



OPEN

# Artificial sounds following biological rules: A novel approach for non-verbal communication in HRI

Beáta Korcsok<sup>1</sup>✉, Tamás Faragó<sup>2</sup>, Bence Ferdinandy<sup>3</sup>, Ádám Miklósi<sup>2,3</sup>, Péter Korondi<sup>1</sup> & Márta Gácsi<sup>2,3</sup>

Emotionally expressive non-verbal vocalizations can play a major role in human-robot interactions. Humans can assess the intensity and emotional valence of animal vocalizations based on simple acoustic features such as call length and fundamental frequency. These simple encoding rules are suggested to be general across terrestrial vertebrates. To test the degree of this generalizability, our aim was to synthesize a set of artificial sounds by systematically changing the call length and fundamental frequency, and examine how emotional valence and intensity is attributed to them by humans. Based on sine wave sounds, we generated sound samples in seven categories by increasing complexity via incorporating different characteristics of animal vocalizations. We used an online questionnaire to measure the perceived emotional valence and intensity of the sounds in a two-dimensional model of emotions. The results show that sounds with low fundamental frequency and shorter call lengths were considered to have a more positive valence, and samples with high fundamental frequency were rated as more intense across all categories, regardless of the sound complexity. We conclude that applying the basic rules of vocal emotion encoding can be a good starting point for the development of novel non-verbal vocalizations for artificial agents.

With the growing importance of social robots and other artificial agents, the development of adequate communication in Human-Robot and Human-Computer Interaction (HRI and HCI) is becoming imperative. A common approach in developing the communicational signals of social robots and other artificial agents is to base them on human communication<sup>1</sup> e.g., on speech<sup>2</sup> and human-specific gestures<sup>3</sup>. Human-like communication seems to be a natural way of interaction for social robots, as human languages can convey high complexity in sharing information<sup>4</sup>, and e.g., facial gestures can express a wide variety of affective states<sup>5</sup>. However, this approach is frequently undermined by technological limitations relating to the perceptive, cognitive, and motion skills implemented in the agent, which can become more obvious during the course of interaction, leading to disappointment<sup>6,7</sup>. Overt similarity can also cause aversion towards human-like robots (Uncanny Valley<sup>8,9</sup>). Furthermore, the proposed functions of specific robots do not always require the level of complexity found in human communication<sup>6</sup>, or their capabilities and functions are not in line with that of humans (e.g., no need for head-turning with 360° vision<sup>9</sup>, no morphological limitations in sound production). To avoid these issues, another approach is to consider HRI as interspecific interaction in which the artificial agent is regarded as a separate species, and only has to be equipped with a basic level of social competence and communicational skills that are aligned with its function<sup>9</sup>. In this framework formation of non-verbal communicational signals of artificial agents rely heavily on the foundations of biological signalling and are based on the behaviour of social animals. A plausible example for such a basis could be the dog (*Canis familiaris*), with which humans have an interspecific bond that is, in many aspects, functionally analogous to the relationship needed in HRI<sup>6,9,10</sup>.

Upholding this approach, features of non-verbal communication not only show common aspects across human cultures e.g., in facial expressions<sup>11</sup> and non-verbal vocalizations<sup>12</sup>, but we can also find similarities with the communicational signals of non-human animals<sup>13,14</sup>, for a review see<sup>15</sup>. These similarities allow the use of communicational signals that are based on general rules observed across multiple taxa<sup>16</sup> or on the behaviour of specific animal species, e.g., dogs<sup>6,17</sup> in artificial agents.

<sup>1</sup>Department of Mechatronics, Optics and Mechanical Engineering Informatics, Faculty of Mechanical Engineering, Budapest University of Technology and Economics, Budapest, Hungary. <sup>2</sup>Department of Ethology, Eötvös Loránd University, Budapest, Hungary. <sup>3</sup>MTA-ELTE Comparative Ethology Research Group, Budapest, Hungary. ✉e-mail: korcsok@mogi.bme.hu

In case of emotionally expressive vocal signals, the similarities between taxa emerge due to the evolutionarily conservative processes of sound production and vocal tract anatomy in terrestrial tetrapods<sup>18</sup>, making sound-based communicational signals more conservative than many other. This phenomenon is best described by the source-filter framework, connecting the physiological processes and anatomical structures to the acoustic features of vocalizations<sup>19</sup>. The source-filter framework also explains the physiological connection between the inner state of an animal and the related vocalizations<sup>15</sup>. Vocalizations are thought to have developed from involuntary sounds of exhalation (e.g., due to quick movements when escaping a predator) usually connected to specific inner states, which through ritualization can lead to communicative sounds<sup>20</sup>. These sounds contain acoustic features influenced by the original physiological changes related to inner states, e.g., the tenseness of respiratory muscles and stretching of vocal folds via laryngeal muscles is increased during high arousal, leading to increased call length and pitch<sup>15,21</sup>. Chimpanzee (*Pan troglodytes*) screams emitted during a severe attack from conspecifics are higher in frequency and longer than screams produced under less aggressive attacks<sup>22</sup>. Similarly, in baboons<sup>23</sup> calls produced in high arousal contexts featured longer individual call lengths and higher average fundamental frequency (among other parameter changes).

Previous studies have proven that emotionally expressive human non-verbal vocalizations are easily recognizable across cultures<sup>24</sup>. Humans are also able to recognise animal vocalizations as emotionally expressive sounds<sup>14</sup>, and rate their valence and intensity similarly to humans' based on acoustic parameters such as the fundamental frequency ( $f_0$ ) or call length (e.g., dogs<sup>25</sup>; domestic pigs (*Sus scrofa*)<sup>26,27</sup>). Animal and human vocalizations with higher fundamental frequencies are perceived as more intense, while vocalizations consisting of short calls are rated as more positive in valence<sup>25,26</sup>. Thus, fundamental frequency and call length can serve as acoustic cues for the listeners, informing them about the inner state of the vocalizing individual even in interspecies communication<sup>28,29</sup>. The aforementioned acoustic cues are consistent with the existence of simple coding rules of affective vocalizations that are shared in mammals and that are the result of homologous signal production and neural processing<sup>30</sup>. These simple coding rules, namely that higher fundamental frequency is connected to higher intensity and shorter call length is connected to positive valence, are substantiated by studies on multiple mammalian species (for reviews see<sup>14,15,21,31</sup>) and in connection with various acoustic parameters (e.g. low harmonic-to-noise ratio is connected to higher arousal in baboons<sup>32</sup>, dogs<sup>25</sup> and bonnet macaques (*Macaca radiata*)<sup>33</sup>).

