

Rapid Detection of Emotion from Human Vocalizations

Disa Anna Sauter and Martin Eimer

Abstract

■ The rapid detection of affective signals from conspecifics is crucial for the survival of humans and other animals; if those around you are scared, there is reason for you to be alert and to prepare for impending danger. Previous research has shown that the human brain detects emotional faces within 150 msec of exposure, indicating a rapid differentiation of visual social signals based on emotional content. Here we use event-related brain potential (ERP) measures to show for the first time that this mechanism extends to the auditory domain, using human nonverbal vocalizations, such as screams. An early fronto-central positivity

to fearful vocalizations compared with spectrally rotated and thus acoustically matched versions of the same sounds started 150 msec after stimulus onset. This effect was also observed for other vocalized emotions (achievement and disgust), but not for affectively neutral vocalizations, and was linked to the perceived arousal of an emotion category. That this timing, polarity, and scalp distribution of this new ERP correlate are similar to ERP markers of emotional face processing suggests that common supramodal brain mechanisms may be involved in the rapid detection of affectively relevant visual and auditory signals. ■

INTRODUCTION

Humans are highly proficient communicators, and we use a wide range of social signals in our interactions, including facial, vocal, and postural cues. It has been postulated that affective processing is subserved by a functionally specialized network in the human brain (see Adolphs, 2003). An early step in the processing of emotional information is a differentiation of affective from nonaffective signals. Given that emotional signals provide important information about our environment and the likely behaviors of those around us, affective information needs to be detected and distinguished rapidly and reliably from other nonemotional stimuli.

In the visual domain, research using event-related brain potentials (ERPs) has found that the early processing of emotional signals differs from that of neutral stimuli, indicating that at least some of the mechanisms that are specifically involved in the processing of emotional signals are engaged rapidly. For pictures of emotional scenes, effects often start around 200–300 msec after exposure but peak substantially later, after approximately 1 sec (e.g., Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000). Although some studies have reported earlier effects during the processing of affective pictures, these components are sensitive to low-level perceptual features, which have not consistently been controlled for (see Olofsson, Nordin, Sequeira, & Polich, 2008).

The processing of pure human signals in the form of emotional facial expressions occurs faster and shows a differentiation between emotional and neutral faces as early

as 120–150 msec after stimulus onset and peaking before 200 msec (Eimer & Holmes, 2007). A number of studies have demonstrated an enhanced frontal positivity for fearful as compared with neutral faces, occurring around 150 msec after stimulus onset (e.g., Ashley, Vuilleumier, & Swick, 2004; Holmes, Vuilleumier, & Eimer, 2003; Eimer & Holmes, 2002). Similar differential effects have been demonstrated for a number of other facial expressions, including anger, happiness, disgust, sadness, and surprise (Eimer, Holmes, & McGlone, 2003). This suggests that this rapid differentiation is not caused by neural activation involved specifically in fear processing but rather is likely to reflect a more emotion-general mechanism.

Rapid processing of emotional facial expressions can even occur in the absence of viewers' conscious awareness of the emotional faces (Liddell, Williams, Rathjen, Shevrin, & Gordon, 2004). In a recent study, Kiss and Eimer (2008) demonstrated a differential ERP effect to fearful as compared with neutral faces. The faces were presented for 8 msec and then immediately followed by a masking stimulus, so that participants were unable to judge whether the faces they saw were neutral or emotional. This indicates that the neural differentiation of emotional from neutral human signals is not dependent on conscious perception of the emotional stimulus.

Analogous results have also been found in research using single-cell recordings in human pFC. This work has found emotion-specific responses to visual affective stimuli at comparative latencies to the effects seen using ERP (Kawasaki et al., 2001). However, humans also use other modalities than vision to communicate emotions. The voice is a crucial channel of communication for humans as well as other primates, and we are highly sensitive to

a range of vocal cues from the very beginning of life. For example, infants show preference for their mother's voice over other speakers (Mehler, Bertoncini, & Barriere, 1978) and for infant-directed over adult-directed speech (Fernald, 1985). Adult human listeners are also able to identify affective contents in vocal signals (Bryant & Barrett, 2007; Sauter & Scott, 2007).

