OXFORD

Salience in a social landscape: electrophysiological effects of task-irrelevant and infrequent vocal change

Ana P. Pinheiro, Carla Barros, and João Pedrosa

Neuropsychophysiology Laboratory, School of Psychology, University of Minho, Braga, Portugal

Correspondence should be addressed to Ana P. Pinheiro, Neuropsychophysiology Laboratory, CIPsi, School of Psychology, University of Minho, Campus de Gualtar, 4710-957 Braga, Portugal. E-mail: ana.pinheiro@psi.uminho.pt.

Abstract

In a dynamically changing social environment, humans have to face the challenge of prioritizing stimuli that compete for attention. In the context of social communication, the voice is the most important sound category. However, the existing studies do not directly address whether and how the salience of an unexpected vocal change in an auditory sequence influences the orientation of attention. In this study, frequent tones were interspersed with task-relevant infrequent tones and task-irrelevant infrequent vocal sounds (neutral, happy and angry vocalizations). Eighteen healthy college students were asked to count infrequent tones. A combined event-related potential (ERP) and EEG time–frequency approach was used, with the focus on the P3 component and on the early auditory evoked gamma band response, respectively. A spatial-temporal principal component analysis was used to disentangle potentially overlapping ERP components. Although no condition differences were observed in the 210–310 ms window, larger positive responses were observed for emotional than neutral vocalizations in the 310–410 ms window. Furthermore, the phase synchronization of the early auditory evoked gamma oscillation was enhanced for happy vocalizations. These findings support the idea that the brain prioritizes the processing of emotional stimuli, by devoting more attentional resources to salient social signals even when they are not task-relevant.

Key words: salience; attention; voice; event-related potentials; neural oscillations

Introduction

The ability to extract regularities and to rapidly detect change in a dynamically changing environment is critical for survival. Surprise and uncertainty represent a cost in terms of energy and processing resources (Kveraga *et al.*, 2007). In the context of social communication, this capacity involves predicting what is going to happen next in the social scene based on memory-driven regularity representations, detecting frequently unexpected changes that might signal an alteration in the emotional state of the speaker, as well as quickly reacting to those unexpected events (e.g. Jessen and Kotz, 2013).

The voice is plausibly the most important sound category in a social environment (e.g. Belin *et al.,* 2004). By representing a dynamically changing auditory stimulus, the listener needs to rapidly integrate multiple cues, such as pitch, intensity and duration (e.g. Schirmer and Kotz, 2006). Vocal emotional cues represent biologically relevant signals that require rapid detection, evaluation and response (e.g. Schirmer and Kotz, 2006). For that

reason, it is not surprising that human beings are tuned to quickly discriminate between emotionally salient and neutral stimuli. Picture a speaker who is describing a life event and suddenly starts shouting out loud angrily. The listener needs to quickly detect the change in voice intonation and to accurately identify if that same change signals any sort of danger in the environment. Alternatively, if the speaker starts jumping and using a vibrant and enthusiastic tone of voice, this change may represent a positive event and it is well known that positive social events play a critical role in promoting social bonding (e.g. Johnstone *et al.*, 2006). These examples demonstrate that attending to changes in the acoustic environment is necessary to evaluate and modify behavior based on perceived threats or opportunities.

Event-related potential (ERP) studies suggest that the differentiation between neutral and emotional vocal cues occurs already at early processing stages (e.g. Paulmann and Kotz, 2008; Pinheiro *et al.*, 2013, 2014). Furthermore, there is a

substantial body of knowledge indicating that the recognition of emotion in voices is intricately associated with attentional (e.g. Sander *et al.*, 2005) and prediction mechanisms (e.g. Jessen and Kotz, 2013). Although attention may lead to increased precision in information processing, prediction biases attention in favor of anticipated stimuli (e.g. Chennu *et al.*, 2013). Within a limited capacity system, emotionally salient information conveyed through the voice modulates the allocation of attentional resources, receiving preferential and prioritized processing in the competition for access to selective attention (e.g. Fenske and Raymond, 2006). Furthermore, emotion may lead to stronger predictive signals than neutral stimuli (e.g. Jessen and Kotz, 2013).

An experimentally controlled way to probe the brain mechanisms associated with vocal change detection is to use an oddball paradigm. Typically, in the 'active' oddball task, participants are instructed to pay attention to a stream of stimuli and to detect rare target (deviant) events in a series of frequent (standard) stimuli. In ERP studies, the P300 component is elicited by deviant stimuli and can be considered an index of the allocation of processing resources (e.g. Wickens *et al.*, 1983), particularly attention (e.g. Polich, 1989), or a manifestation of a mechanism that updates a model of the environment or context in working memory (e.g. Donchin, 1981; Donchin and Coles, 1988).

Rather than representing a unitary phenomenon, the P300 is composed of dissociable and independent components that reflect the involvement of attentional and memory mechanisms (e.g. Goldstein *et al.*, 2002). This observation highlights the importance of using spatiotemporal decomposition methods to disentangle potentially overlapping ERP components (e.g. Spencer *et al.*, 1999, 2001). The classical P300 (or P3b) has a parietal topography and is elicited by rare task-relevant stimuli (e.g. Sutton *et al.*, 1965). The P3a (or Novelty P3) is a positive component with a frontocentral distribution occurring in the same latency range as the P300, and it is elicited by task-irrelevant rare or novel stimuli inserted randomly in a sequence of repetitive (standard) and target (deviant) events (e.g. Courchesne *et al.*, 1975; Squires *et al.*, 1975). This component is a neurophysiological index of attentional orienting towards a stimulus, with its amplitude increasing as a function of stimulus salience (e.g. Nittono, 2006). The elicitation of a P3a component tells us that an unexpected change in the sensory environment engages attention involuntarily and is sufficiently salient to enter into awareness. Therefore, this component is especially relevant for the understanding of how the salience of a change in vocal intonation may differently influence attention mechanisms. Because the P3a is elicited by a task-irrelevant and unexpected change in a sound sequence, it more closely resembles a real-world social communication context.

Very few studies have investigated how the salience of vocal sounds modulates the allocation of attentional resources, using an oddball task (Thierry and Roberts, 2007; Campanella *et al.*, 2010). Two major observations have arisen from emotional studies in the visual modality: first, emotionally salient stimuli engage more attentional resources (e.g. Schupp *et al.*, 2003; Olofsson *et al.*, 2008; Kissler *et al.*, 2009); second, aversive stimuli seem to engage more attention than positive stimuli (negativity bias), even though this is not a consistent finding (e.g. Ito *et al.*, 1998; Carretié *et al.*, 2001). However, fewer studies probed these effects in the auditory modality. Thierry and Roberts (2007) found that the P3a amplitude was significantly more positive for unpleasant relative to neutral sounds. Campanella *et al.* (2010) reported earlier P300 latencies for happy than for sad prosodic speech deviants (word 'paper' spoken with happy or sad prosody). These

studies keep with the notion that attention is oriented faster to an emotional stimulus, and additionally suggest that the pleasantness or unpleasantness of a stimulus (i.e. its valence) may differently engage attentional resources.

In an attempt to probe the brain underpinnings of vocal emotional processing, the information provided by ERP components such as the P300 and the P3a may be complemented with the analysis of neural oscillations in the time–frequency domain. The importance of this type of analysis has come into focus in recent years (e.g. Roach and Mathalon, 2008). In particular, probing the phase of high-frequency oscillatory activity may provide critical insights into the brain mechanisms underlying emotionally salient vocal change detection. Many recent studies suggest that phase synchronization of neural activity plays a critical role in sensory prediction and change detection (e.g. Fell *et al.*, 2003; Kveraga *et al.*, 2007). These studies have highlighted the special role played by gamma-band oscillations in these processes. Gamma phase synchronization was found to be increased for novel relative to standard sounds (Garcia-Garcia *et al.*, 2010) and to play a critical role in selective attention (Fell *et al.*, 2003). Specific prediction and attention effects have been observed in the early auditory evoked gamma-band response that occurs at 30–60 Hz, within 150 ms after stimulus onset (e.g. Pantev *et al.*, 1991, 1995). This response is characterized by precise phase-locking to stimulus onset and is detected by averaging EEG single trials (e.g. Pantev, 1995; Schadow *et al.*, 2007). An increase in the phase synchronization of the early gamma-band response has been associated with increased attention (Fries *et al.*, 2001), re-allocation of attentional resources (Garcia-Garcia *et al.*, 2010), enhanced sensory information processing (e.g. Herrmann *et al.*, 2004), and the match between bottom-up signals and top-down expectations (e.g. Debener *et al.*, 2003; Roye *et al.*, 2010). Nonetheless, few studies to date have examined the effects of salience on the synchronization of gamma oscillations and those that have present a mixed picture. For example, Garcia-Garcia *et al.* (2010) reported increased gamma phase synchronization for novel sounds occurring in a negative visual context relative to a neutral one, and Domínguez-Borràs *et al.* (2012) found increased gamma phase synchronization for novel sounds presented in a positive visual context relative to a neutral one. These findings suggest that gamma oscillations may additionally play a role in the association of the perceptual properties of stimuli with their emotional significance (Oya *et al.*, 2002).