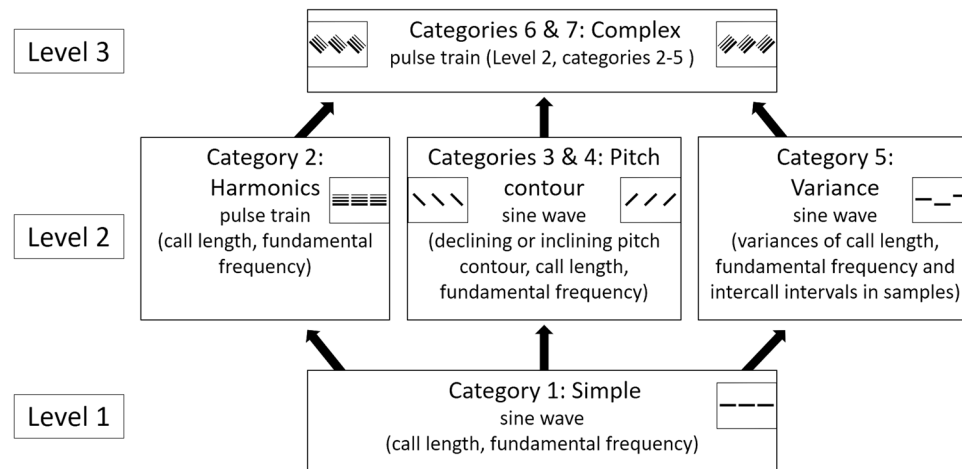
A comparative fMRI study conducted on humans and dogs<sup>34</sup> has shown that the acoustic cues connected to the emotional valence of vocalizations are processed similarly in both species. Another study by Belin *et al.*<sup>35</sup> showed that the cerebral response of human participants differed between hearing positive or negative valence vocalizations of rhesus monkeys (*Macaca mulatta*) and cats (*Felis catus*), and one brain region (the ventrolateral orbitofrontal cortex) showed greater activation to negative valence vocalizations from both animal species and also humans. These conserved neural systems might help to decode emotions from vocalizations and sounds, providing a basis for emotionally expressive cross-species communication<sup>15</sup>. As interaction with social robots and other artificial agents can be viewed as cross-species communication<sup>9</sup>, exploring the processes and behavioural expressions of emotional states in humans and non-human animals can advance the development of artificial communication systems.

The use of non-verbal acoustic features for emotion expression has been studied in HRI<sup>36,37</sup>. Yilmazyildiz *et al.*<sup>38</sup> provided an extensive review of the research area of Semantic-Free Utterances (SFU), which includes the research of gibberish speech, e.g., creating affectively expressive speech with no semantic content via manipulating the vowel nuclei and consonant clusters of syllables of existing languages<sup>39</sup>, non-linguistic utterances and utterances based on musical theory, e.g., melodies comprised of synthesized notes with modification of multiple acoustic parameters (fundamental frequency, fundamental frequency range, pause length and ratio, envelope, pitch contour and musical major and minor modes)<sup>40</sup>, auditory icons (e.g., decreasing pitch representing a falling object<sup>38</sup>); and paralinguistic utterances, e.g., human laughter<sup>41</sup>.

A detailed review by Juslin and Laukka<sup>42</sup> found that basic emotional categories are expressed similarly in human non-verbal emotion expressions as in musical performance, following the same acoustic cues, drawing further attention to the evolutionary background of acoustic emotion expression in mammals. Although multiple SFU studies investigate acoustic parameters that have biological backgrounds, most of the research is focused on signals derived from human communication or culture, and therefore draw from a higher-level system. We propose that establishing simple coding rules for the emotional effect of artificial sounds based on interspecies similarities of sound production and signal processing represents a more fundamental approach with a strong evolutionary basis, which can serve as a complementary general principle to other viewpoints.

As human and animal vocalizations are acoustically complex signals, we follow a systematic approach to reveal which parameters of the vocalizations contribute to basic coding rules, and whether other acoustic parameters affect them. The vocalizations of mammals contain characteristic acoustic parameters due to the similar processes of sound production, e.g., formants, which are generated when the source sound is filtered through the vocal tract, attenuating certain frequencies<sup>31</sup>. Conversely, vocalizations of artificial agents are not bound to morphological structures and can be freely adjusted to the function of the robot. However, biological features increase the similarities of the artificial agents to living beings, which is generally desirable in their interactions with humans<sup>6,16</sup>. While the fundamental frequency and call length can be modified according to the general coding rules found in animal vocalizations<sup>14,25,26</sup> even in simple artificially generated sine-wave sounds that do not depict commonplace terrestrial mammal vocalizations, adding other parameters characteristic of animal vocalizations can increase their perceived animacy.

Following the previously outlined concepts, we created artificial sounds based on general acoustic features of emotionally expressive vocalisations of humans and non-human animals<sup>25,31</sup>. The specific ranges of the various acoustic parameters were mostly based on dog vocalizations, as these have been previously studied with a similar methodology to our own, providing an insight into how humans rate them on valence and intensity in a questionnaire study<sup>25</sup>, while we have less comparable results in e.g., primates. The sound samples were generated



**Figure 1.** The categories of the artificial sounds across three levels of complexity. In each category the basis of the sound (sine wave or pulse train) is followed by the changed parameters in parenthesis. Level 1 category 1: Simple sine wave; Level 2 category 2: Pulse train; category 3: Sine wave sounds with pitch contour down; category 4: Sine wave sounds with pitch contour up; category 5: Variable sine wave; Level 3 category 6: Complex pulse train sounds with pitch contour down; category 7: Complex pulse train sounds with pitch contour up.

in multiple categories. We used sine-wave sounds for the simplest sound category, as these are single-frequency sounds that rarely occur naturally<sup>43</sup>, but which are frequently used in artificial signals of machines. Then, starting with the simple sine-wave sounds we added new acoustic features (pitch contour changes, harmonics, variations of call properties within a sound sample, formants) that are characteristic of animal vocalizations to make more complex and biologically more congruent samples. In each category, we systematically changed the fundamental frequency and call length of the sounds to cover the relevant acoustic ranges of these parameters (for more details see Fig. 1 and Table 1).

## Questions and Hypotheses

Our main question was whether the simple coding rules of fundamental frequency and call length of vocalizations are also applied to artificially generated sounds.

Our hypothesis was:

H0: Simple coding rules do not exist, the direction of the effects of the acoustic parameters on the emotion ratings are different on the distinct complexity levels.

H1: Simple coding rules apply to artificial sounds as well, the direction of the effects of the acoustic parameters on the emotion ratings are the same.




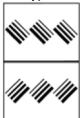
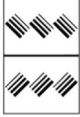
In this latter case we expect that human listeners perceive artificial sounds with higher fundamental frequency as more intense and sounds with longer calls as having more negative valence, just like in case of human and animal sounds, as we can already find the simple coding rules in complex biological sounds evolved to communicate inner states. In parallel, neural systems are present to process these basic acoustic features. Moreover, if the presence of the features that are inherent consequences of the voice production system are inevitable for accepting a sound as biological and thus being a communicative signal encoding emotional states, the simple coding rules could have a stronger effect (a.k.a. stronger association between acoustic features and emotional scales) in more complex sounds.

## Method

**Subjects.** All subjects were unpaid volunteers from various nationalities recruited via online advertisements. The number of participants in the final analysis were 237, from which 95 chose to fill the questionnaire in Hungarian (60 female, 35 male, mean age =  $36.3 \pm \text{SD } 11.8$  years) and 142 in English (122 female, 20 male, mean age  $39.9 \pm \text{SD } 11.7$  years). Questionnaire answers were discarded if the participant was under the age of 18. Part of our sample abandoned the survey before finishing (95 individuals) but as the sample presentation was random, these unfinished responses are unlikely to cause any bias, thus they were included in the analysis. Subjects gave their informed consent to participate in the study, which was carried out in accordance with the relevant guidelines and with the approval of the Institutional Review Board of the Institute of Biology, Eötvös Loránd University, Budapest, Hungary (reference number of the ethical permission: 2019/49).