The current study is the first to investigate whether the neural mechanism engaged in the rapid detection of visual affective signals extends into the auditory domain. Our aim was to establish whether emotional vocalizations would be processed differently from neutral sounds. Furthermore, this study sought to establish whether a differentiation between emotional and neutral sounds would occur quickly and whether it would be similar to that of emotional facial expressions, which elicit a frontal positivity around 150 msec after stimulus onset. This study used nonverbal vocalizations of emotions, such as screams and retching sounds. These kinds of signals are reliably recognized by human listeners (Sauter & Scott, 2007; Sauter, 2006; Schröder, 2003) but differ from emotionally inflected speech and nonsense speech in that they do not have the segmental structure of speech. They can thus be considered more analogous to emotional facial expressions than emotional speech and provide relatively "pure" vocal expressions of emotion (Scott et al., 1997).

A number of studies have investigated electrophysiological correlates of processing emotional speech, often looking at how prosodic and semantic cues interact (for a review, see Schirmer & Kotz, 2006). Several recent studies have shown a differentiation between emotional and neutral speech occurring around 200 msec after stimulus onset (Paulmann & Kotz, 2008; Schirmer, Simpson, & Escoffier, 2007; Schirmer, Striano, & Friederici, 2005). This work suggests that the detection of affect from vocal signals can also occur rapidly. However, the stimuli used in these studies were spoken sentences or isolated syllables, likely to engage systems involved in the decoding of the speech signal concurrently with affective processing. In contrast, the current study uses nonverbal vocalizations, which are not overlaid on the segmental structure of speech, and thus have their own acoustic-phonetic structure (Sauter, Calder, Eisner, & Scott, n.d.; Sauter, 2006).

Only one previous study has used ERP to examine the neural processing of nonverbal vocalizations (Bostanov & Kotchoubey, 2004). Using an oddball design, they found that context-incongruent stimuli elicited an increased negativity starting 300 msec after stimulus onset. They interpreted this as analogous to the incongruity effect commonly seen for semantically inappropriate words, which occurs after around 400 msec (Kutas & Hillyard, 1980). However, because Bostanov and Kotchoubey's (2004) study only included positive (joy) and negative (woe) sounds, it did not allow for a direct comparison between emotional and neutral sounds. The current experiments aimed to compare the processing of emo-

tional sounds to that of neutral control stimuli to examine whether emotional sounds are, like emotional faces, rapidly differentiated from affectively neutral stimuli.

EXPERIMENT 1

This experiment sought to establish whether and when a differential neural response to fearful vocalizations as compared with affectively neutral sounds can be observed in ERP waveforms. Given previous work in the domain of facial expressions (e.g., Eimer & Holmes, 2002), it is conceivable that fearful sounds are rapidly differentiated from neutral sounds, perhaps even within the first 200 msec after stimulus onset. In the current study, spectrally rotated versions of the emotional stimuli were used as neutral control sounds. Spectral rotation preserves amplitude envelope and duration information, pitch, and pitch variation while distorting spectral information (Blesser, 1972), thus providing a baseline condition matched for low-level acoustic features while lacking the affective perceptual quality of the original sounds (Warren et al., 2006).

Methods

Participants

Ten right-handed participants (eight women) with a mean age of 26.4 years took part in the experiment. All had self-reported normal hearing.

Stimuli

The stimuli were taken from a previously validated set of nonverbal vocalizations, with the critical stimuli consisting of 10 sounds expressing fear (Sauter et al., n.d.; Sauter & Scott, 2007; Sauter, 2006; Warren et al., 2006). Spectrally rotated versions of these fear sounds were also included (see Figure 1). Spectral rotation is a technique that can be considered analogous to inversion of facial stimuli in that the same physical information is present, but the global configuration is radically altered (Blesser, 1972). Thus, spectrally rotated sounds are acoustically well matched to the original sounds in terms of amplitude envelope, duration, and pitch but are perceived as affectively neutral (Warren et al., 1972) and do not sound like human vocalizations. The full stimulus set consisted of a total of 110 sounds (11 different sound categories with 10 sound tokens per category). The critical experimental stimuli included in the analyses were 10 tokens of fear sounds and spectrally rotated versions of the same 10 sounds, with an average duration of 0.8 sec. These stimuli were presented among other (distractor) sounds, with 10 tokens each communicating achievement, amusement, anger, disgust, neutral, pleasure, relief, and sadness. In addition, reversed versions of each of the 10 fear sounds were also included to provide an additional cate-

