Motivation and Emotion
https://doi.org/10.1007/s11031-019-09783-9

ORIGINAL PAPER

Decoding emotions from nonverbal vocalizations: How much voice signal is enough?

Paula Castiajo[1] · **Ana P. Pinheiro**[1,2]

Abstract
How much acoustic signal is enough for an accurate recognition of nonverbal emotional vocalizations? Using a gating paradigm (7 gates from 100 to 700 ms), the current study probed the effect of stimulus duration on recognition accuracy of emotional vocalizations expressing anger, disgust, fear, amusement, sadness and neutral states. Participants (n = 52) judged the emotional meaning of vocalizations presented at each gate. Increased recognition accuracy was observed from gates 2 to 3 for all types of vocalizations. Neutral vocalizations were identified with the shortest amount of acoustic information relative to all other types of vocalizations. A shorter acoustic signal was required to decode amusement compared to fear, anger and sadness, whereas anger and fear required equivalent amounts of acoustic information to be accurately recognized. These findings confirm that the time course of successful recognition of discrete vocal emotions varies by emotion type. Compared to prior studies, they additionally indicate that the type of auditory signal (speech prosody vs. nonverbal vocalizations) determines how quickly listeners recognize emotions from a speaker's voice.

Keywords Nonverbal vocalizations · Emotion · Duration · Gate · Recognition

Introduction

The voice is likely the most important sound category in a social environment. It carries not only verbal information, but also socially relevant cues about the speaker, such as his/her identity, sex, age, and emotional state (Belin et al. 2004; Schirmer et al. 2004). Vocal emotions can be communicated either through suprasegmental modulations of speech prosody (Schirmer et al. 2005) or short nonverbal vocalizations (Schröder 2003), also known as affective bursts (Belin et al. 2008). Both speech prosody (Banse and Scherer 1996; Juslin and Laukka 2001) and nonverbal vocalizations (Sauter et al. 2010) rely on a shared acoustic code (e.g., duration, F0, intensity) serving the expression of emotional meaning. Nonetheless, compared to speech prosody, nonverbal vocalizations are considered to represent more primitive expressions of emotions and an auditory analogue of facial emotions (Belin et al. 2004). The ability to accurately infer vocal emotions of social partners—both prosody and nonverbal vocalizations—is critical during communication, specifically in predicting the intentions and behaviors of others (Juslin and Laukka 2003). Whereas the time needed to accurately recognize vocal emotions from speech prosody has been specified (Pell and Kotz 2011), it remains to be clarified how much temporal information is necessary for an accurate decoding of emotions from nonverbal vocalizations, the focus of the current study.

Decoding of emotions from nonverbal vocalizations versus speech prosody

Vocal expressions are inherently dynamic and their temporal structural determines how emotional meaning is decoded. A robust body of evidence demonstrates that emotions are perceived and recognized in a categorical manner during voice perception (Cowie and Cornelius 2003; Juslin and Laukka 2003; Laukka 2005), supporting categorical approaches to

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11031-019-09783-9) contains supplementary material, which is available to authorized users.

✉ Ana P. Pinheiro
appinheiro@psicologia.ulisboa.pt

[1] Psychological Neuroscience Lab, CIPsi, School of Psychology, University of Minho, Braga, Portugal

[2] Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal

Published online: 20 July 2019

emotion (Ekman 1992; Scherer and Ellgring 2007). Several studies probed how accurately listeners decode vocal emotions either from speech prosody (Van Bezooijen 1984; Juslin and Laukka 2001; Pell 2002; Scherer 1989) or nonverbal vocalizations (Belin et al. 2008; Lima et al. 2013; Sauter et al. 2010; Schröder 2003; Vasconcelos et al. 2017). These studies showed that accuracy varies by emotion category. Specifically, speech prosodic cues expressing fear and happiness are commonly less accurately recognized than anger and sadness, whereas disgust tends to be associated with the lowest recognition accuracy (Banse and Scherer 1996; Juslin and Laukka 2003; Van Bezooijen 1984). In the case of nonverbal vocalizations, sadness and disgust are associated with the highest accuracy (Hawk et al. 2009; Lima et al. 2013; Sauter et al. 2010; Schröder 2003), whereas fear, anger and happiness tend to be the least accurately decoded vocal emotions (Sauter et al. 2010; Schröder 2003—but see Vasconcelos et al. 2017, in which amusement [laughter] was the most successfully recognized emotion). Although scarce, studies that have directly compared recognition accuracy in emotional prosody and nonverbal vocalizations revealed that fear, happiness, anger, sadness, contempt and disgust are more easily recognized in vocalizations than in speech prosody (Hawk et al. 2009).

The time course of emotion recognition in speech prosody

Emotion-specific differences in voice processing may be explained by differences in their underlying acoustic properties (Juslin and Laukka 2001; Sauter et al. 2010). The expression and comprehension of vocal emotions rely on a complex configuration of acoustic cues, such as duration, fundamental frequency (F0), and intensity (Schirmer and Kotz 2006). For instance, happy prosody is characterized by high F0 and intensity and fast speech rate, whereas sad prosody is characterized by low F0 and intensity, as well as by slow speech rate (Banse and Scherer 1996; Juslin and Laukka 2003; Sobin and Alpert 1999). The event-related potential (ERP) studies that probed the time course of vocal emotional perception indicate that neutral and emotional cues are rapidly differentiated within 200 ms after stimulus onset (Liu et al. 2012; Paulmann and Kotz 2008; Pinheiro et al. 2013, 2014; Schirmer et al. 2007). By examining emotional priming effects (typically reflected in faster and/or more accurate responses to targets preceded by congruent rather than incongruent primes—Murphy and Zajonc 1993), some of these ERP studies demonstrated that the implicit recognition of emotional prosody occurs very rapidly (i.e., 300 ms after prime-target onset: Bostanov and Kotchoubey 2004; 400 ms after prime-target onset: Paulmann and Pell 2010; Schirmer et al. 2002, 2005). Despite an early differentiation of neutral and emotional acoustic cues at the brain level, listeners may need more or less acoustic information (i.e., longer vocal samples) to explicitly categorize specific discrete emotions. The studies of Pell and Kotz (2011), Rigoulot et al. (2013) with prosodic pseudo-utterances demonstrated that neutral expressions (510 ms and 654 ms, respectively), fear (517 ms and 427 ms, respectively), sadness (576 ms and 612 ms, respectively) and anger (710 ms and 672 ms, respectively) are recognized at shorter gate intervals than happiness (977 ms and 811 ms, respectively) and disgust (1486 ms and 1197 ms, respectively). Using a restricted number of emotion categories, Cornew et al. (2010) observed similar findings with neutral (444 ms), angry (723 ms) and happy (802 ms) prosodic stimuli. The remarkable capacity to recognize negative emotions such as fear, sadness and anger with shorter amounts of acoustic information may be explained by the relevance that a prompt response to vocal signals of threat, loss, and aggression plays in terms of survival (Pell and Kotz 2011). Nonetheless, changes in speech rate may affect the recognition of negative prosodic cues: whereas a slow speech rate (associated with a longer stimulus duration) was associated with a more accurate recognition of sadness, a faster speech rate (associated with a shorter stimulus duration) led to a more accurate recognition of anger (Bergmann et al. 1988).

