Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ECHO Corpus: Cross-cultural database of 15000 nonverbal vocalisations from speakers of 22 languages

Aitana García Arasco

aitana.arasco@gmail.com

Université Lumière Lyon 2

Théophile Turco

Université Jean Monnet

Urara Isa

The University of Tokyo

Hiroki Koda

The University of Tokyo

Isao Tokuda

Ritsumeikan University

David Reby

Université Jean Monnet

Katarzyna Pisanski

Université Jean Monnet

data-descriptor

Keywords: voice production, acoustic communication, human vocal repertoire, emotion communication, cross-cultural research

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10956932/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

ECHO Corpus: Cross-cultural database of 15000 nonverbal vocalisations from speakers of 22 languages

Aitana García Arasco[1,2]*, Théophile Turco[2], Urara Isa[3], Hiroki Koda[4], Isao Tokuda[5], David Reby[2,6] & Katarzyna Pisanski[1,2,7]*

[1] *Dynamique Du Langage (UMR 5596, CNRS et Université Lyon 2), Lyon, France*
[2] *ENES Bioacoustics Research Laboratory, University Jean Monnet of Saint-Etienne, CRNL, CNRS, Saint-Etienne, France*
[3] *Graduate School of Education, The University of Tokyo, Tokyo, Japan*
[4] *Graduate School of Arts and Sciences, University of Tokyo, Tokyo, Japan*
[5] *Graduate School of Science and Engineering, Ritsumeikan University, Kusatsu, Shiga, Japan*
[6] *Institut universitaire de France*
[7] *Institute of Psychology, University of Wrocław, Wrocław, Poland*

*corresponding authors: A. Garcia Arasco (aitana.arasco@gmail.com) and K. Pisanski (katarzyna.pisanski@cnrs.fr)

Abstract

Research on human non-verbal vocalisations such as screams, cries and laughter is rapidly expanding, yet most studies remain restricted to single cultures and heavily overrepresent English speakers. Here, we introduce the ECHO corpus – a large-scale cross-cultural database of non-verbal vocalisations produced by nearly 1050 vocalisers aged 18-84 years from 52 countries, including native speakers of 22 different languages spanning 10 language families. The corpus comprises over 15000 volitional vocalizations expressing sixteen affective states, from pain, aggression and alarm to amusement, relief and sensual pleasure, together with baseline speech from the same speakers. We provide extensive metadata on vocalisers (e.g., age, sex, countries of birth and residence, spoken languages, and media usage) and validate the dataset through acoustic quality assessment, analyses of fundamental frequency, and cross-cultural perception experiments demonstrating high recognition of vocalisation contexts. This

first-of-its-kind resource provides a representative foundation for investigating how humans vocally encode and decode emotions and motivations across diverse cultures, revealing the mechanisms and functions of human vocal communication. The ECHO corpus is openly available at ECHO Corpus.

(https://zenodo.org/records/22754938?token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjMxODdjY2ZiLWJjNTktNDU0MC04NGI1LWVkNDJjMWU2ZTM2MSIsImRhdGEiOnt9LCJyYW5kb20iOiIwZmZiOGNmZDM4MTE2MjU2MzY0MGE4NTFmMzhhMTY4YiJ9.vKnNZEc9dGBZ9pnnWtf8z1Vju3rof-Hs6hBSKBwzmWwSfZCOGq5QLCzTyqTpjIGDqTJn8wmbekrWorw0jkOBSg)

Keywords: voice production; acoustic communication; human vocal repertoire, emotion communication, cross-cultural research

INTRODUCTION

Speech is the primary form of human linguistic expression, and for most of our evolutionary history, it has been the predominant mode of language, having evolved 100 to 200 thousand years ago [1]. However, speech is just the tip of the iceberg of what we humans can achieve with our vocal apparatus [2]. Non-verbal vocalisations, which likely emerged before language itself, are ubiquitous in everyday human interactions. These non-linguistic sounds such as cries, laughter and moans represent an important communicative channel in our species as they can convey information about both static traits such as identity, age, strength and body size [2–7], and dynamic states such as emotions and motivations of the speaker [7–12].

Over the last decade, there has been a growing interest in the study of human vocalisations, particularly with regard to their evolutionary origins and social functions [13 for review]. Several studies have shown that the acoustic characteristics of human vocalisations often map onto their communicative functions [4,7,14–16], paralleling the form-function relationships

observed in other species' calls [17–20]. For instance, distress calls across various non-human juveniles tend to be high-pitched, frequency modulated and rich in harsh nonlinear phenomena [19,21], supporting their function in eliciting aid. Human baby cries share these same acoustic features [22–24], seemingly to grab attention and motivate parents to provide care. While aggressive roars are also characterised by such rough nonlinear phenomena, they are relatively low pitched, serving to communicate threat in humans and numerous other species [2,4,7,20,25]. In contrast, spontaneous human laughter is described as homologous to the play-like calls of other animals [26], characterised by rhythmicity and variations in pitch and intensity [27–30]. This form-function approach provides a robust framework for making predictions about cross-cultural universals and variations in both the production and perception of human non-verbal vocalisations [13,31]. Yet, to test predictions stemming from this and other frameworks, researchers require representative samples of vocalisations spanning geographic, linguistic, and cultural dimensions, along with the many social contexts in which they can potentially appear.

Importantly, previous cross-cultural research on vocal emotion perception has demonstrated that listeners across the globe can often identify the intended emotions and motivations conveyed by non-verbal vocalisations [11,28,32–36]. This suggests that the relationships between the forms and functions of non-verbal vocalisations may be grounded in a shared system of vocal behaviour that is highly conserved across human societies, and even across species [37]. However, there are reasons to also expect cross-cultural acoustic variation in the structures of human non-verbal vocalisations, at least for volitional vocalisations – those that are produced voluntarily using vocal control. Unlike spontaneous vocalisations which are largely reflexive and, in their purest form, driven by deep primitive regions of the brain [38,39], volitional or voluntary non-verbal vocalisations exploit more higher-order cortical pathways shared with speech [38–41]. Recent evidence suggests that human volitional vocal expressions

may thus be learned, as their production requires not only auditory experience [42,43] but also pragmatic knowledge and a well-developed control of the vocal apparatus. If volitional vocalisations are indeed even partly learned, we can expect some variation in both their acoustic production and perception across cultures [44–46].

Although few studies have measured the acoustic properties of human vocalisations across cultures, one important clue that their acoustic forms and functions may vary cross-culturally comes from perception studies of volitional vocalisations. While listeners can recognise the intended emotions of these vocalisations from other cultures, they recognise emotions more easily when judging vocalisations produced by speakers from the same cultural background as their own [11,35]. In these studies, culture is often defined by the language that participants speak or by the country in which they were born. Could acoustic differences in the structure of these vocalisations across cultures explain this in-group advantage in voice perception? Could non-verbal vocalisations – particularly those produced voluntarily – be less 'hard-wired' than previously thought? We cannot answer these questions yet, as little is known about how the production of non-verbal vocalisations can vary across human societies. Our large cross-cultural database provides a unique resource for the scientific study of these outstanding issues.

Several smaller-scale corpora of vocalisations exist, most of them containing volitional calls produced on demand either by professional actors [34,47,48] or non-professionals [33,49–52]. Critically, in most cases, vocalisers are native speakers of only one shared language, and rarely come from small-scale societies. Only two corpora have applied a cross-cultural approach by recording non-verbal vocalisations across multiple cultures: the VENEC corpus, with 640 non-verbal vocalisations [53] and Cowen et al. corpus, with 1607 non-verbal vocalisations [8]. Together, these two corpora include more than 2000 voluntarily produced non-verbal vocalisations expressing 24 different intended emotional states from approximately 100 professional actors. But, vocalisers all belong to cultures in which English is an official

language (the USA, India, Kenya, Singapore, and Australia). This focus on English speakers and on populations traditionally considered WEIRD (Western, Educated, Industrialised, Rich and Democratic [54]), dramatically underrepresents the true variety of human cultures and the impact of sociocultural factors on human behaviour, including vocal behaviour.