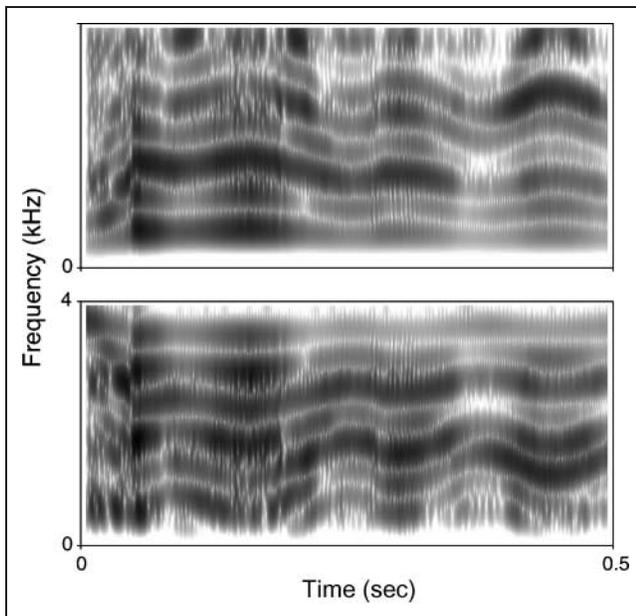


Figure 1. Spectrograms of a fear stimulus (above) and its spectrally rotated counterpart (below) from Experiment 2. All stimuli were low-pass filtered at 4 kHz. Spectral rotation was performed using the method described by Blesser (1972), in which the filtered signal is convolved with a sinusoid at 4 kHz, followed by low-pass filtering at 3.8 kHz. This acoustic manipulation produced unintelligible sounds that lacked the human vocal quality of the original stimuli but maintained a comparable level of acoustic complexity. The rotated sounds were matched to the original sounds in root mean square amplitude and long-term average spectrum.

gory of nonvoice stimuli, created by reversing the sampling points of the original waveforms in the time domain. Acoustic measurements of the sounds were made with PRAAT (Boersma & Weenink, 2005). A series of *t* tests were used to confirm that each emotional category and its spectrally rotated counterpart did not differ in terms of amplitude (root mean square), spectral center of gravity, or number of onsets, in the first 500 msec (all $p > .1$).

Procedure

E-Prime software was used for presentation and behavioral response collection (Psychology Software Tools, Pittsburgh, PA). Participants listened to sounds that were presented via Sennheiser headphones (model PMX 100). A small visual fixation point was displayed on a computer screen in front of the participants throughout the study. Participants performed an emotional one-back task, where they were required to respond with a button press if the current sound was of the same emotional category as the immediately preceding sound. Participants performed 10 blocks of 121 trials. On 11 trials per block, the stimulus presented in the preceding trial was repeated, once for each of the 11 stimulus types. On the remaining 110 nonrepetition trials, participants heard each of the 110 stimuli once in each block, played in a random order. Repetition and nonrepetition trials oc-

curred randomly within each block. Response accuracy was 94%, and false alarm rate was below 3%. One thousand milliseconds was allowed for responding after the end of each stimulus, with the next trial starting 300 msec later.