As all studies mentioned above (Bergmann et al. 1988; Cornew et al. 2010; Pell and Kotz 2011; Rigoulot et al. 2013) probed the effects of duration on vocal emotional recognition using speech prosody, it remains to be clarified how much temporal information is necessary for the accurate decoding of emotions from nonverbal vocalizations. As for speech prosody, it is reasonable to expect that different amounts of voice information are required to recognize discrete emotional states expressed through vocalizations without verbal content. Similarities in the temporal patterns of vocal emotional recognition in speech prosody and nonverbal vocalizations would support a general mechanism involved in emotional decoding irrespective of stimulus type.

Rapid decoding of emotions in nonverbal vocalizations versus speech prosody

A more recent ERP study shed light on whether the processing of emotional prosody versus nonverbal vocalizations relies on the same neurocognitive mechanisms. Pell et al. (2015) showed that the perception of nonverbal vocal emotions takes precedence over prosodic emotional cues (Pell et al. 2015). Earlier N100 and P200 peak latencies were observed for happy (laughter) nonverbal vocalizations compared to happy prosody, as well as earlier P200 peak latencies for nonverbal vocalizations expressing anger compared to angry prosody. In contrast, no differences were observed in the N100 and P200 peak latencies for sadness

as a function of stimulus type (Pell et al. 2015). Latency differences between emotion categories were also observed in the study of Pell et al. (2015): the N100 peaked earlier for happy (laughter) compared to both angry and sad nonverbal vocalizations, whereas the P200 peaked earlier for happy ($M = 216$ ms), followed by angry ($M = 224$ ms) and sad vocalizations ($M = 235$ ms). The earlier ERP response to laughter relative to other emotional categories corroborates the high distinctiveness of this vocal expression (Edmonson 1983; Kipper and Todt 2001). Despite the contributions of these ERP studies, they do not specify the minimum duration of vocal information that is necessary for above-chance emotional recognition of nonverbal vocalizations. Whereas ERPs track the time course of emotional cue processing from stimulus onset until a response is made (i.e., evaluating both bottom-up sensory and higher-order cognitive processing stages), behavioral measures tap into the higher-order evaluation processes that lead to a response (e.g., Paulmann and Pell 2010).

The current study and hypotheses

How much voice signal is enough for a listener to successfully recognize discrete emotional meaning? Does the time course of emotional recognition differ when the voice signal carries (speech prosody) or not (nonverbal vocalizations) verbal content? Using a gating paradigm, the current study examined the minimum stimulus duration necessary for an accurate recognition of emotions in nonverbal vocalizations, extending previous studies (Pell and Kotz 2011). Nonverbal vocalizations were selected from the Montreal Affective Voices (MAV, Belin et al. 2008; validation for the Portuguese population by Vasconcelos et al. 2017) as the MAV is a standardized battery widely used in studies of emotion recognition (e.g., Collignon et al. 2010; Naranjo et al. 2011; Paquette et al. 2013). For a more direct comparison with previous studies testing speech prosody (Pell and Kotz 2011), MAV vocalizations conveying basic emotions, i.e. anger, disgust, fear, amusement, sadness and neutral states were selected to determine the earliest time window at which emotions are accurately recognized. Auditory gates were defined and represented specific time increments: stimuli were divided into seven gate intervals ranging from 100 to 700 ms (i.e., the first gate started at the beginning of the vocalization and the last corresponded to the entire stimulus duration). Furthermore, participants were asked to categorize the vocalizations based on a six-choice classification system (neutral, amusement, sadness, disgust, anger and fear). Two types of analyses were performed: first, recognition accuracy was computed for each type of vocalization at each gate duration; second, the gate at which the emotion was accurately recognized (without classification changes at longer gate durations) was identified for each type of vocalization (i.e., emotion identification points were computed).

As for speech prosody (Pell and Kotz 2011), we hypothesized that emotional recognition accuracy would increase as a function of increased gate duration (i.e., increased stimulus length). Moreover, we hypothesized that an accurate identification of discrete emotions would occur at different gate intervals. Specifically, two alternative hypotheses were formulated. Considering EEG evidence showing an earlier latency of ERP components within 200 ms post-stimulus onset in response to laughter compared to both angry and sad vocalizations (Pell et al. 2015), the recognition of laughter was expected to require less acoustic information than the recognition of negative vocalizations, reflected in an earlier identification point. However, if emotion recognition in nonverbal vocalizations follows a similar time course to speech prosody (Cornew et al. 2010; Pell and Kotz 2011; Rigoulot et al. 2013), negative vocalizations should be identified at shorter gates compared to positive vocalizations. This would be consistent with an evolutionary approach to emotions (e.g., Ito et al. 1998; Nesse 1990; Pell and Kotz 2011): negative vocal expressions are considered highly salient cues since they might communicate possible threats in the environment (Baumeister et al. 2001). Therefore, a rapid detection and recognition of negative emotions is critical for survival (Pell and Kotz 2011). Given the lower acoustic variation of neutral vocalizations, we hypothesized they would be recognized at shorter gates compared to emotional vocalizations (Pell and Kotz 2011). These hypotheses were tested with mixed-effects models since they avoid spurious effects and have more power compared to traditional methods such as the analysis of variance (ANOVA; Jaeger 2008).