The current study and hypotheses

Although the studies reviewed before support the strong modulatory role of stimulus salience on attention, these studies do not directly address whether and how the salience of an unexpected change in an auditory sequence has an impact upon involuntary orienting of attention indexed by the P3a and on the synchronization of early evoked gamma oscillations. The temporal resolution of the EEG methodology is especially suitable to tackle these questions as it may provide critical hints about the speed at which vocal emotional information is decoded and at the processing stages involved (e.g. Paulmann and Kotz, 2008).

In this study, we presented frequent tones (standard) interspersed with task-relevant infrequent tones (target) and task-irrelevant infrequent vocal sounds. Vocal sounds were short neutral, happy and angry non-verbal vocalizations. Emotional vocalizations were matched in perceived arousal. We selected an exemplar of negative (angry) and positive (happy) social signals. Both happy and angry vocal stimuli have similar acoustic

profiles (e.g. high intensity and variable F0) and are recognized above chance (Schröder, 2003; Hawk *et al.*, 2009; Laukka *et al.*, 2013). Although angry vocalizations represent a negative social signal that may convey aversive meaning and have survival value (avoiding danger), happy vocalizations represent a positive social signal with social bonding value (e.g. Compton, 2003). Emotional non-verbal vocalizations represent a more primitive expression of emotion (e.g. Scherer, 1995; Scherer *et al.,* 2001; Belin *et al.,* 2004) and can be viewed as the auditory equivalent of facial emotional expressions (e.g. Belin *et al.,* 2004). Of note, studies probing accuracy differences in the recognition of emotion expressed through different types of auditory stimuli (e.g. prosodic speech, pseudospeech, non-verbal vocalizations) demonstrated that emotions are decoded more accurately through non-verbal vocalizations than through speech-embedded prosody (e.g. Hawk *et al.,* 2009). Therefore, the use of nonverbal emotional vocalizations *vs* speech prosody in experimental research may optimize the recognition of emotional content and avoid confounds associated with concurrent phonological and lexical-semantic information (e.g. Warren *et al.*, 2006; Belin *et al.*, 2011).

We used a modified version of the novelty oddball paradigm. Instead of unique vocal stimuli, low probability vocalizations differing in valence were presented in different experimental blocks to ensure a sufficient number of EEG segments that would allow the comparison of different valence types to probe the role of stimulus salience on attentional orienting. In that sense, vocalizations were familiar (each was presented more than once) but unexpected task-irrelevant stimuli given the context in which they occurred. Because previous studies suggest that after a pre-attentive evaluation of the emotional significance of sounds, those that are evaluated as emotionally significant receive prioritized processing and access to selective attention (e.g. Robinson, 1998; Compton, 2003; Vuilleumier, 2005), we hypothesized that emotional vocalizations would lead to a greater engagement of involuntary attention than neutral vocalizations, reflected in increased P3a amplitude.[1]

Furthermore, we explored whether emotional vocalizations elicit greater synchrony of neural activity across trials. We focused our analysis on phase-locking factor (PLF) values of the early auditory evoked gamma-band response. PLF measures the variance of the EEG phase across single trials and, as such, is a measure of temporal stability of a specific evoked response (e.g. Roach and Mathalon, 2008). Considering the survival value of angry vocalizations and the affiliative value of happy vocalizations, we hypothesized that emotional vocalizations would be associated with enhanced phase synchronization of the early auditory-evoked gamma oscillation, reflecting increased attention towards unexpected salient stimuli relative to unexpected neutral stimuli (e.g. Fries *et al.,* 2001; Garcia-Garcia *et al.,* 2010).

Method

Participants

Participants were 18 healthy college students (9 females; mean age = 23.67, SD = 3.53 years). The inclusion criteria were: European Portuguese as first language; right handedness

Table 1. Affective normative ratings of the experimental stimuli

Type of vocalization	Affective dimension		
	Valence	Arousal	Dominance
	Mean (SD)	Mean (SD)	Mean (SD)
Neutral	4.83 (0.73)	3.21 (2.00)	5.09 (2.56)
Happy	7.54 (1.14)	5.09 (2.19)	6.28 (2.08)
Angry	3.22 (1.54)	5.63 (2.10)	4.67 (2.28)

Note: Mean ± SD are given for each variable.

(Oldfield, 1971); no history of neurological illness; no history or current psychiatric disorder, as assessed by the Brief Symptom Inventory (BSI—Derogatis, 1993; Portuguese version—Canavarro, 2007); no present medication for medical disorders that could have deleterious effects on EEG morphology, as well as neurological, and/or cognitive functioning consequences; no hearing, vision or upper body impairment—relevant for neuropsychological function.

Before the EEG experiment, a brief neurocognitive assessment aimed to assess verbal intelligence quotient (Wechsler, 2008; M = 119.4, SD = 12.21), and a brief clinical assessment ruled out the presence of psychopathological symptoms (BSI—Canavarro, 1999; M = 1.26, SD = 0.26). None of the participants had total scores that indicated the presence of clinical symptoms.

Participants were given course credit for their participation in the study. All participants provided written informed consent for the experimental protocol approved by a local Ethical Committee.

Stimuli

Stimuli were two sinusoidal tones of different frequencies (tone 1–350 Hz, 10 ms rise and fall times; tone 2–500 Hz, 10 ms rise and fall times) but same duration (336 ms), and three exemplars of a female vocalization (neutral, happy and angry) selected from the 'Montréal Affective Voices' battery (Belin *et al.*, 2008). The duration of the vocalizations was shortened so that they lasted only 500 ms, but their emotional content was preserved.[2] Vocalizations were normalized in mean intensity (70 dB) using a Praat script (Boersma & Weenink, 2013). The mean fundamental frequency (F0) was 212.58 Hz for the neutral vocalization, 350.61 Hz for the happy vocalization, and 414.36 Hz for the angry vocalization.

In order to rule out any language or culture-specific factors (Scherer *et al.*, 2001), the sounds were first validated with a sample of European Portuguese participants (n = 155; 90 females; mean age = 21.00, SD = 3.62 years) who did not participate in the ERP study. They rated the sounds in the affective dimensions of valence (indexing the unpleasantness or pleasantness of a stimulus, ranging from unpleasant to pleasant), arousal (reflecting the degree of activation elicited by the stimulus, and ranging from calm to arousing), and dominance (indexing the degree of perceived control that a stimulus evokes, ranging from out of control to in control), using the Self-Assessment Manikin 9-point scale (SAM, Bradley and Lang, 1994). Three vocalizations were selected based on the normative ratings. The affective properties of the three vocalizations are presented in Table 1.

---

1 Throughout the text, we will use the term P3a instead of Novelty P3 since the low-probability vocalizations used in our study did not represent truly novel or unique stimuli, as they were repeated in the experiment.

2 In order to reduce stimulus duration, the most steady-state portion of the vocal sounds was selected and individual waves were manually cut at zero crossings (where the wave crosses zero amplitude) to avoid clicks. Additionally, pauses were reduced.

Procedure

The experiment took place in a sound-attenuated and electrically shielded room. Participants performed an oddball task in which they counted infrequent target sinusoidal tones. The experiment included two oddball sequences: in sequence 1, 800 exemplars of tone 1 (350 Hz; standard—$P = 0.089$), 45 exemplars of tone 2 (500 Hz; target—$P = 0.005$), and 18 exemplars of each vocalization (neutral, happy and angry—$P = 0.002$ each) were pseudorandomized. In sequence 2, 800 exemplars of tone 2 (500 Hz), 45 exemplars of tone 1 (350 Hz) and 18 exemplars of each vocalization (neutral, happy and angry) were presented. The experiment was divided into eight blocks, i.e. two training blocks and six experimental blocks (see Figure 1). The order of these blocks was counterbalanced across participants. No vocalizations were presented during the training blocks. The presentation of sounds in six experimental blocks accomplished three main goals: (i) to minimize working memory demands associated with the task of counting the target sounds (participants were asked to report the number of targets counted at the end of each block); (ii) to avoid an excessive number of vocalizations presented in a single block; (iii) to minimize participants' fatigue and distraction.

Simultaneous to auditory stimuli presentation, an unblinking fixation cross was presented on a monitor 100 cm from the participant. All auditory stimuli were presented in a pseudorandomized order, with the rule that each target was followed by at least two standards and each vocalization was followed by at least 10 non-vocal sounds, with no vocal stimuli appearing in the first 15 sounds. Auditory stimuli were separated by an inter-stimulus interval (ISI) of 1000 ms (see Figure 1), and presented via Sennheiser CX 300-II headphones at 70 dB sound pressure level. The presentation and timing of the stimuli were controlled through Presentation software (version 16.3; Neurobehavioral Systems, Inc., Albany, NY). Participants were instructed to silently count the number of infrequent (target) tones.

At the end of the EEG session, participants were instructed to rate the valence, arousal and dominance of the three vocalizations using the SAM (Bradley and Lang, 1994).

