When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing

César F. Lima[1,2], São Luís Castro[1], and Sophie K. Scott[2]

[1]Faculty of Psychology and Education, University of Porto, Portugal

[2] Institute of Cognitive Neuroscience, University College London, UK

December 2012

Author Note

This research was supported by grants from the Portuguese Foundation for Science and Technology (C.F.L. and S.L.C.), Bial Foundation (S.L.C. and C.F.L.), and Wellcome Trust (S.K.S.). We thank Ana Catarina Monteiro, for helping in data collection, and the four speakers, who generously agreed to produce the vocal stimuli and granted permission to include them in this database. Correspondence to: César F. Lima, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK. E-mail: c.lima@ucl.ac.uk

Abstract

Nonverbal vocal expressions, such as laughter, sobbing and screams, are an important source of emotional information in social interactions. However, the investigation of how we process these vocal cues entered the research agenda only recently. Here we introduce a new corpus of nonverbal vocalizations, which we recorded and submitted to perceptual and acoustic validation. It consists of 121 sounds expressing 4 positive emotions (achievement/triumph, amusement, sensual pleasure, and relief) and 4 negative ones (anger, disgust, fear, and sadness), produced by 2 female and 2 male speakers. For perceptual validation, a forced-choice task was used (n = 20), and ratings were collected for the eight emotions, valence, arousal, and authenticity (n = 20). We provide these data, detailed for each vocalization, for use by the research community. High recognition accuracy was found for all emotions, 86% on average, and the sounds were reliably rated as communicating the intended expressions. The vocalizations were measured for acoustic cues related to temporal aspects, intensity, fundamental frequency (F0), and voice quality. These cues alone provide sufficient information to discriminate between emotion categories, as indicated by statistical classification procedures; they are also predictors of listeners' emotion ratings, as indicated by multiple regression analyses. This set of stimuli seems a valuable addition to currently available expression corpora for research on emotion processing. It is suitable for behavioral and neuroscience research, and might as well be used in clinical settings for the assessment of neurological and psychiatric patients. The corpus can be downloaded from [insert link for supplementary materials].

Keywords: acoustic cues; auditory processing; expression corpus; emotion recognition; nonverbal vocalizations.

When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing

In social interactions we get information about others' affective states through a multitude of nonverbal signals, including facial expressions, body postures, touch and voice cues. The ability to perceive effectively these signals is a crucial component of our emotional competence (Scherer & Scherer, 2011), and it correlates with indicators of personal and social adjustment (e.g., Hall, Andrzejewski, & Yopchick, 2009). Understanding the behavioral, socio-cognitive and neural underpinnings of emotion perception in different channels is thus a topic of central importance. Because research tradition over the past decades has favored facial expressions, less is known about auditory emotion processing. This has been changing, with significant advances made in the last years in the study of auditory emotions, but mostly in what respects to cues communicated through emotionally inflected speech, i.e., emotional speech prosody (e.g., Bach, Grandjean, Sander, Herdener, Strik, & Seifritz, 2008; Banse & Scherer, 1996; Grandjean et al., 2005; Juslin & Laukka, 2001; Pell & Leonard, 2003; Scherer, 2003; Schirmer & Kotz, 2006). However, in addition to speech prosody, in social contexts we also use quite often a wide range of nonverbal expressions, such as laughter, sobs, moans, or screams. These non-linguistic sounds are very unlike speech regarding the underlying production mechanisms (Scott, Sauter, & McGettigan, 2010). They express rich emotional information, in what constitutes a primitive and universal form of communication (Sauter, Eisner, Ekman, & Scott, 2010), which closely

parallels the use of voice by many other animal species (e.g., Belin, Fecteau, & Bédard, 2004; Juslin & Laukka, 2003; Scherer, 1995). The systematic study of nonverbal emotional vocalizations has started only recently, and future research in this domain will benefit greatly from the availability of well-validated corpora of vocal expressions. Several corpora have been developed for facial expressions (e.g., Ekman & Friesen, 1976; Lundqvist, Flykt, & Öhman, 1998), body postures (e.g., de Gelder & van den Stock, 2011), speech prosody (e.g., Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005; Castro & Lima; Liu & Pell, in press), and multimodal materials (e.g., Bänziger, Mortillaro, & Scherer, 2012), but for nonverbal vocalizations they are rare. In this paper we present a new corpus of positive and negative nonverbal emotional vocalizations[1], which we recorded and submitted to detailed perceptual and acoustic validation.

Nonverbal vocalizations, even when presented without context, are at least as effective as facial expressions and speech prosody at communicating discrete emotional states. For instance, Schröder (2003) examined the ability of listeners to recognize 10 emotion categories in a range of vocalizations (admiration, threat, disgust, elation, boredom, relief, startle, worry, contempt, and hot anger), which included spontaneous nonverbal sounds, such as laughter, and more conventionalized affect emblems, such as “yuck”. Although there was variability across emotion categories, the average recognition accuracy in a forced-choice categorization task was very high, 81%. It was also shown that vocalizations’ orthographic transcriptions are much more variable for spontaneous non-linguistic sounds

[1] Nonverbal vocalizations are sometimes designated in the literature as “affect bursts” (Scherer, 1994). We do not use this expression in this paper because the sounds we recorded do not always display the features denoted by the word “burst” (e.g., rapid onsets, intense expressions, very brief durations). Furthermore, we focus on vocal cues alone, not in the co-occurrence of facial and vocal expressions, as the original definition of affect bursts implied.

than for conventionalized emblems. Furthermore, a second group of listeners was able to categorize emotions based solely on orthographic transcriptions of the vocalizations, i.e., without hearing the original vocalizations, suggesting that segmental structure may play a role in emotion recognition. This study analyzed mostly negative emotions, as typically occurs in emotion research, but nonverbal vocalizations are also effective at communicating a range of different positive emotional states. Sauter and Scott (2007) had participants from two language groups (English and Swedish) performing forced-choice and rating tasks on vocalizations intended to express five positive emotions – achievement/triumph, amusement, contentment, sensual pleasure, and relief. Both groups recognized these emotions accurately and rated them consistently as expressing the intended emotions. More recently, it was demonstrated that vocalizations' low-level acoustic attributes predict the way listeners perceive these five positive emotions in vocalizations, as well as the way they perceive negative emotions (anger, disgust, fear, and sadness) and surprise (Sauter, Eisner, Calder, & Scott, 2010). In a series of multiple regression analyses, it was observed that specific combinations of acoustic cues related to temporal aspects, amplitude, pitch and voice quality significantly predicted listeners' responses in an emotion rating task. Different constellations of predictors were found for different emotions, indicating that listeners make use of cues in emotion-specific manner, as had been previously shown in the context of speech prosody (Banse & Scherer, 1996; Juslin & Laukka, 2001). For instance, higher ratings on achievement/triumph were associated with lower minimum pitch, higher mean pitch, and more spectral variation, while higher ratings for disgust were associated with longer durations, lower spectral center of gravity, and more spectral variation. It was further observed that acoustic cues alone provide sufficient information to automatically categorize the vocalizations' emotion in discriminant analyses.