Method

Participants

The best procedure to determine sample size when using mixed-effects modeling remains to be specified (Maas and Hox 2005; McNeish and Stapleton 2016). Hence, and since the current study relies on a similar gating paradigm to the one described by Pell and Kotz 2011 with speech prosody (n = 49), a similar number of participants was included in the current study. Fifty-two participants (mean age = 23.42, $SD = 7.80$ years, age range 18–49 years; 27 females) completed the study. The education level ranged from 12 to 20 years ($M = 13.87$, $SD = 1.99$). All participants reported no hearing difficulties and were native speakers of European Portuguese. They provided written informed consent and received course credit for their participation. The study was approved by a local Ethics Committee (University of Minho, Braga, Portugal).

Stimuli

Fifty-eight nonverbal vocalizations expressing anger (n = 10), disgust (n = 10), fear (n = 8), amusement[1] (n = 10), sadness (n = 10) and neutral states (n = 10) were selected from the MAV. The MAV includes 90 nonverbal emotional vocalizations (e.g., laughter, growls, cries or screams) uttered by five male and five female speakers (five female and five male vocalizations of each emotional category). These speakers were instructed to produce short emotional interjections using the vowel/ɑ/.[2] Despite cultural differences in decoding accuracy for different MAV emotional categories (increased for sadness—86%, and decreased for pain—51%—Belin et al. 2008; increased for sadness—79%, and decreased for fear—25%—Koeda et al. 2013; increased for amusement—90%, and decreased for anger—29.5%—Vasconcelos et al. 2017), all these studies yielded high levels of accuracy.[3] In the current study, the number of male and female vocalizations was the same for each emotional category.

First, the duration of the selected MAV stimuli was shortened (n = 50) or increased (n = 8) to 700 ms[4]: whereas shortening stimulus duration involved cutting the endpoint of its waveform, stimulus duration was gradually increased to 700 ms by adding less variable segments of the sound to its waveform ensuring that its emotionality was not changed (based on a pilot study with three experienced voice researchers). Therefore, two vocalizations of fear were excluded as their too short duration prevented further acoustic manipulations. To test whether the manipulated vocal samples still conveyed the intended emotional meaning, they were first judged by a sample of participants who did not participate in the current study (n = 38; mean age = 21.68, *SD* = 3.74 years, age range 18–33 years; 23 females). Hits were defined as the number of times a vocalization received the highest intensity ratings in the corresponding scale. The overall mean recognition accuracy was quite high (68.67% for a chance level of 25%). Recognition accuracy rates were identical to those described for the original (non-manipulated) stimuli of the MAV (Belin et al. 2008; Vasconcelos et al. 2017).

After this pilot experiment, the 700 ms vocalizations were divided into 7 gates with decreasing durations from 700 to 100 ms, yielding a total of 406 vocalizations (see Supplementary Material for examples). To control for differences in the perceived loudness across stimuli, vocalizations were individually normalized to a peak intensity of 70 dB. Manipulation of stimulus duration and intensity was performed with Praat software (Boersma and Weenink 2005, www.praat.org).

Procedure

An auditory gating task was used, in which listeners were presented with increasing (cumulative frequency) amounts of vocal emotional information. Data were collected on desktop computers equipped with headphones. Superlab software (Cedrus Corporation 1991) was used for stimulus presentation and response recording. Stimuli were equally distributed across two sets (each included vocalizations produced by male and female speakers, representing the five emotion categories and a neutral category—203 stimuli in each set) to avoid fatigue effects: 50% of participants rated set 1 and 50% rated set 2. Hence, five samples of anger, amusement, disgust, sadness, and neutral expressions and four samples of fear presented in the seven gate durations were included in each set (29 vocalizations × 7 gates). Whenever one set included three female and two male samples of an emotion category, the other set included three male and two female samples of the same emotion type. This was true for anger, amusement, disgust, sadness and neutral expressions. As for fear, each set included two female and two male samples at each gate duration. Following Pell and Kotz (2011), the order of stimulus presentation began with gate 1 and ended with gate 7. Each gate duration comprised stimuli of the six categories that were randomized across participants. Before the beginning of the experiment, instructions were provided on the computer screen. Participants were instructed to rate the emotion category of each vocalization by choosing one of six possible options (0 = neutral, 1 = amusement, 2 = sadness, 3 = anger, 4 = disgust, and 5 = fear). A training trial with two extra stimuli was provided at the beginning of the experiment. Each experimental session lasted approximately 45 min. Three additional seven-point scales were used to

[1] Based on existing studies arguing for a clear distinction between different types of positive nonverbal vocalizations (e.g., achievement/triumph, amusement, contentment, sensual pleasure and relief—Sauter and Scott 2007; amusement, interest, relief, awe, compassion, sensory pleasure, enthusiasm and triumph—Simon-Thomas et al. 2009), the term 'happiness' used in the MAV was replaced with 'amusement' in the current study as it more accurately matches the stimuli (laughter) included in this battery.

[2] The lower acoustic variability of the MAV sounds, compared to other stimulus batteries (e.g., Lima et al. 2013), is ideal for the study of the effects of stimulus duration on vocal emotional recognition: when presented with stimuli with lower acoustic variation, listeners may rely more on duration for their emotional judgments than on other acoustic properties of the voice.

[3] 68.2% for the chance level at 12.5%—Belin et al. (2008), 51.1% for the chance level at 12.5%—Koeda et al. (2013), 62.8% for the chance level at 11.1%—Vasconcelos et al. 2017.

