Sounds like a fight: Listeners can infer behavioural contexts from spontaneous nonverbal
vocalisations

Roza G. Kamiloglu, Disa A. Sauter

Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

Author Note: Correspondence should be sent to Roza G. Kamiloglu (Orcid ID: 0000-0002-1018-2595), Department of Psychology, University of Amsterdam, REC G, Nieuwe Achtergracht 129 B, PO Box 15900, 1001 NK, Amsterdam, the Netherlands.

Contact: r.g.kamiloglu@uva.nl

Word count: 9005

Abstract

When we hear another person laugh or scream, can we tell the kind of situation they are in – for example, whether they are playing or fighting? Nonverbal expressions are theorised to vary systematically across behavioural contexts. Perceivers might be sensitive to these putative systematic mappings and thereby correctly infer contexts from others' vocalisations. Here, in two pre-registered experiments, we test the prediction that listeners can accurately deduce production contexts (e.g., being tickled, discovering threat) from spontaneous nonverbal vocalisations, like sighs and grunts. In Experiment 1, listeners (total $n = 3120$) matched 200 nonverbal vocalisations to one of 10 contexts using yes/no response options. Using signal detection analysis, we show that listeners were accurate at matching vocalisations to nine of the contexts. In Experiment 2, listeners ($n = 337$) categorised the production contexts by selecting from 10 response options in a forced-choice task. By analysing unbiased hit rates, we show that participants categorised all 10 contexts at better-than-chance levels. Together, these results demonstrate that perceivers can accurately infer contexts from nonverbal vocalisations, suggesting that listeners are sensitive to systematic mappings between acoustic structures in vocalisations and behavioural contexts.

Keywords: behavioural context; evolution; nonverbal communication; vocalisation

When we hear a person scream or laugh, what affective information can we glean from that? One possibility is that we might guess, for instance, that a screaming person is afraid or feeling negative arousal. Such inferences are difficult to confirm since they require an objective way to establish the screaming person's emotional state (Bryant, 2020). In verifying listener's judgements, researchers have commonly asked expresser to indicate their feelings (e.g., "Which emotion did you feel when you were screaming?"), or to produce vocalisations that correspond to a particular emotion (e.g., "make the kind of sound you would make if you were feeling afraid"). Arguably, neither of these approaches allow for objective inferences about the vocaliser (Bryant, 2020). Here, we use an alternative approach that allows for testing objective information regarding the expresser: asking whether listeners are able to infer the specific behavioural context the screaming person is in, like being faced with a threat. This possibility would be consistent with theoretical proposals suggesting that nonverbal expressions primarily function as instrumental actions (Russell, 1994). For example, wrinkling the nose blocks off foul odours (Susskind & Anderson, 2008), and screams are an effective way to grab others' attention (Pisanski et al., 2022). Observers may then interpret nonverbal expressions as responses to particular types of situations (Scarantino, 2017). For instance, we might be able to guess from a person's wrinkled nose that they just smelled a foul odour. Perceivers would thus be expected to be able to infer information about what is happening in the situation that the expresser is in. In the present study, we test this proposal, focusing on the voice, a phylogenetically ancient channel of nonverbal communication (Darwin, 1872; Filippi, 2020; Pisanski et al., 2016). Specifically, we examine whether listeners accurately infer what type of behavioural context real-world nonverbal vocalisations were produced in.

Information inferred from human nonverbal vocalisations

Human listeners can accurately gauge static features like the sex, body size, and physical strength from nonverbal vocalisations, as well as information about transient states, like the emotions and intentions of the vocalising individual (see Pisanski & Bryant, 2019 for a review). Much of what we know about what listeners call infer from vocalisations about transient states comes from studies on the perception of affective information, including discrete emotions and core affect dimensions. Listeners can infer emotions like anger, fear, and interest, as well as arousal and valence dimensions, from decontextualised vocalisations (e.g., Cordaro et al., 2016; Cowen et al., 2019; Laukka et al., 2013; Sauter et al., 2010; Sauter & Fischer, 2018).

One challenge with perceptual judgments of emotions and affective dimensions from nonverbal expressions is that it is difficult to establish a ground truth, that is, what constitutes a ‘correct’ judgment (Bryant, 2020; Russell, 1994). In trying to obtain a ground truth, researchers have validated the emotional state of the expresser by using retroactive self-report (e.g., “Which emotion did you feel when you were laughing?”), capturing the subjective state of the vocalising person (Shiota & Keltner, 2005). Self-report of emotional states and core affect dimensions, however, relies on the expresser’s understanding of emotion/affect labels, rather than merely reflecting the phenomenological contents of the affective experience itself (Frijda, Markam, Sato, & Wiers, 1995). An alternative approach to trying to establish a ground truth is to ask the expresser to imagine a scenario or feeling when they produce a vocalisation to generate performed displays. This approach typically induces low-intensity emotion in the expresser (Laukka, 2004), and are constrained by the expresser’s interpretation of the target emotion terms or the experimenter’s selection of the emotion-eliciting scenarios (Bryant, 2020). Neither self-report judgments nor posed expressions thus provide a straightforward ground truth against which perceptual judgments can be compared.

In order to overcome these concerns, we need to focus on objective information conveyed by the vocalisations. But what objective information relating to the expresser's emotional state could be detected by the perceiver from nonverbal expressions? Theorists have suggested that understanding the information inferred by perceivers from nonverbal expressions requires examination of the multi-componential nature of emotional states (Sauter & Russell, 2020). The perceiver might infer appraisals (i.e., an individual's subjective evaluation of the emotion-eliciting event), subjective experience (e.g., feeling afraid or amused), a physiological reaction (e.g., being aroused), or instrumental behaviours. Theoretical proposals have suggested that among these components, observers might primarily interpret a nonverbal expression as a part of an instrumental action, providing information about the specific situation the expresser is in (Darwin 1872; Eibl-Eibesfeldt, 1973; Fridlund, 1994). Listeners might thus be able to identify what is happening in the situation (Russell, 1994; Scarantino, 2017). For instance, perceivers can infer from raised eyebrows and widened eyes that the expresser is being threatened and therefore alternating their eyes to allow them to see better; listeners may guess from a growl that the expresser is threatening another: "Step back, or I will attack!" Unlike judgements about emotion labels like anger and fear, or subjective evaluations like appraisals, physical situations involve incontrovertible facts that do not depend on the expresser's subjective interpretation. For instance, if we hear someone shrieking, we may infer that they are in a situation in which they are being physically threatened. The judgment that the person is being physically threatened is either correct or incorrect. If listeners are consistently accurate in such judgments, that would demonstrate that perceivers can identify what is happening in the situation in which certain nonverbal vocalisations were produced. What would this tell us?

Form-function mapping in nonverbal vocalisations

One guiding principle in understanding the production and perception of nonverbal vocalisations is that the physical structure of vocal signals (i.e., their *form*) are shaped by their use (i.e., their *function*). Specific vocal forms are thus connected to their outcomes, that is, their biological and social functions. Such form-function mappings have been instructive for understanding vocal communication in nonhuman animals (Endler, 1993; Morton, 1977; Owren & Rendall 2001), as well as in humans (Bryant, 2013; Bryant & Cosmides, 2007). For instance, infants from diverse mammal species, including humans, produce high pitched vocalisations with a pronounced frequency modulation pattern for attracting attention and preventing habituation in listeners (Lingle et al., 2012; Pisanski et al., 2022). Systematicity in associations between forms of vocalisations and specific behavioural contexts in which they are produced might allow listeners to accurately perceive the contexts from vocalisations across societies. For instance, vocal music has been shown to exhibit robust form-function relations which allow listeners to correctly infer which songs are used for different functions, such as soothing babies and healing illnesses (Mehr et al., 2018). Building on that work, we here catalogue a wide range of common behavioural contexts in which nonverbal vocalisations were produced, and investigate the form-function of nonverbal vocalisations by examining if listeners are able to infer the behavioural context from vocalisations.

Elements of behavioural contexts

Which situations are likely to produce functional nonverbal vocalisations? Our environment involves a wide array of information, creating almost unlimited number of aspects which might affect our behaviour. An effective method for understanding situations is to look at basic elements which constitute a situation (Neel, Brown, & Sng, 2020).

Rauthmann, Sherman, and Funder (2015) have proposed that information in situations can be conveyed at three levels: cues, characteristics, and classes. *Cues* include physical stimuli like nonverbal vocalisations in the immediate environment. Cues are interpreted by perceivers, a

process which yields *situation characteristics* (i.e., a psychological meaning of situations). For instance, the cue of laughter present in an environment may signify a particular function, namely an intention to affiliate with another person (Wood, Martin, & Niedenthal, 2017). Together, similar cues and levels of situation characteristics constitute *situation classes*. For instance, a ‘play’ situation would contain similar cues (e.g., laughs) and situation characteristics (e.g., facilitating mutually beneficial interactions).

Identifying specific behavioural contexts in which nonverbal vocalisations are produced requires an analysis of potential situation classes. One way to catalogue such situations is to leverage existing work on non-human animal behaviour (Neel, Brown, & Sng, 2020). This work points to situations that are likely to produce functional behaviour across species. Here, we examined ten situation classes that have been shown to serve specific biological and social functions in other primates and mammals: Being attacked by another person, being refused access to food, being separated from mother, being tickled, discovering a large food source, discovering something threatening, eating high value (strongly preferred) food, eating dispreferred food, having sex, threatening an aggressive person or people.

