Distribution Agreement

Signature:

Jonathan W. M. Engelberg Date

Acoustic, emotional, and perceptual variation in human screams: Exploring the diversity in a basic call type

By

Jonathan W. M. Engelberg
Doctor of Philosophy

Psychology

Harold Gouzoules
Advisor

Frans de Waal
Committee Member

Robert Hampton
Committee Member

Lynne Nygaard
Committee Member

Gordon Ramsay
Committee Member

James Rilling
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Acoustic, emotional, and perceptual variation in human screams: Exploring the diversity in a basic call type

By

Jonathan W. M. Engelberg
M.A., Emory University, 2017

Advisor: Harold Gouzoules, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2021

Abstract

Acoustic, emotional, and perceptual variation in human screams: Exploring the diversity in a basic call type

By Jonathan W. M. Engelberg

The nonlinguistic vocal repertoire predates language in human (*Homo sapiens*) evolution, and call types such as laughs and screams are likely homologous to acoustically similar calls in nonhuman animals. However, most research on nonlinguistic vocalizations centers on vocalizers' abilities to express emotion without accounting for the call types represented among those expressions. This perspective ignores important questions about why some call types are associated with a variety of emotional contexts, whether acoustic variation within types is communicatively significant, and what functions they might serve beyond emotional communication. To address these questions in human screams, I adopted an alternative perspective from the literature on animal communication, wherein researchers first identify the call type(s) of interest before exploring their potential meanings and functions. In Study 1, I investigated whether the variation among screams elicits varied emotional perceptions in listeners. Listeners rated contextually diverse screams on six emotion scales. Their ratings of screams varied along two primary dimensions, one separating the perception of anger and pain from happiness and surprise, and one independently accounting for some perception of fear. Acoustic parameters predicted listener ratings in ways consistent with patterns of emotional variation reported in other human and nonhuman vocalizations. To compare this variation in screams to speech directly, in Study 2, I recorded nine actors' screams and single-utterance speech samples across five emotional contexts. I found similar patterns of acoustic variation by emotion across screams and speech, suggesting that the same mechanisms underlie variation in each vocalization type. Listeners also achieved similar accuracies across vocalization types in an emotion recognition task, although they demonstrated some response biases, such as a greater tendency towards false alarms for pain and fear, that were potentially specific to screams. In an additional task, listeners rated screams on five hypothesis-based perceptual scales. I found that that some perceptual characteristics, such as attention-getting, did not vary significantly by scream emotion, whereas others, such as communicating distress, did vary by emotion, possibly hinting both at general functions that unite screams across disparate emotional contexts as well as more contextually specific subfunctions. In all, these findings suggest that acoustic variation within human call types is informative and functional. I suggest that the influence of language faculties on preexisting call types that are shared with other taxa partly explains the contextual and acoustic diversity of human calls.

Acoustic, emotional, and perceptual variation in human screams: Exploring the diversity in a basic call type

By

Jonathan W. M. Engelberg
M.A., Emory University, 2017

Advisor: Harold Gouzoules, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2020

"Sometimes, a scream is better than a thesis." – Ralph Waldo Emerson

Certainly, there were a couple of times when I would like to have replaced this dissertation with a good, full-throated scream. Thankfully, I chose to express my ideas linguistically instead. I could not have done it without the following people.

I owe the existence of this research—and of any other accomplishment in graduate school—to my advisor, Dr. Harold Gouzoules. Harold, I am eternally grateful for your endless encouragement, your impeccable writing instincts, and your superhuman knowledge and enthusiasm for all things animal behavior. Your mentorship has made me a better scientist, teacher, writer, and person. Thank you.

I have relied immensely on Jay Schwartz, the other core member of our Scream Team. Jay, I've lost track of the number of times I've bothered you for input on everything from acoustic analysis and linear mixed modeling to applying for jobs. Thank you for your wisdom, your patience, and your friendship. Thanks also to the undergraduate researchers who made my research possible: Shiyin Liu, Kyra Mitchell, and Caroline Wei.

Thank you to everyone at the Laboratory of Comparative Primate Cognition for welcoming me into your world at Yerkes. Dr. Robert Hampton, your guidance has broadened my perspectives on cognition, science, and life in general, and I look forward to learning even more from you. To my fellow graduate students, past and present—Dr. Tom Hassett, Dr. Ryan Brady, Jad Nasrini, and Rael Sammeroff—thank you for improving my research through your insights, and for maintaining my sanity through your games and comradeship. Special thanks to Ryan for joining me on the madcap venture that is teaching monkeys to respond to sounds. Infinite thanks to Tara Dove-VanWormer, without whom the lab would probably cease to function, and to all the undergraduates with whom I had the pleasure to work. Finally, thank you to the monkeys, even though you did not make the cut for this manuscript.

Dr. Frans de Waal, Dr. Lynne Nygaard, Dr. Gordon Ramsay, and Dr. James Rilling: Thank you each for sharing your invaluable expertise and your generous feedback throughout the dissertation process.

To my family: Ben and Daniel, your intellectual curiosity is a constant source of inspiration, and your humor and companionship are sources of perspective: both were necessary for me to finish this thing. Mom and Dad, you have supported me unconditionally in everything I've done, and this was no exception. I think you would have supported me if my dream were to become a wandering bard, although I'm sure you're glad I chose graduate school instead. (Even if the pay was not that much better.) Thank you.

To my second family: Tuan, My-Huong, Thuy-Linh, (and Obi!), thank you for embracing me as a Nguyen, and for providing much-needed distractions in the form of beach vacations and ski trips (whether or not the ski lifts were running).

My-Linh, you have stood next to me through this entire journey: marked every milestone, celebrated every achievement, and grounded me through every fit of self-doubt. This dissertation is for you. Thank you for everything.

Table of Contents

CHAPTER I. GENERAL INTRODUCTION ... 1
- Terminology ... 6
- Lessons from animal communication ... 8
  - Signals, senders, and receivers ... 8
  - Function, form, and the role of emotion ... 10
  - Extended example: Emotion and functions of antipredator screams ... 12
  - Call types, contextual diversity, and variation within types ... 14
  - Extended example: Variation, function, and emotion in primate screams ... 22
  - Summary: Applications from animal communication research ... 25
- Nonlinguistic vocalizations in humans ... 28
  - The existence of human call types ... 28
  - Possible effects of language on nonlinguistic vocalizations ... 30
  - Infant cries ... 33
  - Laughter ... 36
  - Summary ... 37
- Human screams and this dissertation ... 39
- References ... 44

CHAPTER II. The emotional canvas of human screams: Patterns and acoustic cues in the perceptual categorization of a basic call type ... 55
- Abstract ... 56
- Methods ... 61
  - Participants ... 61
  - Stimuli ... 62
  - Procedure ... 65
  - Analysis ... 66
- Results ... 70
  - Overall ratings and agreement ... 70
  - Correlations between emotion ratings ... 72
  - Acoustic predictors of emotion ratings ... 73
  - Ratings as a function of source contexts ... 78
- Discussion ... 79
  - Agreement and perceptual dimensions ... 79
  - Acoustic predictors of emotion perception ... 81
  - Accuracy and the case of happiness ... 85
- References ... 88

CHAPTER III. Emotion-related variation in simple speech extends to screams: Implications for human nonverbal communication ... 92
- Abstract ... 93
- Methods ... 107
  - Participants ... 107
  - Stimuli ... 108
  - Experimental procedure ... 110
  - Acoustic analysis ... 113
  - Statistical analysis ... 115
- Results ... 120
  - Emotion Categorization ... 120
  - Acoustic differences between emotions ... 125
  - Perceptual effects ratings ... 131

Listener-classified screams 134
Discussion 137
Emotion categorization of screams and speech 137
Perceptual effects of screams 144
Screams as a call type 149
References 154
CHAPTER IV. General Discussion 161
Contextual diversity and variation within call types 163
Call function and emotional communication 166
Human call types and language 170
References 173
APPENDIX 176

Figures and Tables

Table 1.1 64
Figure 1.1 70
Table 1.2 71
Table 1.3 72
Table 1.4 73
Table 1.5 74
Figure 1.2 75
Figure 1.3 77
Table 1.6 79
Table 2.1 104
Table 2.2 118
Table 2.3 120
Figure 2.1 121
Figure 2.2 123
Table 2.4 124
Table 2.5 125
Figure 2.3 129
Figure 2.4 130
Table 2.6 132
Figure 2.5 133
Table 2.7 134
Table 2.8 136
Table A1 182
Table A2 184
Table A3 186
Table A4 188

CHAPTER I. GENERAL INTRODUCTION

Nonlinguistic vocalizations such as screams, laughs, and cries play an important role in human communication. By definition, nonlinguistic vocalizations contain no conventionalized words. They are closely associated with emotional states and may sometimes convey cues to emotion that speech does not (Hawk et al., 2009; Sauter & Eimer, 2009; Sauter et al., 2013; Schröder, 2003). Many nonlinguistic vocalizations also share striking functional, acoustic, and mechanistic similarities with animal vocalizations (Davila-Ross et al., 2009; Lingle et al., 2012; Owren et al., 2011) and likely derive from the calls that encompassed all vocal communication in our ancestors prior to the advent of language. Research on these vocalizations, then, is indispensable for understanding the evolution of emotional communication in humans.

Darwin devoted his third and final major work on evolution to exploring continuities in emotional expression (Darwin, 1872/1998), including ample consideration of nonlinguistic vocalizations and their evolutionary origins. Screams figured into these ruminations. Darwin took note of both the emotional salience and the phylogenetic prevalence of vocalizations "associated with terror" (p. 83)—their occurrence not limited to primates but also observed in hares and rabbits, cattle and horses. He speculated as to why screams have acquired such striking acoustic features—"loud, prolonged, and high, so as to penetrate to a distance" (p. 91)—and highlighted what he perceived as evolutionary homologies, or common ancestral origins, between screams in humans and other species.

Human screams, however, are not only associated with terror, but also intense joy, anger, pain, and surprise (Anikin et al., 2018), a *contextual diversity* seemingly unparalleled by the screams of any other species (although other, non-scream vocalizations may occur in diverse contexts in other animals, e.g., Clay et al., 2015). The origins and implications of this diversity in screams remain unknown. One important question is whether screams that occur in different contexts are acoustically homogenous, or whether acoustic variation among screams correlates with context, potentially facilitating context-specific functions. A second question concerns what it is that causes acoustically similar vocalizations to occur in such diverse contexts in the first place.

Researchers have studied similar questions regarding other human vocalizations, most notably infant cries (Fuller & Horii, 1988; Wasz-Hockert et al. 1968) and laughter (Bachorowski & Owren, 2001; Szameitat et al., 2009, 2010), often producing ambiguous results. Indeed, as a whole, there remain many unresolved issues about nonlinguistic vocalizations altogether, let alone whether variation within call types is functional or informative. I contend that this paucity of knowledge owes in part to the perspective from which most research in this field is conducted.

There exist at least two lenses through which researchers of assorted disciplines have viewed nonlinguistic vocalizations, each traceable to Darwin's (1872/1998) treatise on the subject. The first and most common, seemingly favored by emotion psychologists, is foremost as vocal expressions of emotion, ones that are especially "pure" (Pell et al., 2015, p. 4) in that they might more closely track the physiological components of emotion than speech and are supposedly uncontaminated by confounding linguistic content (Bostanov & Kotchoubey, 2004; Cordaro et al., 2016; Laukka et al., 2013; Lima et al., 2013; Pell et al., 2015; Sauter et al., 2010). Researchers embracing this emotion-first approach begin by specifying emotions as the categories of interest and might proceed either by asking actors to portray those emotions or by finding naturalistic samples representative of each category (Scherer, 2003). Usually, stimuli are limited to those containing no identifiable linguistic structure (although the qualifications for "linguistic" vary between research teams; Cordaro et al., 2016, for example, allowed such phonemically distinct sounds as 'Woohoo' and 'Wow'), but no attempt is otherwise made to define or delimit their acoustic forms. Thus, a typical study might include screams, laughter, cries, and other types of vocalizations among its stimuli, and the task for participants is to categorize or rate the emotions that each of these vocalizations conveys.

The second perspective, adopted from the field of animal behavior, is to treat nonlinguistic vocalizations as natural call types comprising the human vocal repertoire (Anikin et al., 2018; Schwartz et al., 2020a), the acoustic structures of each partly deriving from some biological basis and shaped by natural selection (Ackermann et al., 2014; Owren et al., 2011). By this call-first approach, researchers first specify the call type(s) of interest before asking questions about how those types function and what they communicate. This approach thus entails additional constraints on acoustic structure, for example, by requiring actors to produce a certain call type (e.g., Szameitat et al., 2009). Experimenters are not precluded from investigating the emotional content of call types (e.g., Anikin et al., 2018; Szameitat et al., 2009), but other factors relevant to call function are often investigated as well, such as listeners' perceptions of the caller's attributes (e.g., sex or body size; Raine et al., 2017, 2018) or the effects of acoustic variation on social judgments (e.g., affiliation or dominance; Wood et al., 2017).

Neither perspective is inappropriate. Nonlinguistic vocalizations, like other communicative displays in humans, can both express emotions and constitute the basic, evolved repertoire of our species. Each approach, however, prioritizes different research questions and stipulates different considerations for experimental design. In this introductory chapter, I will argue that some gaps in our understanding of nonlinguistic vocalizations owe to the relative predominance of research in humans that neglects consideration of call types. In humans as in nonhuman animals, research conducted blindly to call types will not answer all the questions that motivated Darwin (1872/1998)—how our emotional expressions function, why they have acquired their contemporary forms, and why, altogether, our nonlinguistic repertoire evolved. We could not surmise, for example, why human screams seem to occur in positive emotional contexts (or even verify that they do) without defining screams as the object of research and attempting to uncover shared functional or emotional characteristics that lead to their occurrence across contexts.

I will first turn to the literature on animal communication to elaborate on the rationale that underpins attempts to partition vocal repertoires into call types, and to demonstrate how this perspective is useful for addressing questions related to the functions, acoustic forms, and evolution of vocalizations. One question, central to this thesis, concerns contextual diversity and its possible relationship with acoustic and functional variation within call types. A brief review of the literature on animal screams, including those of nonhuman primates, will illustrate these points while doubling to provide the evolutionary background for our research in human screams. In conveying the importance of acoustics and call function, I will initially (intentionally) deemphasize the role of emotion in vocalizations for the exact reason that emotion has received the lion's share of attention in human nonlinguistic research, but it must be said that emotion plays a central role in these calls. Indeed, I will later argue that attention to call types can enhance our understanding of the role of emotion.

Next, I will advocate that future research should embrace this perspective when investigating nonlinguistic vocalizations in humans. Given the robust evidence for homology between nonhuman and human vocalizations, it is appropriate to ask similar questions of each and use similar methodologies to answer them. This approach may implicate not only the evolutionarily basal properties of human call types, those shared with other species, but also the changes that have occurred in this communicative system during human evolution, particularly with the emergence of language. I will refer to the fairly extensive literatures on infant cries and laughter, which exemplify the kinds of insights that research on call types can yield, but which also leave unanswered many questions regarding contextual diversity and its relationship with functional and acoustic variation in human calls. I will argue that research on human screams, although comparatively sparse, represents a potentially illuminating area for expanding our knowledge about nonlinguistic communication in humans. The studies in this dissertation constitute one such endeavor towards understanding variation among contextually diverse human screams.

Terminology

Most of the terminology in this dissertation is adopted from the animal communication literature and is therefore defined in the section below. Here, I will introduce a few key phrases that are not as commonly used or formalized in the literature.

I use *emotional context* to refer to the externally observable socioecological conditions in which a call occurs, and in which the caller plausibly experiences a particular associated emotion. For example, one might label a predator-prey interaction as a fear-associated context for the prey. This distinction entails no assumptions regarding each vocalizer's actual emotion or internal state, or about the precise nature thereof, in every instantiation of that context. Instead, it is an operationalizable construct by which researchers can characterize the situations in which calls occur and describe any correlated acoustic or behavioral variation without necessarily measuring or positing underlying states. This approach is common in free-ranging or semi-naturalistic animal research in which it is often difficult to access more direct measures of emotional states. We emphasize that emotion is doubtlessly central to the contexts we describe, but we adopt the broader *emotional context* terminology to acknowledge precisely the limitation that it is technically the context and observable behavior, and not the underlying state, that researchers often correlate with signals.

Another justification for cautious nomenclature is that the concept of emotion itself is notoriously contentious. For the purposes of this dissertation, I adopt the view that emotions comprise temporary responses to biologically relevant stimuli involving the organization of multiple, interrelated subsystems that drive synchronous changes in physiology and behavior and potentiate adaptive action (Adolphs, 2017; de Waal, 2011; Paul & Mendl, 2018; Russell et al., 2003; Scherer, 1986; Schartz, 2020). Emotions function as internal interfaces between environmental stimuli and behavioral responses, such that a variety of stimuli can result in similar behavioral outputs, and similar stimuli can result in variable outputs depending on the context. Thus, although emotions are biologically based (Panksepp & Watt, 2011), it is important to note that they are also flexible and malleable by experience, such that the emotional significance of different stimuli and the actions they induce can change with learning and context. Another important distinction is between emotions and subjective feelings: emotions as defined above do not necessarily entail conscious experiences, although these are components of emotions in humans. We therefore can study emotions in nonhuman animals (as well as other humans) without postulating conscious experiences (Adolphs, 2017; de Waal, 2011). Debates regarding specific models of emotion—for example, concerning whether different emotions are best conceptualized as discrete entities (Ekman, 1992; Izard, 1992; Panksepp & Watt, 2011) or as different nexuses in the same multidimensional space (Russell, 2003)—are not addressed in the present research.

In this framework, *contextual diversity* refers to the seemingly disparate emotional contexts in which many call types are observed. Of late, it has become increasingly clear that, at least in primates, one-to-one links between an emotional context and the occurrence of a particular call type are often not readily apparent (Fischer & Price, 2016). For example, bonobos produce “peep” calls during varied activities and in contexts spanning negative, neutral, and positive valences (Clay et al., 2015). A related concept is Oller et al.’s (2013) *functional flexibility*, referring to the capacity for a single vocal type to express a variety of emotions, although I use *contextual diversity* here so as not to presume the underlying state. Indeed, it is possible that some underlying emotion or other unifying factors, not obvious to outside observers, are common to each occurrence of a call type. What exactly it is that might unite a call’s contextually diverse uses is an empirical question, and one that Study 2 of this dissertation starts to address regarding human screams and their possible common effects on listeners.

Lessons from animal communication

Signals, senders, and receivers

Communication refers to the process by which one animal, the *sender*, influences the probability of some behavior in another, the *receiver*. A *signal* is a feature or behavior that was at least partially selected for its function in communication, i.e., because its effects on receiver behavior were, on average, evolutionarily advantageous to both parties (Maynard Smith & Harper, 1995). Whereas this definition focuses on natural selection as the mechanism by which signals are acquired, Smith (1980, p. 10) allows that signals might also be shaped or acquired by convention.

One fundamental concept is that the evolutionary and mechanistic factors regulating sender versus receiver behavior are not necessarily equivalent (Seyfarth & Cheney, 2003). As a strictly hypothetical example offered by Premack (1972), a sender might emit a call that reflects an emotional reaction to a stimulus. However, provided the sender reacts consistently in this way, receivers might learn to associate the call with the correlating stimulus, even if senders had no intention to inform listeners about it. Thus, what could be, in principle, entirely emotion-based variation from the sender's perspective could drive receiver behavior because of its contingency with external events, facilitating associative learning. It should be emphasized that the actual dynamics of primate signaling involve much more complex interactions of emotion, cognition, and present and past context, somewhat elaborated on below. However, the point stands that it is useful to consider the mechanisms of communication separately for senders and receivers (Seyfarth & Cheney, 2003), as I endeavor to do in this chapter.

A second fundamental and more general distinction in biology is that between *proximate* and *ultimate* causation (Mayr, 1961). A proximate account explains a phenomenon at a mechanistic level, such as the physiological and cognitive processes that might govern sender and receiver behavior. An ultimate explanation, in contrast, accounts for behavior in the evolutionary sense—its benefits conferred to senders and receivers, its ancestral precursors, and the selective forces shaping its evolution. As a classic example, vervet monkeys (*Chlorocebus pygerythrus*) produce acoustically distinct alarm calls according to the type of predator they encounter (Seyfarth et al., 1980, Struhsaker, 1967; but see Price et al., 2015, for a more recent analysis and interpretation of vervet data). Proximate explanations for these calls might invoke—for senders and receivers both—the involvement of emotion, the cognitive processes underlying decision-making, and other factors that regulate the communicative interaction. An ultimate explanation might posit that receivers have benefited from predator-specific alarm calls because qualitatively different behavioral strategies were adaptive for responding to different predator types (Townsend & Manser, 2013). A behavior that is well-suited for avoiding a leopard, such as retreating to the terminal tips of tree branches that cannot support a large cat, is not as advantageous for avoiding an eagle, in which case the inner branches are likely safer. Senders plausibly benefit in turn, albeit indirectly, because receivers are often kin, and their survival enhances the probability that shared genetic material is passed onto the next generation (Hamilton, 1964). Thus, by this ultimate account, selection might have favored increasingly perceptually distinct alarm calls because clear communication of predator type facilitated predator-specific behaviors by receivers, yielding fitness advantages to both parties.

Function, form, and the role of emotion

What the vervet monkey example illustrates is that, to understand why animal signals ultimately take the forms that they do, or why they exist at all, requires an appreciation of that signal's past and present *functions*, i.e., how its effects on receiver behavior benefit (or historically have benefited) each party. These dynamics in turn require consideration of many interconnected factors, such as the signaling environment and the perceptual capacities and biases of receivers. For example, early ethologists described the process of *ritualization* whereby signals might evolve by co-opting existing cues to become more perceptually salient, better enabling receivers to discriminate the signal and respond in a way that benefits both parties (Tinbergen, 1952). For example, tidbitting displays in galliforms—wherein a male presents food to a hen, usually accompanied by a suite of vocal and visual behaviors—likely derived historically from male feeding behavior. As females began approaching males with food, males that signaled more conspicuously enjoyed greater reproductive success, as did females who detected the displays (Evans & Smith, 2011), and so more elaborate and conspicuous displays were favored. Conversely, if a signal benefited only the receiver, selection would favor the signaler's capacities not to produce it, or to produce it with characteristics less conspicuous to conspecifics.

The example above illustrates one way in which a signal, taking a particular form, can become initially associated with a particular function because of its behavioral antecedents, following which its evolutionary trajectory is shaped due to this association. Another example, dealing with animal screams, is presented in the next subsection. The takeaway is that the origins of signals and the evolution of signal structure are processes intimately related to their communicative functions (Bradbury & Vehrencamp, 1998, pp. 376-377). Many human signals represent an exception is this regard. For example, the *symbolic* relationship between most words and their referents is often considered one distinguishing feature of language (but see Nygaard et al., 2009). However, as discussed later in this introduction, many signals in our species—including several nonlinguistic vocalizations—are also likely at least partly non-arbitrary in structure and function.

Emotion is not irrelevant to this framework, but for many signals, emotional communication *per se*—in the sense that listeners infer the caller's emotional state or associated behavior—probably only partly accounts for signal function. In taxa such as mammals, it is often presumed that emotions are proximately involved in signal production (Briefer, 2012; Fischer & Price, 2016; Owren et al., 2011), but signals rarely function as absolute read-outs of the sender's internal states. Ethologists understood this point when they described a process of *decoupling* between a signal and its associated condition that can occur during ritualization (Bradbury & Vehrencamp, 1998, p. 377). Although some signals might originate as automatic and tightly emotion-linked cues, to the extent that producing those signals is deleterious to the sender's fitness, selection is expected to favor mechanisms severing the link. This decoupling might indeed lead to an *emancipation* between the signal and condition, such that a call, no longer so tied to the underlying state, could take on novel acoustic qualities that better influence receivers to the sender's benefit. One end result of these evolutionary forces is that we should not necessarily expect a one-to-one correspondence between signals and emotional states.

For receivers, the potential details conveyed by a signal and possibly determining behavior are multifarious, including cues not only related to emotion, but also the sender's body size (Fitch, 1997), quality (Fischer et al., 2004), individual identity (Cheney & Seyfarth, 1980; Seyfarth & Cheney, 2015), location (Brown et al., 1979), and correlating external events (Macedonia & Evans, 1993; Seyfarth et al., 1980; Slocombe & Zuberbühler, 2006), to name a few possibilities. Many of these details derive from biological constraints on vocal production, yielding *indexical* acoustic cues by which receivers can infer vocalizer characteristics. Additionally, many signals may function via mechanisms that involve no receiver inference at all. For example, Owren and Rendall's (1997, 2001) *affect conditioning* model posits that animal calls drive receiver behavior primarily by eliciting an emotional state in listeners. An aversive sound might induce an unconditioned, negatively-valenced state by virtue of its acoustics, and not require receivers to derive any meaning (in the sense of an explicit mental representation) from the call. The point here is not that emotion is not pertinent—indeed, receiver emotion is central to the mechanism described above—but that a receiver's recognition of a sender's emotion is not always the most functionally significant aspect of animal signaling, and it therefore does not necessarily make sense to place primacy on emotional communication alone.

Extended example: Emotion and functions of antipredator screams

Across phylogenetically diverse taxa, prey animals produce remarkably similar vocalizations—loud, high-pitched, scream-like calls—when captured by a predator (Collias, 1960; Moller & Nielsen, 2010; Rohwer, Fretwell, & Tuckfield, 1976; Högstedt, 1983; Toledo et al., 2011). Given the likelihood of imminent death in this context, high-arousal fear is probably causal to the production of some of these antipredator screams, but this emotion by itself may not fully account for their acoustic structures. A likely evolutionarily basal function of screams was to startle the predator, enabling the caller some chance to escape (Driver & Humphries, 1969). What might have originated ancestrally as a spontaneous gasp of fear might therefore have undergone selection for louder and more acute or abrupt acoustic forms, as these characteristics are particularly effective at inducing acoustic startle (Kaltwasser, 1990). This process could entail selection for more frightened callers, thus callers who produced louder screams, but it also likely entailed selection for callers who generated louder screams even without concomitant increases in fear. In other words, a partial emancipation might have occurred between the emotional state and the signal's acoustic structure. Regardless, it was their capacity to startle, more so than their capacity to communicate fear, that likely initially shaped scream acoustics. (It is instructive to note that some species produce acoustically analogous sounds in the same context, but presumably absent any state resembling or communicating fear as we know it; Brehm et al., 2015—for example, the walnut-sphinx caterpillar that 'screams' by expelling air from its abdomen; Fullard & Napolenone, 2001. Human-like emotion is not a prerequisite to scream-like acoustics.)

In many species, screams exhibit acoustic characteristics such as longer durations, broadband frequencies, and frequency modulations that are not easily explained by the startle hypothesis, leading researchers to propose additional, likely more evolutionarily derived functions for antipredator screams. Depending on the species, screams might function altruistically to warn nearby kin, elicit distracting or mobbing behavior from conspecifics (Rohwer et al., 1976), and/or attract the attention of a second predator whose attempts to steal the prey might open an opportunity for escape (Högstedt, 1983). Each of these hypotheses has generated testable predictions about the acoustic features of screams and the behaviors they elicit. For example, to test the predator attraction hypothesis, Högstedt (1983) conducted a playback experiment in which he broadcast the screams of European starlings (*Sturnus vulgaris*) over a speaker, absent other contextual cues. Several species of predators were attracted to the speaker, consistent with this function for antipredator screams. Moreover, this and other functions invoking the approach of distant receivers can explain characteristics such as extended durations and frequency modulations, which would enhance the likelihood that other animals can detect and localize the caller.

Conceivably, the transmittance of fear could play a proximate role in some of these functions—for example, if screams induce fear in conspecifics via emotional contagion, or a low-level transfer of emotional states (de Waal, 2008), in turn motivating mobbing or avoidance behavior—but again, listeners' recognition of fear in the caller probably does not, nor does this mechanism alone account for every acoustic or functional aspect of antipredator screams. It would not, then, be sufficient to ask whether or not fear or some other state is expressed, that is, to partition the communicative system by underlying emotion. We must also ask what *screams* communicate—emotion or otherwise—and how listeners respond to them in ways revealing of their functions. The latter perspective furnishes deeper insights about the evolution of screams and the many factors that have driven their acoustic structures.

Call types, contextual diversity, and variation within types

Implicitly, an investigation of screams represents a partitioning of the communicative system by *call types*, or the canonical, species-typical vocalizations that make up a species' *vocal repertoire.* In chimpanzees (*Pan troglodytes*), for example, screams are one recognizable call type, comprising high frequency, high energy vocalizations produced mainly in the context of agonistic aggression (de Waal & van Hooff, 1981; Slocombe & Zuberbühler, 2005, 2009). These calls are acoustically and perceptually dissociable from the other call types constituting the chimpanzee vocal repertoire, which includes pant grunts (Kojima et al., 2003; Laporte & Zuberbühler, 2010), pant hoots (Clark & Wrangham, 1993; Kojima et al., 2003), alert hoos (Crockford et al., 2012/2018), and laugh-like play calls (van Hooff, 1972), among others, although note that acoustic grading exists between some types (Marler, 1976).

Different call types are presumed to reflect different mechanisms of production; that is, some acoustic variation between types is thought to derive from bottom-up differences in underlying neural control (Anikin et al., 2018). For example, different central pattern generators in the midbrain might activate laryngeal and supralaryngeal motoneurons in a sequence distinct to each call type (Jürgens & Ploog, 1970; Larson, 1991). Practically speaking, however, it is often not possible for primatologists to study the neural mechanisms involved in natural call production, much less disentangle those uniquely implicated in each call type. A more common approach to identifying call types is based on the externally identifiable contexts in which they occur (e.g., alarm calls produced when a predator is near) and/or acoustically, an effort often made challenging by acoustic variation within each type as well as the gradations between types (Fischer et al., 2016). Viewed in acoustic space, the latter task is a clustering problem—because of similarities in production, calls belonging to the same type are thought to cluster around a statistical acoustic centroid and fall closer together than calls between types—and is solved by comparing the acoustic forms of calls, either by ear, by spectrogram (a visual representation of sounds plotting the distribution of energy over frequency and time), or by formal statistical clustering techniques (Wadewitz et al., 2015). The principal limitation of identifying call types by acoustics, however, is that this method does not necessarily reveal the call types that are perceptually salient or ecologically relevant to receivers.

Thus, a final method of identifying call types is to test how conspecifics categorize calls, for example, by determining which vocalizations cause dishabituation after a series of one repeating call type (Cheney & Seyfarth, 1988; Fischer, 1998). This approach is also not without its limitations, as the perceptual categories of conspecifics do not necessarily map directly onto acoustic categories, nor to the productive mechanisms that underlie different call types (Kerhsenbaum et al., 2016). However, an advantage of this approach is that it may reveal which groupings or partitioning of calls are biologically relevant for receivers; see, for example, Jhang and Oller's (2017) justification for identifying infant sounds by ear, based on the sounds that parents would perceive and respond to, rather than acoustically. Overall, the project of describing a species' vocal repertoire benefits from combined research efforts exploring both emergent acoustic clusters and perceptually relevant categories, as well as the relationship between the two.

Regardless of how call types are identified, a takeaway is that different-sounding call types such as laugh-like calls and screams in chimpanzees occur in different contexts, serve different functions (i.e., elicit different responses), and derive from different evolutionary antecedents or trajectories, just as is the case with discrete anatomical adaptations. (Which, of course, is not to suggest that the evolution of one call type cannot interact with or constrain another, just as one physical adaptation might constrain another.) To recapitulate the broader terminology of animal communication research, these call types are effectively semi-distinct signals, the existence of which merits separate lines of investigation devoted to unraveling the functions and evolution of each.

One advantage of organizing research this way, effectively starting with the observable contexts and acoustic forms, is that it enables recognition that call types themselves are rarely homogenous. Acoustic subtypes or gradations *within* call types are reported in widespread taxa (e.g., in primates, Crockford & Boesch, 2003; Crockford, Gruber, & Zuberbühler, 2018; Fischer et al., 2001a; Fischer et al., 2001b; Gouzoules, Gouzoules, & Marler, 1984; Gouzoules & Gouzoules, 1989; Rendall, Seyfarth, Cheney, & Owren, 1999) and are likely often functional, in the sense that receivers are perceptually sensitive to this variation and alter their behavior according to the variants they hear (Gouzoules et al., 1984, Slocombe et al., 2009). Thus, variation *within* types (e.g., among screams) and not only *between* types (e.g., screams versus laughs) is potentially relevant to communication.

This within-type versus between-type distinction maps onto a more traditionally-drawn distinction between *vocal production*, or processes determining the spectro-temporal structure of a canonical call type, and *vocal usage*, or processes determining the occurrence of one call type versus another (or the occurrence of a call versus no call) (Seyfarth & Cheney, 2010). Importantly, the proximate sources of these two types of variation, hence the information made available and their potential functions, are not likely equivalent. For example, nonhuman primates and most other mammals seem to command greater voluntary control over call usage than over actions determining the specific, spectro-temporal composition of each call (Seyfarth & Cheney, 2010). Appropriate call usage seems partially learned (Seyfarth et al., 1986), whereas the structure of species-typical calls is largely biologically determined, developing naturally even in animals with atypical or impaired auditory feedback (Owren et al., 1992; Winter et al., 1973), although the literature here is sparse. Moreover, call usage is commonly subject to *audience effects*, where the occurrence of calls depends on the presence and makeup of nearby listeners

(Brosnan & de Waal, 2000; Slocombe et al., 2010a), and at least in some great apes, usage is also potentially contingent on listeners' attentional and possibly even epistemological states (Crockford et al., 2012; Hopkins et al., 2007). Evidence of such flexibility and volitional control over the spectral structure of any given call type, on the other hand, is far rarer (Ackermann et al., 2014; Fitch & Zuberbühler, 2013).

These findings suggest the possibility that, compared to the selective use of one call type versus another, the acoustic variation *within* call types might involve different sets of acoustic parameters and signal information along different dimensions, potentially including some more reliable, indexical cues of emotion (Fischer & Price, 2016; Manser, 2001; Schwartz et al., 2020b). The added level of volitional, learned, and perhaps strategic regulation apparent in call usage may reduce the influence of emotion over the selection of a call, i.e., between-type variation, whereas the variation within types might embody the physiological effects of emotion in ways that callers cannot as easily control. A fearful primate could theoretically withhold a call (Cheney & Seyfarth, 2007, pp. 225-226), or perhaps even choose between some call types, but not as easily mask acoustic cues to fear in a call it produces. Thus, attention to the variation within call types could, in some cases, reveal more about correlations between call acoustics and emotion than an approach that ignores call types altogether.

Here it is relevant to place these arguments within the broader literature on categorization. Of particular relevance is the concept of categorical perception, whereby continuous variation is perceived as discrete categories, such that listeners more readily discriminate acoustic stimuli on different sides of a category boundary while neglecting the variation within categories (Goldstone & Hendrickson, 2010). For example, when English speakers distinguish /b/ and /p/ predominantly based on a continuously varying feature called *vocal onset time* (VOT), there exists a boundary in VOT below which they perceive /b/ and above which they perceive /p/, but on either side of which they ignore VOT variation (Green, Brandley, & Nowicki, 2020). Although this phenomenon—seen in nonhuman animals (e.g., (Fischer et al., 2001) as well as humans—might seem to diminish the importance of within-type variation for receivers, there are a number of reasons why within-type variation is still potentially communicatively significant. First, production-based categories might exist, in the sense that different neural mechanisms underlie the production of different call types, that listeners do *not* treat categorically. For example, acoustic variation within call types might result in partly overlapping acoustic profiles, with listeners attending to and inferring information from what they perceive as continuous variation. Alternatively, listeners might treat one dimension of variation categorically while still attending to other variation within those categories. In the categorical perception of speech, for example, filtering out within-category variation in one parameter (e.g., variation in VOT that falls on the same side of a category boundary) does not preclude within-category variation in other parameters, such as those related to pitch and timber, from informing in other dimensions, such as the speaker's emotion. Nonhuman animals likewise may use some acoustic parameter(s) to distinguish call types while inferring further details about emotion from different variation within each type (as has been described in meerkat alarm calls; Hollen & Manser, 2007). The grouping of calls as belonging to the same type does not imply that they are treated equivalently in every aspect.

This perspective could have especially significant implications for the contextually diverse calls that constitute many primate repertoires (Fischer & Price, 2016). First, it seems probable that contextual diversity is partly the result of the control that callers exert over call usage, as vocalizers may voluntarily generate calls across a variety of emotional, situational, and environmental contexts. Second, acoustic variation within these calls, perhaps proximately generated by variation in emotion, may then provide more nuanced contextual information (Anikin et al., 2018; Smith, 1980, p. 184) or at least allow listeners to infer details related to emotion-correlated aspects such as intensity or urgency (Manser, 2001) that are not conveyed by call selection. For example, chimpanzee (*Pan troglodytes*) “hoo” calls occur in assorted contexts of travel, rest, and alert (e.g., upon seeing a snake; Crockford et al., 2012). Contextual details available to the listener likely play an important role in their interpretations (Smith, 1980, p. 189), but these calls also vary acoustically with the caller’s context (Crockford et al., 2018), opening the possibility that within-type variation helps listeners disambiguate their meanings (assuming the variation is communicatively relevant and not filtered out).

One way to conceptualize these and other within-type variants, articulated by Schamberg et al. (2018), is that they likely share some general functions while certain variants enable more specific functions that rely on slightly deviating patterns of response depending on the emotional context. Thus, whereas all chimpanzee hoos might function to coordinate group movements, acoustically dissociable variants elicit different movement-related behaviors (e.g., approaching versus staying in place) to both the sender’s and receiver’s benefit. Indeed, it seems reasonable to hypothesize that variation within call types is more likely meaningful to receivers, and perhaps even favored by selection, when call usage is so general as to present ambiguous meanings (i.e., in contextually diverse calls), when other contextual cues are lacking for receivers (e.g., in visually occluded environments), and/or when it is of particular ecological relevance that receivers respond in more contextually specific ways than usage alone might permit (Bouchet, Blois-Heulin, & Lemasson, 2013; Crockford et al., 2018; Schamberg et al., 2018). This conception is not to suggest that the relevance of within-type variation is for receivers to distinguish call types. One might rather consider that, if call usage signals the caller's goal or correlates with some general function (e.g., coordinating movement), within-type variation might signal along a different dimension (e.g., emotion) in such a way that integrating these semi-separate sources of acoustic information enables more specific and behaviorally advantageous responses by receivers.

Put another way, one might conceive of between-type and within-type variation as conveying semi-independent dimensions of information, where between-type variation (at least in primates) may better reflect strategically regulated, cognitive processes and within-type variation better reflects more automatic, emotional processes. (We suggest *semi*-independent as it is probable that emotion also correlates with usage of vocalizations—e.g., Pollick & de Waal, 2007—and cognition affects production.) In this sense, some within-type variation is arguably not technically within-type at all, as we would rather expect the same patterns of emotion-related variation to apply across all types. Terminology aside, however, this perspective may further highlight the importance of identifying call types before studying the variation "within" them so as not to confound these different sources of variation. If a researcher is interested in arousal-related variation in rhesus macaques (*Macaca mulatta*), for example, rather than studying these cues across different call types such as coos and screams in one dataset, they are better off studying these cues in coos alone so as not to conflate the effects of arousal with more general acoustic differences between the call types (Briefer, 2012; Schwartz, 2020). Moreover, it then remains an interesting empirical question to study the effects of arousal within screams, as it is not a given that the patterns established in coos will fully apply to a call type that generally occurs in more intense emotional contexts and imposes different biological constraints on vocal production.

It should be noted that acoustic variation within a call type is not the only potential source of information by which receivers might appraise contextually diverse calls. Listeners' own present and past context (e.g. memories of prior interactions with a caller) likely lends even greater specificity to their interpretations (Cheney et al., 1995; Smith, p. 189). For example, a baboon's response to an ambiguous bark—one that could function either to maintain contact or to signal the presence of a predator—might depend not only on variation among barks but also on her assessment of the immediate dangers in her environment (Fischer et al., 2001). More work is needed to delineate the general functional significance of variation within call types, the likely selective forces that shape it, and the processes by which listeners integrate this information with other sources of contextual information to attribute meaning to calls.

Extended example: Variation, function, and emotion in primate screams

In primates, screams were selected for more specialized acoustic structures as they acquired novel and more complex social functions. Many species produce screams during agonistic conflicts to recruit aid from allies, typically kin (Cheney, 1977; de Waal & van Hooff, 1981, Gouzoules et al., 1984). Patterns of aid play an important role in determining agonistic outcomes, which are vital to the acquisition and maintenance of rank in the dominance hierarchy (Cheney, 1977). Rank, in turn, bears widespread implications for reproductive success (Silk, 2002, but see Gouzoules et al., 1982), and kin who provide aid benefit from the caller's success indirectly by virtue of their shared genetics (Hamilton, 1964). Screams are thus the potential targets of strong socioecological selective pressures (Gouzoules & Gouzoules, 2011), as their effects on receiver behavior might yield considerable advantages for senders and receivers.

For several primate species, in the context of these pivotal social roles, screams have acquired variable acoustic forms, such that the acoustic variation within this call type correlates with relevant, contextually diverse aspects of the encounter. For example, classes of rhesus macaque screams associated with potentially severe aggression (e.g., with a high-ranking opponent or involving physical contact, “noisy” screams), and those associated with dominance challenges from lower-ranked individuals, “arched” screams, are acoustically dissociable and elicit more robust responses from kin compared to other classes of screams (Gouzoules et al., 1984; 1986). Contextual details similarly correlate with acoustic variation in the screams of other macaque species (Gouzoules & Gouzoules, 1989, 2000), vervet monkeys (Mercier et al., 2019), and chimpanzees (Slocombe & Zuberbühler, 2005), although the patterns of acoustic differentiation and the information made available are somewhat specific to each species (Gouzoules & Gouzoules, 2000). In every case, it is plausible that this acoustic diversification is partly the result of selection for enhanced communicative efficacy. The evolutionary starting point for these vocalizations was likely more acoustically homogenous, as this is true of the more phylogenetically widespread antipredator screams from which primate screams probably derived. However, selection may have favored the evolution of meaningful variants in the agonistic context as senders and receivers benefited by communicating more nuanced details about the agonistic event.

