Introducing the Oxford Vocal (OxVoc) Sounds database: a validated set of non-acted affective sounds from human infants, adults, and domestic animals

Christine E. Parsons[1,2], Katherine S. Young[1,2], Michelle G. Craske[3], Alan L. Stein[1] and Morten L. Kringelbach[1,2]*

[1] Section of Child and Adolescent Psychiatry, Department of Psychiatry, University of Oxford, Oxford, UK
[2] Center of Functionally Integrative Neuroscience, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[3] Department of Psychology, University of California, Los Angeles, CA, USA

Sound moves us. Nowhere is this more apparent than in our responses to genuine emotional vocalizations, be they heartfelt distress cries or raucous laughter. Here, we present perceptual ratings and a description of a freely available, large database of natural affective vocal sounds from human infants, adults and domestic animals, the Oxford Vocal (OxVoc) Sounds database. This database consists of 173 non-verbal sounds expressing a range of happy, sad, and neutral emotional states. Ratings are presented for the sounds on a range of dimensions from a number of independent participant samples. Perceptions related to valence, including distress, vocalizer mood, and listener mood are presented in Study 1. Perceptions of the arousal of the sound, listener motivation to respond and valence (positive, negative) are presented in Study 2. Perceptions of the emotional content of the stimuli in both Study 1 and 2 were consistent with the predefined categories (e.g., laugh stimuli perceived as positive). While the adult vocalizations received more extreme valence ratings, rated motivation to respond to the sounds was highest for the infant sounds. The major advantages of this database are the inclusion of vocalizations from naturalistic situations, which represent genuine expressions of emotion, and the inclusion of vocalizations from animals and infants, providing comparison stimuli for use in cross-species and developmental studies. The associated website provides a detailed description of the physical properties of each sound stimulus along with cross-category descriptions.

Keywords: emotion perception, infant crying, distress calls, parent-infant, caregiving, auditory perception, infant vocalization

INTRODUCTION

Darwin (1872) was the first to describe and classify human emotional facial expressions and vocalizations as biologically determined expressions of inner emotions. Human facial expressions have since been the subject of a vast body of research, demonstrating that the basic categories of emotional facial expressions are highly recognizable and present across cultures (for review see Ekman, 1993). Although studied to a lesser extent, vocal expressions of emotion can also be readily categorized and occur across cultures (e.g., Scherer et al., 2001; Sauter et al., 2010a). Furthermore, both facial expressions and non-verbal emotional vocal expressions are ubiquitous across the life span, and apparent in infants from birth (Parsons et al., 2010, 2013a).

Facial expressions and vocal expressions clearly differ in the manner through which they convey emotion. In facial expressions, emotion is communicated via configurations of movements or positions of facial muscles (Ekman, 1993). In vocalizations, emotion can be communicated in two forms: verbal prosody and non-verbal affective bursts (Scherer, 1994). Verbal prosody provides information about the affective state of the speaker through the intonation and melody of speech. Non-verbal affective bursts, by contrast, have no linguistic content and tend to be more spontaneous, less constrained expressions of emotion, such as crying, screams, laughter, and sighing. These vocalizations typically accompany intense emotional states, and closely parallel affective vocalizations from other species (Scherer, 1995).

The majority of research into human emotional vocalizations to date has examined verbal prosody (for review see Scherer, 2003). Sound databases for this purpose generally contain vocalizations from actors speaking semantically neutral sentences or strings of "pseudo-words" with different types of prosody [e.g., the Danish Emotional Speech Database (Engberg and Hansen, 1996); Berlin Database of Emotional Speech, (Burkhardt et al., 2005)]. Controlling for the semantic content in this manner ensures that there is no interaction between meaning and prosody. However, such sentences have the disadvantage of being unnatural because there is typically a correspondence between semantic and prosodic features (Scherer et al., 1984). Another common method, using "pseudo-sentences" (sentences made up of non-words) again avoids semantic confounds, but is limited in

terms of how natural the vocalizations sound. A second issue is that even pseudo-words often contain features that are language-specific, such as legal phoneme groups. In summary, the study of verbal prosody is, by necessity, entangled with features of the language being studied.

As a consequence, there has been increased interest in the use of non-verbal affective bursts as a means of studying emotion processing. These bursts constitute a primitive and universal mode of communication (Sauter et al., 2010a) and parallel the vocalizations of other species (e.g., Scherer, 1995; Juslin and Laukka, 2003; Belin et al., 2004). Some categories of vocal burst, such as laughs, cries and screams, are thought to reflect pure, "raw" emotion and therefore are powerful stimuli to use when investigating responses to emotion (Scherer, 1995).

Affective bursts are fundamentally different to speech in their underlying production mechanisms (Scott et al., 2010). Speech requires that specific physical parameters are produced, whereas affective bursts do not, allowing for the presence of a greater range of vocal features. The relative lack of language specific features means that sounds can be used to study universal or cross-cultural perception of emotion. Studies on cross-cultural perception of non-speech sounds are only beginning to emerge. Those that exist suggest that basic emotions can be recognized across diverse cultures (Sauter et al., 2010b; Koeda et al., 2013; Laukka et al., 2013) but perception of more elaborate emotional expressions, such as "triumph" (Sauter et al., 2010b), "disgust" (Koeda et al., 2013), or "shame" (Laukka et al., 2011) show cross-cultural differences. Like for faces, there may be universally recognized categories and also culture-specific signals (Elfenbein and Ambady, 2002).

While there are numerous well-established databases of faces and facial expressions, the study of vocal emotion processing has traditionally lagged behind. The available databases of affective bursts contain sounds generated by adult actors [e.g., The Montreal Affective Voices, (Belin et al., 2008); The International Affective Digitized Sounds, (Stevenson and James, 2008); Corpus of Non-verbal Vocalizations, (Lima et al., 2013)]. For instance, the Montreal Affective Voices database comprises vocalizations from ten actors, all of whom produced the vowel sound /a/ in discrete emotional categories. The Corpus of non-verbal vocalizations comprises eight emotional sound burst categories (four negative, four positive) from adults asked to imagine an emotion and make a corresponding vocal sound. A major advantage of these databases is that they include sounds with a range of emotions, which overall are highly recognizable (Belin et al., 2008; Lima et al., 2013).

However, the use of actors to obtain affective vocalization bursts is not without difficulty. Acted vocalizations are, by definition, not spontaneous emotional expressions. Reliance on acted vocalizations assumes that acted and authentic vocalizations are the same, or that differences in authenticity are imperceptible. However, there is growing evidence to suggest that authenticity in emotional vocalizations can be readily detected by adult listeners (Barker, 2013). Acted emotional expressions have been shown to be perceived as sounding more stereotyped and exaggerated (Laukka et al., 2011) and more extreme (Barkhuysen et al., 2007) compared to authentic ones. A number of physical differences between acted and authentic vocalizations have also been

described, such as acted vocalizations having a higher and more variable pitch (Audibert et al., 2010; Jürgens et al., 2011). Finally, distinct patterns of neuronal activity have been found when listening to authentic compared to acted speech segments (Drolet et al., 2012), suggesting that differential processing occurs as a function of affective authenticity.

