



Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody



M.D. Pell^{a,b,*}, K. Rothermich^a, P. Liu^a, S. Paulmann^c, S. Sethi^a, S. Rigoulot^b

^a School of Communication Sciences and Disorders, McGill University, Montreal, Canada

^b International Laboratory for Brain, Music, and Sound Research, Montreal, Canada

^c Department of Psychology and Centre for Brain Science, University of Essex, Colchester, United Kingdom

ARTICLE INFO

Article history:

Received 17 March 2015

Received in revised form 4 August 2015

Accepted 19 August 2015

Available online 22 August 2015

Keywords:

Emotional communication

Vocal expression

Speech prosody

ERPs

Anxiety

ABSTRACT

This study used event-related brain potentials (ERPs) to compare the time course of emotion processing from non-linguistic vocalizations versus speech prosody, to test whether vocalizations are treated *preferentially* by the neurocognitive system. Participants passively listened to vocalizations or pseudo-utterances conveying anger, sadness, or happiness as the EEG was recorded. Simultaneous effects of vocal expression type and emotion were analyzed for three ERP components (N100, P200, late positive component). Emotional vocalizations and speech were differentiated very early (N100) and vocalizations elicited stronger, earlier, and more differentiated P200 responses than speech. At later stages (450–700 ms), anger vocalizations evoked a stronger late positivity (LPC) than other vocal expressions, which was similar but delayed for angry speech. Individuals with high trait anxiety exhibited early, heightened sensitivity to vocal emotions (particularly vocalizations). These data provide new neurophysiological evidence that vocalizations, as evolutionarily primitive signals, are accorded precedence over speech-embedded emotions in the human voice.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Consider the last social gathering you attended where different groups of people were talking and circulating from room to room. At one point, you might overhear someone behind you speaking in a notably animated, joyful tone of voice; then, later in the evening, you hear someone crying or perhaps shouting at another guest. How does the brain respond to these different auditory signals of emotion? Which cues take precedence as we hear them? And how are these expressions elaborated over time to determine their meaning in the social context? These questions form the basis of the present study which compared the temporal dynamics of neural responses evoked by emotional auditory cues embedded in speech and in human vocalizations of a non-linguistic nature, based on their effects on “early” and “late” components in the event-related brain potential (ERPs). As a secondary goal, we also monitored whether individual variables such as anxiety influence how vocal signals are processed in the brain, given certain indications in the literature (Martin-Soelch, Stöcklin, Dammann, Opwis, & Seifritz, 2006; Sander et al., 2005; Schirmer & Escoffier, 2010).

1.1. Non-linguistic vocalizations versus speech-embedded emotions

While less empirical attention has been paid to how humans convey emotion in the auditory modality, our example underscores the rich functional significance of vocal signals as a means for communicating emotions, with clear advantages over visual signals in their *detectability* by conspecifics over broad distances and in the absence of joint visual attention (Hawk, van Kleef, Fischer, & van der Schalk, 2009). In fact, vocal emotions can serve a critical role in *directing* visual attention to emotionally relevant events in the environment such as faces (Brosch, Grandjean, Sander, & Scherer, 2008; Paulmann, Titone, & Pell, 2012; Rigoulot & Pell, 2014). In terms of composition, vocal emotion expressions are characterized by ongoing changes in several acoustic parameters that are meaningful to listeners *in combination*—principally, these involve differences in pitch, loudness, rhythm, and voice quality (Banse & Scherer, 1996; Juslin & Laukka, 2003). Critical to this study, vocal correlates of emotion can be operationalized according to the extent that they have been “tamed” or socially- adapted to the context of spoken language (Bolinger, 1978; Scherer, 1994); speech- embedded emotions, or *emotional prosody*, must be encoded by a speaker while preserving fine-level segmental (phonemic) contrasts necessary for linguistic communication and speech intelligibility. In contrast, *non-linguistic vocalizations*, such as cries, moans, growls,

* Corresponding author at: School of Communication Sciences and Disorders, McGill University, Montreal, Canada.

E-mail address: marc.pell@mcgill.ca (M.D. Pell).

or laughter, are described as ‘raw’, ‘pure’, or ‘primitive’ forms of vocal expressions that are largely unconstrained by linguistic structure (Crystal, 1969; Sauter & Eimer, 2009; Scott et al., 1997).

Non-linguistic vocalizations (or simply “vocalizations” hereafter) are considered ‘pure’ in the sense that they more closely reflect concomitant (psycho) physiological and autonomic changes affecting the respiratory and articulatory musculature (Scherer, 1986). Indeed, even when vocalizations take the form of interjections that have a certain speech-like quality (e.g., *huh!* *Aahh.* ...), they are characterized as more reflexive than emotional prosody, produced primarily at the glottal/sub-glottal level with little or no active control of vocal tract configurations (Trouvain, 2014). The lack of physiological constraints while expressing emotion through vocalizations is associated with greater acoustic variability across speakers (Jessen & Kotz, 2011; Lima, Castro, & Scott, 2013).

Based on forced-choice tasks that evaluate the end-point of the recognition process, it is clear that basic emotions can be identified in an accurate, differentiated manner from both speech prosody (Jaywant & Pell, 2012; Pell & Kotz, 2011; Scherer, Banse, Wallbott, & Goldbeck, 1991) and vocalizations (Lima et al., 2013; Sauter & Eimer, 2009). When directly compared (which is rare), listeners tend to be superior at labeling vocalizations over prosody, especially for *anger*, *joy*, *sadness*, *disgust* and *fear* (Hawk et al., 2009). In terms of the recognition time course, behavioral data on speech-embedded emotions using the auditory gating technique show that acoustic patterns differentiate rapidly to reveal basic emotional meanings to listeners after hearing approximately 400–800 ms of an utterance (Jiang, Paulmann, Robin, & Pell, 2015; Rigoulot, Wassiliwizky, & Pell, 2013). However, it is unclear in the behavioral literature whether emotional vocalizations are recognized over a similar time course; moreover, the temporal properties of *on-line* recognition processes when comparing emotional prosody and vocalizations are largely unknown.

1.2. ERP studies of vocal emotion expressions

A more sensitive comparison of the real-time temporal processing characteristics of emotional prosody and vocalizations can be achieved by examining ERPs evoked as listeners are exposed to each stimulus type. To date, there has been no focused attempt to directly compare electrophysiological responses to each stimulus type in an a priori, controlled manner, although research has combined speech and non-speech vocal stimuli in the same experiment to differentiate these stimuli from other sound categories such as non-human sounds or music (Charest et al., 2009; Rigoulot, Pell, & Armony, 2015). In studies involving speech, accumulating data show that the P200 amplitude is modulated by (pseudo)-utterances or words encoding different vocal emotion meanings (Paulmann, Bleichner, & Kotz, 2013; Schirmer & Escoffier, 2010; Schirmer, Kotz, & Friederici, 2005) as well as differences in vocally-expressed confidence (Jiang & Pell, 2015). According to Paulmann et al. (2013), the emotional salience of prosodic information is robustly encoded around 200ms after exposure to speech-embedded emotions, with increased P2 amplitudes for *motivationally significant* events; this explanation may account for reported discrepancies in how emotionality influences the P2 wave for specific expression types (e.g., Paulmann & Kotz, 2008; Schirmer, Chen, Ching, Tan, & Hong, 2013), since attentional shifts relevant to significance detection are likely to vary as a function of task goals. The idea that the P2 indexes early processing of semantically-relevant acoustic parameters and associative details, and not just sensory details of the acoustic input, remains debatable but fits with other data highlighting emotion-specific congruency effects of vocal cues on a face when only 200 ms of acoustic input was presented to listeners (Paulmann & Pell, 2010).

Similar modulations in the P200 time window have been observed in response to emotional interjections and non-linguistic vocalizations (Charest et al., 2009; Jessen & Kotz, 2011; Sauter & Eimer, 2009), a pattern that Sauter & Eimer attributed to differences in stimulus *arousal* rather than emotionality. In fact, Paulmann et al.’s (2013) data on speech prosody imply that both discrete emotional qualities and general arousal could contribute to rapid detection of “motivationally-salient” acoustic signals, depending on the processing environment. The possibility that even earlier effects of emotionality are encoded through differences in the amplitude and latency of the N100 response is also suggested by some data on vocalizations (Jessen & Kotz, 2011; Liu et al., 2012). These findings imply that specific emotion effects in the ERP data could emerge at an earlier time point for vocalizations than for speech prosody, although this hypothesis has not been tested comprehensively. Neural responses to different types of emotional auditory stimuli could also vary according to individual differences such as anxiety (Schirmer & Escoffier, 2010), neuroticism (Brück, Kreifelts, Kaza, Lotze, & Wildgruber, 2011), cultural background (Liu, Rigoulot, & Pell, 2015), and/or social orientation (Ishii, Morita, Shouji, Nakashima, & Uchimura, 2010). These and related data (Rigoulot et al., 2015; Stekelenburg & Vroomen, 2007) motivate a concerted look at how ERP components associated with attention, acoustic-sensory processing, and early meaning detection (N1–P2 complex) are modulated by “pure” versus speech-embedded vocal cues that convey basic emotions, and how these patterns are informed by listener characteristics.