Of course, contextual diversity is not the only factor relevant to the functions and acoustics of primate screams. It is probably equally important that scream variants convey cues to kinship and/or individual identity, such that listeners can determine whether or not the individual soliciting aid is an ally. Suggestions to the contrary notwithstanding (Owren & Rendall, 2003), prodigious evidence shows that primates not only can recognize kin from screams (Cheney & Seyfarth, 1980; Fugate et al., 2008; Gouzoules et al., 1986; Kojima et al., 2003; Seyfarth & Cheney, 2015), but also recognize unrelated callers and infer complex details about an agonistic event based on the sequence of vocalizations and their knowledge of the individuals involved (Bergman et al., 2003; Slocombe et al., 2010b). Screams might also play a secondary role in averting the caller's agonistic opponent, thus favoring harsh and unpleasant acoustic characteristics, per the affect-conditioning model (Owren & Rendall, 1997), although empirical evidence for this position is lacking. At the same time, given that callers must often communicate with distant and/or visually occluded allies, long-distance propagation and localizability remain paramount to the functions of agonistic screams, hence conserved elements of their acoustic structures, such as high amplitudes and broadband frequencies, that may protect against sound attenuation and ensure localizability (Brown et al., 1979). Thus, the acoustic structures of primate screams likely reflect an intricate balance of pressures related to several interwoven functions and constraints.

In this process, emotion likely plays a fundamental part, given its general, hypothesized causal role in primate vocal production (Owren et al., 2011) and the special emotional salience of agonistic encounters. One proximate mechanism by which correlations between context and scream acoustics occur is probably a predictable variation in internal state and its consequent effects on the voice. For example, the acoustic variation within some classes of rhesus macaque screams correlates with the intensity of the agonism (as operationalized by the occurrence of threat and/or contact aggression), following patterns consistent with broadly reported effects of arousal on mammalian vocal production (Briefer, 2012), although acoustic differences between scream classes are not as readily explained by variation in arousal (Schwartz et al., 2020b). A similar explanation might account for severity-correlated acoustic variation in chimpanzee screams produced by the victims in an encounter (Slocombe & Zuberbühler, 2005).

Again, however, a repeated lesson from research on animal screams is that even in acutely emotional contexts, the communication of emotion does not fully account for call functions or forms. Primate screams are not, as once theorized (Rowell, 1962), straightforward read-outs of fear or signals of submission. From an evolutionary perspective, additional factors, such as selective pressures favoring the conveyance of context- and identity-related details, are responsible for the heterogenous acoustic profiles of screams. Even at the proximate level, other sources likely contribute to acoustic variation among screams. For example, some evidence suggests that chimpanzees might intentionally "exaggerate" the severity of a conflict when a high-ranking potential ally is in the audience (Slocombe & Zuberbühler, 2007). Nor do strictly emotion-based mechanisms seem to readily account for the variable patterns of receiver behavior contingent on who is producing which scream type, and in what sequence (Bergman et al., 2003). The variation within screams, and the responses that different variants elicit, evince a complexity of function that only extensive research focused on this call type could uncover.

Summary: Applications from animal communication research

To summarize some of the salient points from the discussion above, many investigations of animal signals—or, indeed, of animal behaviors in general—begin with a focus on externally observable phenomena, e.g., the acoustically or behaviorally identifiable call types in a vocal repertoire and the contexts in which they occur. Efforts are made to account for these signals functionally, that is, with reference to elicited receiver behaviors and the benefits potentially rendered by communication. The communication of emotion is relevant to signal evolution insofar as it pertains to these functional considerations, but it is not, by itself, always sufficient for explaining the existence of animal calls, nor the precise composition of their acoustic structures. Finally, after establishing call types as categories, it has proven fruitful for researchers to zero in on the acoustic and behaviorally relevant variation within those types. Such considerations may be especially important for calls that occur in diverse contexts, the usage of which may therefore potentially produce ambiguity, provided the listener's own context is not sufficient to disambiguate meaning (Smith, 1980, p. 189).

To apply this perspective to human research, I endorse the following considerations. First, researchers should attempt to define the call type(s) represented in their studies. At a minimum, identifying call types by ear and/or by visual inspection of spectrograms (Struhsaker, 1967)—the most convenient and therefore most commonly applied methods in the animal communication literature—may suffice to enable some focus on those types and then, the functions and communicative significance of each. More sophisticated if labor-intensive methods from the animal literature include asking conspecific listeners to categorize types (Fischer, 1998) and/or demarcating types by acoustic clustering techniques (Fischer et al., 2016).

Second, researchers should investigate the potential functions of calls other than, or in addition to, the communication of emotion. In humans, this approach might entail asking listeners to derive information other than the caller's emotional state, for example, regarding attributes such as sex, body size, or individual identity. Other designs can measure cognitive, behavioral, physiological, and neural responses that more directly test a hypothesized function. Studies correlating these functional measures with acoustics can provide insights regarding the possible evolution of call structure (or the influence of other non-arbitrary processes).

Finally, of particular relevance to this dissertation, once a call type of interest is characterized, researchers should also ask questions about the significance of acoustic variation within that type. Authors in the animal communication literature have long appreciated the prevalence and potential relevance of this kind of variation (Marler, 1976; Schamberg et al., 2018), but with some exceptions, the failure to identify species-typical call types in humans has precluded investigation of within-type variation. This omission is particularly problematic given recent data suggesting that human listeners' recognition of call types and their attribution of the caller's emotion are separate (but complementary) processes: that is, when listeners classify calls as belonging to the same type, they are nonetheless liable to attribute different emotions (Anikin et al., 2018; they also tended to classify call type first and emotion secondly, further underscoring the perceptual salience of these categories). It is probable, therefore, that acoustic variation within human call types enables diverse interpretations and functions just as it does in nonhuman animals. Of course, as noted above, the receiver's context likely also plays an important role in disambiguating these signals. Context is well-known to influence the perception (and production) of both low-level and high-level aspects of speech (e.g., Holt, 2006) as well as facial emotional expressions (Barrett, Mesquita, & Gendron, 2011). It is therefore necessary not only to study call types in isolation but under controlled manipulations of listener context.

I am far from the first to call for or conduct ethologically-styled research on human vocalizations, even if the perspective is still underrepresented. The sentiment, of course, dates back to Darwin (1872/1998). Many authors since have treated human call types as the objects of study and characterized them acoustically (Bachorowksi et al., 2001; Green et al., 2011; Kipper & Todt, 2003; Scheiner et al., 2002/2006), with some attempting to delineate the human vocal repertoire (Anikin et al., 2018). Isolated research programs throughout the years have investigated the functions of human calls beyond emotional communication (Bachrowoski & Owren, 2001; Kersken et al., 2017; McCune et el., 1996; Raine et al., 2018), and some have proposed theoretical frameworks that center on social functional effects of calls on receivers as a primary determinant of call acoustics (Bachorowski & Owren, 2003; see also Fridlund's, 2002, behavioral ecology theory for a prominent model emphasizing the social functional effects of facial expressions). To a more limited degree (with some exceptions described in the section below), researchers have also examined the informational and functional significance of variation within call types (e.g., Engelberg et al., 2019; McCune et al., 1996; Schwartz & Gouzoules, 2019).

At the core of many such efforts is an assumption that human call types are indeed reconcilable to the ethological approach: that is, that they are homologous to animal calls and that similar communicative and evolutionary principles might apply. In the section below, I point to the major supporting evidence for this claim. However, I also caution added deliberation concerning the evolution of language and its potential effects on the human call repertoire, as our language capacities have likely shaped the production and usage of our nonlinguistic vocalizations.

Nonlinguistic vocalizations in humans

The existence of human call types

Mechanistic and acoustic homologies suggest that human cries, screams, laughs, and other call types were likely present in our pre-hominin ancestors (Bryant & Aktipis, 2014; Lingle et al., 2012; Owren et al., 2011), and some prior to primate evolution. The production of (at least some) human nonlinguistic vocalizations and calls in other mammals implicates areas of the brainstem, most notably the reticular formation, which coordinates vocal activity through direct projections to laryngeal motoneurons (Simonyan & Horwitz, 2011), and the periaqueductal gray (PAG), which may play a role in vocal initiation (Jürgens, 2009), and the total destruction of which can result in mutism in humans and nonhumans alike (Esposito et al., 1999; Jürgens, 2002). Limbic areas including the amygdala and hypothalamus project into this pathway, supporting claims that these vocalizations are largely emotion-driven (Owren et al., 2011). Additionally, both human and nonhuman primate call types seem to develop reliably in individuals with limited or atypical auditory experience, or even absent auditory experience altogether (Provine & Emmorey, 2006; Scheiner et al., 2006; Hammerschmidt et al., 2001; Owren et al., 1992; Winter et al., 1973), suggesting that pattern generators in this pathway enable calls with somewhat biologically predetermined structures. Finally, because this pathway primarily controls laryngeal but not the supralaryngeal motoneurons that innervate the tongue and lips, both human and nonhuman call types are often only crudely articulated (Bryant & Aktipis, 2014; Sauter et al., 2010b), resulting in coarse, cross-species acoustic similarities (Young et al., 2017)—to say nothing of specific human call types such as laughter, whose acoustic likenesses to calls in other primates, and possibly more taxonomically distant mammals (Knutson et al., 1998), seem directly traceable to a shared ancestral origin (Davila-Ross et al., 2009).

Thus, there do exist some biologically natural, universal human call types in the sense that their distinctive acoustic forms and the patterns of neural activity involved in their production were at least partly selected and shaped by natural selection (with infant cries being a quintessential example; Newman, 2007). Delineating the human vocal repertoire, however, is a complex and ongoing research challenge, particularly given the flexibility of human vocal production and the enormous variability of our vocalizations, as detailed further below. What we and other researchers are in the process of testing (Anikin et al., 2018; Schwartz et al., 2020a) is the idea that there exist meaningful communicative categories and subcategories within this variability, e.g., by asking listeners to classify call types and respond to variation among them.

The extent to which these categories map onto basic, production-based call types is an empirical question, bolstered by such evidence as universal acoustic forms, distinctive neural underpinnings, and evolutionary homologues; at least in some cases, such as screams, we suspect there is a level of correspondence.

Of course, in humans especially, much of the acoustic variation between and within types and their associations with particular contexts probably also reflects a great deal of learning and flexibility by both senders and receivers. This fact does not preclude the emergence of reliable (but possibly culturally specific) and meaningful call types nor variants thereof, nor does a role of learning preclude the possibility that some acoustic variation is functional in ways beyond communicating emotion. The categories identified by listeners likely reflect a mixture of acoustic clusters resulting from distinct, biologically basic production mechanisms, acoustic clusters resulting from learned vocal patterns, and even some categorical perception of acoustic gradients where acoustic clusters are not present at all.

Possible effects of language on nonlinguistic vocalizations

After humans diverged from other apes, language effected an epochal change in our communication system and its neurocognitive underpinnings. Unlike many nonlinguistic vocalizations, speech production necessarily entails the activity of vast networks spanning frontal and temporal cortices as well as subcortical structures (Simonyan & Horwitz, 2011). Critical among these is the laryngeal motor cortex, which (perhaps uniquely in humans; Fitch & Zuberbühler, 2013) directly and densely innervates the brainstem neurons involved in vocal control (Jürgens, 2009), a connection to which some authors have ascribed our unprecedented caliber of vocal flexibility within the primate order (Ackermann et al., 2014; Fitch & Zuberbühler, 2013). Lesions of the laryngeal motor cortex are therefore observed to compromise finely-controlled vocalizations, speech included, while leaving intact a miscellany of fullyformed cries, laughs, and groans (Groswasser, Korn, Groswasser-Reider, & Solzi, 1988; Jürgens, 2009; lesions of homologous areas in squirrel monkeys also leave their calls intact; Jürgens, 2002). Thus, the neuroarchitecture implicated in human vocal production is commonly portrayed as consisting of semi-dissociable dual pathways, one that is phylogenetically older and supports a repertoire of relatively more pre-programmed nonlinguistic calls, and one that represents a novel innovation enabling learned and flexibly-produced speech (Ackermann et al., 2015; Owren et al., 2011; Simonyan & Horwitz, 2011).

It is highly unlikely, however, that an adaptation as neurally and psychologically consequential as language would not leave indelible marks on our already existing forms of communication. For example, the added flexibility and the further weakening of links between emotion and vocal production that occurred with language (McCune et al., 1996; Oller et al., 2013) probably have influenced the nonlinguistic system. Indeed, there is evidence that the same motor cortical areas involved in volitional speech also participate in the production of some nonlinguistic vocalizations (Wattendorf, 2013). In addition to terminating at the same motoneurons, motor cortical inputs feed into the nonlinguistic pathway directly via afferents to the reticular formation, as well as indirectly via projections to the anterior cingulate cortex (which in turn projects to the PAG; Simonyan & Horwitz, 2011). These connections might explain why, although patients with motor cortical lesions are still capable of emitting nonlinguistic sounds, these spontaneous vocalizations are described as acoustically “bizarre” (Grosswasser et al., 1988; it is also unclear that these vocalizations function as they would in unimpaired individuals). Additionally, although cross-cultural recognition studies indicate that certain call types convey emotions across language groups (Bhatara et al., 2016; Cordaro et al., 2016; Laukka et al., 2013; Sauter et al., 2010), considerable in-group advantages are commonly reported (Gendron et al., 2014; Koeda et al., 2013; Sauter & Scott, 2007; Sauter et al., 2010b), perhaps implying that their structures and meanings are not entirely fixed but are instead somewhat malleable by experience. Findings like these begin to illustrate how the faculties enabling vocal learning and control in speech may also have shaped our production of nonlinguistic call types as well (Pisanski et al., 2016).

The evidence reviewed above amounts to a compelling start, but the full implications of language and its effects on our evolutionarily pre-existing vocal repertoire are yet to be understood. One intriguing possibility is that some of the contextual diversity and acoustic variation in human call types occurs not only because of our enhanced vocal motor capacities, but also because of the psychological faculties enabling the invention and transmission of novel sound-to-meaning correspondences. Speech prosody, for example, sometimes conveys externally referential details, particularly when the linguistic content is ambiguous (Tzeng et al., 2018), which could suggest intentional (or at least non-automatic) disambiguation by speakers. The possibility that vocalizers might similarly use and acoustically modulate nonlinguistic vocalizations to disambiguate meaning has not received much attention. Additionally, the functional flexibility of nonlinguistic vocalizations in adults, that is, the extent to which they occur across differently-valenced emotional contexts, deserves more systematic study. In infants, the protophones that precede speech in ontogeny are marked in part by the greater flexibility of their usage across emotional contexts compared to call types such as laughs and cries (Jhang & Oller, 2017; Oller et al., 2013). It would be informative, however, to learn how the flexibility of laughs, cries, and other vocalizations might expand with language development.

Thus, to the already-outlined questions about call types—their functions and the significance of variation within them—it is worthwhile to add questions about how language has interacted with and altered the functions, perceptions, and patterns of variation within each type. Continued neurological work with language-impaired patients is invaluable here, as are imaging studies that reveal more of the extent to which the neural substrates of language and nonlinguistic calls overlap. However, any psychological study that directly compares the production and/or processing of these two modes of vocal communication is potentially informative (see Laukka et al., 2013; Sauter et al., 2010 for authors who have invoked similar calls for research). Finally, where appropriate, the field could benefit from more directly comparative cross-species work. Evidence is mounting that nonhuman primates may vocalize with more flexibility and intentionality than previously thought (e.g., Crockford et al., 2012), but interactions between linguistic and nonlinguistic systems in humans would predict even more flexible and intentional use of shared call types in our species.

To reiterate, it is not the case that research on human vocalizations has altogether neglected these research questions. Arguably, however, only two call types—infant cries and laughter—have undergone robust (and ongoing) inquiry sufficient to address their functions and acoustic forms. Below, I briefly review some relevant methods and findings from these literatures, with focus on the variation among each call type, by way of illustrating the value of this kind of research.

Infant cries

Infant cries represent the best-studied human call type, likely because of clear applications to childcare and early clinical outcomes (LaGasse et al., 2005). Cry-like vocalizations occur in similar contexts across mammals (Lingle et al., 2012), and research on the neural bases of cry communication has revealed homologies not only in production mechanisms—largely implicating the same mammalian vocal pathway described above—but also the brain areas and neurotransmitters involved in cry processing by parents (reviewed in Newman, 2007). Infant cries therefore likely comprise a highly conserved call type in the mammalian lineage (Lingle et al., 2012; Newman, 2007).

Decades of empirical work have revealed the communicative relevance of acoustic variation among cries. Historically, a prominent debate concerned whether or not distinct cries convey information about discrete internal states and/or eliciting contexts such as hunger, pain, or birth (Fuller & Horii, 1988; Wasz-Hockert et al., 1968; Wolff, 1969; Protopapas & Eimas, 1997; Müller et al., 1973). The bulk of evidence now suggests that cries grade acoustically with the infant's level of distress, and that listeners perceive these varied distress levels (Boero et al., 1998; Gustafson & Harris, 1990; Porter et al., 1986; Wood & Gustafson, 2001). The occurrence of other, non-cry call types might communicate more nuanced contextual information (Kersken et al., 2017; Scheiner et al., 2002). This variation in cries is likely related to the effects of the autonomic stress response on vocal production (Porges et al., 1994), an idea supported by correlating physiological measures (Porter, Porges, & Marshall, 1988).

Many researchers have also investigated the relationship between the acoustics of a cry, the information conveyed, and its functional effects on listeners. Variation among cries seems to communicate important information about the infant's health (Wasz-Hockert, Michelsson, & Lind, 1985) and identity (Green & Gustafson, 1983; Gustafson et al., 2013). It is typical for studies of cry perception to include ratings related to the call's potential functions, e.g., judgments of aversiveness, urgency, or other subjective qualities that might impel quicker responses from listeners (LaGasse, Neal, & Lester, 2005; Li et al., 2018; Zeskind & Marshall, 1988; Zeskind et al., 1992). Other clever designs have measured relevant behavioral or physiological responses more directly, such as Wolff's (1969) classic "playback" experiment that recorded mothers' latencies to respond to different cry types, or research that has explored the effects of cries on measures of autonomic arousal (Boukydis & Burgess, 1982; Out, Pieper, Bakermans-Kranenburg, & Van Ijzendoorn, 2010) and motor responses (Parsons, Young et al., 2012). Supplementing this work is a vast literature detailing the perceptual "cry network" constituting the neural substrates of cry processing (reviewed in Newman, 2007), as well as variation in responsiveness at key nodes that might underlie differences in cry perception (e.g., Li et al., 2018). Research on cries has, in essence, broken down the entire communicative event, from the proximate sources of acoustic variation to the effects of that variation on listeners' perceptual, behavioral, and neural responses.

That said, there are a few notable limitations to extrapolating from infant cry data. The evidence is limited to describing a developmentally immature age group in the singular context of offspring-parent communication (Zeifman, 2001). It is unclear how findings from infant cries might extend to other age groups or contexts, particularly because, whatever faculties the infant brain possesses to make it language-ready, infants are nonetheless prelinguistic. Infant cries, then, are not well-suited for investigating all questions about the interaction of language and nonlinguistic calls, even if some cries potentially already evince certain, subtle hallmarks of their native language (Mampe et al., 2009). A related issue regarding emotional expression is that young infants might not experience, and/or might be physically incapable of vocally expressing the full ambit of emotions that older children or adults do. Thus, this literature, as rich as it is, is not by itself sufficient to understand human vocal behavior and should rather stand as a blueprint for equally intensive research on other call types.

Laughter

Laughter is another vocalization that bears conspicuous similarities to calls in other species, and as such has long provoked speculation (Gervais & Wilson, 2005; Darwin, 1872/1998) and empirical work (Bachorowski & Owren, 2001; Bryant & Aktipis, 2014) on its functions and evolution. Compared to the guttural, laugh-like 'play' vocalizations of nonhuman apes, human laughter is far more acoustically variable and contextually diverse (Gervais & Wilson, 2005; Provine, 2000; Ruch & Ekman, 2001). Thus, in addition to the necessary research establishing its distinguishing acoustic features (Bachorowski et al., 2001; Kipper & Todt, 2003; Todt & Kipper, 2001), much work is devoted to exploring the communicative significance of variation within laughter.

As with infant cries, the question of whether contextually distinct laughs are acoustically and perceptually discriminable has proven somewhat contentious (Owren & Bachorowski, 2003; Edmonson, 1987; Nwokah et al., 1993; Szameitat et al., 2009a). Some authors have argued that laughs are too acoustically variable within each emotional context to exhibit consistent correlations between contexts (Owren & Bachorowski, 2003). However, a more recent spate of evidence suggested that laughs associated with joy, tickle, schadenfreude, and taunting contexts are sometimes discriminable (Szameitat et al., 2009a; Szameitat et al., 2011), that these laughs evince partially distinct acoustic profiles (Szameitat et al., 2009b), and that they elicit different neural responses (Szameitat et al., 2010). Contra the evidence for infant cries, these findings do not seem entirely reducible to differences in emotional intensity or autonomic arousal between these contexts.

This acoustic variation in laughter is also associated with variation in laughter's effects on listeners. A commonly made distinction is that between voiced laughter, produced with vocal fold oscillation and sounding tonal and vowel-like, and unvoiced laughter that does not involve the vocal folds and sounds noisier as a result (Grammer & Eibl-Eibesfeldt, 1990; Bachorowski & Owren, 2001). Bachorowski and Owren (2001) found that voiced laughs evoked more positive judgments than unvoiced laughs and suggested that the primary function of the former is to induce a positive affect in listeners. Additional variation is linked to listeners' interpretations of reward, affiliation, and dominance (Wood et al., 2017), judgments which might further affect listeners' behaviors.

The issue of potential interactions between laughter and language has received some attention, albeit indirectly, in research on volitional versus spontaneous laughter. Some of laughter's diverse uses and acoustic forms are thought possible because of our unique capacities among primates for flexible vocal production, in turn reflecting adaptations for speech (Bryant, 2020). We can and often do produce laughter at will, absent any physical or salient emotional elicitors (Bryant & Aktipis, 2014), which seems to suggest greater flexibility and intentionality compared to more context-specific laugh-like calls in nonhuman primates (Provine, 1996). However, a line of research has revealed that this laughter is not evidently produced nor perceived identically to spontaneously occurring laughter (Bryant & Aktipis, 2014; Bryant et al., 2018; Lavan, Scott, & McGettigan, 2016). Such differences could imply that the added flexibility afforded by our speech production mechanisms nonetheless does not enable perfect, voluntary facsimiles of every nonlinguistic call type. It is not clear, however, that the function of volitional laughs relies on sounding indistinguishable from spontaneous ones.

Summary

The literatures on infant cries and laughter, each only touched upon here, demonstrate the utility of research dedicated to dissecting individual call types and the variation within. A strictly emotion-first approach, one that might have treated cries as generalized signals of distress, and laughter as signals of joy, might have ignored the informative and functionally significant acoustic variation among each of these call types.

At the same time, however, these literatures leave unresolved a great many questions about nonlinguistic vocal communication more generally. The extent to which acoustic variation within call types does convey information about emotional context is unclear, especially given seemingly dissimilar findings between infant cries and laughter, where the former grade more consistently with distress or arousal level, and the latter might show more acoustic differences between qualitatively distinct contexts. Similarly underappreciated are the patterns by which acoustic variation might enable varied communicative functions in a contextually diverse call type. The influence of language remains most elusive, given that, of the two extensively investigated call types, one is mostly limited to prelinguistic infants. (The adult form of crying has received far less attention; Zeifman, 2001.)

These examples speak to the broader imperative for more research of an ethological approach in studying human call types. It is difficult, at this time, to draw generalizable conclusions about the evolution of nonlinguistic vocal communication in humans because, for any given question or hypothesis, we seem to operate with an *N* of 1 or 2 call types in our species: that is, we commonly have data on infant cries and/or laughter but little on other vocalizations. For example, is the contextual diversification of laughter, relative to its nonhuman primate homologue (Davila-Ross et al., 2009), and its apparently concomitant acoustic diversification (Szameitat et al., 2009a) a phenomenon specific to that call type, or does it represent a broader trend that occurred in human evolution? The latter possibility, more so than the former, might suggest a general influence of language on our nonlinguistic calls, but lacking research on other adult call types to equal the depth of research on laughter, it is difficult to test this hypothesis. If collecting comparable data across species is a bedrock of evolutionary research, then we need more data on our own call types to track the evolution of vocal communication in our species.

Human screams and this dissertation

Until quite recently (Arnal et al., 2015; Engelberg & Gouzoules, 2019; Engelberg et al., 2020; Schwartz et al, 2020a; Frühholz et al., 2021; Schwartz & Gouzoules, 2019), research on human screams has been noticeably lacking. Knowledge here may have particularly suffered from the over-prioritization of emotion and under-prioritization of call types. In a typical study of nonlinguistic vocal expression, an actor is provided no constraints or instructions regarding call type, and so he might produce a scream to portray fear, a laugh for happiness, and a growl or yell for anger. It is unsurprising, then, that many researchers still treat screams synonymously with fear (e.g., Armony et al., 2007; Arnal et al., 2015; Cordaro et al., 2016) and that their hypotheses about screams' effects on listeners therefore narrowly concern the capacity for screams to convey alarm or induce fear in listeners (Arnal et al., 2015). What this kind of research fails to reveal is why screams in fact occur in all three of these emotional contexts—fear, happiness, and anger, among others (Anikin & Persson, 2017)—and how acoustic variation among screams might enable any discrimination between these contexts by listeners.

Screams may, however, represent the most promising empirical targets for advancing our knowledge of nonlinguistic vocal communication, especially regarding the acoustic, perceptual, and functional variation within a call type. Although characterizing the human vocal repertoire is a thorny endeavor, screams have been established in perceptual and acoustic research as a readily identifiable call type (Anikin et al., 2018; Schwartz et al., 2020a). Thus, there is strong evidence that screams are at least a conceptually meaningful category for listeners. However, we strongly suspect this category reflects some biological basis—that is, that screams and their general neural underpinnings were specifically shaped by selection—because there is some evidence that they are recognized across language groups (Anikin et al., 2018) and because highly acoustically similar vocalizations occur in (some) similar contexts in nonhuman primates and other species (Collias, 1960; Møller & Nielsen, 2010; Slocombe & Zuberbühler, 2005). Of course, these data do not suggest that variation among screams is fixed or predetermined, only that we likely can trace an evolutionary trajectory for some basic form(s) of screams.

The distinguishing acoustic features of human screams are not yet entirely defined, but recent work from our lab and others suggests they are marked by high pitch and harmonic elements, wide pitch variability (Anikin et al., 2018; Green et al., 2011; Schwartz et al., 2020a), and the acoustic quality of roughness (Arnal et al., 2015). These features seem to reflect the elevated intensity or arousal levels associated with screams, which by itself is a potentially interesting constraint, given indications that humans sometimes have difficulty discriminating emotions from high-intensity vocalizations (Patel et al., 2011) and facial expressions (Aviezer et al., 2012), or from expressions with otherwise similar arousal profiles (Russell et al., 2003). Nonetheless, listeners somehow identify a variety of emotions from vocalizations they classify as screams (Anikin et al., 2018).

Our understanding of nonhuman primate screams and the significance of variation among them is rich (Bergman et al., 2003; Gouzoules et al., 1984, 1986; Gouzoules & Gouzoules, 2000; Slocombe & Zuberbühler, 2005; Slocombe et al., 2009) and still expanding (Mercier et al., 2019), constituting a well-grounded basis for comparative, hypothesis-driven work on human screams. Additionally, like screams in nonhuman primates and other taxa, human screams often reach distant listeners who might lack visual contextual cues. Human screams, however, are evidently more contextually diverse, not only compared to screams in other animals, but perhaps even compared to other human call types such as laughs and cries (Anikin & Persson, 2017). For example, human children scream often during play (Sherman, 1975) whereas the same is not true for nonhuman primate juveniles. Play therefore represents one context in which the occurrence of screams seems to be a more recent development, but why this development occurred and its effects on listeners' interpretations of screams remain unclear.

The occurrence of a human scream, then, is potentially an event as ambiguous as it is intense, and one in which a failure for listeners to respond appropriately could conceivably bear dire consequences. I have postulated that variation within call types is likely to be relevant when the occurrence of a call might offer ambiguous information, and when eliciting more contextually specific responses from listeners is plausibly significant. If this is the case, then variation among screams would seem particularly likely to play an important communicative role.

Finally, screams allow consideration of interactions with language. Screams are observed at least in young children (Green et al., 2011) and continue into adulthood (Anikin et al., 2018), albeit likely undergoing yet-undocumented changes in usage with age. We can therefore study scream production in subjects with fully developed linguistic faculties, adding a datapoint to those call types whose functions and perceptions might change with language. Other interesting opportunities may present themselves down the line. For example, we can track the trajectory of scream functions and acoustics longitudinally in individuals whose language abilities are still developing; such data would lend themselves to comparison with similar ontogenetic studies of screams in nonhuman primates (Gouzoules & Gouzoules, 1995). Finally, screamed speech, evidently comprising an acoustic combination of the linguistic and nonlinguistic domains, seems an interesting area for exploring the interactions between these two modes of vocal communication.

The studies in this dissertation represent a continuation of work in our lab that has approached human screams as a basic call type of our species. This dissertation focuses on perceptual and acoustic differences between screams associated with disparate emotional contexts, thus comprising an investigation of how the contextual diversity in screams is matched by diversification of their forms and functions.

In Study 1, we investigate whether acoustic variation in screams associated with diverse emotional contexts results in consistent differences in ratings on six emotion prompts. We characterize the primary dimensions along which listeners' perceptions of screams' emotional content vary, and we describe some acoustic variation among screams that accounts for ratings on each emotion prompt. Note, however, that the design of this study only enabled exploration of the acoustic cues predicting different emotional *perceptions*, not of the acoustic differences between screams associated with different emotions, which is addressed in Study 2. The aim of this study was rather to better establish and more fully characterize the heterogeneity of emotional perceptions elicited by screams.

In Study 2, we expand on this line of research by asking participants to categorize screams into one of five different emotional contexts, enabling a more direct assessment of hits and false alarms. We directly examine how screams produced so as to reflect different emotions differ acoustically, and we compare participants' patterns of response and predictive acoustic cues to an analogous task using short prosodic speech samples, possibly shedding more light on the mechanisms underlying variation in screams. Finally, we ask participants to rate screams on five different perceptual scales, thus exploring potentially functionally relevant effects of screams in addition to the communication of emotion.

References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences, 37*(6), 529.

Adolphs, R. (2017). How should neuroscience study emotions? By distinguishing emotion states, concepts, and experiences. *Social cognitive and affective neuroscience, 12*(1), 24-31.

Anikin, A., & Persson, T. (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods, 49*(2), 758-771.

Anikin, A., Bååth, R., & Persson, T. (2018). Human Non-linguistic Vocal Repertoire: Call Types and Their Meaning. *Journal of Nonverbal Behavior, 42*(1), 53-80.

Armony, J. L., Chochol, C., Fecteau, S., & Belin, P. (2007). Laugh (or cry) and you will be remembered: influence of emotional expression on memory for vocalizations. *Psychological Science, 18*(12), 1027-1029.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology, 25*(15), 2051-2056.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science, 338*(6111), 1225-1229.

Bachorowski, J. A., & Owren, M. J. (2001). Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science, 12*(3), 252-257.

Bachorowski, J. A., & Owren, M. J. (2003). Sounds of emotion: Production and perception of affect-related vocal acoustics. *Annals of the New York Academy of Sciences, 1000*(1), 244-265.

Bachorowski, J. A., Smoski, M. J., & Owren, M. J. (2001). The acoustic features of human laughter. *The Journal of the Acoustical Society of America, 110*(3), 1581-1597.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science, 20*(5), 286-290.

Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science, 302*(5648), 1234-1236.

Bhatara, A., Laukka, P., Boll-Avetisyan, N., Granjon, L., Elfenbein, H. A., & Bänziger, T. (2016). Second language ability and emotional prosody perception. *PLOS ONE, 11*(6), e0156855.

Boero, D. L., Bianchi, C., Volpe, C., Marcello, A., & Lenti, C. (1998). Newborns crying in different contexts: discrete or graded signals?. *Perceptual and Motor Skills, 86*(3_suppl), 1123-1140.

Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology, 41*(2), 259-268.

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2013). Social complexity parallels vocal complexity: a comparison of three non-human primate species. *Frontiers in Psychology, 4*, 390.

Boukydis, C. Z., & Burgess, R. L. (1982). Adult physiological response to infant cries: Effects of temperament of infant, parental status, and gender. *Child Development, 53*(5), 1291-1298.

Bradbury, J. W., & Vehrencamp, S. L. (2011). *Principles of animal communication (2nd ed.)*. Sunderland, MA: Sinauer Associates.

Brehm, G., Fischer, M., Gorb, S., Kleinteich, T., Kühn, B., Neubert, D., ... & Wurdinger, S. (2015). The unique sound production of the Death's-head hawkmoth (Acherontia atropos (Linnaeus, 1758)) revisited. *The Science of Nature, 102*(7-8), 43.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology, 288*(1), 1-20.

Brosnan, S. F., & de Waal, F. B. (2000). Regulation of vocal output by chimpanzees finding food in the presence or absence of an audience. *Evolution of Communication, 4*(2), 211-224.

Brown, C. H., Beecher, M. D., Moody, D. B., & Stebbins, W. C. (1979). Locatability of vocal signals in Old World monkeys: Design features for the communication of position. *Journal of Comparative and Physiological Psychology, 93*(5), 806.

Bryant, G. A. (2020). *Evolution, Structure, and Functions of Human Laughter.* The Handbook of Communication Science and Biology.

Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evolution and Human Behavior, 35*(4), 327-335.

Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Amir, D., Chávez, B., ... & Fux, M. (2018). The perception of spontaneous and volitional laughter across 21 societies. *Psychological Science, 29*(9), 1515-1525.

Cheney, D. L. (1977). The acquisition of rank and the development of reciprocal alliances among free-ranging immature baboons. *Behavioral Ecology and Sociobiology, 2*(3), 303-318.

Cheney, D. L., & Seyfarth, R. M. (1980). Vocal recognition in free-ranging vervet monkeys. *Animal Behaviour, 28*(2), 362-367.

Cheney, D. L., & Seyfarth, R. M. (1988). Assessment of meaning and the detection of unreliable signals by vervet monkeys. *Animal Behaviour, 36*(2), 477-486.

Cheney, D. L., & Seyfarth, R. M. (2008). *Baboon metaphysics: The evolution of a social mind.* University of Chicago Press.

Cheney, D. L., Seyfarth, R. M., & Silk, J. B. (1995). The role of grunts in reconciling opponents and facilitating interactions among adult female baboons. *Animal behaviour, 50*(1), 249-257.

Clark, A. P., & Wrangham, R. W. (1993). Acoustic analysis of wild chimpanzee pant hoots: Do Kibale Forest chimpanzees have an acoustically distinct food arrival pant hoot?. *American Journal of Primatology, 31*(2), 99-109.

Clay, Z., Archbold, J., & Zuberbühler, K. (2015). Functional flexibility in wild bonobo vocal behaviour. *PeerJ, 3*, e1124.

Collias, N. E. (1960). An ecological and functional classification of animal sounds. *Animal Sounds and Communication, 7*, 368-391.

Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion, 16*(1), 117.

Crockford, C., & Boesch, C. (2003). Context-specific calls in wild chimpanzees, Pan troglodytes verus: analysis of barks. *Animal Behaviour, 66*(1), 115-125.

Crockford, C., Gruber, T., & Zuberbühler, K. (2018). Chimpanzee quiet hoo variants differ according to context. *Royal Society Open Science, 5*(5), 172066.

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology, 22*(2), 142-146.

Darwin, C., & Prodger, P. (1998). *The expression of the emotions in man and animals.* Oxford University Press, USA.

Davila-Ross, M., Owren, M., & Zimmermann, E. (2009). Reconstructing the evolution of laughter in great apes and humans. *Current Biology, 19*.

de Waal, F. B. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology, 59*, 279-300.

de Waal, F. B. (2011). What is an animal emotion?.*Annals of the New York Academy of Sciences, 1224*(1), 191-206.

de Waal, F. B., & van Hooff, J. A. (1981). Side-directed communication and agonistic interactions in chimpanzees. *Behaviour, 77*(3), 164-198.

Driver, P. M., & Humphries, D. A. (1969). The significance of the high-intensity alarm call in captured passerines. *Ibis, 111*(2), 243-244.

Edmonson, M. S. (1987). Notes on laughter. *Anthropological Linguistics*, 23-34.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3-4), 169-200.

Engelberg, J. W., & Gouzoules, H. (2019). The credibility of acted screams: Implications for emotional communication research. *Quarterly Journal of Experimental Psychology, 72*(8), 1889-1902.

Engelberg, J. W., Schwartz, J. W., & Gouzoules, H. (2019). Do human screams permit individual recognition?. *PeerJ, 7*, e7087.

Evans, C., & Smith, C. (2011). Exaggeration of display characteristics enhances detection of visual signals. *Behaviour, 148*(3), 287-305.

Esposito, A., Demeurisse, G., Alberti, B., & Fabbro, F. (1999). Complete mutism after midbrain periaqueductal gray lesion. *Neuroreport, 10*(4), 681-685.

Fischer, J. (1998). Barbary macaques categorize shrill barks into two call types. *Animal Behaviour, 55*(4), 799-807.

Fischer, J., & Price, T. (2016). Meaning, intention, and inference in primate vocal communication. *Neuroscience & Biobehavioral Reviews, 82*, 22-31.

Fischer, J., Hammerschmidt, K., Cheney, D. L., & Seyfarth, R. M. (2001). Acoustic features of female chacma baboon barks. *Ethology, 107*(1), 33-54.

Fischer, J., Kitchen, D. M., Seyfarth, R. M., & Cheney, D. L. (2004). Baboon loud calls advertise male quality: acoustic features and their relation to rank, age, and exhaustion. *Behavioral Ecology and Sociobiology, 56*(2), 140-148.

Fischer, J., Metz, M., Cheney, D. L., & Seyfarth, R. M. (2001). Baboon responses to graded bark variants. *Animal Behaviour, 61*(5), 925-931.

Fischer, J., Wadewitz, P., & Hammerschmidt, K. (2016). Structural variability and communicative complexity in acoustic communication. *Animal Behaviour, 134*, 229-237.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America, 102*(2), 1213-1222.

Fitch, W. T., & Zuberbühler, K. (2013). Primate precursors to human language: beyond discontinuity. In E. Altenmuller, S. Schmidt, & E. Zimmermann (Eds.), *The evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man* (26-48). Oxford, UK: Oxford University Press.

Fridlund, A. (1992). The behavioral ecology and sociality of human faces. In M. S. Clark (Ed.), *Review of personality and social psychology* (pp. 90-121). Newbury Park, CA: Sage.

Frühholz, S., Dietziker, J., Staib, M., & Trost, W. (2021). Neurocognitive processing efficiency for discriminating human non-alarm rather than alarm scream calls. *PLoS biology, 19*(4), e3000751.

Fugate, J. M., Gouzoules, H., & Nygaard, L. C. (2008). Recognition of rhesus macaque (Macaca mulatta) noisy screams: evidence from conspecifics and human listeners. *American Journal of Primatology: Official Journal of the American Society of Primatologists, 70*(6), 594-604.

Fullard, J. H., & Napoleone, N. (2001). Diel flight periodicity and the evolution of auditory defences in the Macrolepidoptera. *Animal Behaviour, 62*(2), 349-368.

Fuller, B. F., & Horii, Y. (1988). Spectral energy distribution in four types of infant vocalizations. *Journal of Communication Disorders, 21*(3), 251-261.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science, 25*(4), 911-920.

Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *The Quarterly Review of Biology, 80*(4), 395-430.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(1), 69-78.

Gouzoules, H., & Gouzoules, S. (1989). Design features and developmental modification of pigtail macaque, Macaca nemestrina, agonistic screams. *Animal Behaviour, 37*, 383-401.

Gouzoules, H., & Gouzoules, S. (1995). Recruitment screams of pigtail monkeys (Macaca nemestrina): ontogenetic perspectives. *Behaviour, 132*(5), 431-450.

Gouzoules, H., & Gouzoules, S. (2000). Agonistic screams differ among four species of macaques: the significance of motivation-structural rules. *Animal Behaviour, 59*(3), 501-512.

Gouzoules, H., & Gouzoules, S. (2011). The conundrum of communication. In C. J. Campbell, A. Fuentes, K. C. MacKinnon, R. Stumpf, & S. K. Bearder (Eds.), *Primates in Perspective* (2nd ed., pp. 626–636). New York: Oxford University Press.

Gouzoules, H., Gouzoules, S., & Marler, P. (1986). Vocal communication: a vehicle for the study of social relationships. In R. Rawlins & M. Kessler. (Eds.), *The Cayo Santiago macaques: History, behavior, and biology* (pp. 111-129). Albany, NY: SUNY Press.

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (Macaca mulatta) screams: representational signalling in the recruitment of agonistic aid. *Animal Behaviour, 32*(1), 182-193.

Grammer, K., & Eibl-Eibesfeldt, I. (1990). The ritualisation of laughter. In W. Koch (Ed.), *Natürlichkeit der Sprache und der Kultur* (pp. 192-214). Bochum, Germany: Brockmeyer.

Green, P. A., Brandley, N. C., & Nowicki, S. (2020). Categorical perception in animal communication and decision-making. *Behavioral Ecology, 31*(4), 859-867.

Green, J. A., & Gustafson, G. E. (1983). Individual recognition of human infants on the basis of cries alone. *Developmental Psychobiology, 16*(6), 485-493.