EEG Data Acquisition and Preprocessing

EEG data were recorded and digitized at a sampling rate of 500 Hz using an amplifier with a 0.1- to 40-Hz band-pass filter. Signals were recorded from 23 scalp electrodes mounted in an elastic cap at scalp sites (Fpz, F7, F3, Fz, F4, F8, FC5, FC6, T7, C3, Cz, C4, T8, CP5, CP6, P7, P3, Pz, P4, P8, PO7, PO8, and Oz). Horizontal eye movements were measured from two electrodes placed at the outer canthi of the eyes. All impedances were kept below 10 k Ω . Scalp electrodes were referenced to the left earlobe and re-referenced off-line to the average of both ears. No additional filters were applied after EEG recording. The continuous EEG was epoched off-line relative to the onset of an auditory stimulus. Epochs with activity exceeding ± 30 μ V in the HEOG channel (reflecting horizontal eye movements) or ± 60 μ V at Fpz (indicating eye blinks or vertical eye movements) were excluded from further analysis, as were epochs with voltages exceeding ± 80 μ V at any other electrode. On average, 28.1% of trials were removed due to the presence of artifacts. Waveforms were then averaged separately for each stimulus type. Only nontarget trials where no manual responses were recorded were included in the EEG analyses to avoid overlap of emotion-sensitive and response-related ERP components.

Results and Discussion

ERP mean amplitudes for fear sounds and their spectrally rotated counterparts were computed for the time window 150–300 msec after stimulus onset. Figure 2 shows grand-averaged ERPs measured at anterior and central electrodes, together with a topographic map illustrating the scalp distribution of the differential brain response to fearful versus rotated fearful sounds in the 150- to 300-msec poststimulus time window. To ascertain that rapid ERP responses to fear sounds were localized over anterior scalp areas, we conducted an initial exploratory omnibus repeated measures ANOVA for the factors area [five levels: fronto-polar (F7, Fpz, F8), frontal (F3, Fz, F4), central (C3, Cz, C4), posterior (P7, Pz, P8), and occipital (Po7, Oz, Po8)], site (three levels: left, midline, and right), and rotation (two levels: voiced and rotated). A significant main effect of rotation was found, $F(1, 9) = 6.4$, $p < .05$, reflecting differential processing of fear sounds as compared with rotated control sounds. An interaction between rotation and area, $F(4, 36) = 12.2$, $p < .001$, was also present, indicating that this effect was not evenly distributed. Follow-up analyses conducted separately for fronto-central (fronto-polar, frontal, and