Even though research on nonverbal vocalizations is still in its infancy, the potential of this communicative channel to inform knowledge on auditory emotion processing, and on emotion communication more generally, has been confirmed in many behavioral (e.g., Bestelmeyer, Rouger, DeBruine, & Belin, 2010; Hawk, Kleef, Fischer, & Schalk, 2009), cross-cultural (e.g., Sauter et al., 2010), clinical (e.g., Dellacherie, Hasboun, Baulac, Belin, & Samson, 2011; Jones et al., 2011), electrophysiological (e.g., Jessen & Kotz, 2011; Sauter & Eimer, 2009), neuroimaging (e.g., Blasi et al., 2011; Peelen, Atkinson, & Vuilleumier, 2010; Warren et al., 2006; Banissy et al., 2006), and developmental studies (Hunter, Phillips, & MacPherson, 2010; Sauter, Panattoni, & Happé, 2012). Because nonverbal vocalizations are very effective at expressing different and recognizable positive affective states, they may provide a unique tool to counteract the bias towards negative emotions that characterizes most emotion research. They may **as well** be useful to shed new light on the functions and mechanisms of positive emotions. Moreover, the non-linguist nature of emotion vocalizations makes them appropriate to be used in different countries, on listeners with varied linguistic and cultural backgrounds. **Although there is an in-group advantage in the processing of vocalizations, with accuracy rates being higher for stimuli produced by members of our own culture vs. another, there is also evidence that emotions are recognizable across languages and cultures (Sauter et al. 2010; Sauter & Scott, 2007).**

We devised and validated a set of nonverbal vocalizations, which we make available to the research and clinical communities. This corpus of vocal stimuli is suitable for behavioral and neuroscience studies on auditory emotion processing, as well as to be included in neuropsychological assessment batteries to inspect higher-order pragmatic abilities. To the best of our knowledge, The Montreal Affective Voices (MAV) is the only published corpus of vocalizations available so far (Belin, Fillion-Bilodeau, & Gosselin,

2008)[2]. The MAV includes 90 short emotional sounds, consisting of the French vowel "ah", recorded to express 5 negative emotions (anger, disgust, fear, pain and sadness), 2 positive ones (happiness and pleasure), plus surprise and neutrality. Inter-participant reliability was very high, and all emotions were recognized well above chance level, 68% on average, as indicated by a measure of accuracy derived from intensity ratings. Notwithstanding, the fact that same linguistic sound ("ah") was used to record all the vocalizations may have favored a stimulus consistency that does not make justice to the highly variable structure that we encounter in real-life vocalizations. Additionally, the very different number of negative and positive emotions prevents systematic analyses of valence effects and limits the exploration of differentiated positive affective states. In the present study, we asked female and male speakers to produce nonverbal vocalizations on the basis of emotion labels and scenarios, with no instruction regarding the structure of the sounds they should produce. Four positive and 4 negative emotions were included: achievement/triumph, amusement, sensual pleasure, relief, anger, disgust, fear, and sadness. These emotion categories were previously shown to elicit high categorization accuracy and consistent ratings (Sauter & Scott, 2007; Sauter et al., 2010). The final corpus includes 121 sounds that were perceptually validated on the basis of the two most frequently used tasks in emotion literature: a forced-choice categorization, and a rating paradigm (between-subjects design). We collected ratings for the eight emotions, and also for valence, arousal, and authenticity. The vocalizations were acoustically measured for a battery of cues, and we examined the extent to which these cues can be used to automatically identify the stimulus emotion category and to predict listeners' judgments.

---

[2] In its original form, The Geneva Multimodal Expression Corpus (Bänziger & Scherer, 2010; Bänziger, Mortillaro, & Scherer, 2012) included brief emotion expressions consisting of sustained "aaa", but these were excluded from the final set of stimuli made available for researchers.

Method

Recording

Speakers

Two female (aged 27 and 33 years) and two male (aged 28 and 34 years) speakers produced the vocalizations. They were European Portuguese native speakers and did not have formal acting training. Two of them (one male and one female) had training in music, including singing lessons.

Procedure

The speakers were invited to participate in one recording session. They were provided with a list of the emotions they had to express, as well as with a list of illustrative real life scenarios typically associated with the experience of each emotion (see Appendix). After an initial briefing, the speakers read the emotion words and the corresponding scenarios and were asked to produce the vocal sounds they would make if they were experiencing that emotion. No guidance was provided as to the specific kind of sounds they should make, apart from general examples (e.g., some people laugh when they feel amused, or sob when they feel sad). They were told that they should not produce sounds with verbal content (e.g., “yuck”, “yippee”, “phew”), only nonverbal vocalizations. After a short familiarization phase, several different exemplars of the same category were recorded from each speaker (approximately 7). Extra recordings were made whenever the vocalizations were deemed to be unrecognizable (as exemplars of the intended emotions) by the experimenter (first author). They were told to try to sound as natural and spontaneous as possible. It has been acknowledged that some emotion categories can be expressed in distinct manners – emotion

families –, and this variation might be linked with distinct acoustic profiles (e.g., Banse & Scherer, 1996; Ekman, 1992; Scherer, 2003; Scherer et al., 2003); for instance, anger can be expressed in a hot explosive manner (rage), or in a cold controlled way. In our stimulus set there is variability regarding this issue: anger was produced mainly in a hot rather than in a cold manner, but there are also exemplars of more sustained anger; sadness vocalizations vary between a quiet and a mild form; and for fear, vocalizations vary between milder states and panic.

The vocalizations were recorded in the sound-insulated booth of the Speech Laboratory at University of Porto, using Pro Tools LE version 5.1.1 (Digidesign, Avid Technology) software and a high-quality microphone attached to an Apple Macintosh computer. Digitization was done at a 48-kHz sampling rate and 16-bit resolution. Individual files were prepared for each vocalization from each speaker – some exemplars were discarded at this phase, either because they were not appropriately recorded or because they were judged by the experimenters to be unrecognizable. This resulted in the production of 170 files; from these, a selection of the best ones was made on the basis of a small number of judges. A final set of 130 exemplars was selected and submitted to further perceptual validation (n = 16 for fear and pleasure, and n = 17 for the remaining categories). The sound files were normalized for maximum peak intensity using Sound Studio (version 4.2).

