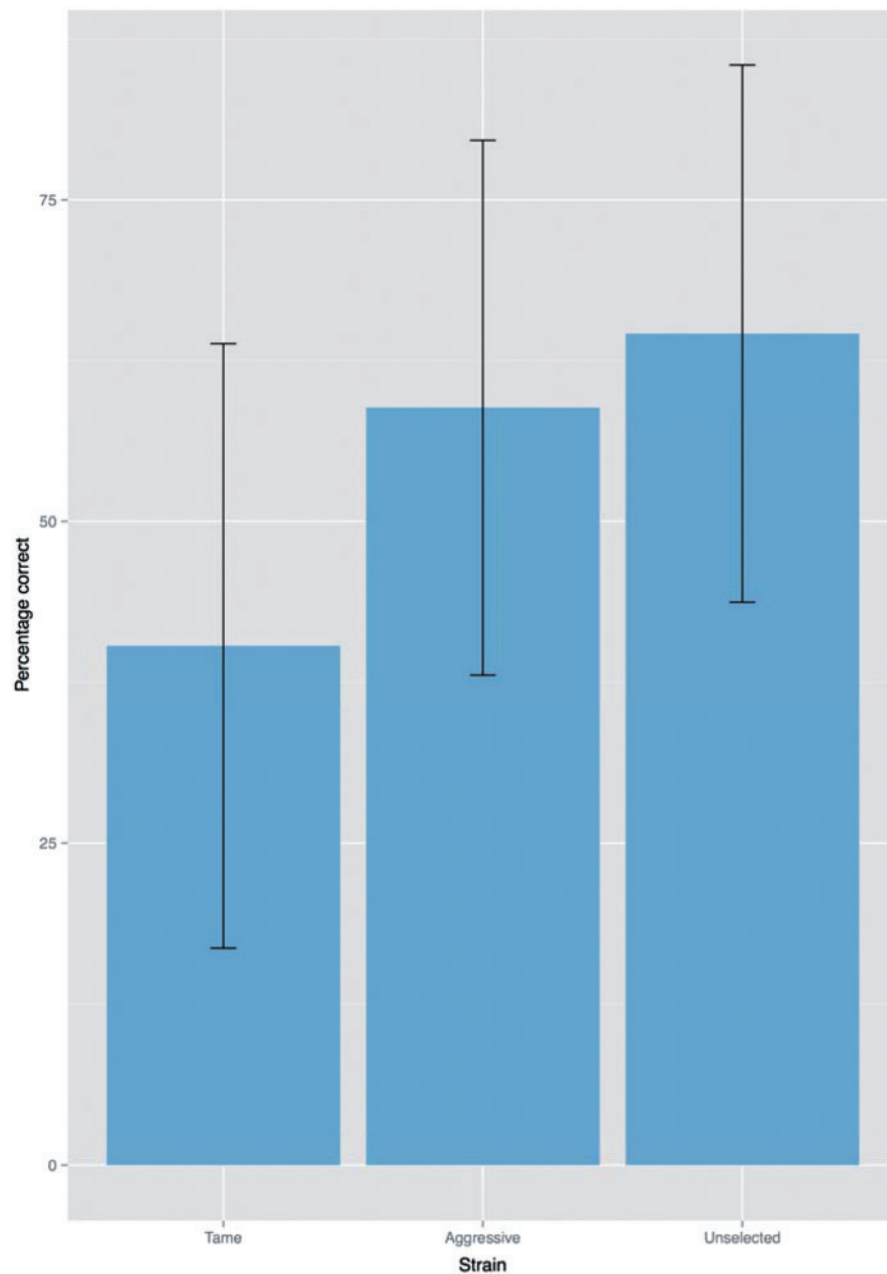


Figure 2. Mean percentage of correct responses for stimuli belonging to Tame, Aggressive, and Unselected silver foxes, averaged across participants. Error bars represent 95% confidence intervals.

were normally distributed ($W(54) = 0.981$, $P = 0.541$). To assess whether our model met the assumption of multicollinearity, we obtained the variance of inflation factor (VIF) value for the best predictors, and found that it was not substantially greater than 1 for any of them (SCOG: 1.662; F0: 1.613; HNR: 1.044). In addition, tolerance values were not below 0.2 (SCOG: 0.602; F0: 0.620; HNR: 0.958). Thus, we can exclude that collinearity is a problem for this model (Bowerman and O'Connell 1990). Finally, a Durbin-Watson test revealed that the residuals are not linearly auto-correlated ($d = 2.301$).