To broaden the scope of research on the human vocal repertoire, here we present the ECHO (Emotional Cross-cultural Human Vocalisations) Corpus – a large-scale cross-cultural corpus of volitional non-verbal vocalisations, along with acoustic measures and survey data from 1051 speakers (aged 18-84) of 22 different languages. In addition to a collection of demographic, linguistic, and social media measures for all vocalisers, we have included audio WAV files of volitional non-verbal vocalisations produced by each consenting vocaliser in 16 hypothetical scenarios traditionally included in emotion recognition research (e.g., Cordaro et al., 2016; Laukka et al., 2013; Sauter et al., 2010). These scenarios range from negative contexts such as aggression, alarm, competition, fear, frustration, pain, sadness, grief, disgust and freezing to positive contexts such as amusement, achievement, relief, sensual pleasure and surprise, as well as a neutral social context (Dingemanse et al., 2013). For each consenting vocaliser, we have also included audio recordings of five vowel sounds embedded in words as modal, baseline speech, that may be used for comparison or control in acoustic analyses or perception experiments. By introducing a novel methodology of collecting standardised vocal recordings remotely (i.e., online) from geographically dispersed populations, alongside field data collection, the ECHO Corpus overcomes traditional logistical barriers to cross-cultural voice research and allows unprecedent sampling breadth. We show the feasibility of large-scale, high-quality online data collection for cross-cultural vocal research.

Alongside audio recordings, this database includes automatically extracted acoustic descriptors such as fundamental frequency (f_0, perceived as voice pitch), amplitude (loudness), peak frequency, call duration and percentage of voiced frames, spectral centroid and other voice

quality measures such as Cepstral peak prominence, harmonic-to-noise ratio, entropy and roughness, all known to be relevant in terms of encoding emotional valence and arousal [2,9,14–16,55–58]. To ensure the reliability and scientific utility of this resource, we performed extensive technical validation of the corpus. This included systematic assessment of recording quality, examination of fundamental frequency across affective contexts and language families, and cross-cultural perception experiments demonstrating that vocalisations convey recognisable affective information.

METHODS

Ethical considerations

Ethical approval for the acoustic recording of human subjects was provided by the Comité d'Ethique du CHU de Saint Etienne for vocalisers recorded via the Prolific platform or in Mongolia (IRBN692019/CHUSTE), and by the Ethical Review Committee for Experimental Research involving Human Subjects at the Graduate School of Arts and Sciences and the College of Arts and Sciences, University of Tokyo, for vocalisers recorded in Japan (# 962). All vocalisers confirmed that they were at least 18 years old and provided informed and written consent prior to their involvement in the study. Vocalisers acknowledged that their anonymized voice and demographic data would be analysed and disseminated in scientific reports and papers. Anonymity and confidentiality were guaranteed to vocalisers as well as the voluntary nature of their participation.

Vocalisers

The final dataset included data from 1028 vocalisers (515 Women, 493 Men, 10 non-binary, 10 unspecified) from 52 countries (See Figure 1), excluding those who withdrew from the study without completing it (n=635), produced words or interjections instead of non-verbal vocalisations (n=20), or did not complete the study before the time limit of 90 minutes (n=101).

Of the 1028 vocalisers who completed the study, 292 (28%) did not consent to publish their voice recordings, and therefore only acoustic measures and survey data are provided for those participants. Basic demographic characteristics of vocalisers in the final dataset are presented in Table 1 (see also Supplementary Table S1 for full details).

As our selection criteria prioritised linguistic diversity among vocalisers, the main objective was to recruit vocalisers from a variety of targeted languages spread across several major language families. To also achieve a broad global distribution, we chose ten language families: Indo-European, Uralic, Afro-Asiatic, Atlantic-Congo, Turkic, Sino-Tibetan, Austronesian, Dravidian, Mongolic and Japonic. In most cases, we obtained vocalisers from two to three different languages within each language family (Table 1).

A second selection criterion was that vocalisers who spoke a target language were born and had lived in the same country speaking that target language until at least the age of 18. In three languages where this was not possible due to reduced online availability of vocalisers (Kazakh, Burmese, and Malay), we included vocalisers whose native language was the target language, regardless of their place of birth. For English, Spanish, and Arabic, we deliberately recruited vocalisers from various countries to achieve greater representation of distinct dialects. For example, we recruited English speakers from Canada, the United States of America, and the United Kingdom. To further increase diversity, we included vocalisations of Darkhad and Dukha people – two small-scale societies in northern Mongolia. Most members of these nomadic groups have zero or very reduced access to the internet and western media, and have never travelled outside of Mongolia.

The voice recording study was mainly carried out online, except for Japanese and Mongolian vocalisers who were recorded onsite during fieldwork. Online vocalisers were recruited through the data collection platform Prolific (https://www.prolific.com/) and

compensated at the recommended rate of ~£7/h for their time. In Japan, we took advantage of an opportunistic sample of students and citizens recruited by word of mouth at Ritsumeikan University and the University of Tokyo. Mongolian recordings were part of a mission to reach the isolated communities of Darkhad and Dukha in the northern region of the country. We also recruited, by word of mouth, Mongolian citizens in the capital Ulan-Bataar and urban dwellers in the Bulgan province. On-site vocalisers provided informed consent to take part in the research and were monetarily compensated for their participation (Japanese: 1000 JPY, Darkhad: 25000 MNT).

Table 1: Gender and age distributions of vocalisers for each language family and language in the ECHO Corpus. Supplementary Table S1 provides full survey data including country of birth, mean number of spoken languages, mean number of languages at a proficient level, hours per year spent using social media, hours per year watching television or listening to the radio, and the number and percentage of vocalisers who agreed to publish their vocalisations online.

Language Family	Language	Gender	Age
Afro-Asiatic	Standard Arabic (N=93)	55 Men, 38 Women	29.7 ± 7.3
	Hebrew (N=42)	20 Men, 22 Women	33.9 ± 9.9
Atlantic-Congo	Swahili (N=42)	18 Men, 24 Women	27.4 ± 7.8
Austronesian	Malay (N=44)	19 Men, 25 Women	32.4 ± 9.4
	Filipino-Tagalog (N=40)	13 Men, 26 Women, 1 Non-binary	37.0 ± 10.3
Dravidian	Telugu (N=32)	15 Men, 15 Women, 2 Preferred not to say	30.1 ± 8.2
	Tamil (N=28)	9 Men, 19 Women	33.7 ± 8.7
	Malayalam (N=34)	20 Men, 14 Women	29.1 ± 9.1
Indo-European	English (N=115)	56 Men, 56 Women, 3 Non-binary	42.4 ± 13.9
	Spanish (N=56)	27 Men, 29 Women	28.9 ± 8.1
	French (N=46)	21 Men, 23 Women, 2 Non-binary	30.6 ± 8.6
Japonic	Japanese (N=48)	25 Men, 23 Women	21.2 ± 1.8
Mongolic	Mongolia (N=91)	51 Men, 40 Women	38.0 ± 15.2
Sino-Tibetan	Mandarin Chinese (N=48)	23 Men, 24 Women, 1 Non-binary	31.7 ± 8.3
	Cantonese Chinese (N=43)	17 Men 23 Women, 1 Non-binary, 2 Preferred not to say	33.5 ± 10.7
	Burmese (N=17)	3 Men, 7 Women, 1 Non-binary, 6 Preferred not to say	25.2 ± 6.1
Turkic	Turkish (N=47)	24 Men, 23 Women	31.2 ± 6.4
	Kazakh (N=26)	8 Men, 18 Women	25.1 ± 4.6
	Uzbek (N=4)	3 Men, 1 Woman	23 ± 4.3
Uralic	Estonian (N=47)	23 Men, 23 Women, 1 Non-binary	32.6 ± 7.9
	Hungarian (N=47)	25 Men, 22 Women	33.4 ± 12.1
	Finnish (N=38)	18 Men, 20 Women	35.4 ± 9.6
Total	**N=1028**	**493 Men, 515 Women, 10 Non-binary, 10 Preferred not to say**	**32.7 ± 11.2**