The reasons why authentic and acted vocalizations differ are not yet clear, but are likely to be variable and dependent on the quality of acting. It has been suggested that some fundamental features that support the identification of acted vocalizations (such as intensity of the emotion) may be unavoidable (Scherer, 2003). Other suggestions center around the idea that emotions are accompanied by physiological reactions (Kreibig, 2010) not under full voluntary control, that influence emotional expression in a way that is difficult to imitate (Scherer, 1986; Juslin and Laukka, 2001). Whatever the source of the differences, it is clear that future research on emotion processing would benefit from the availability of a database of authentic affective vocalizations.

Current understanding of auditory emotion processing has largely been derived from studying responses to adult emotional vocalizations. There are at least two other classes of emotional sounds that could inform our knowledge: vocalizations from younger humans and those from other species. For instance, it is hard to imagine sounds with greater biological salience than the vocal cues from infants. In the early postnatal period, vocalizations communicate degrees of distress, providing vital information for adults about an infant's needs (Soltis, 2004). Initially, the range of vocalizations an infant can produce is constrained by both the physiology of the developing vocal tract and the degree of muscular control over the vocal chords. As the infant develops, the range of vocal expressions becomes more sophisticated, incorporating more positive expressions such as laughter at around 4 months of age (Darwin, 1872; Sroufe and Wunsch, 1972; Nwokah et al., 1994). While adults respond to infant crying in the hope of terminating the sound, laughter promotes positive, rewarding social interactions between parent and infant. Around the same age, infants begin to "babble" (Oller and Eilers, 1988). Babbling comprises phoneme-like sounds, often repeated in meaningless strings, and is thought to be the precursor to language (Petitto and Marentette, 1991). When infants "babble," they tend to be neither distressed nor excited, but in an emotionally neutral state (Rothgänger, 2003).

Vocalizations from animals, particularly of the domestic kind, share a number of similarities with infant vocalizations. They convey information about an animal's current needs and often serve to initiate caregiving responses. In particular, domestic cat meows and dog whines are familiar, easily recognizable sounds that tend to elicit sympathetic and caregiving responses from human hosts (Pongrácz et al., 2005). For instance, domestic cats can effectively use their characteristic vocalizations, purrs, to solicit human care (McComb et al., 2009). Despite some functional overlap between animal and infant vocalizations, these sounds are clearly distinctive and easily differentiated from human vocalizations.

Here, we present a database of authentic affective bursts from all three of these sound categories: human adults, human infants and domestic animals. The major aim of this database is to make

available a set of multi-purpose, high quality, recognizable sounds for research across a broad range of domains. This new database has a number of advantages. First, the sounds are natural, from non-acted sources and therefore not affected by issues of authenticity. Second, in addition to adult vocalizations, this database includes a range of infant and domestic animal vocalizations. These sounds allow questions to be addressed such as, "What is special about human conspecific vocalizations?" and "How do infant vocalizations uniquely elicit parental care?" Third, categories of affective bursts considered to be especially effective in provoking emotion in the listener, such as laughs and cries (Scherer, 1995) along with less intense and emotionally neutral sounds, are included. These categories were included to allow comparison across a range of emotional vocalizations universally produced by infants and adults. In addition, these sound categories appear to be most readily recognized across cultures (e.g., Laukka et al., 2013). This paper describes the acquisition and standardization of this database of spontaneous, natural emotional vocalizations, referred to as the Oxford Vocal Sounds, "OxVoc" database (http://www.kringelbach.org/oxvoc).

METHODS

STIMULI

Infant vocalizations

Infant vocalizations were obtained from video recordings of infants filmed in their own homes in the UK during a play and feeding session with their primary caregiver. This play and feeding session included periods of interactive play, a mealtime, and brief separation of the infant and caregiver. These video recordings were collected as part of a previous project with the approval of the Oxford Psychiatric Research Ethics Committee. In addition, permission for a validation study testing responses to these sounds was obtained from the Central University Research Ethics Committee. Recordings from nine male and female infants were used for the current database (five male).

Exemplars of different types of vocalizations were selected from video recordings. Cry vocalizations ($n = 21$) occurred primarily when infants were separated from their caregiver. Laughter vocalizations ($n = 18$) occurred primarily when the infant played with their caregiver. Neutral "babbles" ($n = 25$) were selected from periods of calm interaction between infant and caregiver, often during mealtime or play. All infants were full-term, healthy, and aged between 6 and 8 months at the time of recording ($M = 6.7$ months, $SD = 0.9$). Audio recordings were extracted from these videos and clips of vocalizations, free from background noise, were selected.

Adult vocalizations

Adult vocalizations were obtained from video diary blogs and video product reviews consisting of individuals speaking directly to a camera (primarily sourced through YouTube.com). The individuals included were all from the UK or the Unites States. Individuals who uploaded this content were contacted in order to obtain their permission for the anonymous use of short clips of vocalizations for research purposes. Video blogs were manually searched for instances of emotional and non-emotional vocalizations. For the distress vocalizations ($n = 19$), clips were taken from videos posted by females aged approximately 18–30 years. The cause of crying was variable, but often occurred when describing upsetting life events. No similar available videos of male adults crying were found. Laughter and neutral vocalizations were obtained from both males and females, aged approximately 18–30 years, occurring naturally during conversation. Thirty exemplars of laughter vocalizations (15 male and 15 female) and thirty exemplars of neutral vocalizations (15 male and 15 female) were collected.

Animal distress vocalizations

Animal distress vocalizations were obtained from freely available internet sources (e.g., www.freesound.org). Recordings of pet cats and dogs posted by their owners were manually searched, and relevant segments extracted.

PHYSICAL MATCHING METHODS

The duration of stimuli was standardized using Audacity software (version 1.3.4-Beta, http://audacity.sourceforge.net/). Matching of all other physical parameters was completed using Adobe® Audition® software (CS5 v4.0, Adobe System Corporation, San Jose, CA). All selected stimuli were free from background noise and were cropped to 1500 ms in length. The onset and offset amplitude envelope was standardized across all stimuli by applying linear rise and fall times of 150 ms to the start and end of each clip ("Volume Envelope" function). Stimuli were matched for total root-mean-square (RMS) amplitude for each clip to -25 dBFS (decibels full scale). This method matched the overall intensity of the clip, while retaining variations in amplitude envelope across time.

Basic physical properties of vocalization stimuli including fundamental frequency, duration of vocal "bursts" within each stimulus and number of vocal bursts per stimulus, are presented in **Table 1**. Full details of physical parameters for each individual stimulus are included in the Supporting Information.

