HAL
open science

# Form follows function in human nonverbal vocalisations

Katarzyna Pisanski, Gregory A Bryant, Clément Cornec, Andrey Anikin, David Reby

## To cite this version:

Katarzyna Pisanski, Gregory A Bryant, Clément Cornec, Andrey Anikin, David Reby. Form follows function in human nonverbal vocalisations. Ethology Ecology & Evolution, In press, 10.1080/03949370.2022.2026482 . hal-03501113

**HAL Id: hal-03501113**

**https://hal.science/hal-03501113v1**

Submitted on 5 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Review

# Form follows function in human nonverbal vocalisations

KATARZYNA PISANSKI[1,2,*], GREGORY A. BRYANT[3], CLÉMENT CORNEC[1], ANDREY ANIKIN[1,4] and DAVID REBY[1]

[1]*ENES Sensory Neuro-Ethology lab, CRNL, Jean Monnet University of Saint Étienne, UMR 5293, 42023 St-Étienne, France*
[2]*CNRS French National Centre for Scientific Research, DDL Dynamics of Language lab, University of Lyon 2, 69007 Lyon, France*
[3]*Department of Communication, Center for Behavior, Evolution, and Culture, University of California, Los Angeles, USA*
[4]*Division of Cognitive Science, Lund University, 22100 Lund, Sweden*

*Corresponding author: Katarzyna Pisanski, ENES lab, 23 Rue du Dr Paul Michelon, 42023 Saint-Étienne, France (Email: katarzyna.pisanski@cnrs.fr).

Highlights

1. Converging evidence suggests that the acoustic *forms* of human vocalisations, like those of other animals, reflect their evolved social *functions.*
2. A much-needed proliferation in cross-cultural studies points to possible universals in the production and perception of human nonverbal vocalisations.
3. The human capacity to voluntarily control vocalisations greatly exceeds that of other primates, and likely played a role in the evolution of speech.
4. New digital technologies for experimental manipulation or complete synthesis of human and animal vocalisations open a new world of causal hypothesis testing.

Until recently, human nonverbal vocalisations such as cries, laughs, screams, moans, and groans have received relatively little attention in the human behavioural sciences. Yet these vocal signals are ubiquitous in human social interactions across diverse cultures and may represent a missing link between relatively fixed nonhuman animal vocalisations and highly flexible human speech. Here, we review converging empirical evidence that the acoustic structure ("forms") of these affective vocal sounds in humans reflect their evolved biological and social "functions". Human nonverbal vocalisations thus largely parallel the form-function mapping found in the affective calls of other animals, such as play vocalisations, distress cries, and aggressive roars, pointing to a homologous nonverbal vocal communication system shared across mammals, including humans. We aim to illustrate how this form-function approach can provide a solid framework for making predictions, including about cross-species and cross-cultural universals or variations in the production and perception of nonverbal vocalisations. Despite preliminary evidence that key features of human vocalisations may indeed be universal and develop reliably across distinct cultures, including small-scale societies, we emphasise the important role of vocal control in their production among humans. Unlike most other terrestrial mammals including nonhuman primates, people can flexibly manipulate vocalisations, from conversational laughter and fake pleasure moans to exaggerated roar-like threat

displays. We discuss how human vocalisations may thus represent the cradle of vocal control, a precursor of human speech articulation, providing important insight into the origins of speech. Finally, we describe how ground-breaking parametric synthesis technologies are now allowing researchers to create highly naturalistic, yet fully experimentally controlled vocal stimuli to directly test hypotheses about form and function in nonverbal vocalisations, opening the way for a new era of voice sciences.

KEY WORDS: voice, acoustic communication, speech evolution, nonverbal vocalisations, vocal control, cross-cultural, laughter, emotion.

Running head: Form follows function in human nonverbal vocalisations

## INTRODUCTION

Compared to speech, nonverbal vocalisations such as laughter, screams, roars, moans, and grunts remain remarkably understudied in humans. Yet these nonverbal vocal signals occupy a unique place in the human vocal repertoire (Anikin et al. 2018), differing from speech in important ways. Without the constraint of communicating intelligible linguistic information, nonverbal vocal signals can exploit a much broader acoustic soundscape. Screams, for instance, can reach extraordinarily high frequencies (Pisanski et al. 2020; Engelberg et al. 2021) and occupy a distinctive niche of acoustic roughness not observed in modal speech (Arnal et al. 2015). This makes screams excellent candidates for attention grabbing (Fitch et al. 2002). Human nonverbal vocalisations also involve evolutionarily conserved neural mechanisms that are not implicated in modal speech production: for example, unlike neutral speech, the production of spontaneous affective vocalisations like crying or laughter in humans involves the anterior cingulate cortex (ACC) with connections to the "emotional" limbic system, a homologous system shared by all mammals (Ackermann et al. 2014; Pisanski et al. 2016). Finally, human vocalisations show remarkable parallels both in their acoustic structure (form) and in their use (function) with the calls of other mammals, including infant distress cries (Lingle et al. 2012; Kelly et al. 2017; Koutseff et al. 2018), vocal threat displays (Morton 1977; Fitch et al. 2002; Raine et al. 2019), and play vocalisations in a range of species (Winkler & Bryant 2021), including our closest primate relatives (Davila Ross et al. 2009; Bryant & Aktipis 2014; Scott et al. 2014).

Most animal vocalisations are not arbitrary, but rather follow a form-function mapping. Form-function mapping refers to a systematic relationship between the acoustic form of a vocalisation and its ostensible social functions. Take for example a lion's roar. The low frequency, high amplitude, and perceptually rough features of a roar are designed to communicate threat and formidability, and a similar form-function mapping has recently been shown in the roar-like threat displays of humans (Raine et al. 2019; Kleisner et al. 2021). In contrast, distress vocalisations in a range of species are more often characterised by a comparatively higher fundamental frequency ($f_o$, perceived as pitch) and a pronounced frequency modulation pattern (Lingle et al. 2012). Here too, human distress vocalisations appear to be no exception (Koutseff et al. 2018; Pisanski et al. 2020; Engelberg et al. 2021).

Converging evidence thus strongly suggests that the acoustic "forms" of human vocalisations, like those of other animals, reflect their putative social communicative "functions", often adhering to general motivation-structural rules (Morton 1977) and sound symbolic associations (Ohala 1984; Hinton et al. 2006; reviewed in Owren & Rendall 2001; Pisanski & Bryant 2019). Such form-function mappings can provide critical clues into the evolutionary origins of nonverbal vocalisations, which emerged in the vocal repertoires of our ancestors before spoken language (Pisanski et al. 2016). Yet, despite their clear evolutionary relevance and ubiquity in our everyday social lives, human nonverbal vocalisations have received far less attention from researchers in the voice sciences than speech has. In speech studies, following the popularisation of the source-filter theory of speech production (Fant 1960), source-related and prosodic features of the voice such as $f_o$ (pitch) and filter-related features,

namely formant frequencies (resonances of the vocal tract), have been intensively studied (Fitch 2018; Pisanski & Bryant 2019 for recent reviews; Aung & Puts 2020). Yet these same important acoustic parameters have been largely ignored in the context of human nonlinguistic vocal sounds. Fortunately, this is quickly changing. The past decade has seen a new wave of researchers exploring the production and perception of nonverbal vocalisations in our own species, from mechanisms involved in their production and perception, to their evolved social functions.