EEG data acquisition and analysis

EEG data acquisition. EEG data were recorded using a 64-channel Active Two Biosemi system (Biosemi, Amsterdam, The Netherlands), in a continuous mode at a digitization rate of 512 Hz, with a bandpass of 0.01–100 Hz, and stored on disk for later analysis. Eye blinks and movements were monitored through electrodes placed on both temples (horizontal
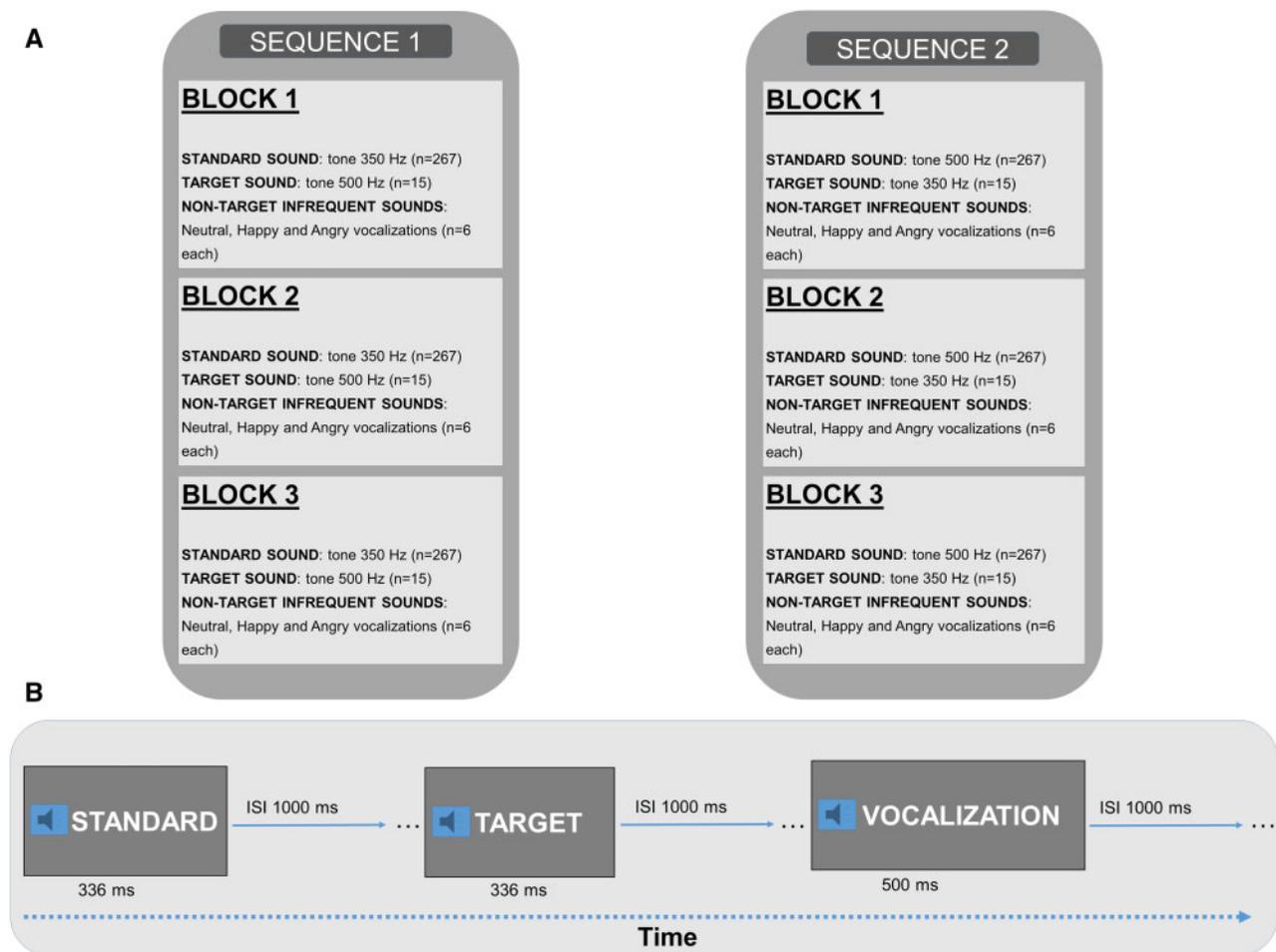
Fig. 1. Schematic illustration of the experimental procedure. **(A)** The experiment included two oddball sound sequences. The order of these blocks was counterbalanced across participants. Before each sequence, a training block took place. This allowed participants to identify which sound would be the standard and which sound would be the target and to which attention should be paid for the following three blocks. No vocalizations were presented during the training blocks. **(B)** Standard and target sounds had the same duration. Non-target infrequent vocalizations had 500 ms duration. The auditory stimuli were separated by a 1000 ms ISI (1000 ms).

electrooculagram) and another one below the left eye (vertical electrooculogram [EOG]).

The EEGLAB 13.1.1b software (Delorme and Makeig, 2004) was used for data analysis. Data were referenced offline to the average of the left and right mastoids and high-pass filtered with a 0.1 Hz filter. Individual ERP epochs were created for each stimulus type, with −200 ms pre-stimulus baseline and 1000 ms post-stimulus epoch. The EEG was baseline corrected using the −200 to 0 ms pre-stimulus interval. The vertical EOG was derived by subtracting the activity measured at an electrode positioned below the left eye from an electrode positioned above it. The horizontal EOG was derived by subtracting the activity measured between electrodes placed at the outer canthi of the eyes. The EEG channels were then corrected for vertical and horizontal eye movements using the method of Gratton *et al.* (1983). Segments were also semiautomatically screened for eye movements, muscle artifacts, electrode drifting and amplifier blocking. EEG epochs exceeding $\pm 100\,\mu v$ were not included in individual ERP averages. After artifact rejection, at least 85% of the segments per condition per subject entered the analyses. The number of trials in each vocal condition did not differ ($P > 0.50$). Individual ERP averages were derived for each stimulus condition for all the participants.

EEG data analysis

ERP data. Spatiotemporal analyses. In order to most accurately capture the spatial topography of evoked responses over time and to functionally separate distinct events that occur simultaneously in time, a spatio-temporal Principal Component Analysis (PCA) method was used. The procedures used here are described in detail in Spencer *et al.* (1999, 2001). The data set used in the PCA consisted of the ERP averages at each electrode site in all stimulus conditions for each participant, using the 0–700 ms portion of the epoch (359 time-points). The data matrix input to the spatial PCA consisted of 61 (oculars excluded) electrode sites × 45234 observations (359 time-points × 7 stimulus conditions[3] × 18 participants). Based on the scree test (Cattell, 1966), 3 spatial factors (SFs) explained most of the variance in the data set and were extracted for rotation with the Varimax procedure (see Figure 3). In this first step, as described by Spencer *et al.* (1999, 2001), SF loadings are computed (variables = electrode sites; observations = time-points, experimental conditions and participants). The covariance between electrode sites is analyzed across the time-points of the averages for each stimulus condition and participant. As a result, each SF provides information about the correlations between the EEG data recorded at each electrode site (original variable) and the factor in the new data space (SFs) with reduced dimensionality. 'Virtual electrodes' represent the SFs that inform about the spatial variance in the data, representing scalp distributions and not typical ERP components. As such, they are not comparable to voltage maps. SF scores can be plotted as time series for each condition and participant and, therefore, considered 'virtual ERPs' (see Spencer *et al.*, 1999, 2001). Topographical maps of the voltage data and SF loadings were plotted using spherical line interpolation.

A temporal PCA was then run on the SF scores, using Varimax rotation (variables = SF scores or 'virtual ERPs' associated with the time-points; observations = 'virtual electrodes',

---

3 The conditions were: Standard 350 Hz, Target 500 Hz, Standard 500 Hz, Target 350 Hz, Neutral Vocalization, Happy Vocalization and Angry Vocalization.

conditions and participants). In this second step, the covariance between time-points for all the virtual electrodes, participants and experimental conditions was analyzed (see Spencer *et al.*, 1999, 2001). Following Spencer *et al.* (1999, 2001), the resulting temporal factor (TF) scores for each SF were used to measure the activity in the ERP component with the temporal morphology and scalp distribution of interest. Factor scores represent a unitless dimension (see Spencer *et al.*, 1999, 2001 for further details).

Habituation effects of non-target vocalizations. We explored whether adaptation (measured as ERP amplitude attenuation) differed as a function of stimulus valence. For that purpose, mean amplitude at the electrode Cz was exported in the 210–310 ms and 310–410 ms latency windows for each trial in the neutral, happy and angry conditions.

Time–frequency EEG data. PLF. PLF values were extracted by wavelet-based spectral decomposition of single-trial EEG epochs time-locked to non-target vocalizations. It was calculated as 1 minus the phase variance, for the frequencies between 3 and 60 Hz. Time–frequency decomposition was performed using the Morlet wavelet transform in EEGLAB software, which was applied in 0.25 Hz steps from 4 to 60 Hz at each time-point to yield time–frequency maps. Before that, the retained artifact-free epochs were re-referenced to the average reference (Dien, 1998). The wavelet frequency/duration ratio f_0/σ_f was $\frac{12\pi}{5\sqrt{2}} \cong 5.33$ and its multiplication factor was $\frac{15}{2\sqrt{2}} \cong 5.30$. Baseline activity was subtracted from each PLF map. Baseline periods were calculated from −200 to 0 ms relative to stimulus onset. The analysis of the TF maps revealed effects in the gamma-band range in an early latency window that corresponds to the early auditory-evoked gamma-band response. PLF values for task-irrelevant infrequent vocalizations were extracted in the 30–50 Hz frequency range in the 75–120 ms latency window, and measured at frontocentral electrode sites (FCz/1/2, and Cz/1/2).

The SPSS 22.00 (SPSS, Corp., USA) software package was used for statistical analyses. Analyses were corrected for non-sphericity using the Greenhouse–Geisser method (the original *df* is reported). All significance levels are two-tailed with the preset significance alpha level of $P < 0.05$. Main effects were followed up with pairwise comparisons between conditions using the Bonferroni adjustment for multiple comparisons. Only significant results are presented in the following section ($P < 0.05$).