Green, J. A., Whitney, P. G., & Potegal, M. (2011). Screaming, yelling, whining, and crying: categorical and intensity differences in vocal expressions of anger and sadness in children's tantrums. *Emotion, 11*(5), 1124-1133.

Groswasser, Z., Korn, C., Groswasser-Reider, I., & Solzi, P. (1988). Mutism associated with buccofacial apraxia and bihemispheric lesions. *Brain and Language, 34*(1), 157-168.

Gustafson, E., Levréro, F., Reby, D., & Mathevon, N. (2013). Fathers are just as good as mothers at recognizing the cries of their baby. *Nature Communications, 4*, 1698.

Gustafson, G. E., & Harris, K. L. (1990). Women's responses to young infants' cries. *Developmental Psychology, 26*(1), 144.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology, 7*(1), 17-52.

Hammerschmidt, K., Freudenstein, T., & Jürgens, U. (2001). Vocal development in squirrel monkeys. *Behaviour, 138*(9), 1179-1204.

Hawk, S., van Kleef, G., Fischer, A., & van der Schalk, J. (2009). "Worth a thousand words": absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion, 9*(3), 293.

Hogstedt, G. (1983). Adaptation unto death: function of fear screams. *The American Naturalist, 121*(4), 562-570.

Hollén, L. I., & Manser, M. B. (2007). Motivation before meaning: motivational information encoded in meerkat alarm calls develops earlier than referential information. *The American Naturalist, 169*(6), 758-767.

Holt, L. L. (2006). Speech categorization in context: Joint effects of nonspeech and speech precursors. *The Journal of the Acoustical Society of America, 119*(6), 4016-4026.

Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour, 73*(2), 281-286.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review, 99*(3), 561-565.

Jhang, Y., & Oller, D. K. (2017). Emergence of functional flexibility in infant vocalizations of the first 3 months. *Frontiers in psychology, 8*, 300.

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews, 26*(2), 235-258.

Jürgens, U. (2009). The neural control of vocalization in mammals: a review. *Journal of Voice, 23*(1), 1-10.

Jürgens, U., & Ploog, D. (1970). Cerebral representation of vocalization in the squirrel monkey. *Experimental Brain Research, 10*(5), 532-554.

Kaltwasser, M. T. (1990). Startle-inducing acoustic stimuli evoke ultrasonic vocalization in the rat. *Physiology & behavior, 48*(1), 13-17.

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., ... & Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biological Reviews, 91*(1), 13-52.

Kersken, V., Zuberbühler, K., & Gomez, J. C. (2017). Listeners can extract meaning from non-linguistic infant vocalisations cross-culturally. *Scientific Reports, 7*(1), 1-7.

Kipper, S., & Todt, D. (2003). The role of rhythm and pitch in the evaluation of human laughter. *Journal of Nonverbal Behavior, 27*(4), 255-272.

Knutson, B., Burgdorf, J., & Panksepp, J. (1998). Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *Journal of Comparative Psychology, 112*(1), 65.

Koeda, M., Belin, P., Hama, T., Masuda, T., Matsuura, M., & Okubo, Y. (2013). Cross-cultural differences in the processing of non-verbal affective vocalizations by Japanese and Canadian listeners. *Frontiers in Psychology, 4*, 105.

Kojima, S., Izumi, A., & Ceugniet, M. (2003). Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates, 44*(3), 225-230.

LaGasse, L. L., Neal, A. R., & Lester, B. M. (2005). Assessment of infant cry: acoustic cry analysis and parental perception. *Mental Retardation and Developmental Disabilities Research Reviews, 11*(1), 83-93.

Laporte, M. N., & Zuberbühler, K. (2010). Vocal greeting behaviour in wild chimpanzee females. *Animal Behaviour, 80*(3), 467-473.

Larson, C. R. (1991). On the relation of PAG neurons to laryngeal and respiratory muscles during vocalization in the monkey. *Brain Research, 552*(1), 77-86.

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Iraki, F. K. E., ... & Thingujam, N. S. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology, 4*, 353.

Lavan, N., Scott, S. K., & McGettigan, C. (2016). Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter. *Journal of Nonverbal Behavior, 40*(2), 133-149.

Li, T., Horta, M., Mascaro, J. S., Bijanki, K., Arnal, L. H., Adams, M., ... & Rilling, J. K. (2018). Explaining individual variation in paternal brain responses to infant cries. *Physiology & Behavior, 193*, 43-54.

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: a corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods, 45*(4), 1234-1245.

Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., & Romanow, C. A. (2012). What makes a cry a cry? A review of infant distress vocalizations. *Current Zoology, 58*(5), 698-726.

Macedonia, J. M., & Evans, C. S. (1993). Essay on contemporary issues in ethology: variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology, 93*(3), 177-197.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*(23), 1994-1997.

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 268*(1483), 2315-2324.

Marler (1976). Social organization, communication and graded signals: the chimpanzee and gorilla. In P. Bateson & R. Hinde (Eds.), *Growing Pains in Ethology* (239–277). Cambridge, UK: Cambridge University Press.

Maynard Smith, M. J., & Harper, D. G. (1995). Animal signals: models and terminology. *Journal of Theoretical Biology, 177*(3), 305-311.

Mayr, E. (1961). Cause and effect in biology. *Science, 134*(3489), 1501-1506.

McCune, L., Vihman, M. M., Roug-Hellichius, L., Delery, D. B., & Gogate, L. (1996). Grunt communication in human infants (Homo sapiens). *Journal of Comparative Psychology, 110*(1), 27-37.

Mercier, S., Déaux, E. C., van de Waal, E., Bono, A. E., & Zuberbühler, K. (2019). Correlates of social role and conflict severity in wild vervet monkey agonistic screams. *PloS one, 14*(5), e0214640.

Møller, A. P., & Nielsen, J. T. (2010). Fear screams and adaptation to avoid imminent death: effects of genetic variation and predation. *Ethology Ecology & Evolution, 22*(2), 183-202.

Müller, E., Hollien, H., & Murry, T. (1974). Perceptual responses to infant crying: Identification of cry types. *Journal of Child Language, 1*, 89–95

Newman, J. D. (2007). Neural circuits underlying crying and cry responding in mammals. *Behavioural Brain Research, 182*(2), 155-165.

Nwokah, E., Davies, P., Islam, A., Hsu, H., & Fogel, A. (1993). Vocal affect in three-year-olds: A quantitative acoustic analysis of child laughter. *The Journal of the Acoustical Society of America, 94*(6), 3076-3090.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112*(1), 181-186.

Oller, D., Buder, E., Ramsdell, H., Warlaumont, A., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences, 110*(16), 6318-6323.

Out, D., Pieper, S., Bakermans-Kranenburg, M. J., & Van IJzendoorn, M. V. (2010). Physiological reactivity to infant crying: a behavioral genetic study. *Genes, Brain and Behavior, 9*(8), 868-876.

Owren, M. J., & Bachorowski, J. A. (2003). Reconsidering the evolution of nonlinguistic communication: The case of laughter. *Journal of Nonverbal Behavior, 27*(3), 183-200.

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews, 10*(2), 58-71.

Owren, M., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocal signaling. In D. H. Owings, M. D. Beecher, & N. S. Thompson (Eds.) *Perspectives in ethology: Volume 12: Communication* (pp. 299-346). New York, NY: Plenum.

Owren, M., & Rendall, D. (2003). Salience of caller identity in rhesus monkey (Macaca mulatta) coos and screams: perceptual experiments with human (Homo sapiens) listeners. *Journal of Comparative Psychology, 117*(4), 380-390.

Owren, M., Amoss, R., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology, 73*(6), 530-544.

Owren, M., Dieter, J., Seyfarth, R., & Cheney, D. (1992). 'Food' calls produced by adult female rhesus (Macaca mulatta) and Japanese (M. fuscata) macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behaviour, 120*(3), 218-231.

Panksepp, J., & Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion review, 3*(4), 387-396.

Parsons, C. E., Young, K. S., Parsons, E., Stein, A., & Kringelbach, M. L. (2012). Listening to infant distress vocalizations enhances effortful motor performance. *Acta Paediatrica, 101*(4), e189.

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology, 87*(1), 93-98.

Paul, E. S., & Mendl, M. T. (2018). Animal emotion: Descriptive and prescriptive definitions and their implications for a comparative perspective. *Applied Animal Behaviour Science, 205*, 202-209.

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology, 111*, 14-25.

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*(4), 304-318.

Pollick, A. S., & de Waal, F. B. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences, 104*(19), 8184-8189.

Porges, S. W., Doussard-Roosevelt, J. A., & Maiti, A. K. (1994). Vagal tone and the physiological regulation of emotion. *Monographs of the Society for Research in Child Development*, 167-186.

Porter, F. L., Miller, R. H., & Marshall, R. E. (1986). Neonatal pain cries: effect of circumcision on acoustic features and perceived urgency. *Child Development,* 790-802.

Porter, F. L., Porges, S. W., & Marshall, R. E. (1988). Newborn pain cries and vagal tone: Parallel changes in response to circumcision. *Child Development*, 495-505.

Premack, D. (1972). Concordant preferences as a precondition for affective but not for symbolic communication (or how to do experimental anthropology). *Cognition, 1*(2-3), 251-264.

Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports, 5.*

Protopapas, A., & Eimas, P. D. (1997). Perceptual differences in infant cries revealed by modifications of acoustic features. *The Journal of the Acoustical Society of America, 102*(6), 3723-3734.

Provine, R. R. (1996). Laughter. *American scientist, 84*(1), 38-45.

Provine, R. R. (2000). The science of laughter. *Psychology Today, 33*(6), 58-62.

Provine, R. R., & Emmorey, K. (2006). Laughter among deaf signers. *Journal of Deaf Studies and Deaf Education, 11*(4), 403-409.

Raine, J., Pisanski, K., & Reby, D. (2017). Tennis grunts communicate acoustic cues to sex and contest outcome. *Animal Behaviour, 130*, 47-55.

Raine, J., Pisanski, K., Oleszkiewicz, A., Simner, J., & Reby, D. (2018). Human listeners can accurately judge strength and height relative to self from aggressive roars and speech. *iScience, 4*, 273-280.

Rendall, D., Seyfarth, R. M., Cheney, D. L., & Owren, M. J. (1999). The meaning and function of grunt variants in baboons. *Animal Behaviour, 57*(3), 583-592.

Rohwer, S., Fretwell, S., & Tuckfield, R. (1976). Distress screams as a measure of kinship in birds. *American Midland Naturalist*, 418-430.

Rowell, T. E. (1962). Agonistic noises of the rhesus monkey (Macaca mulatta). In *Symp. Zool. Soc. Lond* (Vol. 8, pp. 91-96).

Ruch, W., & Ekman, P. (2001). The expressive pattern of laughter. In A. W. Kaszniak (Ed.), *Emotion, Qualia, and Consciousness* (pp. 426-443). Tokyo, Japan: Word Scientific.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145-172.

Russell, J., Bachorowski, J., & Fernández-Dols, J. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*(1), 329-349.

Sauter, D., & Eimer, M. (2009). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience, 22*(3), 474-481.

Sauter, D., & Scott, S. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion, 31*(3), 192-199.

Sauter, D., Eisner, F., Calder, A., & Scott, S. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology, 63*(11), 2251-2272.

Sauter, D., Eisner, F., Ekman, P., & Scott, S. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, 107*(6), 2408-2412.

Sauter, D., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology, 31*(1), 97-113.

Schamberg, I., Wittig, R. M., & Crockford, C. (2018). Call type signals caller goal: a new take on ultimate and proximate influences in vocal production. *Biological Reviews, 93*(4).

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of Voice, 16*(4), 509-529.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2006). Vocal expression of emotions in normally hearing and hearing-impaired infants. *Journal of Voice, 20*(4), 585-604.

Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1), 227-256.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication, 40*(1), 99-116.

Schwartz, J. W. (2020). Vocal emotion expression across contexts, vocalization types, and species: Implications for general processes of vocal evolution. (Doctoral dissertation). Retrieved from https://etd.library.emory.edu/concern/etds/jw827c636

Schwartz, J. W., & Gouzoules, H. (2019). Decoding human screams: perception of emotional arousal from pitch and duration. *Behaviour, 156*(13-14), 1283-1307.

Schwartz, J. W., Engelberg, J. W., & Gouzoules, H. (2020a). Was that a scream? Listener agreement and major distinguishing acoustic features. *Journal of Nonverbal Behavior*, 1-20.

Schwartz, J. W., Engelberg, J. W. M., & Gouzoules, H. (2020b). Evolving views on cognition in animal vocal communication: Contributions from scream research. *Animal Behavior and Cognition, 7*(2), 192-213.

Seyfarth, R. M., & Cheney, D. L. (1986). Vocal development in vervet monkeys. *Animal Behaviour, 34*(6), 1640-1658.

Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language, 115*(1), 92-100.

Seyfarth, R. M., & Cheney, D. M. (2015). The Evolution of Concepts about Agents: Or, What Do Animals Recognize When They Recognize an Individual? In E. Margolis, & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 57-75). Cambridge, MA: MIT Press.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science, 210*(4471), 801-803.

Seyfarth, R., & Cheney, D. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology, 54*(1), 145-173.

Silk, J. B. (2002). Kin selection in primate groups. *International Journal of Primatology, 23*(4), 849-875.

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist, 17*(2), 197-208.

Slocombe, K., & Zuberbühler, K. (2005). Agonistic screams in wild chimpanzees (Pan troglodytes schweinfurthii) vary as a function of social role. *Journal of Comparative Psychology, 119*(1), 67-77.

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences, 104*(43), 17228-17233.

Slocombe, K. E., Kaller, T., Call, J., & Zuberbühler, K. (2010). Chimpanzees extract social information from agonistic screams. *PLoS One, 5*(7), e11473.

Slocombe, K. E., Townsend, S. W., & Zuberbühler, K. (2009). Wild chimpanzees (Pan troglodytes schweinfurthii) distinguish between different scream types: evidence from a playback study. *Animal Cognition, 12*(3), 441-449.

Slocombe, K., & Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour, 72*(5), 989-999.

Sherman, L. W. (1975). An ecological study of glee in small groups of preschool children. *Child Development,* 53-61.

Smith, W. J. (1980). *The behavior of communicating.* Harvard University Press.

Struhsaker, T. T. (1967). Social structure among vervet monkeys (Cercopithecus aethiops). *Behaviour*, 83-121.

Szameitat, D. P., Alter, K., Szameitat, A. J., Darwin, C. J., Wildgruber, D., Dietrich, S., & Sterr, A. (2009). Differentiation of emotions in laughter at the behavioral level. *Emotion, 9*(3), 397.

Szameitat, D. P., Darwin, C. J., Wildgruber, D., Alter, K., & Szameitat, A. J. (2011). Acoustic correlates of emotional dimensions in laughter: arousal, dominance, and valence. *Cognition and Emotion, 25*(4), 599-611.

Szameitat, D., Alter, K., Szameitat, A., Wildgruber, D., Sterr, A., & Darwin, C. (2009). Acoustic profiles of distinct emotional expressions in laughter. *The Journal of the Acoustical Society of America, 126*(1), 354-366.

Szameitat, D., Kreifelts, B., Alter, K., Szameitat, A., Sterr, A., Grodd, W., & Wildgruber, D. (2010). It is not always tickling: distinct cerebral responses during perception of different laughter types. *Neuroimage, 53*(4), 1264-1271.

Tinbergen, N. (1952). "Derived" activities; their causation, biological significance, origin, and emancipation during evolution. *The Quarterly Review of Biology, 27*(1), 1-32.

Todt, D., & Kipper, S. (2001). Variation of sound parameters affects the evaluation of human laughter. *Behaviour, 138*(9), 1161-1178.

Toledo, L. F., Sazima, I., & Haddad, C. F. (2011). Behavioural defences of anurans: an overview. *Ethology Ecology & Evolution, 23*(1), 1-25.

Townsend, S. W., & Manser, M. B. (2013). Functionally referential communication in mammals: the past, present and the future. *Ethology, 119*(1), 1-11.

Tzeng, C. Y., Duan, J., Namy, L. L., & Nygaard, L. C. (2018). Prosody in speech as a source of referential information. Language, *Cognition and Neuroscience, 33*(4), 512-526.

Van Hooff, J. A. (1972). A comparative approach to the phylogeny of laughter and smiling. In R. A. Hinde (Ed.), *Non-verbal communication*. Oxford, UK: Cambridge University Press

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing vocal repertoires—Hard vs. soft classification approaches. *PloS one, 10*(4), e0125785.

Wasz-Höckert, O., Michelsson, K., & Lind, J. (1985). Twenty-five years of Scandinavian cry research. In *Infant crying* (pp. 83-104). Springer, Boston, MA.

Wasz-Hockert, O. (1968). The infant cry: A spectrographic and auditory analysis. *Clinics in Developmental Medicine*, 1-42.

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

Winter, P., Schott, D., Ploog, D., & Handley, P. (1973). Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour, 47*(3-4), 230-239.

Wolff, P. H. (1969). The natural history of crying and other vocalizations in early infancy. *Determinants of Infant Behavior*, 81-111.

Wood, A., Martin, J., & Niedenthal, P. (2017). Towards a social functional account of laughter: Acoustic features convey reward, affiliation, and dominance. *PloS one, 12*(8), e0183811.

Wood, R. M., & Gustafson, G. E. (2001). Infant crying and adults' anticipated caregiving responses: Acoustic and contextual influences. *Child Development, 72*(5), 1287-1300.

Young, K. S., Parsons, C. E., LeBeau, R. T., Tabak, B. A., Sewart, A. R., Stein, A., ... & Craske, M. G. (2017). Sensing emotion in voices: Negativity bias and gender differences in a validation study of the Oxford Vocal ('OxVoc') sounds database. *Psychological Assessment, 29*(8), 967.

Zeifman, D. (2001). An ethological analysis of human infant crying: answering Tinbergen's four questions. *Developmental Psychobiology, 39*(4), 265-285.

Zeskind, P. S., & Marshall, T. R. (1988). The relation between variations in pitch and maternal perceptions of infant crying. *Child Development*, 193-196.

Zeskind, P. S., Klein, L., & Marshall, T. R. (1992). Adults' perceptions of experimental modifications of durations of pauses and expiratory sounds in infant crying. *Developmental Psychology, 28*(6), 1153.

CHAPTER II. The emotional canvas of human screams: Patterns and acoustic cues in the perceptual categorization of a basic call type

Jonathan W. M. Engelberg[1], Jay W. Schwartz[1,2], and Harold Gouzoules[1]

[1]Emory University

[2]Western Oregon University

A version of this manuscript was published in *PeerJ*.

Abstract

Screams occur across taxonomically widespread species, typically in antipredator situations, and are strikingly similar acoustically, but in nonhuman primates, they have taken on acoustically varied forms in association with more contextually complex functions related to agonistic recruitment. Humans scream in an even broader range of contexts, but the extent to which acoustic variation allows listeners to perceive different emotional meanings remains unknown. We investigated how listeners responded to 30 contextually diverse human screams on six different emotion prompts as well as how selected acoustic cues predicted these responses. We found that acoustic variation in screams was associated with the perception of different emotions from these calls. Emotion ratings generally fell along two dimensions: one contrasting perceived anger, frustration, and pain with surprise and happiness, roughly associated with call duration and roughness, and one related to perceived fear, associated with call fundamental frequency. Listeners were more likely to rate screams highly in emotion prompts matching the source context, suggesting that some screams conveyed information about emotional context, but it is noteworthy that the analysis of screams from happiness contexts (n = 11 screams) revealed that they more often yielded higher ratings of fear. We discuss the implications of these findings for the role and evolution of nonlinguistic vocalizations in human communication, including consideration of how the expanded diversity in calls such as human screams might represent a derived function of language.

Keywords: screams, emotional expression, nonlinguistic vocal communication

The emotional canvas of human screams: Patterns and acoustic cues in the perceptual categorization of a basic call type

Screams show remarkable evolutionary conservation across a broad range of species, including humans. Acoustically similar vocalizations occur in phylogenetically diverse taxa, usually when an animal is captured by a predator and faces imminent death (Högstedt, 1983). These screams might function to startle the predator, elicit mobbing behavior from conspecifics, warn kin, and/or attract other predators to increase the probability of the caller’s escape (Collias, 1960; Högstedt, 1983; Møller & Nielsen, 2010; Rohwer et al., 1976; Toledo et al., 2011). It is likely that certain, widely conserved acoustic characteristics of screams were selected for in this context. Thus, screams across species share high amplitudes and wide frequency ranges, enhancing their propagation over distances and listeners’ abilities to locate them (Högstedt, 1983).

In several primate species, the contexts in which screams occur have diversified, shifting from limited predator-prey interactions to more complex and socially nuanced agonistic conflicts among conspecifics, where screams function to recruit aid from allies (Bergman et al., 2003; Bernstein & Ehardt, 1985; Cheney, 1977; de Waal & van Hooff, 1981). In the process, selection has also promoted acoustic diversification, such that different classes of screams tend to correlate with contextually relevant elements of the encounter, such as whether an attack comes from a dominant group member or from a lower-ranking animal involved in a rank challenge (macaques: Gouzoules et al., 1984, 1986; Gouzoules et al., 1998; Gouzoules & Gouzoules, 2000; vervets: Mercier et al., 2019; chimpanzees: Slocombe & Zuberbühler, 2005, 2007; Slocombe et al., 2009), likely because clearer communication of these details was evolutionarily advantageous for both screamers and listeners. More recently, we have reported acoustic variation within some classes of rhesus monkey screams that is more directly attributable to differences in arousal level (or activation/alertness; Schwartz et al., 2020b), suggesting that variation in emotion is one proximate mechanism through which acoustic diversification has occurred. Intriguingly, screams have seen even greater contextual diversification in humans, where they are associated with a variety of emotional contexts including fear, anger, surprise, and even happiness (Anikin & Persson, 2017). This feature of human screams, suggesting novel functions in this evolutionarily conserved call type, render them a fascinating subject for understanding nonlinguistic emotional communication in our species.

Despite the diversity of contexts in which screams occur (Anikin & Persson, 2017), an overwhelming majority of studies have treated human screams synonymously with fear (e.g., Arnal et al., 2015; Aubé et al., 2014; Sauter et al., 2010). In an earlier study, we found that transformations of a scream's pitch and duration affected listeners' perceptions of its emotional intensity: specifically, longer and/or higher-pitched screams were consistently identified as more intense (Schwartz & Gouzoules, 2019), indicating that listeners do not perceive all screams as identical with respect to emotional context. More recently, Frühholz et al. (2021) reported that screams reflecting different emotions varied acoustically, resulting in different categorizations of emotion and assessments of alarm by listeners. However, that study used a corpus of acted screams produced entirely by amateurs and did not explore the acoustic predictors of these different perceptions. The acoustic and perceptual patterns by which listeners perceive different emotions from contextually diverse screams remain to be fully explored.

In the present study, participants listened to a series of screams and rated their agreement with prompts corresponding to six different emotions: (1) anger or aggression (hereafter, anger); (2) fear; (3) frustration or annoyance (hereafter, frustration); (4) happiness or excitement (hereafter, happiness); (5) pain; and (6) surprise or startle (hereafter, surprise). These scales were selected on an exploratory basis to represent a variety of contexts in which human screams are naturally observed (Anikin & Persson, 2017). We favored a multidimensional ratings design over a forced-choice categorization because we predicted that different scream variants might express not only very disparate emotions, but possibly also variegated blends of emotion (Cowen et al., 2019). Our design enabled assessment of how listeners' perceptions reflected this potentially nuanced emotional canvas communicated by human screams.

Our primary aim was to determine whether screams yield consistent judgments about emotion across participants. This entailed, first, characterizing participants' levels of agreement; if screams convey information about the emotions typically associated with different experiential contexts, listeners should achieve a non-trivial consensus. Second, we examined the correlational structure between ratings on the emotional prompts, enabling some assessment of how parallel or orthogonal were listeners' judgments on each of the scales. Two non-correlated scales would suggest that listeners made a distinction between those two emotional contexts. On the other hand, two highly correlated scales might suggest they treated those contexts equivalently or judged them based on correlated cues.

An additional aim was to characterize the acoustic variation among screams that accounted for listeners' ratings of emotion. Although many studies have examined the acoustic predictors of emotional judgments from speech (Scherer et al., 2003), the extent to which these findings apply to variation within nonlinguistic vocalizations is less clear. For example, Szameitat et al. (2009b) found that patterns of emotional expression in laughter were generally similar to, but not equivalent to, those reported in speech. Aside from our earlier study examining the roles of scream pitch and duration in judgments of emotional intensity (Schwartz

& Gouzoules, 2019), to our knowledge, there are no published investigations of the acoustic cues predicting varied emotional perceptions from human screams.

We conducted a targeted acoustic analysis based on a set of six parameters that we hypothesized might influence the perception of emotion from screams. These included the mean and range of the fundamental frequency (F0; a measure of vocal fold oscillation rate that is typically perceived as pitch), as well as duration, mean harmonics-to-noise ratio (HNR; a measure of periodicity, where lower values correspond to noisier sounds and higher values correspond to “purer”, tonal percepts), loudness, and roughness. Variation in F0, duration or similar temporal parameters, and HNR are important cues for distinguishing emotions in prosodic speech (reviewed in Juslin & Laukka, 2003; Scherer et al., 2003) and nonlinguistic vocalizations (Raine et al., 2018; Sauter et al., 2010; Szameitat et al., 2009b; Szameitat et al., 2011; Schwartz & Gouzoules, 2019) and are also consistently linked to emotional variation in nonhuman mammals (Briefer, 2012; Briefer et al., 2015; Briefer et al., 2019; Friel et al., 2019; Morton, 1977). The perceived loudness of a vocalization also potentially correlates with the perception of emotional variation (Scherer, 2003), but some of the variation in the loudness of our stimuli originated from factors other than the vocalizations themselves (e.g., different recording conditions). Thus, we included this measure to account for possible effects of loudness unrelated to the acoustic variation of interest. Finally, roughness corresponds acoustically to rapid amplitude modulation and perceptually to a harsh, “buzzing” quality in sounds (Arnal et al., 2015; Vassilakis, 2007). Arnal et al. (*ibid.*) proposed that roughness is a defining characteristic of screams that is absent from regular speech (but perhaps not from all other nonlinguistic vocalizations; Schwartz et al., 2020a; Li et al., 2018), and showed that rough screams are perceived as more fearful than screams artificially filtered to remove rough modulation. However, it is unknown how natural variation in roughness among screams, such as that documented by Schwartz et al. (2020a), affects emotional perception.

A final aim of this study was to investigate whether ratings on emotion prompts varied as a function of the scream's original source context. Specifically, we hypothesized that if screams have the potential to convey information about emotional context, participants should tend to rate them more highly on trials where the emotion prompt matched the source context. Due to a small number of stimuli in some contexts (**Table 1**), we could not formally investigate the effects of all specific source contexts on ratings, with the exception of happiness, a case of special interest given its potential uniqueness to human screams.

Methods

Testing took place over a two-year period at Emory University's Bioacoustics Laboratory in the Department of Psychology. This research was conducted in compliance with Emory's Institutional Review Board under IRB00051516, approved July 26, 2011.

Participants

182 participants from Emory University took part in this study (124 female, 58 male; Age *M* = 19.418, *SD* = 2.041). We aimed for a relatively large sample size because, for the questions explored here, the literature was sparse and thus precedents were few. Variation within screams remains largely undocumented and we therefore wanted to uncover even small effects (while appropriately noting when effect sizes are small). Participants were recruited via an online portal system and received class credit for completing the study. All participants provided their voluntary and informed written consent.

Stimuli

Stimuli were selected from an in-lab corpus of screams collected from movies, television programs, advertisements, YouTube videos, and commercial sound banks (Human Sound Effects, Partners In Rhyme, Inc., Santa Monica, CA; The Nightingale Voice Box, Nightingale Music Productions, North York, Ontario, Canada). Stimuli comprised a mixture of acted and spontaneous screams, categories which a previous study revealed listeners could not reliably distinguish (Engelberg & Gouzoules, 2019).

Screams were selected on the basis of sound quality (i.e., minimal noise; no overlapping sounds), and to represent a variety of emotional contexts. Original source contexts were identified (by HG) based on the surrounding socioecological circumstances and situational cues. These contexts are conventionally associated with a particular emotional state (hence the term *emotional context*), and it is likely that most vocalizers on average experienced the associated emotion, but we do not assert, nor for the purposes of this study is it necessary, that every instance of that context ineluctably entails that emotion (for example, in acted renditions). We chose instead simply to identify the context because of the greater certainty and objectivity in making these distinctions. Indeed, that professional actors' screams are sometimes indistinguishable from naturally occurring ones (Engelberg & Gouzoules, 2019) might suggest that an authentically experienced emotional state is not a prerequisite to produce a credible scream. That said, listeners themselves were asked to make judgments about the likely associated emotion, and not the surrounding production context. Thus, our references to context in this article concern only the source categorizations and not the perceptual judgments made by participants.

The six contexts to which screams were categorized were the basis for the six prompts listeners were presented in the study. These contexts were deemed appropriate for *best* capturing the emotional situations in which each scream occurred. We reasoned and predicted that participants' responses would reflect the emotions expressed by the screams and thus the situational contexts in which they were produced, but we recognize that some of the screams might involve contexts associated with multiple emotions, a hypothetical example being the fear and pleasurable excitement associated with a rollercoaster ride. Although we considered including screams from additional contexts (e.g., embarrassment or sex), we limited this study to six so as not to induce fatigue or task disengagement in participants, who were asked to rate all screams on all six scales for a fully-crossed experimental design.

Stimuli were processed as previously described in Engelberg et al. (2019). Specifically, online videos were captured or downloaded using Total Recorder version 8.0 (High Criteria, Inc., Richmond Hill, Ontario, Canada) and WinXHD Video Converter Deluxe (Digiarty Software, Inc., Chengdu, China), while DVD media were extracted using WinXDVD Ripper Platinum (Digiarty Software, Inc., Chengdu, China). All source videos were saved, converted to the MPEG file format, and cropped at timestamps surrounding the target vocalizations.

Audio files were extracted and converted to 16-bit 22.05 kHz WAV files using Adobe Audition CC (Adobe Systems, San Jose, CA) and Audacity version 2.1.2 (http://audacity.sourceforge.net). Edits were applied when necessary to delete any clicks and pops (Owren & Bachorowski, 2007) or mitigate noise without distorting or interfering with the acoustics of the screams themselves (as determined by listening and by visual inspection of spectrograms). Additionally, in the case of some DVD sources, separate tracks containing background music were removed. Any screams that would have required more extensive editing were not used in the stimulus set.

The final stimulus set consisted of 30 screams from 26 different vocalizers. Again, the total number of stimuli was intentionally limited to reasonably obtain ratings on each scale from every participant and for every scream, enabling a fully crossed design (i.e., one observation per prompt for each combination of participant and stimulus) and a large number of raters per stimulus. Female vocalizers produced 22 of the screams while males produced 8. The imbalance in gender representation was the consequence of our goal to include an ample number of screams from happy emotional contexts, which proved difficult to find for males (an observation we suggest is noteworthy). No differences were found in ratings on any prompt between male and female participants across all screams, nor between their ratings across only female screams or only male screams (Independent sample *t*-tests, $p > .07$ in every case, with the exception of ratings of frustration on male screams, which were not significant when corrected for the number of tests). The distribution of original source contexts is presented in **Table 1** along with means for each measured acoustic parameter (see Acoustic Analysis below).

Table 1. Scream descriptives by original emotional context

Original source context	n	Duration (s)	Mean F0 (kHz)	F0 Range (kHz)	Mean HNR	Mean Roughness	Loudness (LUFs)
Anger	2	2.112	0.353	0.085	2.240	39.135	-22.505
Fear	5	1.319	1.426	0.963	5.908	34.374	-20.408
Frustration	1	0.296	1.940	1.062	7.610	30.068	-13.980
Happiness	11	1.522	2.165	1.244	12.909	26.702	-9.780
Pain	6	1.301	0.677	0.588	7.910	27.776	-18.662
Surprise	5	0.298	1.000	0.609	6.784	27.889	-19.448

Procedure

The experiment was conducted on E-Prime 2.0 software (Psychology Software Tools, Inc., Pittsburgh, PA) running on a Sony VAIO Pentium 4 computer (model PCV-RS311). Participants were instructed to listen to screams and indicate their level of agreement with statements appearing on-screen.

The start of a trial was indicated by the word "Ready" appearing in the center of the screen. After a period of 0.50 s, a stimulus was delivered through headphones (JVC G-Series model HA-G55, JVCKENWOOD USA Corporation, Long Beach, CA) while the screen displayed a prompt related to the emotional state of the vocalizer (e.g., *"Rate your agreement with this sentence. This person is FRIGHTENED.").* Participants used the computer mouse to indicate their agreement with the prompt by clicking on-screen buttons labelled with the numbers 1 through 5, with 1 indicating strong disagreement and 5 indicating strong agreement. (Listeners may have interpreted 3 as neutral, undecided, or uncertain, indicating neither disagreement nor agreement, although it was not labelled as such.) After participants selected a response, the screen displayed their choice (e.g., "You chose 3.") and an interval of 1 s proceeded before the next trial.

Each scream was presented 6 times throughout the experiment, resulting in 180 total trials (6 presentations x 30 screams), with each presentation requiring the participant to judge the scream on a different emotion prompt: (1) aggressive or angry; (2) annoyed or frustrated; (3) excited or happy; (4) frightened; (5) in great pain; and (6) surprised or startled. As noted above, these prompts were chosen to reflect the emotional contexts that were represented among the stimuli. We distinguished anger versus frustration, and therefore tested these in separate prompts, based on whether the eliciting event was other-caused (anger), and thus more associated with a possibility of physical threat, or circumstance-caused (frustration; Roseman et al., 1996). Similarly, we distinguished fear from surprise because fear is necessarily negatively-valenced, but not necessarily short in duration, whereas the reverse is true for surprise (Kreibig, 2010), distinctions that could conceivably result in differences in vocal production and perception. All participants judged every scream on all 6 of these prompts, except for very rare cases in which a participant accidentally proceeded by clicking on a part of the screen that did not correspond to a number (n = 148 out of 32,760 trials across all participants). Stimuli and questions were presented in a fully randomized order, with the exception that no individual scream was ever presented consecutively.

Participant information was collected from a short questionnaire following the experiment. This included information about gender and age, as well as information about native language, handedness, and experience with screams in media that was used in concurrently administered studies (Engelberg et al., 2019).

Analysis

Acoustic analysis. Six acoustic parameters were measured for each scream: duration, mean F0, F0 range, mean HNR, mean roughness, and Loudness Unit Full Scale (LUFS). Analyses were conducted on spectrograms generated from the waveforms by Fast-Fourier Transform (FFT) on Praat (Boersma & Weenink, 2013). Duration was measured with Praat's selection tool by highlighting the spectrogram from vocalization onset to end, excluding any clear reverberation. Mean F0, F0 range, and mean HNR were measured using the Quantify Source command in the GSU Praat Tools script package (Version 1.9, Owren, 2008). This script estimates the sound's F0 contour from a selected spectrogram segment using Praat's To Pitch autocorrelation function, and estimates HNR using Praat's To Harmonicity autocorrelation function. Measurements were made using a 50 ms analysis window, 75-Hz pitch floor, and 3500-Hz ceiling. The automatically generated F0 contour was checked manually for errors (e.g., spurious voiced segments or octave jumps) by visual and auditory comparison to the original spectrogram before obtaining measurements of mean F0, mean HNR, F0 minimum, and F0 maximum; F0 range was calculated as the difference between the latter two. Roughness was measured using the modulationSpectrumFolder function in the *soundgen* package in R (Anikin, 2019). Finally, loudness was measured as LUFS using the Amplitude Statistics function in Adobe Audition (Build 13.0.8.43; ITU-R BS.1770-4 loudness algorithms). LUFS is a measurement of loudness employing algorithms adopted by the International Telecommunication Union (ITU-R BS.1770). The scale controls for the perception of loudness and includes weighting to account for differences in frequency response (humans are much more sensitive to volume changes in mid-range frequencies compared to the high frequencies found in screams). The loudness unit is equivalent to a decibel, except that it is weighted to the human perception of audio rather than just measuring the electrical signal.

Descriptive statistics for these parameters are presented in **Table 1**. We note that representation of each source context in our stimulus set was not equal, nor the representation of potentially salient vocalizer characteristics (e.g., male vs. female; child vs. adult) within each source context. Thus, the acoustic means in Table 1 are meant only to serve as descriptors for the present stimulus set and not to draw conclusions about any true acoustic differences between the contexts.

Statistical analysis. Statistical analyses were performed using SPSS Statistics Version 25 (IBM Corp., Armonk, NY) and the R statistics environment (R Core Team, 2018).

Levels of participant agreement were estimated using the intraclass correlation statistic (ICC, 2-way random model, single-score consistency). The ICC is an interrater reliability index applicable to datasets of more than 2 raters. The ICC model used here takes into account systematic biases between raters in use of the scales (Koo & Li, 2016). It is calculated based on the ratio of the variance of interest (here, the mean square between subjects minus the mean square of errors, divided by the number of raters) over the total variance in a data matrix (Liljequist et al., 2019). A separate ICC was calculated for each emotion prompt, i.e., based on participants' ratings of all 30 items in that prompt. Participants who did not judge all 30 items due to input error were not included in the calculation of that prompt's ICC. Additionally, a separate ICC value was calculated from only the 30 items in which the prompt matched the original context ("match trials"). ICCs were compared by referencing their estimated 95% confidence intervals (Koo & Li, 2016).

Correlations between screams' mean ratings on the emotion prompts were examined using a Pearson's correlation matrix. To characterize this correlational structure further, a principal component analysis (PCA) was conducted on the screams' mean ratings. The extracted components with eigenvalues > 1 are described.

To analyze the effects of scream acoustics on ratings, six separate cumulative link mixed models (CLMMs) with a logit link were fitted using the clmm function in the *ordinal* package in R (Christensen, 2019), with ratings on each prompt as an ordinal outcome variable, the mean-centered acoustic parameters as fixed effects, and participant and stimulus as crossed random effects. Mean HNR and roughness were highly correlated ($r = -.703$) and their inclusion in the same models resulted in high measures of collinearity (Variance Inflation Factors, VIFs > 6), which leads to inflated standard error estimates (Jaeger, 2008). Between the two variables, HNR yielded higher VIFs. Thus, to reduce collinearity, roughness was included and HNR was excluded from the models reported in this paper (see Lima et al., 2013, for a similar approach to correlated predictors). For confirmation, however, we repeated all analyses with HNR instead of roughness, which resulted in nearly identical significant effects as those reported for roughness (except in opposite directions, given their negative correlation). Maximum likelihood estimates and standard errors for each parameter were estimated using the Laplace approximation. The significance of each effect was determined using likelihood-ratio tests via the Anova function in the *car* and *RVAideMemoire* R packages (Fox & Weisberg, 2009; Hervé, 2020). Additionally, to determine the acoustic correlates of the principal components described above, two separate multiple linear regressions were conducted using the acoustic parameters as predictor variables and screams' scores on each principal component as outcome variables. Again, due to high collinearity, mean HNR was excluded from these models.

To analyze whether a source context matching the rating prompt increased the likelihood of higher ratings, a CLMM was fitted with ratings as the outcome variable and match versus non-match coded as a binary predictor variable, with random effects for participant and stimulus. The significance of the match parameter was determined using a likelihood-ratio test. Finally, for the source context best represented among our stimuli, happiness (n = 11), we explored how participants' ratings varied as a function of emotion prompt. To conduct this analysis, a CLMM was fitted using only trials with happiness screams (12,012 trials), with ratings as the outcome variable, emotion prompt as the predictor variable, and participant and stimulus as random effects. Post-hoc pairwise comparisons of least-square-means were then conducted to estimate individual effects of each prompt relative to each other, using the Kenward-Roger method with Tukey p-value adjustment via the *lsmeans* package (Lenth, 2016).

Results

Overall ratings and agreement

Figure 1 depicts the means, bootstrapped confidence intervals, medians, and distributions of screams' ratings on each prompt. Overall ratings varied by prompt (Repeated Measures ANOVA, *F*(3.856, 697.944) = 211.667, $p < .001$). In general, the data suggest that, independent of original context, participants tended to rate stimuli highest in fear, followed by surprise, pain, and frustration. Participants tended to rate screams lowest in anger and happiness (differences significant at $p < .001$). In the Discussion we consider several possible explanations for differences in overall ratings between prompts.