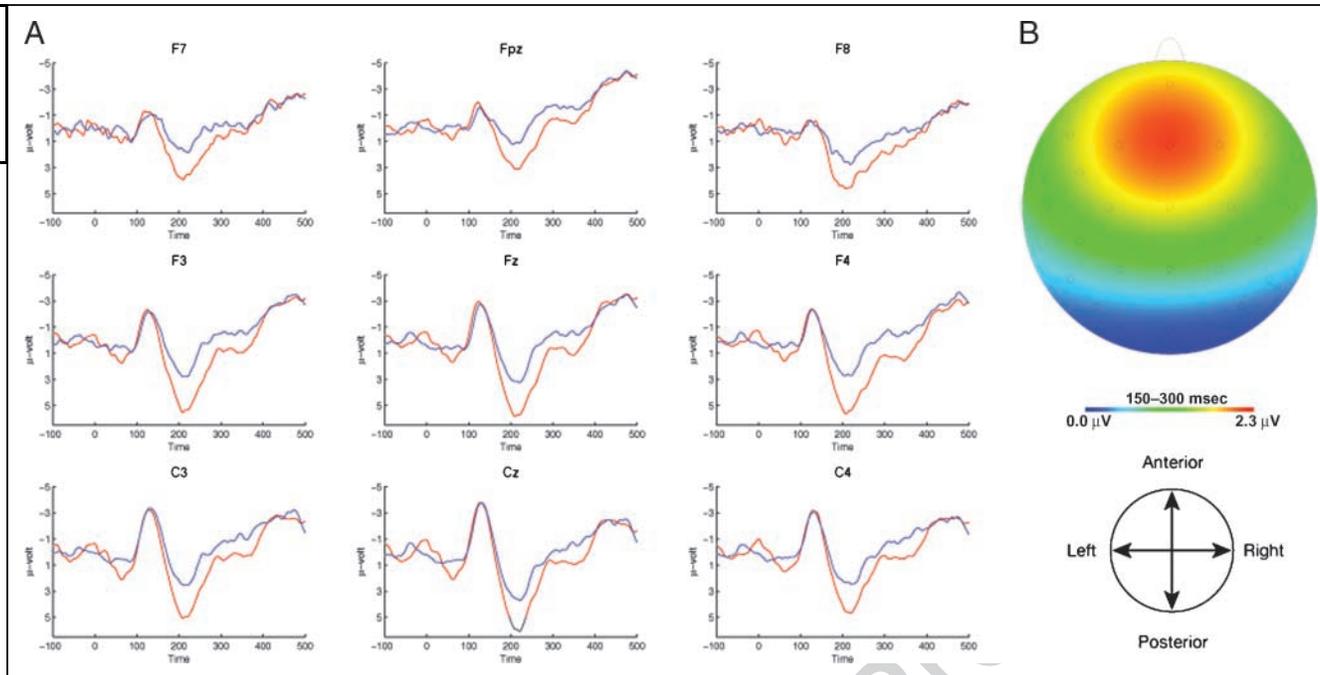


Figure 2. (A) Grand-averaged ERPs elicited in the -100 - to 300 -msec interval relative to sound onset (correct rejection trials only) in Experiment 1 at fronto-central electrodes for fear (red lines) and rotated fear sounds (blue lines). (B) Topographical map showing scalp distributions of differential effect to fear sounds in Experiment 1, obtained by subtracting ERPs to fear minus rotated fear sounds, in the 150 - to 300 -msec latency window. Red and yellow colors indicate an enhanced positivity for fearful relative to rotated fear sounds.

central) and posterior (posterior and occipital) confirmed that this effect was localized over anterior areas. No effect of rotation was present at posterior electrodes, $F(1, 9) = 1.1, p = .32$, but this effect was highly significant in the fronto-central area, $F(1, 9) = 22.5, p < .001$. The amplitude of this early positivity for fear sounds relative to spectrally rotated control sounds did not differ reliably between frontopolar, frontal, and central electrodes, $F(2, 18) = 2.5, p = .11$, suggesting that this effect was broadly distributed across the fronto-central area (see Figure 2).

In sum, Experiment 1 showed that the processing of fearful sounds is rapidly differentiated from acoustically matched neutral sounds. This enhanced fronto-central positivity effect starts as early as 150 msec after sound onset and is similar in terms of timing, polarity, and scalp distribution to the pattern found in previous studies of emotional facial expressions (see Eimer & Holmes, 2007).

EXPERIMENT 2

The results of Experiment 1 suggest that a rapid detection of the emotional quality of vocalizations takes place in anterior brain regions, resulting in an enhanced fronto-central positivity for emotional sounds as compared with their affectively neutral, spectrally rotated counterparts. However, it could be argued that this effect might simply reflect a differential brain response to rotated versus unrotated sounds that is entirely unrelated to their emotional content. If the effect observed in Experiment 1 was indeed due to an early differentiation between emotional and

nonemotional vocalizations, it should be absent in response to neutral as compared with rotated neutral vocalizations. To demonstrate this, we conducted a second experiment, which included neutral vocalizations as well as four emotional conditions (achievement, disgust, fear, and relief), each with its own acoustically matched spectrally rotated baseline. The inclusion of other emotions allowed us to explore whether the differential effect seen for fear would generalize to other emotion categories and whether it would be related to particular aspects of the emotions, such as perceived arousal. This design also ensured that the rotated sounds were no longer infrequent oddball stimuli, as half of the stimuli in each emotion category were spectrally rotated.

Methods

Participants

Seventeen participants (8 women, 14 right-handed; mean age = 24.0 years) with self-reported normal hearing took part in Experiment 2.

Stimuli

The stimuli in Experiment 2 were taken from the same corpus as Experiment 1. They consisted of 10 tokens each communicating achievement, fear, disgust, or relief, and affectively neutral vocalizations, as well as spectrally rotated versions of all of these sounds, making a total of 100 sounds, with an average duration of 0.9 sec. The stim-

ulus set was matched for peak amplitude. Acoustic measurements of the sounds were made with PRAAT (Boersma & Weenink, 2005). The *t* tests confirmed that each emotional category and its spectrally rotated counterpoint did not differ in terms of amplitude (root mean square), spectral center of gravity, or number of onsets in the first 500 msec (all $p > .3$). In addition, a beep sound (a 370-Hz sinusoid tone with a duration of 1 sec) was included.