RATINGS OF STIMULI

A number of independent samples of participants were asked to listen to and rate the presented sounds on multiple dimensions. First, three independent samples of participants rated a large set of stimuli on three dimensions related to the emotional experience of the sounds (Study 1). Participants rated the distress of each sound, the mood of the individual producing the sound and also their own emotional reaction to the sound. Second, a further independent sample of participants rated the sounds on three additional dimensions: arousal of the sound producer, the valence of the sound, and the perceiver's motivation to respond to the sound (Study 2). The valence and arousal scales were selected to allow direct comparisons between ratings of the current stimuli and those from acted databases (Belin et al., 2008; Koeda et al., 2013; Lima et al., 2013). This latter "motivation to respond" rating was obtained because experiencing an emotional cue can be divided into its subjective appraisal ("liking") and also its incentive salience ("wanting"; Berridge and Kringelbach, 2008). This is of relevance to how the present sounds are perceived because previous work has demonstrated gender differences in "wanting" and "liking" of infant cues (Parsons et al., 2011).

Table 1 | Basic physical parameters of vocalization stimuli within the OxVoc database.

Stimulus category	Number of stimuli	F_0 (Hz), M (SD)	Number of vocal bursts, M (SD)	Mean burst duration (s), M (SD)
Infant cry	21	445.54 (84.81)	1.90 (1.09)	1.02 (0.50)
Infant neutral	25	347.34 (122.34)	1.88 (0.97)	0.93 (0.46)
Infant laugh	18	348.31 (87.95)	3.22 (1.40)	0.41 (0.22)
Adult cry	19	368.22 (94.83)	2.11 (0.32)	0.55 (0.130)
Adult neutral	30	228.13 (57.88)	1.00 (0.00)	0.91 (0.20)
Adult laugh	30	348.83 (104.12)	4.27 (1.78)	0.37 (0.29)
Animal distress	30	439.28 (101.53)	1.63 (0.76)	1.01 (0.42)
Total	173			

Values presented are averaged across stimuli within each stimulus category individual category.

PARTICIPANTS

Participants were recruited from the student and general UK population via posters and online advertisement. Inclusion criteria for participation were: currently resident in the UK, normal hearing, not currently taking medication affecting the brain and no self-reported mental health problems.

STUDY 1: THREE DIMENSIONS OF EMOTIONAL EXPERIENCE OF SOUND

Infant vocalizations

Sixty-one participants (30 male; aged $M = 27.58$ years, $SD = 6.87$) rated 64 infant vocalization stimuli (21 cries, 18 laughs, 25 babbles). For each stimulus, participants rated "How distressed do you think the baby was?" on a 7-point Likert scale, with 1 as "not distressed" and 7 as "very distressed." Participants also responded to "Please rate the mood of the baby" and "How did you find the sound?" on 9-point Likert scales, with -4 as "very negative" and $+4$ as "very positive." Participants listened to each stimulus once, after which they completed all three ratings before moving on to the next stimulus. The order of stimuli was pseudo-randomized across participants. Stimuli were presented using WinAmp software (Winamp v 5.61 Nullsoft, Inc., http://www.winamp.com), through Sony in-ear earphones (MDR-EX77LP). Responses were made by circling the appropriate value on the Likert scale using pen and paper. This procedure took approximately 15 min in total. For ease of comparison, data were converted to standardized scores (z-scores).

Adult and animal distress vocalizations

An independent sample of 24 participants (eight male; age $M = 27.67$ years, $SD = 11.72$) rated 49 adult and animal distress vocalizations. This set of stimuli consisted of female adult distress cries ("sobs" as opposed to screams, $n = 19$) and domestic animal distress vocalizations (cat meows, $n = 15$; dog whines, $n = 15$). A similar set of ratings to that described above for the infant vocalizations was collected. Participants responded to: (i) "how distressed do you think the adult/animal was?" from "very distressed" to "not distressed"; (ii) "please rate the mood of the animal," from "very negative" to "very positive"; and (iii) "how did you find the sound?" from "very negative" to "very positive."

Participants responded using computerized visual analog scales (VAS), implemented as vertical onscreen bars. A gray bar indicated the total range of possible responses, while a slightly narrower white bar, centered on top of the gray bar indicated the participants' current response. The height of the white bar could be altered by using the "UP" and "DOWN" arrows on the keyboard. After each stimulus was presented, individual scales appeared on the screen along with the question to be rated (see **Figure 1**). This task was self-paced; participants moved from one rating to the next by pressing the spacebar and the task lasted approximately 12 min in total. Participants' responses were again converted to standardized scores.
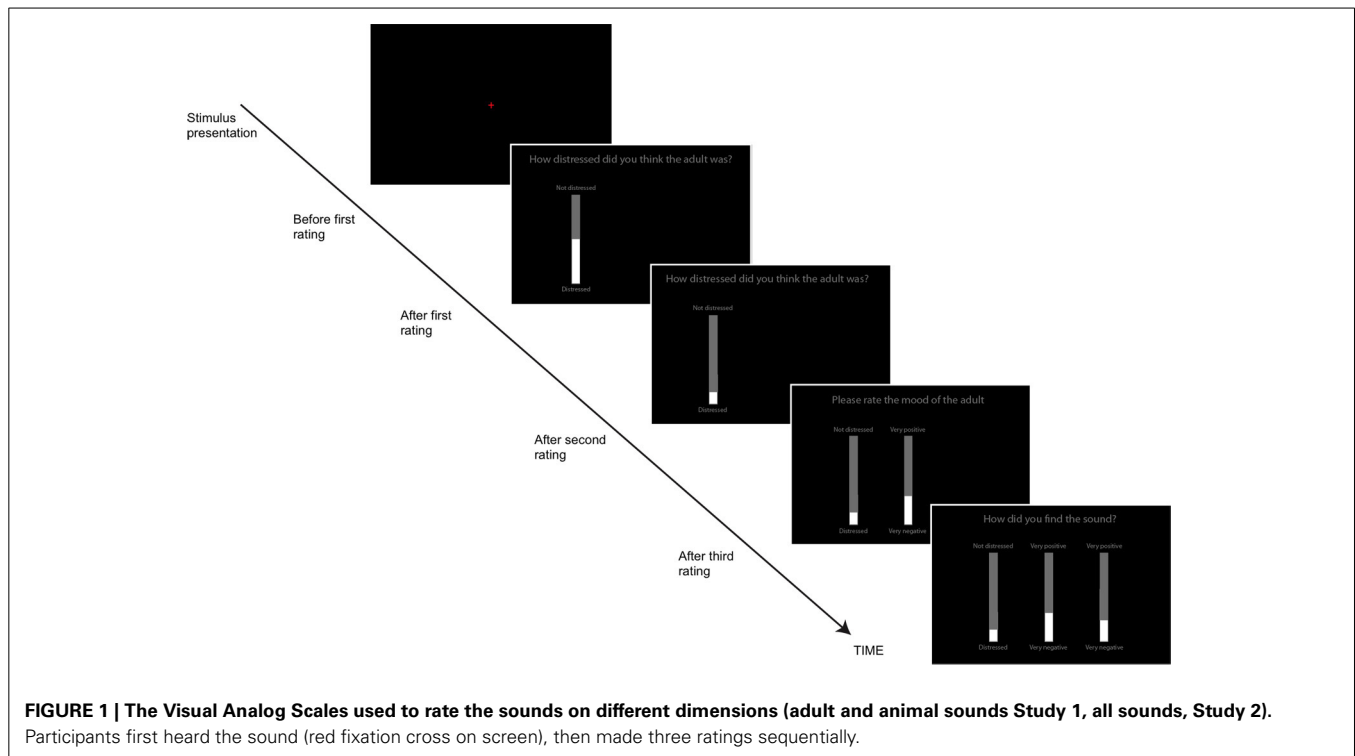
Other adult vocalizations: happy and neutral