[4] The maximum duration of the gate (700 ms) was chosen to allow the use of the current vocalizations in ERP studies probing the time course of vocal emotional processing. In studies using this methodology, differences in stimulus duration across conditions may affect sensory ERP components such as the N1 (Stapells 2002), and potentially confound the interpretation of later processing stages involved in the cognitive evaluation of the stimulus.

assess the approachability (1 = totally avoiding, 7 = totally approaching the stimulus), intensity (1 = not at all intense, 7 = extremely intense) and authenticity (1 = not at all authentic, 7 = extremely authentic) of each vocalization. As the focus of the current study was on the effects of stimulus duration on recognition accuracy, these additional ratings are presented as Supplementary Material.

Data analyses

Recognition accuracy Accuracy was calculated as the proportion of correct categorizations for each emotion category in each gate. For example, for anger, a proportion of 0.50 indicates that 5 out of 10 vocalizations of anger were correctly classified as anger.

Emotion identification points An emotional identification point was defined as the first gate duration at which an emotion was accurately recognized if (1) it was again correctly identified at the next gate duration; and if (2) it was not more than once incorrectly identified at all following (longer) gate durations (Pell and Kotz 2011; Salasoo and Pisoni 1985). Emotion identification points ranged from gates 1 to 6. Trials that did not fulfill the abovementioned criteria were discarded. For example, gate 2 was assigned to the following categorization of amusement: neutral, amusement, amusement, amusement, sadness, amusement and amusement. Emotion identification points were separately estimated for each vocalization when judged by each participant (5 categories × 5 items + 1 category × 4 items × 52), yielding a total of 1508 identification points or 260 identification points for anger, disgust, amusement, sadness and neutral categories, and 208 identification points for fear. Subsequently, we calculated the distribution of correct identification points and discarded trials for each emotion type at each gate duration (see Table 2).

Accuracy data were analyzed with linear mixed-effects models using the lmer4 (Bates et al. 2014) and lmerTest (Kuznetsova et al. 2016) packages in the R environment (R3.4.3. GUI 1.70), which were used to estimate fixed and random coefficients. This type of statistical analysis allows examining the effects of variables that may vary either between subjects, within subjects, or both between and within subjects, representing a more efficient way to account for variance at different levels (Hoffman and Rovine 2007).

Results

Recognition accuracy

Table 1 displays the overall recognition accuracy for each type of vocalization at each gate duration. The overall mean recognition accuracy was 0.34 at gate 1, 0.39 at gate 2, 0.71 at gate 3, 0.78 at gate 4, 0.80 at gate 5, 0.84 at gate 6, and 0.85 at gate 7. All vocalizations were recognized with above-chance accuracy from gates 1 to 7 (the chance-level performance was 16.66% for a 6-alternative forced choice task), except sadness at gate 1.

The hypothesis that emotional recognition accuracy would increase as a function of increased gate duration was first tested with a mixed-effects model including recognition accuracy as outcome, participants as random effects, and emotion category and gate duration as fixed effects. A significant interaction effect between emotion category and gate duration (estimated difference = .969, $p < .001$) indicated that accuracy across gates differed significantly according to emotion category. To account for the effect of set (sets 1 and 2), this dichotomous variable was also added to the model as a fixed effect. The set had no significant effect on the recognition accuracy of all types of vocalizations, except anger (estimated difference = .978; $p = .035$) and disgust (estimated difference = − .094; $p = .006$). The analysis was also replicated using approachability, intensity and authenticity as covariates: the pattern of results for the interaction did not change with the inclusion of these variables.

Subsequently, the effects of gate on recognition accuracy were tested for each emotion category separately. The model included recognition accuracy as outcome, participants as random effect, and gate as fixed effect. The recognition of each type of vocalization improved significantly at successive gate intervals (see Table 1): recognition of amusement improved from gates 1 to 2 (estimated difference = .065, $p = .017$), 2 to 3 (estimated difference = .142, $p < .001$) and 3 to 4 (estimated difference = .065, $p = .017$); recognition of sadness improved from gates 1 to 2 (estimated difference = .112, $p < .001$), 2 to 3 (estimated difference = .565, $p < .001$) and 3 to 4 (estimated difference = .158, $p < .001$); recognition of anger improved from gates 2 to 3 (estimated difference = .135, $p < .001$) and 5 to 6 (estimated difference = .081, $p = .025$); recognition of disgust improved from gates 2 to 3 (estimated difference = .327, $p < .001$) and 3 to 4 (estimated difference = .112, $p < .001$); recognition of fear improved from gates 2 to 3 (estimated difference = .313, $p < .001$) and 6 to 7 (estimated difference = .077, $p = .049$); recognition of neutral vocalizations improved from gates 2 to 3 (estimated difference = .446, $p < .001$).

Emotion identification points

This set of analyses aimed to estimate the exact gate duration at which each vocalization was first recognized by each participant, with no identification changes at subsequent gates. Table 2 illustrates the distribution of correct identification points and discarded trials for each vocalization at each gate. Figure 1 displays the minimum duration, expressed in

Table 1 Proportion of correct responses for each type of vocalization at each gate duration

Emotion category	Recognition accuracy						
	Gate duration (ms)						
	G1[a]	G2[b]	G3[c]	G4[d]	G5[e]	G6[f]	G7
	100	200	300	400	500	600	700
Anger	0.36 (0.24)	0.38 (0.25)	0.52[b] (0.24)	0.52 (0.25)	0.53 (0.21)	0.61[c,d,e] (0.22)	0.62[c,d,e] (0.27)
Disgust	0.33 (0.23)	0.33 (0.26)	0.66[b] (0.17)	0.77[c] (0.17)	0.82[c] (0.15)	0.87[c,d] (0.18)	0.83[c,d] (0.22)
Fear	0.29 (0.23)	0.34 (0.25)	0.65[b] (0.24)	0.69 (0.24)	0.65 (0.26)	0.70 (0.28)	0.78[c,d,e,f] (0.24)
Amusement	0.25 (0.22)	0.32[a] (0.20)	0.88[b] (0.14)	0.95[c] (0.11)	0.97[c] (0.08)	0.98[c] (0.07)	0.98[c] (0.06)
Sadness	0.17 (0.20)	0.28[a] (0.18)	0.73[b] (0.18)	0.89[c] (0.13)	0.91[c] (0.13)	0.95[c,d] (0.10)	0.94[c] (0.09)
Neutral	0.62 (0.31)	0.68 (0.35)	0.82[b] (0.29)	0.88 (0.25)	0.91[c] (0.22)	0.90[c] (0.21)	0.92[c] (0.22)