Finally, our analyses did not reveal any effect of order on correct response ($z = 0.486$, $P = 0.627$). No effect of number of sound replays on participants' correct response was detected ($z = 1.208$, $P = 0.227$).

Discussion

We found that humans are able to identify high arousal in vocalizations (namely, whine calls) of Aggressive and Unselected silver foxes. Intriguingly, in Tame silver foxes human participants identified low-arousal calls as expressing high arousal. Thus, participants' accuracy was significantly lower for whine calls produced by tame silver foxes, which have positive attitudes toward people, in comparison with Aggressive and Unselected foxes, which have negative attitude toward humans. In addition, we found that F0, HNR, and SCOG ratios predicted human accuracy in identifying high-arousal calls across all silver fox strains. Separate analyses revealed that different acoustic parameters affect human accuracy in identifying high-

Table 2. Relative rating task: (A) Values of the GLMMs computed across and within silver fox strains. We assessed acoustic predictors of humans' ability to identify vocalizations expressing higher levels of arousal across and within silver fox strains. Bold type indicate $P \leq 0.05$; degrees of freedom = 1 for all fixed factors. (B) Outcome of model selection procedure based on AICc. Degrees of freedom = 6 for all models. Bold type indicates models with the strongest support based on log likelihood (logLik), akaike weights and the difference between the AICc values of two models ($\Delta AICc \leq 2.0$)

(A) Generalized linear mixed models			(B) Model selection procedure				
Across strains							
Fixed effect	χ^2	P(χ^2)	Model	logLik	AICc	$\Delta AICc$	Akaike weight
duration	4.279	0.039	excluding duration	-422.123	856.362	0.000	0.592
DF	6.258	0.012	excluding DF	-423.113	858.342	1.979	0.220
F0	6.570	0.010	excluding F0	-423.269	858.654	2.291	0.188
HNR	21.582	<.001	excluding HNR	-430.775	873.666	17.303	0.000
SCOG	21.936	<.001	excluding SCOG	-430.952	874.020	17.658	0.000
Aggressive silver foxes							
Fixed effect	χ^2	P(χ^2)	Model	logLik	AICc	$\Delta AICc$	Akaike weight
duration	0.011	0.916	excluding duration	-135.460	283.277	0.000	0.438
HNR	1.541	0.214	excluding HNR	-136.225	284.807	1.530	0.204
SCOG	1.619	0.203	excluding SCOG	-136.264	284.884	1.607	0.196
DF	2.730	0.099	excluding DF	-136.820	285.995	2.718	0.112
F0	4.356	0.037	excluding F0	-137.633	287.621	4.345	0.050
Unselected silver foxes							
Fixed effect	χ^2	P(χ^2)	Model	logLik	AICc	$\Delta AICc$	Akaike weight
duration	0.278	0.598	excluding duration	-121.507	255.370	0.000	0.551
DF	0.808	0.369	excluding DF	-121.772	255.901	0.530	0.423
SCOG	7.126	0.008	excluding SCOG	-124.931	262.218	6.847	0.018
F0	9.381	0.002	excluding F0	-126.059	264.473	9.103	0.006
HNR	10.942	0.001	excluding HNR	-126.839	266.034	10.664	0.003
Tame silver foxes							
Fixed effect	χ^2	P(χ^2)	Model	logLik	AICc	$\Delta AICc$	Akaike weight
SCOG	0.010	0.920	excluding SCOG	-116.492	245.340	0.000	0.399
F0	0.095	0.757	excluding F0	-116.535	245.425	0.085	0.383
DF	1.691	0.193	excluding DF	-117.333	247.021	1.681	0.172
Duration	4.336	0.037	excluding duration	-118.655	249.666	4.325	0.046
HNR	81.549	<0.001	excluding HNR	-157.262	326.879	81.539	0.000