In this review, although we discuss emerging research on various vocalisation types including screams, roars, and grunts, we focus largely on crying and laughter. First, this is because cries and laughs represent, by far, the two most extensively studied types of human nonverbal vocalisations. Second, both cries and laughs provide excellent examples of form mapping onto function. Third, both call types also beautifully demonstrate how humans can voluntarily manipulate nonverbal vocalisations to communicate a range of social nuances, such as in conversational laughter, or when children vocally embellish their level of distress with an exaggerated cry ('crying wolf'). We explain how this rare capacity to control our vocal output sets humans apart from most other terrestrial mammals. This includes nonhuman primates who show some vocal flexibility, but much less vocal control compared to humans. In a similar vein we review a small but growing body of literature examining similarities and differences in the production and perception of nonverbal vocalisations across diverse human cultures, further underscoring how a comparative approach, across species but also across human societies, can provide important insights into the origins and social functions of vocal behaviour. Finally, we conclude by highlighting how recent tools enabling the parametric manipulation and creation of synthetic yet naturalistic nonverbal vocalisations are likely to revolutionise experimental research in the voice sciences.

## THE CRY: A BUILT-IN SURVIVAL MECHANISM

Studies on a wide range of mammal species have investigated how the affective dimensions of arousal and valence are encoded in human distress signals, often focusing on infant cries (Lingle et al. 2012; Kelly et al. 2017). These studies provide compelling evidence for cross-species similarities in the acoustic form of cries, and also in listeners' sensitivity to them. For example, two studies by Lingle and colleagues show that infant distress vocalisations from diverse mammal species share a similar chevron shape and frequency modulation pattern (Lingle et al. 2012), and that deer mothers respond to the distress cries of infants from a range of species owing to this shared acoustic structure (Lingle & Riede 2014). Infant distress vocalisations in mammals appear to share a critical function: to elicit aid from caregivers.

Is this also the case for the human baby cry? Numerous studies have shown that human infant cries affect the neural (Laurent & Ablow 2012; Venuti et al. 2012; Messina et al. 2016; Bornstein et al. 2017; Witteman et al. 2019), physiological (Frodi et al. 1981; Boukydis & Burgess 1982; Fleming et al. 2002) cognitive and behavioural (Gustafson & Harris 1990; Bornstein et al. 1992; Yoo et al. 2019) responses of human adult listeners in ways that align with their ostensible function to elicit aid. For example, in a large comparative study, Bornstein et al. (1992) show cross-culturally shared

behavioural responses among mothers exposed to the cries of their own babies, namely a shared tendency to pick-up and speak to the crying infants. The authors also show concordant activation in brain areas linked to movement, speech processing, and care-directed cognition including the supplementary motor area (SMA), inferior frontal cortex, and superior temporal regions. Also notable among neural responses to infant cries, in vivo functional neuroimaging studies reveal activation in concordant brain structures related to cognitive control and attention (Swain et al. 2007; Swain 2011). Indeed, reduced performance in conflict tasks (Dudek et al. 2016), mental calculation tasks (Morsbach et al. 1986; Chang & Thompson 2011), and memory tasks (Hechler et al. 2015) effectively shows how the infant cry can engage the attention of the listener, reduce concentration, and ultimately disrupt behavioural performance in parents and in non-parents alike (Chang & Thompson 2011; Dudek et al. 2016).

Often triggered by physical discomfort, pain, hunger, or separation from parents, human baby cries have traditionally been studied as distinct types of vocalisations associated with a specific context (Gustafson & Harris 1990 for review). Although specific cry types may exist (e.g., 'siren cry', a persistent and periodic cry resembling a siren alarm sound; Bellieni et al. 2004), emerging evidence suggests that the human infant cry is a graded signal whose acoustic features vary dynamically along a continuum according to the intensity of the negative arousal state of the infant, from mild discomfort to pain (Gustafson et al. 1999 for review). The $f_o$ of human babies' cries can increase sharply once a threshold of pain is exceeded (Bellieni et al. 2004), but information about the ostensible distress level of the infant appears more reliably encoded in the nonlinear acoustic phenomena of the cry (Fitch et al. 2002) such as sidebands, subharmonics, frequency jumps, vibrato-like frequency modulation, and the most perceptually aversive type of all, deterministic chaos (Anikin et al. 2020), as illustrated in Fig. 1. These nonlinear phenomena, caused by irregular or chaotic vocal fold vibration, give the cry a perceptually rough and harsh quality, and appear to increase with arousal or distress in human infant cries (Leger et al. 1996; Tiezzi et al. 2004; Facchini et al. 2005; Koutseff et al. 2018; Yoo et al. 2019). Indeed, not only are nonlinear phenomena more prevalent in babies' pain cries (e.g., produced during a vaccination event) than in mild discomfort cries (e.g., produced during a bath), nonlinearities also increase with distress level within each of these distinct contexts (Koutseff et al. 2018). In addition to their roughness and high pitch, cries are typically long, loud, and high or 'bright' in timbre – all hallmark features of salient acoustic events (Kaya & Elhilali 2014; Huang & Elhilali 2017).

There is thus strong evidence that human preverbal infant cries are acoustically 'designed' to directly impair normal cognitive functioning and, in turn, elicit behavioural responses from caregivers who are motivated to stop the aversive cry signal, thereby attending to the crying infant. Most importantly, cries do this by being extraordinarily variable and unpredictable. The 'unpredictability hypothesis' posits that the presence of nonlinear phenomena in mammalian vocalisations makes them less predictable and therefore difficult to habituate to and ignore (Fitch et al. 2002). While this hypothesis has found support in research on nonhuman mammals (Blumstein & Recapet 2009; Townsend & Manser 2011; Reby & Charlton 2012), it remains to be directly tested in human crying. Nevertheless, with the advent of resynthesis techniques that now allow for direct experimental

manipulation of nonlinear phenomena in human vocalisations, as described below, this will soon be investigated. More broadly, infant cries constitute a great experimental model for studying a general principle in vocal communication: high-intensity vocalisations such as cries and screams (Arnal et al. 2015; Engelberg et al. 2021) appear to be "designed" to attract attention and prevent habituation in listeners by means of exploiting basic properties of auditory perception (Anikin 2020).

<< Fig. 1 in separate attachment >>

Fig. 1. – Nonlinear acoustic phenomena in human baby cries. Top of each panel: spectrogram (x-axis = time in seconds; y-axis = frequency in kilohertz); bottom of each panel: oscillogram (x-axis = time; y-axis= relative amplitude). (A) Cry recorded during a bath (mild discomfort context). (B-D) Cries recorded during vaccination events (painful context). While nonlinear phenomena are observed in both discomfort and pain cries, pain cries are characterised by a higher proportion, including of sidebands, periodic vibrato-like frequency modulation (VFM), subharmonics, frequency jumps, and deterministic chaos, the latter of which represents the most severe and acoustically salient type of nonlinear phenomenon. Figure prepared with the R package *seewave* (Sueur et al. 2008).