Results

ERP data

Grand average waveforms for the three types of vocalizations (see Figure 2) revealed two major effects in the P300 latency window (210–410 ms). The spatiotemporal PCA provided support for the independency of these effects and revealed their spatial distribution.

Spatial-temporal analysis. The spatial PCA revealed three SFs (see Figure 3): SF1 has a frontocentral distribution and accounted for 41.7% of variance; SF2 has a parietal/occipital distribution and accounted for 27.88% of variance; SF3 has a frontal distribution and accounted for 21.93% of variance.

The temporal PCA identified 8 TFs (see Figure 4). We focused our analysis on the TFs that were observed in the latency
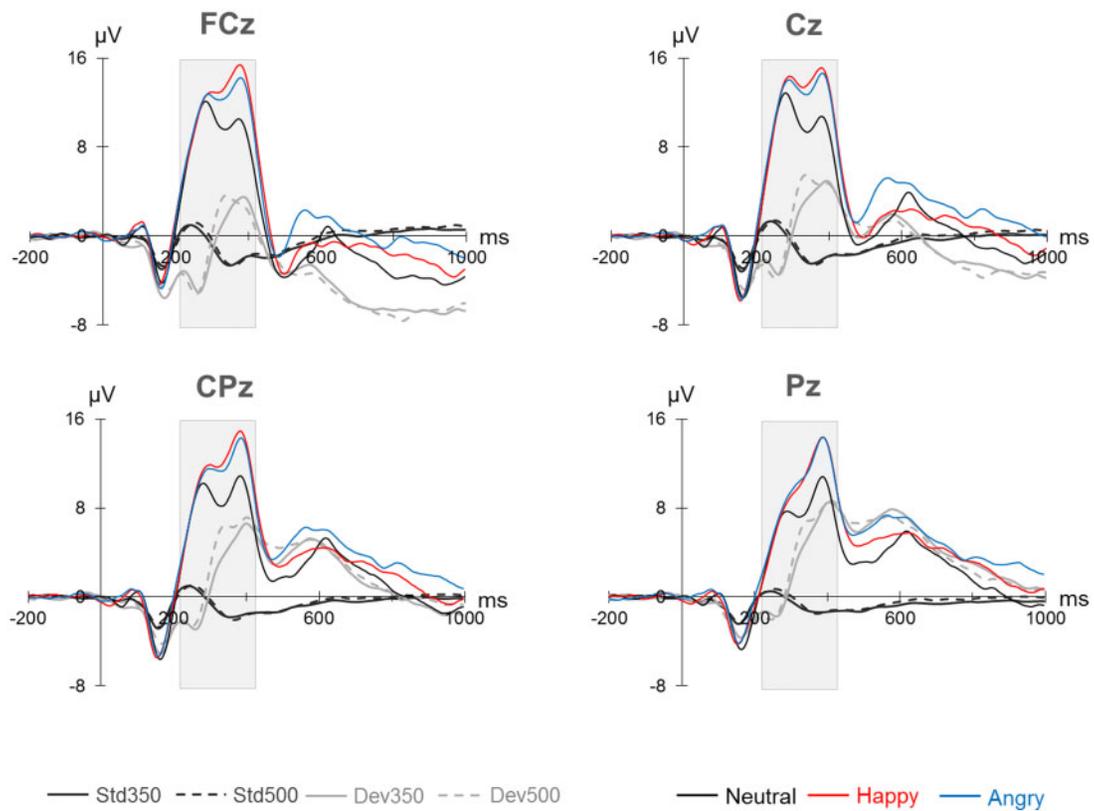
Fig. 2. Grand average ERP waveforms for neutral, happy and angry vocalizations, at midline electrodes (FCz, Cz, CPz, Pz). *Note:* Std, standard; Dev, deviant (target). Grand-averaged waveforms were filtered with a 8-Hz low-pass filter for graphical display only.
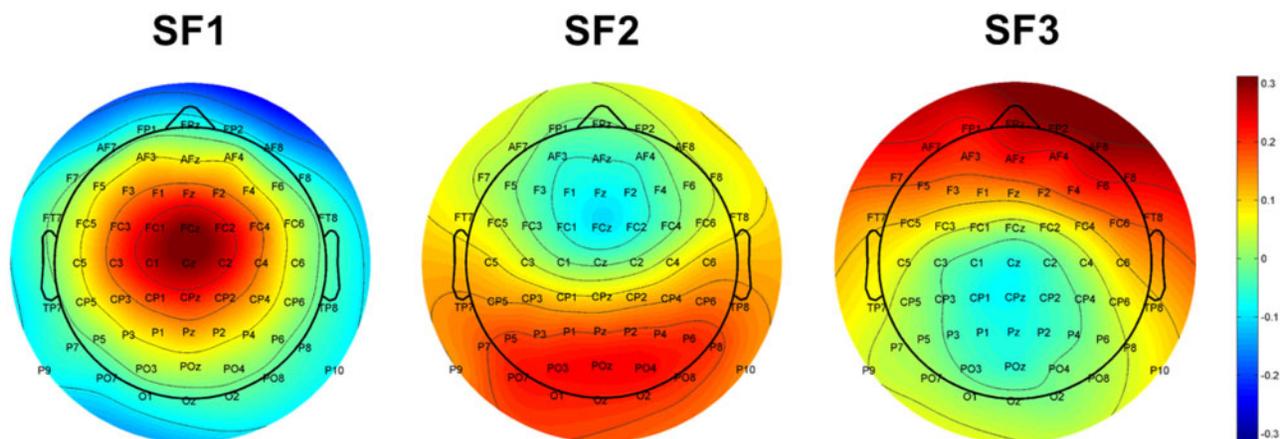

Fig. 3. Topographic maps of the factor loadings for the SFs ('virtual electrodes') accounting for most of the variance after rotation.

window around 300 ms: TF3 and TF4. Corroborating previous studies (e.g. Escera *et al.*, 1998), the P3a to task-irrelevant unexpected vocalizations had a double peak over frontocentral scalp locations, which suggested two different subcomponents. TF4 accounted for 14.03% of the variance and occurred in the latency between 210 and 310 ms. TF3 accounted for 20.53% of the variance, with an latency between 310 and 410 ms.

The factor scores represented by TF4 and TF3 were separately analyzed through Repeated Measures Analyses of Variance (ANOVA) with condition (neutral, happy and angry) and SF (1, 2 and 3) as within-subject factors. No significant

condition effect was found for the early P3a peak or TF4 ($P > 0.05$). The factor scores were significantly enhanced at frontocentral sites represented by SF1 (main effect of SF—$F(2, 34) = 18.108$, $P < 0.001$, $\eta_p^2 = .516$; SF1 *vs* SF2—$P < 0.001$; SF1 *vs* SF3—$P < 0.001$) (see Figure 5).

Although no condition effects were found in the earlier peak of the P3a, a significant effect of condition was observed in the later part of the P3a window or TF3 [$F(2, 34) = 11.183$, $P < 0.001$, $\eta_p^2 = 0.397$]: larger positive responses were observed for both happy and angry vocalizations than for neutral vocal stimuli (neutral *vs* happy—$P = 0.001$; neutral *vs* angry—$P < 0.003$)
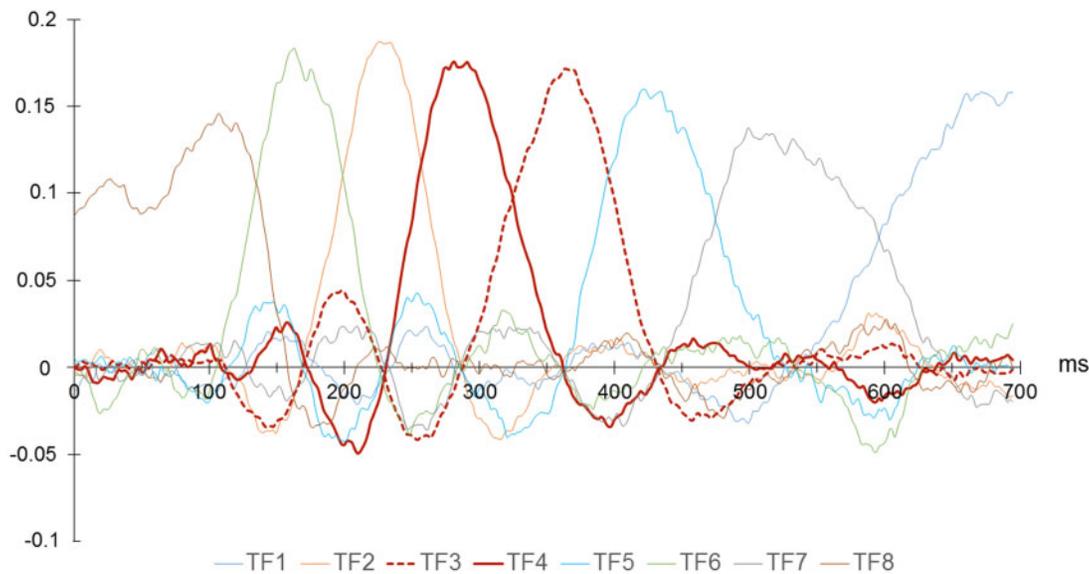
Fig. 4. TF loadings ('virtual epochs'). *Note.* The percentage of variance accounted for by each factor is: TF1-13.1677%; TF2-13.1958%; TF3-20.5298%; TF4-14.0251%; TF5-10.6438%; TF6-4.7668%; TF7-15.5136%; TF8-3.1795%.
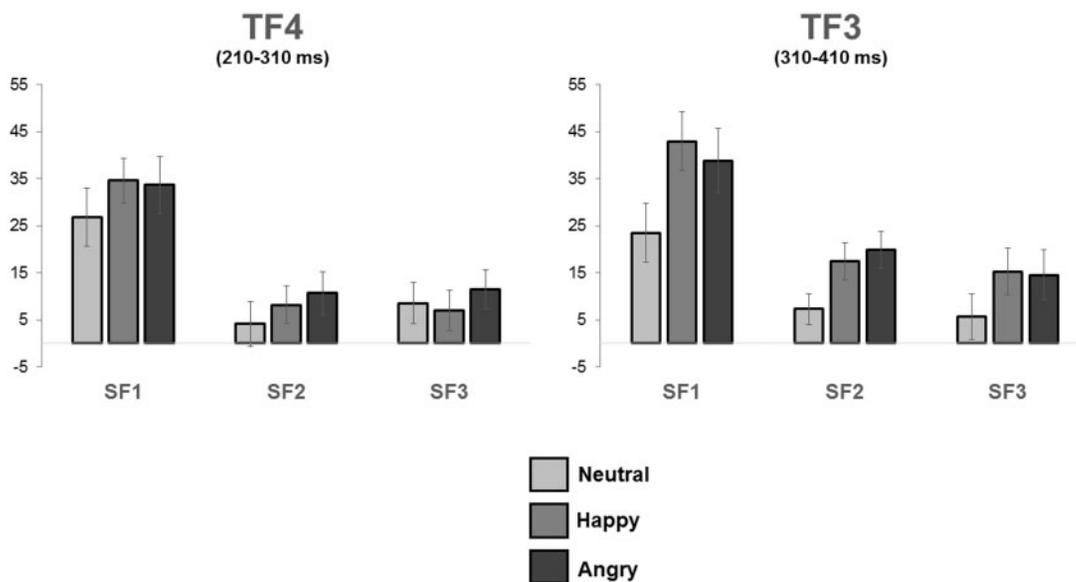