Validation

Participants

A total of 40 undergraduate students took part in the study. Twenty were assigned to the forced-choice task (mean age = 19.9; SD = 1.4; 19 females) and the other 20 to the rating task (mean age = 20.3; SD = 2; 19 females). They were recruited from University of Porto

and received course credits for their participation. Five participants in the forced-choice task had some degree of formal musical training, including instrumental practice (average years = 3.2; *SD* = 1.9); three participants in the rating experiment also had some degree of musical training (average years = 3.7; *SD* = 4.6). All participants were native speakers of European Portuguese and reported no hearing impairments or speech disorders, no psychiatric or neurological illnesses, and no head trauma or substance abuse.

Procedure

Forced-choice task. In this task, participants assigned one out of nine possible categories to each vocalization: achievement/triumph, amusement, anger, disgust, fear, sensual pleasure, relief, sadness, or "none of the above". Before starting the task, emotion labels were introduced alongside a hypothetical real life scenario (the scenarios were the same used for the recoding sessions, see Appendix). Participants were instructed to select the most appropriate category for each vocalization and to select the option "none of the above" every time the vocalization did not express any of the eight possible emotions. This response category was used to avoid producing artificially high recognition rates; this may happen when the task forces participants to use only the emotion categories predetermined by the researcher (for discussions on the limitations of forced-choice response formats, see e.g., Russell, 1994; Banse & Scherer, 1996; Scherer, 2003). Vocalizations were presented only once in random order through headphones and no feedback was given concerning response accuracy; participants underwent a short familiarization phase before starting the task (4 trials). Response options were presented in a fixed order – alphabetical order with the option "none of the above" in the end. They were always on the computer screen throughout the task and participants responded by moving and pressing the mouse button on the intended

category on the screen. An Apple MacBook Pro running SuperLab version 4.0 (Abboud et al., 2006) was used to control the presentation of the stimuli and to record responses.

Rating task. In this task, participants used 7-point scales, from 0 (minimum) to 6 (maximum), to rate how much each stimulus expressed each of the 8 emotions, and also to rate valence, arousal and authenticity. The full set of stimuli was presented 11 times in random order, and on each time participants rated the vocalizations on a single scale: there were 11 scales in total, 8 for each of the 8 emotions, and 3 for valence, arousal and authenticity. The rating scales for emotion categories were completed in a different order for each participant, and the rating scales for valence, arousal and authenticity were always completed before or after all the emotion scales (order balanced across participants). For the emotion scales, participants were instructed to judge how much a given emotion was expressed by each vocalization, from **0** to 6; as in the forced-choice task, scenarios were provided along with the emotion labels. For valence, participants indicated the extent to which each vocalization denoted a negative and unpleasant experience (**0** on the scale) or a positive and pleasant one (6 on the scale); for arousal, they indicated the extent to which each vocalization was produced by someone who was feeling sleepy and with no energy (**0** on the scale) or by someone who was feeling very alert and energetic (6 on the scale); for authenticity, participants were asked to evaluate the extent to which each vocalization was authentic, in the sense that it resembled the ones we encounter in our daily life. Just as accuracy rates, this criterion may be relevant for the actual use of the corpus in research, e.g., for stimulus selection. It is close to the "believability" index used by Bänziger and Scherer (2010), in which participants rated the "capacity of the actor to communicate a natural emotional impression" (p. 12). This index elicited relatively low levels of agreement across participants in this study, probably reflecting the high degree subjectivity involved in making

such a judgment. Responses were provided by pressing the appropriate button on a seven-button response pad, model RB-730, from Cedrus Corporation, attached to a computer running SuperLab; numbers from **0** to 6 were assigned to each button.

Selection

Mean accuracy rates were computed for each vocalization, as were the mean ratings provided for the intensity scales. The vocalizations were included in the final corpus only if (1) the percentage of categorizations was higher for the intended emotion than for all the non-intended ones and if they (2) were rated higher in the intended scale vs. all the non-intended ones. These criteria lead us to discard 9 stimuli; 121 were included in the final database. These were submitted to detailed perceptual and acoustic analyses, as presented below. The mean number of stimuli per emotion is 15 (*SD* = 1.7).

Results and discussion

The database that we provide here consists of 121 nonverbal emotional vocalizations expressing four positive emotions – achievement/triumph, amusement, pleasure and relief – and four negative ones – anger, disgust, fear and sadness –, **as** recorded by four speakers, two women and two men. Detailed perceptual and acoustic characteristics for each vocalization can be found in the supplementary materials. This information and the set of vocalizations itself are available for download at [insert link for supplementary materials].

Recognition accuracy

Inter-participant reliability in categorizations was very high (Cronbach's α = .972), suggesting that these vocalizations produce reliable behavioral responses. Table 1 presents

average accuracy rates for each emotion category in diagonal cells in bold, as well as the distribution of inaccurate categorizations in rows. Emotion recognition accuracy was high, 86% on average, ranging between 70% (for fear) and 97% (for disgust). For all emotions, one sample *t* tests confirmed that accuracy rates were higher than what would be expected by chance alone (12.5%): achievement, $t(12) = 20.48$; amusement, $t(15) = 73.3$; pleasure, $t(16) = 18.76$; relief, $t(15) = 29.75$; anger, $t(11) = 12.09$; disgust, $t(15) = 71.05$; fear, $t(14) = 15.89$; sadness, $t(15) = 22.5$ (all significant after Bonferroni correction, ps $< .00001$). This is evidence that our set of vocalizations was effective at communicating the intended emotions. The obtained accuracy rates are close to or higher than those reported in other studies on nonverbal vocalizations. For instance, Schröder (2003) obtained 81% correct on average for 10 emotion categories; Sauter et al. (2010) obtained 70% correct, also for 10 categories; Bänziger and Scherer (2010) obtained 40% correct for 18 emotions. The very high recognition rates observed for disgust is consistent with previous findings on nonverbal vocalizations (e.g., Belin et al., 2008; Sauter et al., 2010), and it stands in sharp contrast to the difficulties usually found for this emotion in the context of speech prosody (e.g., Banse & Scherer, 1996; Castro & Lima, 2010; Scherer et al., 1991). The differential ease with which disgust is recognized in nonverbal vocalizations and prosody suggests that there might be dissociations in the mechanisms supporting emotion processing, even within the auditory modality (for a dissociation between emotion recognition in prosody and music, Lima, Garrett, & Castro, under review).