Procedure

Online experiments were designed in the Labvanced platform (Scicovery GmbH, Paderborn, Germany) [59]. Japanese and Ulan-bataar citizens conducted the study in an isolated soundproof room to ensure privacy. Darkhad, Dukha and Bulgan vocalisers conducted the study with the principal investigator and an assistant who provided real-time translation assistance. All vocalisers were informed that they could withdraw from the study at any time without repercussion. The study was divided into two parts. First, the participants completed a short survey. Then, they recorded their voices for the voice recording task. In the University of Tokyo study, the voice recording was done first and the survey was completed afterwards.

i) Recording task

The voice recording procedure was similar for online and on-site vocalisers: vocalisers were randomly presented with a vignette for each affective context and instructed to produce a non-verbal vocalisation (without using words or interjections) as convincingly as possible to express the given emotion or context. These contexts, based on situations proposed in previous literature [16,33,60–62], were deliberately selected to be biologically relevant and to represent specific situations that most people, regardless of culture, would either have experienced or would be able to easily imagine experiencing. For instance, vocalisers were asked to imagine the following scenario to elicit a volitional vocalisation of amusement: "*A close family member begins to tickle you. Produce a vocalisation to express your amusement*".

The aim of these vignettes was to elicit vocalisations expressing sixteen different affective states ranging from negative states such as *aggression, alarm, competition, fear, frustration pain, sadness, grief, disgust*, or *freezing* to positive states such as *amusement, achievement, relief, sensual pleasure* and *surprise*. Additionally, we added a neutral context (i.e. *misheard*, see Table 2 for context definitions), which is considered a 'repair initiator',

typically produced when a listener has not clearly heard what someone just said [63]. The sexual pleasure context was excluded for Japanese vocalisers due to ethical concerns and for Dukha, Darkhad and Bulgan vocalisers on the recommendation of the assistant guide. Presentation order of the contexts was randomised for each vocaliser.

Scenario-based vignettes provide a common, standardized and ethically feasible method for eliciting affective vocalisations under controlled conditions [52,53,56]. Unlike single emotion or context labels, which may be interpreted differently by different individuals or across cultures, vignettes provide a shared contextual framework that helps standardize both the intended affective state or context and its intensity. This reduces variability arising from differences in interpretation while enabling direct comparisons across individuals, languages and cultures. Importantly, such vignettes elicit *volitional* vocalisations, which reflect the exceptional degree of voluntary control humans exert over their vocal production [42]. Such voluntarily produced vocalisations are central to everyday social communication, allowing speakers to intentionally convey or modulate affective states [13], and are therefore an important complement to studies of spontaneous emotional vocalisations.

To keep the distance from the microphone as consistent as possible, online vocalisers were asked to position themselves 70 to 80 cm from their microphone (about an arm's length away) before the recording task, and to maintain this distance throughout the task. They were asked to avoid using words or interjections (e.g., yuck, ouch; [64], which we later verified. Additionally, we asked vocalisers to pronounce between 3 and 5 words containing the main /a/, /e/, /i/, /o/, /u/ vowels. To ensure consistency in the vowels used across all languages, we worked with translators to select meaningful words containing these vowel sounds for each language. This baseline speech is also included in the database. However, it is important to note that due to differences in pronunciation across different regions, there may be variations in vowel sounds across cultures. In addition, we included attention tests throughout the study to verify that

vocalisers were actively performing the task. These attention checks consisted of producing an imitation of a dog barking or a cat meowing. Vocalisers who did not complete the attention check (N=2) were excluded.

The procedure was similar for on-site vocalisers, except that attention tests were not needed. Mongolian vocalisers based in Ulaanbaatar and Japanese vocalisers were recorded alone in a quiet room and completed the study on a computer provided by the principal investigator. Darkhad and Dukha vocalisers were recorded inside their yurts, while Bulgan vocalisers were recorded in quiet outdoor spaces far from the town, within walking distance but outside of hearing range from other citizens to ensure privacy. Japanese participants from the Ritsumeikan University and Mongolian vocalisers were recorded using a Sennheiser E835 microphone connected to an H4N Essential Zoom positioned 100 centimetres from the vocaliser's mouth. A second Zoom recorder was placed at 130 metres away from the vocaliser in case of audio clipping. Japanese participants from Tokyo were recorded with a PCM Recorder DR-05 connected to a Tascam DR05 recorder.

Table 2: Scenario-based vignettes used to elicit volitional non-verbal vocalisations (original English version).

Valence	Context	Description
Positive	Achievement	You have won something and you want to celebrate with your family/friends. Produce a vocalisation to express your achievement.
	Amusement	A close family member begins to tickle you. Produce a vocalisation to express your amusement
	Pleasure	You are having sex and you are enjoying it very much. Produce a vocalisation to express your pleasure.
	Relief	You almost dropped your food on the floor. Produce a vocalisation to express that you feel relieved.
	Surprise	Someone gives you a gift and you didn't expect it. Produce a vocalisation to express your surprise.
Negative	Aggression	Someone unknown tries to hit your close family member. Produce a vocalisation to express that you are feeling angry about it.
	Alarm	You realize a friend is going to be attacked by a wild animal but he/she has not noticed. Produce a vocalisation to warn him/her.
	Competition	You are in a competition with a rival. Produce a vocalisation to discourage your adversary
	Disgust	You have eaten some rotten food. Produce a vocalisation to express your disgust.
	Sadness	

	Fear	You wake up and you realize your house is burning. Produce a vocalisation to express your fear.
	Freezing	You are soaked and it is very cold. Produce a vocalisation to express that you feel cold.
	Frustration	You were carrying a bottle of water and suddenly it fell on the floor and it spills out. Produce a vocalisation to express your frustration.
	Grief	You find out that a member of your family has died. Produce a vocalisation to express your grief.
	Pain	You burn your hand. Produce a vocalisation to express your pain.
	Sadness	Someone you care about is leaving the country and you will not see each other regularly. Produce a vocalisation to express your sadness.
Neutral	Misheard	Someone says something to you but you have not clearly heard what he/she just said. Produce a vocalisation to indicate that you have not heard what he/she has said.

ii) Survey

Prior to voice recording, we asked all vocalisers to complete a survey divided into three main sections: demographics, linguistic knowledge and media use (see Supplementary Methods for the online survey).

1- **Demographic data** included sex, age, height, professional experience in voice acting or singing, country of birth, current country of residence, other countries of residence when applicable as well as the time spent in each of them, reasons for living abroad and time spent with other citizens in those countries.

2- **Language knowledge** questions included number of spoken languages and proficiency in each (Native/Bilingual, Proficient, Intermediate or Basic), plus preferred language for communicating with family, friends, work colleagues and on social media.

3- **Media** questions included the average daily, weekly, monthly or annual time spent on social media. We also asked vocalisers to specify their preferred social platforms, and time spent watching television and listening to the radio.

4- **Technical questions** included the computer model as well as the microphone used to complete the study.

Data processing

All audio files were saved as uncompressed WAV files to a laptop computer. Full baseline words containing key vowels and volitional non-verbal vocalisations were extracted from the audio recordings using PRAAT software (version 6.3.03) [65]. When the vocalisations included several syllables, we extracted the vocalisation from the beginning of the first syllable to the end of the last syllable. Silences of 500ms were added before and after each vocalisation or string of baseline words to facilitate further sound processing. We obtained a total of 1028 baseline recordings (words containing five key vowels) and 14730 non-verbal vocalisations. Vocalisations were manually reviewed by two investigators to ensure that they were nonverbal (did not contain speech). A total of 84 non-verbal vocalisations were removed from the database because they contained words or identifiable primary interjections (words like *ouch* or *wow*; [64]. We note that because the investigators were not native speakers of many of the languages, it is possible that some primary interjections were retained. Thus, if we exclude the 3916 vocalizations whose speakers (N=262) did not give consent for online publication, there are a total of 10992 nonverbal vocalizations and 766 baselines available as WAV files.