Procedure

The same stimulus presentation set-up as in Experiment 1 was used, but participants performed a beep detection task, where they were required to respond with a button press every time they heard the beep. This is because using an emotional one-back task (“respond to immediate repetitions of the same emotion”) would have produced a confound between the voiced and the rotated stimuli because rotated sounds are typically not perceived as emotional (Warren et al., 2006). Participants would likely judge all rotated stimuli as affectively similar (i.e., neutral) and thus respond on most rotated trials preceded by another rotated trial. These trials would have to be eliminated from EEG analyses, resulting in a different trial numbers contributing to rotated and unrotated ERPs. Participants performed 10 blocks of 110 trials, with 10 beeps in each block, and 10 trials of each stimulus type per block. Response accuracy was 94%, with less than 1% false alarms. As in Experiment 1, 1000 msec was allowed for responding after the end of each stimulus, with the next trial starting 300 msec later.

EEG Data Acquisition and Preprocessing

Data acquisition and preprocessing were identical to Experiment 1. Any trials during which participants responded to an emotional or rotated sound were excluded. Trials during which participants heard the beep were also excluded from all ERP analyses. On average, 19.1% of trials were removed due to artifacts.

Results and Discussion

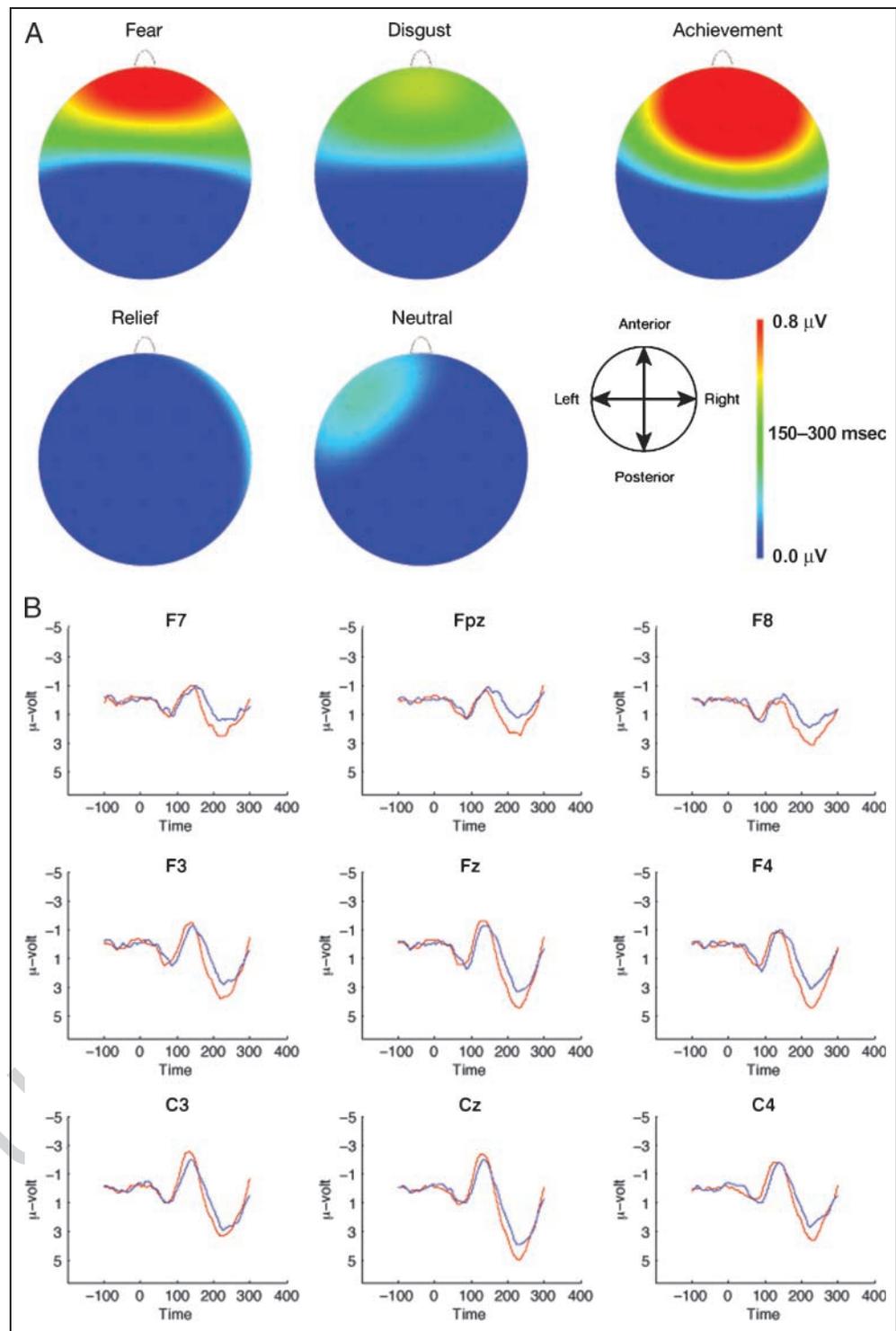
Figure 3A shows topographic maps of differential brain responses to unrotated sounds as compared with their matched rotated control sounds in the 150- to 300-msec poststimulus time window, for each of the four emotion conditions, as well as for neutral sounds. To illustrate the time course of the ERP differences between unrotated and rotated sounds, Figure 3B also shows ERPs obtained for one emotion condition (achievement). Similar to the results observed in the first experiment (Figure 2), an enhanced anterior positivity for emotional vocalizations can be seen for fear, achievement, and disgust but not for relief. Importantly, there was also no sign of any differential processing of neutral versus rotated neutral sounds. To examine these effects statistically, we carried