Standard deviation is shown in parentheses

[a]Significantly different from gate 1

[b]Significantly different from gate 2

[c]Significantly different from gate 3

[d]Significantly different from gate 4

[e]Significantly different from gate 5

[f]Significantly different from gate 6

Table 2 Distribution of correct identification points and discarded trials for each type of vocalization at each gate duration

Emotion	Identification points							
	Gate duration (ms)							
	G1	G2	G3	G4	G5	G6	Total	Total
	100	200	300	400	500	600	Correct	Incorrect
Anger	40 (15.38%)	16 (6.15%)	39 (15%)	17 (6.54%)	15 (5.77%)	36 (13.85%)	163 (62.69%)	97 (37.31%)
Disgust	50 (19.23%)	28 (10.77%)	85 (32.70%)	24 (9.23%)	16 (6.15%)	23 (8.85%)	226 (86.93%)	34 (13.07)
Fear	32 (15.38%)	17 (8.17%)	57 (27.40%)	11 (5.29%)	11 (5.29%)	34 (16.35%)	162 (77.88%)	46 (22.12%)
Amusement	47 (18.08%)	30 (11.53%)	150 (57.69%)	21 (8.08%)	4 (1.54%)	4 (1.54%)	256 (98.46%)	4 (1.54%)
Sadness	22 (8.46%)	40 (15.38%)	116 (44.62%)	42 (16.16%)	16 (6.15%)	8 (3.08%)	244 (93.85%)	16 (6.15%)
Neutral	130 (50%)	33 (12.70%)	43 (16.54%)	18 (6.92%)	9 (3.46%)	5 (1.92%)	238 (91.54%)	22 (8.46%)

Mean percentage values are shown in parentheses

milliseconds, required to identify each category of nonverbal vocalizations, without further changes at subsequent gates.

The hypothesis that the accurate identification of discrete emotions would occur at different gate intervals was tested with a mixed-effects model including identification points as outcome, participants as random effect, and emotion category as fixed effect. The effect of set was also tested but, as no significant effect was found, it was not considered in the final model. Three participants who failed to identify the intended meaning from any of the angry vocalizations, and two participants who responded incorrectly to all neutral vocalizations were excluded from these analyses as the means for both types of vocalizations could not be estimated. Of note, after excluding these participants from the statistical model, we still ensured an equal number of participants per condition. A total of 47 participants (mean age = 23.23, $SD = 7.18$ years, age range 18–46 years; 23 females) was included in the analysis of identification points.

The amount of acoustic information needed to recognize emotions from nonverbal vocalizations differed significantly by emotion category. Neutral vocalizations were identified with the shortest amount of acoustic information

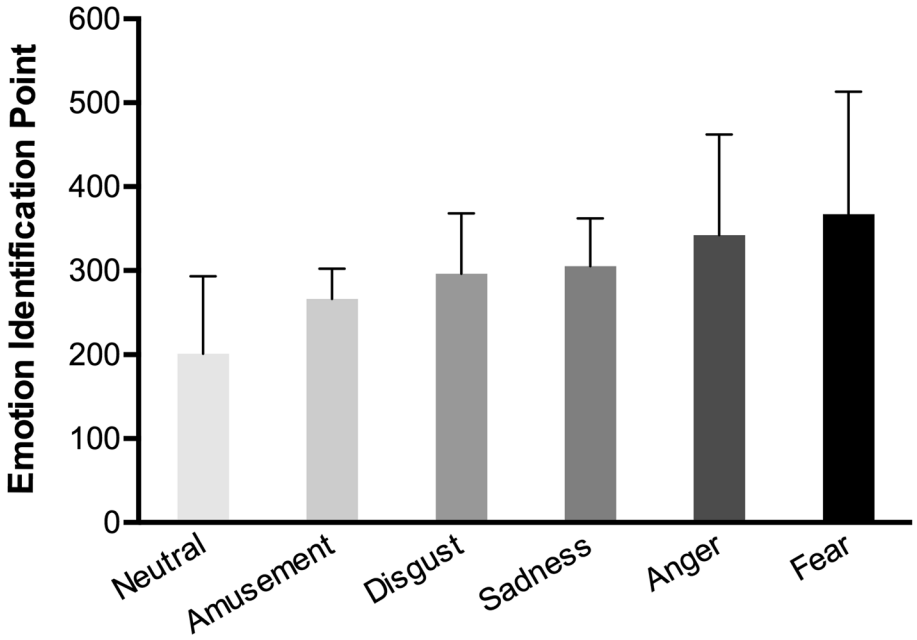

Fig. 1 Mean time values (ms) required to recognize each type of vocalization. The mean values considered only correctly identified vocalizations for each emotion category. Therefore, incorrect judgments were not considered (five exemplars of neutral, anger, sadness, amusement, and disgust and four exemplars of fear). Error bars represent standard deviations

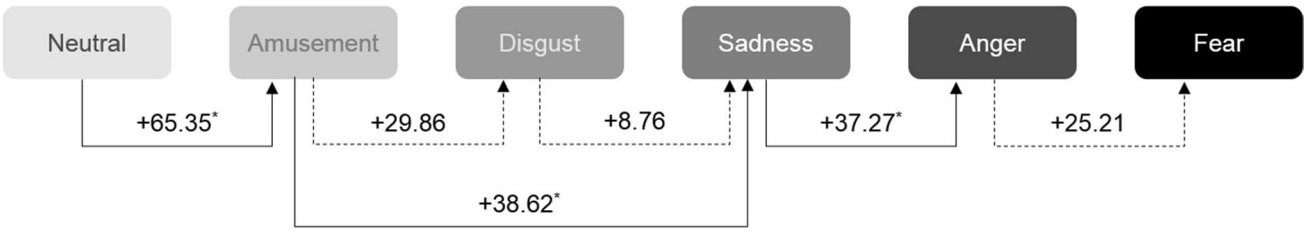

Fig. 2 Schematic diagram representing the estimated differences between emotions. Solid and dashed lines represent significant and non-significant p values, respectively. *$p \leq .05$

relative to all types of vocal emotions (fear: estimated difference = 166.45, $p < .001$; anger: estimated difference = 141.24, $p < .001$; sadness: estimated difference = 103.97, $p < .001$; disgust: estimated difference = 95.21, $p < .001$; amusement: estimated difference = 65.35, $p < .001$). Positive vocalizations (amusement) required a shorter acoustic signal than negative vocalizations (fear: estimated difference = 101.10, $p < .001$; anger: estimated difference = 75.89, $p < .001$ and sadness: estimated difference = 38.62, $p = .041$). Figure 2 illustrates the estimated differences between emotion categories.