Perception of behavioural contexts from nonverbal vocalisations

Prior research on human listeners’ perception of behavioural contexts from nonverbal vocalisations are limited to investigations of vocalisations produced by infants and non-human animals. This research has shown that listeners can match infant vocalisations to production contexts like requesting food and giving an object (Kersken, Zuberbühler, & Gomez, 2017), and parents can infer contexts like interaction with the caregiver (play) and satisfaction after feeding from vocalisations of infants (Lindová, Špinka, & Nováková, 2015). Human listeners can also accurately infer situational information from vocalisations of other species, including domestic piglets (Tallet et al., 2010), dogs (Pongrácz, Molnár, Miklósi & Csányi, 2005; Silva et al., 2021), cats (Nicastro & Owren, 2003), macaques (Linnankoski et

al., 1994), and chimpanzees (Kamiloğlu, Slocombe, Haun & Sauter, 2020). For instance, listeners can accurately judge from barks whether dogs were alone in a park, playing with their owner, or preparing to go for a walk (Pongrácz et al., 2005). Humans thus seem to correctly infer the context of vocalisations from human infants and various species, suggesting the presence of an acoustic form-function relationship in nonverbal vocalisations.

Research on the perception of adult human nonverbal vocalisations has tested listeners' ability to gauge dimensions like valence and intensity from vocalisations produced in different contexts (Anikin, Pisanski, K., & Reby., 2020; Atias et al., 2018), and whether listeners can infer the sex of the expresser and the outcome of a contest (winning vs. losing a match) from grunts produced during tennis matches (Raine et al., 2017). However, it is not known to what extent people can accurately perceive the behavioural context itself from nonverbal vocalisations. Investigating this idea allows us to test verifiable inferences of richer information regarding the production context from nonverbal vocalisations, rather than subjectively assessed general emotion states or core affect dimensions. Addressing this question will help elucidate whether there are perceptual mechanisms through which listeners infer instrumental behaviours from nonverbal vocalisations, and may point to form-function relationships in a wide range of human nonverbal vocalisations.

We also sought to test the prediction that accuracy would be higher for vocalisations produced in negative, as compared to positive, contexts. This prediction draws on accounts emphasising the evolutionary significance of negative information in ancestral environments (Cacioppo & Gardner, 1999). Failing to infer information from vocalisations that were produced in a negative context may be particularly costly for the perceiver, and survival chances might thus have been helped by the ability to infer contextual information in negative contexts in particular (Nesse, 1990). Previous research has consistently found that perceivers are more accurate at inferring affective information (e.g., valence and arousal)

from vocalisations produced in negative, as compared to positive contexts (Filippi et al., 2017; Scheumann, Hasting, Zimmermann, & Kotz, 2017). Moreover, recognising emotions from vocalisations of negative emotions requires less auditory information as compared to vocalisations of positive emotions (Pell & Kotz, 2011). We therefore predicted that listeners would be more accurate in inferring behavioural contexts from vocalisations produced in negative, as compared to positive, contexts.

The present study

In two experiments, we tested the prediction that listeners accurately infer in what behavioural context a vocalisation was produced. In Experiment 1, listeners were asked to decide whether a vocalisation was produced in a particular context or not in a yes/no match-to context task. We predicted that participants would be able to match vocalisations to corresponding behavioural contexts at better-than-chance levels. This simple yes/no task offers a feasible way to conduct an initial test of mappings between vocalisations and behavioural contexts. It does not, however, allow comparisons across different contexts (see Russell, 1994). In Experiment 2, participants were therefore asked to make judgments about vocalisations by selecting from 10 context categories on each trial. We predicted that listeners would be able to categorise all behavioural contexts at better-than-chance levels. In both experiments, we also tested the prediction that performance would be better for vocalisations produced in negative as compared to positive contexts. The hypotheses, methods, and data analysis plan were pre-registered on the Open Science Framework before data collection was commenced (Experiment 1: <https://osf.io/v87fg/>; Experiment 2: <https://osf.io/sefdg>).

Experiment 1: Behavioural context matching task

We employed a modified version of the match-to-context task of Kamiloğlu et al. (2020). Each participant was asked to match vocalisations to a single context (“Does this vocalisation match Context A?”) with yes/no response options on each trial.

Participants

A total of 3120 participants (1648 women, 1440 men, 20 other, 12 preferred not to say; $M_{\text{age}} = 29.32$ years, $Sd_{\text{age}} = 6.82$, range = 16-42 years old) were recruited through the online data collection platform Cint (<https://www.cint.com>). Sample size was predetermined by a power analysis (G*Power 3.1; Faul et al., 2007) based on a t-test given power = 0.80, $d = 0.2$ (Cohen’s d : small effect size), and $\alpha = 0.05$. This revealed that 156 judgments per behavioural context would be needed to detect a small effect size. Each participant was given a context matching task concerning only one context (“Does this vocalisation match Context A?”) with yes/no response options on each trial. To avoid possible learning effects during the context detection task, each participant was displayed half of the vocalisations from the matching context category. This constitutes one fourth of the vocalisations in each context condition; the other three fourths were from the non-matching categories (see Experimental Procedure). This yielded a total number of 312 participants per context category. Given that we sought to test 10 behavioural contexts, the total sample size was thus set to 3120. All participants reported having normal hearing. The experimental session lasted around 15 minutes, and participation was compensated with a monetary reward. All participants provided digital informed consent before participation and were free to stop at any point during the experiment.

Materials and Procedure

Human nonverbal vocalisations. The spontaneous nonverbal vocalisations were collected from www.youtube.com and other online websites. Twenty videos were selected from each of the following behavioural contexts: Eating high value (strongly preferred) food, eating

dispreferred food, having sex, discovering a large food source, being tickled, being separated from mother, being refused access to food, being attacked by another person, threatening an aggressive person or people, and discovering something threatening. The first five contexts were considered positive, the other five negative. The behavioural contexts, as well as the valence classifications, were determined by experts in a previous study, based on observations of chimpanzee behaviour, including social exchanges between individuals (see Kamiloğlu et al., 2020). In addition to these findings, vocalisations produced in alarm, food-associated, and specific social contexts have been shown to serve specific functions in other primates, mammals, and birds (Clay, Arhbold, & Zuberbühler, 2015; Price et al., 2015; Scarantino & Clay, 2015; Townsend & Manser, 2013). The behavioural contexts were thus selected to sample a wide range of contexts necessarily entailing different possible functions.

Three naive research assistants searched for videos of the behavioural contexts using relevant search terms. Inclusion of vocalisations was based exclusively on 1) the eliciting situation matching the target behavioural context; 2) the presence of a nonverbal vocalisation that was clearly audible; and 3) only one person vocalizing. The research assistants also considered suddenness, clarity, and certainty (see Anikin & Persson, 2017): Sudden events offer minimal time for conscious posing or impression management, clear (unambiguous) situations minimize the risk of misunderstanding the target context, and the assistants selected videos that they were maximally certain reflected the target contexts.

A total of 200 vocalisations were collected. Representative vocalisations for each context can be found at <https://emotionwaves.github.io/BehaviouralContexts/>. For the practice trials, two vocalisations produced during a friendly interaction context were taken from www.findsounds.com. Recordings were digitalized at a 44 kHz sampling rate (16 bit, mono) and normalized for peak amplitude using Audacity (<http://audacity.sourceforge.net>). The acoustic characteristics of the recordings were extracted using Praat (Boersma &

Perception of Behavioral Contexts

Weenink, 2017); they are illustrated in Figure 1 and can be found in Table 1. Details of the recordings can be found in Supplementary Materials Table 1S.

Figure 1. Acoustic characteristics of nonverbal vocalisations produced in 10 behavioural contexts. Acoustic measures are broken down by speaker sex. Min. = minimum, Max. = maximum, SCoG = Spectral center of gravity, dB = decibel, Hz = Hertz, sec = seconds.

Table 1. *Acoustic characteristics of vocalisations as means per behavioural context with standard deviations in brackets*

Behavioural context	Duration	Intensity	Pitch Mean	Pitch Sd	Pitch Min	Pitch Max	SCoG Mean	SCoG Sd
Being attacked by another person	1.39 (0.63)	78.76 (3.49)	314.63 (65.34)	84.77 (39.34)	157.98 (83.80)	443.94 (68.39)	1106.73 (382.41)	765.08 (367.54)
Being refused access to food	2.17 (2.20)	77.80 (3.83)	311.51 (68.47)	69.67 (33.94)	150.54 (85.39)	426.63 (106.93)	1756.13 (1059.63)	1482.46 (761.99)
Being separated from mother	2.27 (0.76)	77.48 (2.40)	355.23 (49.60)	78.88 (20.52)	193.05 (62.61)	493.43 (55.30)	1820.51 (521.60)	1223.73 (391.97)
Being tickled	2.20 (1.33)	79.88 (3.77)	349.84 (45.73)	77.97 (32.21)	183.79 (84.72)	473.72 (36.26)	1604.90 (526.47)	861.51 (378.72)
Discovering a large food source	1.34 (1.06)	79.54 (4.80)	265.97 (102.60)	53.27 (33.38)	152.93 (98.88)	342.36 (128.17)	1240.08 (1426.76)	1157.13 (1504.70)
Discovering something threatening	1.17 (1.32)	82.61 (2.71)	344.75 (88.11)	62.63 (44.51)	235.37 (110.67)	453.30 (61.40)	1445.75 (322.43)	758.78 (355.17)
Eating high value (strongly preferred) food	1.04 (0.33)	78.54 (2.60)	193.83 (57.11)	58.99 (33.23)	110.19 (38.91)	306.49 (116.38)	494.99 (188.99)	619.05 (285.86)
Eating dispreferred food	0.92 (0.17)	79.70 (1.52)	189.62 (78.79)	37.86 (36.04)	131.07 (60.07)	253.61 (122.60)	425.78 (130.55)	611.02 (294.58)
Having sex	3.89 (4.68)	75.68 (4.75)	242.85 (90.21)	35.48 (33.77)	175.64 (89.29)	305.21 (94.39)	1302.06 (1464.05)	1356.65 (1279.13)

Perception of Behavioral Contexts

Threatening an aggressive person or people	1.24 (0.86)	78.18 (3.41)	298.90 (82.72)	64.76 (25.98)	175.34 (96.34)	391.70 (81.41)	1129.02 (684.97)	938.86 (627.98)
--	-------------	--------------	----------------	---------------	----------------	----------------	------------------	-----------------

Note. Min. = minimum, Max. = maximum, SCoG = Spectral center of gravity.