## LAUGHTER: FROM REFLEX TO SOCIAL TOOL

While we are born crying, laughter does not emerge until around 4 months of age in human infants (Sroufe & Wunsch 1972). Human laughter evolved from social play vocalisations, and thus like the human baby cry, the acoustic structure of human spontaneous laughter shares many commonalities with the laughter-like calls of other animals (Vettin & Todt 2005). Homologous versions of vocal play signals exist across numerous extant species, with a recent survey of the literature counting at least 65, including three species of birds (Winkler & Bryant 2021). Play vocalisations are thought to serve the important function of communicating benign intent during social play interactions in which animals practice skills needed in adult life, such as predator-prey simulations and play fighting. During these play encounters, individuals act in ways that might appear to be threatening. Biting, growling, and lunging behaviours can quickly escalate into aggression unless there is a reliable indication, through visual or auditory signals, that the intent is non-threatening. Evidence is growing that human laughter is derived from such a signal, but has also evolved into a suite of inter-related signals that are much more complicated, helping people communicate subtle emotional meanings and intentions using a variety of laughter types that are highly sensitive to context, and can even be negatively valenced (Scott et al. 2014; Winkler & Bryant 2021). Indeed, human social life is rich and multifaceted, involving cooperation in the absence of kinship, extended networks of affiliation, and complex social cognition wherein laughs may be used to communicate

anything from friendship and empathy to sarcasm and malice (Scott et al. 2014). Our navigation of the social landscape necessitates sophisticated social cognitive machinery and complex ostensive signalling often involving language (Scott-Phillips 2015). This is where the evolution of laughter takes an interesting turn.

Most uses of language occur in conversation. During conversation, interlocutors use a variety of multimodal behaviours to help coordinate their talk: we gesture with our hands and bodies, modulate our voices prosodically, and make facial expressions. But we also inject nonverbal vocal signals into our speech in systematic ways. Laughter is a prime example of such behaviour. We laugh in ways that "punctuate" our speech (Provine 1993), such as during conversational turns, or just after a particular utterance that might require additional signalling to accurately convey intent. This type of 'conversational' laughter, which humans produce volitionally, sets human laughter apart from the more spontaneous play signals of other animals. It allows humans to voluntarily produce laughs, even 'fake' laughs, to intentionally communicate a range of meanings, motivations, and emotions, from social cohesion to spite (Scott et al. 2014; Bryant 2020). While spontaneous human laughter is perceptually indiscriminable from nonhuman primate vocalisations when slowed down and pitch adjusted (Bryant & Aktipis 2014), voluntarily produced laughter in humans has a unique structure: it is typically shorter, with slower individual calls, more voiced, and lower in pitch (Bryant & Aktipis 2014; Lavan et al. 2015). In other words, volitional laughter is more speech-like compared to spontaneous laughter.

Volitionally controlled vocalisations likely emerged to fulfil new functions in the context of more recently evolved communicative behaviours, such as conversational contexts that involve language. For example, when talking, laughter can function as a type of play signal that works in largely the same way that play signals function in nonhuman social interactions. We may verbally "attack" with a taunt or joke, and then indicate our non-serious intent with a laugh. In this way, laughter has retained its original function shared across many species, but it now occurs in a much more complex pragmatic context, and potentially imparts rich meaning as it interacts with other complex social communicative behaviours and intentions. Human laughter also takes on additional functions in turn-taking, backchanneling, and various discourse functions (Bryant 2020).

Other vocalisations can be analysed similarly, such as the way we cry, scream, and moan during discourse or in various social or intimate interactions. In humans, such vocalisations are often produced volitionally and used strategically as social tools in ways that are frequently different from their spontaneous counterparts. This hints at the possible evolutionary roots of vocal control, as described below, but also opens up the possibility that, while basal acoustic features of nonverbal vocalisations may have evolved similarly across the human lineage, those readily under vocal control may be more susceptible to sociocultural variation. In other words, while affective vocalisations are likely to retain broadly similar features across human populations, people might nevertheless volitionally modulate their nonverbal vocalisations in different ways depending on varying social norms and cultural experiences.

DO HUMAN VOCALISATIONS SHOW CULTURAL UNIVERSALS?

The form-function approach provides a solid framework for making predictions about the universality and cultural variation in vocal signals (Bryant 2021a). If the acoustic structure of a vocalisation automatically affects receivers' physiology and cognition, as in the case of a baby's cry affecting parental attention or an abrupt yell to stop the unwanted behaviour of a young child, then we might expect this relationship in acoustic form and communicative function to occur across cultural boundaries (Fernald 1992; Bryant & Barrett 2007). Conversely, the highly variable sociocultural and pragmatic rules associated with how people express emotion to one another could cause notable variation in the frequency and manner in which these expressions manifest themselves, particularly given how easily people can control vocal output.

An important development in the study of human behaviour across cultures is the recognition of a need for diversity in human participant samples (Rad et al. 2018). Most studies in the behavioural sciences have traditionally relied on WEIRD (Western, Educated, Industrialized, Rich, Democratic) samples, and there are many reasons to be suspicious of the generalizability of findings based solely on such participants (Henrich et al. 2010). Recently the WEIRD concept has received much attention, and there are early efforts across many disciplines to ameliorate the problem. Yet this has caused a new set of issues to emerge, such as: (i) prioritising exotic, small-scale societies that represent an extremely small proportion of current living humans, (ii) perpetuating long standing misunderstandings about the "ancestral slot" that indigenous people often fill for Western researchers, including erroneously treating them as proxies for ancestral humans, and (iii) neglecting populations that represent the bulk of people living today (i.e., rural, uneducated people living on the edge of large socioeconomic environments) (Barrett 2021).

Much work exploring the production and perception of nonverbal vocal characteristics across languages and cultures has focused on WEIRD-like samples in European, North American and Asian societies (e.g., Pell et al. 2009). For example, many studies have examined how emotion is perceived from vocal signals, often recorded from actors manipulating prosodic features of their speech like the pitch, loudness, rhythm and timbre (Banse & Scherer 1996; Pisanski & Bryant 2019 for reviews; Bryant 2021b). Overall, this research shows that people can often accurately identify emotion categories across languages and cultures from the nonverbal parameters of speech, though there is a well-documented in-group advantage in which accuracy decreases as a function of cultural similarity between speaker and listener (Laukka & Elfenbein 2021). In the past decade researchers have also examined nonverbal voice perception in small-scale societies that have little exposure to Western media (Bryant & Barrett 2007, 2008; Sauter et al. 2010; Bryant et al. 2012; Gendron et al. 2014). While most of these studies focused on emotional speech, Sauter et al. (2010), in a seminal paper, showed that Himba people living in isolated Namibian villages could correctly classify several emotions from nonverbal vocalisations produced by British adults, and vice versa, particularly negative emotions, or joy signalled by laughter.