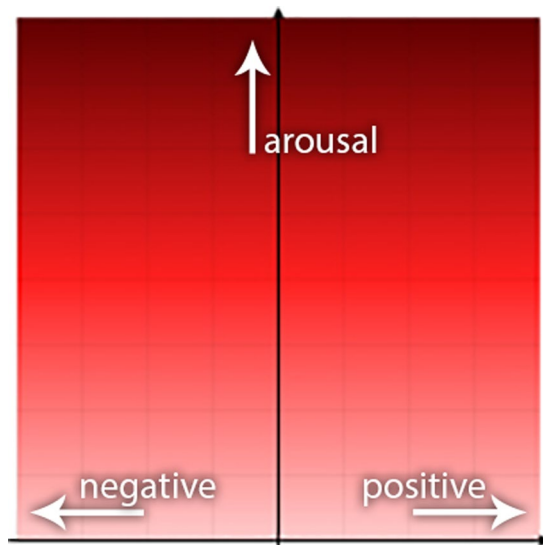
**Stimuli.** The artificial sounds were generated using a custom Praat (version 6.0.19) script (developed by TF and BK, see Supplementary Methods). The sound samples consisted of calls separated by mute intercall periods forming bouts. We varied both the lengths of the calls ( $cl$ ) and the fundamental frequency ( $f_0$ ) in all cases. The range of most parameters was set in accordance with the non-verbal human and dog vocalizations used in<sup>25</sup>.

The range of fundamental frequency varied between 65 Hz to 1365 Hz with 100 Hz steps. There were multiple samples at each frequency step, with differing call lengths; sound samples were generated at every 0.03 s

Parameters	Value or Range (across all samples)	Variance in categories 1.sin; 2.pulse; 3.pitch_d; 4.pitch_u	Variance in categories 5.var_sin; 6.comp_d; 7.comp_u	Reference
Fundamental frequency ( $f_0$ )	65 Hz – 1365 Hz			~50–1600 Hz <sup>25</sup>
Total length (call length + interval length)	~2 s (+ silence until 3 s total duration)			2 s <sup>25</sup>
Call length	0.07; 0.16; 0.46; 0.76; 1.06; 1.96 s			0.11–2 s <sup>25</sup>
Intercall interval length	0.2 s	uniformly distributed random value, $\pm 25\%$ of interval length	uniformly distributed random value, $\pm 50\%$ of interval length	0.05–1.7 s <sup>50</sup>
Pitch contour change in categories 3.pitch_d; 4.pitch_u; 6.comp_d; 7.comp_u	uniformly distributed random value, $\pm 10\%$ of $f_0$			71
Vocal tract length in categories 6.comp_d; 7.comp_u	20 cm			Modelling medium sized dog <sup>72</sup>
Number of formants in categories 6.comp_d; 7.comp_u	10			
First formant ( $f_1$ ) in categories 6.comp_d; 7.comp_u	550 Hz			

**Table 1.** Parameters of sound samples. Categories: 1.sin: Simple sine wave; 2.pulse: Pulse train; 3.pitch\_d: Sine wave sounds with pitch contour down; 4.pitch\_u: Sine wave sounds with pitch contour up; 5.var\_sin: Variable sine wave; 6.comp\_d: Complex pulse train sounds with pitch contour down; 7.comp\_u: Complex pulse train sounds with pitch contour up. More variance was implemented in the sounds of categories 5.var\_sin, 6.com\_d and 7.comp\_u, than in the other categories. Pitch contour changes were only present in categories 3.pitch\_d, 4.pitch\_u, 6.comp\_d and 7.comp\_u, and formants were only modelled in the categories 6.comp\_d and 7.comp\_u.

call length step. Following this, sounds with specific call lengths were selected (call length fell between 0.07 and 1.96 sec, for details see Table 1). The number of calls in a sound sample depended on the length of the calls, as the complete sound samples were consistently 3 s long and contained only complete calls, meaning that calls starting after 2 s, or calls that would have ended after the 3 s were muted, using Adobe Audition. Therefore, all sound samples consisted of a ~2 s part containing calls and ended with a ~1 s silent part. The intercall interval length was varied in all sound samples. The generated sounds showed variation in loudness, which we included in our analysis as a further acoustic parameter (mean loudness  $79.4 \pm \text{SD } 4.8$  dB).



**Figure 2.** The intensity (scale from 0 to 100) and valence (scale from  $-50$  to  $50$ ) axes of the questionnaire. Image first published in<sup>25</sup>.

We created seven categories of artificial sounds with three levels of complexity. Figure 1 presents the characteristics of each category, while Table 1 shows a summary of the acoustic parameters of the sound samples.

Level 1 sounds (category 1) are based on sine waves in which only the call length and the fundamental frequency were varied. In Level 2 sounds (categories 2, 3, 4, 5) we systematically changed one aspect of the original simple sounds in each category. In category 2 we used pulse train sounds instead of sine waves, in which the consecutive non-sinusoid waves model the vibrations of the vocal folds, creating harmonics<sup>19,44,45</sup>. In categories 3 and 4 we implemented pitch contour changes with either decreasing or increasing pitch, while in the category 5 sounds we included variances in call length, in intercall interval length and in fundamental frequency (see Table 1). Level 3 sounds contained all the previously varied parameters (call length, fundamental frequency, harmonics, variances and pitch contour changes), as well as formants based on vocal tract modelling. The physical parameters of this model were defined as a hypothetical vocal tract for a  $\sim 70$  cm tall social robot. The total number of created stimuli consisted of 588 sound samples, 84 in each category.

**Online questionnaire.** The final questionnaire used in the study is accessible online at <http://soundratingtwo.elte.hu>. First, the participants were asked to provide demographic data on their nationality, gender and age, and were asked to answer the question whether they currently owned a dog at the time of the test or owned one in the past. The online page also provided the instructions for the questionnaire, explaining how to indicate the perceived valence and emotional intensity. The participants were asked to use headphones instead of loudspeakers to minimise the differences in the quality and the frequency range of sound production of built-in loudspeakers (e.g., laptops). Participants also had the opportunity to check if their headphones worked correctly and at an optimal volume by playing a non-relevant sound.

The questionnaire used a modified version of the two-dimensional model of emotions by Russell<sup>46</sup>, which had already been successfully used for measuring the perceived emotions associated with dog and human vocalizations<sup>25</sup>. The questionnaire measured the values the participants gave for the sounds on the valence and intensity axes. We used the same questionnaire design in this study. After a sound was played, the participants had to indicate the valence on a horizontal axis and the intensity on the vertical one with one click (Fig. 2). Due to the high number of sound stimuli, each participant received only 11.9% of the samples (70 sound stimuli) after the 4 demo sounds, and received an equal number of samples (10) from all categories. The samples and their listening order were determined randomly.

**Data analysis.** Statistical analysis was conducted in the R statistical environment.

We excluded responses slower than 20 seconds to avoid artefacts caused by network errors and possible lags in the stimuli presentation. Long response time might also indicate high uncertainty in the answer. We used Linear Mixed Modeling (lmer function from the lme4 package, version 1.1-21<sup>47</sup>) fit with backward elimination (drop1 function) to find the best model. The fixed effects were the fundamental frequency, call length, sound category, gender, age, query language (Hungarian or English), and the participants' status of dog ownership, loudness of sound samples, as well as the two-way interactions of category and acoustic parameters; language, acoustic parameters and complexity category. The participant's age, gender and dog ownership status were included as background variables, as these have been found to influence the perception of emotions in vocalizations in some cases<sup>48–50</sup>. The targets were the intensity and the valence values (respectively), and the random effects were the subjects and the ID of the sounds (see also in Table 2). We used a normal probability distribution with an identity link function and all covariates (fundamental frequency, call length, age, loudness) were scaled and centered. Loudness, and the interaction of loudness and category were included in the model after backward elimination.

		fixed effects	random effects
Intensity	intensity ~ cat + f0 + cl + age + lang + dog + gender + cat:f0 + cat:cl + lang:f0 + lang:cl + cat:lang + loud + cat:loud + (1 testid) + (1 soundid)	cat, f0, cl, age, lang, dog, gender, loud	testid, soundid
Valence	valence ~ cat + f0 + cl + age + lang + dog + gender + cat:f0 + cat:cl + lang:f0 + lang:cl + cat:lang + loud + cat:loud + (1 testid) + (1 soundid)	cat, f0, cl, age, lang, dog, gender, loud	testid, soundid

**Table 2.** The linear mixed models used for statistical analysis. Cat: category, f0: fundamental frequency, cl: call length, age: age of the participant, lang: language of the query (English or Hungarian), dog: participants' dog ownership status, gender: gender of the participant, loud: loudness of sound samples, testid: participant ID, soundid: ID of the sound samples.