Table 3. Absolute rating task: (A) Values of the multiple linear regression model computed on human rating of arousal in silver fox calls. We assessed acoustic predictors of humans' perceived arousal in silver fox calls, as reported using a rating scale spanning from 1 (very subdued) to 7 (very excited). Bold type indicate $P \leq 0.05$; degrees of freedom = 1 for all fixed effects. (B) Outcome of model selection procedure based on AICc. Degrees of freedom = 6 for all models. Bold type indicates models with the strongest support based on loglikelihood (logLik), akaike weights, and the difference between the AICc values of two models ($\Delta AICc \leq 2.0$)

(A) Multiple linear regression model			(B) Model selection procedure				
Across strains							
Fixed effect	F	P(F)	Model	logLik	AICc	$\Delta AICc$	Akaike weight
DF	0.150	0.451	excluding DF	-37.297	88.380	0.000	0.558
duration	0.633	0.125	excluding duration	-38.314	90.416	2.036	0.202
HNR	0.727	0.101	excluding HNR	-38.506	90.800	2.419	0.166
F0	1.215	0.035	excluding F0	-39.491	92.770	4.389	0.062
SCOG	2.091	0.007	excluding SCOG	-41.169	96.125	7.745	0.012

arousal calls within each strain. Specifically, our analyses suggest that DF and F0 ratios affect human accuracy in identifying high-arousal calls of Aggressive silver foxes. Furthermore, we found that SCOG, F0, and HNR ratios affect human accuracy in identifying high-arousal calls of Unselected silver foxes, and that duration and HNR ratios affect human accuracy in identifying high-arousal calls of Tame silver foxes. Finally, our analyses suggest that F0 and SCOG are reliable predictors of humans' absolute ratings of arousal in our stimuli, although duration and HNR should not be discounted.

Our findings are consistent with previous research showing that humans are able to perceive arousal in vocalizations of cats (Nicastro and Owren 2003; McComb et al. 2009), dogs (Pongrácz et al. 2005; Faragó et al. 2014; Albuquerque et al. 2016), and piglets (Tallet et al. 2010; Maruščáková et al. 2015). Our results confirm findings from Filippi et al. (2016), suggesting that humans are able to recognize arousal in vocalizations emitted by members of species varying in size, social structure, and ecology. However, we have extended this line of research by disentangling human sensitivity to high-arousal calls with negative valence from human sensitivity to high-arousal calls with positive valence in silver foxes. In line with these studies, our work confirms that acoustic parameters associated with pitch perception play a key role in affecting human participants' ability for both identifying high arousal with negative valence in whine calls across the 3 strains of silver fox, and in assessing their absolute level of arousal on a rating scale. In addition, in line with research on vocal expression of arousal in mammals, which have identified in HNR an acoustic correlate of high arousal (Briefer 2012; but see Blumstein and Chi 2012), our analyses, which examines human perception of vocalizations with arousal content, suggest that HNR facilitates recognition of negative arousal in vocalizations of Unselected silver foxes and positive arousal in Tame silver foxes. Moreover, we found that HNR may have a role in affecting humans' absolute rating of arousal in silver foxes. One interesting finding is that duration affects humans' accuracy in identifying high-arousal calls with positive valence, that is, high-arousal calls emitted by Tame silver foxes. Previous research identified duration as a predictor of human perception of valence in animal calls, without controlling for arousal level (Faragó et al. 2014; Maruščáková et al. 2015). Our work complements this research, and particularly the study conducted by Faragó et al. (2014) on human perception of dog calls, in that we show that duration predicts humans' accuracy in identifying high arousal in silver fox calls with positive valence. In addition, in line with findings reported in Faragó et al. (2014), we found that duration may be used for the absolute rating of arousal in calls across all strains. Crucially, further experimental investigations are needed to estimate the effect of perceived loudness of the calls on correct assessment of their relative level of arousal.

It is possible that humans adopt the same kind of changes in perceived frequency-related parameters to assess arousal levels across silver foxes strains. However, recent research has provided evidence for the use of shared mechanisms in arousal perception across phylogenetically distant species (Belin et al. 2008; Altenmüller et al. 2013; Faragó et al. 2014; Filippi 2016; Filippi et al. 2016; Song et al. 2016). These findings suggest that mechanisms underlying perception and plausibly also expression of emotional arousal—which are related to stress induced higher effort in vocalization—may have emerged in the early stages of animal evolution as a result of selection pressures.