Later processing stages permit a fine-grained analysis of the emotional meaning of evolving vocal expressions, with enriched cognitive elaboration of their social relevance in reference to memory representations (Kotz & Paulmann, 2011; Schirmer & Kotz, 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009). These cognitive stages have been tied to alterations in N300 (Bostanov & Kotchoubey, 2004) or N400 (Liu et al., 2015; Paulmann & Pell, 2010; Schirmer & Kotz, 2003) responses to non-linguistic and speech-embedded emotions, respectively, using priming or conflict paradigms that pair emotional voices with a related or unrelated stimulus (word or face). Recent data have also linked ongoing semantic analysis of vocal emotion expressions to changes in the late positive component (LPC), which exhibits differences according to emotion type (Jessen & Kotz, 2011; Paulmann et al., 2013; Schirmer et al., 2013). For example, it was reported that LPC amplitudes differ significantly for six basic emotions, and a stronger positivity was generally observed for expressions high in arousal, at central-posterior brain regions approximately 400–800 ms post-onset of emotional prosody (Paulmann et al., 2013). These findings converge with behavioral data arguing that discrete emotional representations underlying speech prosody are first registered in this time window (Pell & Kotz, 2011; Pell & Skrup, 2008; Pell, 2005b). Corresponding data showing LPC modulations for non-linguistic vocalizations are still largely unavailable (however, note that Jessen and Kotz (2011) found no significant LPC amplitude differences between anger, fear, and neutral vocalizations in the “audio only” condition of their study). These various findings, along with claims that the LPC reflects deeper analysis of emotionally relevant *visual* stimuli (Hinojosa, Carretié, Méndez-Bértolo, Míguez, & Pozo, 2009; Kanske & Kotz, 2007), act as a foundation to evaluate changes in the LPC to index enhanced cognitive processing of emotions expressed by vocalizations when compared to speech.

1.3. The current study

Building on related work from functional neuroimaging (e.g., Fecteau, Belin, Joannette, & Armony, 2007; Frühholz, Trost, & Grandjean, 2014), our goal was to shed new light on the *temporal* dynamics of processing two functionally related

auditory signals of emotion–emotional speech prosody and non-linguistic vocalizations—to investigate potential differences in neural response sensitivity at early and late on-line processing stages identified in previous research. Based on the literature, we hypothesized that differences in emotion would modulate P200 amplitudes, for both vocal expression types, and that “raw” emotions encoded by vocalizations may promote stronger, and possibly earlier (N100), effects in the ERP data than speech-embedded emotions. We also speculated that differences due to emotion and/or expression type observed at early processing stages, which reflect differences in motivational salience (Paulmann et al., 2013), would persist at later stages of enhanced emotion processing (Schirmer et al., 2013) with increased LPC amplitudes for highly arousing or socially salient expressions that require sustained analysis for adaptive behavior. Finally, as a corollary to our main goal, we explored how individual differences affect the processing of vocal emotions encoded by each stimulus type. There are especially strong indications that cognitive biases present in highly anxious individuals promote differential sensitivity to the type and quality of vocal emotion cues they encounter (Kreifelts et al., 2014; Martin-Soelch et al., 2006; Peschard, Maurage, & Philippot, 2014; Sander et al., 2005; Schirmer & Escoffier, 2010). To build on these insights, we therefore monitored the relationship between core measures of (state or trait) anxiety and personality characteristics of our participants in the context of processing emotional vocalizations and emotional speech.

2. Methods

2.1. Participants

Twenty-four native speakers of English completed the study (12 female/12 male, all right-handed according to self-report). Participants had a mean age of 22.9 (± 3.5) years and were all students at McGill University, Montréal, Canada. Participants volunteered to enter the study after responding to on-line campus advertisements; individuals with reported hearing difficulties were excluded from the experiment. Informed written consent was obtained before initiating the study, and each participant completed questionnaires to assess their state of alertness, anxiety (STAI, Spielberger, Jacobs, Russell, & Crane, 1983), and main personality traits (NEO-FFI-3, Costa & McCrae, 2010). Each participant received \$50CAD at the end of the study as an incentive to take part and to compensate for their time. The study was ethically approved by the Institutional Review Board of the Faculty of Medicine at McGill University.

2.2. Materials

Stimuli consisted of two types of vocal expressions, speech-embedded emotions (prosody condition) and non-linguistic vocalizations (vocalization condition), which communicated one of three basic emotions: anger, sadness, and happiness. We focused on these emotions because they are well studied in the immediate literature and have discrete expressive forms for each expression type (e.g., Hawk et al., 2009), while allowing potential differentiation in our results according to the valence, arousal, and discrete emotional qualities of the stimuli. All stimuli were taken from published inventories based on simulated expressions elicited from actors. Incidental to this report, all vocal expressions were followed by static pictures of facial expressions to investigate the contextual effects of different vocal expressions on face processing, the subject of a forthcoming companion study.

Stimuli in the prosody condition were seven-syllable pseudo-utterances selected from a database of recordings produced by male and female speakers of Canadian English (Pell et al., 2009; Pell,

2002). Pseudo-utterances are language-like stimuli frequently used in published work that concentrates on prosodic effects of speech independent of language content; here, these speech stimuli contained appropriate phonological and morpho-syntactic markers of English but were devoid of linguistic cues about emotion (e.g., *He placktered the tozz*). These items can be produced by speakers with relative ease to express basic emotions in a natural and culturally appropriate manner (Liu & Pell, 2012; Pell et al., 2009; Scherer, Banse, & Wallbott, 2001). Pseudo-utterances produced by 10 different speakers (6 female/4 male) to convey each of the three emotions were chosen (3 emotions \times 10 speakers = 30 total speech items). Mean group accuracy for these items in the respective validation studies, based on a seven emotion forced-choice recognition task, was 93% correct for anger, 91% for sadness, and 89% for happiness. Stimuli in the vocalization condition were non-linguistic vocalizations free of linguistic content, such as growls (anger), sobbing (sadness), or laughter (happiness), selected from Belin, Fillion-Bilodeau, & Gosselin (2008). Vocalizations of each emotion were produced by 10 speakers (5 female/5 male), yielding 30 unique vocalizations in the study (3 emotions \times 10 speakers) with an average recognition accuracy of 75% for anger, 77% for sadness, and 81% for happiness in the respective listener validation study (Belin et al., 2008). Apart from differences in how emotion recognition was characterized in each original dataset, the mean duration of selected recordings varied considerably due to differences in how speech and vocalization stimuli were initially constructed by the respective authors, as well as inherent differences in the temporal encoding of discrete emotions in the vocal channel (Juslin & Laukka, 2003). To control for effects of encountering varying amounts of acoustic information for our companion study on face processing, each pseudo-utterance was individually paired with a vocalization and edited to the same duration (mean duration for anger = 924 ms, sadness = 1990 ms, happiness = 1435 ms). Stimuli were also individually normalized to a peak intensity of 75 dB to mitigate gross differences in perceived loudness across stimuli/datasets. Basic acoustic parameters of the selected and edited stimuli are shown in Table 1.

To directly gauge how major perceptual features of our stimuli may have differed when the two databases are combined, all vocalizations and edited speech stimuli, as well as an equal number of neutral tokens selected from each inventory for the same speakers, were intermixed and presented in random order to a group of 14 listeners who did not take part in the subsequent EEG study (10 M/4F, native English speakers, undergraduate students at McGill University). Stimuli were presented in two separate tasks to gather descriptive data about core affective features of the stimuli (arousal/valence) and how well the items could be emotionally categorized. In the first task, participants rated “how aroused the speaker sounds” along a scale from 1 to 5 (where 1 = not at all, 5 = very much); in the second, they heard the same tokens presented in a different random order and made a two-part judgement of (a) “how pleasant the speaker sounds” (1 = not at all, 5 = very much), followed immediately by (b) “what emotion is the speaker expressing” (choices = angry, sad, happy, and neutral). These three measures, reported in Table 1, were then analyzed in a separate 2×4 (voice type \times emotion) ANOVA, including the filler category “neutral” which was only used in the pilot study (and not the EEG study) to help specify perceptual features of the emotional stimuli.

For accuracy, results showed that emotion categorization was significantly better overall for vocalizations ($M = 91\%$) than for speech prosody ($M = 74\%$, voice type main effect: $F(1,13) = 16.69$, $p = .001$). Accuracy also differed independently by emotion type, $F(3,11) = 8.24$, $p < .01$. Happiness ($M = 91\%$) was recognized significantly better than anger ($M = 72\%$) irrespective of voice type; expressions of sadness ($M = 82\%$) did not differ from either happiness or anger (neutral utterances, $M = 84\%$, were also recognized

Table 1

Perceptual and acoustic features of emotional vocalizations and speech stimuli at different time points in the stimulus. Emotions conveyed by vocalizations vs. speech prosody were produced by distinct groups of speakers ($n = 10/\text{voice type}$).

Measure	Stimulus type					
	Vocalization			Speech prosody		
	Angry	Sad	Happy	Angry	Sad	Happy
Full stimulus						
Emotion ID (%)	88	79	99	57	84	83
Arousal (1–5)	3.2	2.8	3.7	3.0	2.2	3.2
Pleasantness (1–5)	1.9	1.3	4.4	2.5	1.8	3.7
F0 mean (Hz)	297	284	269	242	208	257
F0 SD (Hz)	67	56	54	38	39	42
Intensity mean (dB)	73	66	65	71	72	73
Intensity SD (dB)	6	11	12	9	8	7
To end of N100 time window (0–160 ms)						
F0 mean (Hz)	317	289	224	214	230	250
F0 SD (Hz)	41	21	30	17	15	12
Intensity mean (dB)	74	71	73	69	72	75
Intensity SD (dB)	4	3	2	7	3	4
To end of P200 time window (0–300 ms)						
F0 mean (Hz)	328	305	252	239	237	265
F0 SD (Hz)	44	32	53	34	17	27
Intensity mean (dB)	76	73	73	70	71	74
Intensity SD (dB)	4	5	6	10	7	5

ID = identification, F0 = fundamental frequency, Hz = hertz, SD = standard deviation, dB = decibel.

better than anger). The interaction of voice type and emotion was not significant for accuracy. Ratings of perceived arousal yielded significant effects of voice type ($F(1,13) = 5.55, p < .05$), emotion ($F(3,11) = 23.61, p < .001$), and the interaction of these variables, $F(3,11) = 7.85, p < .01$. In the case of speech, anger and happiness were perceived as significantly more aroused than sadness (and neutral), whereas for vocalizations arousal ratings did not differ for anger, happiness, or sadness (all emotions were perceived as more aroused than neutral vocalizations). Comparing between expression types, only happiness differed significantly in perceived arousal, with higher ratings for vocalizations (laughter) than for speech prosody (joyful speech tone).