	Intensity				Valence			
		Est.	t	p		Est.	t	p
Predictive model (based on 1.sin)	Intercept	41.0812	31.8	<2.2e-16	Intercept	-8.2916	-7.121	<0.001
	f0	9.5927	16.2	<2.2e-16	f0	-5.4035	-7.945	<0.001
					cl	-3.2561	-4.777	<0.001
	r	df	p value	t	r	df	p value	t
1.sin	0.71	1670	<2.2e-16	41.268	0.70	1670	<2.2e-16	40.088
2.pulse	0.49	1690	<2.2e-16	23.345	0.46	1690	<2.2e-16	21.223
3.pitch_d	0.55	1672	<2.2e-16	26.595	0.46	1672	<2.2e-16	21.203
4.pitch_u	0.53	1657	<2.2e-16	25.225	0.53	1657	<2.2e-16	25.191
5.var_sin	0.60	1681	<2.2e-16	30.587	0.58	1681	<2.2e-16	29.015
6.comp_d	0.53	1675	<2.2e-16	25.465	0.47	1675	<2.2e-16	21.986
7.comp_u	0.50	1685	<2.2e-16	23.493	0.48	1685	<2.2e-16	22.491

**Table 3.** Comparison of predicted and actual ratings of valence and intensity. Predictive models are based on a Linear Mixed Effects Model of category 1 (Simple sine wave) sounds. 1.sin: Simple sine wave; 2.pulse: Pulse train; 3.pitch\_d: Sine wave sounds with pitch contour down; 4.pitch\_u: Sine wave sounds with pitch contour up; 5.var\_sin: Variable sine wave; 6.comp\_d: Complex pulse train sounds with pitch contour down; 7.comp\_u: Complex pulse train sounds with pitch contour up, f0: fundamental frequency, cl: call length.

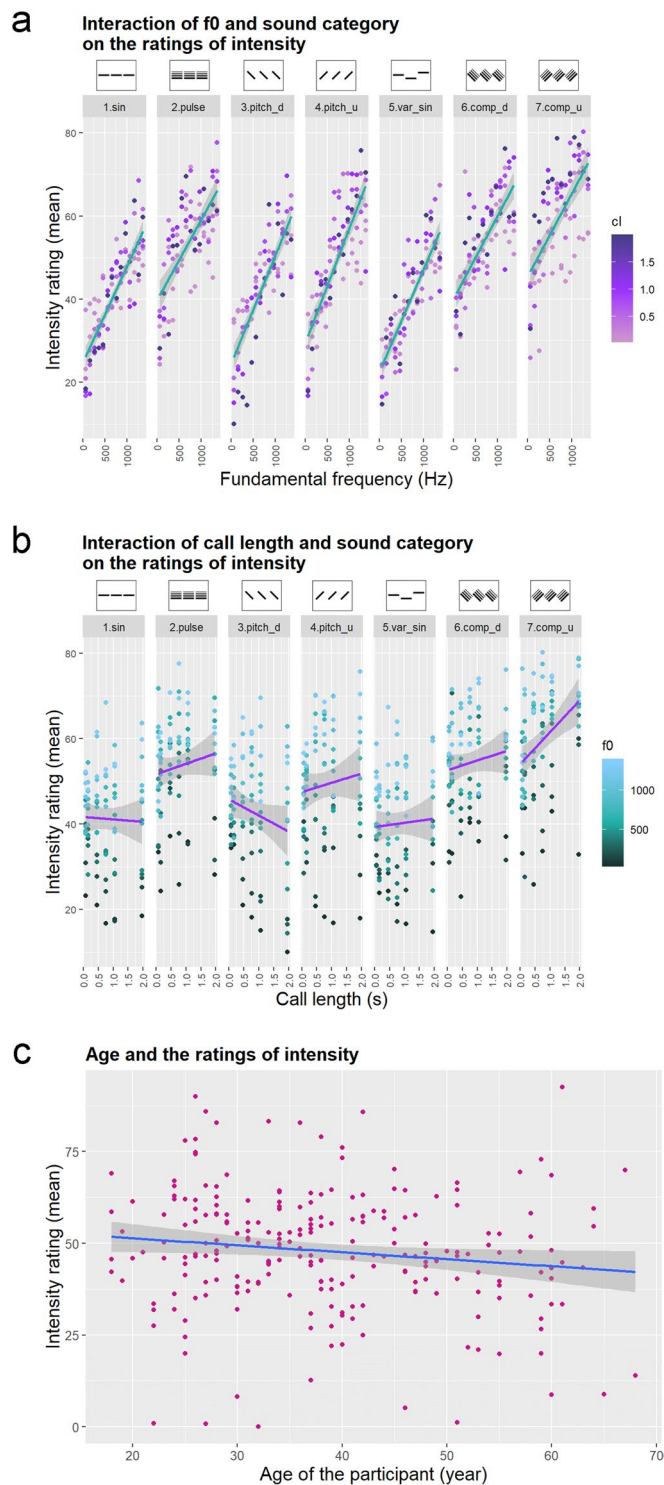
Tukey post-hoc tests (emmeans package, version 1.3.3<sup>51</sup>, emmeans and emtrends functions) were used for pairwise and trend comparisons.

To compare the effects of call length and fundamental frequency in different complexity categories, we created a Linear Mixed Effects Model of category 1 (Simple sine wave), in which the fundamental frequency and the call length were fixed effects, subjects and sound ID were random effects, and the target was the valence or the intensity ratings. We used the models of category 1 to predict the valence and intensity ratings of the other categories. We compared the predicted and actual valence and intensity ratings with Pearson's correlation.

## Results

**Intensity.** We found that the simple Linear Mixed Effects Model fitted on the sinus category sounds predicted the intensity ratings of the other categories quite well based on the correlation between the real and predicted values ( $R = 0.49-0.60$ ). The comparison of the predicted valence and intensity ratings of the categories is in Table 3.

In the Linear Mixed Model, both the fundamental frequency and call length were in interaction with the sound category and the language. According to the post-hoc tests, the fundamental frequency had a similar positive effect on intensity ratings in all categories: the sounds with higher fundamental frequency were rated as more intense, however this effect was stronger in category 4 (Sine wave up), 3 (Sine wave down), 5 (Variable sine wave) and 1 (Simple sine wave) while weaker in 2 (Pulse train), 6 and 7 (Complex pulse train down and up) (Fig. 3a). We see a similar pattern within both the English and the Hungarian responses, although stronger in the former. Call length had a negative effect in categories 1 (Simple sine wave) and 3 (Sine wave down): shorter calls were rated as more intense. In contrast, samples with longer calls were rated more intense in categories 2 (Pulse train), 4 (Sine wave up), 6 and 7 (Complex pulse train down and up) (Fig. 3b). In the Hungarian responses the post-hoc test showed a negative trend (short calls are more intense), compared to the English where the long calls were rated as more intense. The sound category was also in interaction with the language and loudness. In general English speaking respondents in most categories rated the samples as more intense compared to the Hungarian sample with the exception of categories 3 and 4 (Sine wave down and up) where we found no difference. In both languages categories 1 (Simple sine wave), 3 (Sine wave down) and 5 (Variable sine wave) got the lowest ratings, while 2 (Pulse train) the highest. Louder sounds were rated more intense in categories 2 (Pulse train), 7 (Complex pulse train up), 4 (Sine wave up), and 6 (Complex pulse train down). Age had a main effect, as older participants rated sounds as less intensive (Fig. 3c). The participants' gender and dog-owner status had no effect on the intensity rating, thus were excluded from the final model. The results of the Linear Mixed Model are summarized in Table 4, and the post-hoc tests are summarized in Supplementary Tables S1 and S2.