Concerning the pattern of errors, the most common ones included achievement vocalizations categorized as expressing amusement, and anger vocalizations categorized as expressing disgust. Such confusions occurred probably because these emotion pairs are highly similar both in terms of valence and arousal (see Table 3). There are also acoustic

similarities that may have contributed to these confusions: achievement and amusement are close in terms of intensity standard deviation and F0 maximum and range; anger and disgust are close in terms of intensity standard deviation, F0 minimum, and spectral center of gravity (see Table 4). **The reserve was not observed, i.e., amusement and disgust were not confused with achievement and anger, respectively, probably because amusement and disgust are highly distinctive and unambiguous vocal emotions, as indicated by the very high categorization rates obtained, both above 95% correct.** For the remaining emotions, there were no salient trends in the distribution of errors. As can be seen on Table 1, only a small proportion of responses were provided for the option "none of the above" (5%, on average). This suggests that, in most cases, participants found that one of the eight emotion categories reflected appropriately the communicated emotion. "None of the above" responses were highest for fear (13%), probably **due to** the relative ambiguity in the recognition of this emotion; it elicited the lowest accuracy rates.

Table 1. Distribution of responses (percentages) for each emotion category. Diagonal cells in bold indicate accurate categorizations (standard errors in parenthesis).

Stimulus type	Response								
	Achievement	Amusement	Pleasure	Relief	Anger	Disgust	Fear	Sadness	None
Achievement	**77.7** (3.2)	15.8	1.2	2.3	0	0.4	0	0	2.7
Amusement	0	**95.9** (1.1)	0.6	0.6	0	0	0.3	1.9	0.6
Pleasure	2.1	0.3	**85.9** (3.9)	2.9	0	1.5	1.5	0.6	5.3
Relief	1.9	0	5.6	**86.3** (2.5)	0.6	0	0.3	0.3	5
Anger	0.8	0.8	0.4	0	**78.3** (5.4)	8.3	5.8	0	5.4
Disgust	0	0	0	0	0.3	**96.7** (1.2)	0.3	0.3	2.5
Fear	1.7	0.3	1	6	1.7	6	**70** (3.6)	0.3	13
Sadness	0	1.9	0	0.6	0.3	0.3	5	**89.7** (3.4)	2.2

Ratings

Emotion scales

Inter-participant reliability in ratings on emotion scales was also very high (Cronbach's α = .966), further indicating that our set of vocalizations produces reliable responses. Table 2 depicts the average ratings provided on each of the 8 emotion scales, for each stimulus category (for ease of interpretation, raw ratings 0-6 were converted to 0-100). As can be seen in diagonal cells in bold, all emotion categories were rated higher on the intended scale than on the other scales. This is an expected finding, given that only vocalizations rated higher on the intended vs. non-intended scales were included in the database. Statistical support for this was provided by a series of ANOVAs, one for each emotion scale (stimulus category as between-subjects factor), and by planned comparisons contrasting the ratings on the intended scale with the ratings on the other 7 scales [main effect of category, F (7,113) = 252.29 for achievement, 357.28 for amusement, 412.12 for pleasure, 329.61 for relief, 251.9 for anger, 352.06 for disgust, 145.37 for fear, and 288.34 for sadness; all ps < . 0001; all planned contrasts were significant after Bonferroni correction, $p < .001$]. We found a significant correlation between intensity ratings on the intended scales and accuracy rates on the forced-choice task (r = .58, p <. 0001), suggesting that the more intense are the vocalizations, the better they are recognized. Ratings on non-intended scales were in general highest for emotion scales of the same valence as the "correct" one (see Table 2).

To investigate if a smaller number of variables could significantly explain variability in participants' ratings, we computed a principal components analysis on the mean ratings provided for each stimulus on the 8 emotion scales. This analysis revealed two factors with

eigenvalues over 1, which accounted for 43.1% and 15.2% of the ratings' variance, respectively. These two factors are likely to reflect the affective dimensions valence and arousal: Factor 1 correlated strongly with participants' valence ratings (r = .98, p < .0001), and Factor 2 correlated moderately with arousal ratings (r = -.31, p < .01).

We also extracted a derived measure of accuracy from the raw ratings. For each vocalization, when the highest of the 8 ratings was provided on the "correct" scale, the response was considered as a correct categorization; otherwise, the response was considered as an incorrect categorization. These rates are depicted in the last column of Table 2. Such a derived accuracy index has been used in previous studies on emotion recognition (e.g., Adolphs, Damasio, & Tranel, 2002; Lima & Castro, 2011; Vieillard et al., 2008; Belin et al., 2008; Gosselin et al., 2005). Average derived accuracy was 70%, ranging between 45% (for achievement) and 86% (for disgust). These rates are as high as ones obtained for the MAV, 68% (Belin et al., 2008). They correlate with accuracy rates in the forced-choice task (r = .53, p < .001).

Table 2. Intensity ratings (scaled 0-100) and derived accuracy for each emotion category. Diagonal cells in bold indicate ratings on the intended emotion scale (standard errors in parenthesis).

Stimulus type	Rating Scale								Derived Accuracy
	Achievement	Amusement	Pleasure	Relief	Anger	Disgust	Fear	Sadness	
Achievement	**76.3** (3.5)	58.7	30.8	33.8	1	2.8	2.8	1.2	45 (2.5)
Amusement	34.2	**78.5** (3.1)	25.7	22.2	1.3	1.7	1.3	2.7	64.1 (2.8)
Pleasure	18.5	16.7	**81.2** (2)	17.7	5	4.3	4.5	5.2	82.4 (3.3)
Relief	15.8	10	23.3	**77.1** (2.1)	13.8	7.8	14.8	9	75.9 (2.9)

Anger	7.2	3.3	3.7	4.2	**79.5** (3.5)	33.7	22	15.5	65.8 (3.8)
Disgust	2.7	2.5	2.3	3	17.3	**84.6** (1.9)	11.3	10.5	85.9 (1.9)
Fear	11.7	7	12.7	17.7	19.2	25.3	**70.5** (2.9)	21.3	63.7 (3.8)
Sadness	4.8	3.5	3.3	7	10.7	17.2	33	**74.6** (2.5)	70.9 (3.7)

Valence, arousal, and authenticity

Inter-participant reliability was very high for valence (α = .982) and arousal ratings (α = .95); it was lower, though satisfactory, for authenticity ratings (α = .836). Lower agreement rates for authenticity ratings probably reflect the higher degree of subjectivity involved in defining and evaluating this dimension. Relatively low alpha values were also obtained by Bänziger and Scherer (2010) for believability (.67), and by Bänziger and colleagues (2012) for authenticity (.50) and plausibility (.48). Table 3 displays the average valence, arousal, and authenticity ratings for each emotion category (raw ratings 0-6 were converted to 0-100). Concerning valence, as expected, vocalizations expressing achievement, amusement, pleasure and relief were rated as being positive (values > 50), whereas vocalizations expressing anger, disgust, fear and sadness were rated as being negative (values < 50). Concerning arousal, vocalizations communicating achievement, amusement and anger were rated as being the most arousing, disgust and fear as intermediately arousing, and pleasure, relief and sadness as relatively less arousing. These results are consistent with the ones obtained by Sauter et al. (2010) for a different set of vocalizations. Significant variability on valence and arousal scales across emotion categories was confirmed by two ANOVAs [main effect of category for valence, $F(1,113) = 356.36$, $p < .001$, $\eta_p^2 = .96$; and for arousal, $F(1,113) = 37.84$, $p < .001$, $\eta_p^2 = .7$].