Finally, ratings of perceived pleasantness (valence) of selected stimuli differed significantly by emotion ($F(3,11) = 51.26, p < .001$) and the combined effects of voice type and emotion ($F(3,11) = 25.70, p < .001$). In general, expressions of happiness ($M = 4.04$) were perceived as significantly more pleasant than expressions of anger ($M = 2.23$) and neutral ($M = 2.65$), which were both rated as significantly more pleasant than sadness ($M = 1.57$). Interestingly, the perceived pleasantness of positive emotions (i.e., happiness) was significantly *greater* when encoded by vocalizations versus speech, whereas the pleasantness of negative emotions (i.e., anger and sadness) was always significantly *lower* for vocalizations versus speech, implying that vocalizations tended to amplify the perceived positive or negative valence characteristics of emotional vocalizations over corresponding speech signals. There was no perceived difference in the pleasantness of neutral expressions encoded by vocalizations and speech.

2.3. Task and procedures

For the purpose of a companion study on how different types of auditory cues influence face processing, each vocal emotion expression was individually paired with a face for cross-modal presentation in the facial affect decision task (Pell, 2005a,b). The general structure of each trial was as follows: a prime stimulus (speech or vocalization) was presented, followed immediately by a static target face. Participants made a yes/no decision about whether the “facial expression represents an emotion”, where faces portrayed one of the three emotions of interest or a non-emotional

grimace. The processing of discrete emotions in this task is therefore considered implicit for both the unattended vocal stimulus (prime) and the subsequent face that is the object of explicit attention; participants are never required to name an emotion, only to access emotion knowledge to judge whether faces conform to stored prototypes, similar to a lexical decision (Pell, 2005a,b). Thus, while our task demands center on emotion evaluation, it is assumed that vocal emotion expressions encountered in the current study were processed in a relatively implicit manner, outside the focus of attention.

Participants were tested individually in an electrically shielded, sound-attenuated booth seated in front of a computer screen. Vocal emotion expressions were presented through insert-style earphones at a comfortable listening level that was individually adjusted at the onset of the experiment. Participants were told they would hear different types of sounds that might sound like “gibberish” at times, but that they should attend closely to the facial expression that follows the voice to decide whether or not it represents an emotion. No instruction was given to ignore the vocal expression, only that these would occur prior to the face. After participants judged the face (yes/no button press), there was a 2500 ms inter-trial interval when participants were encouraged to blink before the next vocal stimulus was played. In total, 1080 trials were presented during the experiment to allow for different voice-face pairings in a full factorial design; half of the trials involved speech stimuli ($n = 540$) and half involved vocalizations ($n = 540$). Each vocal stimulus (speech or vocalization) was repeated 18 times in the experiment, paired with different face targets, to create identical combinations of trials beginning with either speech stimuli or vocalizations (2 vocal expression types \times 3 emotions \times 10 speakers \times 18 different face conditions = 1080 total trials). Stimuli were presented in six blocks of 180 trials, with an equal proportion of trials containing speech and vocalizations pseudo-randomized within blocks (ensuring that no auditory stimulus was repeated across 12 consecutive trials within the same block). Participants always began with two practice blocks followed by the six experiment blocks counterbalanced for presentation order across participants. The entire experiment, including mandatory breaks and EEG preparation, lasted around 3 h with compensation of \$40CAD. Data for one participant who performed very poorly in the behavioral task

(44% errors across trials, where the group mean was 91% correct, $SD = 12\%$) were removed from further analysis as this individual may not have understood task goals.

2.4. EEG recording

The EEG was recorded by 64 active Ag/AgCl electrodes mounted in an elastic cap (actiCAP, Brain products); we gathered scalp readings from the whole head to analyze ERPs evoked by vocal as well as facial expressions (in our companion report), associated with different spatial distributions. For artifact rejection, horizontal and vertical electrooculogram recordings were taken from two additional electrodes placed at the outer canthus of each eye, and two more placed above and below the right eye. The signal was recorded continuously in reference to FCz with a band pass between DC and 125 Hz, digitized at a sampling rate of 250 Hz, maintaining impedance of all channels below 5 K Ω . The continuous EEGs were first visually inspected and signals with excessive movement artifact, alpha activity or amplifier saturation were manually excluded from analysis. The subsequent data were re-referenced offline to the average of the electrodes and then filtered with a band pass of 0.01 and 30 Hz using EEGLab (Delorme & Makeig, 2004). ERPs were baseline-corrected, time-locked to the onset of the vocal expression and averaged to 800 ms post-stimulus onset (approximating the shortest vocal stimulus) with a 200 ms pre-stimulus baseline. Rejection of ocular artifacts (e.g., blinks) and drifts was performed on the epoched data by automatically rejecting VEOG-artifacts above 75 μ V, and segments in other channels with voltage deflections exceeding 200 μ V within a 1000-ms sliding window, followed by manual inspection of the data. Approximately 38% of trials were rejected in this manner across participants. The remaining artifact-free trials were then averaged separately for each experimental condition and participant.

2.5. ERP analysis

Recall that participants listened passively to vocal expressions before rendering an emotional judgment about a face that could not be predicted from the preceding auditory context; thus, no behavioral measures were relevant to our current analyses of ERP responses evoked by the auditory stimulus. For the ERP analyses, after rejections the mean number of trials per experimental condition and participant was 112 (VocAng = 115, VocSad = 112, VocHap = 114, SpeAng = 111, SpeSad = 110, SpeHap = 111). Mean amplitude and mean peak latency of components in the N100–P200 complex, as well as mean amplitude of the late positive component (LPC), evoked by different auditory stimuli were of a priori interest (Paulmann et al., 2013; Schirmer et al., 2013). Based on visual inspection of the averaged data and consistent with previous work (Jessen & Kotz, 2011; Paulmann et al., 2013), changes in the N1–P2 complex were analyzed in the fronto-central region and surrounding electrodes (19 electrodes: F3, F1, Fz, F2, F4, FC1, FC3, FC2, FC4, C3, C1, Cz, C2, C4, CP1CP3, CPz, CP2, and CP4). The N100, which showed a mean peak latency of 140 ms in the averaged data, was analyzed between 90–160 ms, whereas the P200 was analyzed in the 170–300 ms time window (mean peak latency of 220 ms in the averaged data). Mean LPC amplitudes were analyzed in the 450–700 ms post-onset window at typical central posterior electrode sites (O1, Oz, O2, PO3, POz, PO4, P5, P3, P1, Pz, P2, P4, and P6). Mean amplitude and latency measures were entered into a series of repeated measures ANOVAs involving voice type (speech and vocalization) and emotion (anger, sadness, and happiness). The significance of main or interactive effects for mean amplitude (N100, P200, and LPC) referred to Greenhouse–Geisser corrected values, whereas analyses of peak latency (N100 and P200) used a jackknife-based correction procedure (Ulrich & Miller, 2001).

3. Results

3.1. N100

The N100 amplitude was significantly reduced overall for vocalizations when compared to speech (voice type effect, $F(1,22) = 19.12, p < .001$). There was no significant main or interactive effect of emotion for these data (F 's $< 1.0, p$'s $> .84$). In addition, N100 latencies varied according to differences in voice type and emotion, $F(2,44) = 3.99, p < .05$. Pairwise comparisons showed that the interaction was driven primarily by happy vocalizations (i.e., laughter), which were associated with reduced N1 latencies relative to sad vocalizations (and marginally reduced relative to anger). There were no emotion-related differences in N100 latency in the context of speech. For each expression type, N100 latencies differed only for happy (and not anger or sad) expressions, which were shorter for vocalizations/laughter than for happy prosody.

3.2. P200

The amplitude of the P2 wave was significantly larger for vocalizations than speech (voice type, $F(1,22) = 43.88, p < .001$), although these patterns varied by emotion in the form of a significant interaction ($F(2,44) = 9.82, p < .001$). Pairwise comparisons revealed an increased positivity for vocalizations over speech for anger and sadness ($p < .001$), with similar but marginal differences for happiness ($p = .06$). When each voice type was examined separately, P2 amplitudes evoked by vocalizations were significantly reduced for sadness when compared to other emotions (sad vs. happy: $p = .015$; sad vs. anger: $p = .011$). For speech, the P2 amplitude was reduced for both sadness and anger when compared to happiness (happy vs. sad: $p < .001$; happy vs. anger: $p < .001$).

Vocalizations were further marked by shorter P200 latencies than speech overall (voice type, $F(1,22) = 17.34, p < .001$). An interaction of voice type \times emotion ($F(2,44) = 41.10, p < .001$) indicated that this pattern was only significant for anger and happiness (not sadness). For vocalizations, the P2 latency was significantly different for all three emotional expressions, with shortest latencies for happiness ($M = 216$ ms), followed by anger ($M = 224$ ms), followed by sadness ($M = 235$ ms). No emotion-related differences in P200 latency were evoked by angry, sad, or happy speech prosody ($M = 234, 231, 234$ ms, respectively). The general influence of voice type on the N1–P2 complex, as well as the combined influence of voice type and emotion, is illustrated in Fig. 1.