Emotion recognition is just one domain of vocal signalling that has been explored across disparate cultures, and is perhaps one of the most difficult given the complexities of emotion, culture, and vocal production. Yet, despite the obvious affective basis of most nonverbal vocalisations, few studies have sought to test for culturally universal relationships between acoustic features of nonverbal vocal stimuli and listeners' perceptual judgments (i.e., form-function mapping), with a few notable exceptions. In two large-scale studies, Bryant and colleagues demonstrated that spontaneous and volitional laughter can be distinguished by listeners across more than 20 societies (Bryant et al. 2018), as can detecting friends and strangers from colaughter (Bryant et al. 2016). In these laughter studies, listeners across quite different societies relied on similar acoustic features to make their judgments, namely features associated with speaker arousal. Human aggressive roar-like vocalisations also appear to share a common function of maximising signals of physical formidability such as strength, as shown in three distinct samples of British drama students (Raine et al. 2019), urban-dwelling Cameroonian adults and nomadic Hadza hunter-gatherers living in the Tanzanian bushlands (Kleisner et al. 2021). Data recently collected in the Democratic Republic of the Congo by author C. Cornec, currently being prepared for publication, also shows that distress information encoded and perceived in babies' cries shares extraordinary similarities between two very distinct cultures, Congolese and French. Taken together this body of work represents a promising rise in large-scale cooperation between researchers in which data from multiple geographic sites with diverse participants are combined to provide new insight into the complex ways that vocal properties relate to signalling functions.

## VOCAL CONTROL: THE MISSING LINK?

Emerging evidence for cross-cultural universals in form-function mappings as described above suggests that, like in nonhuman animal calls, there may exist strong fixed components of acoustic structure in human nonverbal vocalisations. Yet, as seen with volitional human laughter, people can adeptly alter the nonverbal components of their vocalisations 'on demand' to communicate or even exaggerate traits and motivational states. This advanced capacity for volitional vocal control in humans is possible owing to direct monosynaptic connections between the motor cortex (M1) and brainstem motoneurons that control the laryngeal muscles and vocal articulators such as the jaw, tongue and lips; neural connections that are thought to be indirect (via the reticular formation/brainstem) in nonhuman primates. In contrast, according to the dual-pathway model of vocal control, the production of spontaneous affective vocalisations in humans and other mammals more readily involves the limbic pathway, including the periaqueductal gray (PAG) and anterior cingulate cortex (ACC), brain regions that are part of an evolutionarily older and more basal neural system (reviewed in Ackerman et al. 2014; Fitch 2018; Pisanski et al. 2016).

While the neuroanatomical mechanisms involved in voluntary and involuntary vocal production in humans have been intensively studied, it remains less clear when and why the capacity to produce volitional vocal signals evolved in our lineage. From a functional perspective, volitional

vocal control could have conferred definite advantages in prelinguistic vocalisers. In the absence of linguistic conventions, the ability to mimic animal, environmental or even affective vocalisations could have been a highly functional means to exchange information for coordinating foraging, hunting and social interactions in our ancestors. The rich communicative potential of vocal iconicity could thus have provided evolutionary pressures for increasing vocal control, and there is evidence for such functionality in present-day humans. For instance, Ćwiek et al. (2021) show that novel vocalisations representing a variety of actions (e.g., cook), objects (e.g., water) and living things (e.g., tiger) can be understood across a diverse range of 28 human cultures and 12 language groups, suggesting that in the absence of shared language, people can use volitional nonverbal vocalisations to communicate meaning. In the context of hunting, deceptive mimicry of animal calls is used by a variety of hunter-gatherer tribes such as the Mbendjele Pygmies in northern Congo (Lewis 2009; Knight & Lewis 2017). Whereas in social contexts, the ability to exaggerate (or even completely fake) the expression of emotional states like pain or pleasure could likewise have clear functional benefits if it effectively elicits a favourable response from listeners. As an example, volitionally produced roar-like vocalisations in both European and African samples of men and women increase the perceived body size and physical strength of vocalisers relative to screams and distressed speech (Raine et al. 2019) or neutral speech (Kleisner et al. 2021), and this may be beneficial in competitive contexts. Similarly, volitional pain vocalisations of increasing intensities elicit corresponding pain ratings in listeners (Raine et al. 2018), suggesting that people may indeed be able to effectively exaggerate their pain level to attract aid from others.

Humans thus clearly possess a capacity to readily modulate nonverbal vocalisations in ways that could be functionally beneficial. Vocal control is also observed in songbirds and some other mammal species, including cetaceans and seals (Fitch 2000), but can other primates voluntarily control their vocal output? On one hand, there is mounting evidence that nonhuman primates may have more control over their vocal output than previously thought (reviewed in Pisanski et al. 2016; Seyfarth & Cheney 2018). For example, wild chimpanzees appear capable of voluntarily inhibiting various vocalisations and preferentially or flexibly producing vocalisations in specific contexts, such as alarm calls or food grunts (see e.g., Laporte & Zuberbühler 2010; Crockford et al. 2012; Schel et al. 2013). Flexibility in vocal production has also been observed in other great ape species, including wild bonobos (Clay et al. 2015; Cornec et al. in press) and both wild and enculturated orangutans (Lameira et al. 2013a, 2013b). Research emerging in the past decade is also challenging the traditional dual-pathway model of vocal control in mammals, pointing to a greater degree of cross-talk between the cortical and limbic pathways typically thought to be respectively involved in volitional and spontaneous vocal production (e.g., Belyk & Brown 2016; Belyk et al. 2016; Ludlow 2015; Wattendorf et al. 2013; reviewed in Ackermann et al. 2014, Pisanski et al. 2016, and Scott 2021).

On the other hand, humans are arguably the only primate that can easily voluntarily modulate the acoustic structure of our vocalisations “on demand”, frequently producing them in the complete absence of endogenous or exogenous stimuli that would normally trigger their production in nonhuman mammals (Ackermann et al. 2014; Pisanski et al. 2016; Fitch 2018). Pisanski et al. (2016)

argue that selection pressure for these clearly functional vocal abilities in humans could have played a crucial role in the emergence of displacement and vocal control in our ancestors, two essential precursors of human speech (Fitch 2018). This approach positions volitional human nonverbal vocalisations as the "missing link" between nonhuman animal calls and full-blown human speech, further underscoring the importance of their empirical investigation. Specifically there is a strong need for comparative cross-disciplinary studies that combine phylogenetic analysis with neural and behavioural measures of vocal control, particularly manipulation of the source and filter, to clarify the extent of vocal control abilities across species, including but not limited to nonhuman primates.

## (RE)SYNTHESIS: A NEW ERA OF VOICE SCIENCES

Increasingly the study of human nonverbal vocalisations spans a broad range of disciplines, including bioacoustics, ethology, computer sciences, psychology and linguistics, with a rise in comparative studies examining vocal interactivity within and among humans, nonhuman animals, and machines (reviewed in Moore et al. 2016). This is both an opportunity, because tools and techniques can be borrowed and adapted from several disciplines, and a challenge, because none of those tools necessarily fit perfectly. The prevalent approach has been to acoustically analyse a large number of vocalisations in order to establish which acoustic characteristics correlate with particular meanings (e.g., with high arousal or perceived authenticity). While this correlational approach is valuable for initial exploration of the acoustic space, allowing researchers insight into which acoustic parameters might be most relevant, the critical problem is that many acoustic features co-vary, and large sample sizes are necessary to tease apart their relative roles. A more powerful approach is to systematically manipulate acoustic characteristics one at a time, or in methodical combinations, to test their independent causal effects on listeners and thus gain deeper insights into their putative evolved functions. Here we discuss new tools for experimental manipulation or complete synthesis of human vocalisations that promise to open a new world of hypothesis testing.