As for authenticity, average ratings were 54, ranging between 37 (for anger) and 78 (for amusement). Significant variability across categories was revealed by an ANOVA,

$F(1,113) = 38.36$, $p < .001$, $\eta_p^2 = .7$. The average authenticity ratings obtained here are roughly comparable to the plausibility (59/100), authenticity (63/100) and believability (66/100) ratings obtained by Bänziger and colleagues in their multimodal expression corpus (Bänziger & Scherer, 2010; Bänziger et al., 2012). As can be seen in Table 3, positive emotions were generally rated as being more authentic than negative ones. Indeed, a Pearson's correlation analysis unveiled a significant association between authenticity and valence ratings, so that higher authenticity ratings were observed for more positive vocalizations ($r = .57$, $p < .0001$). It might be that our index of authenticity partly reflects the frequency with which participants encounter the vocalizations in daily life and, arguably, positive vocalizations are more frequently encountered in normal communicative interactions than negative ones (the instruction was "evaluate the extent to which each vocalization was authentic, in the sense that it resembled the ones we encounter in our daily life"). The public expression of strong negative emotions is constrained by social norms and self-regulation. It is relatively uncommon that we get, for example, angry or disgusted to the point of not being able to inhibit the production of a corresponding emotional vocalization. However, although they may be less frequent, the fact the all negative emotions were very well recognized is clear-cut evidence that they communicate socially-relevant meanings which we are able to perceive. In fact, in two event-related potential studies, Sauter and Eimer (2010) showed that the brain is very quick at processing negative vocalizations, namely fearful ones. A moderate but significant correlation was also found between authenticity ratings and accuracy in the forced-choice task ($r = .32$, $p < .0001$), such that the more accurately vocalizations are recognized, the more they are perceived as authentic. Bänziger and Scherer (2010) also found a positive correlation between believability and accuracy. They took it as evidence against the argument often made that highly recognizable acted stimuli are very stereotypical and do not

produce a realistic or authentic impression. In contrast, authenticity ratings did not correlate with arousal; they also did not correlate with ratings on the intended emotion scale nor with derived accuracy scores (rs < .16, ps > .08). In future studies, it might be fruitful to split authenticity judgments into two different dimensions: frequency/familiarity (how frequently we encounter the vocalizations in our daily life) and realism (how much it seems that the person who produced the vocalization was experiencing the corresponding emotion).

Table 3. Arousal, valence and authenticity ratings (scales 1-100) for each emotion category (standard errors in parenthesis).

Stimulus type	Valence	Arousal	Authenticity
Achievement	87.4 (2.3)	88.4 (2.1)	47.7 (2.4)
Amusement	84.8 (1.6)	79.5 (2.7)	77.9 (2.1)
Pleasure	74.3 (1.5)	38.8 (2.3)	57.3 (2.1)
Relief	55.8 (1.3)	36.7 (2)	64.5 (1.7)
Anger	16.7 (1.6)	70.8 (4.3)	36.7 (2.5)
Disgust	13.8 (0.9)	56.6 (3.3)	42.9 (2.3)
Fear	31.1 (2.2)	62.2 (5)	46.1 (1)
Sadness	15.4 (1.8)	43.1 (2.5)	49.2 (2.6)

Acoustic analyses

Acoustic characteristics of the 121 vocalizations included in the database were extracted using Praat (Boersma & Weenink, 2009). Each vocalization was measured regarding major voice cues related to temporal aspects, intensity, fundamental frequency (F0) and voice quality. Specifically, we analyzed 12 acoustic parameters: duration (in ms); intensity mean and standard deviation (dB); number of amplitude onsets; F0 mean, standard deviation, minimum, maximum, and range (Hz); spectral center of gravity and standard deviation (Hz); and harmonics-to-noise ratio (dB). The number of amplitude onsets gives an

estimation of the number of "syllables" (that is, separate perceptual centers) in a vocalization (Morton, Marcus, & Frankish, 1976). We counted them using an algorithm which detects local rises in the smoothed amplitude envelope (Cummins & Port, 1998; Scott, 1993). First, the signal of each vocalization was band-pass filtered (Hanning filter centered at 2.2 kHz with a band-width of 3.6 kHz), full-wave rectified, and smoothed (Hanning low-pass filter with an 8-Hz cut-off), and then the first derivative of the smoothed envelope was obtained. Onsets were defined as points in time at which (1) a defined threshold in the amplitude envelope was exceeded, and (2) the derivative curve had a positive value. The acoustic parameters, averaged across the four speakers, are depicted for each emotion category in Table 4. All cues showed significant variability across emotions, as indicated by ANOVAs computed for each cue [main effect of category, $F(7,113) = 3.71$ for duration, 11.62 for intensity mean, 5.38 for intensity standard deviation, 14.1 for amplitude onsets, 11.45 for F0 mean, 3.51 for F0 standard deviation, 10.09 for F0 minimum, 3.84 for F0 maximum, 2.89 for F0 range, 11.35 for spectral center of gravity, 14.94 for spectral standard deviation, and 58.68 for harmonics-to-noise ratio; all significant after Bonferroni correction, ps < .009]. This confirms that the intended emotions in our set of vocalizations were communicated through variations in a wide range of acoustic cues.

Table 4. Acoustic characteristics of nonverbal vocalizations for each emotion category.