3.3. Late positive component (LPC)

LPC amplitudes (450–700 ms) tended to be larger for vocalizations than for speech overall (voice type, $F(1,22) = 25.90, p < .001$). Also, a more positive-going response in this time window was associated with anger when compared to both sadness and happiness (emotion, $F(2,44) = 5.87, p < .01$). A marginally significant voice type \times emotion interaction ($F(2,44) = 2.72, p = .08$) suggested that the enhanced LPC response to anger over sadness and happiness was largely driven by vocalizations, as the type of emotional expression did not significantly affect the LPC amplitude in the context of speech. Further inspection of these data (Fig. 2A–B) implied that emotion effects on the LPC for speech may have occurred somewhat later than for vocalizations; to explore this possibility, a sliding window analysis (2 voice type \times 3 emotion \times 7 time window) was performed on the mean amplitudes at 50 ms-intervals in a larger time window terminating with the auditory stimulus (450–800 ms, Fig. 2C). Results pointed to a marginally significant interaction of voice type \times time window ($F(6,132) = 2.41, p = .08$); this reflected the overall trend for larger LPC ampli-

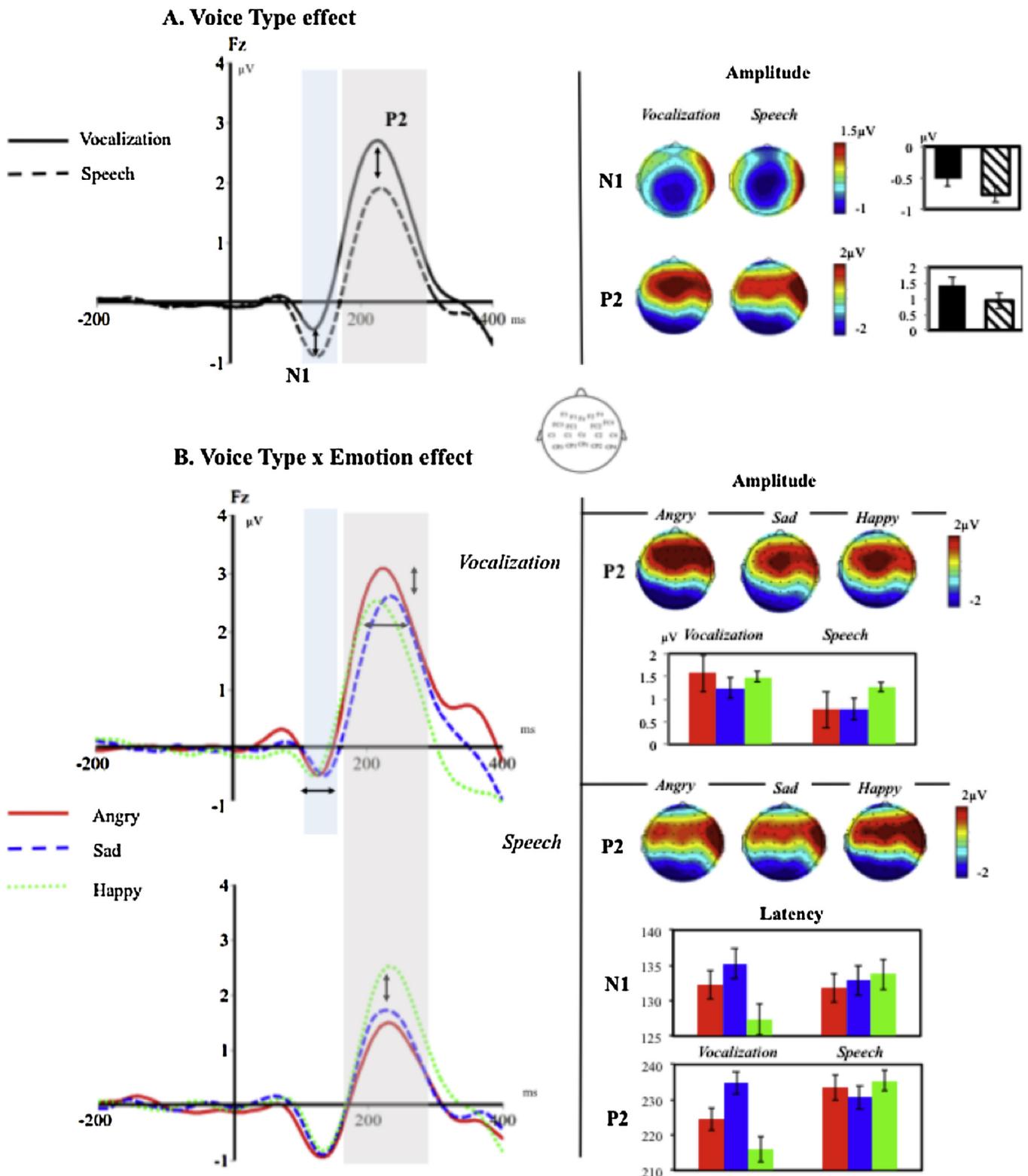


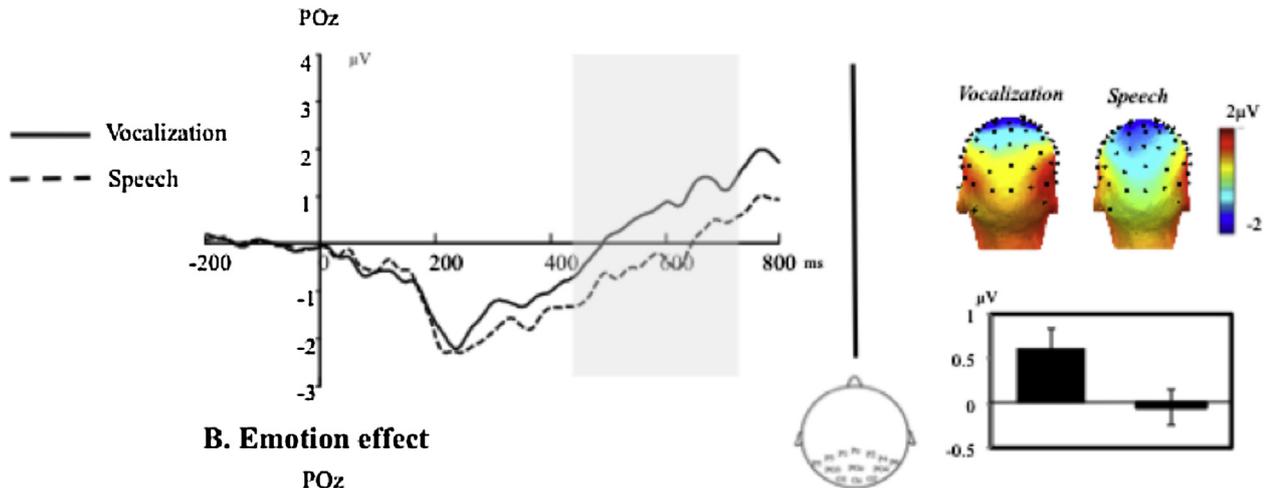
Fig. 1. Illustration of the N100 (90–160 ms) and P200 (170–300 ms) response as a function of (A) vocal expression type (non-linguistic vocalizations vs. speech prosody); and (B) voice type × emotion. Positive is plotted up. Bar graphs show the mean amplitude or latency values (plus standard error) per experimental condition.

tudes to vocalizations versus speech beginning immediately from the first analysis window (450 ms) to the end of the epoch. The interaction of emotion × time window was significant for this analysis ($F(12,264) = 31.31, p < .001$). Beginning at the 550–600 ms time window and at all subsequent intervals up to 800 ms, more

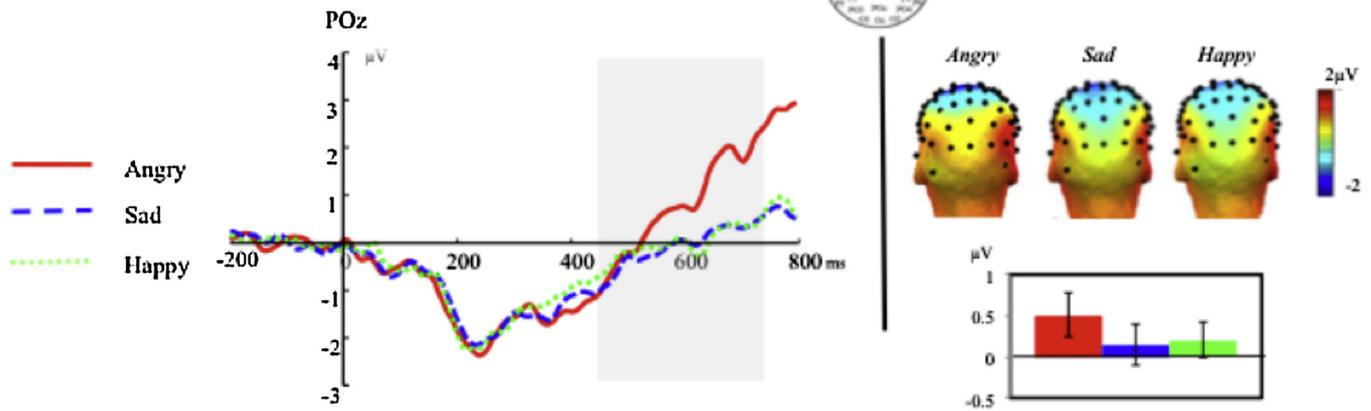
positive waveforms were observed for anger compared to sad and happy expressions, irrespective of voice type when the LPC time analysis window was lengthened. The three-way interaction of voice type, emotion, and time window did not reach statistical significance.

LPC: 450 – 700ms

A. Voice Type effect



B. Emotion effect



C. Extended Time window analysis

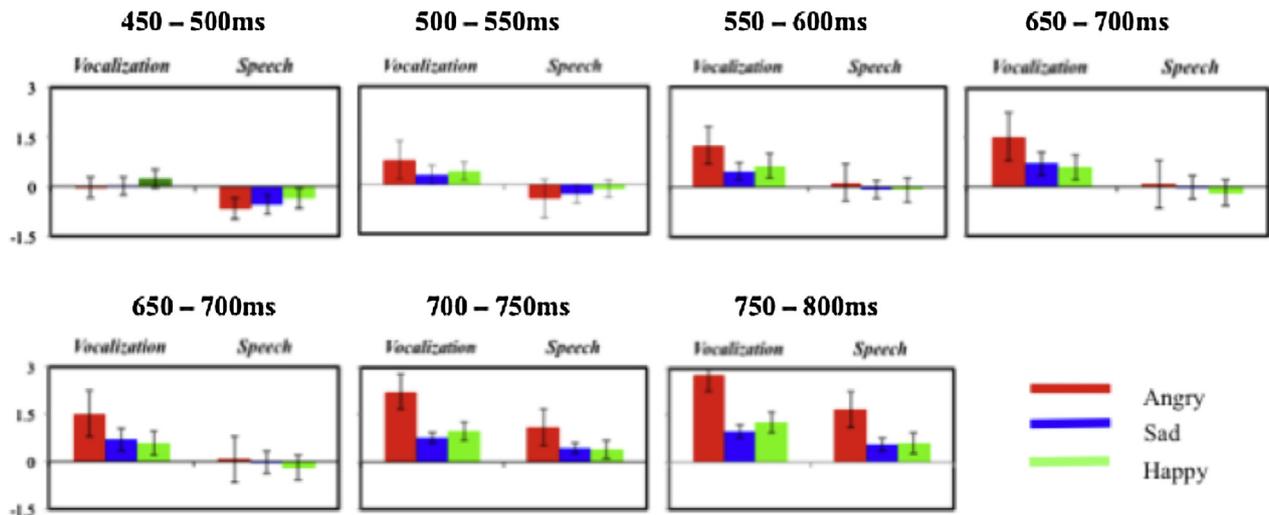


Fig. 2. Illustration of the late positive component (LPC, 450–700 ms) as a function of (A) vocal expression type (non-linguistic vocalizations vs. speech prosody); and (B) emotion type (anger, sadness, and happiness). Positive is plotted up. Figure C illustrates changes in mean amplitude values (plus standard error) over an extended LPC time window (450–800 ms), divided into seven consecutive 50 ms analysis windows, for each voice type and emotion.