Independent digital manipulation of the vocal source ($f_o$, pitch) and filter (formants) in speech and later in nonhuman animal vocalisations greatly contributed to the advancement of animal communication research. Indeed, this step was critical to establishing the independent contribution of pitch and formants in conveying various biosocial traits like body size, dominance, or mate-quality in humans (reviewed in Pisanski & Bryant 2019; Aung & Puts 2020) and other animals (reviewed in Taylor & Reby 2010; Taylor et al. 2016; Charlton et al. 2020) including nonhuman primates (Fitch & Hauser 1995). Digitally manipulating these voice frequencies to study their effects on listeners is often achieved with a technique known as PSOLA, implemented in many popular software packages such as Praat (Boersma & Weenink 2021). While most audio manipulation techniques like PSOLA were first developed for speech or music, many are fully applicable to nonverbal vocalisations. Frequency-domain methods of source-filter separation with phase vocoders, while mostly developed in the music industry, potentially have several advantages over PSOLA for manipulating nonverbal vocalisations because they are not dependent on accurate pitch tracking, can deal with biphonation, and are capable of separating voices from percussive and other non-harmonic noises. Some phase

vocoders can perform pitch shifting and filter manipulations in real time, as well as change some aspects of voice quality by means of sophisticated techniques such as spectral warping and pitch-synchronous amplitude modulation (Burred et al. 2019; Arias et al. 2020). This is a very promising method to apply to nonverbal vocalisations in future studies. Another interesting technique is morphing, or gradually changing one sound into another. Morphing was used to study categorical perception and individual recognition in macaque vocalisations (Chakladar et al. 2008; Furuyama et al. 2017) with the STRAIGHT algorithm (Kawahara 2006). There are already some reports of morphing human emotional vocalisations for the purpose of neuroimaging (Salvia et al. 2014) and testing for categorical perception (Woodard et al. 2021), but morphing vocalisations remains quite challenging and uncommon compared to the ubiquitous use of morphing techniques in the visual domain. This is also an important gap to fill in future studies.

For ultimate control, nonverbal vocalisations can be synthesized de novo. Regrettably, the abundant tools for speech synthesis are largely unsuitable because they seldom provide a simple and transparent mechanism for changing acoustic parameters one at a time or even for synthesising non-speech sounds (Anikin 2019; Arias et al. 2020). There have been a few attempts to adapt concatenative speech synthesis to create synthetic laughs and other affect bursts (Urbain et al. 2013; El Haddad et al. 2016), but this method relies on splicing together pre-recorded samples and is not a viable method for testing acoustic hypotheses. Some intrepid researchers manually programmed sine-wave synthesis to create pure-tone mammalian (Snowdon & Pola 1978; DiMattina & Wang 2006) and avian (Margoliash 1983) calls, but most vocalisations are too complex to be synthesised without dedicated software.

Fortunately, such software is finally becoming available. Some solutions are developed for specific applications such as real-time synthesis of simple animal-like vocalisations in social robots (Moore & Mitchinson 2017) or songbirds (Zúñiga & Reiss 2019), others are more general-purpose, such as the R packages *seewave* (Sueur et al. 2008) and *soundgen* (Anikin 2019) or the Matlab-based *SynSing* (Tanner et al. 2020). These emerging sound resynthesis platforms allow researchers to test the causal effects of a broad range of acoustic parameters, many of which traditionally could not be experimentally manipulated. For example, nonlinear acoustic phenomena are difficult not only to synthesize, but even to detect automatically in audio recordings. As a result, most evidence of their perceptual effects has been indirect. Using *soundgen* (Anikin 2019), researchers can now experimentally add different nonlinearities to synthetic nonverbal vocalisations, as illustrated in Fig. 2. Using this method, the authors have recently shown that nonlinearities directly enhance the perceived intensity of negative emotions in synthesized human vocalisations, such as roars and screams (Anikin et al. 2020), and also cause vocalisers to sound more aggressive and physically larger (Anikin et al. 2021). The availability of specialised tools for parametric synthesis of non-speech sounds thus offers exciting new opportunities for researchers interested in testing specific hypotheses about the acoustic code in animal and human nonverbal vocalisations.

<< Fig. 2 in separate attachment >>

Fig. 2. – Workflow for modifying nonlinear phenomena in speech or nonverbal vocalisations with a hybrid synthesis-resynthesis method implemented in *soundgen* (Anikin 2019). The original recording (panel 1) is separated into an amplitude envelope, a smoothed spectral envelope representing formants, and a pitch contour (panel 2). The pitch contour is used to synthesize a "buzz" with the original intonation and any required combination of nonlinear phenomena (in the above example, we synthetically added deterministic chaos) (panel 3 top). Smoothed spectral and amplitude envelopes are then "transplanted" onto the synthetic "buzz", thereby recreating the original formant and temporal structure, but now with a modified voice quality (panel 3 bottom). Code and audio examples are available in Anikin et al., (2021). Note that frequency (y-axis 0.5-10 kHz) is plotted on a logarithmic scale for improved visualisation of the pitch contour.

## CONCLUSION

In this review, we describe growing, converging evidence suggesting that the acoustic forms of human nonverbal vocalisations, like those of other animals, reflect their evolved functions. Human infant cries elicit care by exploiting the perceptual sensitivities of parents, aggressive roar-like vocalisations maximize physical strength to rivals, and aroused co-laughter between friends communicates their companionship to bystanders across human cultures. These examples illustrate how the acoustic structures of human vocalisations vary systematically and predictably with their intended communicative function, and thus appear largely homologous to those of other mammals. At the same time, human vocalisations differ in critical ways from the affective calls of our closest living primate relatives, most notably in terms of vocal control. On one hand, cross-cultural studies point to possible universals in the production and perception of vocalisations across human societies. On the other hand, the human capacity to voluntarily modulate our vocal output, which is much more advanced in our own species than in any other extant primate, introduces some degree of cultural variability, and is likely to have played a role in the early evolution of speech abilities.

This review demonstrates the broad insight that can be gained by adopting a comparative inter-disciplinary framework in the voice sciences. We show how the cross-cultural ubiquity and unique ontogenetic and phylogenetic positioning of human nonverbal vocalisations, which emerge before speech both in human development and in our ancestral past, makes them an excellent candidate for answering key questions about the evolution of vocal communication, and we show how new digital technologies enabling previously impossible experimental manipulation of these vocalisations will increasingly support this endeavour by allowing researchers to causally test predicted form-function mappings. Yet despite the steady improvements in the range of available digital manipulations and the authenticity of modified or synthesized vocalisations described here, the adoption of these tools remains sluggish as they require considerable expertise in both acoustics and programming. To fully realise the potential of the powerful new tools for voice manipulation and synthesis, it is essential to close the technological gap by improving the usability of the often arcane software solutions while strengthening cross-disciplinary links between the computer and life sciences.

## DISCLOSURE STATEMENT

The authors report no conflicts of interest.


## FUNDING

D. Reby was supported by the French National Research Agency (ANR) under IDEXLYON fellowship (ANR-16-IDEX-0005); A. Anikin was supported by grant 2020-06352 from the Swedish Research Council; C. Cornec was supported by a fellowship funded by the ANR (ANR-19-CE28-0014-01).