An additional sample of 25 participants (seven male; age $M = 29.12$ years, $SD = 9.7$) rated 60 adult vocalizations. This set of stimuli consisted of equal numbers of male and female adult neutral vocalizations (15 each) and equal numbers of males and female happy (laugh) vocalizations (15 each). Participants again responded to: (i) "how distressed do you think the adult was?" from "very distressed" to "not distressed"; (ii) "please rate the mood of the adult," from "very negative" to "very positive"; and (iii) "how did you find the sound?" from "very negative" to "very positive." The ratings procedure was identical to that described for the adult and animal distress vocalizations.

RESULTS

Across all participant ratings, reliability was found to be very high, with intraclass correlations (ICC) > 0.90. Therefore, data is combined across participants throughout (similar to the procedure reported in Tottenham et al., 2009). The relationship between ratings on the three dimensions (distress, stimulus mood, participant mood) across all stimuli was assessed using Pearson correlations. Clear correlations in ratings emerged across all three rating dimensions. There was a strong positive correlation between perceived distress and perceived stimulus mood [$r_{(173)} = 0.92$, $p < 0.001$], a strong positive correlation between perceived distress and participant mood [$r_{(173)} = 0.94$, $p < 0.001$] and a strong positive correlation between stimulus mood and participant mood [$r_{(173)} = 0.97$, $p < 0.001$].

Given the high correlations between all ratings, a factor analysis was performed to investigate whether these measures loaded onto a single underlying variable. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.73 and Bartlett's test of sphericity was significant [$\chi^2_{(3)} = 858.64$, $p < 0.001$],

FIGURE 1 | The Visual Analog Scales used to rate the sounds on different dimensions (adult and animal sounds Study 1, all sounds, Study 2). Participants first heard the sound (red fixation cross on screen), then made three ratings sequentially.

justifying dimension reduction. A principal components analysis revealed the first factor explained 96.15% of the variance across ratings. This component is referred to as "valence."

Differences in "valence" across types of emotional vocalization were then assessed using a One-Way ANOVA, using stimulus type (positive, negative, or neutral) as a factor. This analysis demonstrated a significant main effect of stimulus type on perceived valence [$F_{(2, 172)} = 418.25$, $p < 0.001$, $r = 0.84$]. *Post-hoc* Bonferroni comparisons demonstrated significant differences in perceived valence in pairwise comparisons of positive (infant and adult laughter), negative (infant, adult, and animal distress vocalizations), and neutral (infant and adult babbles) stimuli (all $p < 0.001$). Comparisons of valence between individual stimulus categories were carried out in Study 2 where all stimuli were rated by a single group of participants.

STUDY 2: PERCEPTIONS OF AROUSAL, VALENCE, AND MOTIVATION TO RESPOND TO THE SOUNDS
Participants
Thirty-four participants (18 male) aged between 18 and 22 years ($M = 20$, $SD = 0.9$) took part in Study 2. Participants were recruited from a UK university population using online advertisements. All participants reported having normal hearing, were not parents, were not currently taking medication affecting the brain and had no self-reported mental health problems.

Method
Participants listened to all 173 stimuli and made three ratings for each sound using computerized Visual Analog Scales. Sounds were presented in a random order. The three ratings obtained were: "arousal," "motivation to respond," and "valence." For the

arousal measure, participants were asked, "Please rate how excited or calm you thought the sound was," from, "very excited" to "very calm" (consistent with Bradley and Lang, 2007). For the motivation to respond measure, participants were asked "How much do you want to respond to the sound?" from "very much" to "not at all." For the valence measure, participants were asked, "Please rate how happy or sad you thought the sound was," from "very happy" to "very sad." As for Study 1, participants listened to each stimulus once, after which they completed all three ratings before moving on to the next stimulus.
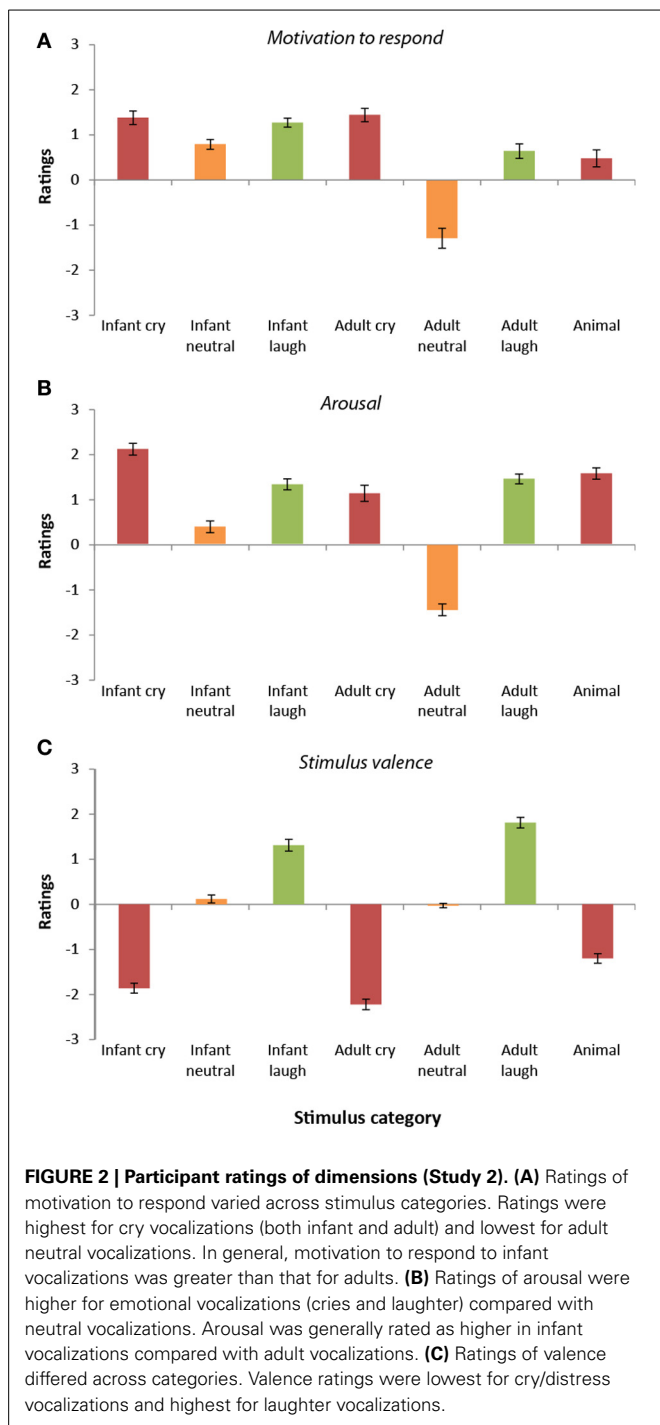
Results
Ratings of motivation to respond to the sounds, arousal, and valence of the sounds were examined using One-Way repeated measures ANOVAs. Differences between the specific stimulus categories were examined using paired samples *t*-tests with Bonferroni adjusted alpha levels of 0.007 (0.05/7).