Fig. 5. TF3 and TF4 scores associated with SF1, SF2 and SF3, for each vocalization type. *Note.* The value of the factor scores is a unitless dimension.

(see Figures 2 and 5). No differences between angry and happy vocalizations were observed ($P > 0.05$). The factor scores were significantly enhanced at frontocentral sites, represented by SF1 (main effect of SF—$F(2, 34) = 12.622$, $P < 0.001$, $\eta_p^2 = 0.426$; SF1 *vs* SF2—$P = 0.009$; SF1 *vs* SF3— $P < 0.001$) (see Figures 2 and 5).

Habituation effects of vocalizations. In order to explore whether the repetition of the non-target vocalization led to differences in adaptation (i.e. amplitude reduction) as a result of stimulus valence, we ran a repeated-measures ANOVA on single-trial amplitude data for each vocal stimulus in the two P3 latency windows (210–310 ms; 310–410 ms). The single trial data were grouped as beginning (trials 1–10), middle (trials 11–20) and end (trials 21–30) of the sequence of vocalizations. The within-subject factors of time (beginning, middle and end of the sequence) and condition (neutral, happy and angry) were included

in the statistical model. No significant interaction between time and condition was observed ($P > 0.05$).

In order to better illustrate amplitude variation across trials for each condition, the EEGLab function 'erpimage()' was used (see Figure 6). This function allows the trial-by-trial visualization of ERP amplitude for a specific channel. First, the group average of ERP activity for each trial and condition (neutral, happy and angry vocalizations) was computed. Then the 'erpimage()' function was run on those data. As a result, the group average of trial-by-trial ERP activity for each condition at the Cz electrode was illustrated in a 2D representation, in which time (from −200 to 1000 ms) was plotted against trials (from first to last). Cz was chosen given the frontocentral/central distribution of the P3 effects. The trials were imaged (bottom-to-top) based on their order of occurrence during the experiment (see Figure 6—Panel A). The data matrix input included 33156 observations (614 time-points × 3 stimulus conditions × 18 participants).
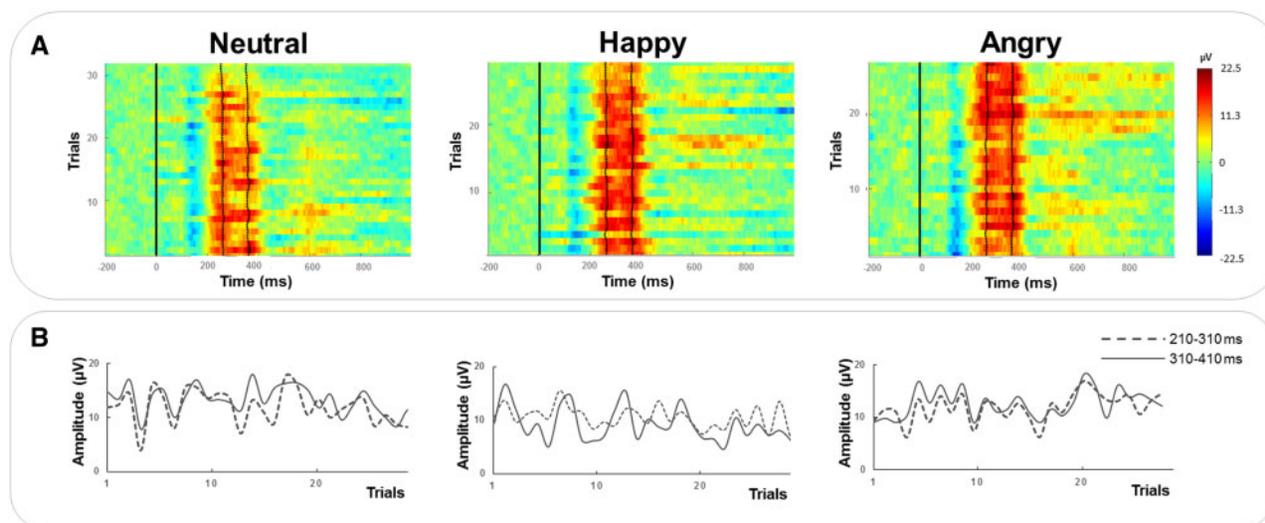
Fig. 6. Trial-by-trial ERP amplitude for neutral, happy and angry vocalizations. **(A)** 2D representation of the group average of trial-by-trial ERP activity for each condition at the Cz electrode. Time (−200 to 1000 ms) was plotted against trials (from first to last), based on their order of occurrence in the experiment. The figure shows darker red color (i.e. increased positive amplitude) in the P300 latency window for happy and angry vocalizations compared with the neutral vocalization, consistently from the first until the last experimental trial. **(B)** Line graphs showing trial-by-trial mean amplitude in two time windows: 210–310 ms; 310–410 ms. Voltage scores represent the group average of mean amplitude for each experimental trial in the specified latency intervals. The figure shows trial-by-trial variability in ERP amplitude for the three conditions.

Using the same EEGlab function, two line graphs were plotted showing the trial-by-trial group average of mean amplitude in the latency intervals of interest: 210–310 ms and 310–410 ms, respectively (see Figure 6—Panel B). Based on the inspection of Figure 6, the ERP response across trials for each condition is characterized by variability, instead of trial-by-trial amplitude reduction that would suggest neural adaptation.

Time–frequency analysis: PLF

The repeated-measures ANOVA included condition (neutral, happy and angry) and electrodes (FCz/1/2, Cz/1/2) as within-subject factors.

The PLF of the early auditory evoked gamma response was enhanced for happy relative to both neutral ($P = 0.027$) and angry ($P = 0.020$) task-irrelevant vocalizations, as demonstrated by a significant effect of condition [$F(2, 34) = 5.920$, $P = 0.006$, $\eta_p^2 = 0.258$) (see Figure 7).

Correlational analyses

Two-tailed Pearson correlation analyses were conducted in an exploratory analysis of the relationship between significant ERP/PLF effects and participants' affective ratings of the vocal stimuli (presented in Table 2). No significant associations were found ($P > 0.05$).

Discussion

Emotion has long been known to modulate attention, leading to an enhancement of processing resources. However, it remains to clarify to what extent and how salient vocal stimuli, such as happy or angry vocalizations, presented as task-unrelated and unexpected stimuli, engage automatic *vs* voluntary aspects of attention. In this study, we probed whether non-target infrequent vocalizations—representing complex deviations from stimulus predictability imposed by task-irrelevant high-probability tones—were differentially processed on the basis of

their emotional salience and how this had an impact on the orienting response around 300 ms after stimulus onset. Because emotional significance evaluation guides the allocation of attention, it stands to reason that both infrequent and task-irrelevant angry and happy vocalizations would automatically elicit stronger attention orienting as they represent stimuli that might be of the highest relevance to the organism's goals, when compared with neutral vocal stimuli. We observed two dissociable components in the P3 window, corroborating previous studies with infrequent non-target sounds (e.g. Escera *et al.*, 1998, 2000) and the hypothesis that these components reflect different cognitive processes (e.g. Friedman *et al.*, 2001). No condition-dependent amplitude differences were observed in the 210–310 ms latency window. These differences emerged only in a later latency window (310–410 ms) and consisted of enhanced amplitude for emotional relative to neutral vocalizations, with no difference between pleasant and unpleasant vocal stimuli.