Extraction of Acoustic Features

Automatic extraction of the acoustic features was performed in soundgen v.2.7.0 [66]. The acoustic structure of each non-verbal vocalisation and baseline was characterised by 10 key acoustic variables shown to be relevant descriptors of emotional content, valence, and arousal, as well as speaker identity and physical traits (e.g., sex, size) in human and other animal vocalisations [67,68].

These acoustic features included: mean fundamental frequency (f_0, perceived as voice pitch), peak frequency (the frequency with maximum spectral energy, in Hz), spectral centroid (median spectral centre of gravity of voiced segments, in Hz), loudness (subjective loudness, given in sones to control for variance in microphone distance), entropy (median Wiener

entropy, scaling from 0 to 1), harmonics-to-noise ratio (HNR, measure of harmonicity, in decibels), cepstral peak prominence (CPP, in dB), roughness (mean proportion of modulation spectrum of voiced segments within the roughness range of amplitude modulation 30–150 Hz, given as a percentage), duration (in seconds) and voicing (percentage of voiced frames, scaling from 0 to 1).

These spectro-temporal parameters were fully automatically extracted with the *pitch app()* function in soundgen v.2.7.0 [66]. Pitch tracking was manually checked for each audio file and corrected when the automatic algorithm made errors. However, consistent with standard practice, manual human checks were not performed on the other automated acoustic measures. We encourage researchers to perform and validate their own acoustic measurements using the provided WAV files. The CSV containing acoustic measures can be found on the project's repository (ECHO Data).

Translations

Study instructions, the survey, and recording tasks were designed in English and translated to each of the 21 other target languages by native speakers using a forward-back translation method [69]. Native bilingual speakers translated the text from the original language (i.e. English) into each of the target languages (e.g., Spanish, Malay, Tamil, etc.) and secondly, back translated from the target language to the original language. Finally, to ensure that the original meaning of each context vignette was fully preserved, the original and back-translated versions were compared and adjusted for any resulting differences. In the case of Spanish, French, Hungarian, Japanese, Mongolian and Filipino-Tagalog, two different bilinguals performed both translations, with the second bilingual (who back-translated into the source language) blind to the original source version. The English version of the survey and vignettes, along with all 21

translated linguistic versions, can be accessed on the project's online repository (ECHO Supplementary Material).

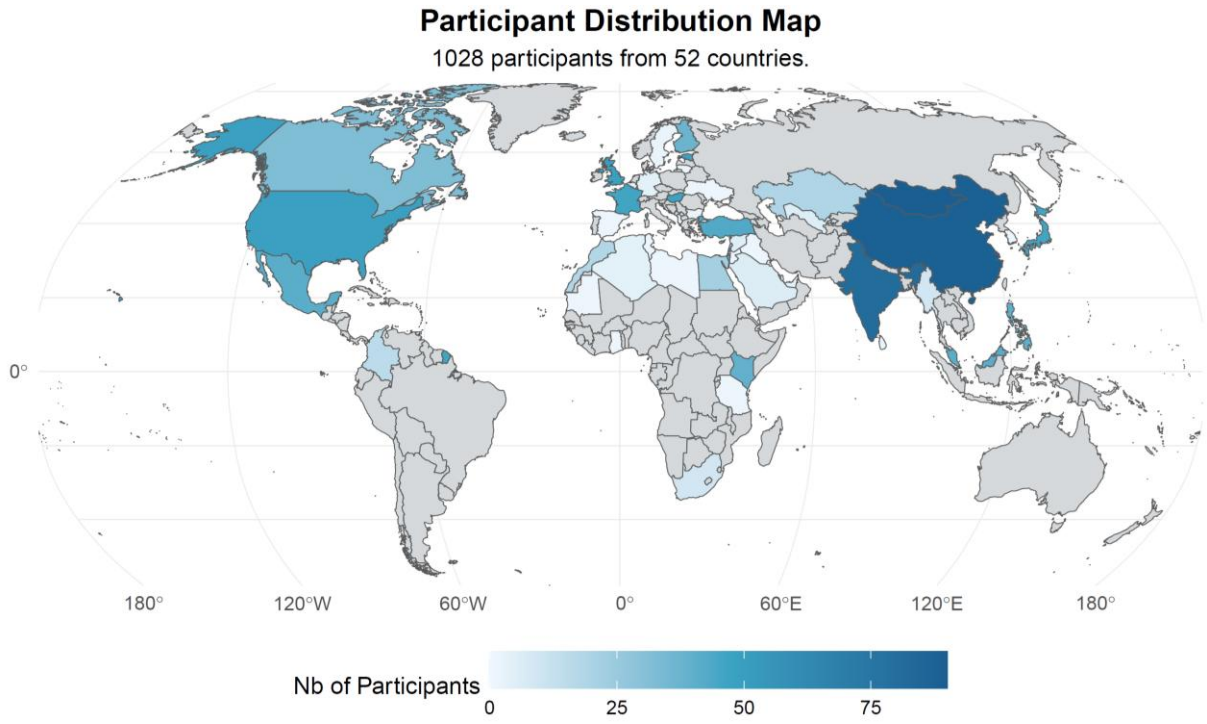


Figure 1. A world map showing the 52 countries of birth of the 1028 vocalisers whose data were included in the study. The colour scale represents the sample size for each country, with dark blue colours featuring higher sample sizes and light blue colours featuring smaller samples. Grey areas represent countries not covered by the data. From the 52 coloured countries, 22 countries had more than 10 vocalisers and 18 countries had at least 30 vocalisers.

DATA RECORDS

The complete ECHO Corpus is publicly accessible via an unrestricted repository in Zenodo (ECHO Corpus) and consists of WAV audio files, TXT text documents, CSV and XLSX data sheets and PNG image files. Voice recordings are organised into folders by language. The details of each folder's contents are provided below:

Acoustic Recordings. This main folder contains a total of 11758 WAV files produced by the participants who gave their consent, organised into sub-folders by language family and then language. Each language folder contains:

Baseline recordings. This folder will contain a WAV file for each vocaliser containing the baseline speech (words including the /a/, /e/, /i/, /o/ and /u/ vowels; e.g., EN_#846957_M_Baseline.wav).

Non-verbal vocalisations. This folder contains the WAV files of the non-verbal vocalisations for each of the 16 different contexts produced by each given vocaliser (e.g., EN_#846957_M_Achievement_NV.wav).

Data. This folder contains two CSV files and a README text file with Metadata.

- **Acoustic parameters** is a CSV file that provides detailed automated acoustic measurements of each vocalisations (e.g., fundamental frequency, duration, loudness, etc).

- **Survey data** is a CSV file that provides metainformation for each participant (e.g., gender, age, country of birth, etc).

- **README** file giving further information about the acoustic parameters and survey data files.

Supplementary Material. This main folder contains:

- **Translations.xlsx.** is a document listing all the linguistic items translated into the various languages.
- **Study.pdf** is a PDF version of the original full study in English
- **Quality assessment** is a CSV file based on the evaluations provided by two observers regarding the quality of the stimulus.

- **Affective context recognition** is a CSV file containing the results of the perception test on the affective context recognition of 5787 non-verbal vocalizations.
- **PNG** files includes spectrograms of each vocalisation.

TECHNICAL VALIDATION

The value of a large cross-cultural vocalisation corpus depends not only on its size, but also on its representativeness, recording quality, and the extent to which it captures biologically meaningful patterns. We therefore performed a series of complementary validation analyses. First, we characterise the composition of the ECHO corpus by quantifying the distribution of vocalisers across languages, language families, countries, and affective contexts. Second, we assess the acoustic quality of every recording and provide quality labels to facilitate future analyses. Third, we evaluate whether the vocalisations reliably communicate their intended affective contexts by conducting cross-cultural perception experiments with listeners from eight language groups on a subset of nearly 40% of the vocalisations. Finally, because fundamental frequency (perceived as voice pitch) is one of the most salient and extensively studied acoustic properties of the human voice [70], we examine whether its variation across affective contexts and language families is consistent with established theoretical predictions.