out an initial omnibus ANOVA, which investigated effect of emotion (five levels: fear, disgust, achievement, neutral, and relief), rotation (two levels: voice and rotated), area (five levels: fronto-polar, frontal, central, parietal, and occipital), and site (three levels: left, right, or center) on ERP mean amplitudes in the 150- to 300-msec post-stimulus interval. A significant interaction between area, rotation, and emotion, $F(16, 256) = 2.2, p < .01$, was found, indicating that emotion-sensitive differential brain responses were unevenly distributed across anterior and posterior areas. An interaction between emotion and rotation was only found for fronto-central electrodes, $F(4, 64) = 2.7, p < .05$, but not for posterior electrodes ($F < 1$), confirming that differential effects of rotation for the different emotion conditions were localized over anterior brain areas.

Further ANOVAs were conducted to specifically evaluate the effects of rotation separately for each of the five each emotion conditions at anterior electrodes, with area (fronto-polar, frontal, and central) as additional factor. For neutral vocalizations, there was no indication of any effect of rotation or interaction between rotation and area (both $F < 1$). This observation is important because it shows that the differential ERP response to fear as compared with rotated fear sounds observed in Experiment 1 is not simply the effect of being presented with spectrally rotated versus unrotated sounds, irrespective of their emotional quality. In contrast, systematic effects of rotation were found for different types of emotional vocalizations. For fear, there was no significant main effect of rotation, $F(1, 16) = 1.3, p = .28$. However, an interaction between rotation and area, $F(2, 32) = 8.7, p < .001$, suggested that the differential brain response to voiced as compared with rotated vocalizations of fear was more localized than in the first experiment (see Figure 3). This was confirmed by follow-up analyses, which revealed a reliable effect of rotation at fronto-polar electrodes, $F(1, 16) = 4.5, p < .05$, but not at frontal or central sites. For achievement, a main effect of rotation, $F(1, 16) = 7.9, p < .05$, in the absence of an interaction between rotation and area, $F(2, 32) = 1.7, p = .19$, demonstrated a differential effect of voiced sounds, relative to rotated sounds, that was broadly distributed across anterior scalp sites. Although Figure 3 suggests the presence of a focal fronto-polar brain response to vocalizations of disgust as well, this was not reflected by a significant main effect of rotation or an interaction between rotation and area (both $F < 2.4$). However, a localized differential effect of rotation for disgust was found in the fronto-polar region for a narrower poststimulus time window of 180–220 msec, $F(1, 16) = 6.4, p < .05$. In contrast to the results obtained for fear, achievement, and disgust, there was no sign of any early anterior positivity when relief sounds were compared with their spectrally matched rotated counterparts, main effect of rotation, $F(1, 16) = 2.0$.