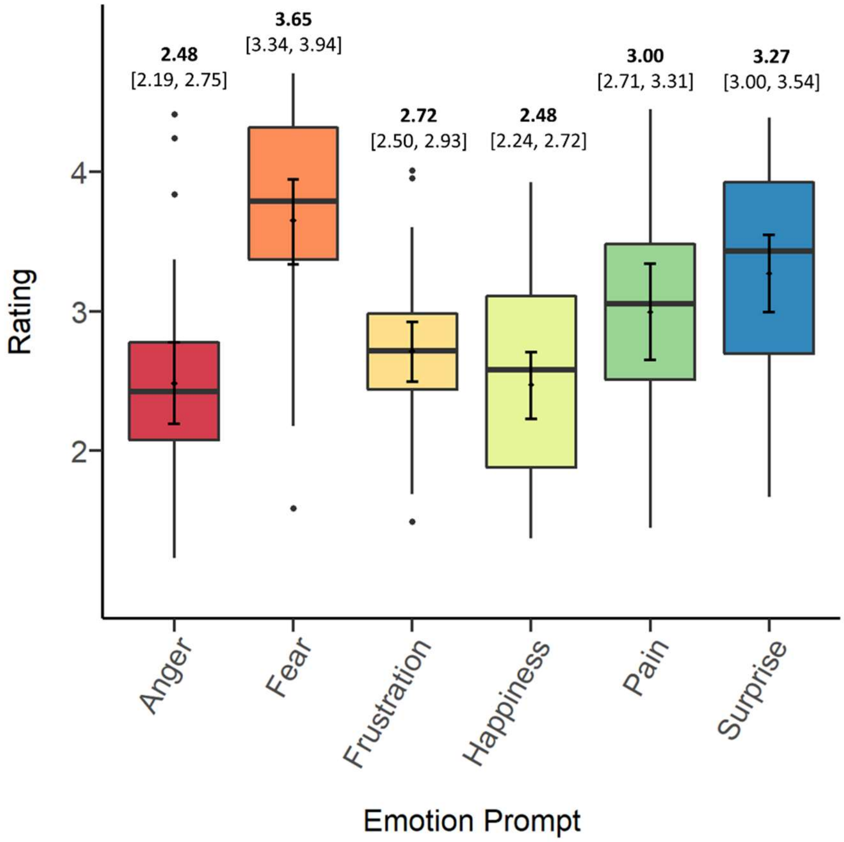


Figure 1. *Means (bold) and 95% confidence intervals (in brackets), medians and distributions of screams' ratings on each of the emotion prompts.*

Table 2 presents participants' intraclass correlations (ICCs) for each of the emotion prompts. All ICCs significantly exceeded chance (*F*-test, $p < .001$), indicating consensus among different raters. Per commonly used guidelines and descriptives for interpreting psychological data (Cicchetti, 1994; Hallgren, 2012), only ratings on fear and pain reached a "fair" level of agreement across all screams. However, overlapping confidence intervals warrant caution in the interpretation that participants agreed significantly more on any of the prompts.

Table 2. Participant agreement statistics

Emotion prompt	ICC	95% C.I.
Anger	0.396	[0.292, 0.544]
Fear	0.453	[0.343, 0.601]
Frustration	0.241	[0.166, 0.367]
Happiness	0.322	[0.230, 0.464]
Pain	0.451	[0.341, 0.598]
Surprise	0.373	[0.272, 0.520]
Match	0.441	[0.332, 0.589]

We suspected that participants might have reached a greater consensus on match trials, those in which the emotion prompt matched the original source context. To investigate this possibility, we calculated the ICC on match trials only and compared this value to the ICCs for each emotion prompt, each of which were mostly calculated from trials not matching the original context (for example, 25/30 screams that participants rated on fear were not from a fear context). The match ICC was higher than the ICCs for anger, frustration, happiness, and surprise, but lower than the ICCs for fear and pain, and again, confidence intervals overlapped in every case. Thus, it appears that participants did not tend to agree more on prompts matching the original source contexts.

Correlations between emotion ratings

Pearson's correlations were conducted on the screams' mean ratings on each of the emotion prompts, indicating significant correlations between the scales (**Table 3**). Ratings of anger, frustration, and pain were highly positively correlated with one another and negatively correlated with ratings of happiness and surprise, which were positively correlated with each other. Ratings of fear did not correlate with ratings on any other emotion prompt.

Table 3. Pearson's correlation matrix between mean ratings for the emotion prompts

Emotion prompt	Anger	Fear	Frustration	Happiness	Pain	Surprise
Anger	1					
Fear	0.117	1				
Frustration	0.962*	0.136	1			
Happiness	-0.696*	-0.061	-0.675*	1		
Pain	0.768*	0.161	0.735*	-0.804*	1	
Surprise	-0.639*	0.255	-0.597*	0.709*	-0.860*	1

** Significant at p < .01*

To further characterize this correlational structure, we conducted a principal component analysis (PCA) on the correlation matrix, yielding two principal components with eigenvalues greater than 1 that cumulatively accounted for 85.6% of the variance (**Table 4**). The first PC, accounting for 66.4% of the variance, loaded positively with ratings of anger, frustration, and pain, and negatively with ratings of happiness and surprise. Happiness and surprise did not load positively onto any PC that explained substantial variance. The second PC, accounting for 19.1% of the variance, loaded positively with fear. Overall, this correlational structure suggests at least two dimensions of variation in participants' ratings of emotions from screams, one separating the perception of (non-fear) negatively-valenced states (anger, frustration, pain) from positively or potentially neutrally-valenced states (happiness, surprise), and the other primarily separating the perception of fear from non-fear. Again, in the Discussion, we consider multiple explanations that might account for these general perceptual correlations.

Table 4. Results of PCA on mean ratings for the emotion prompts

	Principal Component	
Variance Explained	1	2
Eigenvalue	3.986	1.147
% Variance	66.427	19.124
% Cumulative Variance	66.427	85.551
Component Matrix		
Anger	**0.915**	0.127
Fear	0.073	**0.971**
Frustration	**0.894**	0.158
Happiness	**-0.869**	0.023
Pain	**0.936**	0.035
Surprise	**-0.844**	0.403

PC Loadings > .8 are bolded.

Acoustic predictors of emotion ratings

To investigate the effects of selected acoustic parameters on emotion perception, CLMMs were fitted using the parameters as fixed effects, subject and stimulus as crossed random effects, and ratings on each emotion prompt as outcome variables in six separate models. **Table 5** presents the results for the acoustic models. Every parameter except F0 range and loudness significantly affected ratings on at least one emotion prompt and, when taking into account the direction of effects, most emotion percepts were influenced by a unique combination of parameters. Examination of these effects is facilitated with reference to **Fig. 2**, which depicts waveforms and spectrograms of the five screams with the highest mean ratings on (a) Anger and frustration, (b) Fear, (c) Happiness, (d) Pain, and (e) Surprise. Perception of anger and frustration was associated with screams of higher roughness (Figure 2a). Perception of fear was associated with higher mean F0 (Figure 2b; note the high F0 relative to the other spectrograms). Perception of happiness was associated with shorter duration and lower roughness (Figure 2c). Perception of pain, like anger and frustration, was associated with higher roughness as well as longer duration (Figure 2d). Finally, perception of surprise was associated with screams of shorter duration (Figure 2e; note the short timescale).

Table 5. Acoustic predictors of emotion ratings

Emotion prompt	Duration (s)	Mean F0 (kHz)	F0 Range (kHz)	Roughness	LUFS
CLMM estimates					
Anger	0.329	-0.178	-0.012	**0.210*****	0.046
	[.182]	[.372]	[.462]	**[0.038]**	[0.042]
Fear	0.265	**1.037****	0.698	0.037	-0.015
	[0.194]	**[0.398]**	[0.494]	[0.040]	[0.045]
Frustration	0.155	-0.150	0.088	**0.141*****	0.028
	[0.128]	[0.261]	[0.324]	**[0.027]**	[0.029]
Happiness	**-0.638*****	0.292	-0.010	**-0.107*****	0.037
	[0.137]	[0.281]	[-0.348]	**[0.029]**	[0.032]
Pain	**0.892*****	-0.419	0.018	**0.130****	0.059
	[0.203]	[0.416]	[0.516]	**[0.042]**	[0.047]
Surprise	**-0.854*****	0.531	0.142	-0.054	-0.019
	[0.142]	[0.289]	[0.359]	[0.029]	[0.033]
Principal components					
PC1	**0.443*****	-0.254	-0.0001	**0.104*****	0.018
	[0.118]	[0.241]	[0.299]	**[0.024]**	[0.027]
PC2	-0.073	**0.597***	0.445	0.040	0.003
	[0.139]	**[0.284]**	[0.353]	[0.029]	[0.032]

*Note. Brackets indicate standard errors of the estimates. Asterisks indicate significance at * $p < .05$, ** $p < .01$, and *** $p < .001$.*

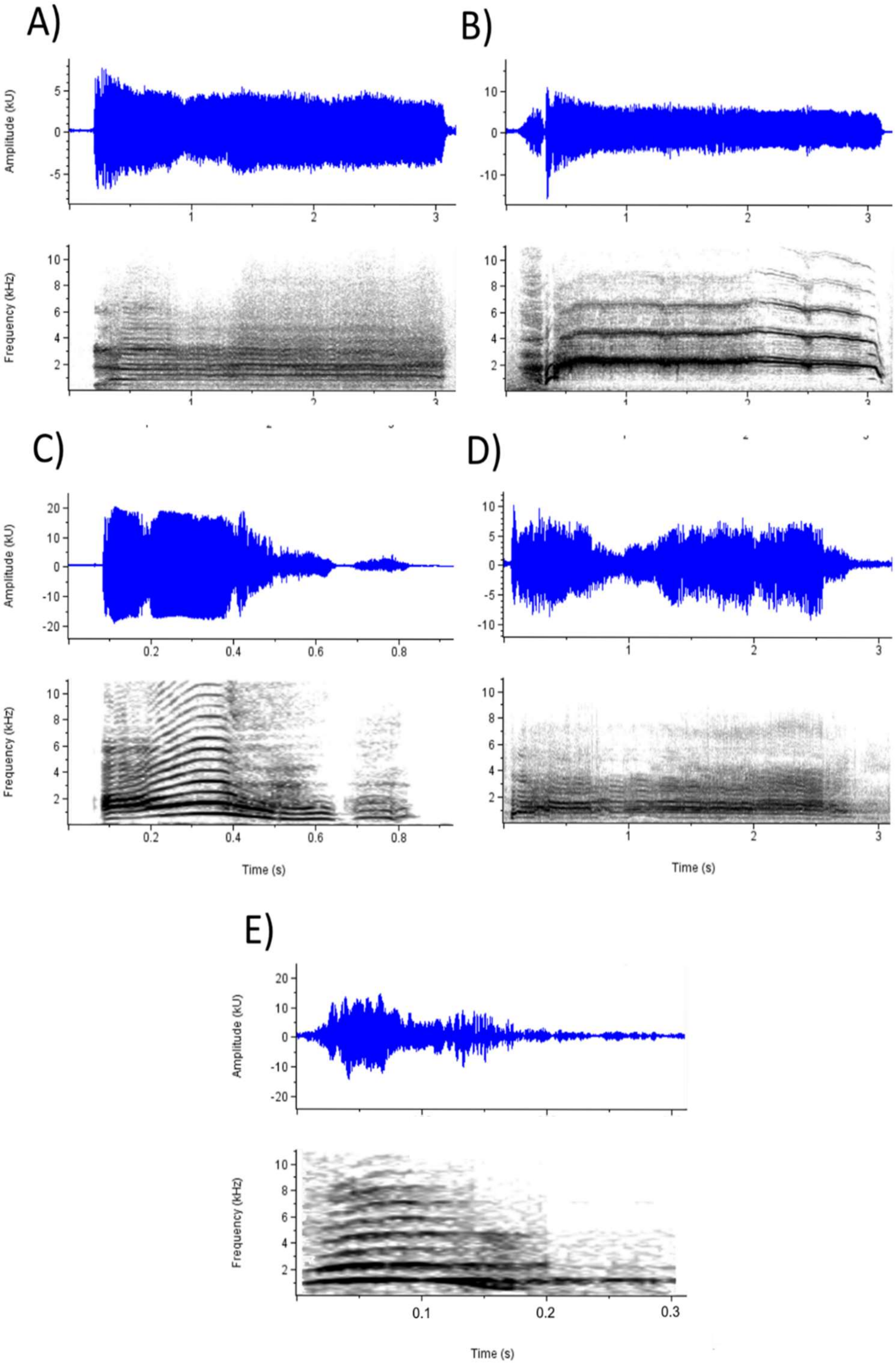


Figure 2. *Waveforms and spectrograms depicting the screams with the highest ratings on each emotion prompt. (A) Anger (Mean rating = 4.41) and Frustration (Mean rating = 4.01); (B) Fear (Mean rating = 4.70 (C) Happiness (Mean rating = 3.92); (D) Pain (Mean rating = 4.45); and (E) Surprise (Mean rating = 4.39).*

The earlier referenced principal components from the perceptual ratings map substantially well to these acoustic results: the emotion perceptions that loaded onto PC1 generally shared acoustic predictors, whereas fear perception, solely loading onto PC2, was also singular in its acoustic predictors. To test this, we conducted two multiple linear regressions using the acoustic parameters as predictor variables and the principal component scores from PC1 and PC2 as outcome variables in two separate models (**Table 5**). These data showed that high PC1 scores, corresponding to higher ratings of anger, frustration, and pain and lower ratings of excitement and surprise, were predicted by longer duration and higher roughness. Higher PC2 scores, corresponding to higher ratings of fear independent of ratings on other emotion prompts, were predicted by higher mean F0s.

We note two important considerations when interpreting these data. First, although these acoustic results characterize the perception of emotions in our study, they might not necessarily correspond to the actual acoustic properties of screams expressing each emotion. For example, although only fear perception was associated with higher F0s, the screams with the highest F0s in our study were those produced in a happy emotional context (*Mean* = 2.17 kHz; see Table 1), which, perhaps, accounts for the finding that screams produced in contexts associated with happiness positively predicted the perception of fear (see section below).

Second, although both the acoustic results and the correlation matrix between ratings on each emotion prompt suggest a broad, perceptual continuum with anger, frustration, and pain on one side, and happiness and surprise on the other, these relationships are not absolute or canonical, and the exceptions are worth examining (**Fig. 3**). Figure 3a depicts a scream with low ratings on anger despite simultaneously low ratings on happiness and surprise (with which anger ratings are negatively correlated) and high ratings on pain (with which anger ratings are positively correlated). This exemplar's particularly long duration might explain its ratings of pain, happiness, and surprise. However, this scream is also low in roughness, and anger ratings were predicted by roughness to a greater extent than other emotions; thus the combination of long duration and low roughness might partially account for the exemplar's idiosyncratic ratings. Similarly, figure 3b depicts a scream with somewhat lower ratings on happiness and higher ratings on surprise, despite the positive correlation between the two prompts. This scream is short (predicting higher ratings on happiness and surprise) but also rough and noisy (significantly predicting lower ratings on happiness only). Thus, although the perceptions of anger and pain, or happiness and surprise were correlated, listeners seemed to differentiate between these concepts and, in some cases, use acoustic information to differentiate their ratings between these scales. In other words, acoustic variation within screams likely allows for more heterogenous perceptions of emotion than the correlation matrix might suggest.

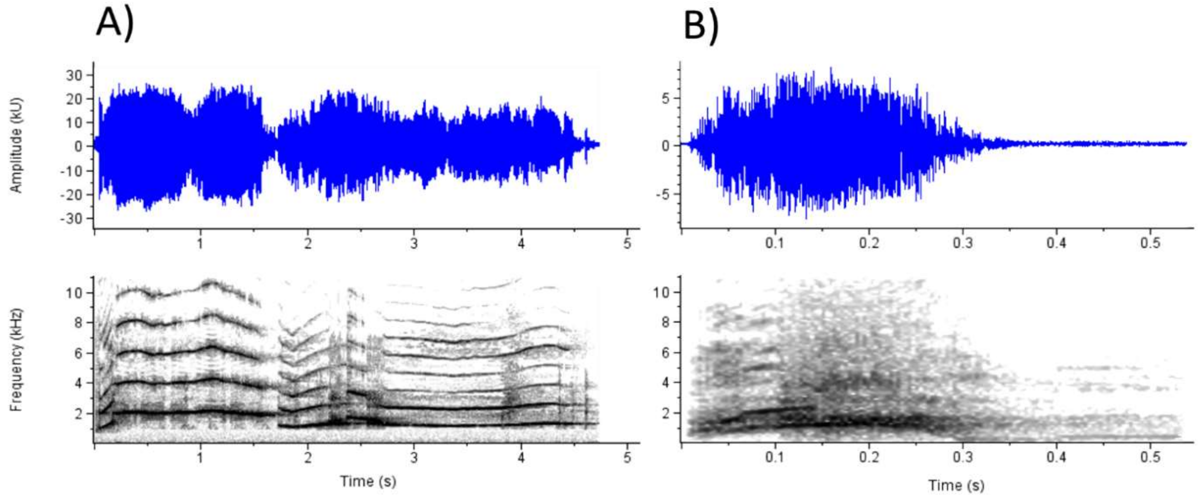


Figure 3. *Two exemplars for which relationships between mean ratings across emotions do not adhere to the general correlations between scales. (A) A scream with low ratings on anger, happiness, and surprise, despite the generally negative correlation between anger with happiness and surprise. This scream also has rather different ratings for anger and pain, despite their strong positive correlation. (B) A scream with low ratings on happiness and high ratings on surprise, despite their strong positive correlation.*

Ratings as a function of source contexts

To investigate whether participants tended to provide higher ratings when the prompt matched the source context, we coded a binary match variable (match = 1 if the prompt matched the context, 0 if not) and used this as a predictor variable in a CLMM with ratings as the outcome variable, and stimulus and participant as random effects. This model revealed a significant positive effect of a matching emotion prompt ($B = 0.869$, $SE = 0.0275$, $p < .001$; i.e., significantly better goodness-of-fit compared to a model without this effect), indicating that participants generally provided higher ratings when a prompt matched the source context.

More close-grained analyses of ratings relative to most original source contexts were precluded because of the limited number of screams from each context (but see Study 2 for these analyses with different stimuli). However, happiness screams were considered a point of special interest, given their potential uniqueness in humans, and they were the best represented source context in our stimulus set ($n = 11$). Therefore, in a qualified assessment of how participants interpreted happiness screams, we fitted a CLMM using the emotion prompt as a categorical fixed effect, subject and stimulus as crossed random effects, and happiness ratings as the outcome variable. This model revealed an overall effect of the prompt on ratings of happiness-associated screams (likelihood ratio test, $p < .001$). Post-hoc pairwise comparisons of least-square means with Tukey *p*-value adjustment were then used to examine the effects of each prompt relative to each other **(Table 6).** Although ratings on the happiness prompt were significantly higher than on the anger prompt, they were lower than ratings on fear, pain, and surprise prompts; fear prompts in particular tended to yield much higher ratings. These data suggest that happiness was not well recognized among our stimuli, and that fear perception included a subset of false positive assessments of screams associated with excitement.

Table 6. CLMM effects of prompts on ratings for screams from happiness-associated contexts

	Anger	Fear	Frustration	Pain	Surprise
Happiness	**0.45***	**-2.352***	**0.042**	**-0.551***	**-0.897***
Anger		-2.802*	-0.408*	-1.001*	-1.347*
Fear			2.394*	1.801*	1.455*
Frustration				-0.593*	-0.939*
Pain					-0.346*

Asterisks indicate significance at p < .001.

Discussion

Screams are phylogenetically widespread, and although they are primarily associated with anti-predation across most taxa, the range of emotional contexts in which they occur has diversified over the course of primate evolution, especially for humans. With few exceptions (e.g., Schwartz & Gouzoules, 2019), potential relationships between listeners' perceptions of emotion and the acoustic variation among screams have remained largely unexamined. In the present study we investigated listeners' responses to screams associated with different emotional contexts using six emotion prompts. We found that listeners derived varied emotional meanings from the acoustic variation among screams. We also identified several acoustic parameters that significantly predicted perceptions on each emotion prompt. These results suggest that the contextual breadth of human scream usage is, to some extent, mirrored by an acoustic diversification that enables meaningful differentiation among scream variants.

Agreement and perceptual dimensions

On every prompt, participants showed higher agreement than expected by chance. Depending on the prompt, levels of agreement ranged from "poor" to "fair" per the guidelines and terminology for interpreting reliability statistics (Cicchetti, 1994). We note, however, that these guidelines were developed for validating psychological assessment instruments (i.e., tests designed for reliability) and thus are likely to be especially stringent. Agreement on our prompts was comparable to those in other studies involving ratings of vocal emotional stimuli (e.g., Bänziger et al., 2014). Given the constraints that participants here rated exemplars all belonging to the same basic call type, screams (Schwartz et al., 2020a), we suggest that their level of agreement is meaningful. It points to the significance of within-call type variation that, for screams, listeners reached non-trivial levels of agreement as to more specific emotional meanings.

On average, participants tended to rate stimuli highest in fear and lowest in anger and happiness. These differences in ratings might reflect that participants more commonly attributed fear to screams, and/or that they were more confident in their assessment of fear and thus tended to register higher ratings. Anger was one of the least represented emotions among screams' original source contexts, so it is plausible that participants were accurate in less frequently perceiving anger among our stimuli. However, happiness was the most commonly represented original source context in our stimulus set, yet participants tended to perceive it less than other emotions. The section below, discussing ratings on each prompt for happiness-associated screams, explores these findings in greater detail.

Correlations between mean ratings on each prompt revealed that listeners perceived emotions from screams at least along two perceptual dimensions. First, their perceptions of anger, frustration, and pain generally grouped together while diverging from their perceptions of happiness or surprise, implying, at a minimum, basic differentiation between the perception of some clearly negatively-valenced states from non-negative states. Recall, however, that these dimensions describe variance in screams' mean ratings, not their acoustic features, and that ratings of some exemplars suggest more fine-grained perceptual distinctions (Figure 3). With the possible exception of anger and frustration, which likely have conceptual overlap, participants apparently differentiated between each emotion but judged some according to similar (but not perfectly correlated) cues, discussed below. The second dimension roughly mapped exclusively onto the perception of fear. Thus, some variance in the perception of fear appeared orthogonal to the perception of the other emotions, an interesting observation given the likely evolutionary significance of fear with respect to the origin of screams (although we note that even in fear contexts, the communication of fear itself might not have been the call's primary adaptive function, which likely centered on startling, distracting, or driving away a predator).

An interpretation of these dimensions must take into account that our selection of original source contexts and prompts was preliminary and exploratory. Our goal in this study was to sample a variety of commonly recognized scream-associated contexts. We do not expect that our selected contexts comprise discrete or pure representations of each emotional state. Certain states such as anger and pain may often co-occur, further contributing to listeners' correlated perceptions of these emotions. It is very likely that in natural contexts, vocalizers convey, and listeners perceive, blends of emotion (Cowen et al., 2019), a feature that would fit our notion of the emotional *canvas* of human screams.

Acoustic predictors of emotion perception

We investigated the role of six acoustic parameters (mean F0, F0 range, duration, mean HNR, roughness, and loudness) in participants' judgments on each emotion prompt. Specifically, we analyzed how these parameters affected listeners' ratings without assuming these parameters are those that categorically delineate scream variants produced in different emotional contexts. An analysis of the latter would require a larger sample of exemplars from each context. Overall, mean F0, duration, and roughness all significantly influenced participants' ratings on at least one prompt; as explained earlier, mean HNR was excluded from acoustic models to avoid collinearity. Only F0 variation and loudness had no effect.

Many of the acoustic effects on perception in our study were consistent with findings in prosodic speech and other nonlinguistic vocalizations. For example, higher scream F0 predicted higher fear ratings and, in other vocalizations, fear is often among the emotions associated with the highest F0 (Belin et al., 2008; Juslin & Laukka, 2003). It is notable, however, that this pattern holds in the perception of variation within screams, whereas in prior studies the correlation might have emerged because the fear stimuli, and only the fear stimuli, consisted of screams, which are characterized by higher pitch than other call types (Schwartz et al., 2020ba). On the other hand, the lack of a role of F0 in the perception of any other emotion, or of any significant effect of F0 variation, contrasts with more general predictions (Scherer, 1986) and findings (Juslin & Laukka, 2003) that these parameters are broadly important for vocal emotion perception. F0 is most commonly implicated as a correlate of arousal (likely reflecting increased tension in the laryngeal musculature; Briefer, 2012), and pitch does contribute to listeners' judgments of scream intensity (Schwartz & Gouzoules, 2019), but a single factor arousal framework does not seem to account for the present results: fear-associated contexts are not obviously linked to higher arousal than the other contexts in our study. It is perhaps relevant that fear perception in screams is not only semi-independent from the perception of other emotions, but also linked to such a salient vocal characteristic as F0. In keeping with the idea of emotion blends, listeners seemed to use F0 as a cue by which they sometimes perceived or did not perceive fear in addition to the concurrent perception of other emotions, a point highlighted by the fact that F0 was the strongest predictor not only of fear ratings but also of screams' scores on the principal component mapping onto the perception of fear independently of other perceived emotions.

The effects of duration aligned reasonably well with the broader literatures on human and nonhuman emotional expression. Pain perception was associated with longer duration, whereas happiness and surprise perception were characterized by shorter duration. These results add further support to the positive relationship between duration and negatively-valenced states, which is one of the more consistently reported correlations across species and call types (Briefer et al., 2015; Briefer et al., 2019; Friel et al., 2019; Raine et al., 2018; Scheiner et al., 2002; albeit with exceptions, e.g., predator alarm calls are negative but characterized by short duration; Caro, 2005). Note that the results here, however, only concern perception: an actual correlation between duration and valence in screams remains to be tested.

Roughness has garnered particular interest of late (Belin & Zatorre, 2015; Frühholz et al, 2021) because of Arnal et al.'s (2015) contention that, among human vocalizations, this feature is unique to screams, affording screams a special acoustic niche for communicating alarm. Partially consistent with this hypothesis, we found that roughness positively predicted the perception of negative states, including anger, frustration, and pain, and negatively predicted the perception of happiness, although it did not significantly affect the perception of fear. These results suggest that, beyond its putative role in distinguishing screams acoustically from other call types, variation in roughness correlates with listeners' perceptions of emotion among screams. Given that roughness variation in infant cries correlates with the vocalizer's level of pain (Koutseff et al., 2018), it seems likely that roughness is an important cue not only for distinguishing call types associated with alarm, but also for interpreting relevant emotional variation within those call types. That said, in our study, roughness and mean HNR were highly correlated and accounted for similar variance in ratings, so it is difficult to discern the effects of one independent of the other. It is possible that the perception of noisiness, and not roughness *per se,* contributed to listeners' perceptions of negative valence. Noisiness is sometimes, but not always, associated with increasing arousal (Briefer, 2012) or pain (Koutseff et al., 2018), but negatively correlated with pain or anger perception from other human nonlinguistic vocalizations (Lima et al., 2013; Raine et al., 2018). Nonlinear phenomena, such as chaotic noise produced by aperiodic vocal fold vibration, also result in less tonal sounds (Fitch et al., 2002) and can lead to the perception of negative valence (Anikin, 2020), so the presence of these phenomena in our stimuli may have also played a role in our results.

Interestingly, our results show some parallels with the motivation-structural (MS) rules that Morton (1977) developed to account for variation in animal calls: namely, that hostile calls should converge upon a "harsh" (wideband, noisier) acoustic structure whereas calls functioning to appease conspecifics (e.g., those associated with friendly or fearful contexts) should sound more tone-like (as well as higher-pitched, consistent with the correlation between F0 and fear perception). Another study from our lab, examining acoustic variation in screams across macaque species, likewise provided nuanced support for MS rules (Gouzoules & Gouzoules, 2000).

Overall, our acoustic results are largely consistent with patterns seen across other mammalian vocalizations (Briefer, 2012; Briefer et al., 2019), human nonlinguistic vocalizations (Belin et al., 2008; Sauter et al., 2010), and speech prosody (Juslin & Laukka, 2003). These findings might suggest either that homologous mechanisms account for emotion-based variation in every case, leading to similar effects on perception, and/or that participants judged variation in screams based on their experience with other types of vocal expression, whether or not those judgments were accurate for screams.

Accuracy and the case of happiness

Overall, participants were more likely to provide higher ratings on emotion prompts matching a scream's original source context, circumspectly suggesting some degree of accuracy in listeners' interpretations of screams. That said, this measure of accuracy did not necessarily entail precision, i.e., ruling out all "incorrect" emotions (see the confusion of happiness-associated screams for fear, discussed in detail below), a determination made complicated in any event by the existence of emotion blends.

For the most part, our stimulus set was not sufficient to enable more detailed examination of accuracy by each source context. However, we provisionally investigated the interpretation of happiness screams as a special case, given that these were relatively well-represented among our stimuli and are potentially unique to humans. Strikingly, participants overall tended not to perceive happiness from our stimuli. Listeners instead provided high ratings of fear to happiness-associated screams, a finding that replicates earlier reports that both listeners (Anikin & Persson, 2017) and acoustic classifiers (Patel et al., 2011) often confuse intense joy for fear.

In the General Discussion, I offer a broader functional (and speculative) account for this confusion, but here we suggest a possible proximate factor: namely, general response biases towards fear and against happiness might suggest that listeners sometimes fell back on familiar stereotypes regarding the contexts in which screams occur. If this is the case, studies investigating scream perception across cultures or participant groups with relevant backgrounds (e.g., parents versus non-parents), whose expectations might vary or have changed through personal experience, could reveal somewhat more robust percepts of happiness from screams. It is additionally likely that other contextual information (e.g., the social context; Wood, 2019) would modulate listeners' interpretations of screams, and perhaps enable a confident assessment of a scream as joyful even if, absent contextual details, listeners might interpret it as fearful.

Listeners' poor identification of happiness in this study contrasts with the findings of Frühholz et al. (2021), who reported that screams associated with joy were processed more efficiently and with greater perceptual sensitivity relative to other scream variants (although they also reported common misclassification of non-alarm screams as alarm screams). This difference might relate to the source of scream stimuli, as that group used entirely acted screams produced by amateurs, whereas we used a combination of natural screams and screams produced by professional actors. Our prior study reporting that actors are capable of producing credible screams notably only included acted screams produced by trained professionals, and even still we observed variation in the convincingness of their renditions (Engelberg & Gouzoules, 2019). Additionally, Frühholz et al. (2021) only observed an advantage in processing positive screams during speeded rather than self-paced tasks. Given our discrepant findings, further investigation will be needed to explore the tasks and conditions under which an advantage for processing happiness screams might exist.

In conclusion, our findings indicate that listeners can perceive different emotional content from human screams. Many of these perceptions were evidently based on similar acoustic cues as those involved in other types of vocal emotional communication, but the effects of some parameters hint at potentially interesting aspects of variation in screams that require further investigation. Listeners tended to provide higher ratings on emotion prompts matching screams' original source context, suggesting that acoustic variation among screams conveys some relevant information about variation in context. However, an examination of happiness screams revealed that they were not accurately perceived, and more commonly yielded higher ratings of fear. In all, our findings suggest that acoustic variation among screams is likely relevant to their communicative roles.

Acknowledgements

We thank Caitlin Clark, Alexander Gouzoules, Leah Friedman, Elizabeth Harlan, and NooRee Lee for assistance with stimulus collection, as well as Anna Duncan for assistance with data collection. We also thank Dr. Pantelis Vassilakis for his generous assistance regarding the measurement and interpretation of roughness, as well as Dr. Andrey Anikin and an anonymous reviewer for their helpful comments on the manuscript.

References

Anikin, A. (2019). Soundgen: an open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods, 51*(2), 778-792.

Anikin, A. (2020). The perceptual effects of manipulating nonlinear phenomena in synthetic nonverbal vocalizations. *Bioacoustics, 29*(2), 226-247.

Anikin, A., & Persson, T. (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods, 49*(2), 758-771.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology, 25*(15), 2051-2056.

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2014). Fear across the senses: brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience, 10*(3), 399-407.

Bänziger, T., Patel, S., & Scherer, K. R. (2014). The role of perceived voice and speech characteristics in vocal emotion communication. *Journal of nonverbal behavior, 38*(1), 31-52.

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*(2), 531-539.

Belin, P., & Zatorre, R. J. (2015). Neurobiology: Sounding the alarm. *Current Biology, 25*(18), R805-R806.

Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science, 302*(5648), 1234-1236.

Bernstein, I. S., & Ehardt, C. L. (1985). Agonistic aiding: kinship, rank, age, and sex influences. *American Journal of Primatology, 8*(1), 37-52.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.51). Retrieved January 9, 2018 from http://www.praat .org/.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology, 288*(1), 1-20.

Briefer, E. F., Maigrot, A. L., Mandel, R., Freymond, S. B., Bachmann, I., & Hillmann, E. (2015). Segregation of information about emotional arousal and valence in horse whinnies. *Scientific Reports, 4*, 1–11. https://doi.org/10.1038/srep09989

Briefer, E. F., Vizier, E., Gygax, L., & Hillmann, E. (2019). Expression of emotional valence in pig closed-mouth grunts: Involvement of both source- and filter-related parameters. *The Journal of the Acoustical Society of America, 145*(5), 2895–2908. https://doi.org/10.1121/1.5100612

Caro, T. M. (2005). *Antipredator defenses in birds and mammals*. Chicago: University of Chicago Press.

Cheney, D. L. (1977). The acquisition of rank and the development of reciprocal alliances among free-ranging immature baboons. *Behavioral Ecology and Sociobiology, 2*(3), 303-318.

Christensen, R. H. (2019). Ordinal—Regression Models for Ordinal Data. R package version 2019. 12-10. https://CRAN.R-project.org/package=ordinal.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284.

Collias, N. E. (1960). An ecological and functional classification of animal sounds. Animal sounds and communication, *7*, 368-391.

Cowen, A. S., Laukka, P., Elfenbein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature human behaviour, 3*(4), 369-382.

de Waal, F. B., & van Hooff, J. A. (1981). Side-directed communication and agonistic interactions in chimpanzees. *Behaviour, 77*(3), 164-198.

Engelberg, J. W., & Gouzoules, H. (2019). The credibility of acted screams: Implications for emotional communication research. *Quarterly Journal of Experimental Psychology, 72*(8), 1889-1902.

Engelberg, J. W., Schwartz, J. W., & Gouzoules, H. (2019). Do human screams permit individual recognition?. *PeerJ, 7*, e7087.

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour, 63*(3), 407–418. https://doi.org/10.1006/anbe.2001.1912

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression, Third edition*. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Friel, M., Kunc, H. P., Griffin, K., Asher, L., & Collins, L. M. (2019). Positive and negative contexts predict duration of pig vocalisations. *Scientific Reports, 9*(1), 2062. https://doi.org/10.1038/s41598-019-38514-w

Frühholz, S., Dietziker, J., Staib, M., & Trost, W. (2021). Neurocognitive processing efficiency for discriminating human non-alarm rather than alarm scream calls. *PLoS biology, 19*(4), e3000751.

Gouzoules, H., & Gouzoules, S. (2000). Agonistic screams differ among four species of macaques: the significance of motivation-structural rules. *Animal Behaviour, 59*(3), 501-512.

Gouzoules, H., Gouzoules, S., & Marler, P. (1986). Vocal communication: a vehicle for the study of social relationships. In M. J. Kessler & R. G. Rawlins (Eds.), *The Cayo Santiago Macaques: History, Behavior and Biology* (pp. 111-129). SUNY Press.

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (Macaca mulatta) screams: representational signalling in the recruitment of agonistic aid. *Animal Behaviour, 32*(1), 182-193.

Gouzoules, H., Gouzoules, S., & Tomaszycki, M. (1998). Agonistic screams and the classification of dominance relationships: are monkeys fuzzy logicians? *Animal Behaviour, 55*(1), 51-60.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23.

Hervé (2020). *Testing and plotting procedures for biostatistics.* R package version 0.9-77. https://CRAN.R-project.org/package=RVAideMemoire

Högstedt, G. (1983). Adaptation unto death: function of fear screams. *The American Naturalist, 121*(4), 562-570.

ITU-R BS.1770-4 (2015). Algorithms to measure audio programme loudness and true-peak audio level. Tech. Rep., International Telecommunications Union.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language, 59*(4), 434-446.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine, 15*(2), 155-163.

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.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394-421.

Lenth, R. V. (2016). Least-Squares Means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Li, T., Horta, M., Mascaro, J. S., Bijanki, K., Arnal, L. H., Adams, M., ... & Rilling, J. K. (2018). Explaining individual variation in paternal brain responses to infant cries. *Physiology & Behavior, 193*, 43-54.

Liljequist, D., Elfving, B., & Roaldsen, K. S. (2019). Intraclass correlation–A discussion and demonstration of basic features. *PloS one, 14*(7).

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: a corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods, 45*(4), 1234-1245.

Mercier, S., Déaux, E. C., van de Waal, E., Bono, A. E., & Zuberbühler, K. (2019). Correlates of social role and conflict severity in wild vervet monkey agonistic screams. *PloS one, 14*(5).

Møller, A. P., & Nielsen, J. T. (2010). Fear screams and adaptation to avoid imminent death: effects of genetic variation and predation. *Ethology Ecology & Evolution, 22*(2), 183-202.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist, 111*(981), 855-869.

Owren, M. J. (2008). GSU Praat tools: Scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods, 40*(3), 822–829. https ://doi.org/10.3758/BRM.40.3.822.

Owren, M. J., & Bachorowski, J. A. (2007). Measuring emotion-related vocal acoustics. *Handbook of emotion elicitation and assessment,* 239-266.

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology, 87*(1), 93-98.

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. Retrieved October 15, 2018 from http://www.r-proje ct.org/.

Raine, J., Pisanski, K., Oleszkiewicz, A., Simner, J., & Reby, D. (2018). Human listeners can accurately judge strength and height relative to self from aggressive roars and speech. *iScience, 4*, 273-280.

Rohwer, S., Fretwell, S., & Tuckfield, R. (1976). Distress screams as a measure of kinship in birds. *American Midland Naturalist*, 418-430.

Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion, 10*(3), 241-278.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology, 63*(11), 2251-2272.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of Voice, 16*(4), 509-529.

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin, 99*(2), 143.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. *Handbook of affective sciences*, 433-456.

Schwartz, J. W. (2020). Vocal emotion expression across contexts, vocalization types, and species: Implications for general processes of vocal evolution. (Doctoral dissertation). Retrieved from https://etd.library.emory.edu/concern/etds/jw827c636

Schwartz, J. W., & Gouzoules, H. (2019). Decoding human screams: perception of emotional arousal from pitch and duration. *Behaviour, 156*(13-14), 1283-1307.

Schwartz, J. W., Engelberg, J. W. M., & Gouzoules, H. (2019a). Was that a scream? Listener agreement and major distinguishing acoustic features. *Journal of Nonverbal Behavior.*http://doi.org/10.1007/s10919-019-00325

Schwartz, J. W., Engelberg, J. W. M., & Gouzoules, H. (2020b). Evolving views on cognition in animal vocal communication: Contributions from scream research. *Animal Behavior and Cognition, 7*(2), 192-213.

Slocombe, K., & Zuberbühler, K. (2005). Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role. *Journal of Comparative Psychology, 119*(1), 67-77.

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences, 104*(43), 17228-17233.

Slocombe, K. E., Townsend, S. W., & Zuberbühler, K. (2009). Wild chimpanzees (*Pan troglodytes schweinfurthii*) distinguish between different scream types: evidence from a playback study. *Animal Cognition, 12*(3), 441-449.

Szameitat, D. P., Alter, K., Szameitat, A. J., Darwin, C. J., Wildgruber, D., Dietrich, S., & Sterr, A. (2009). Differentiation of emotions in laughter at the behavioral level. *Emotion, 9*(3), 397.

Szameitat, D. P., Darwin, C. J., Wildgruber, D., Alter, K., & Szameitat, A. J. (2011). Acoustic correlates of emotional dimensions in laughter: arousal, dominance, and valence. *Cognition and Emotion, 25*(4), 599-611.

Toledo, L. F., Sazima, I., & Haddad, C. F. (2011). Behavioural defences of anurans: an overview. *Ethology Ecology & Evolution, 23*(1), 1-25.

Vassilakis, P. N. (2007). SRA: A web-based research tool for spectral and roughness analysis of sound signals. In *Proceedings of the 4th Sound and Music Computing (SMC) Conference* (pp. 319-325).

Wood, A. (2020). Social context influences the acoustic properties of laughter. *Affective Science, 1*(4), 247-256.

CHAPTER III. Emotion-related variation in simple speech extends to screams: Implications for human nonverbal communication

Abstract

Research on vocal emotional expression has traditionally focused on speech prosody. It is unclear whether patterns of emotion-related variation in speech extend to basic call types such as screams, implicating similar sources of variation. We recorded 9 actors' screams and single-utterance speech samples associated with five different emotions. For each vocalization type, we analyzed the emotion-related variation in 14 acoustic parameters as well as listeners' abilities to recognize emotions in a categorization task. Patterns of emotion-related acoustic variation were similar across screams and speech, suggesting that similar mechanisms are responsible for variation in each case. The effects of acoustic variation on listeners' responses were also similar between vocalization types, although some response tendencies were potentially specific to screams, including relatively higher false alarm rates for pain and fear. Listeners completed an additional task wherein they rated screams on scales related to hypothesized functions, such as capturing attention or communicating distress. Ratings of attention did not differ overall between emotions, suggesting a possible general function of screams, but certain scales, including distress, showed emotion-related variation in ratings. These findings suggest that the acoustic variation in human call types parallels the variation in speech, and that this variation results in perceptual effects that are somewhat specific to certain emotional contexts.

Emotion-related variation in simple speech extends to screams: Implications for human nonverbal communication

The expression of emotions is an important component of vocal communication in humans and other mammals (reviewed in Briefer, 2012). Most of our knowledge in this area in humans comes from research on emotional speech prosody, referring to the spectro-temporal and inflectional cues that express emotions in speech. Less is known about cues to emotion in nonlinguistic vocalizations such as screams and laughs, but broad commonalities with emotional speech prosody have been posited (Ackermann et al., 2014). A leading hypothesis is that patterns of emotion-related variation in the voice largely reflect the automatic effects of emotion on vocal production mechanisms (Juslin & Scherer, 2005; Scherer, 1986). For example, emotional arousal, or alertness, is thought to involve a sympathetically-driven increase in subglottal pressure leading to an increase in fundamental frequency (F0, corresponding to pitch) and amplitude (Scherer, 1986). According to this hypothesis, the global effects of emotion on the vocal musculature should manifest similarly regardless of the presence or absence of linguistic content. Indeed, this hypothesis implies an evolutionary origin of emotion-related acoustic variation common not only to speech and nonlinguistic vocalizations, but also to the calls of other species (Filippi, 2016; Gruber & Grandjean, 2017; Lehoczki et al 2020; Schwartz, 2020; Zimmerman et al., 2013).

There are many reasons, however, to question empirically the extent to which both the patterning and perception of emotions in the voice might differ between nonlinguistic vocalizations and speech. Our evolutionary ancestors communicated vocally via a system of calls, the precursors to our modern-day nonlinguistic vocalizations, long before speech and language developed, and it is reasonable to assume that speech evolved to solve social and ecological problems that this prelinguistic system did not (Owren et al., 2011). In the process, we evolved numerous neural and cognitive adaptations that made language possible (Simonyan & Horwitz, 2011), resulting, among other things, in semi-separate pathways for linguistic versus nonlinguistic vocal production (Ackermann et al., 2014).