An important function of the orienting response is to prepare the organism to deal with an unexpected and potentially salient stimulus. Previous studies suggest that the early phase of the P3a reflects processes related to automatic attention orienting that takes place after the unexpected stimulus has been detected (e.g. Escera *et al.*, 2000). The later phase of the P3a reflects the effective orienting of attention towards the salient stimuli, involving categorization processes (Friedman *et al.*, 2001) that are critical for preparing the organism to goal-directed action selection (Yamaguchi and Knight, 1991; Escera *et al.*, 1998, 2001; Debener *et al.*, 2005; Escera and Corral, 2007). The absence of condition differences in the 210–310 ms window suggests that the mismatch signal elicited by the three types of non-target unexpected stimuli exceeded the threshold for unattended deviance detection in a similar way. Therefore, the early part of the P3a responded to stimulus probability or surprise (a factor that was constant for the three vocal conditions in our study) but not to stimulus salience: the three types of vocalizations automatically triggered a similar attention switch, despite differences in the physical properties and emotional salience of the vocal stimuli.
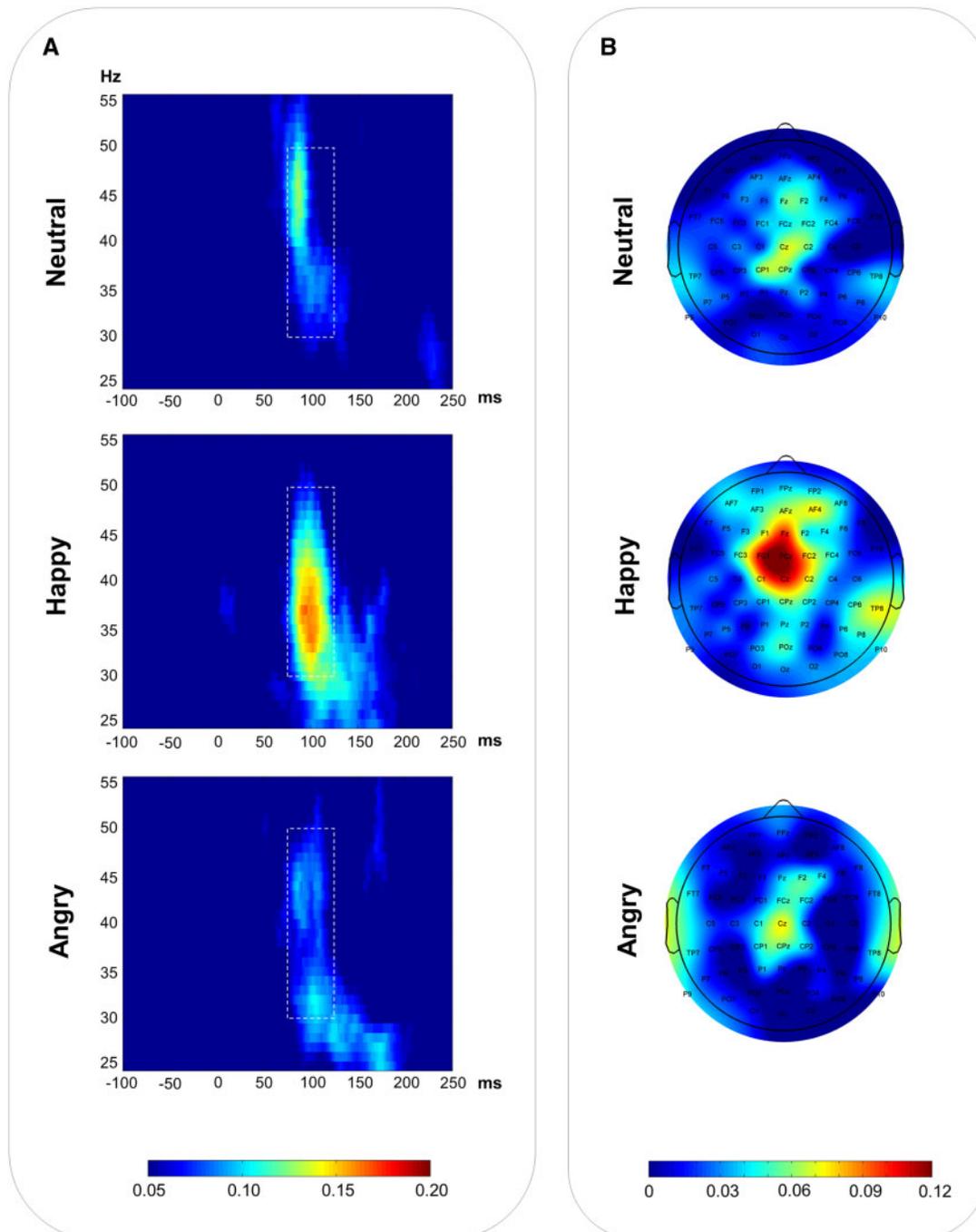
Fig. 7. (A) Time–frequency maps of PLF at FCz and Cz electrodes for neutral, happy and angry vocalizations. **(B)** Topographic maps of gamma-band PLF (30–50 Hz) in the 75–120 ms window for each condition.

Nonetheless, in a later window (310–410 ms), P3a amplitude was enhanced for emotional vocalizations, irrespective of valence. Because the later response in the P3 window seems to reflect the effective orienting of attention towards stimuli that are categorized as salient or emotionally relevant (e.g. Friedman *et al.*, 2001), our finding provides further support for the notion that emotional information cannot be ignored even when it is not task-relevant (e.g. Liu *et al.*, 2012), that unexpected emotional events are rapidly detected and categorized (e.g. Schirmer *et al.*, 2005; Liu *et al.*, 2012), and that emotional cues have a

preferential access to attentional resources (e.g. Thierry and Roberts, 2007). The enhanced mobilization of higher-order attentional resources for emotionally salient compared with neutral vocalizations plausibly has an evolutionary source, since it is advantageous to pay more attention to a potential danger or source of joy communicated through the voice in the social environment. For example, the potential threat carried by an angry voice may require a behavioral response with a protective function ('fight' or 'flight'). Nonetheless, contrary to what has been reported by studies on visual emotional processing, we did

Table 2. Affective ratings of the experimental stimuli by the participants of the study

Type of vocalization	Affective Dimension		
	Valence	Arousal	Dominance
	Mean (SD)	Mean (SD)	Mean (SD)
Neutral	4.73 (0.70)	3.07 (1.75)	6.80 (1.78)
Happy	8.27 (0.88)	6.07 (1.98)	6.53 (1.60)
Angry	2.27 (1.10)	6.93 (1.71)	3.07 (1.22)

Note: Mean ± SD are given for each variable.

not find a negativity bias (e.g. Ito *et al.*, 1998; Carretié *et al.*, 2001), i.e. enhanced P3a amplitude for angry relative to happy and neutral vocalizations. It is plausible that both the modality of stimulus presentation (visual *vs* auditory) and task (task-relevant *vs* task-irrelevant change detection; change detection when attention is focused on the auditory stimuli *vs* elsewhere) have an impact upon how the behavioral relevance of a stimulus is assessed. The ERP amplitude differences between emotionally salient and neutral vocalizations in this study cannot be explained by differences in stimulus probability, a factor that affects the P3, since the three types of vocalizations were presented with the same probability. Also, it is unlikely that the surprise or unusualness of the sounds were the only factor explaining the effects, since if that were the case we would have observed similar ERP amplitude for the three conditions in the later P3a peak. Furthermore, these findings do not seem to be explained by increased habituation to neutral relative to emotional vocal sounds, since no significant interaction between moment and condition was found when analyzing the ERP amplitude in the two P3 latency windows. In other words, the saliency of task-irrelevant stimuli, more than their novelty, seems to affect the physiological response in a later stage of the P3a.

Another important finding from our study is the increased phase synchronization for happy relative to both angry and neutral vocalizations in the early auditory-evoked gamma-band response that is typically observed in the first 150 ms after stimulus onset (e.g. Pantev *et al.*, 1991). The PLF differentiation between positive and negative vocalizations occurred even though the two types of non-target vocalizations elicited a similar P3a response. The early phase-locked auditory gamma-band response, observed within 100 ms after stimulus onset, reflects the synchronization of neural assemblies involved in the perceptual processing of sensory input (Hall *et al.*, 2009) that is critical for the integration of stimulus features and for the assignment of emotional significance to those features (e.g. Garcia-Garcia *et al.*, 2010). In particular, the PLF represents a measure of temporal consistency in neural synchrony, i.e. the event-related phase synchronization across trials (e.g. Hall *et al.*, 2009). The current findings suggest that the neural response to unexpected and task-irrelevant happy vocalizations was characterized by increased temporal synchrony compared with neutral or angry vocal stimuli.