Stimulus type	Acoustic cue

	Duration (ms)	Int_M (dB)	Int_{SD} (dB)	Amp Onsets	$F0_M$ (Hz)	$F0_{SD}$ (Hz)	$F0_{MIN}$ (Hz)	$F0_{MAX}$ (Hz)	$F0_{RANGE}$ (Hz)	$Spectral_{COG}$ (Hz)	$Spectral_{SD}$ (Hz)	H/N_{RATIO} (dB)
Achievement	1222	81.9	10	1.6	451.3	114.4	229.2	635.2	406	835.2	645.3	22.3
Amusement	982	73.3	9.7	4.1	327.4	133.4	161.8	603.9	442.1	1046	1206.3	6
Pleasure	1257	79.6	7.7	2.3	178.7	62.6	109.1	369.1	260.1	238.6	233.9	21.7
Relief	1034	69.6	10.2	1.4	468.5	135.6	274.4	673.6	399.2	952.3	1349.2	5.8
Anger	931	77.7	8.3	1.7	244.3	106.7	105.7	462.8	357.1	1019.5	1017.4	6.6
Disgust	1136	74.8	8.7	3.9	313.1	157.1	121.4	575.6	454.3	1011.6	1375.1	10.2
Fear	876	72	13	1.4	420.1	63.1	322.2	537.5	215.3	948.6	1116.1	9.3
Sadness	1087	70.4	9.3	4.6	351.8	123.1	197.7	653.9	456.2	802.3	1150.6	9.1

Note. Int = Intensity, Amp = Amplitude, COG = center of gravity, H/N = harmonics-to-noise.

Acoustic cues predict vocalizations' category membership

We used statistical classification procedures to inspect whether acoustic cues alone provide sufficient information to predict the category membership of vocalizations in our database. The dependent variable of the models was the vocalization's emotion category, and the independent variables were the acoustic cues. To keep the set of independent variables small and to avoid collinearity, we tried as much as possible to exclude cues that were strongly intercorrelated ($r > .6$). The following cues were included: duration; intensity mean and standard deviation (dB); number of amplitude onsets; F0 mean and standard deviation; spectral center of gravity; and harmonics-to-noise ratio (F0 minimum, maximum and range, and spectral standard deviation were excluded). In addition to a standard discriminant analyses, we carried out a "jackknife" analysis. The more conservative jackknife method was employed because standard discriminant analyses can inflate the accuracy of the model (Sauter et al., 2010). This procedure predicts each stimulus' category on the basis of discriminant functions derived from all other stimuli whose categories were known to the model; each stimulus is analyzed separately. The standard discriminant analysis was able to correctly classify the emotion category of 71.1% of the stimuli [Wilks's $\lambda = .032$; $F(56, 576) = 9.22$, $p < .0001$], and the jackknife analyses was able to classify correctly 59.5% of the

stimuli (for a chance-level of 12.5%). Concerning specific emotions, prediction accuracy in the discriminant analysis was 69.2% for achievement, 56.3% for amusement, 88.2% for pleasure, 87.5% for relief, 83.3% for anger, 62.5% for disgust, 66.7% for fear, and 56.2% for sadness; in the jackknife analysis, it was 61.5% for achievement, 50% for amusement, 76.5% for pleasure, 68.8% for relief, 83.3% for anger, 31.3% for disgust, 66.7% for fear, and 37.5% for sadness. Both models' overall performance was slightly below human listeners' (86%), but still it was well above what would be expected by chance, as indicated by one sample *t* tests [$t(7) = 12.44$ for the standard discriminant analysis, and 7.22 for the jackknife analysis, ps < .001]. Therefore, the acoustic attributes of our set of vocalizations provide enough detail to automatically categorize emotions accurately. The question of how listeners used acoustic cues to provide subjective emotion ratings remains open, though. This relationship between acoustic cues and perceptual judgments was examined in a series of multiple regression analyses.

Acoustic cues predict listeners' ratings

Previous studies on speech prosody and nonverbal vocalizations have shown that acoustic cues predict subjective emotion judgments (e.g,. Banse & Scherer, 1996; Juslin & Lukka, 2001; Lima & Castro, 2010; Sauter et al., 2010). We conducted one standard (simultaneous) multiple regression analysis for each emotion, and also for valence, arousal, and authenticity. Acoustic cues were taken as predictors (these were the same used for the classification analyses, see above), and the dependent variable was the listeners' raw ratings on the respective scale. These analyses aimed at unveiling whether constellations of acoustic cues, in our set of vocalizations, are able to significantly predict participants' judgments on the different scales. The main findings are presented in Table 5 in terms of beta weights and

proportion of variance explained by the acoustic measures (adjusted R^2). In what respects to emotion scales, all multiple regressions were significant, indicating that listeners' ratings on each emotion category could be reliably predicted from the low-level vocalizations' physical attributes. The proportion of explained variance ranged between .06, for amusement, and .41, for sensual pleasure. A low proportion of explained variance for amusement was also reported by Sauter et al. (2010) for a different set of vocalizations. An inspection of Table 5 suggests that listeners' ratings were driven by many cues – no less than three cues reached significant beta weights for all emotions, except for pleasure, for which two cues were significant predictors. Additionally, the specific constellation of predictors was unique for each emotion, showing that listeners relied on different acoustic profiles to perceive different emotions. This is consistent with what have been described for speech prosody (Banse & Scherer, 1996; Juslin & Laukka, 2001). It thus seems that specific cues are particularly determinant for specific emotions in nonverbal vocalizations. Ratings for achievement/ triumph were predicted by increments in intensity mean and standard deviation, as well as in F0 mean and harmonics-to-noise ratio (indicating decreased noise in the vocalization); for amusement, ratings were predicted by higher intensity mean and standard deviation, and by higher number of amplitude onsets; ratings for pleasure were predicted by increased harmonics-to-noise ratio and lower number of amplitude onsets; ratings for relief were predicted by longer duration, higher F0 mean, and by lower number of amplitude onsets, spectral center of gravity and harmonics-to-noise ratio; for anger, ratings were predicted by higher intensity mean and harmonics-to-noise ratio, and by lower number of amplitude onsets and F0 mean; ratings for disgust were predicted by higher F0 standard deviation and spectral center of gravity, and by lower F0 mean; for fear, ratings were predicted by increased intensity standard deviation and F0 mean, and by shorter duration and lower F0 standard

deviation; finally, ratings for sadness were predicted by increased number of amplitude onsets and higher F0 mean, and by decreased intensity mean. There are both similarities and differences between these emotion-specific profiles of acoustic predictors and the ones described previously by Sauter and colleagues (2010). For instance, as in our set of vocalizations, they found that increased F0 mean predicts significantly ratings for achievement/triumph and relief. On the other hand, they observed that variations in the spectral center of gravity predicted ratings for pleasure and fear, whereas in the present study this cue did not reach significant beta weights for these emotions. It might be that there are slight differences in how vocalizations are produced and perceived by speakers of different languages – in their study, speakers and listeners were native British English speakers, and in the present study they were European Portuguese. In a similar vein, Sauter and Scott (2007) observed that, although both British and Swedish listeners had a broadly similar performance in categorizing and rating vocalizations produced by British speakers, there were also differences, with Swedish listeners showing lower categorization accuracy.