Experimental procedure. The study was run online using the Qualtrics survey tool (Provo, UT). Participants were instructed to complete the experiment in a silent environment and to use headphones during the entire experiment. Before the main experiment, we presented two screening questions in which participants were played a bird sound and a car horn, and were asked to indicate what they heard, with ‘bird tweet’ and ‘car horn’ as response options. The screening questions were used to make sure that participants were paying attention and listening to the stimuli. Only participants who answered both of the screening questions correctly could continue with the experiment. After the screening trials, participants completed two practice trials in which they were asked to indicate whether each of two vocalisations matched the context ‘friendly interaction with others’ by selecting “yes” or “no”. The ‘friendly interaction with others’ context was specific to the practice trials and was not included in the main experiment. During the practice trials, participants were asked to set a comfortable sound level and to then keep it constant for the rest of the study.

After completing the two practice trials, each participant was randomly assigned to one of ten conditions, with each condition focused on one behavioural context. In each condition, participants were asked to give a forced-choice yes/no judgment for each vocalisation, indicating whether they thought it was produced in the target context or not. In total, each participant heard 40 vocalisations. To reduce the risk of learning effects, only a quarter of the stimuli heard by a given participant were from the matching behavioural context. These were a randomly selected subset of the vocalisations from that behavioural context. The other three-fourths of the stimuli were randomly drawn from all of the non-matching behavioural contexts.

The presentation order of vocalisations was randomized for each participant. There were no time constraints, and participants were able to replay each sound as many times as needed to make a judgment for each vocalisation.

Statistical Analyses

Outlier detection and statistical analysis were done following the pre-registered plan. A total of thirty-five participants were determined to be outliers because their performance was 3 Sd or more below the context-specific mean recognition level. The number of outliers and total number of participants per behavioural context are reported in Table 2. Their data was consequently excluded from the dataset before data analysis commenced.

Table 2. *Number of participants who scored 3 Sd or more below the mean for each behavioural context (n), and number of listeners assigned to each context (N) in Experiment 1*

Behavioural context	N	n
Being attacked by another person	317	1
Being refused access to food	300	0
Being separated from mother	305	9
Being tickled	294	6
Discovering a large food source	311	0
Discovering something threatening	310	8
Eating high value (strongly preferred) food	304	5
Eating dispreferred food	315	0
Having sex	305	6
Threatening an aggressive person or people	324	0

Note. N is the number of participants whose data are included in the analysis, and n is the number of participants who scored 3 Sd or more below the mean for each behavioural context (i.e., outliers).

We quantified participants' ability to match vocalisations to behavioural contexts using the sensitivity index d-prime, which controls for individual biases in the use of a particular

response option. D-prime is calculated as z-transformed hit rates minus false alarm rates (Macmillan & Creelman, 2004). Hit rates were calculated as the proportion of Yes trials to which participants responded Yes, and false alarm rates as the proportion of No trials responded to as Yes. Hit and false alarm rates with extreme values (i.e., 0 or 1) were adjusted by replacing rates of zero with $0.5/n$ ($0.5/m$) and rates of 1 with $(n-0.5)/n$ ($[m-0.5]/m$), where n (m) is the number of signal (noise) trials (Macmillan & Kaplan, 1985). Kolmogorov-Smirnov tests indicated non-normality in the distribution of d-prime scores for all behavioural contexts ($ps < 0.001$). In order to test whether human listeners performed better than chance in matching vocalisations to behavioural contexts, d-prime scores for each context were therefore tested against chance (random guessing denoted by a d-prime score of zero) using separate Wilcoxon Signed Rank tests. We employed Bonferroni corrected α levels ($\alpha = .005$). Kolmogorov-Smirnov tests also indicated non-normality in the distribution of d-prime scores for positive and negative contexts ($ps < 0.001$). In order to test whether listeners performed better for negative than for positive behavioural contexts, Mann-Whitney¹ test was used to compare the mean accuracy (d-prime scores) for negative vs. positive contexts.

Results and Discussion

The results showed that participants were able to accurately match all types of vocalisations to the correct behavioural contexts ($ps < 0.001$), except for vocalisations produced while eating dispreferred food (see Table 3). Specifically, participants were able to match vocalisations to being attacked by another person, being refused access to food, being separated from mother, being tickled, discovering a large food source, discovering something threatening, eating high value (strongly preferred) food, having sex, and threatening an

¹ Our pre-registration refers to ANOVA assuming that the data is normally distributed. We instead used Mann-Whitney test, because the data followed a non-normal distribution.

aggressive person or people. Distribution of the data per behavioural context are shown in Figure 3A.

Table 3. *D-prime scores indicating participants' performance in matching vocalisations to behavioural contexts, tested against chance level*

Behavioural context	Mean (Sd)	Mdn	<i>n</i>	<i>Z</i>	<i>r</i>
Being attacked by another person	1.01 (0.83)	1.03	317	15.08*	0.85
Being refused access to food	0.58 (0.52)	0.59	300	13.18*	0.76
Being separated from mother	2.53 (0.83)	2.76	305	15.10*	0.86
Being tickled	2.55 (0.85)	2.76	294	14.73*	0.86
Discovering a large food source	0.54 (0.70)	0.51	311	11.16*	0.63
Discovering something threatening	1.66 (0.63)	1.81	310	15.16*	0.86
Eating preferred food	1.90 (0.87)	2.05	304	14.99*	0.86
Eating dispreferred food	0.10 (1.05)	0.00	315	0.66	0.04
Having sex	2.11 (0.69)	2.12	305	15.12*	0.87
Threatening an aggressive person or people	1.09 (0.79)	1.05	324	14.81*	0.82

Note. * significance level, $p < 0.001$ (Wilcoxon non-parametric test). *n* indicates the number of participants who listened to a particular behavioural context as target category, *r* reflects effect size.

We found that accuracy was not significantly different for vocalisations produced in negative, as compared to positive, contexts (negative: $n = 1547$, $M = 1.44$, $Sd = 0.91$, $Mdn = 1.50$; positive: $n = 1538$, $M = 1.35$, $Sd = 1.3$, $Mdn = 1.22$; $W = 1234000$, $p = 0.07$).

Distribution of the data for positive and negative contexts can be found in Figure 4A.

The results from Experiment 1 show that listeners were able to match nonverbal vocalisations to nine out of ten contexts better than would be expected by chance level. There was no difference in performance accuracy between negative and positive contexts. Listeners failed to accurately match vocalisations to a single context: eating dispreferred food. Not only was performance low, but it was also more variable for this context as compared to all of the others (see Table 3). We used the term “dispreferred” to refer to types of food that are not strongly liked, but not disliked. We suspect that the description of this context may have been ambiguous to participants; some participants may have interpreted it as referring to eating low-preference foods (e.g., cucumber), while others may have thought it meant eating disliked foods (e.g., insects). In Experiment 2, we therefore provided a clarification of the description of this context.

Experiment 2: Categorisation of behavioural contexts from 10 categories

In Experiment 1, listeners were asked to judge whether a nonverbal vocalisation was produced in a particular behavioural context, using a simple yes/no match-to-context task. In Experiment 2, we tested whether listeners could accurately categorise the behavioural context in which vocalisations were produced by selecting from 10 context categories. The second experiment allowed us to 1) test whether results of the Experiment 1 would replicate using a different kind of task, and 2) examine which contexts were confused by participants.

Methods

Participants

Using G*Power (Faul et al., 2009), assuming $d = 0.2$ (Cohen’s d : small effect size), a sample of $N = 337$ was predicted a priori to have power of .80, at $\alpha = 0.005$ (Bonferroni-corrected alpha level based on the context identification task including the recognition of 10 context categories, thus $0.05/10$). Consequently, a total of 337 participants were recruited (170 women, 143 men, 12 non-binary, 1 other, 11 prefer not to say; $M_{age} = 20.26$ years, Sd_{age}

= 3.02; range= 17–38 years old). All participants reported having no hearing impairments or issues. Participants were recruited via the research pool. Participation was compensated with course credit.