**Figure 3.** (a) The interaction of  $f_0$  and sound category on the ratings of intensity. Colouring of the dots shows the call length. (b) The interaction of call length and sound category on the ratings of intensity. Colouring of the dots shows the fundamental frequency. (c) The effect of the participants' age on the ratings of intensity. Categories in (a) and (b): 1.sin: Simple sine wave; 2.pulse: Pulse train; 3.pitch\_d: Sine wave sounds with pitch contour down; 4.pitch\_u: Sine wave sounds with pitch contour up; 5.var\_sin: Variable sine wave; 6.comp\_d: Complex pulse train sounds with pitch contour down; 7.comp\_u: Complex pulse train sounds with pitch contour up. The dots represent the mean intensity ratings of the sounds, while the grey shaded area around the regression line indicates the confidence interval at 95% confidence level.

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
age	2274	2274.3	1	226.8	4.7418	0.030470	*
cat:f0	47594	7932.3	6	545.7	16.5385	<2.2e-16	***
cat:cl	10459	1743.2	6	547.2	3.6345	0.001515	**
f0:lang	4282	4282.1	1	11444.2	8.9279	0.002814	**
cl:lang	29122	29122.1	1	11447.5	60.7184	7.154e-15	***
cat:lang	26478	4413.0	6	11436.8	9.2009	4.462e-10	***
cat:loud	20891	3481.9	6	558.9	7.2595	1.763e-07	***

**Table 4.** Results of the Linear Mixed Model fit of the intensity ratings. Pr(>F): the p-value belonging to the F statistics. Cat: category, f0: fundamental frequency, cl: call length, age: age of the participant, lang: language of the query, loud: loudness of sound samples.

**Valence.** We found that the simple Linear Mixed Effects Model fitted on the sinus category sounds predicted the valence ratings of the other categories quite well based on the correlation between the real and predicted values ( $R = 0.46\text{--}0.58$ ). The comparison of the predicted valence and intensity ratings of the categories is in Table 3.

The fundamental frequency had a significant main effect in the Linear Mixed Model: samples with lower fundamental frequency were rated to be more positive (Fig. 4a). The post hoc test showed that in the sound category and call length interaction the sound samples that consist of longer calls were rated as having a more negative valence in all categories (Fig. 4b). The interaction of sound category and language showed a significant language effect only within the 2nd (Pulse train) and 3rd (Sine wave down) category: Hungarian responses tended to be more positive in the former and more negative in the latter than English ratings. In both languages category 2 (Pulse train), 6 and 7 (Complex pulse train down and up) were the most negatively rated, while category 4 (Sine wave up) was the most positive. Louder sounds were generally rated as more negative, which effect was steepest in category 2 (Pulse train) and less so in categories 4 and 3 (Sine wave up and down). Age had a main effect: older participants rated the sounds as less negative regardless of the complexity category (Fig. 4c). The gender of the participants and their dog-owner status had no effect on the valence ratings, and neither did the interaction of language and call length, the interaction of language and fundamental frequency, and the interaction of complexity category and fundamental frequency. The results of the Linear Mixed Model are summarized in Table 5, and the post-hoc tests are summarized in Supplementary Tables S3 and S4.

## Discussion

The results show that our artificially generated sounds are able to mimic some of the basic coding rules that are present in animal (mammalian) vocalizations. The predictive models based on sinus sound samples explain quite well both the valence and the intensity ratings in all other complexity categories suggesting the presence of the simple rules. The fundamental frequency of the sounds affects the perceived intensity, that is, sounds with higher fundamental frequency were perceived as more intense, while sounds containing longer calls were rated as more negative across all categories. These results align with the findings of previous research on animal and human vocalizations<sup>14,25,26</sup>.

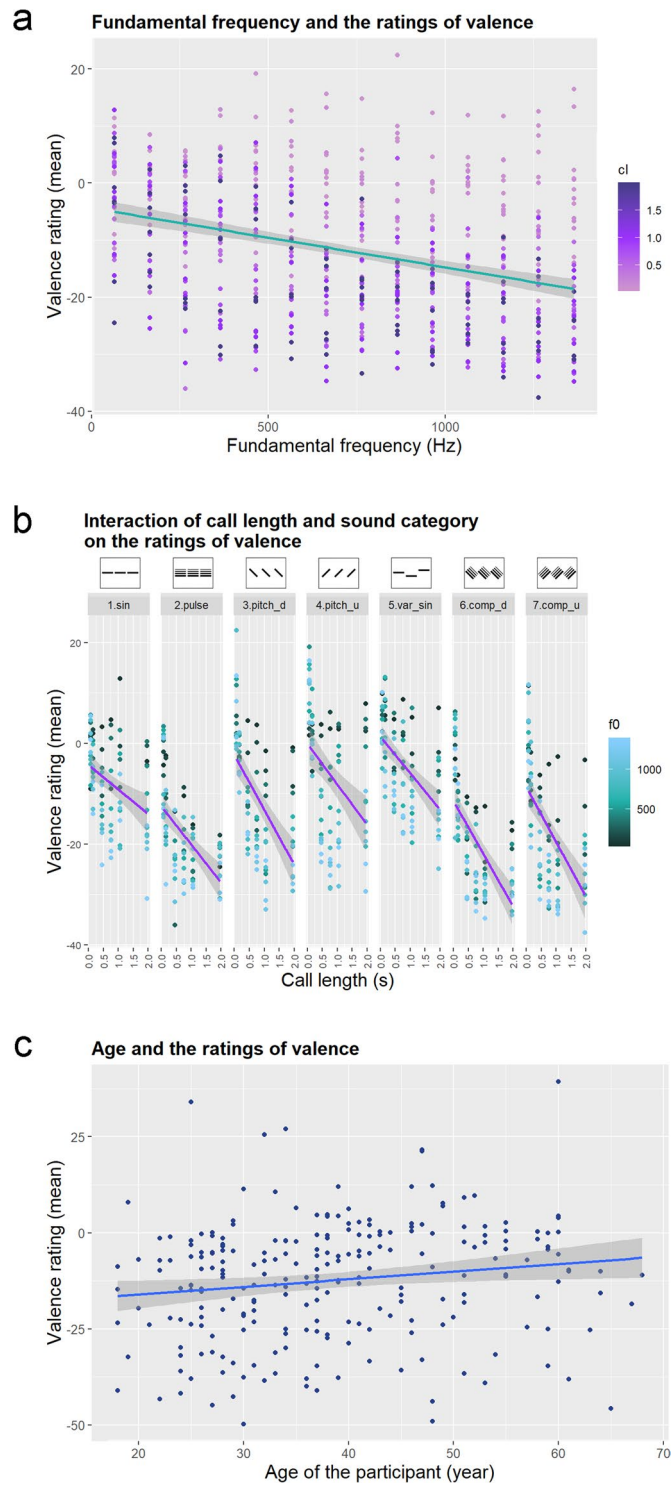
An interesting result was the effect of fundamental frequency on valence: sounds with a higher fundamental frequency were rated as more negative in all categories. Although the fundamental frequency-valence effect was not found by Faragó *et al.*<sup>25</sup> in dog or human vocalizations, the spectral centre of gravity showed a similar pattern in the case of human vocalizations. Multiple other studies also found that higher pitch was associated with negative valence, in e.g., dogs<sup>50</sup>, pigs<sup>26</sup> and wild boars<sup>52</sup>, horses (*Equus caballus*)<sup>53</sup> and bonobos (*Pan paniscus*)<sup>54</sup>. However, high frequency vocalizations in positive contexts can also be found (for a review see<sup>31</sup>), suggesting that the effect of pitch on valence might be non-linear, or can be influenced by other acoustic parameters.