Vocalisers across languages

The distribution of vocalisers by language family varied, with the highest proportion of vocalisers in the Indo-European family (N=217 vocalisers, 21.1% of all vocalisers), followed by the Afro-Asiatic family (N=135, 13.1%), Uralic (N=132, 12.8%), Sino-Tibetan (N=108, 10.5%), Dravidian (N=94, 9.1%), Mongolic (N=91, 8.9%), Austronesian (N=84, 8.2%), Turkic (N=77, 7.5%), Japonic (N=48, 4.7%), and Atlantic-Congo (N=42, 4.1%).

The distribution of vocalisers also varied across the 22 languages. Three of these languages had more than 60 vocalisers: English (N=115), Standard Arabic (N=93), and Mongolian (N=91),

but most (fifteen languages) had between 30 and 60 vocalisers (Telugu, N=32 to Spanish N=56). Only four languages had less than 30 vocalisers: Tamil (N=28), Kazakh (N=26), Burmese (N=17) and Uzbek (N=4).

Overall, the ratio of men to women was relatively balanced (average 48% men across all languages), but ranged from 59.1% men for Standard Arabic to 30% men in Burmese. Vocalisers' ages were overall heterogeneous, with an average age of 32.7±11.2 years, varying across languages from 21.2 (SD=1.8) for Japanese vocalisers to 42.4 (SD=13.9) for English vocalisers.

Additional vocaliser details for each language and language family can be found in Table 1 and Supplementary Table S1.

Nonverbal vocalisations

From a total of 14646 non-verbal vocalisations, 7476 were produced by women (51.1%), 6971 by men (47.6%), 124 by individuals identifying as non-binary (<1%) and 75 by vocalisers who preferred not to specify a gender (<1%).

Importantly, non-verbal vocalisations were distributed similarly across contexts (Figure 2), with around 915 vocalisations per context: aggression (N=926), alarm (N=930), competition (N=869), fear (N=909), frustration (N=912), pain (N=941), sadness (N=905), grief (N=903), disgust (N=967), freezing (N=917), amusement (N=956), achievement (N=896), relief (N=925), positive surprise (N=929), and misheard (N=949). The only exception was sexual pleasure (N=812), as this vocalisation was not produced by Japanese, Darkhad, Dukha or Bulgan vocalisers for ethical reasons (Figure 2).

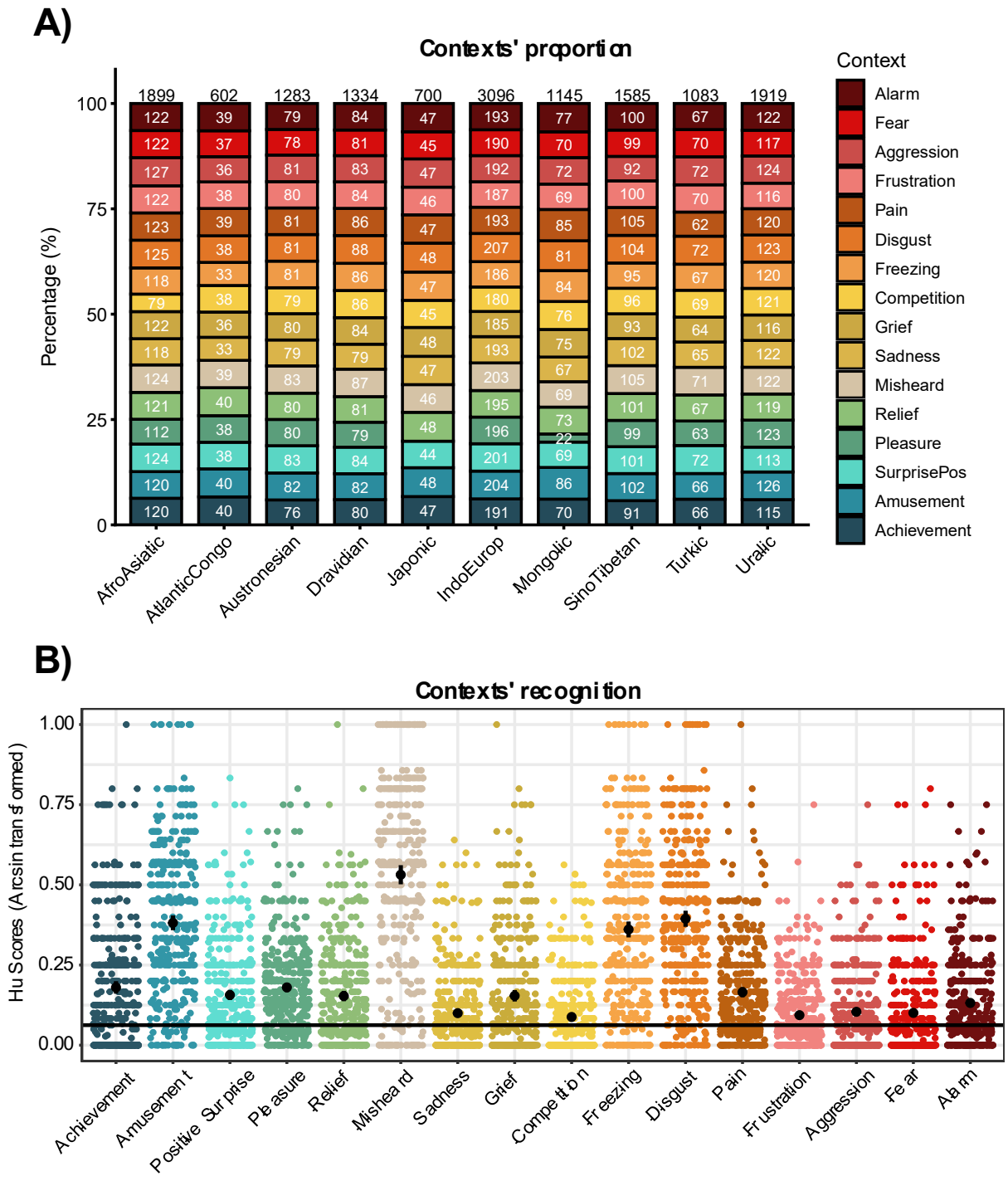


Figure 2: A) Distribution of the total number of vocalisations by language family and context. The x-axis shows each language family, while the y-axis represents the total percentage of vocalisations per context. Numbers within each cell represent the total number of vocalisations for each given context and language family. The number at the top of each bar

corresponds to the total number of vocalizations for each language family. Each colour represents one of the 16 contexts included in the study. The number of vocalisations is highly similar across contexts and this is true across language families, with the exception of sexual pleasure, which is not represented in the Japonic family nor for most Mongolian speakers. **B) Average Hu scores showing listener's judgements for each context controlling for individual perceptual biases.** Points represent each participant's Hu scores for each context, representing the proportion of correct responses. Every context was significantly well-classified above the chance level (0.06, black line).

Acoustic quality

Because nonverbal vocalisations were collected online in variable environments and with different devices (computer and microphone specifications are summarised in Supplementary Table S1), their sound quality naturally varied. Thus, two trained and experienced researchers (AG and TT) visually and acoustically inspected all 14646 vocalisations before dividing them into two categories: The High Quality (HQ) class included vocalisations that were recognised as relatively high quality, with little background noise, reverberation or clipping. The Medium Quality (MQ) class contained vocalisations that included some of these audible disturbances, but whose acoustic properties, such as fundamental frequency, could still be largely reliably measured. To ensure high inter-rater reliability, an inter-observer analysis based on 1416 vocalisations (around 10% of the dataset) revealed 84.8% agreement between the two observers.

Based on this quality check, our freely available acoustic dataset contains a total of 14646 non-verbal vocalisations, of which the 83.4% were identified as high quality and the 16.6% were identified as medium quality. Quality labels are included in the CSV file of acoustic parameters, and thus, data can be filtered according to quality.