Previous studies suggested that early phase synchronization in the gamma-band reflects feature binding underlying stimulus representation (Herrman *et al.*, 2004). More recently, the role of gamma-band phase synchronization in sensory prediction and change detection has been highlighted (e.g. Fell *et al.*, 2003; Herrman *et al.*, 2004; Kveraga *et al.*, 2007). In oddball paradigms, the early auditory evoked gamma-band response has been considered an index of a comparison process involving matching bottom-up signals against top-down expectations or a memory

template of the target (e.g. Debener *et al.*, 2003; Roye *et al.*, 2010): when the perceived sound matches the expected sound, an enhanced early evoked gamma-band response is observed (e.g. Engel *et al.*, 2001; Widmann *et al.*, 2007). In emotional oddball paradigms (e.g. novel sounds presented in a neutral or negative visual context—Garcia-Garcia *et al.*, 2010), the synchronization of the early auditory evoked gamma-band response has been found to index the integration of novelty and emotional salience information (Garcia-Garcia *et al.*, 2010), lending support to the idea that neurons in the primary auditory cortex are tuned to respond to the motivational salience of a stimulus (e.g. Weinberger, 2004).

Even though previous studies suggested that the early gamma-band response mainly reflects sensory processing (Karakas and Basar, 1998), we have strong reasons to believe that this was not the main factor accounting for the PLF differences between the three types of vocalizations. If that were the case, then we would have additionally observed PLF differences between neutral and angry vocalizations. Indeed, the mean F0 characterizing the vocal stimuli was higher for angry (414.36 Hz) than for neutral (212.58 Hz) and happy (350.61 Hz) vocalizations. The acoustic difference between angry and neutral vocalizations was, therefore, the highest (difference angry–neutral = 201.78 Hz; difference happy–neutral = 138.03 Hz; difference angry–happy = 63.75 Hz), but no significant PLF difference was observed between neutral and angry vocalizations. We have additionally analyzed the mean F0 of the first 120 ms of each vocal stimulus type (neutral = 199.66 Hz; happy = 339.55 Hz; angry = 421.74 Hz): the pitch pattern was the same when considering this shorter sample or the total stimulus duration, i.e. increased F0 for angry, followed by happy and then by neutral vocalizations. It is also unlikely that the PLF differences between happy and angry vocalizations were driven by arousal differences, as subjective arousal ratings were very similar for angry and happy stimuli, and no significant associations were found between PLF findings and participants' affective ratings.

Considering previous evidence demonstrating the role played by gamma oscillations in the association of the perceptual properties of stimuli with their emotional significance (e.g. Oya *et al.*, 2002; Garcia-Garcia *et al.*, 2010), it is surprising that the PLF was not increased for angry vocalizations, refuting our initial hypothesis. A plausible explanation is that mood-congruent effects influenced the match between bottom-up stimulation and top-down expectations, accounting for the current pattern of findings. Indeed, since the participants were college students with no history of psychopathological disorder, it seems possible that happy vocalizations represented the vocal stimulus category that was more congruent with their mood, and more frequently encountered in daily life social interactions. As such, the increased top-down expectations for pleasant vocal stimuli may have resulted in a facilitated match with bottom-up signals and, therefore, in increased early temporal consistency of neural synchrony in the gamma-band (see Herbert *et al.,* 2008 for a similar account based on ERP results). This interpretation is in concordance with studies suggesting that healthy subjects present a bias towards pleasant information, when compared with depressive subjects (e.g. Deldin *et al.,* 2001; Herbert *et al.,* 2008), and demonstrating the particularly salient role of happy vocalizations in healthy participants with no neurological, psychiatric or other medical problems (e.g. Johnstone *et al.,* 2006; Warren *et al.,* 2006). In keeping with our results, Roye *et al.* (2010) found that the early gamma-band response was enhanced for participant's own ringtone compared

with the ringtone of others within the first 100 ms after stimulus onset. The authors hypothesized that the auditory system may pre-activate the representation of a personally significant sound. These findings add to more recent evidence showing top-down influences on early sound processing (e.g. Widmann *et al.*, 2007), and extending the role of the primary auditory cortex beyond the analysis of stimulus acoustic properties to include: (i) the analysis of the behavioral significance of those acoustic properties; (ii) and (perhaps more importantly for the interpretation of the current findings) the short-term and long-term retention of specific memory traces about the sounds' behavioral significance (Weinberger, 2004). It is plausible that, because of the absence of any threat-related consequences in the environment, the evoked gamma-band response to angry vocalizations was less temporally stable across trials, in this early processing stage.

Considering the early time window of this effect (75–120 ms after stimulus onset), it is likely that the acoustic analysis of the complex and rapidly changing vocal stimulus properties occurred in parallel with the analysis of other stimulus features (e.g. stimulus motivational value; memory representations) (Roye *et al.*, 2010). In spite of the differences between happy and angry vocal stimuli in the early evoked gamma-band response (i.e. before full access to the emotional meaning of the vocalizations took place), they elicited similar attentional processing in a later stage indexed by the P3a (310–410 ms). Therefore, distinct stages of auditory processing may be characterized by different neurophysiological effects, as a result of the dynamic interplay between bottom-up sensory mechanisms and top-down cognitive processes (e.g. Schirmer and Kotz, 2006). For example, recent models such as the Auditory Event Representation System (Schroger *et al.*, 2014) suggest that after the formation of auditory stimulus representations, regularity representations that predict subsequent sensory input are formed and, subsequently, predictions and sensory input are compared. A final stage is the evaluation of the relevance of the relationship between the incoming sound events and the context. Furthermore, the model of Schirmer and Kotz (2006) proposes that vocal emotional comprehension is a multi-stage process consisting of several stages that are differentially represented in the brain: after the sensory processing of emotionally relevant acoustic cues (within the first 100 ms after stimulus onset), salient acoustic features are categorized as emotional *vs* non-emotional (around 200 ms after stimulus onset) and, in the last stage (approximately after 300 ms), the cognitive evaluation of the emotional significance of the voice signal takes place.

We should note that differences in task requirements and stimulus type may be a critical factor explaining the apparent inconsistency between this and other studies (e.g. Garcia-Garcia *et al.*, 2010). For example, the study of Garcia-Garcia *et al.* (2010) did not include a positive emotional context, and participants were asked to focus their attention on visual stimuli selected from the International Affective Picture System (Lang *et al.*, 2005) and to ignore the sounds (a 700 Hz standard tone and 100 unique environmental sounds). Furthermore, in previous studies from our research team we have observed differences in the pattern of early gamma-band phase synchronization to neutral, happy and angry vocalizations using a passive oddball paradigm (manuscripts in preparation). Future studies should clarify the effects of task requirements on PLF and ERP effects. In addition, in order to further explore the putative role of mood-congruency effects on the early auditory gamma-band response, future studies should investigate how emotional vocal change is detected by participants with mood disorders, such as depression.

Conclusions

Much work remains before we understand vocal emotional processing as well as face perception is currently understood (e.g. Belin *et al.*, 2004). In the context of social communication, it is advantageous to prioritize emotionally salient events, such as angry or happy vocalizations. We observed that the irregularity imposed by the three types of unexpected and task-irrelevant auditory events was similarly detected, but that only emotional vocalizations led to enhanced attention orienting that is necessary for goal-directed action selection, in the later P3a peak. In spite of the fact that ERPs did not differentiate between happy and angry vocalizations, increased synchronized oscillatory activity in the gamma frequency band for happy vocalizations suggests that pleasant vocal signals are associated with early increased synchronous neural activity across trials than unpleasant or neutral ones.

Together, these findings further support the idea that the brain must prioritize the processing of emotionally salient stimuli even when they are not task-relevant. In the context of social communication, this means devoting more attentional resources to angry and happy vocalizations, which represent salient signals in a social landscape.

Funding

Acknowledgements

References

Belin, P., Bestelmeyer, P.E.G., Latinus, M., Watson, R. (2011). Understanding voice perception. *British Journal of Psychology,* **102**(4), 711–25.

Belin, P., Fecteau, S., Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, **8**(3), 129–35.

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), 531–9.

Boersma, P., Weenink, D. (2013). *Praat: Doing phonetics by computer* (Version 5.3.53) [Computer program]. Available: http://www.praat.org/. [July 10, 2013].

Bradley, M.M., Lang, P.J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, **25**(1), 49–59.

Campanella, S., Bruyer, R., Froidbise, S., *et al.* (2010). Is two better than one? A cross-modal oddball paradigm reveals greater sensitivity of the P300 to emotional face-voice associations. *Clinical Neurophysiology*, **121**(11), 1855–62.

Canavarro, M. C. (1999). Inventário de Sintomas Psicopatológicos [Brief Symptoms Inventory]: BSI. In: Simões, M.R. Gonçalves, M., & Almeida, L.S. (Eds.), *Testes e provas psicológicas em Portugal [Psychological Tests in Portugal]-vol. II*. Braga: SHO/APPORT.

Canavarro, M.C. (2007). Inventário de Sintomas Psicopatológicos (BSI): Uma revisão crítica dos estudos realizados em Portugal. In: Simões, M.R., Machado, C., Gonçalves, M., Almeida, L., editors. *Avaliação psicológica. Instrumentos validados para a população Portuguesa*, **vol. III**, pp. 305–31. Coimbra: Quarteto Editora.

Carretié, L., Mercado, F., Tapia, M., Hinojosa, J.A. (2001). Emotion, attention, and the 'negativity bias', studied through event-related potentials. *International Journal of Psychophysiology*, **41**(1), 75–85.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, **1**(2), 245–76.

Chennu, S., Noreika, V., Gueorguiev, D., *et al.* (2013). Expectation and attention in hierarchical auditory prediction. *The Journal of Neuroscience*, **33**(27), 11194–205.