These functional and mechanistic innovations are likely to have altered the relationships between vocal production, emotional expression, and listeners' perceptions of emotion in turn. For example, in the nonlinguistic pathway, subcortical limbic areas may sometimes act as the initial triggers for vocal production, perhaps suggesting a closer connection to emotional activity (Owren et al., 2011), whereas speech often entails relatively more cortical involvement (Simonyan & Horwitz, 2011). Relatedly, speech relies on a level of voluntary control that may afford greater inhibition or regulation of emotional cues in speech prosody than in spontaneous, nonlinguistic vocalizations. It is also worth noting that, in addition to the physiological effects contributing to vocal emotional cues—what Scherer (1989) deems "push" effects—vocal cues in humans also reflect modulations pursuant to learned, socially determined and culturally variable display rules ("pull" effects; Scherer, 1989), which some have suggested may play a more dominant role in speech prosody than in nonlinguistic vocalizations (Patel et al., 2011). All of these factors could conceivably result in a distinct (if overlapping) set of acoustic cues by which emotions manifest in linguistic versus nonlinguistic expression. Listeners' perceptions of emotion from these channels may differ in turn, for example, if some brain regions involved in affective processing are more responsive to certain emotion-related acoustic parameters than others (Leitman et al., 2010).

It therefore cannot be taken as a given that findings regarding the expression and perception of emotion-related variation in speech extend faithfully to nonlinguistic vocalizations.

The reverse is true as well, and although nonlinguistic vocalizations are often used in studies of emotion processing (Belin et al., 2008; Hunter et al., 2010; Warren et al., 2006), researchers should perhaps exercise caution when extrapolating from these results to emotional speech perception. Of course, the theoretically common basis for emotional variation in all vocalizations, not to mention the potentially complex interactions between linguistic and nonlinguistic channels of communication (referred to in the General Introduction), leave open the possibility that much of the variation in nonlinguistic vocalizations is indeed consistent with emotional speech prosody. Independent empirical evidence is needed to compare the two and characterize the extents of their similarities and differences.

One commonly replicated finding so far is that studies of nonlinguistic vocalizations tend to yield higher overall emotion recognition accuracy, many nearing or exceeding 80% (Belin et al., 2008; Dietrich et al., 2006; Hawk et al., 2009; Hunter et al., 2010; Lima et al., 2013; Pell et al. 2015; Sauter et al., 2013; Schröder, 2003; Simon-Thomas et al., 2009), compared to studies of speech prosody which often report rates in the 50-70% range (Banse & Scherer, 1996). Additionally, children’s recognition of vocally expressed emotions shows improvements with age for prosody but not for nonlinguistic vocalizations (Sauter et al., 2013), perhaps corroborating notions that emotions in the latter are easier to recognize. Some data also suggest that certain emotions are more (or less) effectively conveyed nonlinguistically than linguistically. For example, a researcher narrowly focused on emotional speech prosody might conclude, erroneously, that the voice does not communicate disgust, given traditionally poor recognition rates of disgust through this channel (e.g., Banse & Scherer, 1996). However, across several studies, disgust has proven highly recognizable from nonlinguistic vocalizations (Laukka et al., 2013; Lima et al., 2013; Schröder, 2003). A similar story may apply to a handful of variegated positive emotions that are fairly well recognized from nonlinguistic vocalizations but less so in speech (Sauter & Scott, 2007; Simon-Thomas et al., 2009).

The mechanisms responsible for these differences in recognition rate are not fully delineated. To some extent, they may reflect qualitative differences in acoustic patterning. For example, Sauter et al. (2010) found that harmonic spectral cues other than pitch were most correlated with nonlinguistically-expressed fear—represented, in their study, by screams—in contrast with findings in emotional prosody that implicate pitch as a key predictor of fear (Banse & Scherer, 1996; Juslin & Laukka, 2003). ERP findings also have revealed that emotions in nonlinguistic vocalizations produce differential neural signals more rapidly than those same emotions do in prosodic speech (Pell et al., 2015), suggesting differences in processing.

This limited evidence aside, however, interpretations regarding differences in vocal production or perception are made difficult because speech prosody, by definition, entails language-related constraints that nonlinguistic vocalizations do not. Emotional speech prosody is restricted by the phonemic requirements of its carrier phrase (Sauter et al., 2010; Schröder, 2003), not to mention the other functions that prosody fulfills in speech (e.g., prosodic inflections that disambiguate semantic content; Nygaard et al., 2009). In contrast, in most studies on nonlinguistic vocalizations, vocalizers are given the latitude to express emotions without any constraints on acoustic structure, a freedom likely to enhance their expressive capacities. Some evidence indicates that, for certain nonlinguistic vocalizations, participants can identify the emotion from a neutrally-spoken (Dietrich et al., 2006) or even transcribed (Schröder, 2003) version of the call: that is, the phonemic patterning alone is distinctive enough that participants do not require any other acoustic cues. On one hand, these findings underscore the ambiguous boundaries between linguistic and nonlinguistic, as it is unclear whether some of these phonemically identifiable vocalizations, sometimes called *affect emblems* (Scherer, 1994), are meaningfully different than conventionalized words. On the other hand, they highlight the added expressive possibilities for a vocalizer who need not adhere to the segmental structure of carrier phrases (Schröder, 2003). It is therefore difficult to determine to what extent a nonlinguistic advantage might reflect, for example, tighter production links to emotion, versus the lesser influence of phonemic and semantic prosodic constraints.

What most studies that include nonlinguistic vocalizations do not account for are the various call types represented among their stimuli. To allow vocalizers to produce laughs, screams, or cries as needed is to let them draw from many different basic human call types (Anikin et al., 2018). Consider a hypothetical experiment involving a simple discrimination between joy and fear in which, in a nonlinguistic condition, listeners must match laughs to the former and screams to the latter, but in a linguistic condition, they must discriminate these emotions provided speech samples that use the same carrier phrase. Such an approach might prove informative with respect to our communicative capacities provided no acoustic constraints, but it suffers from a number of limitations. First, it makes for an odd comparison of nonlinguistic vocalizations to speech prosody, given the latter's surface-level constraints, and given that the features distinguishing a laugh versus a scream presumably reflect factors additional to their non-overlapping associated emotions. Screams, for example, likely possess certain characteristics that promote long-distance propagation, a functional consideration that might make them easier to distinguish from laughter—for which propagation is probably less important—but not necessarily for reasons intrinsically related to the emotions they convey. Second, this approach may bias the researcher towards overlooking important nuances, not the least of which is that screams sometimes convey joy (Anikin & Persson, 2017), and laughs a variety of negative emotions (Darwin, 1872/1998; Provine, 1996; Russell et al., 2003).

A more interesting point of comparison between linguistic and nonlinguistic expressions of emotion is perhaps the acoustic variation *within* call types. Call types such as laughs and screams bear their own acoustic and perceptual constraints, in that exemplars must possess some combination of acoustic traits for listeners to recognize them as belonging to their respective types (Kipper & Todt, 2002; Schwartz et al., 2020a), and all exemplars of a type likely elicit some shared perceptual effects (Schamberg et al., 2018). Nonetheless, when a listener hears a scream, they might attribute one of several different emotions (Frühholz et al., 2021; Study 1). The question emerges how the acoustic variation leading to heterogeneous attributions of emotions from screams compares to the variation that enables heterogenous interpretations of the same linguistic phrase. To the extent that linguistic and nonlinguistic emotional expression derive from the same global effects of emotion on vocal production, patterns of variation within call types should closely mirror those seen in emotional prosody, and listeners should perceive and judge those patterns similarly. In other words, we might expect that the patterns of emotional expression and perception are mostly independent of call type and largely transcend a particular vocal channel. Points of divergence, on the other hand, may begin to implicate either unique patterns of emotional expression that emerged with speech, or patterns unique to specific call types that are less easily reducible to differential phonemic constraints.

A second complication when interpreting the literature is that only a few studies (none of which accounted for call type; Hawk et al., 2009; Pell et al., 2015; Sauter et al., 2013) have compared nonlinguistic vocal communication and prosody in the same experiment. Patterns of response on emotion recognition tasks are well known to vary with factors such as the researchers' procedures for stimulus selection and the number and nature of response options available to participants (Elfenbein & Ambady, 2002; Russell et al., 2003). As a result, comparing findings across studies and research teams is always a fraught proposition, one that especially hampers any claims regarding differences in one vocal channel versus another. Methodological discrepancies might explain why some researchers have reported much lower recognition rates for nonlinguistic vocalizations, more comparable to those seen in speech. See, for examples, Bänziger & Scherer (2007), whose participants chose from 18 different emotions; Gendron et al. (2014), who used a free-naming task rather than one with provided emotion labels; and Sauter & Fischer (2018), who reported drastically different recognition rates for vocalizations taken from different stimulus sources.

In sum, more work is needed investigating acoustic variation in nonlinguistic vocalizations while accounting for call types—for example, by formally defining and selecting call types in advance and exploring the variation within them, rather than collapsing analyses across many types—and directly comparing this variation to speech prosody given the same experimental parameters.

In Study 1 of this dissertation, listeners rated screams on six emotion categories, revealing that contextually diverse screams elicited heterogenous profiles of emotion ratings from listeners. We also identified some key acoustic parameters involved in these perceptions. The emotion categories from that study—anger, fear, frustration, happiness, pain, and surprise—corresponded to contexts in which screaming naturally occurs (Anikin & Persson, 2017). However, the limited number of stimuli from some emotion categories did not enable analyses of listeners' accuracies at identifying each emotion, nor of the acoustic cues important for distinguishing those emotions from each other. Additionally, we did not compare patterns of variation and perception in screams to those in analogous speech samples.

The design and objectives of this study built on the findings from Study 1 in several ways. First, with the use of a forced-choice emotion recognition task, along with a larger set of exemplars representing each emotion, we aimed to assess listeners' accuracies in each emotion category. We used the same emotion categories from Study 1 with the exception of frustration, which in that study listeners seemed to conceptualize and/or perceive very similarly to anger. Second, the larger stimulus set in this study enabled examination not only of the acoustic cues predicting listeners' *perceptions* (a partial replication of Study 1), but also of the cues important to the *expression* of different emotions; that is, how screams associated with different emotions actually varied acoustically. Third, by collecting comparable data from emotional prosodic speech samples, we aimed to compare the acoustic and perceptual differences between emotions within screams to those within speech.

The final objective of this study was to use listeners' ratings of screams to determine their perceptual effects other than or in addition to emotional communication. As reviewed in the General Introduction, functional perspectives on call types and other emotional expressions, largely adopted from the literature on animal signaling behavior, emphasize their direct behavioral effects on listeners that may or may not involve emotional communication (Fridlund, 1997; Owren & Bachorowski, 2003), and/or the information conveyed that is relevant to listeners' behavior but is not limited to the sender's internal state (e.g., details about external events).

This perspective is not antithetical to the idea that calls convey emotion, but instead complements research on emotional communication by focusing on listeners' responses and the acoustic variation underlying them. Some effects on listeners likely involve judgments of emotion but may entail additional judgments or cognitive processes that play into the listeners' decision-making. For example, an effect of agonistic screams in chimpanzees (*Pan troglodytes*) and other primates is to promote aid from kin or other allies (Cheney, 1977; de Waal & van Hooff, 1981). Aid recruitment in primates might involve emotional communication, but the decision listeners make to ignore or respond to a scream is contingent on additional information such as the caller's identity (Bergman et al., 2003; Gouzoules & Gouzoules, 1990) and contextual cues correlating with scream acoustics (Gouzoules et al., 1984; Slocombe & Zuberbühler, 2007). Other effects may not require emotional communication at all, such as antipredator screams in nonhuman animals that might have originated to startle predators.

Importantly, these different potential effects of screams bear different implications for their acoustics. For example, screams that recruit aid might exhibit acoustic properties correlating with identity cues and context-relevant details that are significant to listeners' decision-making processes, whereas similar acoustic variation is not likely needed in the context of startling predators. Conversely, shared acoustic properties between screams in different contexts might implicate common effects. Thus, an investigation of the effects of human screams might provide insights related to the acoustic variation within the type and help to illuminate common effects that screams share across diverse contexts (Anikin & Persson, 2017).

We developed our hypotheses for screams' perceptual effects based mostly on evidence from the nonhuman primate agonistic screams, as it is very likely that screams in our species derive from homologous evolutionary precursors. We also considered the proposed effects of antipredator screams, which are more phylogenetically widespread vocalizations that prey animals emit when captured by a predator, and which may comprise the evolutionarily starting point for primate (and other species') screams (Högstedt, 1983; Rohwer et al., 1976). Finally, we considered hypotheses more directly positing the effects of screams in humans (Arnal et al., 2015; Belin & Zatorre, 2015), although formally stated hypotheses are relatively scarce.

It is important to reemphasize here that although our hypotheses are largely based on homologous data from the animal literature, functional perceptual effects might emerge among screams without positing strictly evolution-based, invariant acoustic subtypes. Humans clearly modulate their voices in many ways that are not wholly fixed by selection, but some such acoustic variation may nonetheless prove functional. For example, the patterns according to which vocalizers modulate the fundamental frequency of their speech might reflect individual conditioning, social learning, and other somewhat flexible mechanisms. Nonetheless, many of these patterns are potentially non-arbitrary (Pisanski et al., 2016), and regardless, listeners can equally learn to associate reliable patterns of variation with interpretations leading to appropriate behavioral responses. It is significant to know whether some screams, e.g., elicit more tendencies towards providing aid regardless of the extent to which the acoustic variation enabling those responses was learned by vocalizers and listeners.

In all, we tested five different hypotheses using five different perceptual rating scales, presented in **Table 1,** which also lists the shorthand term I will use to refer to each hypothesis in this manuscript**.** These hypotheses are not mutually exclusive and in fact may build upon or facilitate one another. As explained below, hypotheses for some effects might tap into potential mechanisms underlying others.

Table 1. Hypotheses and Rating Scales to Investigate Perceptual Effects of Screams

Hypothesis	Example sources	Rating (Agreement; 1-5)
Elicit aid *Shorthand: Aid*	Blumstein et al., 2008; Cheney, 1977; de Waal, 2008; de Waal and Van Hooff, 1981; Gouzoules et al., 1984; Gouzoules & Gouzoules, 1989; Jovanovic & Gouzoules, 2001; Rohwer, Fretwell, & Tuckfield, 1976	I would help this person.
Avert aggressors *Shorthand: Aversion*	Owren & Rendall, 2001; Owren & Rendall, 2003; Rendall et al., 2009	This vocalization is unpleasant to listen to.
Induce fear or alarm *Shorthand: Fear*	Arnal et al., 2015; Belin & Zatorre, 2015	This vocalization makes me feel afraid.
Attract attention or induce arousal (not specific to aid or fear; also potentially related to startling listeners) *Shorthand: Attention*	Dezecache, Zuberbühler, Davila-Ross, & Dahl, 2017; Driver & Humphries, 1969; Fecteau et al., 2007; Högstedt, 1983; Møller & Nielsen, 2010	This vocalization would capture my attention.
Communicate distress *Shorthand: Distress*	Blumstein & Chi, 2012; Groulx, 2005; Hawk et al., 2009	This person is in trouble.

Our first hypothesis (Aid) was that screams solicit aid from listeners, akin to their role in nonhuman primates (de Waal & van Hooff, 1981; Gouzoules et al., 1984) as well as some species whose antipredator screams attract conspecifics or other prey animals to elicit mobbing behavior towards the predator (Chu, 2001; Stefanski & Falls, 1972). In some primate species, acoustic variation among agonistic screams is linked to different contextual details of the encounter, such as the rank of the agonistic opponent (Gouzoules et al., 1984; Gouzoules & Gouzoules, 1989) and/or the severity of aggression (Mercier et al., 2019; Schwartz et al., 2020b;

Slocombe & Zuberbühler, 2007). Playback experiments have revealed that listeners respond preferentially to screams associated either with more severe aggression (Slocombe et al., 2009) or, at least in rhesus macaques, with potential rank challenges from individuals lower in the dominance hierarchy (Gouzoules et al., 1984). It is doubtlessly the case that human screams sometimes elicit aid, but it is also plausible that acoustic variation among screams might evoke different inclinations towards providing aid. In particular, listeners' interpretations of the vocalizer's emotion (among other factors, such as cues to identity) are likely to influence their decision. If listeners are sensitive to the acoustic variation among screams associated with different emotions, we might predict that certain screams, such as those expressing fear, elicit greater inclinations toward aid than others.

Our second hypothesis (Aversiveness) was that screams might avert or deter aggressors. The most notable version of this account, advanced by Owren & Rendall (2001), holds that characteristics of primate screams, such as their high amplitudes and chaotic noise, evolved because they might be perceptually noxious or aversive and therefore likely to repel aggressors. This hypothesis—an example of this Owren and Rendall's model of *affect conditioning*, which emphasizes the direct effects of vocalizations on receivers' nervous systems over the potential information conveyed—remains untested in nonhuman primates. Nonetheless, we aimed to test whether some acoustic features of human screams are aversive. Again, we predicted that ratings of aversiveness might vary by emotional context, although because these ratings represent more direct judgments of the sound characteristics, varying judgments of emotion do not necessarily underlie variation in perceived aversiveness.

A third hypothesis (Fear) was that screams induce fear or alarm in listeners (as opposed to, or not limited to, expressing fear in the caller). One basis for this hypothesis was the finding by Arnal et al. (2015) that a perceptual attribute known as *roughness* is a defining feature of screams, and that listeners rated rough screams as more fear-inducing than screams from which roughness was artificially reduced. More recently, we observed natural variation in roughness among screams (Schwartz et al., 2020a). Other studies have documented roughness in non-scream call types (Li et al., 2018), suggesting that this feature is not unique to screams, but it is nonetheless possible that activation of fear is a critical component of how screams function. For example, fear induction is one plausible mechanism underlying the two effects described above: listeners might feel more inclined to provide aid (as allies) or withdraw (as an opponent) if they are afraid. This process could occur with explicit emotion recognition, or theoretically without it if our auditory systems are indeed especially attuned to properties such as roughness. Either way, it seemed likely that the acoustic profiles of screams associated with some emotions, such as fear and potentially anger, would induce more fear than others.

Our fourth hypothesis (Attention) was that screams capture attention or induce arousal regardless of whatever other effects they achieve. We included this hypothesis for several reasons. First, screams might function in some ways that involve capturing attention without necessarily invoking the effects hypothesized above. The evolutionarily basal function of startling predators, for example, would not involve attracting conspecifics or evoking fear *per se*, but could conceivably promote a call's attention-getting acoustic characteristics. Moreover, various reports in the nonhuman animal literature suggest that screams elicit general attentional responses, such as prolonged orientation towards the sound source (this is the behavioral basis of most playback experiments; Gouzoules et al., 1984; Slocombe & Zuberbühler, 2005), and some proposed functions of animal screams may rely on this property without necessarily invoking the other functions listed above (e.g., antipredator screams that attract other predators to enhance the caller’s probability of escape; Högstedt, 1983). Additionally, testing this hypothesis enabled some insights regarding the mechanisms underlying other effects of screams. For example, we would predict that eliciting attention is a prerequisite to eliciting aid, but not sufficient by itself: as explained above, other cognitive processes likely contribute to the decision to intervene. If so, then the tendency for screams to elicit aid might prove more specific to some emotional contexts than the tendency for screams to elicit attention.

Our final hypothesis (Distress) was that screams communicate distress regardless of the behaviors consequently elicited. Again, the inclusion of this scale enabled better insight into the mechanisms underlying listeners’ decision-making processes. For example, we could determine if, when listeners indicated they would not provide aid to a caller, they did so because they did not identify the caller as distressed, or if they sometimes attributed distress but did not wish to provide aid for some other reason.

Methods

Testing took place from February through October of 2020. This research was conducted in compliance with Emory’s Institutional Review Board under IRB00051516, approved July 26, 2011.

Participants

The participant pool consisted of 97 subjects (42 male, 53 female, 2 unreported; mean age: 19.19, *SD*: 1.12). This sample size was chosen because not every participant heard every stimulus for each task (see procedure below); a sample size above 90 ensured that every stimulus was rated by a minimum of $n = 30$ listeners. Participants were recruited from introductory psychology courses at Emory via an online portal system and received credit for their participation. All participants provided their voluntary and informed consent.

Stimuli

Stimuli were recorded in the sound-proofed Student Production Studio of the Emory Woodruff Library, a recording facility with sound-absorbing acoustic paneling to mitigate background noise and reduce reverberations. Recordings were made on an iMac computer (model 18,3; Apple Inc., Cupertino, CA) using a Yeti USB microphone (Blue Microphones, Westlake Village, CA) providing audio input to Adobe Audition 3.0 (Adobe Inc., San Jose, CA). Sounds were recorded in stereo using a cardioid recording pattern at a 44.1 kHz sampling frequency and a 32 bit-rate.

Nine professional actors (five female, four male) provided the screams and speech samples for this study. Actors were provided a $150 Amazon gift card as an honorarium. Each actor produced 30 recorded vocalizations in total, including 15 screams and 15 speech samples, each of which included 3 renditions of each emotion (anger, fear, happiness, pain, surprise). Thus, in all, 270 stimuli were recorded (9 actors x 2 vocalization types x 5 emotions x 3 renditions).

No instructions were provided to actors regarding how the screams should sound, except that they should contain no speech; otherwise, they were asked to produce a scream they considered appropriate to the emotion. For speech samples, actors were requested to say a name, "Alex," as if they were feeling the emotion. The use of a short name rather than a longer carrier phrase limits the potential for acoustic variation, but our goal was to choose a semantically neutral utterance of comparable duration to screams (Hawk et al., 2009). The particular name "Alex" was also selected because it could refer to a male or female.

Actors were provided examples of scenarios associated with each emotional context (e.g., for happiness, "Your favorite sports team wins the championship.") These examples were validated in an online questionnaire of 18 participants asked to categorize each scenario according to its emotion. All of the scenarios were categorized to the intended emotion by at least 10 of 18 participants. The actors were not told to act out each specific scenario, but rather to use them collectively as exemplary reference points for the intended emotional state. Actors were told to use whatever theatrical or stage techniques would best help them generate the specified emotion and produce the associated vocalizations, and were given as long as they needed for each sample. They were also permitted to substitute a different vocalization if they were not satisfied with an initial rendition, but no actor requested a substitute.

Screams and speech samples were recorded to individual sound files. Actors produced every scream or every speech sample before proceeding to the next vocalization type; this was done so as not to require adjustments to microphone gain (due to the disparate amplitudes of screams and speech) midway through a file. The first-recorded vocalization type was alternated among actors.

When a recording of either vocalization type began, actors were asked to produce two practice vocalizations corresponding to two emotional contexts not represented among the stimuli (triumph and embarrassment). Adjustments to sound recording level were made after these practice vocalizations if necessary (e.g., due to clipping). The experimental stimuli were then recorded. For each stimulus, the experimenter provided the emotion, which the actor verbally repeated before producing the vocalization. Stimuli for each emotion were recorded in a semi-randomized order in blocks of five, such that actors provided one rendition for each emotion before moving on to the next renditions. No stimuli corresponding to the same emotion were recorded twice in a row. The total recording time for a given vocalization type (i.e., all 15 stimuli) generally ranged from 2-4 minutes and never exceeded 5 minutes.

After a recording session, the following processing steps were taken within Adobe Audition. Recordings were converted to 16-bit mono files (44.1 kHz sampling rate). The Marquee selection tool was used to highlight each vocalization, leaving .3 s silence on either side, and save it as an independent .wav file. Finally, the amplitude envelope of each file was normalized to a peak amplitude of -9 dB by highlighting the whole sound and using the Adjust Amplitude tool.

Experimental procedure

This study consisted of three primary tasks: Emotion categorization, functional ratings, and sound classification. The order of these tasks was counterbalanced between participants. Additionally, to prevent fatigue, participants were not required to listen to all stimuli. Instead, they were assigned to one of three groups determining the subset of stimuli they heard. For the relatively shorter classification and categorization tasks, group A heard the first and second renditions, group B heard the first and third, and group C heard the second and third. For the longer ratings task, group A heard only the first renditions, group B heard the third, and group C heard the second.

Trials in every task followed the same basic procedure. Trials began with the phrase "Please listen now" appearing on-screen for 1.5 s. A stimulus then played through the headphones as the response options appeared on-screen. The response options were grayed out and not selectable until the stimulus was finished playing, at which point they turned green to indicate that the participant could respond. Once a response was provided, the next trial proceeded. All stimuli within each task were presented in a randomized order.

Participants were encouraged to respond as quickly and as accurately as possible, but there were no time limits for responding. Participants were not able to replay stimuli. More specific procedures for each task are provided below.

Experiments were programmed and data were collected via E-Prime 3.0 software (Psychology Software Tools, Inc., Pittsburgh, PA). A subset of participants (*n* = 55) completed the study online due to the Covid-19 pandemic. Modifications pursuant to online data collection are described in a section below. Those participants who completed the study in-person did so on a Dell OptiPlex 755 computer with sounds delivered through Beyerdynamic headphones (model DT 770 PRO 250, Beyerdynamic GmbH & Co. KG, Heilbronn, Germany).

Emotion categorization. Participants were asked to categorize each vocalization according to the emotion they thought it conveyed. Response options were provided via clickable boxes, each with a label corresponding to one of five emotions (anger, fear, happiness, pain, or surprise). Listeners also had the option of selecting a box labeled "Other" if they did not feel a vocalization conveyed any of these emotions. These six response boxes appeared in two rows of three. The order of emotions appearing in these boxes was determined randomly (with the exception of "Other", which always appeared in the last box) but stayed the same throughout the study for each participant.

This task consisted of two blocks, one in which participants heard only nonlinguistic stimuli, and one in which they heard only linguistic stimuli. The order of these blocks was counterbalanced between participants. Participants heard two renditions of each emotion from each actor and for each vocalization type. Thus, each block consisted of 90 stimuli (9 actors x 2 renditions x 5 emotions), and the total task consisted of 180 stimuli.

Perceptual effects ratings. Participants were asked to rate their agreement with a series of statements regarding the vocalizations they heard. When a stimulus was presented, a statement appeared on-screen along with the response options. These statements corresponded to the five presented in **Table 1**. Participants used the mouse to click on boxes labelled 1-5, where 1 indicated strong disagreement with the statement, and 5 indicated strong agreement.

This task was divided into five blocks. In each block, participants heard only screams, including one rendition of each emotion from each actor (45 stimuli per block). Each block included one presentation of each scream, but the statements were presented in randomized order within a block. Thus, in the complete task, participants heard each stimulus five times and rated their agreement regarding every statement.

Sound classification. Participants were asked to classify each nonlinguistic vocalization according to whether or not they considered it a scream. Responses were recorded via an external button response box (Psychology Software Tools, Inc., Pittsburgh, PA) with buttons labelled "1" and "2". On-screen instructions indicated which button corresponded to which response (scream or not). The correspondence of these buttons to responses was counterbalanced between participants. This task consisted of one block in which participants heard only nonlinguistic stimuli, including two renditions of each emotion from each actor (90 stimuli total).

Modifications for online data collection. The Covid-19 pandemic necessitated online data collection for the last 55 participants of the study. This was achieved using E-Prime Go (Psychology Software Tools, Inc., Pittsburgh, PA), a remote data collection software with which we created transferrable executable files for each experiment. Meetings were set up using Zoom video conferencing software (Zoom Video Communications, San Jose, CA) to walk participants through the initial steps of the study. In these meetings, participants were directed to an online form (JotForm, San Francisco, CA) where they provided demographic information. Through this form, they downloaded the consent form, signed it electronically, and uploaded it for the experimenter to sign. They then downloaded the three executable files corresponding to the three tasks. These downloads were listed in an order counterbalanced between participants, and participants were asked to complete them in the listed order. Once participants successfully downloaded the first file, the experimenter signed off Zoom while remaining on-hand via e-mail in case of technical difficulty.

Participants were told to complete the study in one sitting in a quiet environment, and not to adjust their volume once the experiment began. They were also strongly encouraged to use headphones, if possible. The experiments themselves were almost entirely unchanged, except that in the classification task, rather than using an external button response box to provide input, participants responded using the “1” and “2” buttons on their keyboards. After participants completed all three tasks, they uploaded their data through the online form, where the experimenter could then access the data.

Acoustic analysis

For each vocalization, 14 acoustic parameters were measured from spectrograms generated via Fast-Fourier Transform (FFT) on Praat (Boersma & Weenink, 2013) and Raven 1.5 (Center for Conservation Bioacoustics, 2014) software. Duration was measured by using Praat’s selection tool to highlight the spectrogram from vocalization onset to end. Mean F0, F0 range, mean harmonics-to-noise ratio (HNR, a measure of periodicity), jitter (relative average perturbation, a measure of local frequency modulation), and shimmer (local dB, a measure of local amplitude modulation) were measured using the Quantify Source command in the GSU Praat Tools script package (Version 1.9, Owren, 2008). Measurements were made using a 50 ms analysis window, 75-Hz pitch floor, and 2000-Hz default ceiling. The automatically generated F0 contour was checked manually for errors (e.g., spurious voiced segments or octave jumps) by visual and auditory comparison to the original spectrogram before obtaining measurements of mean F0, mean HNR, jitter, shimmer, F0 minimum, and F0 maximum; F0 range was calculated as the difference between the latter two. The time of max F0 was determined manually using the selection tool to mark the timepoint corresponding to the peak in F0 contour. This value was divided by the vocalization's duration to calculate the percent time at which F0 reached its peak. Slope to max F0 was calculated by dividing the difference between max F0 and start F0 by the difference in seconds between these points. One speech sample—the first actor's third rendition of anger—had no voiced elements and therefore was omitted from analyses relying on F0 measures.

Measurements for interquartile range (IQRBW, a measure of the width of energy distribution, where larger values correspond to more diffuse energy across the spectrum), mean peak frequency (PF mean, the frequency with the greatest energy in the spectrum), and DFA50 (or center frequency, the frequency above and below which the spectrum contains equal energy) were made using Raven 1.5. Sounds were highlighted using the Selection tool on the spectrogram using a 23 ms window length. The window length was changed to 46 ms for analyses. Parameter measurements were obtained from each selection using Raven's Choose Measurements tool. Pitch jumps and other nonlinear phenomena (i.e., subharmonics or deterministic chaos) were coded as binary variables based on the presence or absence of these phenomena as determined by visual inspection of the spectrogram (Fitch et al., 2002). Pitch jumps involved a sudden leap up or down in frequency and appeared as complete discontinuities in the F0 contour. Subharmonics were evident when a second fundamental frequency and harmonic stack appeared at half the frequencies of the first. Deterministic chaos was evident as broadband noise overlaid atop some visible periodic bands, the presence of which distinguished it from turbulent noise (Fitch et al., *ibid.*). Subharmonics and chaos often co-occurred in the same sound (unsurprisingly, as variation in subglottal pressure or vocal fold tension across time can lead to transitions, or bifurcations, between the two; Fitch et al., *ibid.*), so these observations were treated as one variable, nonlinear phenomena (NLP), documenting any case in which the voice exhibited a nonlinear vibratory regime. Finally, measurements for roughness were collected using the modulationSpectrumFolder function in the *soundgen* package (Anikin, 2019) in R.

Despite efforts to eliminate clipping during recording, clipping was observed in $n = 20$ sound files, mostly due to microphone issues during the recording session with actor 2. These files were omitted from any analyses involving parameters potentially sensitive to spectral distortions, namely, mean HNR, PF mean, DFA50, IQRBW, and roughness. Means for every acoustic parameter for each actor are presented in **Table A1** in the appendix.

Statistical analysis

The following section provides a general overview of our statistical analyses. The appendix (Statistical Details for Study 2, p. 177 provides greater detail regarding the selection of parameters, tests, and model specifications.

Emotion categorization accuracy. For every participant, the overall percent correct (or raw hit rate, HR) was calculated separately for both linguistic and nonlinguistic stimuli. These HRs were compared using a paired sample *t*-test to determine if, in general, participants were more accurate on one vocalization type relative to the other. In addition, we used a Pearson's correlation to determine if accuracy on one vocalization type correlated with accuracy on the other. We also derived each participant's overall false alarm rate (FA) and unbiased hit rate (*Hu*; Wagner, 1993, a measure akin to *d'* but for judgments with > 2 response options) for each emotion category and vocalization type. The effects of Emotion, Vocalization Type (linguistic or nonlinguistic), and their interaction were analyzed using two-way repeated measures ANOVAs. (Note that we will capitalize terms, e.g., Emotion, Vocalization Type, Actor, and Rendition when referring to them as variables in statistical tests or models.) Finally, we qualitatively examined participants' patterns of errors using confusion matrices for linguistic and nonlinguistic stimuli.

Acoustic differences between emotional contexts. To examine the relationship between acoustic parameters and emotions, 14 separate mixed models were initially fitted with each acoustic parameter as an outcome variable and Emotion, Vocalization Type, and their interaction as predictor variables. Rendition (first, second, or third within an emotion category and vocalization type) was also included as a categorical fixed effect, and Actor was included as a random effect to account for the non-independence of stimuli produced by the same vocalizer and to account for vocalizer-specific baselines for each acoustic parameter. The goal of these initial models was to determine whether the Vocalization Type * Emotion interaction was significant. If not, we refitted the model without Vocalization Type or the interaction term to examine the main effects of Emotion across screams and speech together, whereas if the interaction was significant, we fitted separate models for each vocalization type to explore simple effects of Emotion within speech and within screams. When Emotion was significant, post-hoc tests were used to determine how the acoustic parameter varied by each emotion.

Acoustic effects on listeners' responses. To test the effects of acoustics on listeners' responses while limiting multicollinearity in our models, we selected a subset of relevant acoustic parameters: duration, mean F0, jitter, time to max F0, mean HNR, IQRBW, pitch jumps, and NLP (see appendix for details on parameter selection). To determine how the acoustic parameters predicted listeners' categorizations of emotion, we developed composite variables, for each emotion and within each vocalization type, reflecting the number of listeners who categorized a given stimulus as conveying that emotion. These variables were used as the outcome variables in five separate negative binomial generalized linear mixed models (GLMMs), with all 8 of the selected acoustic parameters as fixed effects and Actor as a random intercept. To account for the possible effects of vocalizer sex on emotion perception, vocalizer sex was also included as a binary predictor variable in these models.

Perceptual ratings. To investigate how a scream's original emotional context affected listeners' perceptual ratings, four cumulative link mixed models (CLMMs) were fitted with ratings on each scale as the ordinal outcome variables, Emotion as a categorical predictor variable, and Participant and Stimulus as random effects. CLMMs with the same random effects were used to examine the effects of the 8 selected acoustic parameters on ratings. Vocalizer sex was also included as a fixed effect in these acoustic models. Due to a very high correlation between Distress and Help ($r = .97$), Help was not included in these analyses.

Sound classification. Results from the sound classification task were used to arrive at a more conservatively defined category of screams. Although actors were asked to provide screams, not every vocalization they produced was necessarily perceived as a scream by listeners. A previous study in our lab documented significant variation in listeners' classifications of sounds as screams versus non-screams (Schwartz et al., 2020a). Here, we developed a binary variable reflecting whether or not > 50% of listeners classified a vocalization as a scream. We chose this cutoff because it reflects an intuitive, *ipso facto* majority while preserving ample representation of each emotion (8 or more stimuli per emotion; see **Table 2**).

Adopting this criterion with the data from our prior classification study (Schwartz et al., *ibid.*) also generated strong agreement with our own assessments of the stimuli: 36/38 sounds that we identified as screams in that study were classified as screams by >50% of those listeners, while only 5/37 sounds that we did not identify as screams were classified as screams by this criterion (i.e., false alarms; *d'* = 2.722); these generally fell into a category we would classify as yells.

Table 2. Stimuli classified screams by actor and emotion.

Actor	Anger	Fear	Happiness	Pain	Surprise	**Total**
1	0	1	1	1	2	**5**
2	0	3	2	2	2	**9**
3	2	0	2	1	0	**5**
4	1	3	2	0	1	**7**
5	1	3	2	2	2	**10**
6	1	3	2	1	1	**8**
7	2	3	3	2	3	**13**
8	1	1	2	0	0	**4**
9	0	1	1	0	1	**3**
Total	**8**	**18**	**17**	**9**	**12**	**64**

Note. Sounds were considered screams if classified as such by over 50% of participants.

In all, *n* = 64 out of 135 nonlinguistic sounds in this study were classified as screams according to this criterion (**Table 2**). This classification was not independent of emotion: fear and happiness stimuli were more represented among sounds classified as screams, whereas anger and pain stimuli were less represented. Similarly, actors were not equally represented among this subset: the number of more strictly classified screams actors produced ranged from 3 to 13 out of 15 nonlinguistic stimuli. (Interestingly, the two actors who produced the most stimuli classified as screams were female, and two who produced the least were male, suggesting a role of vocalizer sex in listener's classification of screams.) The imbalance in actor representation was a potential concern given differences in the acoustic characteristics of their voices. The inclusion of Actor as random intercept in acoustic models accounts for baseline differences, but potentially subtler actor-specific effects—e.g., differences in how actors expressed certain emotions—were too complex to specify in our models and therefore warrant caution.

All analyses described above were repeated on the subset of sounds classified as screams by the majority of listeners. Results reported in this manuscript focus on the analyses of the broader nonlinguistic set, which are consistent with how many studies on human vocalizations define call types (i.e., by study design and without listener input); however, major similarities and differences arising from analyses on listener-defined screams are noted as a point of interest. It should be noted that screams can grade into other call types (Anikin et al., 2018) and (consistent with other call types) are not clearly demarcated by strict acoustic boundaries (Schwartz et al., 2020a). Given our instructions to actors, we suspect that many nonlinguistic vocalizations in our study, if not classified as screams >50% of listeners, are likely nonetheless "scream-adjacent" (rather than, e.g., clearly belonging to a different, unrequested call type like crying). We therefore considered analyses on our broader corpus of nonlinguistic vocalizations as well as on our more conservatively-defined, listener-classified screams as informative with respect to this call type, although this is a position we planned to reevaluate after analyses.

Results

Emotion Categorization

Overall accuracy. Participants' mean raw hit rates (HR), unbiased hit rates (*Hu*), and false alarm rates (FA) for each vocalization type and emotion are shown in **Table 3**. A paired-sample *t*-test on overall HRs for each vocalization type revealed that participants did not perform significantly better with linguistic ($M = .49$, $SE = .01$) compared to nonlinguistic stimuli ($M = .50$, $SE = .01$) ($t(96) = 1.799$, $p = .075$).

Participants' HRs on linguistic stimuli and nonlinguistic stimuli were strongly correlated (Pearson's $r = .656$, $p < .001$), revealing that participants who responded accurately to one vocalization type were also accurate with the other (**Figure 1**).

Table 3. Unbiased hit rates, percent correct, and false alarms

	Expressed emotion				
	Anger	Fear	Happiness	Pain	Surprise
Nonlinguistic					
Mean unbiased hit rate (Hu)	0.53	0.18	0.16	0.33	0.35
S.E.	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Mean raw hit rate (HR)	0.63	0.40	0.21	0.68	0.61
S.E.	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
False alarm rate (FA)	0.03	0.13	0.03	0.18	0.13
S.E.	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
Linguistic					
Mean unbiased hit rate (Hu)	0.53	0.30	0.20	0.25	0.25
S.E.	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Mean raw hit rate (HR)	0.72	0.51	0.38	0.42	0.42
S.E.	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
False alarm rate (FA)	0.07	0.09	0.10	0.08	0.09
S.E.	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)

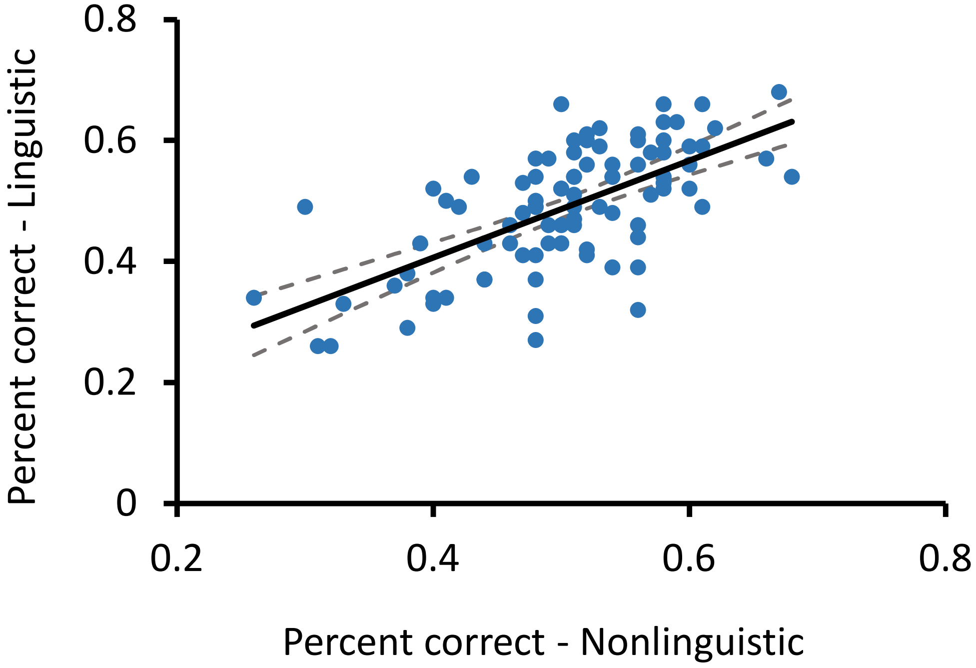


Figure 1. *Correlation between participants' percent correct on linguistic and nonlinguistic stimuli.*

Unbiased hit rates. An ANOVA conducted on the align-rank transformed *Hu* scores revealed a significant effect of the Emotion * Vocalization Type interaction ($F = 45.5$, $p < .001$; **Figure 2A**). Due to this significant interaction, we examined simple effects of emotion within each vocalization type using Friedman's Test with post-hoc Dunn-Bonferroni adjusted pairwise comparisons. For nonlinguistic stimuli, all pairwise comparisons were significant (adjusted $p < .001$) with the exception of happiness versus fear ($p = .146$) and pain versus surprise ($p = .527$). Thus, among nonlinguistic stimuli, participants most accurately categorized anger, followed by pain and surprise, and least accurately categorized happiness and fear. For linguistic stimuli, too, participants categorized anger better than any other emotion ($p < .001$ in every case), and they categorized fear better than pain ($p = .029$) and happiness ($p < .001$), but no other pairwise comparisons were significant.