Affective context recognition

To validate that these non-verbal vocalisations communicate affective states effectively across cultures, we tested how well listeners recognize all 16 emotional contexts from the vocalisations alone. We conducted a series of perception experiments with native speakers of eight different languages belonging to seven language families. A total of 400 listeners were recruited, including English and Spanish (Indo-European), Hungarian (Uralic), Mandarin Chinese (Sino-Tibetan), Filipino-Tagalog (Austronesian), Standard Arabic (Afro-Asiatic), Turkish (Turkic) and Tamil (Dravidian). Listeners were also recruited online through Prolific (Prolific Academic Ltd., London, UK) and compensated at the recommended rate of ~£7/h for their time. Details of the procedure of this study can be found in Supplementary Methods.

Each listener judged a total of 70 non-verbal vocalisations drawn randomly from a specific stimulus set containing vocalisations from four languages: those belonging to the same language as the listener, a different language from the same language family, or two languages from two different language families. In total, the pool of tested vocalisations contained 5787 non-verbal vocalisations (38.6% of the total dataset). We calculated unbiased hit rates (Hu scores) for each listener (Wagner, 1993) to control for individual biases in choosing some contexts more than others. These values were arcsine-transformed and included in a second Bayesian model with the same fixed and random effects. Finally, we ran Bayesian contrast comparisons, comparing arcsine-transformed Hu scores to arcsine-transformed corrected chance levels for each emotional context. Based on LOIC comparisons and parsimony principle, the best model was: brm (Hu score ~ Context + Age + (1|ID)) fitting a hurdle gamma family with a normal prior distribution (mean=0, sd=1).

The overall recognition rate of vocalisation contexts was 38%, which is 6 times higher than expected by chance (6.2%). The misheard context showed the highest recognition overall (73.2%) and competition the lowest (17.3%), however all contexts were recognized significantly above chance (See Supplementary Methods 2 for more details).

454 **Acoustic analysis: fundamental frequency**

455 Research has shown that fundamental frequency values span an exceptionally wide
456 range in human non-verbal vocalisations, from below 100 Hz in some aggressive roars [4,7] to
457 more than 2000 Hz in some fearful screams [62,71]. The range of fundamental frequencies (f_o)
458 in our corpus of non-verbal vocalisations paralleled this wide variety, ranging from 85 Hz to
459 2865 Hz. Importantly, consistent with the form-function framework [13], vocalisations
460 produced in the fear, alarm or achievement contexts typically had the highest mean f_o values
461 compared to the rest of the vocalisations (See Figure 3A). In contrast, as predicted, vocalisations
462 produced in the competition, freezing or frustration contexts systematically had the lowest f_o
463 values across language families.

464 Interestingly, although this form-function pattern was generally consistent across
465 cultures, there was also some variation across language families (Figure 3B). Regarding the
466 alarm context, for example, the mean f_o ranged from 363 ± 170 Hz in Austronesian languages
467 to 277 ± 103 Hz in Mongolic languages. Details of the acoustic measurements for each context
468 and language family can be found in Supplementary Table S2.

469

470

471

472

473

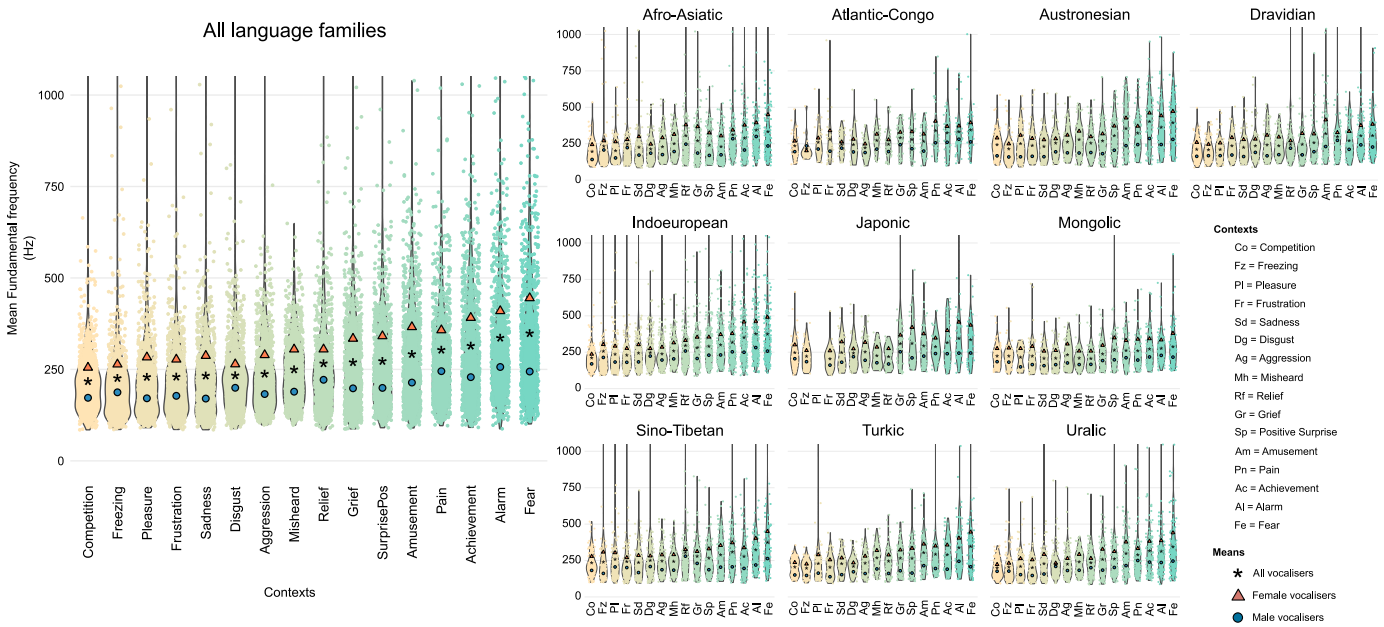


Figure 3. Fundamental frequency varied predictably by affective context, but also by language family: Central panel of violin plots representing mean fundamental frequency values by affective context, from lowest to highest, pooled across all languages (N=14646 vocalisations). Circles represent mean f_0 values for men, triangles for women and asterisks represent global mean f_0 values. Surrounding panels show violin plots representing mean fundamental frequency values for each language family, with the same context order retained as in the central panel. A similar predictable pattern is observed across language families: vocalizations produced in contexts of competition, freezing or frustration tend to have lower values than those produced in contexts such as fear, alarm, or achievement.

Dataset limitations

The online voice recording approach used here represents a novel methodological advance for vocal communication research, enabling large-scale, geographically distributed data collection while preserving experimental control and rich participant metadata. While online platforms

have long been used to obtain large cross-cultural samples for perception experiments, we show that remote methods can also be used to record human voices. At the same time, collecting recordings remotely inevitably introduces variation in recording conditions, which may affect sound quality and, potentially, the naturalness of vocal production. To address these challenges, we assessed the acoustic quality of every recording and labelled vocalisations suitable for robust acoustic analyses and perception experiments. In addition, we collected baseline speech recordings (vowels embedded in words) from every vocaliser. These baseline recordings allow researchers to compare acoustic variation across affective contexts within the same individual while controlling for recording characteristics that remain constant across that vocaliser's recordings. Reassuringly, our technical validation showed that approximately 83% of the vocalisations were of high acoustic quality, and cross-cultural perception experiments confirmed that the vocalisations communicated their intended affective contexts several times above chance.

A second limitation is that most vocalisers and listeners were recruited through the English-language crowdsourcing platform Prolific. As a result, many participants are likely to possess at least some knowledge of English in addition to their native language. To facilitate analyses of multilingualism and its potential effects, detailed information on the linguistic background of both vocalisers and listeners – including all languages spoken and proficiency levels – is provided in the *SurveyData.csv* file.