3.4. Correlation analysis

At a final step, Pearson correlations (two-tailed, $p < .05$) were computed to examine the relationship between participant variables such as anxiety (state or trait, STAI raw scores) and the mean

amplitude of N1, P2, and LPC components as a function of voice type and emotion.¹ The N1 amplitude for vocalizations was signifi-

¹ NEO-FFI-3 personality dimensions (T scores) were entered into an initial analysis but removed for clarity of exposition in the text. In general, trait anxiety of

Table 2

Relationship between participant features (anxiety, personality dimensions) and mean amplitude of ERP response to emotional stimuli at early and late processing stages (collapsed across the three emotion types for expository purposes).

Measure	ERP component					
	N100 (90–160 ms)		P200 (170–300 ms)		LPC (450–700 ms)	
	Vocalization	Speech	Vocalization	Speech	Vocalization	Speech
Anxiety state	.381	.337	.160	.270	-.230	-.129
Anxiety trait	.417*	.287	.578**	.620**	-.231	-.408*
Neuroticism	.093	-.046	.357	.420*	-.002	-.217
Extraversion	.145	.181	.086	-.007	-.087	-.071
Openness	-.150	.004	.028	.052	.296	.166
Agreeableness	-.070	-.033	.235	.116	.103	.173
Conscientiousness	-.178	-.050	-.389	-.453*	-.043	.149

* $p < .05$, ** $p < .01$; LPC = late positive component.

cantly associated with trait (but not state) anxiety ($R = .417, p < .05$); there was no apparent relationship between anxiety scores and the N1 amplitude for speech overall. Closer inspection of these data by emotion type showed that the relationship between trait anxiety and N1 response to vocalizations was driven largely by anger ($R_{\text{angry}} = .509, p = .01$) and happiness/laughter ($R_{\text{happy}} = .383, p = .07$; all other correlations by emotion type were not significant for N1 amplitude, $p > .16$). For the P200, trait anxiety was positively associated with P2 amplitude for both vocalizations ($R_{\text{angry}} = .642, p < .001, R_{\text{sad}} = .596, p < .01, R_{\text{happy}} = .436, p < .05$) and speech prosody ($R_{\text{angry}} = .601, p < .01, R_{\text{sad}} = .558, p < .01, R_{\text{happy}} = .649, p < .001$) for all emotions. Thus, for both the N1 and P2 waves, increased anxiety levels were associated with amplified responses in these early components. For the LPC (450–700 ms), trait anxiety was *negatively* associated with the amplitude of responses evoked by speech, as participants with lower anxiety scores exhibited more positive-going LPC amplitudes for sad ($R = -.474, p < .05$) and happy ($R = -.414, p < .05$) prosody. No association was observed between anxiety and LPC amplitude for vocalizations. Finally, the amplitude of P2 and LPC responses was negatively correlated when processing speech, for angry ($R = -.406, p = .05$) and sad ($R = -.398, p = .06$) prosody. Enhanced P2 to angry and sad speech was associated with reduced LPC waves. There was no relationship between the P2 and LPC amplitude for vocalizations. Correlation results are supplied in full in [Table 2](#).

4. Discussion

Previous studies focusing on how the brain differentiates sound categories (e.g., vocal and environmental sounds, music) imply that human vocalizations and speech prosody are associated with early differences in neuronal activity in the 100–300 ms time window ([Charest et al., 2009](#); [Rigoulot et al., 2015](#)). However, speech and non-speech-embedded emotions were combined in most analyses conducted by previous authors. Here, we uncovered direct evidence that vocalizations are robustly differentiated from speech prosody, and seem to be treated *preferentially*, as early as 100 ms after listeners were exposed to acoustic input. Both amplitude and latency of the N100 and P200 were modulated by vocal expression type; for example, vocalizations were associated with *reduced* N1 amplitude and *enhanced* P2 waveform when compared to emotional speech.

individuals bore a significant positive relationship to neuroticism ($R = .821, p < .001$) and a negative relationship to conscientiousness ($R = -.676, p < .001$). The inter-relationship of these three variables and their impact on specific ERP responses are shown in [Table 2](#). No significant correlations were observed between state anxiety and the five personality variables.

4.1. Early decoding stages: sensory and motivational significance

In the context of auditory processing, the N1 component has been linked to early sensory encoding of physical properties of a stimulus (e.g., [Näätänen & Picton, 1987](#)) and effort associated with the allocation of processing resources to form and maintain a sensory memory trace ([Obleser & Kotz, 2011](#)). Attention effects on early sensory processing of an auditory stimulus (beginning as early as 20 ms post-stimulus) are also known to occur in tasks of focused attention to an auditory target, with increased N1 amplitudes when attended (e.g., [Woldorff et al., 1993](#)). With notable exceptions ([Jessen & Kotz, 2011](#); [Liu et al., 2012](#)), the amplitude of N1 seems to be insensitive to emotional/semantic characteristics of a stimulus, in keeping with our findings, although perceptual processes at this stage robustly differentiate sound categories, such as human and non-human sounds ([Charest et al., 2009](#)) and music versus human vocal sounds ([Kaganovich et al., 2013](#); [Meyer, Elmer, Baumann, & Jancke, 2007](#); [Rigoulot et al., 2015](#) for negative findings).

Our results extend literature on the auditory N1 by establishing that human vocal expressions of emotion encoded as non-linguistic vocalizations (or ‘affect bursts’, [Scherer, 1994](#)), and the same emotional signals embedded in speech, are registered as distinct sound categories as early as 100 ms after acoustic onset. Vocalizations, which are relatively unconstrained by the segmental structure of language, emanate in large part from glottal and sub-glottal activity associated with autonomic/physiological responses to emotion ([Trouvain, 2014](#)). This renders these expressions acoustically distinct and less complex in many ways from emotional prosody ([Meyer et al., 2007](#); [Szameitat et al., 2009](#)), since speech requires fine control of the entire vocal apparatus to preserve specific vocal tract configurations associated with segmental distinctions in linguistic communication. Given that we did not manipulate or direct attention to vocal stimuli in our task (participants always attended to a face), it seems likely that the perceptual-acoustic quality of vocalizations, such as growls and laughs, require fewer resources than speech to form a sensory trace of these events at initial stages of analysis. In this manner, the neurocognitive system appears to register vocalizations as distinct ‘objects’ from speech-based signals as early as 100 ms, analogous to differences between other sound categories that exemplify the functional specialization of the human brain for behaviorally relevant sounds ([Belin, Zatorre, Lafaille, Ahad, & Pike, 2000](#); [Formisano, De Martino, Bonte, & Goebel, 2008](#)).

As was true here, effects of stimulus *emotionality* are routinely observed after processing 200 ms of vocal expressions, with fronto-centrally distributed modulation of the P2 amplitude for different types of speech and non-speech vocal expressions ([Liu et al., 2012](#); [Paulmann et al., 2013](#); [Sauter & Eimer, 2009](#); [Schirmer et al., 2005](#)). The directionality of this effect remains controversial, perhaps owing to task-related variability across studies; whereas some

work reports enhanced P2 amplitudes for neutral over emotional prosody (Paulmann & Kotz, 2008), other studies show a larger P2 for emotional over neutral prosody (Schirmer et al., 2013) or vocalizations (Liu et al., 2012), or for vocalizations associated with high versus low arousal (Sauter & Eimer, 2009). Changes in P2 amplitude have been attributed to attentional shifts that facilitate *preferential processing* of emotionally or motivationally significant events (Paulmann et al., 2013; Paulmann, Ott, & Kotz, 2011). Neuronal activity in the P2 time window has also been tied to the auditory “what” sensory processing stream involved in object classification (Charest et al., 2009; Schirmer, Fox, & Grandjean, 2012), as this response seems to index *motivationally salient* voice qualities beyond emotions, such as voice familiarity, gender, speaker identity, and even speaker confidence (e.g., Beauchemin et al., 2006; Jiang & Pell, 2015).

In a study that presented non-linguistic vocalizations of happiness (*laughter*), anger (*humph*), and neutral (*mmm*), Liu et al. (2012) observed reduced N1 and increased P2 responses to emotional (angry and happy) vocalizations when compared to neutral ones; this pattern is analogous to the one witnessed here with reduced N1 and enhanced P2 responses to vocalizations versus speech prosody. On the basis that motivationally salient stimulus characteristics rapidly evoke sustained attention and preferential processing (Jessen & Kotz, 2011; Paulmann et al., 2013), it can be said that vocalizations elicited a preferential response in our experiment when compared to vocal expressions in speech, just as emotional vocalizations were treated preferentially over neutral sounds in Liu et al. (2012) study. Here, early preferential processing of vocalizations should facilitate initial stages of perceptual processing in the form of reduced N1 amplitudes, where speech is associated with more effortful (complex) demands on acoustic analysis (Meyer et al., 2007). At the same time, P2 amplitudes would be enhanced by rapid deployment of attentional resources to vocalizations at the stage of “salience detection”, where initial encoding of motivational properties of the stimulus takes place (e.g., emotionality, valence, and arousal). This result fits the notion that non-linguistic vocalizations are more primitive and salient signals of emotion than spoken utterances in an evolutionary sense (Scherer, 1994).