In all, these results suggest general similarity between linguistic and nonlinguistic stimuli in unbiased hit rates across emotions, with participants categorizing anger with the highest accuracy, pain and surprise with moderate accuracies, and happiness with the lowest accuracy. A possible notable difference is that, for linguistic stimuli, participants' categorization of fear was more accurate compared to some other emotions, whereas for nonlinguistic stimuli it was among the least accurately categorized emotions.

Unbiased hit rates depend on both a participant's raw hit rate for an emotion as well as false alarms, i.e., the tendency to use that response even when that response is incorrect. Therefore, a lower *Hu* could result from a low raw hit rate and/or a high false alarm rate. To shed some potential light on the causes of different *Hu* values, we examined the false alarm rates for each emotion and vocalization type using an ANOVA conducted on aligned rank transformed FAs. This ANOVA showed that FAs depended on the interaction between Emotion and Vocalization Type ($F = 121.5$, $p < .001$; **Figure 2B**). Thus, Dunn-Bonferroni adjusted pairwise comparisons between emotions were conducted within each vocalization type.

Within nonlinguistic stimuli, pain categorizations showed a higher FA than any other emotions ($p < .001$), i.e., participants more often incorrectly categorized stimuli as pain compared to other emotions. Fear and surprise had higher FAs than happiness and anger ($p < .001$ in every case), but fear did not differ significantly from surprise nor happiness from anger (adjusted $p = 1$ for both comparisons). Within linguistic stimuli, happiness had a significantly higher FA than anger ($p < .001$), pain ($p = .034$), and surprise ($p =.029$), and fear had a higher FA than anger ($p = .001$). Overall, these findings suggest that false alarm rates were relatively diffuse across emotions in the linguistic stimuli, whereas in the nonlinguistic stimuli, participants showed a clear tendency to incorrectly categorize stimuli as pain and not as happiness or anger.

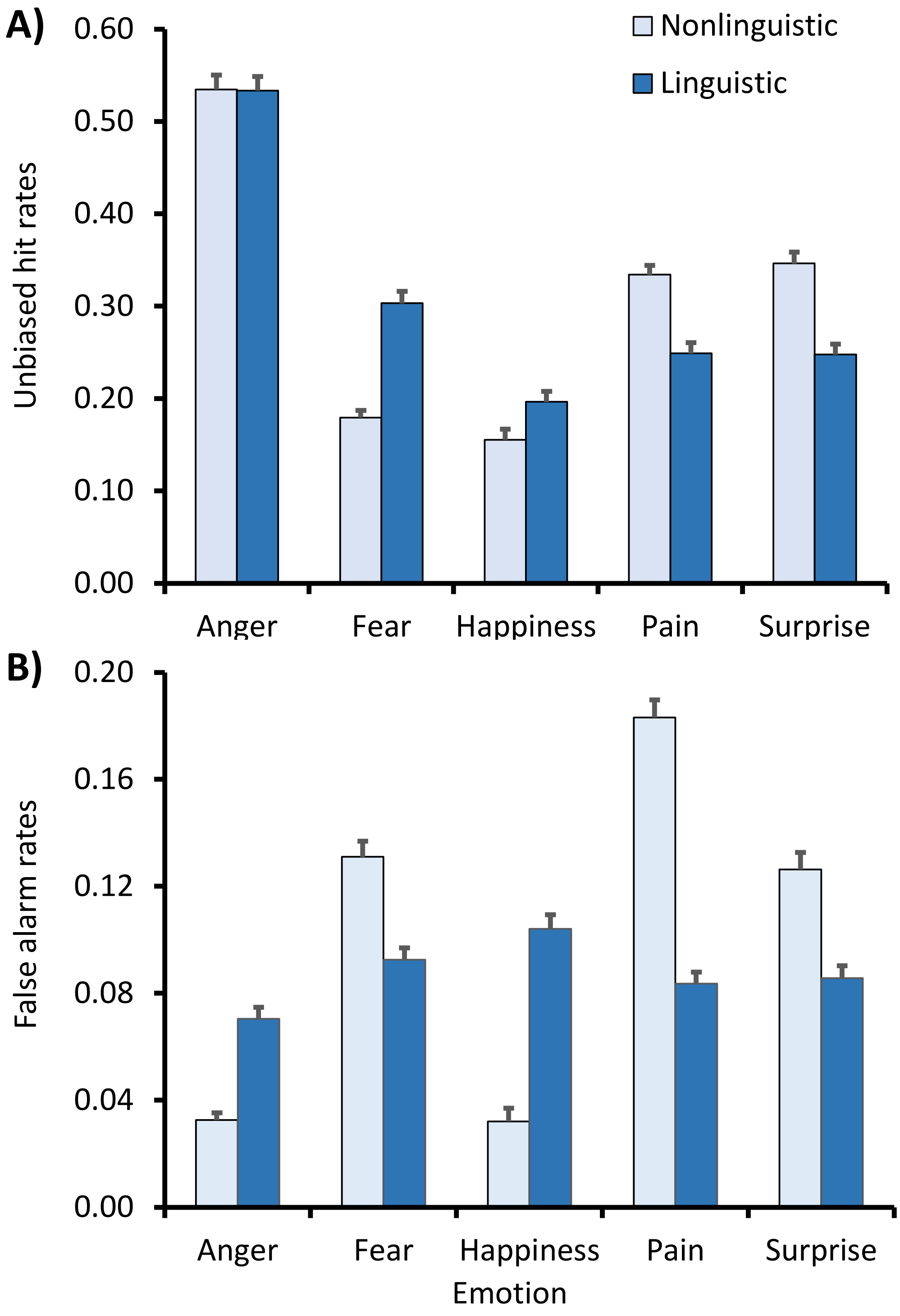


Figure 2. *Participants' (A) mean unbiased hit rates and (B) mean false alarm rates across vocalization types and emotions.*

Patterns of errors. To further examine listeners' patterns of errors, we plotted the confusion matrices for listeners' responses to nonlinguistic and linguistic stimuli (**Table 4)**. The diagonals show participants' mean raw hit rates for each emotion. The remainder of each row shows participants' mean rates of providing each response to a given emotion, e.g., the rate with which they responded "fear" when the stimulus was conveying anger.

An exploratory, qualitative examination suggests some interesting trends. Nonlinguistic anger, fear, and happiness were all commonly mistaken for pain, contributing to its high false alarm rate overall; the same effect was not as pronounced among linguistic categorizations. Nonlinguistic happiness was also commonly mistaken for fear, but fear was very rarely mistaken for happiness. Nonlinguistic surprise and fear were commonly mistaken for one another. Among linguistic categorizations, however, surprise was more commonly mistaken for happiness than for fear.

Table 4. Confusion matrix of participant responses

	Response (%)					
Stimulus	Anger	Fear	Happiness	Pain	Surprise	Other
All nonlinguistic						
Anger	**0.63**	0.02	0.01	0.21	0.04	0.09
Fear	0.03	**0.40**	0.03	0.24	0.24	0.07
Happiness	0.03	0.23	**0.21**	0.24	0.16	0.14
Pain	0.06	0.07	0.03	**0.68**	0.06	0.10
Surprise	0.01	0.21	0.06	0.05	**0.61**	0.06
Linguistic						
Anger	**0.72**	0.03	0.02	0.05	0.03	0.15
Fear	0.03	**0.51**	0.07	0.12	0.15	0.13
Happiness	0.07	0.07	**0.38**	0.11	0.11	0.26
Pain	0.17	0.15	0.05	**0.42**	0.05	0.17
Surprise	0.02	0.11	0.28	0.06	**0.42**	0.11

Note. Bolded diagonals represent hit rate for each emotion.

Acoustic differences between emotions

To investigate the acoustic differences between emotions, we fitted 14 mixed models with each acoustic parameter as the outcome variable, Emotion, Vocalization type, the Emotion * Vocalization type interaction, and Rendition as fixed effects, and Actor as a random intercept. In this initial step, our interest was in the interaction effect, which if significant (as assessed by a likelihood ratio test) would indicate that the effects of emotion on that parameter depended on the vocalization type. The chi-squared statistic and significance of the interaction term for each parameter are presented in **Table 5**.

*Table 5. Significance of the emotion * vocalization type effect on acoustic parameters*

	Emotion * Vocalization type Interaction	
Parameter	χ^2	*p*
Mean F0	2.9	.58
F0 Range	2.5	.64
Duration	**18.1**	**.001**
Time to Max F0	**16.1**	**.003**
Slope To Max	3.2	.52
Jitter	1.8	.77
Shimmer	3.9	.41
Mean HNR	**11.1**	**.025**
PF Mean	7.1	.129
DFA50	4.8	.312
IQRBW	7.9	.094
Roughness	**18.0**	**.001**
Pitch Jump	5.7	.226
NLP	**11.7**	**.02**

Overall, we observed very similar patterns of emotion-correlated acoustic variation between linguistic and nonlinguistic stimuli (**Figure 3;** see also **Figure 4** for sample waveforms and spectrograms of one vocalizer’s screams for each emotion), and for most parameters, the interaction was not significant. For duration, time to max F0, roughness, mean HNR, and NLP, the effects of Emotion depended on Vocalization Type, although the interpretation of HNR and NLP warrants caution as their significance would not survive correction for multiple tests.

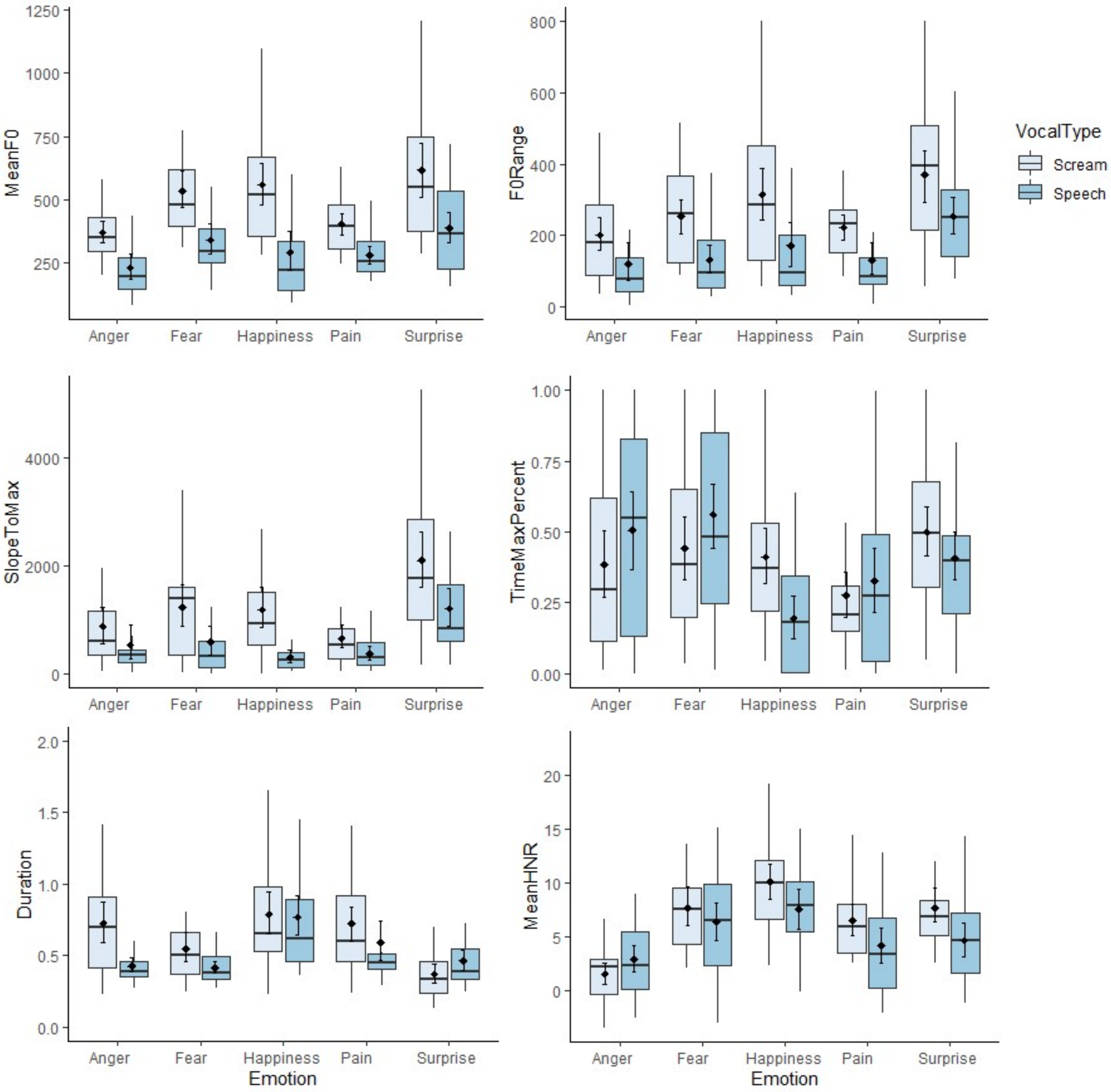
For parameters where the Vocalization Type * Emotion interaction was not significant, we refitted these initial models omitting Vocalization Type and its interaction with Emotion, i.e., examining only the main effects of Emotion on that parameter across all stimuli. These full results are presented in **Table A2** and are summarized below.

F0-related parameters showed similar effects of emotion across vocalization types. Overall, across linguistic and nonlinguistic vocalizations, both anger and pain stimuli were characterized by lower mean F0s and lower F0 ranges, i.e., relatively low, flat pitch, whereas happiness and surprise stimuli exhibited higher F0s and F0 ranges, i.e., higher and wider pitch ranges. Surprise additionally exhibited a greater slope to max F0 compared to all other emotions, indicating that these stimuli were characterized by a dramatic initial increase in pitch over a short period of time. Fear stimuli were characterized by a somewhat unique combination of high F0s but low F0 ranges, that is, higher but relatively less variable pitch.

Other parameters for which Emotion had a significant effect across both vocalization types included shimmer and PF mean (e.g., anger exhibiting the highest shimmer, happiness the lowest). Jitter was significant in the overall model but pairwise comparisons between emotions did not survive corrections for multiple comparisons. Finally, DFA50, IQRBW, and pitch jumps were not significantly affected by emotion.

For parameters in which the Vocalization Type * Emotion interaction was significant, we could examine, cautiously and preliminarily, differences in the effects of Emotion between linguistic and nonlinguistic stimuli. To do so, we fitted a separate mixed model for each parameter within each vocalization type, with the parameter as the outcome variable, Emotion and Rendition as fixed effects, and Actor as a random intercept. In no case was there a significant, directional difference in the effect of an emotion on any parameter: the only differences concerned effect size, hence significance (full results shown in **Table A3**).

Many of the differences concerned the effects of anger on these parameters. For example, for both nonlinguistic and linguistic stimuli, anger and pain stimuli exhibited relatively lower mean HNR (i.e., greater noisiness) and higher roughness compared to other emotions. However, for nonlinguistic stimuli alone, anger was characterized by even lower HNR and greater roughness than pain. Nonlinguistic anger also showed greater incidence of nonlinear phenomena than any other emotion, whereas among linguistic stimuli, pain exhibited the highest incidence of nonlinearities. (Note that, among the nonlinguistic stimuli, the effects of anger on nonlinear phenomena were inestimable because every nonlinguistic anger stimulus exhibited at least some nonlinear phenomena, resulting in quasi-separation; Allison, 2008). Other differences concerned parameters related to the time course of the vocalization, namely, duration and time to max F0. In linguistic but not nonlinguistic vocalizations, anger and fear stimuli were shorter in duration than happiness and pain stimuli. Additionally, in linguistic vocalizations, happiness stimuli exhibited an earlier max F0 relative to other emotions, whereas in nonlinguistic vocalizations, pain reached an earlier max F0 than happiness and surprise stimuli.


MeanF0
1250
1000
750
500
250
F0Range
800
600
400
200
0
VocalType
Scream
Speech
SlopeToMax
4000
2000
0
TimeMaxPercent
1.00
0.75
0.50
0.25
0.00
Duration
2.0
1.5
1.0
0.5
0.0
MeanHNR
20
15
10
5
0
Anger
Fear
Happiness
Pain
Surprise
Emotion

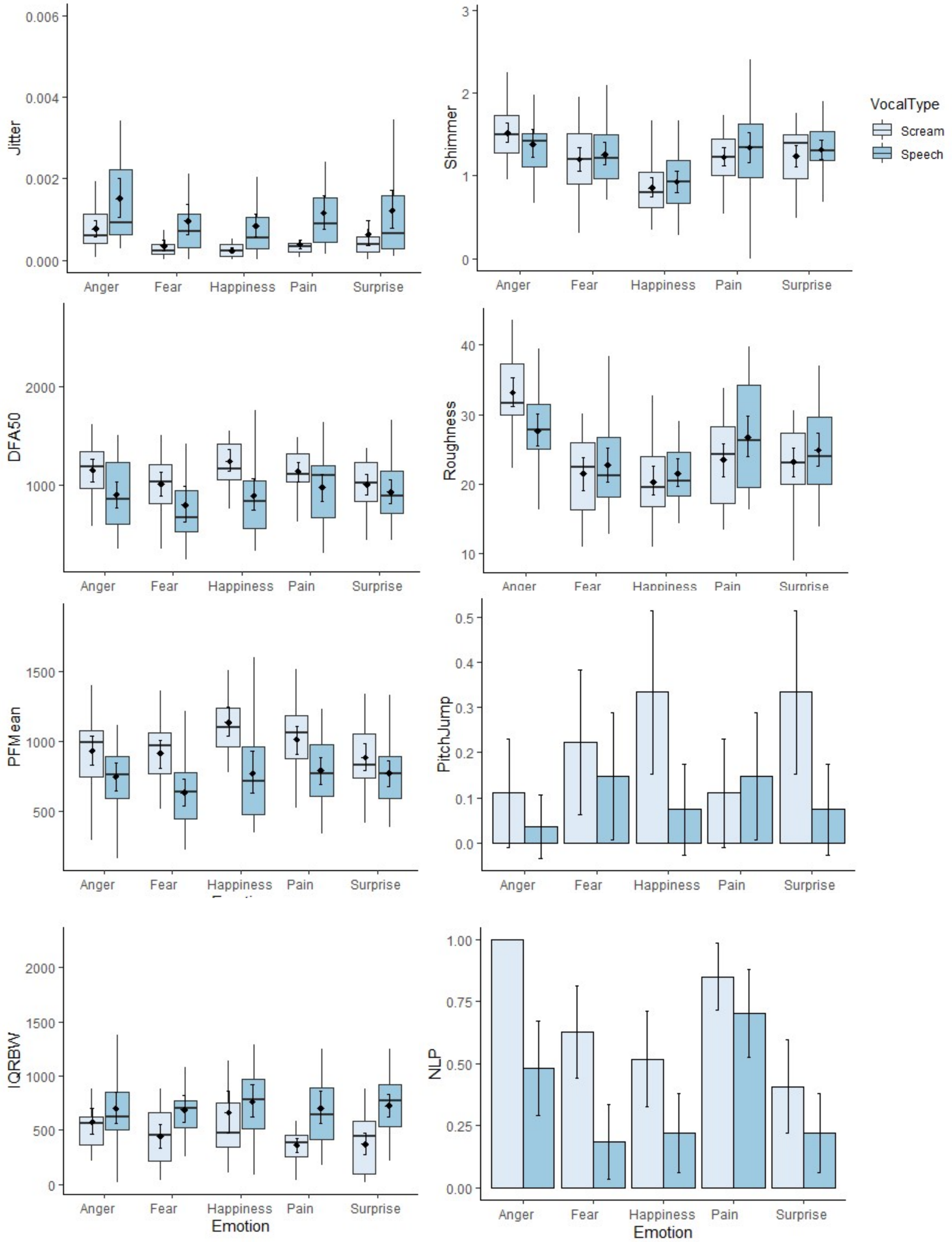


Figure 3. *Means and 95% confidence intervals for each parameter by each vocalization type and emotion. Graphs of binary variables (pitch jump and NLP) portray proportions of each vocalization type and emotion that exhibited each phenomenon.*

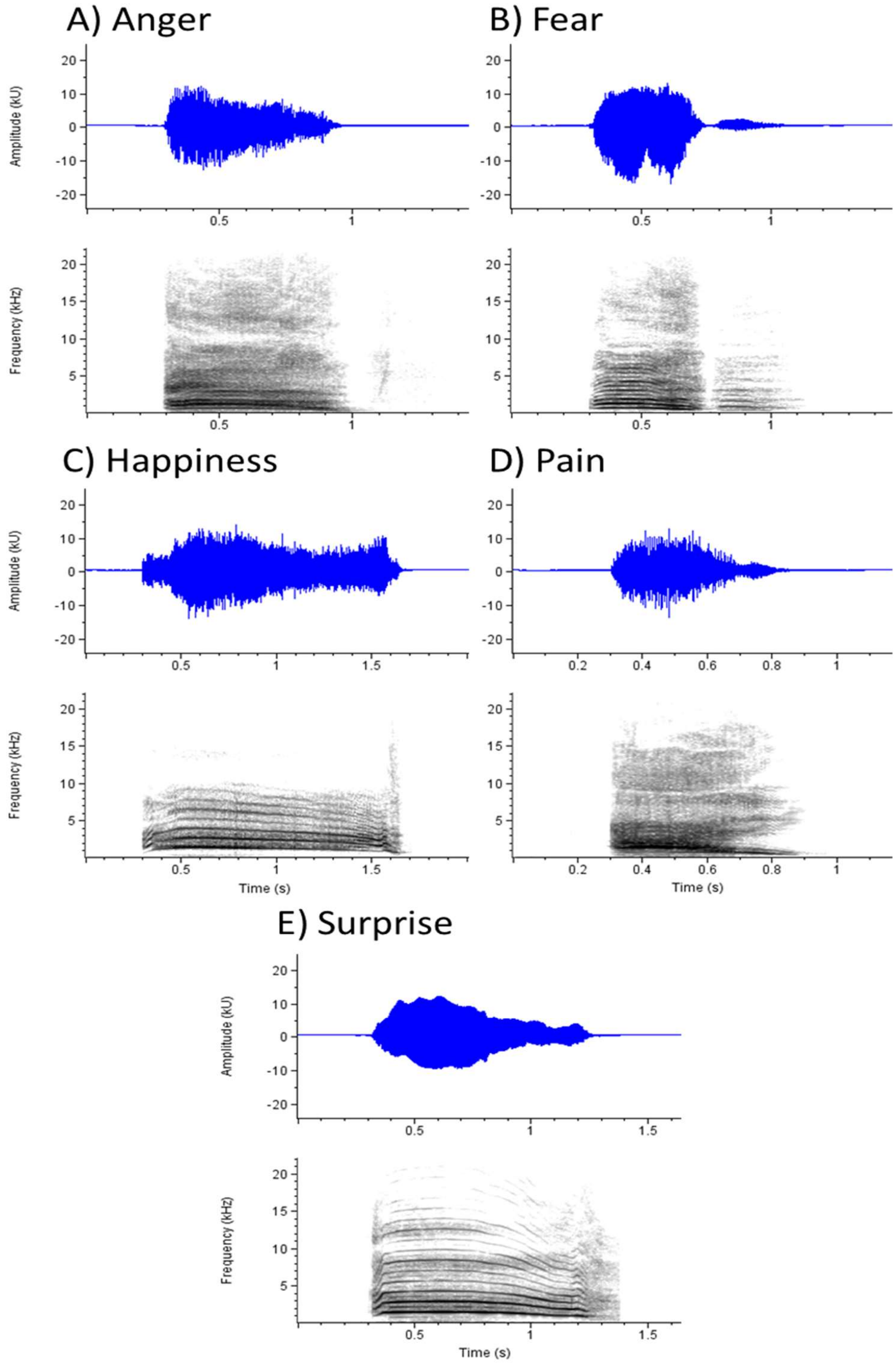


Figure 4. *Waveforms and spectrograms of five screams from one actor (Actor 7) for each emotion, representing the stimuli that the highest proportion of listeners classified as screams for each emotion (Anger 3, Fear 1, Happiness 3, Pain 2, and Surprise 1, all classified as screams by > 70% of listeners; two fear screams with clipping excluded) from the actor who produced the highest number of classified screams (13 out of 15 vocalizations).*

Acoustic predictors of categorization. To examine the effects of acoustic parameters on listeners' categorization of emotions, five separate GLMMs were fitted with the number of listeners categorizing a stimulus as that emotion as the response variable (modeled with a negative binomial response distribution), 8 selected acoustic parameters along with vocalizer sex as fixed effects, and Actor as a random intercept. The effects of each acoustic parameter are shown in the **Table A3**, with significant effects shown in bold.

Every parameter except interquartile bandwidth had a significant effect on listeners' categorizations of at least one emotion, and patterns of acoustic effects on the categorizations showed some similarities between linguistic and nonlinguistic stimuli. Across both vocalization types, anger categorizations were predicted by lower F0 and greater noisiness, fear by shorter duration and greater tonality, pain by an earlier max F0 and greater incidence of nonlinear phenomena, and surprise by shorter duration and higher F0. However, some differences in significant predictors also emerged. Of note, relative tonality or noisiness of a vocalization seemed more predictive of emotion perceptions among nonlinguistic vocalizations, whereas time to max F0 was more predictive of emotion perceptions among linguistic vocalizations.

Perceptual effects ratings

Overall ratings. Mean ratings on every scale were positively correlated with one another (Pearson's correlations, $p < .001$ in every case; **Table 6**). As explained in the Methods section, ratings on the Distress and Help scales, in particular, were almost perfectly correlated with each other ($r = .97$). Thus, Help was dropped from all subsequent analyses on rating scales to reduce the number of overall tests.

Table 6. Correlations between functional rating scales

Rating scale	Aversiveness	Distress	Fear	Help
Attention	0.82	0.68	0.83	0.68
Aversiveness		0.69	0.70	0.64
Distress			0.77	0.97
Fear				0.77

Note. All correlations were significant, p < .001.

Effects of emotion on ratings. Four cumulative link mixed models were fitted with participants' ratings on each scale as the outcome variables, Emotion as a fixed effect, and Participant and Stimulus as random intercepts. Likelihood ratio tests revealed that Emotion had a significant effect on ratings of Aversiveness ($\chi^2 = 28.55$, $p < .001$), Distress ($\chi^2 = 31.29$, $p < .001$), and Fear ($\chi^2 = 13.88$, $p = .007$), but not Attention ($\chi^2 = 5.32$, $p = .256$).

For the rating scales in which emotion was significant, pairwise comparisons were conducted to examine the effects of each emotion relative to others (**Figure 5**). Surprise stimuli were perceived as less averse than stimuli associated with other emotions (anger: $p < .001$, fear: $p < .001$, happiness: $p = .027$, pain: $p < .001$). Vocalizers expressing fear were perceived as more in distress than those expressing anger ($p < .001$), happiness ($p = .006$), or surprise ($p < .001$); vocalizers expressing pain were also perceived as more in distress than those expressing anger ($p = .008$) or surprise ($p = .001$). Finally, fear stimuli were more fear-inducing than stimuli associated with happiness ($p = .006$), pain ($p = .025$), and surprise ($p = .039$).

These data suggest that screams associated with different emotions captured listeners' attention to relatively similar degrees. They were also similar in aversiveness, except that surprise stimuli were somewhat less averse, and in their induction of fear in listeners, except that fear stimuli induced somewhat more fear or alarm. Finally, fear and pain stimuli tended to convey more distress than stimuli associated with anger, happiness, or surprise.

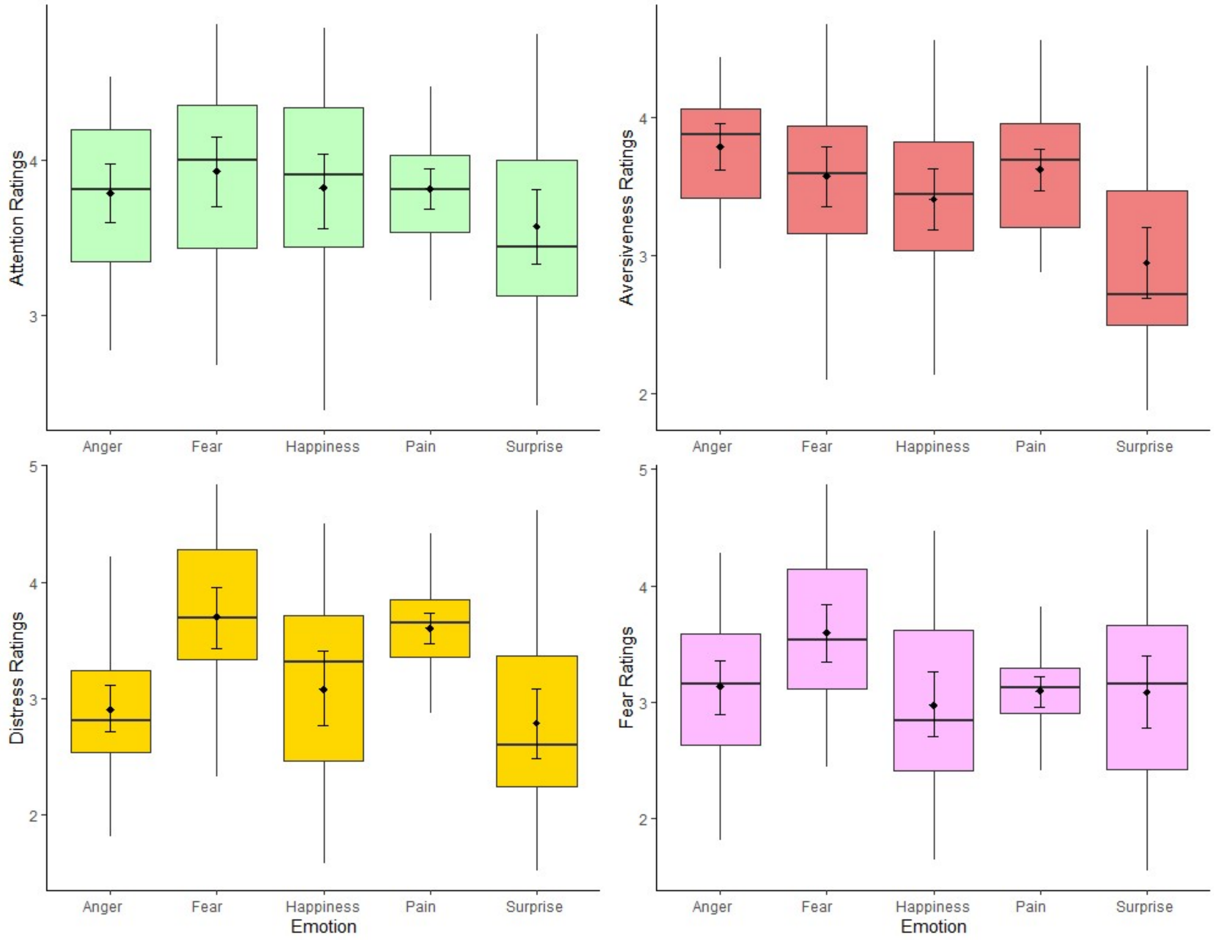


Figure 5. *Means and 95% confidence intervals of ratings on each scale as a function of stimulus emotion.*

Acoustic predictors of ratings. To examine the effects of acoustic parameters on listeners' perceptual ratings, four cumulative link mixed models were fitted with ratings on each scale as the outcome variable, the 8 acoustic parameters and Vocalizer Sex as fixed effects, and Participant and Stimulus as random effects. The effects of each acoustic parameter are shown in **Table 7.**

Acoustic effects on every scale were similar (unsurprisingly, given positive correlations between the scales), with differences only emerging in significance and not in the direction of effects. Stimuli with longer durations, higher F0s, less jitter, and earlier max F0s tended to elicit higher ratings in general. However, some acoustic effects specific to certain perceptual scales are observable, such as noisier stimuli being perceived as more aversive and fear-inducing, and stimuli with more chaos or subharmonics being perceived as more aversive and attention-getting.

Table 7. Acoustic predictors of perceptual effects

	Attention	Aversive	Distress	Fear-inducing
Duration	**1.785** **(0.197)**	**1.969** **(0.188)**	**1.103** **(0.264)**	**1.278** **(0.249)**
Mean F0	**3.508** **(0.46)**	**0.992** **(0.392)**	0.324 (0.571)	**3.374** **(0.642)**
Jitter	-0.214 (0.149)	**-0.459** **(0.140)***	**-1.109** **(0.198)**	**-0.57** **(0.201)***
Mean HNR	-0.028 (0.02)*	**-0.051** **(0.019)**	0.001 (0.026)*	**-0.079** **(0.025)**
Time Max F0	**-0.74** **(0.302)***	-0.434 (0.271)	**-1.468** **(0.41)***	-0.663 (0.372)
IQRBW	-0.187 (0.225)*	-0.226 (0.201)*	**-0.691** **(0.311)**	-0.46 (0.283)*
Pitch Jump	0.141 (0.185)	0.124 (0.167)	**0.589** **(0.261)***	0.102 (0.237)
NLP	**0.506** **(0.178)***	**0.961** **(0.162)***	0.373 (0.256)	0.439 (0.227)
Vocalizer Sex	**0.471** **(0.187)**	0.241 (0.168)	0.205 (0.227)	**0.723** **(0.239)**

Note. Effects in bold were significant at $p < .05$. Asterisks indicate effects that differed in significance when analyses were limited to the subset of vocalizations classified as screams.

Listener-classified screams

We repeated every analysis on the subset of stimuli that the majority of listeners classified as screams, comprising a more conservative, listener-defined category of vocalization. Notable results are briefly summarized below.

Emotion categorization. In contrast to their accuracy on all nonlinguistic stimuli, participants' percent corrects for stimuli classified as screams (*M* = .44, *SE* = .01) was significantly lower compared to linguistic stimuli ($t(96) = 5.618$, $p < .001$). Interestingly, however, the relatively low accuracy for fear stimuli seen among the larger set of nonlinguistic stimuli was not repeated in this subset: here, fear was significantly more accurately categorized than happiness ($p = .001$) and not significantly lower than pain ($p = 1$) or surprise ($p = .246$). This slightly enhanced accuracy for fear seemed driven by an increase in percent correct, and not a decrease in false alarms. In fact, the false alarms for this subset largely resembled those for all nonlinguistic stimuli, except that false alarms for fear increased relative to other emotions ($p < .001$ for every comparison except pain).

Ratings. When analyses of perceptual ratings were conducted only on the stimuli classified as screams, Emotion no longer had a significant effect on ratings of Aversiveness (likelihood-ratio test, $p = .172$) or Fear ($p = .059$) as well as Attention ($p =.749$). Thus, only ratings of Distress still varied with emotion, with vocalizers expressing fear again perceived as more distressed than those expressing anger ($p = .013$), happiness ($p = .019$), or surprise ($p = .024$). It is unclear to what extent this discrepancy from the larger nonlinguistic set is due to reduced power, or because among our more conservatively defined screams, there is less functional variation related to aversiveness or fear induction. One possibility is that variation on these scales is more associated with listeners' classification of vocalizations as screams, i.e., between-type variation, than with variation among screams. Supporting this idea, vocalizations classified as screams on average received higher ratings on every rating scale than vocalizations not classified as screams (**Table 8**).

Table 8. Mean ratings of screams vs. non-screams

	Classified as screams		Not classified as screams			
Rating dimension	*Mean*	*S.E.*	*Mean*	*S.E.*	*t*	*p*
Attention	4.13	0.06	3.47	0.06	-8.29	<.001
Aversiveness	3.70	0.06	3.26	0.08	-4.36	<.001
Distress	3.58	0.09	2.89	0.08	-5.79	<.001
Fear	3.58	0.08	2.82	0.06	-7.70	<.001

Acoustics. Generally, both the real acoustic differences between emotions and the effects of acoustic parameters on responses resembled those reported above for the larger nonlinguistic stimulus set, with the exception that some parameters were no longer significant, noted with asterisks in tables 7, A2, and A3. Again, we cannot say to what extent these differences derive from reduced variation within stimuli perceived as screams versus reduced power due to fewer stimuli. Note, however, that the number of stimuli per emotion in this smaller subset was comparable to the number of total stimuli in some other studies (Lima et al., 2014).

Nonetheless, even among this somewhat more conservatively defined subset, the effects of F0- and noise-related parameters were largely maintained. Thus, anger and pain stimuli generally exhibited lower and flatter pitches but greater noisiness and/or roughness compared to happiness and surprise stimuli; anger stimuli also exhibited greater shimmer while happiness stimuli exhibited lower shimmer. Fear stimuli again exhibited a combination of a relatively high pitch but a lower pitch range. Finally, unique to this subset, happiness stimuli exhibited a wider bandwidth (i.e., greater IQRBW) relative to fear, pain, or surprise stimuli.

The effects of acoustics on listener categorizations were also largely similar to those reported for all nonlinguistic stimuli, with some parameters losing significance. For example, fear categorizations were still predicted by lower jitter and higher F0, but duration, HNR, and pitch jumps were no longer significant predictors. Likewise, in analyses of ratings, time to max

F0, pitch jumps, nonlinear phenomena, and sometimes jitter were no longer significant. However, duration, F0, and HNR continued to have significant effects on ratings in the same direction; in fact, HNR as well as IQRBW were predictive of ratings on more scales among this subset. Thus, at a minimum, variation in these parameters is perceptually relevant even among conservatively defined screams.

Discussion

We investigated whether screams vary acoustically and perceptually based on the emotion actors portrayed, and we directly compared the patterns of variation within this call type to a simple speech exemplar. We found that screams associated with different emotional states are moderately distinct acoustically, such that listeners can recognize the emotion conveyed with an accuracy roughly equivalent to their ability to do so from a short speech segment. We also found that the patterns of listeners' accurate judgments and errors between emotions, as well as the patterns of acoustic variation between emotions in screams largely mirrored those in speech, although we uncovered some potential differences between these vocalization types. We additionally found variation in perceptual ratings between emotionally distinct screams, such that they might elicit different responses from listeners. Finally, we repeated these analyses with a subset of screams classified as such by a simple majority of listeners, adding to the evidence that variation within this call type is informative and potentially significant to communication (Engelberg et al., 2019; Frühholz et al., 2021; Schwartz & Gouzoules, 2019).

Emotion categorization of screams and speech

Accuracy. In our emotion categorization task, we found no evidence that participants more accurately recognized emotions from screams versus speech or vice versa. These results contrast somewhat with the literature on nonlinguistic emotional expression, in which participants have often achieved higher accuracies with nonlinguistic vocalizations than those reported for speech prosody (Hawk et al., 2009; Pell et al., 2015; Sauter et al., 2013; Schröder, 2003). Those studies, however, imposed few if any constraints on the acoustic forms that vocalizers generated, allowing them to produce variable call types, as they themselves determined appropriate. It is likely that, when asked to produce a single call type which nonetheless occurs in varied emotional contexts, vocalizers' expressive capacities are limited in turn (even if they did not always produce sounds that the majority of our participants classified as screams), resulting in capacities of listener recognition more similar to those enabled by prosodic speech. Indeed, the recognition accuracies in our study resembled that reported by Szameitat et al. (2009), in which listeners categorized different types of laughter, i.e., judged the variation within a single call type. The question of why vocalizers might use the same call type to convey different emotions, despite higher communicative accuracy when employing multiple call types, is touched upon later in this discussion.

Listeners who recognized emotions accurately in one vocalization type tended to respond accurately to the other as well, potentially supporting the hypothesis that shared mechanisms underlie the perception of emotion-related variation in screams as in speech prosody. Further similarities emerged in the patterns by which listeners' accuracies varied with emotion. Across both vocalization types, listeners identified anger with the highest accuracy, pain and surprise with moderate accuracy, and happiness with relatively low accuracy. That the effects of emotion largely transcended vocal type would seem to support the notion that they are not, strictly speaking, "within type" at all and are treated by listeners somewhat independently of call type (Anikin et al., 2018), although we do explore some speech- and scream-specific effects below.

One notable divergence between the vocalization types was that when categorizing screams, listeners were equally poor at recognizing fear as happiness, whereas with speech, they categorized fear more accurately than happiness and pain, and equally as accurately as surprise. These data suggest a relative deficit at recognizing fear with precision from screams. This finding is perhaps somewhat surprising, not only because of the intuitive association between fear and screams, but also because recognizing fear would seem as important as recognizing anger; both contexts might imply immediate danger.

Note, however, that our measure of accuracy took into account false alarms as well as hit rates. One factor that likely contributed to listeners' lower accuracies for nonlinguistic fear was their relatively high false alarm rates for fear (as well as surprise and especially pain) that did not occur among the linguistic stimuli, where their false alarm rates were overall more diffuse across emotions. It is possible that, absent contextual cues to the contrary, listeners tend to bias their interpretations of screams towards certain emotions—including fear and pain, the identification of which is potentially important, and which represent potentially ancestral contexts for screams—at the expense of precision, which is perhaps not the most prominent function of communication via screams. (Indeed, as noted at the end of this discussion, absolute communicative accuracy can, in some situations, be detrimental.) Thus, for example, happiness screams were often misclassified as fear, replicating our finding from study 1 (and again contra the findings of Frühholz et al., 2021; see Study 1 Discussion), but the same pattern did not occur in speech. Additionally, when nonlinguistic fear was confused for another emotion, listeners tended to identify it as pain or surprise (possibly because many fear-related contexts do involve a component of surprise as well), but notably *not* as happiness, which would likely elicit a very different response. Overall, then, our data suggest that anger screams are well-recognized, whereas fear and pain screams seemed to produce interpretations that could enable appropriate responses from listeners, even if those interpretations are not quite as precise and may require added contextual cues to disambiguate.