One interesting implication of our finding is that high-arousal calls with positive valence may not be as salient as high-arousal calls with negative valence to the human ear. Further research may analyze the

perceptual saliency of acoustic correlates of negative emotional arousal in vocalizations produced by individuals belonging either to the same or to different species. Humans' ability to recognize arousal in calls with positive valence might have not been selected by evolution because they are not as crucial for survival as arousal calls produced in negative-valenced contexts. Indeed, while the latter are a direct response to different degrees of external threat or danger, arousal calls with positive valence are emitted in contexts that may not be directly linked to survival. Therefore, in contrast to the ability to identify arousal calls with positive valence, humans may have evolved the ability to recognize arousal calls with negative valence as an adaptive trait. Further work is required to establish whether humans are able to identify high-arousal silver foxes calls that were elicited by the approach of another animal species. This analysis would add further support to the investigation of acoustic variation as a correlate of emotional arousal rather than as a function of predator species.

Notably, in the recordings adopted in our experiments, the emotional content of silver fox calls was inferred based on previous findings reporting on hormonal responses to human approach, and on observational indicators of the focal fox, namely on motor activity of the caller in response to the approaching behavior to the human experimenter. However, motor response in the 3 strains within the human-approach experimental setting described in the "Acoustic recordings and emotion classification" section was not statistically quantified. This is a crucial limitation of the present study. Future work should aim to classify arousal calls by combining quantitative assessment of observational correlates of arousal (e.g., motor response) with quantitative assessment of arousal in terms of physiological and/or neural responses of each specific caller during the production of calls. This is an important issue for future research, which should aim to integrate multiple types of data to quantify the exact degree of emotional arousal and valence. Specifically, further research is required to examine how behavioral observations of the contexts in which the vocalization is emitted can be mapped to data on brain activity (Belin et al. 2008; Panksepp 2011; Ocklenburg et al. 2013; Andics et al. 2014) and physiological states of the caller such as heart rate, adrenaline, or stress hormone levels at the time of vocal response to human approach (Paul et al. 2005; Briefer et al. 2015a, 2015b; Stocker et al. 2016). This quantitative assessment of arousal and valence of each caller, recorded during vocal production, may be adopted to fine-tune the examination of responses in listeners, assessing their neurological, physiological, and behavioral activity in response to each call. Crucially, further research in this direction might help to identify which emotion correlates in the focal Tame silver foxes led human participants to categorize low-arousal calls as expressing a high arousal. Generally, within this research framework, the analysis of acoustic correlates of emotional content in nonhuman animal vocalizations may be particularly valuable in assessing and improving animal welfare.

In conclusion, our findings provided empirical evidence for humans' sensitivity to arousal vocalizations across 3 strains of silver fox that have different genetically based predispositions toward approaching humans. We found that humans are able to identify high emotional arousal in whine calls with negative arousal—produced by Aggressive and Unselected silver foxes, but not in calls with positive valence, which were produced by Tame foxes. Our data did not identify in any of frequency-related parameters (F0, SCOG, DF) the best predictors of humans' assessment of arousal content in whine calls of Tame silver foxes, suggesting that these types of acoustic parameters in the calls may be key to recognition of negative arousal in animal calls (Lingle et al. 2012; Teichroeb et al. 2013; Lingle and Riede 2014;

Volodin et al. 2017). This work extends our understanding of vocal communication between species, providing key insights on the effect of acoustic correlates of emotional arousal and valence. Finally, this frame of investigation may enhance our understanding of the adaptive role of “eavesdropping” on heterospecific calls, providing key insights into the evolution of inter-species acoustic communication.

Supplementary material

Supplementary material can be found at <https://academic.oup.com/cz>.