Further evidence that vocalizations were treated preferentially at initial stages of perceptual decoding and salience detection is that P2 *latencies* were significantly reduced overall for vocalizations than speech prosody for most emotions (except sadness). In fact, P2 latencies displayed a highly differentiated pattern *only* for different types of vocalizations: peak latencies occurred earlier on average for happiness (laughter), followed by anger (growls), and then sadness (sobs). In contrast, no differences in P2 latency were noted among emotions for speech prosody. Inspection of major acoustic parameters of the stimuli up to the end of the P2 time window (0–300 ms, Table 1) does not readily explain P2 modulation based solely on mean acoustic differences and/or the time point of local pitch or amplitude peaks in the stimulus. Rather, these results suggest that during early emotion evaluation, vocalizations not only take general precedence in their encoding, but that the brain encodes the distinct qualities of each type of vocalization for rapid identification of these signals, at an earlier time point than when listeners are exposed to speech. Registering the specific emotional qualities of vocalizations in the P200 time window may be viewed as another example of how the neurocognitive apparatus quickly ‘tags’ the motivational salience of this class of signals to facilitate a behaviorally adaptive response, given that the consequences of vocalizations and related action tendencies vary considerably in their biological significance to the organism.

In terms of the impact of specific emotions on early ERP components, vocal expressions of sadness, irrespective of their form of expression, were associated with reduced P2 amplitude when

compared to other emotions (Paulmann et al., 2013) and a delayed P2 response for sad vocalizations. Since events high in arousal tend to be more motivationally salient (Lee et al., 2014; Mather & Sutherland, 2011), it is possible that the P2 response is partly sensitive to general arousal features of vocal expressions, with increased response to stimuli that are high in arousal irrespective of their positive/negative valence (Sauter & Eimer, 2009 for vocalizations; see also trends in Paulmann et al., 2013 for speech prosody). As our sad stimuli were rated as lower in arousal than other emotions (and speech was perceived as less arousing than vocalizations overall), sensitivity to arousal features in the P2 time window could explain some of the ERP patterns we observed, pending new studies that employ a wider range of stimulus types. Another pattern of interest in our data refers to laughter (i.e., joyful vocalizations), which uniquely displayed an early N1 peak and earliest P2 response when compared to all other voice/emotion types. Laughter in our study was rated as notably more aroused *and* more pleasant than all other stimulus types (Table 1), underscoring the distinctiveness of these signals and the likelihood that they elicited positive affect, since different types of laughter (joyful, taunting, etc.) are associated with distinct behavioral responses and neural activation patterns (Bachorowski & Owren, 2001; Wildgruber et al., 2013). Similarly, responses to *spontaneous* or reflexive laughter, which are presumably captured by the stimuli we used (Belin et al., 2008), are known to differ from those evoked by volitional, social forms of laughter (Bryant & Aktipis, 2014; Szameitat et al., 2010). Our data add to this literature by showing that neural responses to laughter are associated with a distinct *time course* at early processing stages, an observation that should be examined in a more concerted manner in future studies.

4.2. Late decoding stages: cognitive elaboration of social significance

A deeper fine-grained analysis of vocal expressions in the context of evolving stimulus properties, stored memory representations, and task demands occurs at a later processing stage, evoking different ERP responses sensitive to vocal cues in the 400–800 ms post-stimulus time window (Bostanov & Kotchoubey, 2004; Jiang & Pell, 2015; Rigoulot, Fish, & Pell, 2014; Schirmer & Kotz, 2003, 2006). Emotion-specific modulation of the late positive component (LPC) has been observed for speech prosody (Paulmann et al., 2013; Schirmer et al., 2013), with a more positive-going wave at central-posterior brain sites for expressions high in arousal (Paulmann et al., 2013; Jessen & Kotz, 2011 who reported no significant LPC amplitude differences for anger, fear, and neutral vocalizations in the “audio” condition of their study). Here, the LPC amplitude between 450–700 ms was strongly influenced by both voice and emotion type; *angry vocalizations* (i.e., growls) exhibited a sustained, more positive-going wave than sad or happy vocalizations in this time frame (beginning around 550 ms). Subsequent analyses pinpointed a similar effect of *angry speech* on the LPC after 650 ms, with increased positivity over sad and happy prosody. Note that these patterns do not align neatly with underlying valence-arousal dimensions of the stimuli as perceived by our pilot raters, who judged anger to be *less* aroused than happiness but *more* pleasant than sadness.

These findings underscore qualitative differences in how discrete expressions of anger influence later processing stages where in-depth (or second pass) analysis of vocal signals is thought to occur (Paulmann et al., 2013), while showing that this neural response can be detected earlier when anger is encoded *non-linguistically*. Of the three emotions studied here, anger is uniquely associated with external signals of aggression and threat, which tend to promote increased psychophysiological responses (Öhman, 1987; Siegman et al., 1990) and capture and *hold* attention to a

greater extent than non-aversive stimuli when facial expressions are presented (e.g., [Koster, Crombez, Van Damme, Verschuere, & De Houwer, 2004](#)). Our ERP data imply that listeners automatically engage in sustained monitoring of anger voices, irrespective of its acoustic form, to fully specify the significance of these unique and potentially threatening events ([Frühholz & Grandjean, 2012](#)) and that some of these operations are indexed by changes in the late positivity. Quite possibly, these patterns are further influenced by task goals, stimulus factors (e.g., predictability) and individual variables that impact on attentional mechanisms and coping potential (e.g., [Mogg & Bradley, 1998](#)). For example, the observation that P2 amplitudes evoked by angry (and sad) speech were associated with significantly reduced LPC waves is informative, as it implies that individual differences in how salient negative prosody is encoded at early processing stages (stronger P2 response) influence the extent of cognitive analysis accorded to these signals “downstream” in the 400–750 time window (reduced LPC; see also [Schirmer et al., 2013](#)). New studies that examine our effects in the context of a broader array of vocal emotion types, and with fewer repetitions of the same expressions, will help to clarify and extend these claims. In addition, it will be important to corroborate our findings in a design where vocal stimuli are not accompanied by emotional faces, to ensure that our effects were not influenced by these stimuli.

4.3. Effects of anxiety on vocal emotion processing

Although not our primary focus, both early and late ERP responses were modulated by the trait anxiety level of participants within our sample (and related personality dimensions such as neuroticism), suggesting that the temporal neural dynamics for processing emotional sounds can be altered at several processing stages in anxiety-related disorders. It is well known that anxiety disorders are associated with maladaptive changes in information processing due to an attentional bias ([Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007](#); [Mogg & Bradley, 1998](#)) characterized by pre-attentive shifts in attention and enhanced processing of threat-related cues ([Koster, Crombez, Verschuere, Van Damme, & Wiersema, 2006](#); [Zvielli, Bernstein, & Koster, 2014](#)). Here, it seemed that high trait anxiety was linked to deeper/more effortful acoustic processing of non-linguistic vocalizations (promoting stronger N1 amplitudes for vocalizations, especially anger and laughter), followed by a broad-based increase in the deployment of attention to all vocal emotion expressions to determine their motivational relevance (i.e., increased P2 amplitudes to both vocalizations and speech in individuals with high versus low trait anxiety).

These results provide new evidence that individuals with trait anxiety preferentially allocate their attention to emotional voices at very early stages of stimulus encoding in the 100–200 ms time window ([Schirmer & Escoffier, 2010](#)), similar to early attentional biases for face stimuli ([Mogg, Philippot, & Bradley, 2004](#)). Moreover, individuals with high trait anxiety continue to show differences in their neural response sensitivity at later cognitive processing stages as they continue to process emotional speech. Our findings justify the opinion that heightened vigilance to emotional voices, not necessarily confined to signals with an overt negative valence, represents an important feature of trait anxiety and perhaps related disorders such as social phobia ([Peschard et al., 2014](#); [Quadflieg, Wendt, Mohr, Miltner, & Straube, 2007](#)), a topic that is ripe for exploration.

4.4. Neurocognitive mechanisms for processing emotion in and out of speech

Vocalizations of a non-linguistic nature (e.g., cries, growls, and laughs) and emotional speech prosody are both natural signals humans exploit to convey emotion, with inherent

communicative value and decoding mechanisms that differ from those used in other forms of communication such as language ([Wilson & Wharton, 2003](#)). However, when processing of these stimuli is examined with fine temporal resolution, our data underscore that vocalizations are processed preferentially by the neurocognitive system as distinct signals that are more motivationally salient than speech-embedded emotions, particularly in the early time course of auditory processing. Thus, while speech and non-speech-based vocal expressions share many core acoustic features that refer to basic emotions ([Juslin & Laukka, 2003](#)), vocalizations appear to have the advantage of conveying rich affective meanings that promote a more immediate, deeper analysis of the motivational properties of these signals when compared to many speech stimuli, measurable at different stages of the neuronal response ([Bestelmeyer et al., 2014](#)). This is consistent with neurophysiological studies on primary auditory cortex in nonhuman primates, which suggest that species-specific vocalizations are highly affective stimuli that can induce greater neural responses than other less behaviorally relevant sounds ([Wang & Kadia, 2001](#)).