Motivation to respond
Ratings of motivation to respond to the sounds differed significantly across the seven categories [$F_{(6, 192)} = 39.12$, $p < 0.0001$, $r = 0.41$]. Overall, infant vocalizations (cries, laughs, and babbles) elicited higher motivation to respond ($M = 1.15$, $SD = 0.54$) than the equivalent adult sounds [$M = 0.26$, $SD = 0.68$; $t_{(33)} = 6.22$, $p < 0.0001$, $r = 0.73$]. The adult neutral vocalizations stood apart from the other stimulus categories, receiving the lowest ratings of motivation to respond of all (all $p < 0.0001$).

Participant ratings of motivation to respond were highest overall for the infant cries, the adult cries, and the infant laughter (see **Figure 2A**). These three sound categories did not differ significantly from one another (all $p > 0.55$). Participants' ratings of

FIGURE 2 | Participant ratings of dimensions (Study 2). (A) Ratings of motivation to respond varied across stimulus categories. Ratings were highest for cry vocalizations (both infant and adult) and lowest for adult neutral vocalizations. In general, motivation to respond to infant vocalizations was greater than that for adults. **(B)** Ratings of arousal were higher for emotional vocalizations (cries and laughter) compared with neutral vocalizations. Arousal was generally rated as higher in infant vocalizations compared with adult vocalizations. **(C)** Ratings of valence differed across categories. Valence ratings were lowest for cry/distress vocalizations and highest for laughter vocalizations.

motivation to respond to infant crying were significantly higher than for the infant neutral sounds [$t_{(33)} = 4.09, p < 0.0001, r = 0.58$], the adult neutral sounds, [$t_{(33)} = 8.43, p < 0.0001, r = 0.83$], the adult laugh sounds [$t_{(33)} = 2.79, p < 0.009, r = 0.44$], and the animal sounds [$t_{(33)} = 8.48, p < 0.0001, r = 0.83$].

The adult cry sounds were also given higher "motivation to respond" ratings in comparison to the two other categories of adult vocalization [neutral, $t_{(33)} = 9.34, p < 0.0001, r = 0.85$; laughter, $t_{(33)} = 3.64, p < 0.0001, r = 0.54$], the animal sounds

[$t_{(33)} = 5.23, p < 0.0001, r = 0.67$] and the infant neutral sounds [$t_{(33)} = 3.99; p < 0.0001, r = 0.57$]. Participants' rated motivation to respond to infant laughter was significantly higher than for the infant neutral sounds [$t_{(33)} = 5.39, p < 0.0001, r = 0.68$], the adult neutral sounds, [$t_{(33)} = 8.43, p < 0.0001, r = 0.83$], the adult laugh sounds [$t_{(33)} = 2.79, p < 0.009, r = 0.44$], and the animal sounds [$t_{(33)} = 8.48, p < 0.0001, r = 0.83$]. Motivation to respond was rated similarly for the other stimulus categories ($p > 0.05$).

Arousal

Ratings of arousal within the sounds also differed significantly across the seven categories [$F_{(6, 192)} = 114.93, p < 0.0001, r = 0.61$, see **Figure 2B**]. Like the "motivation to respond" ratings, infant vocalizations (cries, laughs, and babbles) were rated as having higher arousal levels ($M = 1.27, SD = 0.62$) than the equivalent adult sounds [$M = 0.37, SD = 0.54; t_{(33)} = 9.54, p < 0.0001, r = 0.86$]. Again, in keeping with the "motivation to respond" ratings, arousal ratings were lowest in response to the adult neutral sounds and highest for the infant crying (all comparisons, $p < 0.0001$). Adult crying was rated as higher in arousal than the infant neutral and adult neutral vocalizations ($p < 0.0001$), but was rated similarly to the other vocalization subtypes (all $p > 0.08$). Infant neutral sounds were rated as having higher arousal levels than the adult neutral sounds [$t_{(33)} = 14.98, p < 0.0001, r = 0.93$], but significantly lower levels of arousal than all other stimulus categories (all $p < 0.0001$). Infant laughter, adult laughter and animal sounds were all rated similarly in terms of their arousal levels ($p > 0.05$).

Valence

Ratings of the valence of the sounds differed significantly across the seven categories [$F_{(6, 192)} = 200.00, p < 0.0001, r = 0.71$; see **Figure 2C**]. Adult cries received the most negative ratings of valence (all comparisons $p < 0.0001$), followed by infant cries (again, all comparisons, $p < 0.0001$). In contrast, adult laughter received the most positive ratings of valence (all comparisons, $p < 0.0001$), followed by infant laughter (all comparisons, $p < 0.0001$). Infant neutral vocalizations and adult neutral vocalizations received similar valence ratings [$t_{(33)} = 1.35, p = 0.86, r = 0.23$], and were both rated as significantly less negative than the animal distress vocalizations [infant neutral vs. animal vocalizations; $t_{(33)} = 9.7, p < 0.0001, r = 0.86$; adult neutral vs. animal vocalizations; $t_{(33)} = 8.68, p < 0.0001, r = 0.83$].

Gender differences

Independent samples t-tests showed that there were no gender differences in ratings of motivation to respond to the sounds [$t_{(32)} = 0.71, p = 0.48, r = 0.12$], arousal levels of the sounds [$t_{(32)} = 0.96, p = 0.38, r = 0.17$], or in valence of the sounds [$t_{(32)} = 1.83, p = 0.08, r = 0.31$].

Relationship between ratings of motivation to respond, arousal, and valence

Examining all 173 sound stimuli, there was a significant relationship between participants' ratings of their motivation to respond to the sound and ratings of the arousal of the sound ($r = 0.796, p < 0.0001$; two tailed). There was no significant relationship

between valence and motivation to respond ($r = -0.12$, $p = 0.09$) or between valence and arousal level ($r = -0.13$, $p = 0.11$).