Acknowledgements

We thank Anastasia V. Kharlamova, Lyudmila N. Trut for cooperation and the staff of the experimental fox farm for their help and support during the acoustic data collection. We are grateful to Tim Mahrt for valuable comments on the manuscript. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Funding

P. F. was supported by the following research grants: the European Research Council Starting Grant “ABACUS” [293435] awarded to BdB; ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and ANR-11-IDEX-0001-02 (A*MIDEX); a visiting fellowship from the Max Planck Society. For acoustic data collection from silver foxes at the experimental farm of the Institute of Cytology and Genetics, Novosibirsk, Russia and preparation and the acoustic analyses of the playback stimuli, SSG, EVV and IAV were supported by the Russian Science Foundation, grant [14-14-00237].

Author contribution

P. F. developed the study concept and design. P. F., S. S. G., E. V. V., and I. A. V. 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. All authors approved the final version of the manuscript for submission.

References

Albuquerque N, Guo K, Wilkinson A, Savalli C, Otta E et al., 2016. Dogs recognize dog and human emotions. *Biol Lett* 12:20150883.

Altenmüller E, Schmidt S, Zimmermann E, 2013. A cross-taxa concept of emotion in acoustic communication: an ethological perspective. In: Altenmüller E, Schmidt S, Zimmermann E, editors. *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*. Oxford: Oxford University Press. 339–367.

Anderson DJ, Adolphs R, 2014. A framework for studying emotions across species. *Cell* 157:187–200.

Anderson DR, Burnham KP, 2002. Avoiding pitfalls when using information-theoretic models. *The Jf Wildlife Manage* 66:912–918.

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.

Aubin T, 1991. Why do distress calls evoke interspecific responses. *Behav Process* 23:103–111.

Bakken M, 1998. The effect of an improved man-animal relationship on sex-ratio in litters and on growth and behaviour in cubs among farmed silver foxes *Vulpes vulpes*. *Appl Anim Behav Sci* 56:309–317.

Bakken M, Moe RO, Smith AJ, Selle G-ME, 1999. Effects of environmental stressors on deep body temperature and activity levels in silver fox vixens *Vulpes vulpes*. *Appl Anim Behav Sci* 64:141–151.

Belin P, Fecteau S, Charest I, Nicastro N, Hauser M et al., 2008. Human cerebral response to animal affective vocalizations. *Proc Roy Soc B* 275:473–481.

Belyaev DK, 1979. Destabilizing selection as a factor in domestication. *J Hered* 70:301–308.

Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Sta Soc. Series B (Methodological)* 57: 289–300.

Blumstein DT, Chi YY, 2012. Scared and less noisy: glucocorticoids are associated with alarm call entropy. *Biol Lett* 8:189–192.

Boersma P, 2002. Praat, a system for doing phonetics by computer. *Glott International* 5: 341–345. Version 5.4.02 retrieved from <http://www.praat.org/>.

Bowerman BL, O’connell RT, 1990. *Linear Statistical Models: An Applied Approach*. Boston: PWS-Kent Publishing Company. 106–129.

Briefer EF, 2012. Vocal expression of emotions in mammals: mechanisms of production and evidence. *J Zool* 288:1–20.

Briefer EF, Maigrot AL, Mandel R, Freymond SB, Bachmann I et al., 2015a. Segregation of information about emotional arousal and valence in horse whinnies. *Sc Rep* 4:9989.

Briefer EF, Tettamanti F, McElligott A, 2015b. Emotions in goats: mapping physiological, behavioural and vocal profiles. *Anim Behav* 99:131–143.

Bryant GA, 2013. Animal signals and emotion in music: coordinating affect across groups. *Front Psychol* 4:990.

Cohen J, 1992. Statistical power analysis. *Curr Dir Psychol Sci* 1:98–101.

Cross N, Rogers LJ, 2005. Mobbing vocalizations as a coping response in the common marmoset. *Horm Behav* 49:237–245.

Darwin C, 1872. *The Expression of the Emotions in Man and Animals*. London: John Murray.

Davis PJ, Zhang SP, Winkworth A, Bandler R, Davis P, 1996. Neural control of vocalization: respiratory and emotional influences. *J Voice* 10:23–38.

de Boer B, Wich SA, Hardus ME, Lameira AR, 2015. Acoustic models of orangutan hand-assisted alarm calls. *J Exp Biol* 218:907–914.