Materials and Procedure

Human nonverbal vocalisations. Experiment 2 employed the same stimuli as Experiment 1, both for the screening questions and the main experiment. Two vocalisations were used in the practice trials, taken from www.youtube.com which allowed us to obtain vocalisations produced in the ‘being tickled’ and ‘discovering threat’ contexts. These vocalisations were not used in the main experiment.

Experimental procedure. The study was implemented in Qualtrics. Participants were instructed to complete the experiment in a silent environment and to use headphones during the entire experiment. Before the main experiment, participants were presented with two screening questions, which were the same as in Experiment 1. Participants who answered both screening questions correctly could continue with the practice trials. In the practice trials, participants were asked to indicate the context in which they thought each vocalisation was produced, by selecting from the 10 behavioural context categories. During the practice trials, participants were asked to adjust the sound to a comfortable level and to keep it constant for the rest of the experiment.

For each of the 200 vocalisations, participants were asked to make a forced-choice context categorisation judgement, selecting from the 10 behavioural context categories (“Please select the context in which you think this vocalisation was produced”). The presentation order of vocalisations was randomized for each participant; the context categories appeared in alphabetical order for all participants. There were no time constraints on completing each trial, and participants were able to replay stimuli as many times as they want. Given that the forced-choice context categorisation task is rather repetitive, we sought

to reduce attrition by presenting a soundtrack quiz question after every 20 trials and including dynamic background images that changed after each quiz question. The quiz questions consisted of brief segments of soundtracks (maximally 5 seconds long) from popular arcade games, movies, and cartoons. Participants were asked to listen to the sound clip and to guess which game/movie/cartoon it was taken from using a forced-choice format. We provided immediate feedback for each quiz question, followed by information on how much of the experiment (in percentages) the participant had completed so far.

In Experiment 1, listeners failed to match vocalisations to one of the contexts, *eating dispreferred food*, potentially due to the ambiguity of the label of this context. In order to maximize clarity about the meaning of this context, we included a description of this context in the practice trials, and added a mouse-over description for this category in the main task. In the practice trials, we presented the following description: “this means food that is not particularly well liked, but not food that is disliked”. In the main task, when the cursor was on the eating low-preference food category, a tooltip box was displayed: “refers to types of food that are not strongly liked, but not disliked”.

Statistical Analysis

The analyses followed our pre-registered plan for outlier detection and exclusion, as well as statistical analyses. The data was checked for outlier detection. Specifically, participants whose performance was 3 Sd or more below the mean on the task (in terms of percent correct responses) were excluded, since they were considered outliers. This resulted in exclusion of 10 participants.

The proportion of correct responses was calculated for each participant for each context category, in order to test whether listeners would perform at better-than-chance levels in the recognition of each behavioural context. In order to control for individual biases in the use of particular context categories as response categories, unbiased hit rates (Hu scores, see

Wagner, 1993) were also calculated. Hu scores yield an unbiased recognition score for each participant per context category. Since Hu scores are proportional, they are arcsine-transformed for statistical tests. We tested the normality of the distribution of arcsine-transformed Hu scores using Kolmogorov-Smirnov test. The distribution was normal for the being attacked, being refused, and having sex contexts ($ps > 0.05$), and non-normal for the remaining contexts ($ps < 0.05$). In order to test whether performance was better than chance (random guessing, here 1/10 as there were 10 response options), we used one-sample t-tests for contexts with normal distribution, and one-sample Wilcoxon signed-rank tests for the contexts with non-normal distribution. Bonferroni corrections were applied to correct for multiple tests ($0.05/10$). This analysis was conducted to test the prediction that judgements of each context would be significantly better ($ps < 0.005$, Bonferroni corrected) than the chance level.

We also sought to test whether recognition would be more accurate for negative as compared to positive contexts. To test this prediction, the transformed Hu scores for the recognition of negative contexts were compared with the transformed Hu scores for positive contexts. We used paired samples t-test, because data followed a normal distribution based on a Kolmogorov-Smirnov test ($p = 0.200$).

Results and Discussion

Listeners were able to recognize all behavioural contexts from vocalisations at better-than-chance levels ($ps < 0.001$; see Table 4). Figure 2 shows a confusion matrix for the average recognition accuracies per context, and Figure 3B presents distribution of the data (Hu scores) per behavioural context as raincloud plots. The results support our prediction that all ten behavioural contexts would be accurately recognised from vocalisations.

Table 4. *Hu scores indicating listeners' performance in categorising contexts, tested against chance level*

Perception of Behavioral Contexts

Behavioural context	Mean (Sd)	Mdn	Z	r
*Being attacked by another person	0.45 (0.11)	0.45	59.60	3.30
*Being refused access to food	0.30 (0.13)	0.30	27.75	1.54
Being separated from mother	0.81 (0.17)	0.82	12.80	1.00
Being tickled	1.03 (0.12)	1.03	16.81	1.00
Discovering a large food source	0.49 (0.13)	0.49	6.99	1.00
Discovering something threatening	0.47 (0.24)	0.49	6.72	0.96
Eating preferred food	0.66 (0.09)	0.67	10.04	1.00
Eating dispreferred food	0.31 (0.16)	0.30	3.81	0.96
*Having sex	0.90 (0.11)	0.89	135.53	7.50
Threatening an aggressive person or people	0.65 (0.13)	0.66	9.87	1.00

Note. * one sample t-test was used for testing against chance level given normality in the distribution of the data. For these contexts, the Z column represents *t* values and *r* column represents Cohen's *d* effect size. Wilcoxon-signed rank test was used for the other contexts in order to test the performance against chance level.

Perception of Behavioral Contexts

Figure 2. Heatmap of confusion matrix (%) for behavioural context categorisation.

When recognition accuracy of negative and positive contexts were compared, we found that contrary to prediction, listeners performed better in the recognition of positive ($M = 0.70$, $Sd = 0.05$) as compared to negative contexts ($M = 0.56$, $Sd = 0.08$; $t = 31.19$, $d = 1.73$, $p < 0.001$). Raincloud plots for positive and negative contexts are shown in Figure 4B.

Experiment 2 revealed that listeners accurately inferred all 10 behavioural contexts from vocalisations produced in these contexts, and performance was better for positive compared to negative contexts. Recognition percentages show high variability between the behavioural contexts, and range from 23.55 % for being refused access to food to 88.69 % for having sex (see Figure 3). Listeners performed particularly well in two contexts, being tickled (89.37 %) and having sex, which might explain the better recognition of positive as compared to negative contexts. Laughter produced while being tickled and moans produced while having sex might be highly distinct vocalisation types. These types of vocalisations were rarely confused with the other contexts in this study; whether they are acoustic distinct from a wider range of vocalisations is an open question. Contexts with lower recognition rates were often systematically confused with another context. Most commonly, listeners confused

being refused with being separated, being attacked with threatening another, and eating dispreferred food with eating preferred food context (see Figure 3).

Figure 3. Raincloud graph for (A) d' prime scores (Experiment 1), (B) arcsine Hu scores (Experiment 2). Raincloud graphs combine split-violin plots and boxplots: split-violin plots show the data distribution, and boxplots display the distribution of the data, their full range

from minimum to maximum, the 25–75% range (box), and the median (middle of the box). Horizontal dashed lines indicate the chance level for each experiment. In both experiments, listeners performed better than chance for all contexts, with the exception of eating dispreferred food in Experiment 1 (A). Code adapted from Allen et al. (2019).

Figure 4. Raincloud graph for (A) d prime scores (Experiment 1), (B) arcsine Hu scores (Experiment 2). There was no difference in d scores between positive and negative contexts in Experiment 1 (A). Listeners were better at recognizing positive contexts compared to negative contexts in Experiment 2 (B).

General Discussion

Inference of behavioural contexts from nonverbal vocalisations

In two pre-registered experiments, we tested whether human listeners can accurately infer behavioural contexts from spontaneous nonverbal vocalisations, using a yes/no match-to-context task (Experiment 1) and a 10-way forced-choice context categorisation task (Experiment 2). The results show that listeners can accurately map nonverbal vocalisations to a wide range of behavioural contexts. For instance, whimpers were correctly perceived to indicate separation, while shrieks were accurately judged to reflect the discovery of a threat. The results across the two different kinds of tasks were remarkably consistent (see Figure 3).

The current study takes an alternative approach to the study of nonverbal vocalisations than previous research, by moving the focus away from the judgements of subjective states. Typically, perceptual judgments of emotional vocalisations are compared to expressers' retroactive self-reports, which involves considerable subjectivity in determining the emotional state of the expresser (Shiota & Keltner, 2005). Alternatively, judgments are made of posed vocalisations, which might involve different structural regularities in emotional vocalisations as compared to spontaneously produced vocalisations (Bryant, 2020). Here, we provide evidence that listeners can infer a wide variety of behavioural contexts from vocalisations occurring in real life (e.g., fighting, tickling). The present study thus provides an objective examination of nonverbal vocal behaviour, and demonstrates for the first time the presence of identifiable associations between a range of human adult vocalisations and specific situational features.