Discussion

Clarifying how fast listeners decode vocal emotions is critical for our understanding of the mechanisms underlying emotional voice perception and communication. Vocal expressions change over time and these dynamic changes dictate how emotion-specific meaning is encoded and decoded. The current study probed how accurately listeners recognize emotional meaning communicated through short nonverbal vocalizations and determined the exact amount of acoustic information required for an accurate emotion recognition. A gating paradigm with precise time increments (100 ms) was used to investigate these questions. Our findings demonstrate that emotions are very rapidly decoded even when the voice signal does not contain verbal information, in good agreement with previous studies (Liu et al. 2012; Pell et al. 2015; Sauter and Eimer 2010). Significant improvements in recognition occurred from gates 2 to 3, when listeners were exposed to at least 200 ms voice samples. Further, the current findings support the observation that discrete vocal emotions unfold at different rates and are recognized at different points in time. Specifically, neutral vocalizations were recognized with a shorter acoustic signal than emotional vocalizations. When comparing recognition for the different discrete emotions, we observed that the recognition of amusement (laughter) required the shortest acoustic signal,

whereas anger and fear required the longest acoustic signal for successful recognition.

Decoding emotions from short nonverbal vocalizations

Vocal emotions rely on temporally-unfolding acoustic representations. An incremental increase of recognition accuracy was observed with longer vocal samples for all emotion categories. Specifically, the recognition of nonverbal vocal emotions was found to rapidly improve in the 200–300 time window. Of note, gate 3 (300 ms duration) was enough for the accurate recognition of more than 50% vocal samples of all emotion categories (see Table 1). Hence, the acoustic modulations occurring within the first 300 ms after vocalization onset seem to be the most critical time window for explicit decisions about a speaker's emotional state. This observation agrees with the time window showing effects of emotional salience detection in ERP studies (i.e., within 200 ms after voice onset—P50: Liu et al. 2012; N100 and P200: Pinheiro et al. 2013; P200: Sauter and Eimer 2010). Although not as critical, acoustic changes in later time windows facilitated emotional recognition in general, as more acoustic information was available to listeners.

Extending previous research with speech prosody (Pell and Kotz 2011; Rigoulot et al. 2013), the current results indicate differences in the rate at which the recognition of discrete emotions improves from one gate to the next. Specifically, amusement (266 ms), disgust (296 ms) and sadness (305 ms) were identified at shorter gate intervals compared to anger (342 ms) and fear (367 ms) (see Figs. 1, 2). In addition, significant differences were observed between amusement and sadness, with the first being identified with less acoustic information.

The earliest recognition of amusement at the behavioral level agrees with previous EEG evidence showing an earlier or facilitated discrimination of laughter compared to vocalizations expressing anger and sadness (Pell et al. 2015). The high acoustic distinctiveness of laughter (e.g., frequency parameters, duration or distribution of spectral energy—e.g., Edmonson 1983; Kipper and Todt 2001) may enhance the salience of this vocal emotion (e.g., Pinheiro et al. 2017a, b) and explain its earlier identification point (Kipper and Todt 2001) compared to the other emotion categories.

Following amusement, listeners were also able to decode sadness from shorter voice samples. The social function played by the two vocal emotions (Belin 2006; Latinus and Belin 2011; Meneses and Díaz 2017) may account for the earlier identification points observed in the current study. Particularly, spontaneous laughter (e.g., amused laughter, joyful laughter) facilitates cooperative behavior (Gervais and Wilson 2005; Greatbatch and Clark 2003). Laughter may also elicit emotional contagion (Mehu and Dunbar 2008), which is critical in social bonding (Gervais and Wilson 2005; Vettin and Todt 2004). Sadness (reflected in crying vocalizations) may promote the development of social bonds by stimulating in others the willingness to provide assistance and emotional support (Hendriks et al. 2008; Vingerhoets et al. 2009). Further, listeners seem to be innately predisposed to rapidly respond to crying and laughing sounds (Barr et al. 1996; Caron 2002) as both represent a child's primary means of communication (Barr et al. 1996; Scheiner et al. 2002). The identification of disgust was also possible when shorter vocal samples were presented to the listeners. This could be related to the survival value associated with the fast identification of disgusting sounds that might signal a threat for the organism, such as rotten food (Zimmer et al. 2016). Although the rapid detection of vocal expressions of anger and fear also has an adaptive value, the tendency to misinterpret anger as fear and vice versa (Belin et al. 2008; Vasconcelos et al. 2017) may explain why a longer acoustic signal was required for their accurate identification.

Supporting our hypothesis and a prior study with speech prosody (Pell and Kotz 2011), neutral vocalizations were accurately recognized at gate 1 (100 ms), with accuracy levels significantly higher (62%) than those achieved by all emotional categories (see Table 1). Plausibly, the smaller acoustic variation of neutral vocal cues (Belin et al. 2008) facilitated their identification at shorter gates. Additional acoustic cues provided by longer gates did not seem to add crucial information for its recognition.

Decoding emotions from short nonverbal vocalizations versus speech prosody

Our results also provide support for differences in recognition accuracy according to stimulus type (with vs. without verbal content). In general, the nonverbal vocalizations used in the current study yielded higher accuracy ratings than the prosodic pseudo-utterances in the study of Pell and Kotz (2011). Emotion-specific differences are also observed when contrasting accuracy for speech prosody (from shorter to longer duration: fear, sadness, anger, happiness and disgust—Pell and Kotz 2011) and nonverbal vocalizations (from shorter to longer duration: amusement, disgust, sadness, anger and fear—the current study).