De Kort SR, Carel ten C, 2001. Response to interspecific vocalizations is affected by degree of phylogenetic relatedness in *Streptopelia* doves. *Anim Behav* 61:239–247.

Desrochers AU, Be M, Bourque J, 2002. Do mobbing calls affect the perception of predation risk by forest birds? *Anim Behav* 64:709–714.

Fallow PM, Gardner JL, Magrath RD, 2011. Sound familiar? Acoustic similarity provokes responses to unfamiliar heterospecific alarm calls. *Behav Ecol* 22:401–410.

Fallow PM, Magrath RD, 2010. Eavesdropping on other species: mutual interspecific understanding of urgency information in avian alarm calls. *Anim Behav* 79:411–417.

Faragó T, Andics A, Devecseri V, Kis A, Gácsi M et al., 2014. Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biol Lett* 10:20130926.

Fernald A, 1992. Human maternal vocalizations to infants as biologically relevant signals: an evolutionary perspective. In: Barkow JH, Cosmides L, Tooby J, editors. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press. 391–428.

Filippi P, 2016. Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Front Psychol* 7:1393.

Filippi P, Congdon JV, Hoang J, Bowling DL, Reber SA et al., 2016. Humans recognize vocal expressions of emotional states universally across species. In: Roberts SG, Cuskley C, McCrohon L, Barceló-Coblijn L, Fehér O et al., editors. *The Evolution of Language: Proceedings of the 11th International Conference (EVLANG11)*. Available online: <http://evolang.org/neworleans/papers/91.html>.

Frijda NH, 1987. Emotion, cognitive structure, and action tendency. *Cognition Emotion* 1:115–143.

Gogoleva SS, Volodin IA, Volodina EV, Kharlamova AV, Trut LN, 2009. Kind granddaughters of angry grandmothers: the effect of domestication on vocalization in cross-bred silver foxes. *Behav Process* 81:369–375.

Gogoleva SS, Volodin IA, Volodina EV, Kharlamova AV, Trut LN, 2010a. Sign and strength of emotional arousal: vocal correlates of positive and negative attitudes to humans in silver foxes *Vulpes vulpes*. *Behaviour* 147:1713–1736.