Sounds were then divided into three categories, "infant," "adult," and "animal" to examine whether the relationships between the three different scales (motivation, arousal, valence) varied across categories. For the infant sounds, there was a strong positive correlation between ratings of arousal and motivation to respond ($r = 0.59$, $p < 0.0001$) and a strong negative correlation between arousal and valence ($r = -0.42$, $p < 0.0001$). However, there was no correlation between motivation to respond and valence ($r = -0.15$, $p = 0.25$). For the adult sounds, again there was a very strong positive correlation between ratings of arousal and motivation to respond ($r = -0.90$, $p < 0.0001$). In contrast to the infant vocalizations, there was no significant correlation between ratings of arousal and valence ($r = -0.19$, $p = 0.09$). Finally, as for the infant sounds, there was no significant correlation between ratings of motivation to respond and valence ($r = -0.15$, $p = 0.19$). For the animal sounds, as for the infant and adult sounds, there was a positive correlation between ratings of arousal and motivation to respond ($r = -0.43$, $p < 0.01$). There was also a negative correlation between ratings of arousal and valence ($r = -0.43$, $p < 0.01$), similar to the infant sounds. Finally, consistent with both the infant and adult sounds, there was no correlation between ratings of motivation to respond and valence ($r = -0.35$, $p = 0.06$).

DISCUSSION

The Oxford Vocal ("OxVoc") Sounds database consists of 173 natural, spontaneous non-verbal affective vocalizations from infants, adults, and domestic animals. For the human sounds, the vocalizations are further subdivided into three discrete emotional categories, positive "laughs," neutral "babbles," and negative, "cries." For the animal sounds, familiar distress sounds from two of our most common domestic companions, cats and dogs, are included. A detailed analysis is presented of how adults perceive these sounds on dimensions of distress, mood and listener mood (Study 1) as well as valence, arousal, and how motivated they are to respond to them (Study 2).

Adults rated the emotional content ("valence") of the stimuli in both Study 1 and Study 2 consistent with the experimenter defined categories. For example, while sounds categorized as cries were rated as negatively valenced, laughter was rated as positively valenced. In Study 2, some striking differences emerged across the adult, infant and animal categories. For measures of "valence," the adult stimuli received the most extreme ratings overall. Adult cries were rated more negatively than infant cries, and adult laughter was rated more positively than infant laughter. Both categories of human distress sound were rated more negatively than animal distress sounds. Neutral sounds from infants and adults were indistinguishable in terms of their rated valence.

The cause of the difference in ratings of valence in adult and infant vocalizations was not investigated in this study. It might be hypothesized that these differences are related to individual's attributions regarding the cause of vocalizations. For example, the cause of infant crying is often related to readily controllable environmental factors, such as hunger or separation from a caregiver. The cause of adult crying on the other hand, is often related to less controllable factors, such as loss or failure. It might be anticipated that, in line with attribution theory (Weiner, 1985), the greater "controllability" of infant crying compared with adult crying would impact the appraisal of these sounds. Indeed, previous research has demonstrated that even within infant vocalizations, individuals' perceived efficacy in caregiving, and presumably their ability to terminate infant distress, was related to how negatively infant cries were rated (e.g., Verhage et al., 2013). Despite equivalence in rated valence, adults were significantly more motivated to respond to the infant neutral vocalizations compared to adult neutral vocalizations. Furthermore, while adult positive and negative vocalizations were rated more extremely than the equivalent infant vocalizations, "motivation to respond" was greater overall for the infant sounds. These ratings suggest that sounds from an infant, be they distressed, neutral or positive, can especially motivate a response from an adult listener. This is consistent with a body of literature showing increased autonomic arousal and superior motor performance after listening to infant cries compared with other sounds, including adult cry sounds (Boukydis and Burgess, 1982; Zeskind and Collins, 1987; Bakermans-Kranenburg et al., 2012; Parsons et al., 2012). In addition, the potential for greater "controllability" of an infant's affective state, relative to that of an adult, may at least partly explain greater motivation to respond to these sounds.

Crying, from infants and adults alike, evoked the strongest ratings of motivation to respond. This is consistent with models of human empathy (e.g., Decety et al., 2012), which posit specific sensitivity to distress signals from others. While crying is relatively rare in adulthood (Zeifman, 2001), and often occurs when alone (Frey, 1985), it can serve to induce sympathy, empathy and comforting behavior and can help strengthen bonds between people (Vingerhoets et al., 2000).

The sound categories (infant, adult, and animal) also differed significantly with regards to their rated arousal. As for "motivation to respond" ratings, infant sounds were rated as having higher arousal levels than adult sounds. The relationship between emotion subcategory (e.g., cry) and rated arousal also differed dependent on who produced the sound (adult, infant). Generally the negatively valenced sounds were rated as highest in arousal, followed by the positively valenced sounds and then the neutral sounds. A comparable non-linear relationship between valence and arousal level in sound has been described previously (Bradley and Lang, 2007).

The lack of a linear relationship between ratings of arousal and valence was also evident in the correlational analysis from Study 2. The most robust finding from this analysis, occurring overall and for infant, adult and animal categories, was a positive correlation between arousal and motivation to respond. Sounds rated as "very excited" tended to elicit high "motivation to respond" ratings. For the infant and animal sounds, there was also a correlation between ratings of arousal and valence. However this relationship was not present in ratings of the adult sounds, and did not emerge in analyses of the sounds overall. These findings indicate that there are considerable differences in how adults perceive arousal and valence in similar vocalizations from adults, infants and animals.

We did not find any significant differences in men and women's ratings of the OxVoc sounds. While there is evidence that women show advantages in certain aspects of auditory emotion processing (e.g., emotion identification; for review, see Schirmer and Kotz, 2006), these effects are small in nature (e.g., Bänziger et al., 2012) and certainly not apparent across the board. Women have been shown to be more accurate at emotion recognition (Belin et al., 2008) and some differences in cortical processing of speech prosody have been reported (Schirmer et al., 2004, 2005). However, a body of evidence has shown no differences between men and women in ratings of non-verbal vocalizations (e.g., Hawk et al., 2009; Lima et al., 2013; Sauter et al., 2013), in response to speech prosody (e.g., Paulmann et al., 2008) or even in ratings of infant emotional cues (Parsons et al., 2011). Taken together, these findings suggest that any gender differences in response to emotional vocalizations, if they exist, are small in magnitude.

The OxVoc represents the first database of its kind to compile non-acted affective vocalizations from adults, infants and animals. The database includes human stimuli of variable valence, from distress cries, to laughter, to more neutral "babble" sounds. We have demonstrated significant differences in adults' motivation to respond to these sound categories, and also in perception of arousal and valence within the sounds. As well as a detailed analysis of overall responses to the sound categories, we provide information on adults' responses to each individual sound stimulus and details of the physical properties of the sounds (Supplementary information). OxVoc will be made available, with appropriate ethical approval, for non-commercial research at established academic institutions. Interested researchers can refer to the OxVoc web page (add link here) for detailed information about accessing the material.