In human communication, vocalizations are considered a “raw” form of emotional expression in the vocal channel because they are often the spontaneous product of reflexive (neuro) physiological responses to an emotional event ([Bostanov & Kotchubey, 2004](#)) without precise articulatory targets ([Trouvain, 2014](#)). In contrast, emotional prosody represents the socialization of emotional cues in speech behavior, a context where attention and mental resources are simultaneously deployed in service of linguistic information sharing; in this context the acoustic expression of segmental and suprasegmental features necessary for speech intelligibility must be rigorously preserved. Human systems for producing vocal emotions in their raw form (i.e., non-linguistic vocalizations) are phylogenetically older and share critical properties, including underlying neural organization, with the vocal call systems of other primates and many mammalian species ([Meyer et al., 2007](#); [Owren, Amoss, & Rendall, 2011](#)). However, as human language evolved in the face of an existing vocal system for expressing emotion, a functionally distinct, more volitional capacity for expressing emotions in speech may have simultaneously developed, supported by functionally distinct neural pathways. That is, vocal systems for communicating emotion in humans are likely supported by a dual neural pathway architecture and are partly independent ([Frühholz, Trost, & Grandjean, 2014](#); [Owren et al., 2011](#)).

At this time, our findings must be treated with some caution, given that early ERP effects are highly sensitive to basic differences in the physical (acoustic) properties of auditory events; given that the vocalizations and speech stimuli we presented here were produced by different speakers, and that there was a slight imbalance of male and female speakers in the two groups, it remains possible that some of the effects we report are not specific to vocal communication. Nonetheless, when combined with related behavioral and neuroimaging studies, our results allow speculation that emotional vocalizations are more salient to listeners, or given temporal precedence by the neurocognitive system, because they emanate from a functionally distinct, affectively-triggered neural control system dedicated to vocal production that is not consistently engaged by speech.

In the neuroimaging literature, there is evidence that vocalizations predominantly activate the amygdala and other subcortical structures ([Fecteau et al., 2007](#); [Phillips et al., 1998](#); [Sander & Scheich, 2005](#); [Szameitat et al., 2010](#)), whereas this evidence is less consistent for emotional prosody ([Wiethoff et al., 2009](#)). On the basis that vocalizations selectively trigger limbically-dominated processes that form part of the ancestral subcortical system for regulating emotional responses (e.g., [Frühholz, Trost, & Grandjean, 2014](#); [Jürgens, 2008](#); [Meyer et al., 2007](#); [Sander & Scheich, 2005](#)

Meyer et al., 2007; Sander & Scheich, 2005), activation of this circuitry would strengthen input used to tag the urgency and biological salience of vocal expressions to listeners by dedicated temporo-frontal cortical mechanisms that process voice information (Belin et al., 2000; Schirmer & Kotz, 2006; Schwartz & Kotz, 2013). In terms of the *time course* of emotion processing, our data show that this leads to preferential deployment of attention to the type and quality of different vocal signals in the 100–200 ms time window, even when vocal stimuli are not the subject of attentional focus (Aschliemann et al., 2008; Fecteau, Armony, Joannette, & Belin, 2004; Gädeke, Föcker, & Röder, 2013).

In contrast, processing emotional expressions within speech is likely dominated by species-specific, cortically-based circuitry involved in the volitional production of language and emotions, engaging a distinct (parallel) set of processing routines (Hertrich et al., 2008; Kalberlah et al., 2013; Owren et al., 2011; Owren et al., 2011). Pending new studies, especially those that consider the temporal specificity of these effects, it can be argued that as emotion expression and spoken language co-evolved, the need for greater emotional control in the context of speech behaviour (Jablonka, Ginsburg, & Dor, 2012) became associated with less urgent, more socially-constructed responses to emotional prosody at the neurocognitive level. Over time, this experience-guided learning would yield subtle changes in the nature and time course for processing emotions from vocalizations and speech prosody.

Acknowledgements

This research was supported by a Discovery Grant (RGPIN/203708-2011) awarded to M.D. Pell from the Natural Sciences and Engineering Research Council of Canada. We are grateful for the kindness and expertise of Dr. Xiaoming Jiang for helpful comments on the data analyses, and to Lila Weintraub for help with manuscript preparation.

References

- Bachorowski, J. A., & Owren, M. J. (2001). Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science: A Journal of the American Psychological Society/APS*, 12, 252–257. <http://dx.doi.org/10.1111/1467-9280.00346>
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636.
- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychol. Bull.*, 133(1), 1–24. <http://dx.doi.org/10.1037/0033-2909.133.1.1>
- Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., & Lassonde, M. (2006). Electrophysiological markers of voice familiarity. *European Journal of Neuroscience*, 23(April), 3081–3086. <http://dx.doi.org/10.1111/j.1460-9568.2006.04856.x>
- 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. <http://dx.doi.org/10.3758/BRM.40.2.531>
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403, 309–312. <http://dx.doi.org/10.1038/35002078>
- Bolinger, D. (1978). Intonation across languages. In J. H. Greenberg (Ed.), *Universals of human language*. Phonology (Vol. 2) (pp. 471–524). Stanford: Stanford University Press.
- Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, 41, 259–268. <http://dx.doi.org/10.1111/j.1469-8986.2003.00142.x>
- Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2008). Cross-modal emotional attention: emotional voices modulate early stages of visual processing. *Journal of Cognitive Neuroscience*, 21, 1670–1679. <http://dx.doi.org/10.1162/jocn.2009.21110>
- Brück, C., Kreifelts, B., Kaza, E., Lotze, M., & Wildgruber, D. (2011). Impact of personality on the cerebral processing of emotional prosody. *Neuroimage*, 58(1), 259–268. <http://dx.doi.org/10.1016/j.neuroimage.2011.06.005>
- Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evolution and Human Behavior*, 35(4), 327–335. <http://dx.doi.org/10.1016/j.evolhumbehav.2014.03.003>
- Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., et al. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience*, 10, 127. <http://dx.doi.org/10.1186/1471-2202-10-127>
- Costa, P. T., & McCrae, R. R. (2010). Bridging the gap with the five-factor model. *Personality Disorders: Theory, Research, and Treatment*, 1(2), 127–130. <http://dx.doi.org/10.1037/a0020264>
- Crystal, D. (1969). *Prosodic systems and intonation in English*. London: Cambridge University Press.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9–21. <http://dx.doi.org/10.1016/j.jneumeth.2003.10.009>
- Fecteau, S., Armony, J. L., Joannette, Y., & Belin, P. (2004). Priming of non-speech vocalizations in male adults: the influence of the speaker's gender. *Brain Cognition*, 55, 300–302. <http://dx.doi.org/10.1016/j.bandc.2004.02.024>
- Fecteau, S., Belin, P., Joannette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage*, 36(2), 480–487. <http://dx.doi.org/10.1016/j.neuroimage.2007.02.043>
- Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). Who is saying what? Brain-based decoding of human voice and speech. *Science*, 322(7 November), 970–973.
- Frühholz, S., & Grandjean, D. (2012). Towards a fronto-temporal neural network for the decoding of angry vocal expressions. *Neuroimage*, 62(3), 1658–1666. <http://dx.doi.org/10.1016/j.neuroimage.2012.06.015>
- Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in Neurobiology*, 123, 1–17. <http://dx.doi.org/10.1016/j.pneurobio.2014.09.003>
- Gädeke, J. C., Föcker, J., & Röder, B. (2013). Is the processing of affective prosody influenced by spatial attention? An ERP study. *BMC Neuroscience*, 14, 14. <http://dx.doi.org/10.1186/1471-2202-14-14>
- 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), 293–305. <http://dx.doi.org/10.1037/a0015178>
- Hinojosa, J. A., Carretié, L., Méndez-Bértolo, C., Míguez, A., & Pozo, M. A. (2009). Arousal contributions to affective priming: electrophysiological correlates. *Emotion*, 9(2), 164–171. <http://dx.doi.org/10.1037/a0014680>
- Ishii, Y., Morita, K., Shouji, Y., Nakashima, Y., & Uchimura, N. (2010). Effects of emotionally charged sounds in schizophrenia patients using exploratory eye movements: comparison with healthy subjects. *Psychiatry Clinical Neurosciences*, 64, 10–18. <http://dx.doi.org/10.1111/j.1440-1819.2009.02043.x>
- Jablonka, E., Ginsburg, S., & Dor, D. (2012). The co-evolution of language and emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2152–2159. <http://dx.doi.org/10.1098/rstb.2012.0117>
- Jaywant, A., & Pell, M. D. (2012). Categorical processing of negative emotions from speech prosody. *Speech Communication*, 54(1), 1–10.
- Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage*, 58(2), 665–674. <http://dx.doi.org/10.1016/j.neuroimage.2011.06.035>
- Jiang, X., Paulmann, S., Robin, J., & Pell, M. D. (2015). More than accuracy: nonverbal dialects modulate the time course of vocal emotion recognition across cultures. *Journal of Experimental Psychology: Human Perception and Performance*, 41(3), 597–612. <http://dx.doi.org/10.1037/xhp0000043>
- Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, 66, 9–34. <http://dx.doi.org/10.1016/j.cortex.2015.02.002>
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, 129(5), 770–814. <http://dx.doi.org/10.1037/0033-2909.129.5.770>
- Kaganovich, N., Kim, J., Herring, C., Schumaker, J., Macpherson, M., & Weber-Fox, C. (2013). Musicians show general enhancement of complex sound encoding and better inhibition of irrelevant auditory change in music: an ERP study. *European Journal of Neuroscience*, 37, 1295–1307. <http://dx.doi.org/10.1111/ejn.12110>
- Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, 1148, 138–148. <http://dx.doi.org/10.1016/j.brainres.2007.02.044>
- Koster, E. H. W., Crombez, G., Van Damme, S., Verschuere, B., & De Houwer, J. (2004). Does imminent threat capture and hold attention? *Emotion (Washington, D. C.)*, 4(3), 312–317. <http://dx.doi.org/10.1037/1528-3542.4.3.312>
- Koster, E. H. W., Crombez, G., Verschuere, B., Van Damme, S., & Wiersema, J. R. (2006). Components of attentional bias to threat in high trait anxiety: facilitated engagement, impaired disengagement, and attentional avoidance. *Behaviour Research and Therapy*, 44, 1757–1771. <http://dx.doi.org/10.1016/j.brat.2005.12.011>
- Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, 5(3), 108–125.
- Kreifelts, B., Brück, C., Ritter, J., Ethofer, T., Domin, M., Lotze, M., et al. (2014). They are laughing at me: cerebral mediation of cognitive biases in social anxiety. *PLoS One*, 9(6) <http://dx.doi.org/10.1371/journal.pone.0099815>
- 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, 1234–1245. <http://dx.doi.org/10.3758/s13428-013-0324-3>
- Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: a validated database of Chinese vocal emotional stimuli. *Behavior Research Methods*, 1042–1051. <http://dx.doi.org/10.3758/s13428-012-0203-3>