The contexts employed in the present study were situations that most people encounter in their own lives. It is thus possible that participants' associations between nonverbal vocalisations and contexts were based on individual learning. However, recent research has shown that naive human listeners are also able to match chimpanzee vocalisations to similar contexts to those used here (Kamiloğlu et al., 2020). Given that listeners did not have much familiarity with chimpanzees, it is unlikely that they had learned about associations between chimpanzee vocalisations and specific behavioural contexts. Instead, human listeners' ability to infer specific behavioural contexts from both heterospecific and conspecific vocalisations likely reflect capacities to pick up on phylogenetically conserved acoustic regularities. In many vertebrate species, low pitched, harsh vocalisations with nonlinearities grab perceivers' attention and function to signal hostility towards others (e.g., Ladich, 1989; Mueller, 1971; Wagner, 1989). Such growl-like vocalisations also function as a signal of potential aggressive acts in humans (Tsai et al., 2010). Direct acoustics comparisons of vocalisations produced by different species in equivalent situations have demonstrated shared acoustic forms. For instance, an acoustic comparison of tickling-induced laughter from humans, orangutans, gorillas, chimpanzees and bonobos suggested that laughter produced while being tickled is homologous across great apes (Davila Ross, Owren, & Zimmermann, 2009). In the current study, phylogenetic continuity in mapping of vocalisations to behavioural contexts may thus underlie inherent capacities of perceiving acoustic regularities that are conserved across species. For instance, acoustic structures of vocalisations produced in situations like discovering a large food source or being attacked might be shared across related species, allowing listeners to infer the production context from nonverbal vocalisations from different species. There might be thus preserved acoustic regularities in nonverbal vocalisations that are produced in specific behavioural contexts, mapping to certain functions. In order to assess

such regularities, future research might directly compare the acoustic properties of vocalisations produced in matched behavioural contexts across related species.

Confusions in inferences of behavioural contexts

In Experiment 1, listeners failed to accurately match vocalisations to one of the contexts: eating dispreferred food. However, participants were able to accurately categorise vocalisations from this context in Experiment 2 at better-than-chance rates when a clarification was provided for the description of this context. Having said that, performance was worse than for all of the other contexts (correct classification rate: 24.13 %), and the most common type of categorisation confusion between contexts was that eating dispreferred food was often categorised as eating preferred food (confusion percentage: 64.11 %). These results suggest that listeners accurately inferred from vocalisations that they involved eating, but with little sensitivity to the degree of preference for the food being consumed.

A closer look at the information communicated via the vocalisations produced in these contexts might be helpful for understanding the confusions. In our study, confusions in judgments of certain contexts from vocalisations might reflect them being similar situation classes. For instance, vocalisations produced in situations in which individuals were separated vs refused access were frequently confused, perhaps suggesting that they might have been interpreted by perceivers as indicators of loss of opportunity. Similarly, listeners' confusions in the categorisation of vocalisations produced while being attacked, discovering threat, and threatening another might reflect the perception of these vocalisations as general threat displays, preparing the body for rapid action and signalling urgency and danger in the immediate environment. For instance, individuals from a Melanesian society, Trobrianders of Papua New Guinea, have been shown to interpret 'fear' gasping and 'anger' scowling as displays of threat (Crivelli et al., 2016). This might suggest that being threatened and

threatening someone else might share similarities in the way they are expressed due to common characteristics like urgency and danger.

While performance in the match-to-context task used in Experiment 1 did not differ between positive and negative contexts, listeners performed better in the categorisation task for vocalisations produced in positive, as compared to negative, contexts in Experiment 2. We are hesitant to draw strong conclusion from these unexpected and inconsistent results; it is not clear whether there is a difference in accuracy in judgments of positive vs. negative vocalisations, but the present data provide no evidence of superior performance for negative as compared to positive contexts.

Conclusions

In conclusion, the present study provides empirical evidence for the theoretical proposal that spontaneous nonverbal vocalisations are perceived as responses to particular types of situations like being threatened or tickled (Russell, 1994; Scarantino, 2017). In contrast to studies relying on subjective judgements of listeners' perceptions of emotions, examining responses to verifiable events provides an objective way of investigating the perception of nonverbal vocalisations. Moreover, this approach allows us to better understand vocal signalling behaviour across species. Research from biological sciences demonstrates that the structural forms of vocalisations are shaped by their communicative functions in many vertebrates (e.g., Morton, 1977). Here, we show that human listeners can accurately infer specific behavioural contexts from spontaneous nonverbal vocalisations, suggesting that structural forms in nonverbal vocalisations appear to be systematically shaped by our behavioural context, and our sensitivity to this information in others' vocalisations may reflect a phylogenetically ancient vocal signalling system.

Acknowledgements

We would like to thank Dr. Piera Filippi for useful comments on an earlier version of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

R.G.K. and D.A.S. are supported by ERC Starting grant no. 714977 awarded to D.A.S.

Availability of data

The data that support the findings of Experiment 1 are openly available from <https://doi.org/10.21942/uva.13560374>, and the data that support the findings of Experiment 2 are openly available from <https://doi.org/10.21942/uva.18972962>.

Ethics statement

This study (project no. 2020-SP-11883) was approved by the Faculty Ethics Review Board of the University of Amsterdam, the Netherlands.

References

- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*, 4, 63. <https://doi.org/10.12688/wellcomeopenres.15191.1>
- Andics, A., & Faragó, T. (2018). Voice perception across species. In S. Frühholz, & P. Belin (Eds.), *Oxford handbook of voice perception* (pp. 363-392). Oxford (UK): Oxford University Press. <https://doi.org/10.1093/oxfordhob/9780198743187.013.16>
- Anikin, A., & Persson, T. (2017). Nonlinguistic vocalisations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods*, 49, 758-771. <https://doi.org/10.3758/s13428-016-0736-y>
- Anikin, A., Pisanski, K., & Reby, D. (2020). Do nonlinear vocal phenomena signal negative valence or high emotion intensity?. *Royal Society Open Science*, 7, 201306. <https://doi.org/10.1098/rsos.201306>
- Atias, D., Todorov, A., Liraz, S., Eidinger, A., Dror, I., Maymon, Y., & Aviezer, H. (2019). Loud and unclear: Intense real-life vocalisations during affective situations are perceptually ambiguous and contextually malleable. *Journal of Experimental Psychology: General*, 148, 1842. <https://doi.org/10.1037/xge0000535>
- Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer (version 6.0.30). Retrieved from <http://www.praat.org/>
- Bryant, G. A. (2013). Animal signals and emotion in music: Coordinating affect across groups. *Frontiers in Psychology*, 4, 1–13. <https://doi.org/10.3389/fpsyg.2013.00990>
- Bryant, G. A. (2021). The evolution of human vocal emotion. *Emotion Review*, 13, 25-33. <https://doi.org/10.1177/1754073920930791>

- Bryant, G. A., & Barrett, H. C. (2007). Recognizing intentions in infant- directed speech: Evidence for universals. *Psychological Science*, 18, 746–751.
<https://doi.org/10.1111/j.1467-9280.2007.01970.x>
- Clay, Z., Archbold, J., and Zuberbühler, K. (2015). Functional flexibility in wild bonobo vocal behaviour. *PeerJ*, 3, e1124. <https://doi.org/10.7717/peerj.1124>
- Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion*, 16, 117-128. <https://doi.org/10.1037/emo0000100>
- Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist*, 74, 698.
<https://doi.org/10.1037/amp0000399>
- Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2020). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, 1-7.
<https://doi.org/10.1038/s41586-020-3037-7>
- Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J. M. (2016). The fear gasping face as a threat display in a Melanesian society. *Proceedings of the National Academy of Sciences*, 113(44), 12403-12407. <https://doi.org/10.1073/pnas.1611622113>
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.
- Davila Ross, M., Owren, M. J., & Zimmermann, E. (2009). Reconstructing the evolution of laughter in great apes and humans. *Current Biology*, 19, 1106–1111.
<https://doi.org/10.1016/j.cub.2009.05.028>
- Eibl-Eibesfeldt, I. (1973). The expressive behavior of the deaf-and- blind-born. In M. von Cranach & I. Vine (Eds.), *Social communication and movement* (pp. 163-194). San Diego, CA: Academic Press.

- Endler, J. A. (1993). Some general comments on the evolution and design of animal communication systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 340, 215-225. <https://doi.org/10.1098/rstb.1993.0060>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. <https://doi.org/10.3758/BF03193146>
- Filippi, P. (2020). Emotional voice intonation: A communication code at the origins of speech processing and word-meaning associations?. *Journal of Nonverbal Behavior*, 44, 395-417. <https://doi.org/10.1007/s10919-020-00337-z>
- Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., ... & Newen, A. (2017). Humans recognize emotional arousal in vocalisations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences*, 284, 20170990. <https://doi.org/10.1098/rspb.2017.0990>
- Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.
- Frijda, N. H., Markam, S., Sato, K., & Wiers, R. (1995). Emotions and emotion words. In J. A. Russell, J. Fernandez-Dols, A. S. R. Manstead, & J. C. Wellenkamp (Eds.), *Everyday conceptions of emotion: An introduction to the psychology, anthropology and linguistics of emotion* (pp. 121–143). Dordrecht, The Netherlands: Kluwer Academic. https://doi.org/10.1007/978-94-015-8484-5_7
- Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: Facial expressions of emotions, in context. *Emotion Review*, 5, 60-65. <https://doi.org/10.1177/1754073912451331>
- Kamiloğlu, R. G., Slocombe, K. E., Haun, D. B. M., & Sauter, D. A. (2020). Human