Emotionally expressive vocalizations of terrestrial tetrapods are assumed to have evolved from involuntary sounds emitted due to breathing during aroused emotional states<sup>55</sup>. However, due to the morphological structures and processes of sound production, even simple emotionally expressive vocalizations are acoustically complex, e.g., phonation already appears in frog vocalizations with the appearance of vocal cords, and continues to be present in terrestrial mammals as a result of vocal fold or membrane vibration<sup>56</sup>. As the basic coding rules related to fundamental frequency and call length were also present in the artificial sounds with no added biological features, we can infer that these effects might originate from a more fundamental component of sound processing.

Communicational signals are frequently the result of ritualization, in which a behaviour that carries only involuntary information goes through an evolutionary process in which it becomes specialized and gains a signalling function<sup>57,58</sup>. Ritualization also increases signal complexity, leading e.g., to decreased signal ambiguity or to reproductive isolation via better species recognition<sup>59</sup>. Systematic investigations using generated sounds akin to ours could be used to find common aspects in the ritualized vocal signals of multiple species, aiding in the understanding of how evolutionary pressures affect specific acoustic parameters.

The results also underscore the compatibility of our approach with other SFU methods of emotion expression by showing that the added acoustic parameters did not interfere with the coding rules based on the acoustic cues derived from the call length and fundamental frequency. We found some overall differences in categories with pulse train sounds (categories 2, 6 and 7) as these were generally rated as more intense and more negative than the sounds in sine wave categories (categories 1, 3, 4 and 5). Pulse train sounds can be perceived to be noisier compared to sine wave sounds, which could have resulted in the higher intensity and more negative valence ratings. Furthermore, as pulse train sounds were used to approximate harmonics (category 2) and formants (categories





**Figure 4.** (a) The effect of fundamental frequency on the ratings of valence. Colouring of the dots shows the call length. (b) The interaction of call length and sound category on the ratings of valence. Colouring of the dots shows the fundamental frequency. (c) The effect of the participants' age on the ratings of valence. Categories in (b): 1.sin: Simple sine wave; 2.pulse: Pulse train; 3.pitch\_d: Sine wave sounds with pitch contour down; 4.pitch\_u: Sine wave sounds with pitch contour up; 5.var\_sin: Variable sine wave; 6.comp\_d: Complex pulse train sounds with pitch contour down; 7.comp\_u: Complex pulse train sounds with pitch contour up. The dots represent the mean valence ratings of the sounds, while the grey shaded area around the regression line indicates the confidence interval at 95% confidence level.

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
f0	39154	39154	1	567.0	102.0804	<2.2e-16	***
age	3531	3531	1	220.0	9.2069	0.002701	**
cat:cl	15431	2572	6	569.3	6.7053	7.127e-07	***
cat:lang	21056	3509	6	11348.5	9.1494	5.150e-10	***
cat:loud	31721	5287	6	574.7	13.7834	1.155e-14	***

**Table 5.** Results of the Linear Mixed Model fit of the valence ratings. Pr(>F): the p-value belonging to the F statistics. Cat: category, f0: fundamental frequency, cl: call length, age: age of the participant, lang: language of the query, loud: loudness of sound samples.

6 and 7) of animal and human vocalizations, these might have caused an unintended eeriness, which could have resulted in an uncanny effect (as described in HRI, e.g.<sup>8,9</sup>), near fundamental frequencies that approximate human speech, leading to more intensive and negative ratings.

The call length of the sounds affected the intensity ratings differently in some of the categories, indicating that it does not represent a general coding rule. The effect of call length on intensity was not found in human vocalizations in<sup>25</sup>, only in dogs, which indicates that this association might be species-specific. By including other acoustic parameters in the artificial sounds, further systematic investigations could specify if some rules are species or taxon specific (e.g., in<sup>25</sup> the tonality (harmonic-to-noise ratio, HNR) affected the intensity ratings of only dog vocalizations, as sounds with high HNR were rated as less intense) or if there are other general coding rules based on the added parameters. It could also clarify which parameters can be added to implement further rules with the potential to enrich or refine the range of expressible emotions.

Loudness influenced both the intensity and valence ratings in interaction with the categories: louder sounds were rated as more negative in all categories with varying degrees, while in case of intensity ratings the direction of the effect differed among the categories. As loudness of biological sounds is notoriously hard to measure reliably, especially in field recordings (recording distance and direction highly affects the measurements) this parameter cannot be compared between species, and its role in emotion encoding is uncertain. Although based on physiology and neural control of vocalization we can hypothesise that it can be linked with both higher arousal and negative inner states<sup>31</sup>. Our results partially support this, but it seems that fundamental frequency and call length plays a more crucial role in emotion encoding.

A limitation of the current set of sounds is the low number of sound samples that were rated notably positive. The majority of sounds had a mean rating on the valence axis lower than 0, and only a small number of sounds had a mean rating higher than 20. This presents a problem in the framework of human-robot interactions, as social robots are to exhibit behaviours also associated with positive emotions. However, considering the basis of these sounds, the scarcity of positive valence sounds is not surprising. In animal vocalisations, the expression of positive inner states is less frequent, and their functionality is limited to very specific behaviours or situations, e.g., grooming<sup>60</sup>, greeting<sup>61</sup>, play<sup>62</sup>. Vocalizations of dogs show a similar pattern in their perceived valence in contrast to human non-verbal vocalizations which cover the whole scale<sup>25</sup>. An acoustic parameter which is associated with positive inner states in humans is a steeper spectral slope<sup>63</sup>, which can be incorporated in the next iteration of the artificial sounds.

In some cases, the language of the questionnaire influenced the strength of the effects, and in the call length-intensity connection, its direction. As the effect of the call length on the intensity ratings was only present in interaction with the categories and not as a general rule independent of added acoustic parameters, it can be assumed that a slight difference of interpretation of the word ‘intensity’ by the Hungarian or English speaking participants could have caused this discrepancy. However, this seems to have no major confounding effect in the case of our main questions about the simple encoding rules.

We found that the age of the participant had a significant effect on both the valence and intensity ratings of the sounds, as older participants considered the sound samples to be more positive and less intensive than did younger adults. This could be explained by the neural changes that occur during ageing, which leads to a bias towards positive stimuli found in the elderly (positivity effect), causing increased attention towards<sup>64</sup> and memories of<sup>65</sup> positive stimuli. Elderly people are faster to recognize positive facial expressions than negative ones<sup>66</sup>, while studies have contradictory results on intensity ratings (increased intensity<sup>66</sup>; decreased intensity<sup>67</sup>). Age related hearing loss could have also influenced the answers of elderly participants, as hearing impairment is more prevalent in sounds with higher frequencies, starting from 1000 Hz<sup>68,69</sup>, which could somewhat reduce the effects of higher fundamental frequency on the intensity and valence ratings found on younger adults. However, the associations between the acoustic cues and the intensity and valence ratings persisted, despite the effects of age, and the noise caused by possible sound differences due to the headphone devices of the participants.