DATA AVAILABILITY

Vocalisations, acoustic data and material associated with this corpus is publicly available in Zenodo (https://doi.org/10.5281/zenodo.22754938).

(https://zenodo.org/records/22754938?token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjMxODdjY2ZiLWJjNTktNDU0MC04NGI1LWVkNDJjMWU2ZTM2MSIsImRhdGEiOnt9LCJyYW5kb20iOiIwZmZ

iOGNmZDM4MTE2MjU2MzY0MGE4NTFmMzhhMTY4YiJ9.vKnNZEc9dGBZ9pnnWtf8z1Vju3ro

f-Hs6hBSKBwzmWwSfZCOGq5QLCzTyqTpjIGDqTJn8wmbekrWorw0jkOBSg).

References

1. Pagel M. 2017 Q&A: What is human language, when did it evolve and why should we care? *BMC Biol* **15**, 64. (doi:10.1186/s12915-017-0405-3)

2. Anikin A, Valente D, Pisanski K, Cornec C, Bryant GA, Reby D. 2024 The role of loudness in vocal intimidation. *Journal of Experimental Psychology: General* **153**, 511–530. (doi:10.1037/xge0001508)

3. Amorim M, Anikin A, Mendes AJ, Lima CF, Kotz SA, Pinheiro AP. 2021 Changes in vocal emotion recognition across the life span. *Emotion* **21**, 315–325. (doi:10.1037/emo0000692)

4. Kleisner K *et al.* 2021 Predicting strength from aggressive vocalizations versus speech in African bushland and urban communities. *Phil. Trans. R. Soc. B* **376**, 20200403. (doi:10.1098/rstb.2020.0403)

5. Lavan N, Domone A, Fisher B, Kenigzstein N, Scott SK, McGettigan C. 2019 Speaker Sex Perception from Spontaneous and Volitional Nonverbal Vocalizations. *J Nonverbal Behav* **43**, 1–22. (doi:10.1007/s10919-018-0289-0)

6. Lima CF, Alves T, Scott SK, Castro SL. 2014 In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. *Emotion* **14**, 145–160. (doi:10.1037/a0034287)

7. Raine J, Pisanski K, Bond R, Simner J, Reby D. 2019 Human roars communicate upper-body strength more effectively than do screams or aggressive and distressed speech. *PLoS ONE* **14**, e0213034. (doi:10.1371/journal.pone.0213034)

8. Cowen AS, Elfenbein HA, Laukka P, Keltner D. 2019 Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist* **74**, 698–712. (doi:10.1037/amp0000399)

9. Daunay V, Reby D, Bryant GA, Pisanski K. 2025 Production and perception of volitional laughter across social contexts. *The Journal of the Acoustical Society of America* **157**, 2774–2789.

10. Kamiloğlu RG, Sun R, Bos P, Huber F, Attema JJ, Sauter DA. 2024 Tickling induces a unique type of spontaneous laughter. *Biol Lett* **20**, 20240543. (doi:10.1098/rsbl.2024.0543)

11. Sauter DA, Eisner F, Ekman P, Scott SK. 2010 Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2408–2412. (doi:10.1073/pnas.0908239106)

12. Sauter DA, Eisner F, Ekman P, Scott SK. 2015 Emotional Vocalizations Are Recognized Across Cultures Regardless of the Valence of Distractors. *Psychol Sci* **26**, 354–356. (doi:10.1177/0956797614560771)

13. Pisanski K, Bryant GA, Cornec C, Anikin A, Reby D. 2022 Form follows function in human nonverbal vocalisations. *Ethology Ecology & Evolution* **34**, 303–321. (doi:10.1080/03949370.2022.2026482)

14. Anikin A. 2020 A Moan of Pleasure Should Be Breathy: The Effect of Voice Quality on the Meaning of Human Nonverbal Vocalizations. *Phonetica* **77**, 327–349. (doi:10.1159/000504855)

15. Arnal LH, Flinker A, Kleinschmidt A, Giraud A-L, Poeppel D. 2015 Human Screams Occupy a Privileged Niche in the Communication Soundscape. *Current Biology* **25**, 2051–2056. (doi:10.1016/j.cub.2015.06.043)

16. Raine J, Pisanski K, Simner J, Reby D. 2019 Vocal communication of simulated pain. *Bioacoustics* **28**, 404–426. (doi:10.1080/09524622.2018.1463295)

17. Bryant GA, Aktipis CA. 2014 The animal nature of spontaneous human laughter. *Evolution and Human Behavior* **35**, 327–335. (doi:10.1016/j.evolhumbehav.2014.03.003)

18. Engelberg JW, Schwartz JW, Gouzoules H. 2021 The emotional canvas of human screams: patterns and acoustic cues in the perceptual categorization of a basic call type. *PeerJ* **9**, e10990.

19. Lingle S, Wyman MT, Kotrba R, Teichroeb LJ, Romanow CA. 2012 What makes a cry a cry? A review of infant distress vocalizations. *Curr Zool* **58**, 698–726. (doi:10.1093/czoolo/58.5.698)

20. Morton ES. 1977 On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds. *The American Naturalist* **111**, 855–869. (doi:10.1086/283219)

21. Massenet M, Philippe R, Pisanski K, Arnaud V, Barluet De Beauchesne L, Reynaud K, Mathevon N, Reby D. 2024 Puppy whines mediate maternal behavior in domestic dogs. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2316818121. (doi:10.1073/pnas.2316818121)

22. Arnal LH, Gonçalves N. 2025 Rough is salient: a conserved vocal niche to hijack the brain's salience system. *Philosophical Transactions of the Royal Society B: Biological Sciences* **380**, 1923.

23. Cornec C, Mathevon N, Pisanski K, Entani D, Monghiemo C, Bola B, Planas-Bielsa V, Reby D, Levréro F. 2024 Human infant cries communicate distress and elicit sex stereotypes: Cross cultural evidence. *Evolution and Human Behavior* **45**, 48–57. (doi:10.1016/j.evolhumbehav.2023.08.004)

24. Corvin S, Massenet M, Hardy A, Patural H, Peyron R, Fauchon C, Mathevon N. 2025 Nonlinear acoustic phenomena affect the perception of pain in human baby cries. *Phil. Trans. R. Soc. B* **380**, 20240023. (doi:10.1098/rstb.2024.0023)

25. Owren MJ, 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. (doi:10.1002/EVAN.1014)

26. Winkler SL, Bryant GA. 2021 Play vocalisations and human laughter: a comparative review. *Bioacoustics* **30**, 499–526. (doi:10.1080/09524622.2021.1905065)

27. Bryant GA *et al.* 2016 Detecting affiliation in colaughter across 24 societies. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4682–4687. (doi:10.1073/pnas.1524993113)

28. Bryant GA *et al.* 2018 The Perception of Spontaneous and Volitional Laughter Across 21 Societies. *Psychol Sci* **29**, 1515–1525. (doi:10.1177/0956797618778235)

29. Lavan N, Scott SK, McGettigan C. 2016 Laugh Like You Mean It: Authenticity Modulates Acoustic, Physiological and Perceptual Properties of Laughter. *J Nonverbal Behav* **40**, 133–149. (doi:10.1007/s10919-015-0222-8)

30. Todt D, Vettin J. 2005 Human laughter, social play, and play vocalizations of non-human primates: an evolutionary approach. *Behav* **142**, 217–240. (doi:10.1163/1568539053627640)

31. Bryant GA. 2022 Vocal communication across cultures: theoretical and methodological issues. *Phil. Trans. R. Soc. B* **377**, 20200387. (doi:10.1098/rstb.2020.0387)

32. Anikin A, Bååth R, Persson T. 2018 Human Non-linguistic Vocal Repertoire: Call Types and Their Meaning. *J Nonverbal Behav* **42**, 53–80. (doi:10.1007/s10919-017-0267-y)

33. Cordaro DT, Keltner D, Tshering S, Wangchuk D, Flynn LM. 2016 The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion* **16**, 117–128. (doi:10.1037/emo0000100)