- listeners' perception of behavioural context and core affect dimensions in chimpanzee vocalisations. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20201148. <https://doi.org/10.1098/rspb.2020.1148>
- Kersken, V., Zuberbühler, K., & Gomez, J. C. (2017). Listeners can extract meaning from non-linguistic infant vocalisations cross-culturally. *Scientific Reports*, 7, 1-7. <https://doi.org/10.1038/srep41016>
- Koutseff, A., Reby, D., Martin, O., Levrero, F., Patural, H., & Mathevon, N. (2018). The acoustic space of pain: cries as indicators of distress recovering dynamics in pre-verbal infants. *Bioacoustics*, 27, 313-325. <https://doi.org/10.1080/09524622.2017.1344931>
- Ladich, F. (1989). Sound production by the river bullhead, *Cottus gobio* L. (Cottidae, Teleostei). *Journal of Fish Biology*, 35, 531-538. <https://doi.org/10.1111/j.1095-8649.1989.tb03004.x>
- Laukka, P. (2004). *Vocal expression of emotion: discrete-emotions and dimensional accounts* (Doctoral dissertation, Acta Universitatis Upsaliensis, Uppsala, Sweden).
- Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W.,... Thingujam, N. S. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalisations. *Frontiers in Psychology*, 4, 353. <https://doi.org/10.3389/fpsyg.2013.00353>
- Lindová, J., Špinka, M., & Nováková, L. (2015). Decoding of baby calls: can adult humans identify the eliciting situation from emotional vocalisations of preverbal infants?. *PLoS One*, 10, e0124317. <https://doi.org/10.1371/journal.pone.0124317>
- Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., & Romanow, C. A. (2012). What makes a cry a cry? A review of infant distress vocalizations. *Current Zoology*, 58, 698-726. <https://doi.org/10.1093/czoolo/58.5.698>

- Linnankoski, I., Laakso, M., Aulanko, R., & Leinonen, L. (1994). Recognition of emotions in macaque vocalizations by children and adults. *Language & Communication*, 14, 183-192. [https://doi.org/10.1016/0271-5309\(94\)90012-4](https://doi.org/10.1016/0271-5309(94)90012-4)
- Lutz, C. (1982). The domain of emotion words on Ifaluk. *American Ethnologist*, 9, 113–128. <https://doi.org/10.1525/ae.1982.9.1.02a00070>
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98, 185–199. <https://doi.org/10.1037/0033-2909.98.1.185>
- Mehr, S. A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and function in human song. *Current Biology*, 28, 356-368. <https://doi.org/10.1016/j.cub.2017.12.042>
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111, 855–869. <https://doi.org/10.1086/283219>
- Mueller, H. C. (1971). Displays and vocalizations of the sparrow hawk. *The Wilson Bulletin*, 249-254.
- Neel, R., Brown, N. A., & Sng, O. (2020). Evolutionary Perspectives on Situations. In J. Rauthmann, R. A., Sherman, & D. C. Funder, (Eds.), *The Oxford handbook of psychological situations*, (pp. 112-123). New York, NY: Oxford University Press.
- Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, 1, 261-289. <https://doi.org/10.1007/BF02733986>

- Nicastro, N. & Owren, M. J. (2003). Classification of domestic cat (*Felis catus*) vocalisations by naïve and experienced human listeners. *Journal of Comparative Psychology*, 117, 44-52. <https://doi.org/10.1037/0735-7036.117.1.44>
- Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology*, 10, 58–71.
- Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, 6, e27256. <https://doi.org/10.1371/journal.pone.0027256>
- Pisanski, K., & Bryant, G. A. (2019). The evolution of voice perception. In N. S. Eidsheim & K. L. Meizel (Eds.), *Oxford handbook of voice studies* (pp. 269–300). New York, NY: Oxford University Press.
- Pisanski, K., Bryant, G. A., Cornec, C., Anikin, A., & Reby, D. (2022). Form follows function in human nonverbal vocalisations. *Ethology Ecology & Evolution*, 1-19. <https://doi.org/10.1080/03949370.2022.2026482>
- Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice modulation: A window into the origins of human vocal control? *Trends in Cognitive Sciences*, 20, 304–318. <https://doi.org/10.1016/j.tics.2016.01.002>
- Pongrácz, P., Molnár, C., Miklósi, Á. & Csányi, V. (2005). Human listeners are able to classify dog (*Canis familiaris*) barks recorded in different situations. *Journal of Comparative Psychology*, 119, 136-144. <https://doi.org/10.1037/0735-7036.119.2.136>
- Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports*, 5, 1-11. <https://doi.org/10.1038/srep13220>

- Raine, J., Pisanski, K., & Reby, D. (2017). Tennis grunts communicate acoustic cues to sex and contest outcome. *Animal Behaviour*, 130, 47-55.
<https://doi.org/10.1016/j.anbehav.2017.06.022>
- Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality*, 29, 363-381. <https://doi.org/10.1002/per.1994>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178. <https://doi.org/10.1037/h0077714>
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102-141.
<https://doi.org/10.1037/0033-2909.115.1.102>
- Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107, 2408–2412.
<https://doi.org/10.1073/pnas.0908239106>.
- Sauter, D. A., & Fischer, A. H. (2018). Can perceivers recognise emotions from spontaneous expressions?. *Cognition and Emotion*, 32, 504-515.
<https://doi.org/10.1080/02699931.2017.1320978>
- Sauter, D. A., & Russell, J. A. (Accepted/In press). What do nonverbal expressions tell us about emotion? In A. Scarantino (Ed.), *Handbook of Emotion Theory* Taylor & Francis.
- Scarantino, A. (2017). How to do things with emotional expressions: The theory of affective pragmatics. *Psychological Inquiry*, 28, 165-185.
<https://doi.org/10.1080/1047840X.2017.1328951>

- Scarantino, A., & Clay, Z. (2015). Contextually variable signals can be functionally referential. *Animal Behaviour*, 100, e1-e8.
<https://doi.org/10.1016/j.anbehav.2014.08.017>
- Shiota, M. N., & Keltner, D. (2005). What do emotion words represent?. *Psychological Inquiry*, 16, 32-37. <https://www.jstor.org/stable/20447257>
- Silva, K., Faragó, T., Pongrácz, P., Romeiro, P., & Lima, M. (2021). Humans' Ability to Assess Emotion in Dog Barks Only Slightly Affected by their Country of Residence, a Replication of Pongracz et al.(2005) in a Portuguese Sample. *Animal Behavior and Cognition*, 8, 107-123. <https://doi.org/10.26451/abc.08.02.02.2021>
- Susskind, J. M., & Anderson, A. K. (2008). Facial expression form and function. *Communicative & Integrative Biology*, 1, 148-149.
<https://doi.org/10.4161/cib.1.2.6999>
- Tallet, C., Špínka, M., Maruščáková, I., & Šimeček, P. (2010). Human perception of vocalizations of domestic piglets and modulation by experience with domestic pigs (*Sus scrofa*). *Journal of Comparative Psychology*, 124, 81.
<https://doi.org/10.1037/a0017354>
- Townsend, S. W., & Manser, M. B. (2013). Functionally referential communication in mammals: the past, present and the future. *Ethology*, 119, 1-11.
<https://doi.org/10.1111/eth.12015>
- Tsai, C. G., Wang, L. C., Wang, S. F., Shau, Y. W., Hsiao, T. Y., & Auhagen, W. (2010). Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception*, 27, 209-222.
<https://doi.org/10.1525/mp.2010.27.3.209>

Wagner, W. E. (1989). Fighting, assessment, and frequency alteration in Blanchard's cricket frog. *Behavioral Ecology and Sociobiology*, 25, 429-436.

<https://doi.org/10.1007/BF00300189>

Wood, A., Martin, J., & Niedenthal, P. (2017). Towards a social functional account of laughter: Acoustic features convey reward, affiliation, and dominance. *PloS ONE*, 12, e0183811. <https://doi.org/10.1371/journal.pone.0183811>

Supplementary Materials

Sounds like a fight: Listeners can infer behavioural contexts from spontaneous nonverbal
vocalizations

Roza G. Kamiloğlu, Disa A. Sauter

Table 1S. *Details of human nonverbal vocalizations obtained from online resources*

URL*	Context Category	Speaker Sex	Speaker Age
https://www.youtube.com/watch?v=q2RmWDgO6O8	Being attacked by another person	Woman	18-40
https://www.youtube.com/watch?v=4o9nfu0ncyY	Being attacked by another person	Woman	18-40
https://www.youtube.com/watch?v=siSue4OFDuY	Being attacked by another person	Woman	18-40
https://www.youtube.com/watch?v=cvi-35SyAjM	Being attacked by another person	Woman	18-40
https://www.youtube.com/watch?v=BxA_7XT6qs	Being attacked by another person	Woman	18-40
https://www.youtube.com/watch?v=M34JIvAejWU	Being attacked by another person	Woman	18-40
https://www.youtube.com/watch?v=LTmbJKmcDDg&list=PLHdUfgVj6foDPQ3nyjZbXrWNp7DCApgun	Being attacked by another person	Woman	18-40
https://www.youtube.com/watch?v=SI0L28Nqsxw	Being attacked by another person	Woman	18-40
https://www.youtube.com/watch?v=MAkgkdG0GQc	Being attacked by another person	Woman	40-60
https://www.youtube.com/watch?v=_aknbU0mkg&list=PLtaCF5hE81M7NuR9VK6pA2KoCcjtEUnd&index=2&t=0s&has_verified=1	Being attacked by another person	Man	18-40
https://www.youtube.com/watch?v=ueH_DtPbTYk	Being attacked by another person	Boy	13-18
https://www.youtube.com/watch?v=Uz1KXqGbW6s	Being attacked by another person	Man	18-40
https://www.youtube.com/watch?v=a0GCF9lKgg	Being attacked by another person	Boy	13-18
https://www.youtube.com/watch?v=x3BbIrVFsQo	Being attacked by another person	Man	18-40
https://www.youtube.com/watch?v=HZDakU2xLLM	Being attacked by another person	Man	18-40
https://www.youtube.com/watch?v=eqSODkTf2z8	Being attacked by another person	Boy	5-13
https://www.youtube.com/watch?v=y1wen1MmDD0	Being attacked by another person	Man	18-40
https://www.youtube.com/watch?v=bAtXfleRbL8	Being attacked by another person	Man	18-40
https://www.youtube.com/watch?v=OP3Lx-ruNKA	Being attacked by another person	Man	18-40
https://www.youtube.com/watch?v=HRia9W7Hduc	Being attacked by another person	Man	60+