- Gogoleva SS, Volodin IA, Volodina EV, Kharlamova AV, Trut LN, 2010b. Vocalization toward conspecifics in silver foxes *Vulpes vulpes* selected for tame or aggressive behavior toward humans. *Behav Process* 84:547–554.
- Gogoleva SS, Volodina EV, Volodin IA, Kharlamova AV, Trut LN, 2010c. The gradual vocal responses to human-provoked discomfort in farmed silver foxes. *Acta Ethologica* 13:75–85.
- Gogoleva SS, Volodin IA, Volodina EV, Kharlamova AV, Trut LN, 2011. Explosive vocal activity for attracting human attention is related to domestication in silver fox. *Behav Process* 86:216–221.
- Gogoleva SS, Volodin IA, Volodina EV, Kharlamova AV, Trut LN, 2013. Effects of selection for behavior, human approach mode and sex on vocalization in silver fox. *J Ethol* 31:95–100.
- Hauser MD, 1988. How infant vervet monkeys learn to recognize starling alarm calls: the role of experience. *Behaviour* 105:187–201.
- Johnson FR, Mcnaughton EJ, Shelley CD, Blumstein DT, 2003. Mechanisms of heterospecific recognition in avian mobbing calls. *Aust J Zool* 51:577–585.
- Kitchen D, Bergman T, Cheney D, Nicholson J, 2010. Comparing responses of four ungulate species to playbacks of baboon alarm calls. *Anim Cognition* 13:861–870.
- Kukekova AV, Oskina IN, Kharlamova AV, Chase K, Temnykh SV et al., 2008a. Fox farm experiment: hunting for behavioral genes. *VOGis Herald* 12:50–62.
- Kukekova AV, Temnykh SV, Johnson JL, Trut LN, Acland GM, 2012. Genetics of behavior in the silver fox. *Mamm Genome* 23:164–177.
- Kukekova AV, Trut LN, Chase K, Shepeleva DV, Vladimirova et al., 2008b. Measurement of segregating behaviors in experimental silver fox pedigrees. *Behav Genet* 38:185–194.
- Laukka P, Juslin P, Bresin R, 2005. A dimensional approach to vocal expression of emotion. *Cognition Emotion* 19:633–653.
- Lea AJ, Barrera JP, Tom LM, Blumstein DT, 2008. Heterospecific eavesdropping in a nonsocial species. *Behav Ecol* 19:1041–1046.
- Lingle S, Riede T, 2014. Deer mothers are sensitive to infant distress vocalizations of diverse mammalian species. *Am Nat* 184:510–522.
- Lingle S, Wyman MT, Kotrba R, Teichroeb LJ, Romanow CA, 2012. What makes a cry a cry? A review of infant distress vocalizations. *Curr Zool* 58:698–726.
- Magrath RD, Haff TM, Fallow PM, Radford AN, 2015. Eavesdropping on heterospecific alarm calls: from mechanisms to consequences. *Biol Rev* 90:560–586.
- Maruščáková IL, Linhart P, Ratcliffe VF, 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.
- McComb K, Taylor AM, Wilson C, Charlton BD, 2009. The cry embedded within the purr. *Curr Biol* 19:R507–R508.
- McConnell PB, Baylis JR, 2010. Interspecific communication in cooperative herding: acoustic and visual signals from human shepherds and herding dogs. *Z Tierpsychol* 67:302–328.
- Mendl M, Burman OHP, Paul ES, 2010. An integrative and functional framework for the study of animal emotion and mood. *Proc Roy Soc Lond B* 2010, 277:2895–2904.
- Morton ES, 1977. On the occurrence and significance of motivation: structural rules in some bird and mammal sounds. *Am Nat* 111:855–869.
- Nesse RM, 1990. Evolutionary explanations of emotions. *Hum Nat* 1:261–289.
- Newton-Fisher N, Harris S, White P, Jones G, 1993. Structure and function of red fox *Vulpes vulpes* vocalizations. *Bioacoustics* 5:1–31.
- Nicastro N, Owren MJ, 2003. Classification of domestic cat *Felis catus* vocalizations by naive and experienced human listeners. *J Comp Psychol* 117:44–52.
- Nimon AJ, Broom DM, 2001. The welfare of farmed foxes *Vulpes vulpes* and *Alopex lagopus* in relation to housing and management: a review. *Anim Welfare* 10:223–248.
- Ocklenburg S, Ströckens F, Güntürkün O, 2013. Lateralisation of conspecific vocalisation in non-human vertebrates. *Laterality: Asymm Body, Brain Cognition* 18:1–31.
- Oskina IN, 1996. Analysis of the functional state of the pituitary-adrenal axis during postnatal development of domesticated silver foxes *Vulpes vulpes*. *Scientifur* 20:159–161.
- Oskina IN, Herbeck YE, Shikhevich SG, Plyusnina IZ, Gulevich RG, 2008. Alterations in the hypothalamus-pituitary-adrenal and immune systems during selection of animals for tame behavior. *VOGis Herald* 12:39–49. (in Russian).
- Owings D, Morton E, 1998. *Animal Vocal Communication: A New Approach*. Cambridge: Cambridge University Press.
- Panksepp J, 2011. The basic emotional circuits of mammalian brains: do animals have affective lives? *Neurosci Biobehav Rev* 35:1791–1804.
- Paul ES, Harding EJ, Mendl M, 2005. Measuring emotional processes in animals: the utility of a cognitive approach. *Neurosci Biobehav Rev* 29:469–491.
- Pedersen V, 1991. Early experience with the farm environment and later effects on behaviour in silver *Vulpes vulpes* and blue *Alopex lagopus* foxes. *Behav Process* 25:163–169.
- Pedersen V, 1993. Effects of different post-weaning handling procedures on the later behaviour of silver foxes. *Appl Anim Behav Sci* 37:239–250.
- Pedersen V, 1994. Long-term effects of different handling procedures on behavioural, physiological, and production-related parameters in silver foxes. *Appl Anim Behav Sci* 40:285–296.
- Pedersen V, Jeppesen LL, 1990. Effect of early handling on later behaviour and stress responses in the silver fox *Vulpes vulpes*. *Appl Anim Behav Sci* 26:383–393.
- Peirce JW, 2007. PsychoPy: psychophysics software in Python. *J Neurosci Methods* 162:8–13.
- Pongrácz P, Miklósi A, Molnár C, Csányi V, 2005. Human listeners are able to classify dog *Canis familiaris* barks recorded in different situations. *J Comp Psychol* 119:136–144.
- Pongrácz P, Molnár C, Miklósi A, 2006. Acoustic parameters of dog barks carry emotional information for humans. *Appl Anim Behav Sci* 100:228–240.
- R Development Core Team, 2013. R Foundation for Statistical Computing. Vienna: Austria. Version 3.1.2 retrieved from <http://www.R-project.org>.
- Ratcliff E, 2011. Taming the wild. *Nat Geogr* 219:34–59.
- Russell JA, 1980. A circumplex model of affect. *J Pers Soc Psychol* 39:1161–1178.
- Sauter DA, Eisner F, Calder AJ, Scott SK, 2010. Perceptual cues in nonverbal vocal expressions of emotion. *Quart J Exp Psychol* 63:2251–2272.
- 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.
- Song X, Osmanski MS, Guo Y, Wang X, 2016. Complex pitch perception mechanisms are shared by humans and a New World monkey. *Proc Natl Acad Sci* 113:781–786.
- Specht R, 2002. *Avisoft-Saslab Pro: Sound Analysis and Synthesis Laboratory*. Berlin: Avisoft Bioacoustics.
- Stocker M, Munteanu A, Stöwe M, Schwab C, Palme R et al., 2016. Loner or socializer? Ravens' adrenocortical response to individual separation depends on social integration. *Horm Behav* 78:194–199.
- 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.
- Tallet C, Špinka M, Maruščáková I, Šimeček P, 2010. Human perception of vocalizations of domestic piglets and modulation by experience with domestic pigs *Sus scrofa*. *J Comp Psychol* 124:81–91.
- Taylor AM, Reby D, 2010. The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology* 280:221–236.
- Tchernichovski O, Oller DK, 2016. Vocal development: how marmoset infants express their feelings. *Curr Biol* 26:R422–R424.
- Teichroeb LJ, Riede T, Kotrba R, Lingle S, 2013. Fundamental frequency is key to response of female deer to juvenile distress calls. *Behav Process* 92:15–23.
- Templeton C, Greene E, Davis K, 2005. Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science* 308:1934–1937.
- Trut LN, 1980. The genetics and phenogenetics of domestic behavior. In: Belyaev DK, editor. *Proceedings of the XIV International Congress of Genetics*. Vol. 2, Book 2: *Problems of General Genetic*. Moscow: MIR Publishers. 123–136.