As the participants only rated the sounds on their intensity and valence, some functionally important aspects were not investigated. Based on the current results it is not possible to differentiate between sounds with high intensity and negative valence, as they may be perceived as ‘angry’ or ‘fearful’. However, vocalizations perceived as angry/aggressive or fearful/distressed usually elicit opposing behavioural responses from others, as the first may prompt behaviours to avoid the source of the sound, while fearful or distressed vocalizations may elicit approach<sup>70</sup>. This difference in the behavioural response to sounds is instrumental in HRI, and therefore should be investigated as an added dimension to the valence and intensity.

**Outlook.** In the current study, we have established that humans assess the intensity and emotional valence of artificial sounds according to simple coding rules that are based on acoustic cues of animal vocalizations: sounds with higher fundamental frequency are perceived as more intense, while sounds with shorter call lengths are perceived as being more positive. As these coding rules are considered to be shared at least among mammals, the artificial sounds presumably elicit similar responses in non-human mammalian species that live in the human social environment. In our future work, we are planning on investigating the responses of humans and companion animals to the artificially generated sounds, with comparative fMRI studies on humans and dogs and with behavioural tests on humans, dogs and cats. We are also investigating the approach-avoidance responses of humans to the artificial sounds with a follow up questionnaire study.

Defining basic rules of emotion encoding using comparative approach can be the key to understanding the evolutionary processes of animal vocalizations. We suggest that the presented systematic method of assessing the effects of artificial sounds provides a novel opportunity to investigate the evolution of both the production and perception mechanisms underlying vocal emotion expression.

## Data availability

The dataset generated during the current study is available as a supplementary file (Dataset.csv).

Received: 6 August 2019; Accepted: 11 March 2020;

Published online: 27 April 2020

## References

- Mavridis, N. A review of verbal and non-verbal human-robot interactive communication. *Rob. Auton. Syst.* **63**, 22–35 (2015).
- Bennewitz, M., Faber, F., Joho, D. & Behnke, S. Fritz - A Humanoid Communication Robot. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication* 1072–1077, <https://doi.org/10.1109/ROMAN.2007.4415240> (IEEE, 2007).
- Meena, R., Jokinen, K. & Wilcock, G. Integration of gestures and speech in human-robot interaction. 3rd IEEE Int. Conf. Cogn. Infocommunications, CogInfoCom 2012 - Proc. 673–678, <https://doi.org/10.1109/CogInfoCom.2012.6421936> (2012).
- Pellegrino, F., Coupé, C. & Marsico, E. Across-Language Perspective on Speech Information Rate. *Language (Baltim.)* **87**, 539–558 (2012).
- Ekman, P. & Friesen, W. Unmasking the face: A guide to recognizing emotions from facial clues. (ISHK, 2003).
- Miklósi, Á. & Gácsi, M. On the utilization of social animals as a model for social robotics. *Front. Psychol.* **3**, 1–10 (2012).
- Rose, R., Scheutz, M. & Schermerhorn, P. Towards a conceptual and methodological framework for determining robot believability. *Interact. Stud.* **11**, 314–335 (2010).
- Mori, M. The Uncanny Valley. *Energy* **7**, 33–35 (1970).
- Miklósi, Á., Korondi, P., Matellán, V. & Gácsi, M. Ethorobotics: A New Approach to Human-Robot Relationship. *Front. Psychol.* **8**, 1–8 (2017).
- Faragó, T., Miklósi, Á., Korcsok, B., Száraz, J. & Gácsi, M. Social behaviours in dog-owner interactions can serve as a model for designing social robots. *Interact. Stud. Soc. Behav. Commun. Biol. Artif. Syst.* **15**, 143–172 (2014).
- Ekman, P. *et al.* Universals and cultural differences in the judgments of facial expressions of emotion. *J. Pers. Soc. Psychol.* **53**, 712–717 (1987).
- Anikin, A. & Persson, T. Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behav. Res. Methods* **49**, 758–771 (2017).
- Ehret, G. Common rules of communication sound perception. in *Behaviour and Neurodynamics for Auditory Communication* (eds. Kanwal, J. S. & Ehret, G.) 85–114 (Cambridge University Press, 2006).
- Filippi, P. *et al.* Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proc. R. Soc. London B Biol. Sci.* **284**, 1–9 (2017).
- Andics, A. & Faragó, T. Voice Perception Across Species. in *The Oxford Handbook of Voice Perception* (eds. Frühholz, S. & Belin, P.) 362–392, <https://doi.org/10.1093/oxfordhb/9780198743187.013.16> (Oxford University Press, 2018).
- Korcsok, B. *et al.* Biologically inspired emotional expressions for artificial agents. *Front. Psychol.* **9**, 1–17 (2018).
- Gácsi, M., Szakadát, S. & Miklósi, Á. Assistance dogs provide a useful behavioral model to enrich communicative skills of assistance robots. *Front. Psychol.* **4**, 1–11 (2013).
- Fitch, W. T. & Hauser, M. D. Unpacking “Honesty”: Vertebrate Vocal Production and the Evolution of Acoustic Signals. in *Acoustic Communication* 65–137 [https://doi.org/10.1007/0-387-22762-8\\_3](https://doi.org/10.1007/0-387-22762-8_3) (Springer-Verlag, 2003).
- Fant, G. Acoustic Theory of Speech Production (Mouton, The Hague, The Netherlands). 125–128 (1960).
- Scott-Phillips, T. C., Blythe, R. A., Gardner, A. & West, S. A. How do communication systems emerge? *Proc. R. Soc. B Biol. Sci.* **279**, 1943–1949 (2012).
- Zimmermann, E., Lisette, L. & Simone, S. Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices. In *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man* (eds. Eckart, A., Sabine, S. & Elke, Z.) 116–132 (Oxford University Press, 2013).
- Slocombe, K. E. & Zuberbühler, K. Chimpanzees modify recruitment screams. *Pnas* **104**, 17228–17233 (2007).
- Rendall, D. Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *J. Acoust. Soc. Am.* **113**, 3390 (2003).
- Laukka, P. *et al.* Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Front. Psychol.* **4**, 1–8 (2013).
- Faragó, T. *et al.* Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biol. Lett.* **10**, 20130926 (2014).
- Maruščáková, I. L. *et al.* 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 (2015).
- Tallet, C., Špinka, M., Maruščáková, I. & Šimeček, P. Human Perception of Vocalizations of Domestic Piglets and Modulation by Experience With Domestic Pigs (*Sus scrofa*). *J. Comp. Psychol.* **124**, 81–91 (2010).
- Gácsi, M., Vas, J., Topál, J. & Miklósi, Á. Wolves do not join the dance: Sophisticated aggression control by adjusting to human social signals in dogs. *Appl. Anim. Behav. Sci.* **145**, 109–122 (2013).
- Congdon, J. V. *et al.* Hear them roar: A comparison of black-capped chickadee (*Parus atricapillus*) and human (*Homo sapiens*) perception of arousal in vocalizations across all classes of terrestrial vertebrates. *J. Comp. Psychol.*, <https://doi.org/10.1037/com0000187> (2019).
- Andics, A. & Miklósi, Á. Neural processes of vocal social perception: Dog-human comparative fMRI studies. *Neurosci. Biobehav. Rev.* **85**, 54–64 (2018).