In sum, these findings provide further evidence that screams are, to some degree, informative about the caller's emotional state. Although they did not enable perfect judgments about emotion by any means, listeners' accuracies were comparable to their accuracies attained from a short, semantically neutral speech sample conveying the same emotions. We must, however, emphasize some important limitations regarding the speech sample used in this study. First, in natural contexts, speech is almost always longer and can carry concurrent semantic information about emotion. We therefore do not claim that the communication of emotion in screams is equal to that in naturalistic speech, only that the variation within this call type is perhaps not inherently less informative than the prosodic variation in a speech segment of similar duration, and that the perception of emotion seems to transcend call type.

Additionally, it could be argued that the use of a name still might activate potential associations for listeners in a way that, e.g., a nonsense word or a word in an unfamiliar language might not. We used a name so as not to impose added artificiality to the instructions provided to our English-speaking actors. However, an interesting extension of this study would involve comparing the acoustic variation among screams (and/or other call types) to a greater variety of speech samples, potentially including samples produced by native speakers of a non-English language.

A final caveat, which also applies to the previously cited studies on vocal emotional expression (Hawk et al., 2009; Pell et al., 2015; Sauter et al., 2013; Schröder, 2003), is that our studies relied on acted portrayals rather than spontaneous expressions of emotion. We have shown that (at least some) actors are capable of producing screams that listeners cannot reliably distinguish from acted screams (Engelberg & Gouzoules, 2019), but it remains possible that the artificiality of the recording context would alter the experience, expression, and perception of different emotions among screams as well as speech. Further research can complement this work by exploring the accurate recognition of emotions from naturally occurring screams.

Acoustics. Screams and speech exhibited very similar patterns of acoustic variation by emotion. Of the fourteen acoustic parameters we examined, only five varied by emotion in significantly different ways between screams and speech, and in none of these cases did the effects of emotion on that parameter differ directionally between vocalization types. These findings therefore strongly support the idea that shared mechanisms underlie emotion-based variation across screams and speech.

Patterns of acoustic variation in both vocalization types were largely consistent with those reported in other studies of emotional vocal production. Anger and pain were characterized by similar acoustic profiles, corresponding to a low, flat pitch (i.e., low F0 and F0 range) and a harsh, noisy percept (i.e., low HNR, high roughness, and frequent incidence of nonlinear phenomena). The same acoustic characteristics are commonly reported for expressions of anger, especially in nonlinguistic vocalizations (Lima et al., 2013), as well as in nonhuman animal calls associated with aggression (Morton, 1977). Similarly, pain is often associated with roughness, noise, and nonlinearities (e.g., in infant cries: Koutseff et al., 2018), but some studies have found an increase in pitch with pain, rather than a decrease (Bellieni et al., 2004; Raine et al., 2018). However, those studies examined the acoustic variation between different levels of pain. It is feasible that, as a whole, pain is associated with lower-pitched calls relative to other emotions, whereas differences in pain intensity correlate positively with pitch within a lower frequency bandwidth.

In contrast to anger and pain, both happiness and surprise were marked by high, fluctuating pitches. Happiness stimuli exhibited higher frequency elements in general (including F0 and peak frequency), a wide pitch range, and greater tonality, all features consistent with prior findings (Lima et al., 2013; Scherer, 2003). Of note, Szameitat et al. (2011) reported that among laughs, tonality distinguished joyful calls from other laugh variants, suggesting that relative tonality versus noisiness might play a consistent role in marking positive valence within call types (Anikin et al., 2018; see also section on screams below). In the case of surprise, the wide pitch range reflected a dramatic increase in F0 over a short period of time, suggesting that pitch contour (measures of which are sometimes overlooked in studies of emotional expression) is an important cue for distinguishing this emotion. Finally, fear stimuli exhibited moderately high F0s (resembling happiness and surprise), yet relatively narrow F0 ranges (resembling anger and pain). The combination suggests a relatively high-pitched, sustained vocalization, in line with descriptions not only of fearful speech (Juslin & Laukka, 2003) and nonlinguistic calls (Lima et al., 2013), but also certain nonhuman distress calls (Lingle et al., 2012).

We observed differences between vocalization types in the effects of emotion on five parameters: mean HNR, roughness, nonlinear phenomena, duration, and time to max F0, although it is important to reiterate that none of these differences were directional, and thus that the following interpretations are made with caution. For the first three of these parameters, the general tendency was that anger was more differentiated from other emotions among screams than in speech, such that nonlinguistic anger stimuli were even noisier, rougher, and more reflective of a nonlinear vocal regime than linguistic anger. Given that HNR and roughness were highly correlated (*r* = -.821) and that our count of nonlinear phenomena included chaotic noise, it is likely that all of these differences reflect the same (or related) effects, where the noisiness and roughness of anger relative to other emotions is exaggerated in the case of screams. These characteristics contribute to the perceived harshness of a vocalization (Anikin, 2020), and it is possible that social conventions and/or intelligibility-related constraints impose an upper limit on the harshness of many spoken utterances (e.g., as proposed for roughness: Arnal et al., 2015), limitations conceivably less in effect in the nonlinguistic domain. Several of our nonlinguistic vocalizations, on the other hand, may have approached what some other researchers have deemed 'roars', calls typified by their low pitch and noisy quality (Anikin & Persson, 2017), but it is important to note that these effects generally held even for the subset of vocalizations that listeners classified as screams. Differences between speech and screams in the other two parameters, duration and time to max F0, appeared to be more subtle; for example, happiness and pain speech stimuli were longer relative to the other linguistic utterances, whereas duration was overall less affected by emotion in screams. It may be relevant that parameters related to the duration of voicing and the contour of F0 are regularly and intentionally modulated in the course of speech (Pisanski et al., 2016); conceivably, vocalizers exert more control over these elements in speech than in screams.

Listeners' use of acoustic cues to judge emotions often aligned with the real acoustic differences between them, presumably facilitating their accurate judgments, and showed many similarities between speech and screams. For example, across both vocalization types, lower F0 predicted anger categorizations (as well as pain, among nonlinguistic vocalizations) while higher F0 predicted surprise categorizations (as well as fear, among nonlinguistic vocalizations), all judgments consistent with the differences in F0 between these emotions. Moreover, certain effects suggested that listeners might have been sensitive to some of the interactions between emotion and vocalization type: for example, duration was predictive of happiness categorizations of linguistic but not nonlinguistic stimuli, consistent with duration emerging as a discriminative cue for happiness in speech but not screams. Nonetheless, there were several (almost all non-directional) differences in significant predictors of categorizations between speech and screams that are not explicable on this basis. In addition to HNR and duration, jitter seemed an important parameter especially for judging screams, perhaps more so than for speech. However, not only was jitter not more discriminative of emotions in screams than in speech: it was not useful for discriminating emotions in general. More research will be needed to replicate these findings, particularly using different speech samples.

It should be noted that although we selected a set of acoustic parameters that we hypothesized were relevant to emotion-related variation in screams and speech, we almost certainly did not capture the full scope of significant acoustic variation in these vocalizations. One likely important area of variation relates to the time-varying information across an utterance, which in this study we only examined via one limited measure (time to peak F0). Such information in speech prosody has proven well-diagnostic of specific emotions and perceptually salient to listeners (Meyer, Steinhauer et al., 2004; Rao, Koolagudi, et al. 2013). Given that time to max F0 significantly differed between screams associated with different emotions, this kind of dynamic variation requires further exploration in screams.

Perceptual effects of screams

Perceptual ratings by emotion. Listeners’ ratings on several perceptual scales varied according to the emotion actors were asked to portray in their screams. Most notably, fear and pain screams were perceived as more distressed and likely to elicit aid compared to screams associated with other emotions, an effect likely shared with screams in nonhuman primates (Cheney, 1977; de Waal & van Hooff, 1981; Gouzoules et al., 1984; Gouzoules & Gouzoules, 1989). Arguably, these emotions were the two among our categories for which aid is the most plausibly beneficial response for vocalizers, and by extension, for cooperative listeners. Listeners' interpretations were therefore appropriate, suggesting that acoustic variation between screams from different emotional contexts is not only informative, but also potentially functionally relevant. Notably, listeners' ratings of the vocalizer's distress and of their own willingness to provide aid were highly correlated, suggesting that cases where rare in which they perceived distress but would not intervene. That said, in natural scenarios, it seems likely that additional informational and contextual cues play into listeners' decisions to help distressed vocalizers, as they do in nonhuman primates (Cheney, 1977; Gouzoules & Gouzoules, 1989).

Fear screams were also more likely to induce self-reported fear in listeners compared to other screams. Speculatively, one possible mechanism by which fear screams could promote aid is through emotional contagion (Briefer, 2018), which could facilitate the listener's recognition of emotion in vocalizers (by self-referencing their own emotional state) and/or motivation to help. Such an account could partly explain some cases of aid in nonhuman animals (Preston & de Waal, 2002), although, again, additional processes such as individual recognition and cognitive decision-making are likely also involved in nonhuman primate agonistic aiding. Emotional contagion may also constitute a more general mechanism involved in emotional communication in humans. Certain types of laughter, for example, appear to affect listeners by inducing positive states in them (Bachorowski & Owren, 2001).

In contrast, we did not find significant variation by emotion in ratings of screams' likeliness to capture attention. Directing listeners' attention towards the vocalizer might represent a general perceptual characteristic of screams; that is, they are acoustically suited for capturing attention regardless of the more specific emotional context in which they occur. This finding sheds some light on the mechanisms underlying the differences between emotions described above: it suggests, for example, that fear screams did not induce more fear and more tendencies towards aid only because they were more attention-getting than other screams.

Similarly, screams associated with different emotions were rated as equally aversive, with the exception of surprise screams, potentially because those screams are relatively short and tonal, two characteristics that did not predict aversiveness (see acoustic predictors of ratings, below). Notably, some authors have proposed that, other than soliciting aid, warding off aggressors via aversive acoustic characteristics is an important function of nonhuman primate screams (Owren & Rendall, 2001). Importantly, however, there remains no empirical evidence for this hypothesis in nonhuman primates or other taxa (Gouzoules, 2005), and our evidence here shows only that screams are perceived as aversive, not that they are necessarily aversive enough to repel opponents. More data would be required in both humans and nonhuman primates to support the possibility that the repelling of opponents is a shared function of their screams.

Acoustic predictors of ratings. Strong correlations between mean ratings on each scale suggested that listeners based their ratings on partially shared acoustic cues. Our analysis of the acoustic predictors of ratings was consistent with this prediction: acoustic variation seemed to influence ratings on each scale similarly, with differences primarily evident in the relative weighting of some parameters versus others. Notably, the general trend that longer, higher pitched, and noisier screams received higher ratings accords with widely reported effects of autonomic arousal on vocal production in humans and nonhuman animals, as well as potential effects of negative valence (Briefer, 2012). Similar acoustic effects are also seen in other human call types. For example, in infant cries, the same acoustic characteristics have correlated not only with listeners' judgments of qualities such as distress and aversiveness (LaGasse et al., 2005; Zeskind & Marshall, 1988) but also with the severity of the procedure eliciting the cry, directly implicating autonomic arousal as a source of the variation (Porter et al., 1986). It is conceivable that, in our study, the perception of high arousal (and possibly negative valence) was a common denominator generally underlying the ratings of screams as attention-getting, aversive, distressed, and fear-inducing.

Nonetheless, that some acoustic effects only emerged as significant in certain rating scales suggests that perceptual judgments were not based on an identical set of cues. Differences in the relative importance of certain parameters may hint at the functional relevance of acoustic variation. For example, the presence of nonlinear phenomena significantly influenced ratings of only attention and aversiveness. One hypothesis regarding the function of nonlinearities such as deterministic chaos in animal calls, which has been applied to nonhuman primate screams specifically, is that they render calls less repetitively similar and therefore harder to habituate to or ignore (e.g., due to sudden transitions between vibratory regimes of the vocal folds; Fitch et al., 2002). The effects of nonlinear phenomena here on ratings of attention are consistent with this idea. The occurrence and timing of nonlinear phenomena might also provide cues to caller identity (Fitch et al., *ibid.*), a likely important consideration for listeners' responses to attention-getting calls. Likewise, it has been suggested that nonlinear phenomena may contribute to the aversiveness of a call (Bachorowksi et al., 2001; Owren & Rendall, 2003); most recently, Anikin (2020) demonstrated a causal link between nonlinear phenomena and judgments of aversive states. Note that this effect in our study is likely not specific to noise produced by nonlinearities—greater noise in general yielded higher ratings of aversiveness—but nonlinearities may provide an added source of aversive noise.

We emphasize that our assessments of scream effects here relied on listeners' ratings. Future studies could more closely approximate the potential functional effects of screams using direct behavioral measures in more naturalistic scenarios as well as physiological measures (e.g., of listeners' stress and/or arousal). Even these approaches, however, would not offer a proximate causal account for the acoustic variation among screams from different contexts. We do not know the extent to which the acoustic variation among screams typically emerges because of automatic effects of emotion on vocal production, intentional modulation on the part of signalers, learning of conventionalized emotion-correlated patterns, and other causal mechanisms.

Nonetheless, the salient point is that whatever their proximate causes, acoustic features common to screams might enable common effects, whereas acoustic features specific to subtypes might enable effects befitting more specific situations (Schamberg et al., 2018). For example, if the control of listeners' attentional states represents a general effect of screams, then this common goal could partly explain the preservation of attention-getting characteristics in screams occurring across seemingly disparate emotional contexts. That is, although screams associated with happiness and those associated with fear derive from very different emotional states, they may share the goal of eliciting attention, resulting in similar acoustic profiles that are well-suited to induce this effect on receivers. More specific effects such as the induction of fear or solicitation of aid may account for some acoustic variation within the type and correlate with variation in emotional context. Thus, some acoustic features distinct to, for example, fear-associated screams may facilitate the fulfillment of certain receiver actions that are not common to screams occurring in anger- or pain-associated contexts.

Screams as a call type

Human call types are not purely discrete, but instead grade acoustically and perceptually into other call types—as do call types in nonhuman apes and other species (Anikin & Persson, 2017; Marler, 1976; Schwartz et al., 2020a). As discussed in the General Introduction, this does not imply that call types are not biologically real, or that no acoustic or perceptual clusters exist, but the determination of their boundaries remains a complex subject of ongoing empirical work based on a combination of acoustic and perceptual data. Screams, however, are almost certainly a basic call type in humans. They occur and are recognized across cultures (Anikin et al., 2018; Sauter et al., 2010), they are readily identified by listeners (Schwartz et al., 2020a), and their origin is likely homologous with acoustically similar screams in nonhuman primates.

Here, for analysis purposes, we started by considering every nonlinguistic stimulus in our study a scream, or at least scream-like, because we had specifically asked our actors to produce screams. However, our additional data on listeners' classifications of screams allowed us to repeat analyses, but restrict them to stimuli that were perceived by at least half of participants as screams, thereby asking whether acoustic and perceptual variation exists even among the subset representing relatively more recognizable screams. We note the caveat that this perceptual criterion for screams was necessarily still somewhat lenient, a point expounded upon at the end of this chapter.

Among this subset, overall categorization accuracy fell such that it was now slightly lower than listeners' accuracy on speech, possibly demonstrating the reduced expressive capacities of variation within a call type. Interestingly, although differences in accuracy by emotion resembled those from the larger stimulus set, with anger still the highest and happiness the lowest, fear saw a relative enhancement, such that it was no longer among the most poorly categorized emotions. This was accompanied by an increase in false alarms for fear. In other words, listeners were more likely in general to categorize screams from this subset as fear, but they somewhat compensated for this response bias by more often identifying fear correctly. It is possible that the relative imbalance towards fear stimuli among this subset (*n* = 18, or 28% of the subset) contributed to this tendency, but notably, the same trend did not occur in happiness stimuli which were nearly as heavily represented (*n* = 17, or 27% of the subset). Instead, pending further research, we suggest that both listeners' biases towards fear and their enhanced ability to detect it when expressed point to an important communicative role for screams: Namely, that when fear is conveyed, listeners should interpret it as such. One might also speculate that fear constitutes the ancestral emotional condition for screams, and thus that listeners are especially biased to perceive fear in screams, although this account would also require further evidence.

Both our analysis of perceptual ratings and of acoustics suggested reduced variation among these screams, but nonetheless, some variation remained. Screams no longer significantly varied in ratings by emotion except that fear screams conveyed more distress than screams associated with other emotions. However, screams from this subset received higher ratings on every scale than stimuli not perceived as screams by most listeners. Thus, these data support the idea that each of these perceptual characteristics is fundamental to screams: As calls become more recognizable as screams, they tend to become more attention-getting, aversive, distressed, and fear-inducing. If screams in general are likely to share these characteristics, then there may exist less space for perceptual variation between emotional contexts, although as described below, some relevant acoustic variation remains within this subset (again, with the caveat that this subset was defined somewhat leniently).

Acoustically, many of the differences in pitch, pitch contour, and tonality involved in distinguishing anger and pain from happiness and surprise remained, as did some effects of emotion on roughness and shimmer. Likewise, some but not all effects of duration, pitch, and tonality retained significance in predicting judgments of emotion category and perceptual characteristics; if anything, HNR as well as interquartile range—both features that can influence perceived tonality—played a greater role in predicting perceptual ratings than it did in the larger set. Anikin and Persson (2017) suggested that the relative tonality versus noisiness of a call might be especially meaningful within call types, as opposed to differentiating between call types. Our finding regarding the importance of tonality in more conservatively defined screams compared to the broader nonlinguistic set may offer partial support for this idea. That said, the continued significance of pitch-related parameters, which we previously reported were important for distinguishing screams from other call types (Schwartz et al., 2020a), suggests that the parameters significant to variation between types and within types can be one and the same.

In summary, although variation was perhaps reduced among our restricted subset of more recognizable screams, these screams nonetheless varied acoustically in ways that enabled judgments about emotion and perceptual characteristics to vary as well. Note that we designed our study chiefly to compare the broader set of nonlinguistic stimuli to matched linguistic stimuli, and not to compare more conservatively defined screams to a larger nonlinguistic repertoire. The analyses described in this section necessarily involved fewer stimuli than those conducted on the full nonlinguistic set, and it is therefore difficult at this time to attribute differences in parameter significance to actual reduced variation rather than the reduced power of statistical tests. However, we suggest that the main takeaway here is that *in spite of* reduced power, and in spite of our adoption of a somewhat more conservative, listener-determined criterion for screams, some variation significant to communication exists within this call type.

We conclude with some notes relevant to the interpretation of our findings. First, it is important to emphasize that our operationalization of screams based on a simple majority of listeners' classifications was not based on any established empirical definition. Indeed, a consensus definition does not exist for screams, although we have contributed perceptual and acoustic data towards establishing one (Schwartz et al., 2020a). Instead, here we selected the 50% cutoff as an intuitive criterion that preserved adequate representation of each emotion. We asked actors to produce screams, but we chose not to influence their portrayals by asking them to reproduce vocalizations that we did not consider screams, instead leaving that determination to listeners. This procedure resulted in a large number of stimuli that the majority of listeners did not classify as screams, yielding a modest but still analytically viable subset of exemplars by our 50% criterion, but making stricter cutoffs less viable for analyses. Had we enough stimuli to adopt a stricter criterion, it is likely some of the specific perceptual and acoustic results described in this section may have changed. Incidentally, that over half of our stimuli were not classified as screams by the majority of listeners even though actors were asked to produce screams might bear important implications for future research on human call types. Researchers should exercise caution before assuming their stimuli belong to a given type, particularly when relying on actors whose abilities and interpretations might vary (Engelberg & Gouzoules, 2019).

Finally, although we have laid out an account referring to constraints on expressiveness within a call type, we must reiterate that complete expressiveness or accuracy is not necessarily the most adaptive condition in communication. It may not benefit signalers to always reveal precise information about their emotional states; rather, ambiguity or obfuscation is often advantageous. It is possible, then, that expressiveness in screams is limited not only because their acoustic features are constrained to achieve certain perceptual effects, but also because (at least some) screams are *not* designed to convey emotions clearly. I expand on this point in the General Discussion.

Acknowledgements

We thank Shiyin Liu, Caroline Wei, and Kyra Mitchell for their assistance with data collection. We also thank the nine actors who generously volunteered their skills and time to produce the stimuli for this study.

References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences, 37*(6), 529.

Allison, P. D. (2008, March). Convergence failures in logistic regression. In *SAS Global Forum* (Vol. 360, pp. 1-11).

Anikin, A. (2019). Soundgen: an open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods, 51*(2), 778-792.

Anikin, A. (2020). The perceptual effects of manipulating nonlinear phenomena in synthetic nonverbal vocalizations. *Bioacoustics, 29*(2), 226-247.

Anikin, A., & Persson, T. (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods, 49*(2), 758-771.

Anikin, A., Bååth, R., & Persson, T. (2018). Human Non-linguistic Vocal Repertoire: Call Types and Their Meaning. *Journal of Nonverbal Behavior, 42*(1), 53-80.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology, 25*(15), 2051-2056.

Bachorowski, J. A., & Owren, M. J. (2001). Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science, 12*(3), 252-257.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614-636.

Bänziger, T., & Scherer, K. R. (2007, September). Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. In *International conference on affective computing and intelligent interaction* (pp. 476-487). Springer, Berlin, Heidelberg.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bellieni, C. V., Sisto, R., Cordelli, D. M., & Buonocore, G. (2004). Cry features reflect pain intensity in term newborns: an alarm threshold. *Pediatric Research, 55*(1), 142-146.

Belin, P., & Zatorre, R. J. (2015). Neurobiology: Sounding the Alarm. *Current Biology, 25*(18), R805-R806.

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*(2), 531-539.

Blumstein, D. T., & Chi, Y. Y. (2012). Scared and less noisy: glucocorticoids are associated with alarm call entropy. *Biology Letters, 8*(2), 189-192.

Blumstein, D. T., Richardson, D. T., Cooley, L., Winternitz, J., & Daniel, J. C. (2008). The structure, meaning and function of yellow-bellied marmot pup screams. *Animal Behaviour, 76*(3), 1055-1064.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer (*Version 5.3.51). Retrieved January 9, 2018 from http://www.praat .org/.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology, 288*(1), 1-20.

Briefer, E. F. (2018). Vocal contagion of emotions in non-human animals. *Proceedings of the Royal Society B: Biological Sciences, 285*(1873), 20172783.

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM (2017). "glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling." *The R Journal, 9*(2), 378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html.

Center for Conservation Bioacoustics. (2014). Raven Pro: Interactive Sound Analysis Software (Version 1.5) [Computer software]. Ithaca, NY: The Cornell Lab of Ornithology. Available from http://ravensoundsoftware.com/

Chen, S. T., Xiao, L., & Staicu, A. M. (2019). An Approximate Restricted Likelihood Ratio Test for Variance Components in Generalized Linear Mixed Models. arXiv preprint arXiv:1906.03320.

Cheney, D. L. (1977). The acquisition of rank and the development of reciprocal alliances among free-ranging immature baboons. *Behavioral Ecology and Sociobiology, 2*(3), 303-318.

Christensen, R. (2019). ordinal—Regression Models for Ordinal Data. R package version 2019.12-10. Retrieved from https://CRAN.R-project.org/package=ordinal.

Chu, M. (2001). Heterospecific responses to scream calls and vocal mimicry by phainopeplas (Phainopepla nitens) in distress. *Behaviour, 138*(6), 775-787.

Darwin, C., & Prodger, P. (1998). *The expression of the emotions in man and animals.* Oxford University Press, USA.

de Waal, F. B. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology, 59*, 279-300.

de Waal, F. B., & van Hooff, J. A. (1981). Side-directed communication and agonistic interactions in chimpanzees. *Behaviour, 77*(3), 164-198.

Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2017). Skin temperature changes in wild chimpanzees upon hearing vocalizations of conspecifics. *Royal Society Open Science, 4*(1), 160816.

Dietrich, S., Ackermann, H., Szameitat, D. P., & Alter, K. (2006). Psychoacoustic studies on the processing of vocal interjections: how to disentangle lexical and prosodic information? *Progress in Brain Research, 156,* 295-302.

Driver, P. M., & Humphries, D. A. (1969). The significance of the high-intensity alarm call in captured passerines. *Ibis, 111*(2), 243-244.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin, 128*(2), 203.

Elfenbein, H. A., Mandal, M. K., Ambady, N., Harizuka, S., & Kumar, S. (2002). Cross-cultural patterns in emotion recognition: highlighting design and analytical techniques. *Emotion, 2*(1), 75.

Engelberg, J. W., & Gouzoules, H. (2019). The credibility of acted screams: Implications for emotional communication research. *Quarterly Journal of Experimental Psychology, 72*(8), 1889-1902.

Engelberg, J. W., Schwartz, J. W., & Gouzoules, H. (2019). Do human screams permit individual recognition?. *PeerJ, 7*, e7087.

Engqvist, L. (2005). The mistreatment of covariate interaction terms in linear model analyses of behavioural and evolutionary ecology studies. *Animal Behaviour, 70*(4).

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage, 36*(2), 480-487.

Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Frontiers in Psychology, 7,* 1393.

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour, 63*(3), 407e418.https://doi.org/10.1006/anbe.2001.1912

Fridlund, A. J. (1997). The new ethology of human facial expressions. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The Psychology of Facial Expression* (pp. 103-132). Cambridge University Press.

Frühholz, S., Dietziker, J., Staib, M., & Trost, W. (2021). Neurocognitive processing efficiency for discriminating human non-alarm rather than alarm scream calls. *PLoS biology, 19*(4), e3000751.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science, 25*(4), 911-920.

Gouzoules, H., & Gouzoules, S. (1989). Design features and developmental modification of pigtail macaque, Macaca nemestrina, agonistic screams. *Animal Behaviour, 37*, 383-401.

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (Macaca mulatta) screams: representational signalling in the recruitment of agonistic aid. *Animal Behaviour, 32*(1), 182-193.

Gouzoules, H. (2005). Monkeying Around With Symbolism: Are Vocalizations Simple Symbols...or More Like Cymbals? In L. L. Namy (Ed.), Emory symposia in cognition. Symbol use and symbolic representation: Developmental and comparative perspectives (p. 245–265). Lawrence Erlbaum Associates Publishers.

Groulx, B. (2005). Screaming and wailing in dementia patients (part II). *The Canadian Alzheimer Disease Review, 8*(1), 7-11.

Gruber, T., & Grandjean, D. (2017). A comparative neurological approach to emotional expressions in primate vocalizations. *Neuroscience & Biobehavioral Reviews, 73*, 182-190.

Hawk, S., van Kleef, G., Fischer, A., & van der Schalk, J. (2009). "Worth a thousand words": absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion, 9*(3), 293.

Hogstedt, G. (1983). Adaptation unto death: function of fear screams. *The American Naturalist, 121*(4), 562-570.

Hunter, E. M., Phillips, L. H., & MacPherson, S. E. (2010). Effects of age on cross-modal emotion perception. *Psychology and Aging, 25*(4), 779.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language, 59*(4), 434-446.

Jovanovic, T., & Gouzoules, H. (2001). Effects of nonmaternal restraint on the vocalizations of infant rhesus monkeys (Macaca mulatta). *American Journal of Primatology, 53*(1), 33-45.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 65-135). Oxford, UK: Oxford University Press.

Kipper, S., & Todt, D. (2003). The role of rhythm and pitch in the evaluation of human laughter. *Journal of Nonverbal Behavior, 27*(4), 255-272.

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.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

LaGasse, L. L., Neal, A. R., & Lester, B. M. (2005). Assessment of infant cry: acoustic cry analysis and parental perception. *Mental retardation and developmental disabilities research reviews, 11*(1), 83-93.

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Iraki, F. K. E., ... & Thingujam, N. S. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology, 4*, 353.

Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., ... & Gur, R. (2010). " It's not what you say, but how you say it": a reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience, 4*, 19.

Lehoczki, F., Szenczi, P., Bánszegi, O., Lakatos, K., & Faragó, T. (2020). Cross-species effect of separation calls: family dogs' reactions to pup, baby, kitten and artificial sounds. *Animal Behaviour*, *168*, 169-185.

Lenth, R. V. (2016). Least-Squares Means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Lesnoff, M., & Lancelot, R. (2012). aod: Analysis of Overdispersed Data. R package version 1.3.1. Retrieved from http://cran.r-project.org/package=aod

Li, T., Horta, M., Mascaro, J. S., Bijanki, K., Arnal, L. H., Adams, M., ... & Rilling, J. K. (2018). Explaining individual variation in paternal brain responses to infant cries. *Physiology & Behavior, 193*, 43-54.

Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (2014). In the ear of the beholder: how age shapes emotion processing in nonverbal vocalizations. *Emotion*, *14*(1), 145.

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: a corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods, 45*(4), 1234-1245.

Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., & Romanow, C. A. (2012). What makes a cry a cry? A review of infant distress vocalizations. *Current Zoology, 58*(5), 698-726.

Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2020). performance: Assessment of Regression Models Performance. CRAN. doi: 10.5281/zenodo.3952174, R package, https://easystats.github.io/performance/.

Marler, P. (1976). Social organization, communication and graded signals: the chimpanzee and the gorilla. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing Points in Ethology* (pp. 239-280). Cambridge University Press.

Mercier, S., Déaux, E. C., van de Waal, E., Bono, A. E., & Zuberbühler, K. (2019). Correlates of social role and conflict severity in wild vervet monkey agonistic screams. *PloS one, 14*(5), e0214640.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and language, 89*(2), 277-289.

Møller, A. P., & Nielsen, J. T. (2010). Fear screams and adaptation to avoid imminent death: effects of genetic variation and predation. *Ethology Ecology & Evolution, 22*(2), 183-202.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist, 111*(981), 855-869.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112*(1), 181-186.

Owren, M. J. (2008). GSU Praat tools: Scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods, 40*(3), 822–829. https ://doi.org/10.3758/BRM.40.3.822.

Owren, M. J., & Bachorowski, J. A. (2003). Reconsidering the evolution of nonlinguistic communication: The case of laughter. *Journal of Nonverbal Behavior, 27*(3), 183-200.

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews, 10*(2), 58-71.

Owren, M. J., & Rendall, D. (2003). Salience of caller identity in rhesus monkey (Macaca mulatta) coos and screams: perceptual experiments with human (Homo sapiens) listeners. *Journal of Comparative Psychology, 117*(4), 380.

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology, 73*(6), 530-544.

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology, 87*(1), 93-98.

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology, 111*, 14-25.

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*(4), 304-318.

Porter, F. L., Miller, R. H., & Marshall, R. E. (1986). Neonatal pain cries: effect of circumcision on acoustic features and perceived urgency. *Child Development,* 790-802.

Preston, S. D., & de Waal, F. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences, 25,* 1-72.

Provine, R. R. (1996). Laughter. *American scientist, 84*(1), 38-45.

Rendall, D., Owren, M., & Ryan, M. (2009). What do animal signals mean? *Animal Behaviour, 78*(2), 233-240.

Raine, J., Pisanski, K., Oleszkiewicz, A., Simner, J., & Reby, D. (2018). Human listeners can accurately judge strength and height relative to self from aggressive roars and speech. *iScience, 4*, 273-280.

Rao, K. S., Koolagudi, S. G., & Vempada, R. R. (2013). Emotion recognition from speech using global and local prosodic features. *International journal of speech technology, 16*(2), 143-160.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience, 11*(4), 426-428.

Rohwer, S., Fretwell, S., & Tuckfield, R. (1976). Distress screams as a measure of kinship in birds. *American Midland Naturalist,* 418-430.

Russell, J., Bachorowski, J., & Fernández-Dols, J. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*(1), 329-349.

Sauter, D. A., & Fischer, A. H. (2018). Can perceivers recognise emotions from spontaneous expressions? *Cognition and Emotion, 32*(3), 504-515.

Sauter, D., & Scott, S. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion, 31*(3), 192-199.

Sauter, D., Eisner, F., Calder, A., & Scott, S. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology, 63*(11), 2251-2272.

Sauter, D., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology, 31*(1), 97-113.

Schamberg, I., Wittig, R. M., & Crockford, C. (2018). Call type signals caller goal: a new take on ultimate and proximate influences in vocal production. *Biological Reviews, 93*(4).

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143-165.

Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (Eds.), *Wiley handbooks of psychophysiology: Handbook of social psychophysiology* (p. 165–197). John Wiley & Sons.

Scherer, K. R. (1994). Affect bursts. *Emotions: Essays on Emotion Theory, 161*, 196.

Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1), 227-256.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication, 40*(1), 99-116.

Schwartz, J. W. (2020). Vocal emotion expression across contexts, vocalization types, and species: Implications for general processes of vocal evolution. (Doctoral dissertation). Retrieved from https://etd.library.emory.edu/concern/etds/jw827c636

Schwartz, J. W., & Gouzoules, H. (2019). Decoding human screams: perception of emotional arousal from pitch and duration. *Behaviour, 156*(13-14), 1283-1307.

Schwartz, J. W., Engelberg, J. W., & Gouzoules, H. (2020a). Was that a scream? Listener agreement and major distinguishing acoustic features. *Journal of Nonverbal Behavior*, 1-20.

Schwartz, J. W., Engelberg, J. W. M., & Gouzoules, H. (2020b). Evolving views on cognition in animal vocal communication: Contributions from scream research. *Animal Behavior and Cognition, 7*(2), 192-213.

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion, 9*(6), 838.

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist, 17*(2), 197-208.

Slocombe, K. E., & Zuberbühler, K. (2005). Agonistic screams in wild chimpanzees (Pan troglodytes schweinfurthii) vary as a function of social role. *Journal of Comparative Psychology, 119*(1), 67.

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences, 104*(43), 17228-17233.

Slocombe, K. E., Townsend, S. W., & Zuberbühler, K. (2009). Wild chimpanzees (Pan troglodytes schweinfurthii) distinguish between different scream types: evidence from a playback study. *Animal Cognition, 12*(3), 441-449.

Stefanski, R. A., & Falls, J. B. (1972). A study of distress calls of song, swamp, and white-throated sparrows (Aves: Fringillidae). I. Intraspecific responses and functions. *Canadian Journal of Zoology, 50*(12), 1501-1512.

Szameitat, D. P., Alter, K., Szameitat, A. J., Darwin, C. J., Wildgruber, D., Dietrich, S., & Sterr, A. (2009). Differentiation of emotions in laughter at the behavioral level. *Emotion, 9*(3), 397.

Szameitat, D. P., Darwin, C. J., Wildgruber, D., Alter, K., & Szameitat, A. J. (2011). Acoustic correlates of emotional dimensions in laughter: arousal, dominance, and valence. *Cognition and emotion, 25*(4), 599-611.

Venables W., & Ripley B. (2002). Modern applied statistics with S-PLUS. Springer.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of nonverbal behavior, 17*(1), 3-28.

Warren, J. E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J., ... & Scott, S. K. (2006). Positive emotions preferentially engage an auditory–motor "mirror" system. *Journal of Neuroscience, 26*(50), 13067-13075.

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011, May). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 143-146).

Zeskind, P. S., & Marshall, T. R. (1988). The relation between variations in pitch and maternal perceptions of infant crying. *Child Development*, 193-196.

Zimmermann, E., Leliveld, L., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices. In E. Altenmuller, S. Schmidt, & E. Zimmermann (Eds.), *The evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man* (pp. 116-132). Oxford, UK: Oxford University Press.

CHAPTER IV. General Discussion

The majority of research on nonlinguistic vocalizations has paid little attention conceptually or methodologically to call types, instead focusing on vocalizers' abilities to communicate emotion provided no constraints on the type of calls they produce. To some extent, this omission is unsurprising, as human call types (like those in other apes) are not purely discrete, instead grading acoustically into one another. Nonetheless, listeners readily classify some vocalizations, including screams, as identifiable call types (Anikin et al., 2018; Schwartz et al., 2020a). In the General Introduction, I argued that ignoring these call types risks neglecting important aspects of nonlinguistic communication, including the occurrence of call types across diverse contexts, the acoustic variation within types, and the functions of calls beyond conveying emotion. Important questions have persisted about *why* vocalizers might use the same call across contexts, and whether the variation among them conveys different meanings. These questions, although long deliberated and addressed in the animal communication literature (Smith, 1980, p. 184), have received relatively less attention in humans.

This dissertation presented two studies centered on human screams, a call type that has received little dedicated research outside of recent efforts from our lab (Engelberg & Gouzoules, 2019; Engelberg et al., 2019; Engelberg et al., 2021; Schwartz & Gouzoules, 2019; Schwartz et al., 2020a; see Arnal et al., 2015 and Frühholz et al., 2021 for studies from different research groups). In Study 1, we investigated the perceptual variation elicited by screams from varied emotional contexts, as well as targeted acoustic cues predicting that variation. Using a design in which every participant rated each stimulus on six emotion scales, we found that different screams elicit varied rather than homogenous perceptions of emotion. Listeners perceived differences in emotion primarily along two principal axes of variation, one separating the perception of anger and pain from the perception of happiness and surprise, and one independently accounting for some perception of fear. The acoustic results additionally suggested that more nuanced assessments of emotion might be possible.

In Study 2, we expanded on these findings in a number of ways. First, we used a forced-choice emotion categorization task to enable a more fine-grained analysis of listeners' accuracies and errors when perceiving emotions from screams. Second, we explored the acoustic variation between screams associated with different emotions, an analysis precluded in Study 1 by its relatively smaller stimulus set. Third, we compared patterns of emotion recognition and acoustic variation in screams to short samples of prosodic speech, mostly revealing parallels in both the expression and perception of emotion across these vocalization types. Finally, we obtained participants' ratings corresponding to screams' potential perceptual effects on listeners, revealing some perceptual characteristics potentially common to contextually diverse screams while also showing variation associated with different scream variants.

Together, these studies provide novel evidence suggesting that acoustic variation within human call types allows listeners to infer varied emotional meanings. That these inferences were often accurate further suggests that a single call type, despite spanning many contexts, nonetheless may sometimes convey more specific information about the caller's emotion.

Contextual diversity and variation within call types

Authors in the animal communication literature have pointed out that the less contextually specific a call's usage, the less potential it has (by itself) to convey specific information (Seyfarth & Cheney, 2003). Smith (1980) wrote at length about such generalized behavioral displays, pointing out that listeners can still infer more specific information by integrating contextual cues (p. 189) but also by attending to variants in the signal itself (p. 184). Indeed, the acoustic variation within a call type may expand its informative capacity if that variation correlates with contextually relevant details (Manser, 2001; Schwartz et al., 2020b). To determine the potential meanings of a call, it is therefore necessary to investigate the acoustic variation within the type, especially if the call is used across multiple contexts and does not provide specific information by virtue of its occurrence.

In nonhuman primates, researchers have identified various call types that occur across contexts (Clay et al., 2015; Crockford et al., 2018), and in many cases, they have described potentially informative acoustic variation within those types (e.g., Cheney & Seyfarth, 2007, p. 230; Crockford & Boesch, 2003; Gouzoules et al., 1984; Gouzoules & Gouzoules, 1989; Gouzoules & Gouzoules, 2011; Rendall, Seyfarth, Cheney, & Owren, 1999; Schwartz, 2020). Research on human call types suggests that several occur in even more diverse contexts than their counterparts in nonhuman primates (Anikin & Persson, 2017) and are therefore, on the basis of their usage patterns alone, potentially even more ambiguous. However, the extent to which this contextual diversity is matched by informative acoustic or perceptual variation has remained unclear. Among infant cries, acoustic variation seems to correlate with physiological arousal but may not convey more nuanced details about emotional context (Porter et al., 1986; Wood & Gustafson, 2001). In contrast, the acoustic variation evident among contextually diverse laughs seems less reducible to arousal-level differences (Szameitat et al., 2009).

The present studies on human screams represent an incipient third data point to complement our knowledge on infant cries and laughter, suggesting several broader conclusions about human call types. First, it is safer now to say that variation within human call types is potentially informative—albeit imperfectly—with respect to the caller's emotional context. In Study 2, listeners identified some calls of every included emotional context as screams, i.e., as belonging to the same call type, yet those calls elicited varied and often accurate interpretations of emotion. Second, it is not obvious that the variation among screams is attributable solely to differences in the vocalizer's arousal level. This aspect of our data may more closely resemble the evidence regarding variation in laughs than in infant cries, although verifying this point would require research obtaining direct physiological measures. Preliminarily, however, we suggest that the acoustic variation among human call types may provide some cues to differentiate qualitatively distinct emotional contexts, and not only levels of arousal or emotional intensity.

These findings are especially notable with respect to the treatment of screams in the emotion literature. Researchers have commonly used screams as fear stimuli (Armony et al., 2007; Cordaro et al., 2016) and/or suggested inducing alarm as a primary function (Arnal et al., 2015). Our findings across both studies showing that screams do not uniformly convey fear suggest that this treatment of screams is problematic (see also Frühholz et al., 2021).

The proximate sources of acoustic variation among screams remain to be studied. However, the strong parallels between this variation and the variation in prosodic simple speech presented in Study 2 offer some indirect evidence. One potential explanation for these similarities is that emotion renders the same automatic effects on the vocal musculature across both vocalization types. Many authors have suggested that these effects in speech prosody originate from homologous mechanisms to those that produce emotion-correlated variation in nonhuman animal calls (Filippi, 2016; Zimmerman et al., 2013). Supporting this hypothesis, many of our acoustic results across both studies echoed those correlated with emotion in nonhuman animals, such as increased noisiness in calls associated with negative emotions (Briefer, 2012). Thus, it is possible that the acoustic variation associated with different emotions in speech, screams, and nonhuman calls all partly reflect the same ancestral processes.