## ORCID

*Katarzyna Pisanski* https://orcid.org/0000-0003-0992-2477
*Clément Cornec* https://orcid.org/0000-0002-4110-328X
*Andrey Anikin* https://orcid.org/0000-0002-1250-8261
*David Reby* https://orcid.org/0000-0001-9261-1711
*Gregory A. Bryant* https://orcid.org/0000-0002-7240-4026


## REFERENCES


Ackermann H, Hage SR, Ziegler W. 2014. Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. Behav Brain Sci. 37(06):529–546. doi:https://doi.org/10.1017/S0140525X13003099

Anikin A. 2019. Soundgen: An open-source tool for synthesizing nonverbal vocalizations. Behav Res Methods. 51(2):778–792.

Anikin A. 2020. The link between auditory salience and emotion intensity. Cogn Emot. 34(6):1246–1259.

Anikin A, Bååth R, Persson T. 2018. Human non-linguistic vocal repertoire: Call types and their meaning. J Nonverbal Behav. 42(1):53–80.

Anikin A, Pisanski K, Massenet M, Reby D. 2021. Harsh is large: nonlinear vocal phenomena lower voice pitch and exaggerate body size. Proc R Soc Lond B. 288(1954). doi:https://doi.org/10.1098/rspb.2021.0872

Anikin A, Pisanski K, Reby D. 2020. Do nonlinear vocal phenomena signal negative valence or high emotion intensity? R Soc Open Sci. 7(12):201306.

Arias P, Rachman L, Liuni M, Aucouturier J-J. 2020. Beyond correlation: acoustic transformation methods for the experimental study of emotional voice and speech. Emot Rev. 13:12-24.

Arnal LH, Flinker A, Kleinschmidt A, Giraud A-L, Poeppel D. 2015. Human screams occupy a privileged niche in the communication soundscape. Curr Biol. 25(15):2051–2056. doi:https://doi.org/10.1016/j.cub.2015.06.043

Aung T, Puts D. 2020. Voice pitch: a window into the communication of social power. Curr Opin Psychol. 33:154–161.

Banse R, Scherer KR. 1996. Acoustic profiles in vocal emotion expression. J Pers Soc Psychol. 70(3):614–636. doi:https://doi.org/10.1037/0022-3514.70.3.614

Barrett HC. 2021. Psychology within and without the state. Annu Rev Psychol. doi:https://doi.org/10.1146/annurev-psych-020821-110248

Bellieni CV, Sisto R, Cordelli DM, Buonocore G. 2004. Cry features reflect pain intensity in term newborns: An alarm threshold. Pediatr Res. 55(1):142–146. doi:https://doi.org/10.1203/01.PDR.0000099793.99608.CB

Belyk M, Brown S. 2016. Pitch underlies activation of the vocal system during affective vocalization. Soc Cog Affect Neurosci. 11(7):1078-1088.

Belyk M, Pfordresher PQ, Liotti M, Brown S. 2016. The neural basis of vocal pitch imitation in humans. J Cogn Neurosci. 28(4):621-635.

Blumstein DT, Recapet C. 2009. The sound of arousal: The addition of novel non-linearities increases responsiveness in marmot alarm calls. Ethology. 115(11):1074–1081.

Boersma P, Weenink D. 2021. Praat: doing phonetics by computer [software]. Available from: http://www.praat.org/ [Accessed November 30, 2021].

Bornstein MH, Putnick DL, Rigo P, Esposito G, Swain JE, Suwalsky JTD, Su X, Du X, Zhang K, Cote LR, et al. 2017. Neurobiology of culturally common maternal responses to infant cry. Proc Natl Acad Sci. 114(45):E9465–E9473. doi:https://doi.org/10.1073/pnas.1712022114

Bornstein MH, Tamis-LeMonda CS, Tal J, Ludemann P, Toda S, Rahn CW, Pêcheux M, Azuma H, Vardi D. 1992. Maternal responsiveness to infants in three societies: The United States, France, and Japan. Child Dev. 63(4):808–821.

Boukydis CZ, Burgess RL. 1982. Adult physiological response to infant cries: Effects of temperament of infant, parental status, and gender. Child Dev. 53:1291–1298.

Bryant GA. 2020. Evolution, structure, and functions of human laughter. In: Floyd F, Weber R, editors. The handbook of communication science and biology. New York (NY): Routledge; p. 63–77.

Bryant, GA. 2021a. Vocal communication across cultures: Theoretical and methodological issues. Philos Trans R Soc B. 376:20200387.

Bryant GA. 2021b. The evolution of human vocal emotion. Emot Rev. 13(1):25–33.

Bryant GA, Aktipis CA. 2014. The animal nature of spontaneous human laughter. Evol Hum Behav. 35(4):327–335. doi:https://doi.org/10.1016/j.evolhumbehav.2014.03.003

Bryant GA, Barrett HC. 2007. Recognizing intentions in infant-directed speech: Evidence for universals. Psychol Sci. 18(8):746–751.

Bryant GA, Barrett HC. 2008. Vocal emotion recognition across disparate cultures. J Cogn Cult. 8(1–2):135–148. doi:https://doi.org/10.1163/156770908X289242

Bryant GA, Fessler DMT, Fusaroli R, Clint E, Aarøe L, Apicella CL, Petersen MB, Bickham ST, Bolyanatz A, Chavez B, et al. 2016. Detecting affiliation in colaughter across 24 societies. Proc Natl Acad Sci. 113(17):4682–4687. doi:https://doi.org/10.1073/pnas.1524993113

Bryant GA, Fessler DM, Fusaroli R, Clint E, Amir D, Chávez B, Denton KK, Díaz C, Duran LT, Fanćovićová J. 2018. The perception of spontaneous and volitional laughter across 21 societies. Psychol Sci. 29(9):1515–1525.

Bryant GA, Lienard P, Barrett HC. 2012. Recognizing infant-directed speech across distant cultures. J Evol Psychol. 10(2):47–59. doi:https://doi.org/10.1556/jep.10.2012.2.1

Burred JJ, Ponsot E, Goupil L, Liuni M, Aucouturier J-J. 2019. CLEESE: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition. PLoS ONE. 14(4):e0205943.

Chakladar S, Logothetis NK, Petkov CI. 2008. Morphing rhesus monkey vocalizations. J Neurosci Methods. 170(1):45–55.

Chang RS, Thompson NS. 2011. Whines, cries, and motherese: Their relative power to distract. J Soc Evol Cult Psychol. 5(2):131.

Charlton BD, Pisanski K, Raine J, Reby D. 2020. Coding of static information in terrestrial mammal vocal signals. In: Aubin T, Mathevon, N, editors. Coding strategies in vertebrate acoustic communication. Cham (Switzerland): Spring Nature Switzerland; p. 115-136.

Clay Z, Archbold J, Zuberbühler K. 2015. Functional flexibility in wild bonobo vocal behaviour. PeerJ. 3:e1124.

~~Cornec C. et al. (submitted). Human infant cries communicate distress and elicit sex stereotypes: Cross cultural evidence.~~

Cornec C., Ngofuna M., Lemasson A., Monghiemo C., Narat V., Levréro F. (In press). A pilot study of temporal patterns of vocal behaviours in wild bonobos *Pan paniscus*. Ethol Ecol Evol.

Crockford C, Wittig RM, Mundry R, Zuberbühler K. 2012. Wild chimpanzees inform ignorant group members of danger. Curr Biol. 22(2):142-146.