- Trut LN, 1999. Early canid domestication: the farm-fox experiment. *Am Sci* 87:160–169.
- Trut LN, 2001. Experimental studies of early canid domestication. In: Ruvinsky A, Sampson J, editors. *The Genetics of the Dog*. New York: CABI Publishing. 15–41.
- Trut LN, Oskina IN, Kharlamova AV, 2009. Animal evolution during domestication: the domesticated fox as a model. *BioEssays* 31:349–360.
- Trut LN, Plyusnina IZ, Oskina IN, 2004. An experiment on fox domestication and debatable issues of evolution of the dog. *Russ J Genet* 40:644–655.
- Verhoeven KJF, Simonsen KL, McIntyre LM, 2005. Implementing false discovery rate control: Increasing your power. *Oikos* 108:643–647.
- Volodin IA, Sibiryakova OV, Frey R, Efreмова KO, Soldatova NV et al., 2017. Individuality of distress and discomfort calls in neonates with bass voices: wild–living goitred gazelles *Gazella subgutturosa* and saiga antelopes *Saiga tatarica*. *Ethology* 123:386–396.
- Zhang YS, Ghazanfar AA, 2016. Perinatally influenced autonomic system fluctuations drive infant vocal sequences. *Curr Biol* 26:1249–1260.
- 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: Altenmüller E, Schmidt S, Zimmermann E, editors. *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*. Oxford: Oxford University Press. 116–132.