The earlier recognition points in the current study suggest that shorter amounts of acoustic information are required for an accurate recognition of emotions in vocalizations relative to prosodic pseudo-utterances (between 266 and 367 ms in the current study vs. between 510 and 1486 ms in the study of Pell and Kotz 2011). This finding agrees with previous evidence demonstrating an earlier and preferential detection of emotions from nonverbal vocalizations relative to speech prosody (Pell et al. 2015). This difference might be justified by the fact that nonverbal vocalizations

served a communicative function long before the emergence of language (Belin et al. 2004). Compared to speech prosody, vocalizations seem to rely on less complex (Juslin and Laukka 2003) and more automatic (Lima et al. 2018) decoding mechanisms that could promote a faster behavioral response (Pell et al. 2015). Moreover, whereas speech prosody demands the concurrent processing of segmental and suprasegmental cues (which may require more processing resources), the processing of emotions in vocalizations relies on nonverbal information only (Pell et al. 2015). These features may explain why accurate emotion recognition is achieved with less acoustic information when the listener is exposed to vocalizations compared to speech prosody. Further, they may account for differences in the time course of emotion recognition when the voice signal contains (speech prosody) or not (nonverbal vocalizations) concurrent verbal information. Of note, both neutral vocalizations and neutral speech prosody were identified from shorter acoustic events compared to all vocal emotions. This suggests that stimulus type (speech prosody vs. nonverbal vocalizations) only affects the time course of vocal emotional (not neutral) recognition.

In good agreement with previous evidence (Hawk et al. 2009; Pell et al. 2015), the current findings suggest that emotions are more reliably and rapidly detected from nonverbal vocalizations than from emotional prosody. Research on vocal emotional processing may thus benefit from the use of nonverbal vocal emotions. This could be especially advantageous in the case of emotion recognition training programs. For example, Schlegel et al. (2017) conducted a video-based emotion recognition training program with non-clinical adults by combining distinct types of emotional information (facial, postural, gestural and speech prosody). They found that whereas training improved the performance of young and middle-aged adults, no improvement was observed in older adults (Schlegel et al. 2017). Even though there is some evidence indicating that the recognition of emotional information declines with age (Ruffman et al. 2008), it is plausible that the simultaneous processing of segmental and suprasegmental information (emotional prosody) increased task demands and contributed to the null effects observed in older participants. Therefore, the use of emotional nonverbal vocalizations, which have been found to be universally recognized (Laukka et al. 2013; Sauter et al. 2010), may be of particular relevance to emotion recognition intervention programs.

Limitations

Despite the strengths of the current study, some limitations should be considered. First, the use of a forced-choice task may have inflated recognition accuracy as listeners' judgments were restricted to the number of available options. Future studies should use a free-response format to determine whether spontaneous judgements produce similar accuracy levels for each type of vocalization as a function of gate. Second, besides laughter, no other types of positive vocal emotions (e.g., achievement, sensual pleasure, relief) were examined in the current study. As a result, it remains to be clarified whether the recognition of different types of positive vocalizations relies on different amounts of acoustic information. Third, the generalization of the current results is limited by the cultural specificities of this sample. Future studies should test the role of sociocultural differences in the time course of vocal emotional recognition. Finally, our conclusions are constrained by the gate durations used in the current study, which were defined by 100 ms increments. Further studies are required to examine whether accuracy varies by emotion type with gates shorter than 100 ms.

Conclusion

The current study adds to emotional prosody research by shedding light on the time required to decode basic emotions from nonverbal vocalizations. Acoustic modulations occurring within the first 300 ms after stimulus onset represented the most critical time window for explicit identification of emotions from nonverbal vocalizations. Further, the amount of vocal information that is necessary to decode the emotional meaning of nonverbal vocalizations varied as a function of emotion category: the recognition of amusement relied on the shortest acoustic signal, whereas the recognition of anger and fear required longer acoustic samples. As for speech prosody, these results provide further support for the notion that critical acoustic-based emotional cues unfold over the course of a nonverbal vocalization at different time points. Specifically, they suggest a facilitated decoding of amusement (laughter), which may be related to its social significance and acoustic distinctiveness in nonverbal social communication. The time course for recognition of a speaker's emotional state has received little attention so far. This study provides the starting point for further research aiming to unveil temporal effects on vocal emotional perception.

Acknowledgements The authors are grateful to all participants who took part in this study.

Funding This work was supported by a Doctoral Grant SFRH/BD/92772/2013 awarded to PC, and by Grants IF/00334/2012, PTDC/MHN-PCN/3606/2012, and PTDC/MHC-PCN/0101/2014 awarded to APP. These Grants were funded by the Science and Technology Foundation (Fundação para a Ciência e a Tecnologia - FCT, Portugal) and FEDER (European Regional Development Fund) through the European programs QREN (National Strategic Reference Framework) and COMPETE (Operational Programme 'Thematic Factors of Competitiveness').

Compliance with ethical standards

Conflict of interest No potential conflict of interest was reported by the authors.

References

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

Barr, R. G., Chen, S., Hopkins, B., & Westra, T. (1996). Crying patterns in preterm infants. *Developmental Medicine and Child Neurology, 38*(4), 345–355.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version, 1*(7), 1–23.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370.

Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society of London B, 361*(1476), 2091–2107.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129–135.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40*(2), 531–539.

Bergmann, G., Goldbeck, T., & Scherer, K. R. (1988). Emotionale Eindruckswirkung von prosodischen Sprechmerkmalen (The effects of prosody on emotion inference). *Zeitschrift fur Experimentelle und Angewandte Psychologie, 35,* 167–200.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer. 2009. Computer program available at http://www.praat.org

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

Caron, J. E. (2002). From ethology to aesthetics: Evolution as a theoretical paradigm for research on laughter, humor, and other comic phenomena. *Humor, 15*(3), 245–282.

Cedrus Corporation. (1991). Super Lab, general purpose psychology testing software.