Ćwiek A, Fuchs S, Draxler C, Asu EL, Dediu D, Hiovain K, Kawahara S, Koutalidis S, Krifka M, Lippus P. 2021. Novel vocalizations are understood across cultures. Sci Rep. 11(1):1–12.

Davila Ross M, J Owren M, Zimmermann E. 2009. Reconstructing the evolution of laughter in great apes and humans. Curr Biol. 19(13):1106–1111. doi:https://doi.org/10.1016/j.cub.2009.05.028

DiMattina C, Wang X. 2006. Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. J Neurophysiol. 95(2):1244–1262.

Dudek J, Faress A, Bornstein MH, Haley DW. 2016. Infant cries rattle adult cognition. PLoS ONE. 11(5):e0154283.

El Haddad K, Cakmak H, Sulír M, Dupont S, Dutoit T. 2016. Audio affect burst synthesis: A multilevel synthesis system for emotional expressions. In: Proceedings of the 24th European Signal Processing Conference; p. 1158–1162.

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. doi:https://doi.org/10.7717/peerj.10990

Facchini A, Bellieni CV, Marchettini N, Pulselli FM, Tiezzi EBP. 2005. Relating pain intensity of newborns to onset of nonlinear phenomena in cry recordings. Phys Lett A. 338(3–5):332–337. doi:https://doi.org/10.1016/j.physleta.2005.02.048

Fant G. 1960. Acoustic theory of speech production. The Hague (The Netherlands): Mouton.

Fernald A. 1992. Human maternal vocalizations to infants as biologically relevant signals. In: Barkow JH, et al., editors. The adapted mind: Evolutionary psychology and the generation of culture. New York (NY): Oxford University Press; p. 391–428.

Fitch WT. 2018. The biology and evolution of speech: A comparative analysis. Annu Rev Linguist. 4:255–279.

Fitch WT. 2000. The evolution of speech: a comparative review. Trends Cogn Sci. 4(7):258-267.

Fitch WT, Hauser MD. 1995. Vocal production in nonhuman primates: acoustics, physiology, and functional constraints on “honest” advertisement. Am J Primatol. 37(3):191-219.

Fitch WT, Neubauer J, Herzel H. 2002. Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. Anim Behav. 63(3):407–418. doi:https://doi.org/10.1006/anbe.2001.1912

Fleming AS, Corter C, Stallings J, Steiner M. 2002. Testosterone and prolactin are associated with emotional responses to infant cries in new fathers. Horm Behav. 42(4):399–413.

Frodi AM, Lamb ME, Wille D. 1981. Mothers’ responses to the cries of normal and premature infants as a function of the birth status of their own child. J Res Personal. 15(1):122–133.

Furuyama T, Kobayasi KI, Riquimaroux H. 2017. Acoustic characteristics used by Japanese macaques for individual discrimination. J Exp Biol. 220(19):3571–3578.

Gendron M, Roberson D, van der Vyver JM, Barrett LF. 2014. Cultural relativity in perceiving emotion from vocalizations. Psychol Sci. 25(4):911–920. doi:https://doi.org/10.1177/0956797613517239

Gustafson GE, Harris KL. 1990. Women's responses to young infants' cries. Dev Psychol. 26(1):144–152.

Gustafson GE, Wood RM, Green JA. 1999. Can we hear the causes of infants' crying? Clin Dev Med.(152):8–22.

Hechler C, Beijers R, de Weerth C. 2015. Young adults' reactions to infant crying. Infant Behav Dev. 38:41–48.

Henrich J, Heine SJ, Norenzayan A. 2010. The weirdest people in the world? Behav Brain Sci. 33(2–3):61–83. doi:https://doi.org/10.1017/S0140525X0999152X

Hinton L, Nichols J, Ohala JJ. 2006. Sound symbolism. Cambridge (UK): Cambridge University Press.

Huang N, Elhilali M. 2017. Auditory salience using natural soundscapes. J Acoust Soc Am. 141(3):2163–2176.

Kawahara H. 2006. STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. Acoust Sci Technol. 27(6):349–353.

Kaya EM, Elhilali M. 2014. Investigating bottom-up auditory attention. Front Hum Neurosci. 8:327.

Kelly T, Reby D, Levréro F, Keenan S, Gustafsson E, Koutseff A, Mathevon N. 2017. Adult human perception of distress in the cries of bonobo, chimpanzee, and human infants. Biol J Linn Soc. 120(4):919–930. doi:https://doi.org/10.1093/biolinnean/blw016

Kleisner K, Leongómez JD, Pisanski K, Fiala V, Cornec C, Groyecka-Bernard A, Butovskaya M, Reby D, Sorokowski P, Akoko RM. 2021. Predicting strength from aggressive vocalisations versus speech in African bushland and urban communities. Philos Trans R Soc B. doi:https://doi.org/10.1098/rstb.2020.0403

Knight C, Lewis J. 2017. Wild voices: Mimicry, reversal, metaphor, and the emergence of language. Curr Anthropol. 58(4):435–453.

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(4):313-325. doi:https://doi.org/10.1080/09524622.2017.1344931

Lameira AR, Hardus ME, Kowalsky B, de Vries H, Spruijt BM, Sterck EH, Shumaker RW, Wich SA. 2013a. Orangutan (*Pongo* spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. J Acoust Soc. 134(3):2326-2335.

Lameira AR, Hardus ME, Nouwen KJ, Topelberg E, Delgado RA, Spruijt BM, Sterck EH, Knott CD, Wich SA. 2013b. Population-specific use of the same tool-assisted alarm call between two wild orangutan populations (*Pongopygmaeus wurmbii*) indicates functional arbitrariness. PLoS ONE. 5;8(7):e69749.

Laporte MNC, Zuberbühler K. 2010. Vocal greeting behaviour in wild chimpanzee females. Anim. Behav. 80(3):467–473.

Laukka P, Elfenbein HA. 2021. Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. Emot Rev. 13(1):3–11.

Laurent HK, Ablow JC. 2012. A cry in the dark: depressed mothers show reduced neural activation to their own infant's cry. Soc Cogn Affect Neurosci. 7(2):125–134.

Lavan N, Scott SK, McGettigan, C. 2015. Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter. J Nonverbal Behav. 40(2):133-149.

Leger DW, Thompson RA, Merritt JA, Benz JJ. 1996. Adult perception of emotion intensity in human infant cries: effects of infant age and cry acoustics. Child Dev. 67(6):3238–3249.

Lewis J. 2009. As well as words: Congo Pygmy hunting, mimicry, and play. In: Botha R, Knight C, editors. The cradle of language: Studies in the evolution of langugage. Vol. 12. Oxford (UK): Oxford University Press; p. 236–256.

Lingle S, Riede T. 2014. Deer mothers are sensitive to infant distress vocalizations of diverse mammalian species. Am Nat. 184(4):510–522. doi:https://doi.org/10.1086/677677

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(5):698–726.

Ludlow CL. 2015. Central nervous system control of voice and swallowing. J Clin Neurophysiol. 32(4):294.

Margoliash D. 1983. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. J Neurosci. 3(5):1039–1057.

Messina I, Cattaneo L, Venuti P, de Pisapia N, Serra M, Esposito G, Rigo P, Farneti A, Bornstein MH. 2016. Sex-specific automatic responses to infant cries: TMS reveals greater excitability in females than males in motor evoked potentials. Front Psychol. 6:1909.