34. Hawk ST, Van Kleef GA, Fischer AH, Van Der Schalk J. 2009 'Worth a thousand words': Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion* **9**, 293–305. (doi:10.1037/a0015178)

35. Laukka P, Elfenbein HA, Söder N, Nordström H, Althoff J, Chui W, Iraki FK, Rockstuhl T, Thingujam NS. 2013 Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Front. Psychol.* **4**. (doi:10.3389/fpsyg.2013.00353)

36. Pinheiro AP, Anikin A, Conde T, Sarzedas J, Chen S, Scott SK, Lima CF. 2021 Emotional authenticity modulates affective and social trait inferences from voices. *Phil. Trans. R. Soc. B* **376**. (doi:10.1098/rstb.2020.0402)

37. Fournier F, Briefer EF, Pisanski K, Levréro F. 2026 What do we know about vocal communication of emotion between different species of terrestrial tetrapods? *Biological Reviews* **101**, 944–969. (doi:10.1002/brv.70111)

38. Ackermann H, Hage SR, Ziegler W. 2014 Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences* **37**, 529–546. (doi:10.1017/S0140525X13003099)

39. Belyk M, Brown S. 2017 The origins of the vocal brain in humans. *Neuroscience & Biobehavioral Reviews* **77**, 177–193. (doi:10.1016/j.neubiorev.2017.03.014)

40. Jürgens U. 2002 Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews* , 24. (doi:10.1016/s0149-7634(01)00068-9)

41. Owren MJ. 2011 Human Voice in Evolutionary Perspective. *Acou. Today* **7**, 24. (doi:10.1121/1.3684225)

42. 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.

43. Pisanski K, Reby D, Oleszkiewicz A. 2024 Humans need auditory experience to produce typical volitional nonverbal vocalizations. *Commun Psychol* **2**, 65. (doi:10.1038/s44271-024-00104-6)

44. Matsumoto D, Seung Hee Yoo, Fontaine J. 2008 Mapping Expressive Differences Around the World: The Relationship Between Emotional Display Rules and Individualism Versus Collectivism. *Journal of Cross-Cultural Psychology* **39**, 55–74. (doi:10.1177/0022022107311854)

45. Sauter DA. 2013 The role of motivation and cultural dialects in the in-group advantage for emotional vocalizations. *Frontiers in psychology* **4**, 814.

46. Yoshie M, Sauter DA. 2020 Cultural norms influence nonverbal emotion communication: Japanese vocalizations of socially disengaging emotions. *Emotion* **20**, 513–517. (doi:10.1037/emo0000580)

47. Belin P, Fillion-Bilodeau S, Gosselin F. 2008 The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods* **40**, 531–539. (doi:10.3758/BRM.40.2.531)

48. Maurage P, Joassin F, Philippot P, Campanella S. 2007 A validated battery of vocal emotional expressions. *Neuropsychological Trends* , 5. (doi:10.7358/neur-2007-002-maur)

49. Holz N, Larrouy-Maestri P, Poeppel D. 2022 The variably intense vocalizations of affect and emotion (VIVAE) corpus prompts new perspective on nonspeech perception. *Emotion* **22**, 213.

50. Jiang Z, Long Y, Zhang X, Liu Y, Bai X. 2025 CNEV: A corpus of Chinese nonverbal emotional vocalizations with a database of emotion category, valence, arousal, and gender. *Behav Res* **57**, 62. (doi:10.3758/s13428-024-02595-x)

51. Lima CF, Castro SL, Scott SK. 2013 When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behav Res* **45**, 1234–1245. (doi:10.3758/s13428-013-0324-3)

52. Simon-Thomas ER, Keltner DJ, Sauter D, Sinicropi-Yao L, Abramson A. 2009 The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion* **9**, 838–846. (doi:10.1037/a0017810)

53. Laukka P, Elfenbein HA, Chui W, Thingujam NS, Iraki FK, Rockstuhl T, Althoff J. 2010 Presenting the VENEC corpus: Development of a cross-cultural corpus of vocal emotion expressions and a novel method of annotating emotion appraisals. In *Proceedings of the LREC 2010 Workshop on Corpora for Research on Emotion and Affect*, pp. 53–57. European Language Resources Association Valletta.

54. Henrich J, Heine SJ, Norenzayan A. 2010 The weirdest people in the world? *Behavioral and brain sciences* **33**, 61–83.

55. Bergan CC, Titze IR. 2001 Perception of Pitch and Roughness in Vocal Signals with Subharmonics. *Journal of Voice* **15**, 165–175. (doi:10/dnnwxj)

56. Sauter DA, Eisner F, Calder AJ, Scott SK. 2010 Perceptual Cues in Nonverbal Vocal Expressions of Emotion. *Quarterly Journal of Experimental Psychology* **63**, 2251–2272. (doi:10.1080/17470211003721642)

57. Szameitat DP, Alter K, Szameitat AJ, Wildgruber D, Sterr A, Darwin CJ. 2009 Acoustic profiles of distinct emotional expressions in laughter. *The Journal of the Acoustical Society of America* **126**, 354–366. (doi:10.1121/1.3139899)

58. Valente D, Magnard C, Koutseff A, Patural H, Chauleur C, Reby D, Pisanski K. 2025 Vocal communication and perception of pain in childbirth vocalizations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **380**, 20240009. (doi:10.1098/rstb.2024.0009)

59. Finger H, Goeke C, Diekamp D, Standvoß K, König P. 2017 LabVanced: a unified JavaScript framework for online studies. In *International conference on computational social science (cologne)*, pp. 1–3. University of Osnabrück Cologne.

60. Anikin A, Persson T. 2017 Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behav Res* **49**, 758–771. (doi:10.3758/s13428-016-0736-y)

61. Kamiloglu RG, Sauter D. 2021 Sounds like a fight: Listeners can infer behavioural contexts from spontaneous nonverbal vocalisations. (doi:10.31219/osf.io/rt49v)

62. Pisanski K, Raine J, Reby D. 2020 Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries. *R. Soc. open sci.* **7**, 191642. (doi:10.1098/rsos.191642)

63. Dingemanse M, Torreira F, Enfield NJ. 2013 Is “Huh?” a Universal Word? Conversational Infrastructure and the Convergent Evolution of Linguistic Items. *PLoS ONE* **8**, e78273. (doi:10.1371/journal.pone.0078273)

64. Ponsonnet M, Coupé C, Pellegrino F, Garcia Arasco A, Pisanski K. 2024 Vowel signatures in emotional interjections and nonlinguistic vocalizations expressing pain, disgust, and joy across languages. *The Journal of the Acoustical Society of America* **156**, 3118–3139. (doi:10.1121/10.0032454)

65. Boersma P, Weenink D. 2022 Praat: Doing phonetics by computer.

66. Anikin A. 2019 Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods* **51**, 778–792. (doi:10.3758/S13428-018-1095-7/FIGURES/4)

67. Briefer EF. 2012 Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology* **288**, 1–20. (doi:10.1111/j.1469-7998.2012.00920.x)

68. Charlton BD, Pisanski K, Raine J, Reby D. 2020 Coding of Static Information in Terrestrial Mammal Vocal Signals. In *Coding Strategies in Vertebrate Acoustic Communication* (eds T Aubin, N Mathevon), pp. 115–136. Cham: Springer International Publishing. (doi:10.1007/978-3-030-39200-0_5)

69. Kowal M. 2024 Translation practices in cross-cultural social research and guidelines for the most popular approach: Back-translation. *Anthropological Review* **87**, 19–32.

70. Karthikeyan S, Puts D. 2024 Voice Pitch. In *Encyclopedia of Sexual Psychology and Behavior* (ed TK Shackelford), pp. 1–11. Cham: Springer International Publishing. (doi:10.1007/978-3-031-08956-5_2442-1)

71. Anikin A, Canessa-Pollard V, Pisanski K, Massenet M, Reby D. 2023 Beyond speech: Exploring diversity in the human voice. *iScience* **26**, 108204. (doi:10.1016/j.isci.2023.108204)