Collignon, O., Girard, S., Gosselin, F., Saint-Amour, D., Lepore, F., & Lassonde, M. (2010). Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia, 48*(1), 220–225.

Cornew, L., Carver, L., & Love, T. (2010). There's more to emotion than meets the eye: A processing bias for neutral content in the domain of emotional prosody. *Cognition and Emotion, 24*(7), 1133–1152.

Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication, 40*(1–2), 5–32.

Edmonson, M. S. (1983). Notes on laughter. *Anthropological Linguistics, 29,* 23–33.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169–200.

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

Greatbatch, D., & Clark, T. (2003). Displaying group cohesiveness: Humour and laughter in the public lectures of management gurus. *Human Relations, 56*(12), 1515–1544.

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

Hendriks, M. C. P., Croon, M. A., & Vingerhoets, A. J. J. M. (2008). Social reactions to adult crying: The help-soliciting function of tears. *The Journal of Social Psychology, 148*(1), 22–42.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods, 39*(1), 101–117.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology, 75*(4), 887.

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

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 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*(5), 770.

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

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models. R package Version 2.0-20 [Computer software]. Retrieved April 15, 2016.

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology, 21*(4), R143–R145.

Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion, 5*(3), 277–295.

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 vocalizations. *Frontiers in Psychology, 4,* 353.

Lima, C. F., Anikin, A., Monteiro, A. C., Scott, S. K., & Castro, S. L. (2018). Automaticity in the recognition of nonverbal emotional vocalizations. *Emotion, 19*(2), 219–233.

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

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport, 23*(2), 108–112.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86–92.

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review, 28*(2), 295–314.

Mehu, M., & Dunbar, R. I. (2008). Naturalistic observations of smiling and laughter in human group interactions. *Behaviour, 145*(12), 1747–1780.

Meneses, J. A. C., & Díaz, J. M. M. (2017). Vocal emotion expressions effects on cooperation behavior. *Psicológica, 38,* 1–24.

Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology, 64*(5), 723–739.

Naranjo, C., Kornreich, C., Campanella, S., Noël, X., Vandriette, Y., Gillain, B., et al. (2011). Major depression is associated with impaired processing of emotion in music as well as in facial and vocal stimuli. *Journal of Affective Disorders, 128*(3), 243–251.

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature, 1*(3), 261–289.

Paquette, S., Peretz, I., & Belin, P. (2013). The "Musical Emotional Bursts": A validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in Psychology, 4,* 509.

Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo-and lexical-sentence context. *Brain and Language, 105*(1), 59–69.

Paulmann, S., & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: How much is enough? *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 230–242.

Pell, M. D. (2002). Evaluation of nonverbal emotion in face and voice: Some preliminary findings on a new battery of tests. *Brain and Cognition, 48*(2–3), 499–514.

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE, 6*(11), e27256.

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

Pinheiro, A. P., Barros, C., Dias, M., & Kotz, S. A. (2017a). Laughter catches attention! *Biological Psychology, 130,* 11–21.

Pinheiro, A. P., Barros, C., Vasconcelos, M., Obermeier, C., & Kotz, S. A. (2017b). Is laughter a better vocal change detector than a growl? *Cortex, 92,* 233–248.

Pinheiro, A. P., Del Re, E., Mezin, J., Nestor, P. G., Rauber, A., McCarley, R. W., et al. (2013). Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: An electrophysiological investigation. *Psychological Medicine, 43*(3), 603–618.

Pinheiro, A. P., Rezaii, N., Rauber, A., Liu, T., Nestor, P. G., McCarley, R. W., et al. (2014). Abnormalities in the processing of emotional prosody from single words in schizophrenia. *Schizophrenia Research, 152*(1), 235–241.

Rigoulot, S., Wassiliwizky, E., & Pell, M. D. (2013). Feeling backwards? How temporal order in speech affects the time course of vocal emotion recognition. *Frontiers in Psychology, 4*(367), 1–14.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews, 32*(4), 863–881.

Salasoo, A., & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language, 24*(2), 210–231.

Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience, 22*(3), 474–481.

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

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

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*(3), 192–199.

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

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

Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion, 7,* 158–171.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*(1), 24–30.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14*(2), 228–233.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005a). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research, 24*(3), 442–452.

Schirmer, A., Simpson, E., & Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research, 1176,* 103–112.

Schirmer, A., Striano, T., & Friederici, A. D. (2005b). Sex di¡erences in the preattentive processing of vocal emotional expressions. *NeuroReport, 16*(6), 635–639.

Schirmer, A., Zysset, S., Kotz, S. A., & Von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage, 21*(3), 1114–1123.

Schlegel, K., Vicaria, I. M., Isaacowitz, D. M., & Hall, J. A. (2017). Effectiveness of a short audiovisual emotion recognition training program in adults. *Motivation and Emotion, 41*(5), 646–660.

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

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion, 9*(6), 838–846.

Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research, 23*(4), 347–365.

Stapells, D. R. (2002). Cortical event-related potentials to auditory stimuli. *Handbook of Clinical Audiology, 5,* 378–406.

Van Bezooijen, R. (1984). *Characteristics and recognizability of vocal expressions of emotion* (Vol. 5). Berlin: Walter de Gruyter.

Vasconcelos, M., Dias, M., Soares, A. P., & Pinheiro, A. P. (2017). What is the melody of that voice? Probing unbiased recognition accuracy of nonverbal vocalizations with the Montreal Affective Voices. *Journal of Nonverbal Behavior, 41*(3), 239–267.

Vettin, J., & Todt, D. (2004). Laughter in conversation: Features of occurrence and acoustic structure. *Journal of Nonverbal Behavior, 28*(2), 93–115.

Vingerhoets, A., Bylsma, L., & Rottenberg, J. (2009). Crying: A biopsychosocial phenomenon. *Tears in the Graeco-Roman World* 439–475.

Zimmer, U., Höfler, M., Koschutnig, K., & Ischebeck, A. (2016). Neuronal interactions in areas of spatial attention reflect avoidance of disgust, but orienting to danger. *NeuroImage, 134,* 94–104.