31. Briefer, E. F. Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology* **288**, 1–20 (2012).
32. Fischer, J., Metz, M., Cheney, D. L. & Seyfarth, R. M. Baboon responses to graded bark variants. *Anim. Behav.* **61**, 925–931 (2001).
33. Coss, R. G., McCowan, B. & Ramakrishnan, U. Threat-Related Acoustical Differences in Alarm Calls by Wild Bonnet Macaques (*Macaca radiata*) Elicited by Python and Leopard Models. *Ethology* **113**, 352–367 (2007).
34. Andics, A., Gácsi, M., Faragó, T., Kis, A. & Miklósi, Á. Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. *Curr. Biol.* **24**, 574–578 (2014).
35. Belin, P. *et al.* Human cerebral response to animal affective vocalizations. *Proc. R. Soc. B Biol. Sci.* **275**, 473–481 (2008).
36. Breazeal, C. Emotive qualities in robot speech. In *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium* (Cat. No.01CH37180) **3**, 1388–1394 (IEEE, 2003).
37. Gácsi, M. *et al.* Humans attribute emotions to a robot that shows simple behavioural patterns borrowed from dog behaviour. *Comput. Human Behav.* **59**, 411–419 (2016).
38. Yilmazyildiz, S., Read, R., Belpeame, T. & Verhelst, W. Review of Semantic-Free Utterances in Social Human–Robot. *Interaction. Int. J. Hum. Comput. Interact.* **32**, 63–85 (2016).
39. Yilmazyildiz, S., Verhelst, W. & Sahli, H. Gibberish speech as a tool for the study of affective expressiveness for robotic agents. *Multimed. Tools Appl.* **74**, 9959–9982 (2014).
40. Wolfe, H., Peljhan, M. & Visell, Y. Singing Robots: How Embodiment Affects Emotional Responses to Non-linguistic Utterances. *IEEE Trans. Affect. Comput.* **14**, 1–12 (2017).
41. Becker-Asano, C. & Ishiguro, H. Laughter in Social Robotics—no laughing matter. *Intl. Work. Soc. Intell. Des.* 287–300 (2009).
42. Juslin, P. N. & Laukka, P. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychol. Bull.* **129**, 770–814 (2003).
43. Plack, C. J. Auditory Perception. In *Handbook of Cognition* (eds. Lamberts, K. & Goldstone, R. L.) 71–104 (Sage Publications Ltd, 2005).
44. Klatt, D. H. & Klatt, L. C. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* **87**, 820–857 (1990).
45. Titze, I. R. & Martin, D. W. Principles of Voice Production. *J. Acoust. Soc. Am.* **104**, 1148–1148 (1998).
46. Russell, J. A. A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**, 1161–1178 (1980).
47. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67** (2015).
48. Dupuis, K. & Pichora-Fuller, M. K. Aging Affects Identification of Vocal Emotions in Semantically Neutral Sentences. *J. Speech, Lang. Hear. Res.* **58**, 1061–1076 (2015).
49. Bonebright, T. L., Thompson, J. L. & Leger, D. W. Gender stereotypes in the expression and perception of vocal affect. *Sex Roles* **34**, 429–445 (1996).
50. Pongrácz, P., Molnár, C. & Miklósi, Á. Acoustic parameters of dog barks carry emotional information for humans. *Appl. Anim. Behav. Sci.* **100**, 228–240 (2006).
51. Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. Estimated Marginal Means, aka Least-Squares Means. Available at: <https://www.rdocumentation.org/packages/emmeans> (2019).
52. Maigrot, A. L., Hillmann, E. & Briefer, E. F. Encoding of emotional valence in wild boar (*Sus scrofa*) calls. *Animals* **8**, 1–15 (2018).
53. Briefer, E. F. *et al.* Perception of emotional valence in horse whinnies. *Front. Zool.* **14**, 1–12 (2017).
54. Clay, Z., Archbold, J. & Zuberbühler, K. Functional flexibility in wild bonobo vocal behaviour. *PeerJ* **3**, e1124 (2015).
55. Darwin, C. *The expression of the emotions in man and animals.* (John Murray, 1872).
56. Fitch, T. Production of Vocalizations in Mammals. In *Encyclopedia of Language & Linguistics* 115–121, <https://doi.org/10.1016/B0-08-044854-2/00821-X> (Elsevier, 2006).
57. Tinbergen, N. 'Derived' Activities; Their Causation, Biological Significance, Origin, and Emancipation During Evolution. *Q. Rev. Biol.* **27**, 1–32 (1952).
58. Scott, J. L. *et al.* The evolutionary origins of ritualized acoustic signals in caterpillars. *Nat. Commun.* **1**, 1–9 (2010).
59. Cullen, J. M. E. Ritualization of animal activities in relation to phylogeny, speciation and ecology: Reduction of ambiguity through ritualization. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **251**, 363–374 (1966).
60. Sakura, O. Variability in contact calls between troops of Japanese macaques: a possible case of neutral evolution of animal culture. *Anim. Behav.* **38**, 900–902 (1989).
61. Andrew, R. J. The situations that evoke vocalization in primates. *Ann. N. Y. Acad. Sci.* **102**, 296–315 (1962).
62. Pongrácz, P., Molnár, C., Miklósi, Á. & Csányi, V. Human listeners are able to classify dog (*Canis familiaris*) barks recorded in different situations. *J. Comp. Psychol.* **119**, 136–144 (2005).
63. Goudbeek, M. & Scherer, K. Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *J. Acoust. Soc. Am.* **128**, 1322 (2010).
64. Brassens, S., Gamer, M. & Büchel, C. Anterior Cingulate Activation Is Related to a Positivity Bias and Emotional Stability in Successful Aging. *Biol. Psychiatry* **70**, 131–137 (2011).
65. Mather, M. & Carstensen, L. L. Aging and motivated cognition: The positivity effect in attention and memory. *Trends Cogn. Sci.* **9**, 496–502 (2005).
66. Di Domenico, A., Palumbo, R., Mammarella, N. & Fairfield, B. Aging and emotional expressions: Is there a positivity bias during dynamic emotion recognition? *Front. Psychol.* **6**, 1–5 (2015).
67. Phillips, L. H. & Allen, R. Adult aging and the perceived intensity of emotions in faces and stories. *Aging Clin. Exp. Res.* **16**, 190–199 (2004).
68. Gordon-Salant, S. Hearing loss and aging: New research findings and clinical implications. *J. Rehabil. Res. Dev.* **42**, 9–24 (2005).
69. Van Eyken, E., Van Camp, G. & Van Laer, L. The Complexity of Age-Related Hearing Impairment: Contributing Environmental and Genetic Factors. *Audiol. Neurotol.* **12**, 345–358 (2007).
70. Ehret, G. Infant rodent ultrasounds - A gate to the understanding of sound communication. In *Behavior Genetics* **35**, 19–29 (2005).
71. Morton, E. S. On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds. *Am. Nat.* **111**, 855–869 (1977).
72. Riede, T. & Fitch, T. Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). *J. Exp. Biol.* **202**, 2859–67 (1999).

## Acknowledgements

We would like to thank Viktor Devecseri for creating the first version of the online interactive questionnaire, and for providing access to the source code. The research was supported by the ÚNKP-16-3-I. New National Excellence Program of the Ministry of Human Capacities; the Hungarian Academy of Sciences (MTA 01 031); the Premium Postdoctoral Grant (460002) by the Office for Research Groups Attached to Universities and Other Institutions of the Hungarian Academy of Sciences in Hungary; and the National Research, Development, and Innovation Office grant (K115862) and (K120501); and by the National Research, Development and Innovation Fund (TUDFO/51757/2019-ITM, Thematic Excellence Program). The research reported in this paper was

supported by the Higher Education Excellence Program of the Ministry of Human Capacities within the Artificial Intelligence research area of Budapest University of Technology and Economics (BME FIKP-MI).

### Author contributions

M.G., T.F. and B.K. developed the main concept, B.K. and T.F. created the sounds, B.F. created the online questionnaire, B.K. and T.F. analysed the data, B.K., T.F., M.G., Á.M., B.F. and P.K. wrote the manuscript. All authors contributed to the manuscript and approved the final version.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-63504-8>.

**Correspondence** and requests for materials should be addressed to B.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020