We would not suggest, however, that acoustic variation derives only from automatic effects of emotion. Indeed, a second (non-mutually exclusive) explanation for shared acoustic patterns across vocalization types implicates our *control* over vocal production. Our capacities for vocal flexibility in speech prosody, e.g., to convey emotion intentionally, likely extend to some extent to our production of nonlinguistic vocalizations. It is possible, and perhaps even communicatively efficacious that some of the same expressive conventions have carried over as well, such that linguistic and nonlinguistic patterns of expression complement rather than contradict one another (Anikin & Persson, 2017). That we found potentially more contextually specific variation in screams than in infant cries, mirroring results in adult laughter (Szameitat et al., 2009, 2011), may suggest an influence of speech-related vocal flexibility in this way.

In sum, our findings suggest that acoustic variation among screams enables some disambiguation of an otherwise contextually ambiguous call type, which combined with evidence from laughter (Szameitat et al., 2009) may represent a general feature of nonlinguistic vocal communication in humans. That said, we also documented significant misattributions by listeners, most notably in their tendencies to perceive happiness screams as conveying fear in both studies. Relevant to this finding, it is worth reiterating that under natural conditions, the communication of emotion is only one among many factors that might influence listeners' responses. Below, I expand on this perspective with respect to our results.

Call function and emotional communication

One theme emphasized in the General Introduction was the importance of considering call functions beyond emotional communication, i.e., their effects on listeners and the information conveyed in addition to cues related to caller emotion. I argued that such considerations were highly relevant, for example, to explaining the acoustic structure and evolutionary trajectory of screams in nonhuman animals. Moreover, adopting a framework from Schamberg et al. (2018), I suggested that the broad, shared effect(s) of a call type might partly explain its occurrence across diverse contexts, whereas effects more specific to certain contexts could account for some acoustic variation within the type (provided there is some benefit to contextually specific responses).

Screams were particularly well-suited for an investigation of call function not only to account for their contextual diversity, but because it was plausible that certain contexts would require more specific responses from distant listeners who lacked concurrent visual cues. Our findings are consistent with this idea, both because listeners derived heterogenous (often accurate) emotional meanings from screams—which would enable heterogeneous responses—and because acoustic variation correlated with variation in perceptual ratings.

Study 2, in particular, embodied an initial attempt to explore possibly functionally relevant perceptual effects of human screams along with correlated acoustic variation. We found that screams associated with different emotional contexts were, overall, rated as equally attention-getting. However, they seemed to vary by context in other perceptual aspects, such as their likelihood of conveying distress or eliciting fear in listeners. Capturing attention, then, might represent a general perceptual characteristic of screams that may unite its diverse uses, whereas inducing fear, soliciting aid, and other effects are specific to certain scream variants. Thus, again contra the typical treatment of screams in the literature, it is probably less generally accurate to say that human vocalizers scream to convey fear or induce alarm (Arnal et al., 2015; Belin & Zatorre, 2015), and more accurate to say that vocalizers scream to capture listeners' attention, sometimes conveying or inducing fear depending on the context and acoustic profile of the call.

What is the relationship between *variation in emotion* (or emotional perceptions, as shown in Study 1 and part of Study 2) that correlates with acoustic variation in screams and *variation in perceptual effects* (as shown in Study 2) that correlates with the same phenomenon? One possible answer is that variation in emotion produces acoustic changes in the voice, whether by physiological links, voluntary and/or conditioned fidelity to conventions, or some other causal mechanisms. These acoustic changes in turn might produce varied effects in listeners because they enable different interpretations of emotion, and thus different responses. Potentially, some effects on listeners' behavior are also more direct, such as acoustic features that might be perceived as aversive regardless of the listeners' attributions of emotion to a caller. Regardless, if emotion-correlated acoustic variation does not affect listeners' behavior—or especially if it does affect behavior, but in ways deleterious to the vocalizer—then we might expect this variation to be diminished. However, if this acoustic variation is useful, either because it facilitates cooperative, emotional communication or because it otherwise affects listeners' behavior in ways that (on average) benefit each communicative party, it is likely it will persist. These dynamics are classically applied to the evolution of signals (Searcy & Nowicki, 2005), but in principle, they apply to conventionalized signals as well: a code would not persist if it is not useful.

To reiterate an argument from the General Introduction, the communication of accurate or precise emotional information is not always paramount to how a call functions. Listeners' relatively poor recognition of happiness screams is potentially illuminating. It is possible that callers in happy contexts benefit by producing screams with the same acoustic characteristics that render other screams attention-getting; our perceptual results are consistent with this hypothesis. As for listeners, to offer one illustrative (and speculative) account, it may be pertinent that screams associated with happiness or excitement are so prominent in childhood play (Sherman, 1975). Play might represent a safe context in which children can, in effect, "practice" screaming while parents can become familiar with the acoustic features distinguishing a given child's screams from those of others (Engelberg et al., 2019). It is even perhaps functionally advantageous that these should resemble fear screams closely, such that, e.g., parents might come to recognize their daughter screaming in genuine fear even if they had not often heard her scream when truly frightened.

The acoustic characteristics significant to happy screams, then, would include features that capture attention, that survive propagation across distances, and that convey identity cues across contexts, each of which might limit the acoustic variation specific to happiness contexts. If, as a result of these constraints, listeners sometimes attribute fear to happy screams (as participants in both our studies did), the fitness consequences of this error would likely be relatively minor. Again, this account is speculative. Yet we would have speculated equally to suggest, *a priori*, that the communication of happiness is important to the functioning of happiness screams. The present data are not consistent with this latter account.

Thus, we echo others (Wood et al., 2017; Wood, 2020) who have suggested that explorations of perceptual effects including *but not limited to* emotional communication are fundamental to understanding human call types and the variation among them. We have only explored a subset of potentially significant perceptual considerations in this dissertation and in other recent studies (Engelberg et al., 2019), but our findings that screams might vary perceptually underscore that further explorations of function in human call types could prove productive.

Human call types and language

In light of these studies, I must first reiterate the general appeal for expanded research focusing on human call types. The principal findings from this dissertation were made possible only within this framework. Prior work had established screams as a perceptually recognized call type marked by certain acoustic features (Schwartz et al., 2020a). We acknowledged that calls with these acoustic profiles occurred in diverse contexts—and that listeners classified calls from diverse contexts as screams (Anikin & Persson, 2017)—and we therefore investigated whether these calls exhibited acoustic and perceptual variation. It is unclear how, without identifying screams as a call type, we might have appreciated their contextual diversity let alone any communicatively significant variation within the type.

The effects of language and its underlying cognitive faculties on human call types remains an important question. Deacon (1998, p. 418) suggested that, on the evolutionary scale, the emergence of language likely resulted in a constriction of our nonlinguistic repertoire; that is, as language became the dominant mode of communication, we lost some of our prelinguistic calls (although verifying this claim would require more information on the vocal repertoire of our hominin ancestors). Similarly, in development, the onset of language abilities is marked by a decline in the use of certain call types such as cries (Zeifman, 2001).

What becomes, however, of the several call types persisting as a part of our vocal repertoires even as our faculties for language are fully realized? In screams now as well as laughs (Gervais & Wilson, 2005; Szameitat, 2009; Provine, 2000), we have seen contextual and concomitant acoustic diversification of calls that in nonhuman primates are restricted to relatively more limited contexts (agonistic conflicts for screams, e.g., Cheney, 1977; de Waal & van Hooff, 1981; Gouzoules et al., 1984; tickle/play contexts for laughs, e.g., Davila-Ross et al., 2009; van Hooff, 1972). In other words, there emerges a trend whereby the range of emotional and social contexts in which particular nonlinguistic calls occur has expanded, even if the size of the overall nonverbal repertoire may have decreased (Deacon, 1998, p. 418).

We have suggested that the expanded contextual diversity of human call types in adults may reflect increased control over call usage and weakening ties between calls and emotional states, both abilities important for speech. Oller and colleagues have similarly suggested that *functional flexibility*, or the ability to produce utterances across differently-valenced contexts, is foundational to language (Oller et al., 2016; Oller & Griebel, 2020). Their research group showed that this capacity manifests in infants' prelinguistic protophones (e.g., squeals and vowel-like sounds) by 3 months of age (Oller et al., 2013), and perhaps even within the first month of life (Jhang & Oller, 2017), whereas their laughs and cries are more tightly linked to specific valences. The contextual diversity of screams and laughs in adults, however, suggests that calls that in human infants and in other species are more context-specific have become more functionally flexible as well. Interestingly, the (limited) control over call usage that nonhuman primates exert likely implicates frontal cortical areas, including potential homologs to Broca's area (Taglialatela et al., 2008). Thus, a fascinating possibility is that the benefits of using calls in diverse contexts favored mechanisms for flexible vocal usage that would later become essential and more robust in support of language.

We have additionally argued that the acoustic variation within call types to some degree likely reflects our enhanced abilities of flexible control over the spectro-temporal structures of vocalizations, or call production, also essential to the production of speech (Ackermann et al., 2014; Fitch, 2000). This type of control—not only over call usage but over coordinated respiratory, laryngeal, and articulatory actions that affect specific acoustic features of calls—

seems far rarer among other primates and may implicate neural pathways that are absent or relatively sparse in other species (Fitch & Zuberbühler, 2013). It is this flexibility, for example, that might allow the modulation of acoustic characteristics in accordance with conventional codes as discussed earlier. Indeed, vocalizers might have the capacity to introduce new variation conveying novel meanings in different contexts, and listeners might then learn to derive those meanings and reproduce the variation in their own calls—not dissimilar from the processes involved in inventing and ascribing meaning to conventionalized words.

Some language theorists refer to language's derived functions, or its secondary uses apart from the direct functions for which it was selected (Oesch, 2016). We have, in effect, suggested that the diversification and enhancement of the information provided by nonlinguistic calls might represent a derived or secondary function of the language faculties. Although human call types are homologous to nonhuman calls (Davila-Ross et al., 2009; Lingle et al., 2012; Owren et al., 2011), it is likely they have also undergone fundamental changes in human evolution, such that they operate in novel ways in our species. Many remarkable aspects of human screams—their contextual diversity, and the variation shown in this dissertation—likely reflect these novel developments.

References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences, 37*(6), 529.

Anikin, A., & Persson, T. (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods, 49*(2), 758-771

Armony, J. L., Chochol, C., Fecteau, S., & Belin, P. (2007). Laugh (or cry) and you will be remembered: influence of emotional expression on memory for vocalizations. *Psychological Science, 18*(12), 1027-1029.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology, 25*(15), 2051-2056.

Belin, P., & Zatorre, R. J. (2015). Neurobiology: Sounding the alarm. *Current Biology, 25*(18), R805-R806.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology, 288*(1), 1-20.

Cheney, D. L. (1977). The acquisition of rank and the development of reciprocal alliances among free-ranging immature baboons. *Behavioral Ecology and Sociobiology, 2*(3), 303-318.

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind.* University of Chicago Press.

Clay, Z., Archbold, J., & Zuberbühler, K. (2015). Functional flexibility in wild bonobo vocal behaviour. *PeerJ, 3*, e1124.

Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion, 16*(1), 117.

Crockford, C., & Boesch, C. (2003). Context-specific calls in wild chimpanzees, Pan troglodytes verus: analysis of barks. *Animal Behaviour, 66*(1), 115-125.

Crockford, C., Gruber, T., & Zuberbühler, K. (2018). Chimpanzee quiet hoo variants differ according to context. *Royal Society Open Science, 5*(5), 172066.

Davila-Ross, M., Owren, M., & Zimmermann, E. (2009). Reconstructing the evolution of laughter in great apes and humans. *Current Biology, 19*.

de Waal, F. B., & van Hooff, J. A. (1981). Side-directed communication and agonistic interactions in chimpanzees. *Behaviour, 77*(3), 164-198.

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain* (No. 202). WW Norton & Company.

Engelberg, J. W., & Gouzoules, H. (2019). The credibility of acted screams: Implications for emotional communication research. *Quarterly Journal of Experimental Psychology, 72*(8), 1889-1902.

Engelberg, J. W., Schwartz, J. W., & Gouzoules, H. (2019). Do human screams permit individual recognition?. *PeerJ*, 7, e7087.

Engelberg, J. W., Schwartz, J. W., & 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.

Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Frontiers in Psychology, 7*, 1393.

Fitch, W. T., & Zuberbühler, K. (2013). Primate precursors to human language: beyond discontinuity. In E. Altenmuller, S. Schmidt, & E. Zimmermann (Eds.), *The evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man* (26-48). Oxford University Press.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in cognitive sciences, 4*(7), 258-267.

Frühholz, S., Dietziker, J., Staib, M., & Trost, W. (2021). Neurocognitive processing efficiency for discriminating human non-alarm rather than alarm scream calls. *PLoS biology, 19*(4), e3000751.

Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *The Quarterly Review of Biology, 80*(4), 395-430.

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (Macaca mulatta) screams: representational signalling in the recruitment of agonistic aid. *Animal Behaviour, 32*(1), 182-193.

Gouzoules, H. & Gouzoules, S. (1989). Design features and developmental modification of pigtail macaque, Macaca nemestrina, agonistic screams. *Animal Behaviour, 37*, 383–401.

Gouzoules, H., & Gouzoules, S. (2011). The conundrum of communication. In C. J. Campbell, A. Fuentes, K. C. MacKinnon, R. Stumpf, & S. K. Bearder (Eds.), *Primates in perspective* (2nd ed., pp. 626–636). New York: Oxford University Press.

Jhang, Y., & Oller, D. K. (2017). Emergence of functional flexibility in infant vocalizations of the first 3 months. *Frontiers in psychology, 8*, 300.

Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., & Romanow, C. A. (2012). What makes a cry a cry? A review of infant distress vocalizations. *Current Zoology, 58*(5), 698-726.

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 268*(1483), 2315-2324.

Oesch, N. (2016). Deception as a derived function of language. *Frontiers in psychology, 7*, 1485.

Oller, D., Buder, E., Ramsdell, H., Warlaumont, A., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences, 110*(16), 6318-6323.

Oller, D. K., Griebel, U., & Warlaumont, A. S. (2016). Vocal development as a guide to modeling the evolution of language. *Topics in cognitive science, 8*(2), 382-392.

Oller, D. K., & Griebel, U. (2020). Functionally flexible signaling and the origin of language. *Frontiers in psychology, 11*.

Owren, M., Amoss, R., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology, 73*(6), 530-544.

Porter, F. L., Miller, R. H., & Marshall, R. E. (1986). Neonatal pain cries: effect of circumcision on acoustic features and perceived urgency. *Child development,* 790-802.

Provine, R. R. (2000). The science of laughter. *Psychology Today, 33*(6), 58-62.

Rendall, D., Seyfarth, R. M., Cheney, D. L., & Owren, M. J. (1999). The meaning and function of grunt variants in baboons. *Animal Behaviour, 57*(3), 583-592.

Schamberg, I., Wittig, R. M., & Crockford, C. (2018). Call type signals caller goal: a new take on ultimate and proximate influences in vocal production. *Biological Reviews, 93*(4).

Schwartz, J. W., Engelberg, J. W., & Gouzoules, H. (2020a). Was that a scream? Listener agreement and major distinguishing acoustic features. *Journal of Nonverbal Behavior*, 1-20.

Schwartz, J. W., Engelberg, J. W. M., & Gouzoules, H. (2020b). Evolving views on cognition in animal vocal communication: Contributions from scream research. *Animal Behavior and Cognition, 7*(2), 192-213.

Schwartz, J. W. (2020). Vocal emotion expression across contexts, vocalization types, and species: Implications for general processes of vocal evolution. (Doctoral dissertation). Retrieved from https://etd.library.emory.edu/concern/etds/jw827c636

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication: reliability and deception in signaling systems*. Princeton University Press.

Seyfarth, R., & Cheney, D. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology, 54*(1), 145-173.

Sherman, L. W. (1975). An ecological study of glee in small groups of preschool children. *Child Development,* 53-61.

Smith, W. J. (1980). *The behavior of communicating.* Harvard University Press.

Szameitat, D. P., Alter, K., Szameitat, A. J., Darwin, C. J., Wildgruber, D., Dietrich, S., & Sterr, A. (2009). Differentiation of emotions in laughter at the behavioral level. *Emotion, 9*(3), 397.

Szameitat, D. P., Darwin, C. J., Wildgruber, D., Alter, K., & Szameitat, A. J. (2011). Acoustic correlates of emotional dimensions in laughter: arousal, dominance, and valence. *Cognition and Emotion, 25*(4), 599-611.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2008). Communicative signaling activates 'Broca's' homolog in chimpanzees. *Current Biology, 18*(5), 343-348.

Van Hooff, J. A. (1972). A comparative approach to the phylogeny of laughter and smiling. In R. A. Hinde (Ed.), *Non-verbal communication*. Cambridge University Press.

Wood, R. M., & Gustafson, G. E. (2001). Infant crying and adults' anticipated caregiving responses: acoustic and contextual influences. *Child development, 72*(5), 1287-1300.

Wood, A., Martin, J., & Niedenthal, P. (2017). Towards a social functional account of laughter: Acoustic features convey reward, affiliation, and dominance. *PloS one, 12*(8), e0183811.

Wood, A. (2020). Social context influences the acoustic properties of laughter. *Affective Science, 1*(4), 247-256.

Zimmermann, E., Leliveld, L., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices. In E. Altenmuller, S. Schmidt, & E. Zimmermann (Eds.), *The evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man* (pp. 116-132). Oxford University Press.

APPENDIX

Statistical analysis details for Study 2

Emotion categorization accuracy

Unbiased hit rates and confusion matrices. In addition to raw hit rates and false alarm rates (FAs), we calculated unbiased hit rates (*Hu*) and examined errors via confusion matrices. *Hu* scores are similar to *d'* scores in signal detection theory, i.e., a measure of accuracy that takes into account participants' response biases, but unlike *d'* they apply to judgments with more than two response options. They are calculated, for each category, as the squared number of hits divided by the product of the number of stimuli in that category and the total number of stimuli the participant identified as belonging to the category (correctly or not). Confusion matrices, developed separately for linguistic and nonlinguistic stimuli, plot mean hit rates for every emotion (correct identification of a stimulus as conveying that emotion, e.g., categorizing a fear stimulus as fear) on the diagonal and false alarm rates (incorrect identification of a stimulus that belongs to a different category as conveying that emotion, e.g., categorizing an anger stimulus as fear) on non-diagonal entries (Elfenbein et al., 2002).

ANOVA and post-hoc tests. To analyze the effects of Emotion, Vocalization Type (linguistic or nonlinguistic), and the Emotion*Vocalization Type interaction on measures of accuracy, we conducted two-way repeated measures ANOVAs on the aligned rank transformed *Hu*s and FAs. The aligned rank transform allows the use of *F*-tests on nonparametric, repeated measures factorial data (Wobbrock et al., 2011), which was necessary because the standardized residuals for some groups were non-normal. If the Emotion*Vocalization Type interaction was significant, we then examined the simple effects of Emotion within each Vocalization Type using a Friedman's Test. Post-hoc pairwise comparisons were conducted with Dunn-Bonferroni-corrected *p*-values to examine the effects of each emotion relative to the others.

Acoustic differences between emotions

Model specifications and tests. To examine the relationship between acoustic parameters and emotions, 14 separate mixed models were initially fitted with each acoustic parameter as an outcome variable and Emotion, Vocalization Type, and their interaction as predictor variables, as well as Rendition as a fixed effect and Actor as a random effect. Every continuous outcome variable was first fitted as a linear mixed model (LMM) with a Gaussian distribution and identity link using the lme4 package in R (Bates et al., 2015). Mixed models enable analysis of trial-by-trial data while accounting (via random effects) for the nonindependence of multiple observations associated with the same factor (e.g., multiple trials from the same participant or involving the same stimulus or actor; Barr et al., 2013). Residuals were checked visually for violations of homoscedasticity or normality assumptions. In the event that an assumption was violated, the model was refitted as a penalized quasi-likelihood generalized linear mixed model (GLMM-PQL) with a log link using the glmmPQL function in the MASS package (Venables & Ripley, 2012). Binomial outcome parameters (pitch jump and NLP) were fitted using generalized linear mixed models (GLMMs) with logit links, which are similar to ordinary logistic regressions in that they enable analysis of binary outcomes, but they also allow incorporation of random effects like other mixed models (Jaeger, 2008).

After fitting these models, we first checked the significance of the Emotion * Vocalization Type interaction term. For LMMs and GLMMs, the effect of the interaction term was estimated using a likelihood-ratio test, conducted using the anova function in the stats package, which tests whether the inclusion of a fixed effect significantly improves model fit relative to the same model omitting that term (Barr et al., 2013). For GLMM-PQLs, which do not estimate likelihood directly (Chen, Xiao, & Staicu, 2019), significance was estimated via a Wald test using the wald.test function in the aod package (Lesnoff & Lacelot, 2012), which tests whether the coefficient estimate for the fixed effect significantly differs from 0. If the Vocalization Type * Emotion interaction was insignificant, these models were refitted without the interaction or Vocalization Type fixed effects to investigate the main effect of Emotion on each parameter. If the interaction was significant, separate models were fitted for each parameter within each Vocalization Type, according to the same specifications and procedure described above.

For LMMs, main effects of Emotion and Rendition were estimated using Satterthwaite's method in the anova function of the lmertest package (Kuznetsova et al., 2017). For GLMMs, main effects were estimated using a likelihood-ratio test, and for GLMM-PQLs, using a Wald Test. For both GLMMs and GLMM-PQLs, if Rendition was insignificant, the model was re-fitted without this variable, as GLMM estimates are sensitive to overfitting caused by the inclusion of insignificant variables (Barr et al., 2013); note that Rendition was never significant and was always removed. This step was not necessary for the LMMs because Satterthwaite's method does not necessitate removal of insignificant effects except in the case of nonsignificant interactions (Engqvist, 2005). For every model in which Emotion was significant, post-hoc comparisons between different emotions were conducted using the Kenward-Roger Method with Tukey *p*-value adjustments via the ls_means function in lmertest (for LMMs) or the lsmeans function in the lsmeans package (for GLMMs and GLMM-PQLs).

Acoustic effects on listeners' responses

Parameter selection. Due to correlations between several of the acoustic parameters, the inclusion of all 14 parameters in models predicting listeners' responses would have resulted in multicollinearity and inflated standard errors (Jaeger, 2008). Therefore, we selected the following subset of parameters that were of theoretical interest and that we believed might account for semi-independent vocal-acoustic variation: duration, mean F0, jitter, time to max F0, mean HNR, PFMean, IQRBW, pitch jumps, and NLP. Note that although the speech and scream utterances possess intrinsically different durations, F0s, etc., our focus was on how these variables varied by emotion *within* each vocal type, rather than on overall differences between speech and screams. We conducted a Varimax-rotated principal component analysis (PCA) to characterize the acoustic data's covariance structure and to ensure that none of the selected parameters loaded strongly onto the same component. To check that this selection procedure successfully limited collinearity in our models, variance inflation factors (VIFs) were inspected using the check_collinearity function in the performance package in R (Lüdecke, 2020), revealing that the inclusion of PFMean resulted in concerningly high VIFs. PFMean was subsequently dropped, leaving 8 acoustic parameters.

Model specifications and tests. To determine how the acoustic parameters predicted listeners' categorizations of emotion, we developed composite variables, for each emotion and within each vocalization type, reflecting the number of listeners who categorized a given stimulus as conveying that emotion. These variables were used as the outcome variables in five separate GLMMs, with all 8 of the selected acoustic parameters as well as Vocalizer Sex as fixed effects and Actor as a random intercept. Due to the large number of zeros and negative skew of the response variables, these models were specified with a zero-inflated negative binomial response distribution using the *glmmTMB* package in R (Brooks et al., 2017). The zero-inflation parameter was dropped if a model failed to converge, which can occur if the zero-inflation parameter is near zero and its inclusion therefore overfits the model (Brooks et al., *ibid.*). After fitting the full models, each model was simplified by removing factors with non-significant effects. The significance of each effect in this minimal model was estimated using likelihood-ratio tests. Each non-significant factor was re-added individually to the minimal model to check for type-II errors and to estimate the level of non-significance via a likelihood-ratio test (Barr et al., 2013).

Perceptual ratings

Model specifications and tests. To investigate how a scream's original emotional context affected listeners' perceptual ratings, four cumulative link mixed models (CLMMs, appropriate for ordinal outcome data) with logit links were fitted using the clmm function in the *ordinal* package in R (Christensen, 2019), with ratings on each scale as the ordinal outcome variables, Emotion as a categorical predictor variable, and Participant and Stimulus as random effects. Maximum likelihood estimates for each parameter were computed using the Laplace approximation and post-hoc pairwise comparisons of least-squares-means were conducted using the Kenward-Roger method with Tukey *p*-value adjustment via the lsmeans function in the *lsmeans* package (Lenth, 2018).

To determine how acoustic parameters predicted listeners' ratings on each perceptual scale, we fitted four separate cumulative link mixed models (CLMMs), with the rating on each scale as the outcome variable, the 8 selected acoustic parameters and Vocalizer Sex as fixed effects, and Participant and Stimulus as random effects. Models were then simplified and effects were estimated according to the same procedure described above.

Table A1. Acoustic descriptives of stimuli from Study 2.

		Nonlinguistic stimuli													
Actor	Sex	Duration	F0 Mean	F0 Range	Max Time Percent	Slope to Max	Jitter ($*10^3$)	Shimmer	HNR Mean	DFA50	IQRBW	PF Mean	Roughness	Pitch Jumps	NLP
1	F	0.61	0.48	0.19	0.64	0.70	0.30	1.08	8.37	1.15	0.47	1.01	21.97	6.00	6.00
		(0.48)	(0.08)	(0.11)	(0.26)	(0.65)	(0.29)	(0.35)	(5.42)	(0.18)	(0.4)	(0.15)	(1.28)		
2	F	0.52	0.55	0.28	0.71	1.26	0.25	1.05	8.96	0.88	0.40	0.85	24.58	3.00	11.00
		(0.29)	(0.16)	(0.17)	(0.23)	(1.2)	(0.16)	(0.21)	(3.37)	(0.21)	(0.11)	(0.37)	(5.61)		
3	F	0.63	0.59	0.41	0.43	1.71	0.50	1.37	6.78	1.14	0.58	0.99	24.23	3.00	8.00
		(0.4)	(0.17)	(0.16)	(0.19)	(1.03)	(0.39)	(0.48)	(5.35)	(0.51)	(0.55)	(0.39)	(2.23)		
4	M	1.04	0.31	0.15	0.35	0.50	0.62	1.37	4.22	1.14	0.40	1.04	28.36	3.00	11.00
		(0.62)	(0.06)	(0.08)	(0.29)	(0.43)	(0.48)	(0.25)	(3.67)	(0.13)	(0.15)	(0.1)	(1.96)		
5	F	0.68	0.64	0.37	0.30	2.16	0.30	1.03	8.37	1.28	0.37	1.17	24.57	1.00	8.00
		(0.45)	(0.24)	(0.17)	(0.29)	(1.81)	(0.31)	(0.39)	(5.99)	(0.2)	(0.27)	(0.23)	(2.12)		
6	M	0.61	0.43	0.23	0.36	1.37	0.70	1.37	5.24	1.18	0.65	1.04	25.61	3.00	13.00
		(0.3)	(0.15)	(0.11)	(0.27)	(0.98)	(0.62)	(0.36)	(3.46)	(0.25)	(0.5)	(0.22)	(1.52)		
7	F	0.76	0.77	0.45	0.25	1.28	0.21	1.24	8.23	1.19	0.40	1.07	23.21	3.00	14.00
		(0.29)	(0.29)	(0.15)	(0.12)	(0.78)	(0.19)	(0.52)	(5.9)	(0.24)	(0.32)	(0.28)	(2.22)		
8	M	0.49	0.34	0.13	0.28	0.54	0.47	1.03	6.18	0.90	0.59	0.81	21.67	1.00	8.00
		(0.16)	(0.05)	(0.05)	(0.17)	(0.4)	(0.42)	(0.38)	(3.38)	(0.25)	(0.17)	(0.23)	(1.66)		
9	M	0.54	0.37	0.23	0.31	1.45	0.96	1.32	5.55	0.93	0.44	0.74	25.80	7.00	13.00
		(0.25)	(0.09)	(0.13)	(0.31)	(1.14)	(1.03)	(0.4)	(3.37)	(0.28)	(0.27)	(0.24)	(1.86)		
Overall		**0.65**	**0.50**	**0.27**	**0.40**	**1.22**	**0.48**	**1.21**	**6.63**	**1.11**	**0.49**	**0.98**	**24.45**	**30.00**	**92.00**
		(0.4)	**(0.21)**	**(0.17)**	**(0.28)**	**(1.12)**	**(0.54)**	**(0.4)**	**(4.76)**	**(0.29)**	**(0.36)**	**(0.27)**	**(0.66)**		

Table A1 cont.

		Linguistic stimuli													
Actor	Sex	Duration	F0 Mean	F0 Range	Max Time F0	Slope to Max	Jitter (*10^3)	Shimmer	HNR Mean	DFA50	IQRBW	PF Mean	Roughness	Pitch Jumps	NLP
1	F	0.35	0.33	0.10	0.45	0.63	0.97	1.26	6.11	1.07	0.88	0.90	23.22	0.00	3.00
		(0.08)	(0.13)	(0.08)	(0.34)	(0.76)	(1.39)	(0.57)	(3.83)	(0.28)	(0.39)	(0.3)	(1.57)		
2	F	0.47	0.30	0.13	0.38	0.53	0.96	1.13	7.53	0.76	0.65	0.60	22.29	0.00	4.00
		(0.09)	(0.1)	(0.09)	(0.31)	(0.39)	(1.77)	(0.36)	(3.92)	(0.35)	(0.19)	(0.27)	(1.25)		
3	F	1.06	0.52	0.35	0.40	1.20	0.40	1.22	7.27	1.09	0.68	0.79	23.10	5.00	9.00
		(0.69)	(0.16)	(0.17)	(0.29)	(1.29)	(0.33)	(0.46)	(5.31)	(0.62)	(0.46)	(0.28)	(1.79)		
4	M	0.47	0.19	0.09	0.39	0.30	1.66	1.49	1.94	1.03	0.61	0.72	26.45	0.00	5.00
		(0.37)	(0.04)	(0.05)	(0.27)	(0.23)	(1.02)	(0.43)	(2.17)	(0.25)	(0.21)	(0.17)	(1.23)		
5	F	0.51	0.44	0.27	0.27	1.08	0.85	1.40	5.45	0.96	0.77	0.85	25.78	3.00	8.00
		(0.2)	(0.13)	(0.18)	(0.38)	(1.35)	(1.42)	(0.38)	(4.67)	(0.41)	(0.27)	(0.3)	(1.9)		
6	M	0.38	0.18	0.07	0.53	0.39	2.39	1.37	1.20	0.84	0.85	0.79	30.59	0.00	4.00
		(0.09)	(0.06)	(0.05)	(0.32)	(0.36)	(1.11)	(0.5)	(1.97)	(0.3)	(0.23)	(0.29)	(1.03)		
7	F	0.52	0.37	0.18	0.45	0.59	0.88	1.13	6.86	0.77	0.68	0.69	21.35	1.00	6.00
		(0.23)	(0.23)	(0.15)	(0.36)	(0.63)	(0.74)	(0.45)	(5.22)	(0.52)	(0.58)	(0.43)	(1.71)		
8	M	0.61	0.20	0.08	0.24	0.28	1.13	1.00	6.73	0.65	0.51	0.50	23.72	0.00	3.00
		(0.31)	(0.06)	(0.04)	(0.14)	(0.2)	(0.73)	(0.33)	(5.28)	(0.34)	(0.37)	(0.16)	(2 29)		
9	M	0.55	0.25	0.18	0.49	0.81	1.39	1.20	3.63	0.86	0.80	0.83	25.47	4.00	7.00
		(0.17)	(0.09)	(0.12)	(0.33)	(0.59)	(0.87)	(0.26)	(3.52)	(0.24)	(0.17)	(0.15)	(1.75)		
Overall		**0.55**	**0.31**	**0.16**	**0.40**	**0.63**	**1.18**	**1.24**	**5.11**	**0.90**	**0.72**	**0.75**	**24.76**	**13.00**	**49.00**
		(0.36)	**(0.16)**	**(0.14)**	**(0.32)**	**(0.77)**	**(1.21)**	**(0.44)**	**(4.61)**	**(0.4)**	**(0.36)**	**(0.29)**	**(0.59)**		

Note. Duration is presented in seconds. All frequency parameters are presented in kHz. Slope to Max is presented in kHz/s. Pitch jumps and NLP reflect frequencies of incidence with a maximum of 15 per actor.

*Table A2. Effects of emotion on acoustic parameters with no significant Emotion * Vocalization Type interaction in Study 2.*

Mean F0	Fear	Happiness	Pain	Surprise
Anger	**-0.373**	**-0.378**	-0.118	**-0.543**
	(0.086)	**(0.085)**	(0.094)	**(0.082)**
Fear		-0.006	**0.254**	-0.170*
		(0.068)	**(0.079)**	(0.063)
Happiness			**0.260**	-0.165*
			(0.079)	(0.063)
Pain				**-0.425**
				(0.074)
F0 Range				
Anger	-32.182	**-83.253**	-15.430	**-150.717**
	(24.756)	**(24.756)**	(24.756)	**(24.756)**
Fear		**-51.070**	16.752	**-118.535**
		(24.636)	(24.636)	**(24.636)**
Happiness			**67.822**	**-67.465***
			(24.636)	**(24.636)**
Pain				**-135.287**
				(24.636)
Slope to Max				
Anger	-0.212	-0.279*	0.261	**-0.936**
	(0.2)	(0.206)	(0.261)	**(0.17)**
Fear		-0.067	0.473	**-0.724**
		(0.178)	(0.24)	**(0.135)**
Happiness			0.540	**-0.658**
			(0.245)	**(0.144)**
Pain				**-1.198**
				(0.215)
Jitter				
Anger	0.547	0.744	0.243	0.200
	(0.236)	(0.275)	(0.192)	(0.187)
Fear		0.197	-0.304	-0.347
		(0.321)	(0.253)	(0.249)
Happiness			-0.501*	-0.544
			(0.29)	(0.287)
Pain				-0.043
				(0.208)

Table A2 cont.

Shimmer	Fear	Happiness	Pain	Surprise
Anger	**0.226**	**0.561**	**0.168***	**0.174**
	(0.07)	**(0.07)**	**(0.07)**	**(0.07)**
Fear		**0.335**	-0.058	-0.052
		(0.07)	(0.07)	(0.07)
Happiness			**-0.393**	**-0.387***
			(0.07)	**(0.07)**
Pain				0.006
				(0.07)
PF Mean				
Anger	83.361	-107.238	-52.523	21.571
	(56.885)	(55.999)	(55.978)	(56.677)
Fear		**-190.599**	**-135.884***	-61.789
		(56.865)	**(56.858)**	(57.428)
Happiness			54.715	**128.810***
			(55.954)	**(56.576)**
Pain				74.095
				(56.578)

Note. DFA50, IQRBW, and Pitch Jump not shown due to insignificance of model. Parentheticals show standard errors.

*Table A3. Simple effects of emotion on acoustic parameters for which there was a significant Emotion * Vocalization Type interaction in Study 2.*

	Linguistic				**Nonlinguistic**			
Duration	Fear	Happiness	Pain	Surprise	Fear	Happiness	Pain	Surprise
Anger	0.007	**-0.831**	**-0.527**	-0.134	0.260	0.052	0.120	**0.804***
	(0.152)	**(0.117)**	**(0.124)**	(0.142)	(0.133)	(0.118)	(0.123)	**(0.199)**
Fear		**-0.837**	**-0.534**	-0.141		-0.208	-0.140	0.544
		(0.117)	**(0.125)**	(0.143)		(0.136)	(0.14)	(0.21)
Happiness			**0.304**	**0.697**			0.068	**0.752***
			(0.078)	**(0.105)**			(0.126)	**(0.201)**
Pain				**0.393**				**0.684***
				(0.113)				**(0.204)**
Time to Max F0								
Anger	-0.055	**0.311**	**0.179**	0.098	-0.057	-0.027	0.107	-0.115
	(0.079)	**(0.079)**	**(0.079)**	(0.079)	(0.064)	(0.064)	(0.064)	(0.064)
Fear		0.365	**0.234**	0.152		0.030	**0.164***	-0.058
		(0.078)	**(0.078)**	(0.078)		(0.064)	**(0.064)**	(0.064)
Happiness			-0.132	**-0.213**			**0.135***	-0.088*
			(0.078)	**(0.078)**			**(0.064)**	(0.064)
Pain				-0.081				**-0.222**
				(0.078)				**(0.064)**
Mean HNR								
Anger	**-3.508**	**-4.727**	-1.197	-1.875	**-0.678**	**-0.840**	**-0.559***	**-0.677**
	(1.076)	**(1.076)**	(1.065)	(1.089)	**(0.126)**	**(0.12)**	**(0.128)**	**(0.124)**
Fear		-1.218	**2.312**	1.633		-0.162	0.119	0.001
		(1.075)	**(1.065)**	(1.086)		(0.077)	(0.088)	(0.083)
Happiness			**3.530**	**2.851**			**0.281***	0.163
			(1.065)	**(1.086)**			**(0.08)**	(0.075)
Pain				-0.679				-0.118
				(1.078)				(0.086)

Table A3 continued.

	Linguistic				**Nonlinguistic**			
Roughness	Fear	Happiness	Pain	Surprise	Fear	Happiness	Pain	Surprise
Anger	**5.030**	**6.147**	0.831	2.873	**11.746**	**12.937**	**9.714***	**10.022**
	(1.672)	**(1.672)**	(1.655)	(1.692)	**(1.577)**	**(1.522)**	**(1.538)**	**(1.538)**
Fear		1.117	**-4.199**	-2.157		1.192	-2.032*	-1.724
		(1.67)	**(1.655)**	(1.687)		(1.577)	(1.591)	(1.591)
Happiness			**-5.315**	-3.274			**-3.223**	-2.916
			(1.655)	(1.687)			**(1.538)**	(1.538)
Pain				2.042				0.308
				(1.674)				(1.552)
NLP								
Anger	1.462	1.227	-0.98	1.227	+∞	+∞	+∞	+∞
	(0.642)	(0.616)	(0.586)	(0.616)	*	*	*	*
Fear		-0.236	**-2.442**	-0.236		0.558	-1.423	1.109
		(0.687)	**(0.673)**	(0.687)		(0.611)	(0.728)	(0.624)
Happiness			**-2.207**	0			-1.981	0.551
			(0.648)	(0.664)			(0.731)	(0.608)
Pain				**2.207**				**2.532***
				(0.648)				**(0.749)**

*Unable to estimate effects and standard errors due to quasi-complete separation: Every nonlinguistic anger stimulus contains NLP.

Table A4. Effects of acoustics on listeners' categorizations of emotion in Study 2.

	Linguistic					**Nonlinguistic**				
	Anger	Fear	Happiness	Pain	Surprise	Anger	Fear	Happiness	Pain	Surprise
Duration	-0.025 (0.484)	**-2.644** **(0.525)**	**1.87** **(0.447)**	0.503 (0.427)	**-1.113** **(0.284)**	**1.415** **(0.496)**	**-0.663*** **(0.277)**	-0.41 (0.376)	0.23 (0.26)	**-2.656** **(0.316)**
F0	**-5.572** **(1.224)**	1.207 (1.167)	0.122 (0.97)	1.336 (0.981)	**4.595** **(0.854)**	**-2.629** **(0.935)**	**1.722** **(0.743)**	0.824 (0.742)	**-5.143** **(0.624)**	**2.06*** **(0.609)**
Jitter	**-0.394** **(0.157)**	0.101 (0.119)	-0.016 (0.103)	-0.106 (0.092)	0.113 (0.108)	0.211 (0.312)	**-0.729** **(0.26)**	0.39 (0.27)	**-1.169** **(0.199)**	**0.667*** **(0.206)**
HNR	**-0.224** **(0.049)**	**0.151** **(0.031)**	-0.001 (0.035)	0.034 (0.028)	-0.031 (0.03)	**-0.214** **(0.041)**	**0.113*** **(0.03)**	**0.177** **(0.031)**	**0.068*** **(0.023)**	**0.081*** **(0.027)**
Time Max F0	**0.876** **(0.446)**	0.626 (0.387)	-0.704 (0.448)	**-1.028** **(0.387)**	**0.855** **(0.364)**	0.555 (0.552)	-0.323 (0.449)	0.543 (0.522)	**-1.064*** **(0.358)**	-0.394 (0.47)
IQRBW	-0.173 (0.61)	0.184 (0.357)	0.233 (0.41)	0.43 (0.349)	-0.09 (0.305)	-0.07 (0.415)	0.094 (0.355)	0.269* (0.35)	-0.444 (0.253)	0.159 (0.344)
Pitch Jump	-0.808 (0.588)	**1.297** **(0.47)**	0.513 (0.455)	0.582 (0.42)	-0.446 (0.367)	**-0.897** **(0.329)**	**0.548*** **(0.279)**	0.223 (0.332)	**0.482*** **(0.214)**	0.341 (0.247)
NLP	0.542 (0.306)	-0.155 (0.288)	**-0.92** **(0.251)**	**1.125** **(0.266)**	-0.433 (0.223)	**1.525*** **(0.392)**	-0.051 (0.258)	-0.443 (0.268)	**0.963** **(0.213)**	**-1.072** **(0.207)**
Vocalizer Sex	**-1.181** **(0.389)**	-0.745 (.363)	0.273 (.400)	-0.457 (.258)	**1.212** **(0.261)**	-0.541 (.366)	**0.792** **(0.286)**	-0.428 (.267)	**-0.759** **(0.222)**	0.357 (.290)

Note. Shaded cells indicate an effect whose significance was different from the other vocalization type. Asterisks for the nonlinguistic effects indicate parameters whose significance differed when analyses were restricted to stimuli classified as screams by > 50% of listeners.