Perception of Behavioral Contexts

https://www.youtube.com/watch?v=VKb0FIIBTas	Being refused access to food	Woman	18-40
https://www.youtube.com/watch?v=M6ml7NOnZ4Q	Being refused access to food	Girl	0-5
https://www.youtube.com/watch?v=by2eCYN AJFI	Being refused access to food	Woman	18-40
https://www.youtube.com/watch?v=VzJFucfQWeY	Being refused access to food	Woman	18-40
https://www.youtube.com/watch?v=xrY8Qfjh5t8&t=2s	Being refused access to food	Woman	18-40
https://www.youtube.com/watch?v=RO9o3WJeu_s	Being refused access to food	Boy	0-5
https://www.youtube.com/watch?v=PUEwR2WSNuY	Being refused access to food	Boy	0-5
https://www.youtube.com/watch?v=EDvm-dwvMdo	Being refused access to food	Boy	5-13
https://www.youtube.com/watch?v=by2eCYN AJFI	Being refused access to food	Man	18-40
https://www.youtube.com/watch?v=ZJqY-Sk9tvA	Being refused access to food	Boy	5-13
https://www.youtube.com/watch?v=KZ7TODtULWA	Being refused access to food	Man	18-40
https://www.youtube.com/watch?v=ZupPO9FhA3E	Being refused access to food	Boy	0-5
https://www.youtube.com/watch?v=BZ2sK77-Fl4	Being refused access to food	Boy	0-5
https://www.youtube.com/watch?v=zYhE5pMX9Bc	Being refused access to food	Boy	0-5
https://www.youtube.com/watch?v=QwgIKMwMkLQ	Being refused access to food	Boy	0-5
https://www.youtube.com/watch?v=TVf2PXE2fOE	Being refused access to food	Boy	0-5
https://www.youtube.com/watch?v=QSXsKVZ-V4Y	Being refused access to food	Boy	0-5
https://www.youtube.com/watch?v=amUiXp02sQA	Being refused access to food	Boy	0-5
https://www.youtube.com/watch?v=9tAbhNvWi90	Being refused access to food	Man	18-40
https://www.youtube.com/watch?v=ERIRi_Xnv2A	Being refused access to food	Man	18-40
https://www.youtube.com/watch?v=x2usVu8O2YY	Being separated from mother	Girl	0-5
https://www.youtube.com/watch?v=x2usVu8O2YY	Being separated from mother	Girl	0-5
https://www.youtube.com/watch?v=kUoJ_NfTpvQ	Being separated from mother	Girl	0-5
https://www.youtube.com/watch?v=DRejV6f-Y3c&t=1s	Being separated from mother	Girl	0-5
https://www.youtube.com/watch?v=yxAwOv7BPFY	Being separated from mother	Girl	0-5
https://www.youtube.com/watch?v=yxAwOv7BPFY	Being separated from mother	Girl	0-5
https://www.youtube.com/watch?v=yxAwOv7BPFY	Being separated from mother	Girl	0-5

Perception of Behavioral Contexts

https://www.youtube.com/watch?v=i5MudJ7yxkE&index=5&list=PL4Qr_qW_KkY3uFvoL-cRdSPPIWwwariS-	Being separated from mother	Girl	0-5
https://www.youtube.com/watch?v=i5MudJ7yxkE&index=5&list=PL4Qr_qW_KkY3uFvoL-cRdSPPIWwwariS-	Being separated from mother	Girl	0-5
https://www.youtube.com/watch?v=JqlbvpXp74k	Being separated from mother	Girl	0-5
https://www.youtube.com/watch?v=x2usVu8O2YY	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=4qq1CGBao_M	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=4qq1CGBao_M	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=u9OKCtXWJ3k&index=4&list=PL4Qr_qW_KkY3uFvoL-cRdSPPIWwwariS-	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=0rIXzh5bEGg	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=0rIXzh5bEGg	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=MRfbXbLjwzA&index=4&list=PLFsmYzR8A1fgRs96WqU1fsH4lN2n9kJjq	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=Y6QtuU1L_A8	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=Rdrp-0GLzws&t=17s	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=Y6QtuU1L_A8	Being separated from mother	Boy	0-5
https://www.youtube.com/watch?v=m-ZKFgzioos	Being tickled	Woman	18-40
https://www.youtube.com/watch?v=Ai1GwNeFUKk	Being tickled	Woman	18-40
https://www.youtube.com/watch?v=s43zyzJX308&index=3&list=PLL7K_qE1CdkuujCMIg11nToLREOVHmw_I	Being tickled	Woman	18-40
https://www.youtube.com/watch?v=5GNQcGVLPVc&index=9&list=PLL7K_qE1CdkuujCMIg11nToLREOVHmw_I	Being tickled	Girl	13-18
https://www.youtube.com/watch?v=azII1qfhc4s&index=59&list=PLL7K_qE1CdkuujCMIg11nToLREOVHmw_I	Being tickled	Woman	18-40
https://www.youtube.com/watch?v=kt3afdVgvNg&list=PLL7K_qE1CdkuujCMIg11nToLREOVHmw_I&index=71	Being tickled	Girl	13-18
https://www.youtube.com/watch?v=us-O25lnj1o	Being tickled	Woman	18-40
https://www.youtube.com/watch?v=sdjn1111u5o&list=PL2479B8C9D8E2F6BF&index=11	Being tickled	Woman	18-40
https://www.youtube.com/watch?v=o61a-o62S1U&index=3&list=PL040EB6B8BBFEDD33	Being tickled	Woman	18-40

Perception of Behavioral Contexts

https://www.youtube.com/watch?v=7HmAt1nF_iY&list=RDs43zyzJX308&index=1	Being tickled	Woman	60+
https://www.youtube.com/watch?v=iazF-pbm-h4	Being tickled	Man	18-40
https://www.youtube.com/watch?v=BBGz0pER2Mc&index=67&list=PLL7K_qE1CdkuuJCMlg11nToLREOVHmw_I	Being tickled	Boy	13-18
https://www.youtube.com/watch?v=FQpLNv7uDis	Being tickled	Boy	13-18
https://www.youtube.com/watch?v=CcLaq6zs4Fk	Being tickled	Boy	13-18
https://www.youtube.com/watch?v=Atrph-uhU9M	Being tickled	Man	40-60
https://www.youtube.com/watch?v=VWMvvTuwb4k	Being tickled	Man	18-40
https://www.youtube.com/watch?v=pN0DpRweFTI	Being tickled	Man	60+
https://www.youtube.com/watch?v=KZEhdiBf41s&list=PLxeLUE R8gqauGRcsESUCytn8V5-sHStD9	Being tickled	Boy	13-18
https://www.youtube.com/watch?v=QzUNfuGTju4	Being tickled	Boy	13-18
https://www.youtube.com/watch?v=6ob0m7oG9jU&index=80&list=PLO-ZVz4tLUCbTXxmd7gaJxY6yk9sqOQUA	Being tickled	Man	18-40
https://www.youtube.com/watch?v=XsviZdZYQe0	Having sex	Woman	18-40
https://www.youtube.com/watch?v=3KDDkAR-vpM	Having sex	Woman	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5e07870d72c61	Having sex	Woman	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5e07870d72c61	Having sex	Woman	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5db33259e7a6a	Having sex	Woman	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5df3c6efe730a	Having sex	Woman	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5df3c6efe730a	Having sex	Woman	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5c88049da5899	Having sex	Woman	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5db33259e7a6a	Having sex	Man	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5db33259e7a6a	Having sex	Man	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5df3c6efe730a	Having sex	Man	18-40