There are, of course a number of limitations to the OxVoc database. While an effort was made to include stimuli from a variety of important and underexplored emotional categories (e.g., infant and adult vocalizations, domestic animal calls), the selection process was necessarily one of convenience. We cannot make any claims as to how representative our set of vocalizations is of the entire range that exists. In obtaining multiple ratings on the same sounds, we did not counterbalance the order of rated dimensions. This was done to ensure simplicity for participants, but could be changed in future studies.

Useful future additions to the current database, or indeed any future database, might include: (i) a set of adult male cry vocalizations; (ii) vocalizations from children, adolescents and adults of different ages, (iii) vocalizations from individuals with different language backgrounds. Ratings reported here were from UK residents, so replication in another country and culture would be of considerable interest. Ultimately, it is our intention to add these stimulus categories and ratings in the future and describe both as we have done for the current set of sounds.

In conclusion, there are several strengths to the OxVoc, rendering it a unique and versatile corpus that can be used for many different research purposes. First, these sounds were obtained from non-acted sources, and are as close to spontaneous, genuine expressions of emotion as possible. Second, the inclusion of vocal expressions from animals and infants provides comparison stimuli for use in cross-species and developmental studies. For instance, comparisons of responses to human and non-human vocalizations would allow the investigation of hypotheses regarding selective tuning to conspecific emotion. Third, ratings are available for these stimuli on a range of dimensions. The availability of such information means that the stimuli can be used for many different purposes, dependent on experimenter requirements. For instance, the sounds have already been used to examine the neural processing of infant cues (Parsons et al., 2013a,b), detection of functionally important physical characteristics of vocalizations (Young et al., 2012) and movement in response to sound salience (Parsons et al., 2012). Other possibilities include investigation of physiological responses to the sound categories (e.g., infant sounds) or to the specific dimensions (e.g., high arousal sounds) independent of category.

Fourth, the variability of adults' motivation to respond to these sounds provides intriguing opportunities to investigate the relationship between emotion perception, action and attributes of the vocalizer. Fifth, the associated website (http://www.kringelbach.org/oxvoc) provides a detailed description of the physical properties of each sound stimulus, along with cross-category descriptions.

REFERENCES

Audibert, N., Aubergé, V., and Rilliard, A. (2010). "Prosodic correlates of acted vs. spontaneous discrimination of expressive speech: a pilot study," in *Paper Presented at the 5th International Conference on Speech* (Chicago, IL).

Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., Riem, M. M. E., Tops, M., and Alink, L. R. A. (2012). Oxytocin decreases handgrip force in reaction to infant crying in females without harsh parenting experiences. *Soc. Cogn. Affect. Neurosci.* 7, 951–957. doi: 10.1093/scan/nsr067

Bänziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. *Emotion* 12, 1161–1179. doi: 10.1037/a0025827

Barker, P. L. (2013). *Recognition of Spontaneous and Acted Vocalizations of Emotion.* Masters dissertation, Universiteit van Amsterdam, Amsterdam: Digital Academic Repository.

Barkhuysen, P., Krahmer, E., and Swerts, M. (2007). "Cross-modal perception of emotional speech," in *Paper Presented at the International Congress of Phonetic Sciences* (Saarbrücken).

Belin, P., Fecteau, S., and Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* 8, 129–135. doi: 10.1016/j.tics.2004.01.008

Belin, P., Fillion-Bilodeau, S., and Gosselin, F. (2008). The Montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* 40, 531–539. doi: 10.3758/BRM.40.2.531

Berridge, K. C., and Kringelbach, M. L. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology* 199, 457–480. doi: 10.1007/s00213-008-1099-6

Boukydis, C. F., and Burgess, R. L. (1982). Adult physiological response to infant cries: effects of temperament of infant, parental status, and gender. *Child Dev.* 53, 1291–1298. doi: 10.2307/1129019

Bradley, M. M., and Lang, P. J. (2007). "The International Affective Digitized Sounds (2nd Edition; IADS-2): affective ratings of sounds and instruction

manual," in *Technical Report B-3,* ed Technical Report B-3 (Gainesville, FL: University of Florida).

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., and Weiss, B. (2005). A database of German emotional speech. *Proc. Interspeech* 5, 1517–1520. doi: 10.1.1.130.8506

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals:* London: John Murray. doi: 10.1037/10001-000

Decety, J., Norman, G. J., Berntson, G. G., and Cacioppo, J. T. (2012). A neurobehavioral evolutionary perspective on the mechanisms underlying empathy. *Prog. Neurobiol.* 98, 38–48. doi: 10.1016/j.pneurobio.2012.05.001

Drolet, M., Schubotz, R. I., and Fischer, J. (2012). Authenticity affects the recognition of emotions in speech: behavioral and fMRI evidence. *Cogn. Affect. Behav. Neurosci.* 12, 140–150. doi: 10.3758/s13415-011-0069-3

Ekman, P. (1993). Facial expression and emotion. *Am. Psychol.* 48, 384–392. doi: 10.1037/0003-066X.48.4.384

Elfenbein, H. A., and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol. Bull.* 128, 203–235. doi: 10.1037/0033-2909.128.2.203

Engberg, I. S., and Hansen, A. V. (1996). "Documentation of the Danish emotional speech database Aalborg DES," in *Internal AAU Report* (Center for Person Kommunikation).

Frey, W. H. (1985). *Crying, the Mystery of Tears*. Minneapolis, MN: Winston Press.

Hawk, S. T., van Kleef, G. A., Fischer, A. H., and van der Schalk, J. (2009). "Worth a Thousand Words": absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion* 9, 293–305. doi: 10.1037/a0015178

Jürgens, R., Hammerschmidt, K., and Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Front. Psychol.* 2:180. doi: 10.3389/fpsyg.2011.00180

Juslin, P. N., and Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion* 1, 381–412. doi: 10.1037/1528-3542.1.4.381

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Koeda, M., Belin, P., Hama, T., Masuda, T., Matsuura, M., and Okubo, Y. (2013). Cross-cultural differences in the processing of non-verbal affective vocalizations by Japanese and Canadian listeners. *Front. Psychol.* 4:105. doi: 10.3389/fpsyg.2013.00105

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* 84, 394–421. doi: 10.1016/j.biopsycho.2010.03.010

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W., et al. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizationscross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Front. Psychol.* 4:353. doi: 10.3389/fpsyg.2013.00353

Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., and Elenius, K. (2011). Expression of affect in spontaneous speech: acoustic correlates and automatic detection of irritation and resignation. *Comput. Speech Lang.* 25, 84–104. doi: 10.1016/j.csl.2010.03.004

Lima, C. F., Castro, S. L., and Scott, S. K. (2013). When voices get emotional: a corpus of nonverbal vocalizations for research on emotion processing. *Behav. Res.* 45, 1234–1245. doi: 10.3758/s13428-013-0324-3

McComb, K., Taylor, A. M., Wilson, C., and Charlton, B. D. (2009). The cry embedded within the purr. *Curr. Biol.* 19, R507–R508. doi: 10.1016/j.cub.2009.05.033