- Liu, P., Rigoulot, S., & Pell, M. D. (2015). Culture modulates the brain response to human expressions of emotion: electrophysiological evidence. *Neuropsychologia*, 67, 1–13.
- 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, 108–112. <http://dx.doi.org/10.1097/WNR.0b013e32834ea757>
- Martin-Soelch, C., Stöcklin, M., Dammann, G., Opwis, K., & Seifritz, E. (2006). Anxiety trait modulates psychophysiological reactions, but not habituation processes related to affective auditory stimuli. *International Journal of Psychophysiology*, 61, 87–97. <http://dx.doi.org/10.1016/j.ijpsycho.2005.07.009>
- Meyer, M., Elmer, S., Baumann, S., & Jancke, L. (2007). Short-term plasticity in the auditory system: differential neural responses to perception and imagery of speech and music. *Restorative Neurology and Neuroscience*, 25, 411–431.
- Mogg, K., & Bradley, B. P. (1998). A cognitive-motivational analysis of anxiety. *Behavior Research and Therapy*, 36, 809–848. [http://dx.doi.org/10.1016/S0005-7967\(98\)63-1](http://dx.doi.org/10.1016/S0005-7967(98)63-1)
- Mogg, K., Philippot, P., & Bradley, B. P. (2004). Selective attention to angry faces in clinical social phobia. *Journal of Abnormal Psychology*, 113(1), 160–165. <http://dx.doi.org/10.1037/0021-843X.113.1.160>
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–423.
- Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage*, 55(2), 713–723.
- Öhman, A. (1987). The psychophysiology of emotion: an evolutionary-cognitive perspective. *Advances in Psychophysiology*, 2, 79–127.
- Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: implications for nonhuman and human primates. *American Journal of Primatology*, 73, 530–544. <http://dx.doi.org/10.1002/ajp.20913>
- Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, 4(June), 345. <http://dx.doi.org/10.3389/fpsyg.2013.00345>
- Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport*, 19, 209–213.
- Paulmann, S., Ott, D. V. M., & Kotz, S. A. (2011). Emotional speech perception unfolding in time: the role of the basal ganglia. *PLoS One*, 6(3), 13–17. <http://dx.doi.org/10.1371/journal.pone.0017694>
- 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. <http://dx.doi.org/10.3758/CABN.10.2.230>
- Paulmann, S., Titone, D., & Pell, M. D. (2012). How emotional prosody guides your way: evidence from eye movements. *Speech Communication*, 54(1), 92–107. <http://dx.doi.org/10.1016/j.specom.2011.07.004>
- Pell, M. D. (2002). Evaluation of nonverbal emotion in face and voice: some preliminary findings on a new battery of tests. *Brain Cognition*, 48, 499–504.
- Pell, M. D. (2005a). Nonverbal emotion priming: evidence from the facial affect decision task. *Journal of Nonverbal Behavior*, 29(1), 45–73. <http://dx.doi.org/10.1007/s10919-004-0889-8>
- Pell, M. D. (2005b). Prosody-face interactions in emotional processing as revealed by the facial affect decision task. *Journal of Nonverbal Behavior*, 29(4), 193–215. <http://dx.doi.org/10.1007/s10919-005-7720-z>
- Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS One*, 6(11) <http://dx.doi.org/10.1371/journal.pone.0027256>
- Pell, M. D., Paulmann, S., Dara, C., Allasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: a comparison of four languages. *Journal of Phonetics*, 37, 417–435. <http://dx.doi.org/10.1016/j.wocn.2009.07.005>
- Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication*, 50, 519–530. <http://dx.doi.org/10.1016/j.specom.2008.03.006>
- Peschard, V., Maurage, P., & Philippot, P. (2014). Towards a cross-modal perspective of emotional perception in social anxiety: review and future directions. *Frontiers in Human Neuroscience*, 8(May), 322. <http://dx.doi.org/10.3389/fnhum.2014.00322>
- Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., & Gray, J. A. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings, Biological Sciences/The Royal Society*, 265(July), 1809–1817. <http://dx.doi.org/10.1098/rspb.1998.0506>
- Quadflieg, S., Wendt, B., Mohr, A., Miltner, W. H. R., & Straube, T. (2007). Recognition and evaluation of emotional prosody in individuals with generalized social phobia: a pilot study. *Behaviour Research and Therapy*, 45, 3096–3103. <http://dx.doi.org/10.1016/j.brat.2007.08.003>
- Rigoulot, S., Fish, K., & Pell, M. D. (2014). Neural correlates of inferring speaker sincerity from white lies: an event-related potential source localization study. *Brain Research*, 1565, 48–62. <http://dx.doi.org/10.1016/j.brainres.2014.04.022>
- Rigoulot, S., & Pell, M. D. (2014). Emotion in the voice influences the way we scan emotional faces. *Speech Communication*, 65(June), 36–49. <http://dx.doi.org/10.1016/j.specom.2014.05.006>
- Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience*, 290, 175–184. <http://dx.doi.org/10.1016/j.neuroscience.2015.01.033>
- 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(June), 367. <http://dx.doi.org/10.3389/fpsyg.2013.00367>
- Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *Neuroimage*, 28, 848–858. <http://dx.doi.org/10.1016/j.neuroimage.2005.06.023>
- Sander, K., & Scheich, H. (2005). Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. *Journal of Cognitive Neuroscience*, 17, 1519–1531. <http://dx.doi.org/10.1162/089892905774597227>
- Sauter, D. A., & Eimer, M. (2009). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience*, 22, 474–481. <http://dx.doi.org/10.1162/jocn.2009.21215>
- Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, 99(2), 143–165. <http://dx.doi.org/10.1037/0033-2909.99.2.143>
- Scherer, K. R. (1994). Affect bursts. In S. H. M. van Goozen, N. E. van de Poll, & J. A. Sergeant (Eds.), (pp. 161–193). Hillsdale: Erlbaum.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion Inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32, 76–92. <http://dx.doi.org/10.1177/0022022101032001009>
- Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15(2), 123–148. <http://dx.doi.org/10.1007/BF00995674>
- Schirmer, A., Chen, C.-B., Ching, A., Tan, L., & Hong, R. Y. (2013). Vocal emotions influence verbal memory: neural correlates and interindividual differences. *Cognitive, Affective & Behavioral Neuroscience*, 13, 80–93. <http://dx.doi.org/10.3758/s13415-012-0132-8>
- Schirmer, A., & Escoffier, N. (2010). Emotional MMN: anxiety and heart rate correlate with the ERP signature for auditory change detection. *Clinical Neurophysiology*, 121(1), 53–59. <http://dx.doi.org/10.1016/j.clinph.2009.09.029>
- Schirmer, A., Fox, P. M., & Grandjean, D. (2012). On the spatial organization of sound processing in the human temporal lobe: a meta-analysis. *Neuroimage*, 63(1), 137–147. <http://dx.doi.org/10.1016/j.neuroimage.2012.06.025>
- Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, 15, 1135–1148. <http://dx.doi.org/10.1162/089892903322598102>
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Science*, 10(1), 24–30. <http://dx.doi.org/10.1016/j.tics.2005.11.009>
- Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: sex differences revisited. *Cognitive Brain Research*, 24, 442–452. <http://dx.doi.org/10.1016/j.cogbrainres.2005.02.022>
- Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature*. <http://dx.doi.org/10.1038/385254a0>
- Siegmán, A. W., Anderson, R. A., & Berger, T. (1990). The angry voice: its effects on the experience of anger and cardiovascular reactivity. *Psychosomatic Medicine*, 52, 631–643.
- Spielberger, C., Jacobs, G., Russell, S., & Crane, R. (1983). Assessment of anger: the state-trait anger scale. In J. Butcher, & C. Spielberger (Eds.), *Advances in personality assessment (Vol 2)* (pp. 161–189). Hillsdale: Erlbaum.
- Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, 19, 1964–1973. <http://dx.doi.org/10.1162/jocn.2007.91213>
- Szameitat, D. P., Alter, K., Szameitat, A. J., Darwin, C. J., Wildgruber, D., Dietrich, S., & Sterr, A. (2009). Differentiation of emotions in laughter at the behavioral level. *Emotion (Washington, D. C.)*, 9(3), 397–405. <http://dx.doi.org/10.1037/a0015692>
- Szameitat, D. P., Kreifelts, B., Alter, K., Szameitat, A. J., Sterr, A., Grodd, W., & Wildgruber, D. (2010). It is not always tickling: distinct cerebral responses during perception of different laughter types. *Neuroimage*, 53(4), 1264–1271. <http://dx.doi.org/10.1016/j.neuroimage.2010.06.028>
- Trouvain, J. (2014). Laughing, breathing, clicking—the prosody of nonverbal vocalisations. *Speech Prosody*, 2014, 598–602.
- Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, 38, 816–827.
- Wang, X., & Kadia, S. C. (2001). Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *Journal of Neurophysiology*, 86, 2616–2620.
- Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, 11(4), 277–281. <http://dx.doi.org/10.1080/17549500902943043>
- Wildgruber, D., Szameitat, D. P., Ethofer, T., Brück, C., Alter, K., Grodd, W., & Kreifelts, B. (2013). Different types of laughter modulate connectivity within distinct parts of the laughter perception network. *PLoS One*, 8(5), 10–13. <http://dx.doi.org/10.1371/journal.pone.0063441>
- Woldorff, M. G., Gallen, C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Science*, 90, 8722–8726.
- Zvielli, A., Bernstein, A., & Koster, E. H. W. (2014). Dynamics of attentional bias to threat in anxious adults: bias towards and/or away? *PLoS One*, 9(8), e104025. <http://dx.doi.org/10.1371/journal.pone.0104025>