Multiple regressions for valence and arousal judgments were also significant, as can be seen in Table 5. Concerning valence, the model accounted for .13 of the variance in ratings. The only predictor reaching marginally significant beta weights was harmonics-to-noise ratio, with higher values in this acoustic feature being associated with higher positive valence ratings. As for arousal, .50 of the variance in the ratings was explained by the acoustic predictors: vocalizations were perceived as more arousing the higher was their intensity mean and standard deviation, and their spectral center of gravity. In contrast, authenticity ratings were not predictable from the vocalizations' acoustic features. This scale arguably reflects to a large extent a subjective dimension, possible more related to the

frequency with which different vocalizations are encountered in daily life than with the physical attributes per se.

Table 5. Summary of results of multiple regression analyses for each rating scale (rows), against acoustic cues (columns). Values represent beta weights; adjusted R^2 are also shown.

Scale	Acoustic cue								Adj R^2
	Duration	Int_M	Int_{SD}	Amp Onsets	$F0_M$	$F0_{SD}$	$Spectral_{COG}$	H/N_{RATIO}	
Achievement	.02	.29*	.23*	-.03	.22*	.03	.15	.45*	.38*
Amusement	-.05	.24!	.23*	.21*	.02	.06	.15	.19	.06*
Pleasure	.12	.01	-.07	-.16!	-.06	-.1	-.18	.4*	.41*
Relief	.18!	-.07	-.06	-.35*	.48*	-.06	-.33*	-.29*	.26*
Anger	.09	.43*	-.01	-.35*	-.32*	.04	.09	.14*	.32*
Disgust	-.01	-.02	-.01	.1	-.33*	.24*	.35*	.03	.08*
Fear	-.24*	-.03	.27*	-.0	.29*	-.29*	-.17	-.2	.28*
Sadness	-.09	-.31*	-.15	.38*	.25*	-.16	-.14	-.04	.19*
Valence	.05	.12	.07	-.08	.12	-.04	-.08	.28!	.13*
Arousal	.02	.63*	.49*	.1	-.1	.1	.53*	.06	.50*
Authenticity	.04	-.05	.0	.14	.23!	-.16	-.15	-.16	.00

Note. Int = Intensity, Amp = Amplitude, COG = center of gravity, H/N = harmonics-to-noise, Adj = Adjusted. * $p < .05$; ! $p < .08$

Conclusion

We herein produced and validated a well-controlled corpus of nonverbal vocalizations, which we make available for future research on emotion processing. It includes 121 sounds, recorded by four different speakers, and represents similarly positive and negative emotion categories: achievement/triumph, amusement, sensual pleasure, relief, anger, disgust, fear, and sadness. These vocalizations elicited high emotion recognition accuracy in a forced-choice task, 86% on average, and were consistently rated as communicating the intended emotions in a rating task. Perceptual data includes also details

concerning perceived valence, arousal, and authenticity for each vocalization. Furthermore, we have shown that acoustic cues alone contain sufficient information to automatically classify vocalizations' emotion category and to predict listeners' behavioral responses. **Although the validation procedure was mostly based on young women, we have established that the vocalizations can be accurately recognized by both male and female listeners of different ages: as part of another ongoing project of our team, female (*n* = 52) and male participants (*n* = 34) varying widely in age (18 – 83 years) performed a rating task on a subset of these vocalizations, and derived accuracy rates were generally high and similar across genders (66% for women and 69% for men; Lima, Alves, Scott, & Castro, submitted). Belin et al. (2008) reported that women are more accurate than men at recognizing vocal emotions, but this has not been widely replicated. The absence of consistent gender effects has been repeatedly observed in studies on nonverbal vocalizations (Hawk et al., 2009; Sauter, 2006; Sauter, Panattoni, & Happé, 2012), as well as in studies on other auditory emotion modalities, notably speech prosody (e.g., Paulmann, Pell, & Kotz, 2008).**

We suggest that this corpus is suitable for many different research purposes. For instance, it can be used to explore putative specificities in the behavioral, cognitive and neural mechanisms underlying different positive emotions; the potential impact of neurological, psychiatric and developmental disorders in auditory emotion processing (e.g., Jones et al., 2011); the structural and functional bases of emotion perception (e.g., Omar et al., 2011; Peelen et al., 2010); the processing time-course of different emotions; or the effects of valence and arousal in emotion processing, as the set affords wide variability regarding these dimensions. Because of the nonverbal nature of the sounds, they can be used in different countries and cultural backgrounds. They are appropriate for behavioral paradigms

(e.g., forced-choice, ratings, and reaction time tasks), as well as for studies using different neuroscience techniques, including MRI, EEG/MEG, and transcranial magnetic stimulation (TMS). For example, a recent TMS study using a very similar set of stimuli (Banissy et al., 2010) showed that stimulation in the right postcentral gyrus and in the right lateral premotor cortex disrupts listeners' ability to perceive emotions in voice (amusement, sadness, fear, and disgust), but not the speakers' identity, suggesting that sensorimotor activity might be important for emotion discrimination. Finally, these vocal expressions can also be used in clinical settings to inspect pragmatic skills.

References

Abboud, H., Schultz, W. H., & Zeitlin, V. (2006). SuperLab, Stimulus presentation software (Version 4.0). San Pedro, California: Cedrus Corporation.

Adolphs, R., Damasio, H., & Tranel, D. (2002). Neural systems for recognition of emotional prosody: A 3-D lesion study. *Emotion, 2*, 23-51. doi:10.1037/1528- 3542.2.1.23

Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage, 42*, 919-927. doi:10.1016/j.neuroimage.2008.05.034

Banissy, M. J., Sauter, D., Ward, J., Warren, J. E., Walsh, V., & Scott, S. K. (2010). Supressing sensorimotor activity modulates the discrimination of auditory emotions but not speaker identity. *The Journal of Neuroscience, 30*, 13552-13557. doi:10.1523/ JNEUROSCI.0786-10.2010

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614-636. doi:10.1037/0022- 3514.70.3.614

Bänziger, T., & Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus In K. R. Scherer, T. Bänziger & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 271-294). Oxford: Oxford University Press.

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal Expression Corpus for experimental research on emotion perception. *Emotion, 12,* 1161-1179. doi:10.1037/a0025827

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*, 129-135. doi:10.1016/j.tics.2004.01.008

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

Bestelmeyer, P. E. G., Rouger, J., DeBruine, L. M., & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition, 117*, 217-223. doi:10.1016/j.cognition.2010.08.008

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., … Murphy, D. G. M. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology, 21*, 1-5. doi:10.1016/j.cub.2011.06.009

Boerma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1.05). Retrieved May 1, 2009, from http://www.praat.org/.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., & Weiss, B. (2005). *A database of German emotional speech.* Paper presented at the 9th European Conference on Speech Communication and Technology, Lisbon, Portugal.

Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudo-sentences for research on emotional prosody. *Behavior Research Methods, 42*, 74-81. doi:10.3758/BRM.42.1.74

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26,* 145–171. doi:10.1.1.47.7942

Dellacherie, D., Hasboun, D., Baulac, M., Belin, P., & Samson, S. (2011). Impaired recognition of fear in voices and reduced anxiety after unilateral temporal lobe resection. *Neuropsychologia, 49*, 618-629. doi: 10.1111/j.1469- 8986.2010.01075.x

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169-200. doi: 10.1080/02699939208411068

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists Press.

de Gelder, B. d., & van den Stock, J. (2011). The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Frontiers in Psychology, 2:181*. doi:10.3389/fpsyg.2010.00187

Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, C., Baulac, M., & Samson, S. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain, 128*, 628-640. doi:10.1093/brain/awh420

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8*, 145-146. doi:10.1038/nn1392

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior, 33*, 149-180. doi:10.1007/s10919-009-0070-5

Hawk, S. T., Kleef, G. A. v., Fischer, A. H., & Schalk, J. v. d. (2009). "Worth a thousand words": Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion, 9*, 293-305. doi:10.1037/a0015178

Hunter, E. M., Phillips, L. H., & MacPherson, S. E. (2010). Effects of age on cross-modal emotion perception. *Psychology and Aging, 25*, 779-787. doi:10.1037/a0020528

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage, 58*, 665-674. doi:10.1016/j.neuroimage.2011.06.035

Jones, C. R. G., Pickles, A., Falcaro, M., Marsden, A. J. S., Happé, F., Scott, S. K., … Charman, T. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders. *The Journal of Child Psychology and Psychiatry, 52*, 275-285. doi: 10.1111/j.1469-7610.2010.02328.x

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

Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (submitted). In the hear of the beholder: Age shapes emotion recognition in positive and negative nonverbal vocalizations.

Lima, C. F., & Castro, S. L. (2011). Emotion recognition in music changes across the adult life span. *Cognition and Emotion, 25*, 585-598. doi:10.1080/02699931.2010.502449

Lima, C. F., Garrett, C., & Castro, S. L. (under review). Not all sounds sound the same: Parkinson's disease affects differently emotion processing in music and in speech prosody.

Liu, P., & Pell, M. D. (in press). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods*. DOI 10.3758/s13428-012-0203-3

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces - KDEF*. Stockholm: Karolinska Institute, Department of Clinical Neuroscience, Psychology section.

Morton, J., Marcus, S. M., & Frankish, C. (1976). Perceptual centres (P-centres). *Psychological Review, 8*, 405–408.

Omar, R., Henley, S., Bartlett, J., Hailstone, J., Gordon, E., Sauter, D., … Warren, J. D. (2011). The structural neuroanatomy of music emotion recognition: Evidence from frontotemporal lobar degeneration. *NeuroImage, 56*, 1814-1821. doi:10.1016/ j.neuroimage.2011.03.002

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language, 104*, 262-269. doi:10.1016/j.bandl.2007.03.002

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *The Journal of Neuroscience, 28*, 10127-10134. doi:10.1523/ JNEUROSCI.2161-10.2010

Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective, & Behavioral Neuroscience, 3*, 275-288. doi:10.3758/CABN.3.4.275

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115*, 102-141. doi: 10.1037/0033-2909.115.1.102

Sauter, D. (2006). *An investigation into vocal expressions of emotions: The roles of valence, culture, and acoustic factors* (Unpublished doctoral dissertation). University College London, London.

Sauter, D., & Eimer, M. (2009). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience, 22*, 474-481. doi:10.1162/jocn.2009.21215

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*, 2251-2272. doi:10.1080/17470211003721642

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, 107*, 2408-2412. doi:10.1073/pnas.0908239106

Sauter, D., Panattoni, C., & Happé, F. (2012). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*. Advance online publication. DOI: 10.1111/j.2044-835X.2012.02081.x

Sauter, D. A., & Scott, S. K. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion, 31*, 192-199. doi:10.1007/s11031-007-9065-x

Scherer, K. R. (1994). Affect bursts. In: van Goozen, S., van de Poll, N. E., Sergeant, J. A. (Eds.), Emotions: Essays on Emotion Theory (pp. 161-196). Hillsdale, NH: Erlbaum.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of 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 Communication, 40*, 227-256. doi:10.1016/S0167- 6393(02)00084-5

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*, 123-148. doi:10.1007/ BF00995674

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. Davidson, K. R. Scherer & H. Goldsmith (Eds.), *Handbook of the affective sciences* (pp. 433-456). New York: Oxford University Press.

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the emotion recognition index. *Journal of Nonverbal Behavior, 4*, 305-326. doi:10.1007/s10919-011-0115-4

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*, 24-30. doi: 10.1016/j.tics.2005.11.009

Schröder, M. (2003). Experimental study of affect bursts *Speech Communication, 40*, 99-116. doi:10.1016/S0167-6393(02)0078-X

Scott, S. K. (1993). *P-centers in speech: An acoustic analysis.* Unpublished doctoral dissertation, University College London, London, UK.

Scott, S. K., Sauter, D., & McGettigan, C. (2010). Brain mechanisms for processing perceived emotional vocalizations in humans. In M. B. Stefan (Ed.), *Handbook of Behavioral Neuroscience* (pp. 187-197). London: Academic Press.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition and Emotion, 22*, 720-752. doi:10.1080/02699930701503567

Warren, J. E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J. S., … Scott, S. K. (2006). Positive emotions preferentially engage an auditory-motor "mirror" system. *The Journal of Neuroscience, 26*, 13067-13075. doi:10.1523/JNEUROSCI. 3907-06.2006

Appendix

Illustrative scenarios provided for each emotion and for the dimensions of valence and arousal, as well as instructions for authenticity

Emotions and Dimensions	Scenario
Achievement	You are a football fan and your team wins the most important game of the championship
Amusement	Someone tells you a joke that you find really funny
Pleasure	You are eating your favorite dessert, which you had not eaten for a long time
Relief	You think you lost your wallet but find it again
Anger	Someone is being deliberately very rude to you and you loose all your patience
Disgust	You put your hand in vomit
Fear	Someone suddenly taps on your shoulder in a dark alleyway
Sadness	You find out that someone close to you has died
Arousal	Minimum: You are feeling very sleepy. Maximum: You are feeling very alert and energetic
Valence	Positive: You are having an experience of extreme pleasure. Negative: you are having and extremely unpleasant experience of stress or discomfort
Authenticity	Is the vocalization a realistic representation of what we observe in everyday life?