Nwokah, E. E., Hsu, H. C., Dobrowolska, O., and Fogel, A. (1994). The development of laughter in mother-infant communication: timing parameters and temporal sequences. *Infant Behav. Dev.* 17, 23–35. doi: 10.1016/0163-6383(94)90019-1

Oller, D. K., and Eilers, R. E. (1988). The role of audition in infant babbling. *Child Dev.* 59, 441–449. doi: 10.2307/1130323

Parsons, C. E., Stark, E. A., Young, K. S., Stein, A. S., and Kringelbach, M. L. (2013a). Understanding the human parental brain: a critical role of the orbitofrontal cortex. *Soc. Neurosci.* 8, 525–543. doi: 10.1080/17470919.2013.842610

Parsons, C. E., Young, K. S., Mohseni, H., Woolrich, M. W., Thomsen, K. R., Joensson, M., et al. (2013b). Minor structural abnormalities in the infant face disrupt neural processing: a unique window into early caregiving responses. *Soc. Neurosci.* 6, 1–7. doi: 10.1080/17470919.2013.795189

Parsons, C. E., Young, K. S., Murray, L., Stein, A., and Kringelbach, M. L. (2010). The functional neuroanatomy of the evolving parent-infant relationship. *Prog. Neurobiol.* 91, 220–241. doi: 10.1016/j.pneurobio.2010.03.001

Parsons, C. E., Young, K. S., Parsons, E., Dean, A., Murray, L., Goodacre, T., et al. (2011). The effect of cleft lip on adults' responses to faces: cross-species findings. *PLoS ONE* 6:e25897. doi: 10.1371/journal.pone.0025897

Parsons, C. E., Young, K. S., Parsons, E., Stein, A., and Kringelbach, M. L. (2012). Listening to infant distress vocalizations enhances effortful motor performance. *Acta Paediatr.* 101, e189–e191. doi: 10.1111/j.1651-2227.2011.02554.x

Paulmann, S., Pell, M. D., and Kotz, S. A. (2008). Functional contributions of the basal ganglia to emotional prosody: evidence from ERPs. *Brain Res.* 1217, 171–178. doi: 10.1016/j.brainres.2008.04.032

Petitto, L. A., and Marentette, P. F. (1991). Babbling in the manual mode: evidence for the ontogeny of language. *Science* 251, 1493–1496. doi: 10.1126/science.2006424

Pongrácz, P., Molnár, C., Miklósi, Á., and Csányi, V. (2005). Human listeners are able to classify dog (Canis familiaris) barks recorded in different situations. *J. Comp. Psychol.* 119, 136–144. doi: 10.1037/0735-7036.119.2.136

Rothgänger, H. (2003). Analysis of the sounds of the child in the first year of age and a comparison to the language. *Early Hum. Dev.* 75, 55–69. doi: 10.1016/j.earlhumdev.2003.09.003

Sauter, D. A., Eisner, F., Calder, A. J., and Scott, S. K. (2010a). Perceptual cues in nonverbal vocal expressions of emotion. *Q. J. Exp. Psychol.* 63, 2251–2272. doi: 10.1080/17470211003721642

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

Sauter, D. A., Panattoni, C., and Happé, F. (2013). Children's recognition of emotions from vocal cues. *Br. J. Dev. Psychol.* 31, 97–113. doi: 10.1111/j.2044-835X.2012.02081.x

Scherer, K. R. (1986). Vocal affect expression. A review and a model for future research. *Psychol. Bull.* 99, 143–165. doi: 10.1037/0033-2909.99.2.143

Scherer, K. R. (1994). "Affect bursts," in *Emotions: Essays on Emotion Theory,* eds K. R. Scherer and H. G. Wallbott (Hillsdale, NJ: Erlbaum), 161–193.

Scherer, K. R. (1995). Expression of emotion in voice and music. *J. Voice* 9, 235–248. doi: 10.1016/S0892-1997(05)80231-0

Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Commun.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5

Scherer, K. R., Banse, R., and Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross Cult. Psychol.* 32, 76–92. doi: 10.1177/0022022101032001009

Scherer, K. R., Ladd, D. R., and Silverman, K. E. A. (1984). Vocal cues to speaker affect: testing two models. *J. Acoust. Soc. Am.* 76, 1346–1356. doi: 10.1121/1.391450

Schirmer, A., and Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* 10, 24–30. doi: 10.1016/j.tics.2005.11.009

Schirmer, A., Striano, T., and Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport* 16, 635–639. doi: 10.1097/00001756-200504250-00024

Schirmer, A., Zysset, S., Kotz, S. A., and Von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *Neuroimage* 21, 1114–1123. doi: 10.1016/j.neuroimage.2003.10.048

Scott, S. K., Sauter, D., and McGettigan, C. (2010). "Brain mechanism for processing perceived emotional vocalization in humans," in *Handbook of Mammalian Vocalization: An Integrative Neuroscience Approach,* ed S. M. Brudzynski (London: Academic Press), 187–197.

Soltis, J. (2004). The signal functions of early infant crying. *Behav. Brain Sci.* 27, 443–458. doi: 10.1017/S0140525X0400010X

Sroufe, L. A., and Wunsch, J. P. (1972). The development of laughter in the first year of life. *Child Dev.* 43, 1326–1344. doi: 10.2307/1127519

Stevenson, R. A., and James, T. W. (2008). Affective auditory stimuli: characterization of the International Affective Digitized Sounds (IADS) by discrete emotional categories. *Behav. Res. Methods* 40, 315–321. doi: 10.3758/BRM.40.1.315

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: judgments

from untrained research participants. *Psychiatry Res.* 168, 242–249. doi: 10.1016/j.psychres.2008.05.006

Verhage, M. L., Oosterman, M., and Schuengel, C. (2013). Parenting self-efficacy is associated with cry perception, not autonomic responses, during a cry response task. *Parenting* 13, 253–265. doi: 10.1080/15295192.2013.832570

Vingerhoets, A. J. J. M., Cornelius, R. R., Van Heck, G. L., and Becht, M. C. (2000). Adult crying: a model and review of the literature. *Rev. Gen. Psychol.* 4, 354–377. doi: 10.1037/1089-2680.4.4.354

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychol. Rev.* 92, 548–573. doi: 10.1037/0033-295X.92.4.548

Young, K. S., Parsons, C. E., Stein, A., and Kringelbach, M. L. (2012). Interpreting infant vocal distress: the ameliorative effect of musical training in depression. *Emotion* 12, 1200–1205. doi: 10.1037/a0028705

Zeifman, D. M. (2001). An ethological analysis of human infant crying: answering Tinbergen's four questions. *Dev. Psychobiol.* 39, 265–285. doi: 10.1002/dev.1005

Zeskind, P. S., and Collins, V. (1987). Pitch of infant crying and caregiver responses in a natural setting. *Infant Behav. Dev.* 10, 501–504. doi: 10.1016/0163-6383(87)90046-4