Moore RK, Marxer R, Thill S. 2016. Vocal interactivity in-and-between humans, animals, and robots. Front Robot AI. 3(61):1-17. doi:https://doi.org/10.3389/frobt.2016.00061

Moore RK, Mitchinson B. 2017. A biomimetic vocalisation system for MiRo. In: Proceedings of the Conference on Biomimetic and Biohybrid Systems 2017 Jul 26. Cham (Switzerland): Springer; p. 363-374.

Morsbach G, McCulloch M, Clark A. 1986. Infant crying as a potential stressor concerning mothers' concentration ability. Psychol Int J Psychol Orient. 29(1):18-20.

Morton ES. 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. Am Nat. 111(981):855–869.

Ohala JJ. 1984. An ethological perspective on common cross-language utilization of F0 of voice. Phonetica. 41(1):1–16.

Owren MJ, Rendall D. 2001. Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. Evol Anthropol Issues News Rev. 10(2):58–71. doi:https://doi.org/10.1002/evan.1014

Pell M, Paulmann S, Dara C, Alasseri A, Kotz S. 2009. Factors in the recognition of vocally expressed emotions: A comparison of four languages. J Acoust Soc Am. 37(4):417–435.

Pisanski K, Bryant GA. 2019. The evolution of voice perception. In: Eidsheim NS, Meizel KL, editors. Oxford handbook of voice studies. New York (NY): Oxford University Press; p. 269-300.

Pisanski K, Cartei V, McGettigan C, Raine J, Reby D. 2016. Voice modulation: A window into the origins of human vocal control? Trends Cogn Sci. 20(4):304–318.

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(2):191642.

Provine RR. 1993. Laughter punctuates speech: Linguistic, social and gender contexts of laughter. Ethology. 95(4):291–298.

Rad MS, Martingano AJ, Ginges J. 2018. Toward a psychology of *Homo sapiens*: Making psychological science more representative of the human population. Proc Natl Acad Sci. 115(45):11401–11405.

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(3):e0213034.

Raine J, Pisanski K, Simner J, Reby D. 2018. Vocal communication of simulated pain. Bioacoustics. 28:404-426. doi:https://doi.org/10.1080/09524622.2018.1463295

Reby D, Charlton BD. 2012. Attention grabbing in red deer sexual calls. Anim Cogn. 15(2):265–270. doi:https://doi.org/10.1007/s10071-011-0451-0

Salvia E, Bestelmeyer PE, Kotz SA, Rousselet GA, Pernet CR, Gross J, Belin P. 2014. Single-subject analyses of magnetoencephalographic evoked responses to the acoustic properties of affective non-verbal vocalizations. Front Neurosci. 8:422.

Sauter DA, Eisner F, Ekman P, Scott SK. 2010. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proc Natl Acad Sci. 107(6):2408–2412. doi:https://doi.org/10.1073/pnas.0908239106

Schel AM, Machanda Z, Townsend SW, Zuberbühler K, Slocombe KE. 2013. Chimpanzee food calls are directed at specific individuals. Anim Behav. 86(5):955-965.

Scott S. 2021. The neural control of volitional vocal production – from speech to identity, from social meaning to song. Philos Trans Royal Soc B. 377(1841).

Scott SK, Lavan N, Chen S, McGettigan C. 2014. The social life of laughter. Trends Cogn Sci. 18(12):618–620. doi:https://doi.org/10.1016/j.tics.2014.09.002

Scott-Phillips T. 2015. Speaking our minds: Why human communication is different, and how language evolved to make it special. London (UK): Red Globe Press and Springer Nature Limited.

Seyfarth R, Cheney D. 2018. Pragmatic flexibility in primate vocal production. Curr Opin Behav Sci. 21:56-61.

Snowdon CT, Pola YV. 1978. Interspecific and intraspecific responses to synthesized pygmy marmoset vocalizations. Anim Behav. 26:192–206.

Sroufe LA, Wunsch JP. 1972. The development of laughter in the first year of life. Child Dev. 43:1326–1344.

Sueur J, Aubin T, Simonis C. 2008. Seewave, a free modular tool for sound analysis and synthesis. Bioacoustics. 18(2):213–226.

Swain JE. 2011. The human parental brain: in vivo neuroimaging. Prog Neuropsychopharmacol Biol Psychiatry. 35(5):1242–1254.

Swain JE, Lorberbaum JP, Kose S, Strathearn L. 2007. Brain basis of early parent–infant interactions: psychology, physiology, and in vivo functional neuroimaging studies. J Child Psychol Psychiatry. 48(3-4):262–287.

Tanner JC, Justison J, Bee MA. 2020. SynSing: open-source MATLAB code for generating synthetic signals in studies of animal acoustic communication. Bioacoustics. 29(6):731–752.

Taylor AM, Charlton BD, Reby D. 2016. Vocal production by terrestrial mammals: Source, filter, and function. In: Suthers A, et al., editors. Vertebrate sound production and acoustic communication. Cham (Switzerland): Springer Nature Switzerland; p. 229-259.

Taylor AM, Reby D. 2010. The contribution of source-filter theory to mammal vocal communication research: Advances in vocal communication research. J Zool. 280(3):221–236. doi:https://doi.org/10.1111/j.1469-7998.2009.00661.x

Tiezzi EBP, Pulselli FM, Facchini A. 2004. Determination of pain intensity in newborns by time series analysis. WIT Trans Ecol Environ. 73:443–450.

Townsend SW, Manser MB. 2011. The function of nonlinear phenomena in meerkat alarm calls. Biol Lett. 7(1):47–49.

Urbain J, Cakmak H, Dutoit T. 2013. Evaluation of HMM-based laughter synthesis. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing; p. 7835–7839.

Venuti P, Caria A, Esposito G, De Pisapia N, Bornstein MH, de Falco S. 2012. Differential brain responses to cries of infants with autistic disorder and typical development: An fMRI study. Res Dev Disabil. 33(6):2255–2264.

Vettin J, Todt D. 2005. Human laughter, social play, and play vocalizations of non-human primates: an evolutionary approach. Behaviour. 142:217-240.

Wattendorf E, Westermann B, Fiedler K, Kaza E, Lotze M, Celio MR. 2013. Exploration of the neural correlates of ticklish laughter by functional magnetic resonance imaging. Cereb Cortex. 23(6):1280-1289.

Winkler S, Bryant GA. 2021. Play vocalisations and human laughter: a comparative review. Bioacoustics. 30:499-526. doi:https://doi.org/10.1080/09524622.2021.1905065

Witteman J, Van IJzendoorn M, Rilling J, Bos P, Schiller N, Bakermans-Kranenburg M. 2019. Towards a neural model of infant cry perception. Neurosci Biobehav Rev. 99:23–32.

Woodard K, Plate RC, Morningstar M, Wood A, Pollak SD. 2021. Categorization of vocal emotion cues depends on distributions of input. Affect Sci. doi:https://doi.org/10.1007/s42761-021-00038-w

Yoo H, Buder EH, Bowman DD, Bidelman GM, Oller DK. 2019. Acoustic correlates and adult perceptions of distress in infant speech-like vocalizations and cries. Front Psychol. 10:1154.

Zúñiga J, Reiss JD. 2019. Realistic procedural sound synthesis of bird song using particle swarm optimization. J Audio Eng Soc. 210886639.