Perception of Behavioral Contexts

https://www.pornhub.com/view_video.php?viewkey=ph5df3c6efe730a	Having sex	Man	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5c88049da5899	Having sex	Man	18-40
https://www.pornhub.com/view_video.php?viewkey=ph5c88049da5899	Having sex	Man	18-40
https://www.youtube.com/watch?v=zgGfKc9J5uc	Having sex	Man	18-40
https://www.youtube.com/watch?v=U882EX8AXys	Having sex	Man	18-40
https://www.youtube.com/watch?v=IYQRHBiltek	Having sex	Man	18-40
https://www.youtube.com/watch?v=Y96c03ULZec	Having sex	Man	18-40
https://www.youtube.com/watch?v=jgXZ1u2GuLw	Having sex	Woman	18-40
https://www.youtube.com/watch?v=4_9R5MJHRn4	Having sex	Woman	18-40
https://www.youtube.com/watch?v=Fd7vY-_bK_0	Discovering a large food source	Woman	18-40
https://www.youtube.com/watch?v=tp8yQpQN5l8	Discovering a large food source	Woman	40-60
https://www.youtube.com/watch?v=VwyTC0OpKmg	Discovering a large food source	Woman	40-60
https://www.youtube.com/watch?v=mOWxa3ym4Gc	Discovering a large food source	Woman	18-40
https://www.youtube.com/watch?v=ANT-tgvz0oc	Discovering a large food source	Woman	18-40
https://www.youtube.com/watch?v=1R5uPaL0V-0	Discovering a large food source	Woman	18-40
https://www.youtube.com/watch?v=syH0oN_B4g0	Discovering a large food source	Woman	18-40
https://www.youtube.com/watch?v=_SeMYvpwAzo	Discovering a large food source	Woman	18-40
https://www.youtube.com/watch?v=HbLiJwQObmQ	Discovering a large food source	Woman	18-40
https://www.youtube.com/watch?v=SzvFtRmH0uQ	Discovering a large food source	Girl	5-13
https://www.youtube.com/watch?v=SzvFtRmH0uQ	Discovering a large food source	Girl	5-13
https://www.youtube.com/watch?v=y0jLphbuWzA	Discovering a large food source	Woman	18-40
https://www.youtube.com/watch?v=-uS2o4XKPk4	Discovering a large food source	Woman	18-40
https://www.youtube.com/watch?v=SljBIDOUXyY	Discovering a large food source	Man	18-40
https://www.youtube.com/watch?v=NFr-85NNmFk	Discovering a large food source	Man	18-40
https://www.youtube.com/watch?v=NFr-85NNmFk	Discovering a large food source	Man	18-40

Perception of Behavioral Contexts

https://www.youtube.com/watch?v=toeFNbExKkQ	Discovering a large food source	Man	18-40
https://www.youtube.com/watch?v=RYNdJ-auesc	Discovering a large food source	Man	18-40
https://www.youtube.com/watch?v=3Lq0RT94nzg	Discovering a large food source	Man	18-40
https://www.youtube.com/watch?v=-uS2o4XKPk4	Discovering a large food source	Man	18-40
https://www.youtube.com/watch?v=ifgDyujYbmc	Discovering something threatening	Girl	NA
https://www.youtube.com/watch?v=ifgDyujYbmc	Discovering something threatening	Girl	13-18
https://www.youtube.com/watch?v=ifgDyujYbmc	Discovering something threatening	Woman	18-40
https://www.youtube.com/watch?v=2NZSd1nUoEI	Discovering something threatening	Woman	18-40
https://www.youtube.com/watch?v=hsGI221te2s	Discovering something threatening	Woman	18-40
https://www.youtube.com/watch?v=hsGI221te2s	Discovering something threatening	Woman	NA
https://www.youtube.com/watch?v=hsGI221te2s	Discovering something threatening	Woman	18-40
https://www.youtube.com/watch?v=7PafDNqaM-M	Discovering something threatening	Woman	18-40
https://www.youtube.com/watch?v=6eH5IOiBzPI	Discovering something threatening	Woman	18-40
https://www.youtube.com/watch?v=6eH5IOiBzPI	Discovering something threatening	Woman	18-40
https://www.youtube.com/watch?v=u1cv08ZOHkM	Discovering something threatening	Boy	13-18
https://www.youtube.com/watch?v=ifgDyujYbmc	Discovering something threatening	Man	NA
https://www.youtube.com/watch?v=ifgDyujYbmc	Discovering something threatening	Man	18-40
https://www.youtube.com/watch?v=ifgDyujYbmc	Discovering something threatening	Man	18-40
https://www.youtube.com/watch?v=ifgDyujYbmc	Discovering something threatening	Boy	13-18
https://www.youtube.com/watch?v=2NZSd1nUoEI	Discovering something threatening	Boy	13-18
https://www.youtube.com/watch?v=2NZSd1nUoEI	Discovering something threatening	Man	NA
https://www.youtube.com/watch?v=2NZSd1nUoEI	Discovering something threatening	Man	18-40
https://www.youtube.com/watch?v=hsGI221te2s	Discovering something threatening	Man	NA
https://www.youtube.com/watch?v=7PafDNqaM-M	Discovering something threatening	Boy	13-18
https://www.youtube.com/watch?v=5BLX3Df3sc4	Eating high value (strongly preferred) food	Woman	18-40
https://www.youtube.com/watch?v=QDlbOrOO0_E	Eating high value (strongly preferred) food	Woman	18-40

Perception of Behavioral Contexts

https://www.youtube.com/watch?v=-yMXWp-Fh3g	Eating high value (strongly preferred) food	Woman	18-40
https://www.youtube.com/watch?v=-yMXWp-Fh3g	Eating high value (strongly preferred) food	Woman	18-40
https://www.youtube.com/watch?v=0xdW4xccFkE	Eating high value (strongly preferred) food	Woman	18-40
https://www.youtube.com/watch?v=HdqHZ3LA9pE	Eating high value (strongly preferred) food	Woman	18-40
https://www.youtube.com/watch?v=lfsvLRVwflA	Eating high value (strongly preferred) food	Woman	18-40
https://www.youtube.com/watch?v=lfsvLRVwflA	Eating high value (strongly preferred) food	Woman	18-40
https://www.youtube.com/watch?v=oQGkzsTGMPg	Eating high value (strongly preferred) food	Woman	18-40
https://www.youtube.com/watch?v=oQGkzsTGMPg	Eating high value (strongly preferred) food	Woman	18-40
https://www.youtube.com/watch?v=4E2K3IrhKJE	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=xFo6NI12fjc	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=xFo6NI12fjc	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=xFo6NI12fjc	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=FLmNLrog6dQ	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=FLmNLrog6dQ	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=5r-jxlwpHKs	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=5r-jxlwpHKs	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=HdqHZ3LA9pE	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=HdqHZ3LA9pE	Eating high value (strongly preferred) food	Man	18-40
https://www.youtube.com/watch?v=5BLX3Df3sc4	Eating low value (strongly dis-preferred) food	Woman	18-40
https://www.youtube.com/watch?v=QDIbOrOO0_E	Eating low value (strongly dis-preferred) food	Woman	18-40
https://www.youtube.com/watch?v=QDIbOrOO0_E	Eating low value (strongly dis-preferred) food	Woman	18-40
https://www.youtube.com/watch?v=QDIbOrOO0_E	Eating low value (strongly dis-preferred) food	Woman	18-40
https://www.youtube.com/watch?v=-yMXWp-Fh3g	Eating low value (strongly dis-preferred) food	Woman	18-40
https://www.youtube.com/watch?v=HdqHZ3LA9pE	Eating low value (strongly dis-preferred) food	Woman	18-40

Perception of Behavioral Contexts

https://www.youtube.com/watch?v=HdqHZ3LA9pE	Eating low value (strongly dis-preferred) food	Woman	18-40
https://www.youtube.com/watch?v=ixS3OzzE44E	Eating low value (strongly dis-preferred) food	Woman	18-40
https://www.youtube.com/watch?v=21e4sNc7KdI	Eating low value (strongly dis-preferred) food	Woman	18-40
https://www.youtube.com/watch?v=G8fF-Wt_sGQ	Eating low value (strongly dis-preferred) food	Woman	18-40
https://www.youtube.com/watch?v=HdqHZ3LA9pE	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=HdqHZ3LA9pE	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=HdqHZ3LA9pE	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=HdqHZ3LA9pE	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=lfsvLRVwflA	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=G8fF-Wt_sGQ	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=G8fF-Wt_sGQ	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=-8vonWWx3rY	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=-8vonWWx3rY	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=-8vonWWx3rY	Eating low value (strongly dis-preferred) food	Man	18-40
https://www.youtube.com/watch?v=HBsfO2-z51k	Threatening an aggressive person or people	Woman	18-40
https://www.youtube.com/watch?v=jd8zWzH4xb0	Threatening an aggressive person or people	Girl	5-13
https://www.youtube.com/watch?v=g_YqxB8OT_U	Threatening an aggressive person or people	Woman	18-40
https://www.youtube.com/watch?v=TtC14cpwwXg	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=d4FbHzeCJjM	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=ij3wbx67zwY	Threatening an aggressive person or people	Man	40-60
https://www.youtube.com/watch?v=veVQRNi4QKo	Threatening an aggressive person or people	Man	18-40

Perception of Behavioral Contexts

https://www.youtube.com/watch?v=c-eAozVfBkU	Threatening an aggressive person or people	Man	40-60
https://www.youtube.com/watch?v=YZEbBZ2IrXE	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=WqkFN1hcdv8	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=bJMOoAVbcNk	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=jpB7jrJ7Zds	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=HmmJzRGisLA	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=hboT-iJzgKQ	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=hboT-iJzgKQ	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=cvwPLFsM46o	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=vJg9ycj6ZaU	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=cvwPLFsM46o	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=gcTg9mt4eAk	Threatening an aggressive person or people	Man	18-40
https://www.youtube.com/watch?v=DwDxEfRAQyk	Threatening an aggressive person or people	Man	18-40

Notes. * indicates the URLs from which the vocalisations were taken. They might not remain active.