Compton, R.J. (2003). The interface between emotion and attention: a review of evidence from psychology and neuroscience. *Behavioral and Cognitive Neuroscience Reviews*, **2**(2), 115–29.

Courchesne, E., Hillyard, S.A., Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, **39**(2), 131–43.

Debener, S., Makeig, S., Delorme, A., Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research,* **22**(3), 309–21.

Debener, S., Herrmann, C.S., Kranczioch, C., Gembris, D., Engel, A.K. (2003). Top-down attentional processing enhances auditory evoked gamma band activity. *NeuroReport,* **14**(5), 683–6.

Deldin, P.A., Keller, J., Gergen, J.A., Miller, G. (2001). Cognitive bias and emotion in neuropsychological models of depression. *Cognition and Emotion,* **15**, 787–802.

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**(1), 9–21.

Derogatis, L.R. (1993). *BSI, Brief Symptom Inventory: Administration, Scoring, and Procedures Manual*. Minneapolis, MN: National Computer Systems.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, and Computers*, **30**(1), 34–43.

Domínguez-Borràs, J., Garcia-Garcia, M., Escera, C. (2012). Phase re-setting of gamma neural oscillations during novelty processing in an appetitive context. *Biological Psychology*, **89**(3), 545–52.

Donchin, E. (1981). "Surprise!… surprise?". *Psychophysiology*, **18**(5), 493–513.

Donchin, E., and Coles, M.G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, **11**(3), 357–74.

Engel, A.K., Fries, P., Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience,* **2**(10), 704–16.

Escera, C., Alho, K., Winkler, I., Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, **10**(5), 590–604.

Escera, C., Corral, M.J. (2007). Role of mismatch negativity and novelty-P3 in involuntary auditory attention. *Journal of Psychophysiology*, **21**(3–4), 251–64.

Escera, C., Yago, E., Alho, K. (2001). Electrical responses reveal the temporal dynamics of brain events during involuntary attention switching. *European Journal of Neuroscience*, **14**(5), 877–83.

Escera, C., Yago, E., Polo, M.D., Grau, C. (2000). The individual replicability of mismatch negativity at short and long inter-stimulus intervals. *Clinical Neurophysiology*, **111**(3), 546–51.

Fell, J., Fernandez, G., Klaver, P., Elger, C.E., Fries, P. (2003). Is synchronized neuronal gamma activity relevant for selective attention?. *Brain Research Reviews*, **42**(3), 265–72.

Fenske, M.J., Raymond, J.E. (2006). Affective influences of selective attention. *Current Directions in Psychological Science*, **15**(6), 312–6.

Friedman, D., Cycowicz, Y.M., Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, **25**(4), 355–373.

Fries, P., Reynolds, J.H., Rorie, A.E., Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, **291**(5508), 1560–3.

Garcia-Garcia, M., Yordanova, J., Kolev, V., Domínguez-Borràs, J., Escera, C. (2010). Tuning the brain for novelty detection under emotional threat: the role of increasing gamma phase-synchronization. *Neuroimage*, **49**(1), 1038–44.

Goldstein, A., Spencer, K.M., Donchin, E. (2002). The influence of stimulus deviance and novelty on the P300 and novelty P3. *Psychophysiology*, **39**(6), 781–90.

Gratton, G., Coles, M.G.H., Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology,* **55**(4), 468–84.

Hall, M.-H., Taylor, G., Sham, P., *et al.* (2009). The early auditory gamma-band response is heritable and a putative endophenotype of schizophrenia. *Schizophrenia Bulletin,* **37**(4), 778–87.

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.

Herbert, C., Junghofer, M., Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology,* **45**(3), 487–98.

Herrmann, C.S., Munk, M.H., Engel, A.K. (2004). Cognitive functions of gamma-band activity: memory match and utilization. *Trends in Cognitive Sciences*, **8**(8), 347–55.

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), 887–900.

Jessen, S., Kotz, S. A. (2013). On the role of crossmodal prediction in audiovisual emotion perception. *Frontiers in Human Neuroscience*, **7**, 369.

Johnstone, T., Van Reekum, C.M., Oakes, T.R., Davidson, R.J. (2006). The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience,* **1**(3), 242–49.

Karakas, S., Basar, E. (1998). Early gamma response is sensory in origin: a conclusion based on cross-comparison of results from multiple experimental paradigms. *International Journal of Psychophysiology*, **31**(1), 13–31.

Kissler, J., Herbert, C., Winkler, I., Junghofer, M. (2009). Emotion and attention in visual word processing - An ERP study. *Biological Psychology*, **80**(1), 75–83.

Kveraga, K., Ghuman, A.S., Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, **65**(2), 145–68.

Lang, P.J., Bradley, M.M., Cuthbert, B.N. (2005). *Technical Report A-6: International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual*. Gainesville, FL: University of Florida.

Laukka, P., Elfenbein, H.A., Söder, N., *et al.* (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology*, **4**, 353.

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), 108–12.

Nittono, H. (2006). Voluntary stimulus production enhances deviance processing in the brain. *International Journal of Psychophysiology*, **59**(1), 15–21.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia,* **9**(1), 97–113.

Olofsson, J.K., Nordin, S., Sequeira, H., Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. *Biological Psychology*, **77**(3), 247–65.

Oya, H., Kawasaki, H., Howard, M.A., Adolphs, R. (2002). Electrophysiological responses in the human amygdala discriminate emotion categories of complex visual stimuli. *The Journal of Neuroscience*, **22**(21), 9502–12.

Pantev, C. (1995). Evoked and induced gamma-band activity of the human cortex. *Brain Topography*, **7**(4), 321–30.

Pantev, C., Makeig, S., Hoke, M., Galambos, R., Hampson, S., & Gallen, C. (1991). Human auditory evoked gamma-band magnetic fields. *Proceedings of the National Academy of Sciences*, **88**(20), 8996–9000.

Paulmann, S., Kotz, S.A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport*, **19**(2), 209–13.

Pinheiro, A.P., Del Re, E., Mezin, J., *et al.* (2013). Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: an electrophysiological investigation. *Psychological Medicine*, **43**(3), 603–18.

Pinheiro, A.P., Rezaii, N., Rauber, A., *et al.* (2014). Abnormalities in the processing of emotional prosody from single words in schizophrenia. *Schizophrenia Research*, **152**(1), 235–41.

Polich, J. (1989). P300 from a passive auditory paradigm. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, **74**(4), 312–20.

Roach, B.J., Mathalon, D.H. (2008). Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia Bulletin*, **34**(5), 907–26.

Robinson, M.D. (1998). Running from William James' bear: a review of preattentive mechanisms and their contributions to emotional experience. *Cognition and Emotion*, **12**(5), 667–96.

Roye, A., Schroger, E., Jacobsen, T., Gruber, T. (2010). Is my mobile ringing? Evidence for rapid processing of a personally significant sound in humans. *Journal of Neuroscience,* **30**(21), 7310–3.

Sander, D., Grandjean, D., Pourtois, G., *et al.* (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *Neuroimage*, **28**(4), 848–58.

Schadow, J., Lenz, D., Thaerig, S., Busch, N. A., Frund, I.Herrmann, C. S. (2007). Stimulus intensity affects early sensory processing: Sound intensity modulates auditory evoked gamma-band activity in human EEG. *International Journal of Psychophysiology*, **65**(2), 152–61.

Scherer, K.R. (1995). Expression of emotion in voice and music. *Journal of Voice,* **9**(3), 235–48.

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**(1), 76–92.

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., Striano, T., Friederici, A.D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport*, **16**(6), 635–9.

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

Schroger, E., Bendixen, A., Denham, S.L., Mill, R.W., Bohm, T.M., Winkler, I. (2014). Predictive regularity representations in violation detection and auditory stream segregation: from conceptual to conceptual models. *Brain Topography,* **27**(4), 565–77.

Schupp, H.T., Junghöfer, M., Weike, A.I., Hamm, A.O. (2003). Attention and emotion: an ERP analysis of facilitated emotional stimulus processing. *Neuroreport*, **14**(8), 1107–10.

Spencer, K.M., Dien, J., Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, **36**(3), 409–14.

Spencer, K.M., Dien, J., Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, **38**(2), 343–58.

Squires, N.K., Squires, K.C., Hillyard, S.A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, **38**(4), 387–401.

Sutton, S., Braren, N., Zubin, J., John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, **150**, 1187–8.

Thierry, G., Roberts, M.V. (2007). Event-related potential study of attention capture by affective sounds. *Neuroreport,* **18**(3), 245–8.

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, **9**(12), 585–94.

Warren, J.E., Sauter, D.A., Eisner, F., *et al.* (2006). Positive emotions preferentially engage an auditory-motor "mirror" system. *Journal of Neuroscience,* **26**(50), 13067–75.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: NCS Pearson.

Weinberger, N.M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience*, **5**, 279–90.

Wickens, C., Kramer, A., Vanasse, L., Donchin, E. (1983). Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. *Science*, **221**(4615), 1080–2.

Widmann, A., Gruber, T., Kujala, T., Tervaniemi, M., Schroger, E. (2007). Binding symbols and sounds: Evidence from event-related oscillatory gamma-band activity. *Cerebral Cortex,* **17**, 2696–702.

Yamaguchi, S., Knight, R.T. (1991). Age effects on the P300 to novel somatosensory stimuli. *Electroencephalography and Clinical Neurophysiology*, **78**(4), 297–301.