Thus, Experiment 2 revealed an early differential brain response to emotional vocalizations (i.e., an anterior

Figure 3. (A) Topographical maps showing scalp distributions of differential effect to emotional sounds in Experiment 2, obtained by subtracting ERPs to unrotated minus rotated sounds, in the 150- to 300-msec latency window. Yellow and red colors indicate an enhanced positivity for unrotated relative to rotated sounds. (B) Grand-averaged ERPs elicited in the -100 to 300-msec interval relative to sound onset in Experiment 2 at fronto-central electrodes for one representative emotion condition (achievement). Red lines show ERPs to achievement sounds, and blue lines show ERPs for rotated achievement sounds.



enhanced positivity for voiced sounds relative matched rotated sounds) for fear, achievement, and (although somewhat less reliably) for disgust, whereas this effect was entirely absent for relief and for emotionally neutral vocalizations. We then examined whether the differential ERP effect to the emotional sounds in Experiment 2 was related to the perceived arousal of the different emotional categories. A previous study, which used fMRI to compare the perception of nonverbal vocalizations to a

spectrally rotated baseline, found that increasing arousal was associated with enhanced activation in the presupplementary motor (pre-SMA) area, whereas the valence of the sound was associated with the activation levels in the inferior frontal gyrus (Warren et al., 2006). Ratings for perceived valence and arousal for the stimuli in Experiment 2 have previously been obtained from naive listeners for all of the categories except the neutral sounds, which were excluded from this analysis (Sauter, 2006). A cor-

relation analysis was carried out to examine whether the differential ERP effects of sound rotation observed in Experiment 2 were related to the perceived arousal or valence of the categories. The differential ERP effect was quantified as the difference between ERP amplitudes for unrotated versus rotated baseline sounds, obtained at anterior and posterior electrodes in the 150- to 300-msec poststimulus time window, separately for each emotion category. This analysis showed that there was a close relationship between the categories' arousal and the magnitude of the differential effect in the fronto-central (Pearson's $r = 0.97, p < .05$), but not the posterior area. Follow-up analyses showed that this relationship was significant for fronto-polar (Pearson's $r = 0.97, p < .05$) and frontal (Pearson's $r = 0.97, p < .05$) areas, with a trend in the central area (Pearson's $r = 0.94, p = .063$). There was no significant correlation between the perceived valence of the sounds and the differential effect in either area.

In sum, Experiment 2 demonstrated an enhanced frontal positivity that is elicited rapidly by the processing of emotional, but not neutral, human vocalizations. Specifically, fear, achievement, and (to some degree) disgust vocalizations elicited an enhanced fronto-central positivity starting 150–180 msec after stimulus onset. In addition, these data suggest that the magnitude of differentiation is greater for emotions that are perceived to be more aroused. The differential effect was most pronounced for achievement vocalizations, followed by fear and then disgust sounds. No differential processing was found for relief sounds, which are perceived as low in arousal and relatively neutral in terms of valence (Sauter & Scott, 2007).

GENERAL DISCUSSION

Our results demonstrate a differential response to emotional vocalizations as compared with acoustically matched control sounds, in the form of an enhanced frontal positivity starting as early as 150 msec after stimulus onset. The specificity of this effect to emotional sounds was demonstrated, establishing that it does not occur for neutral vocalizations. This is important because it rules out the possibility that this differentiation is due to neural responses to spectrally rotated versus unrotated sounds regardless of their emotional quality. Rather, our data show that emotional, but not neutral sounds, are differentiated rapidly from matched baseline sounds.

The effect is similar in timing, polarity, and scalp distribution to the effect found with emotional faces: an enhanced fronto-central positivity occurring around 150 msec after stimulus onset (Ashley et al., 2004; Eimer & Holmes, 2002). It is also consistent with a study which, using single-cell recordings in human pFC, found emotion-specific responses to visual affective stimuli starting 120–160 msec after stimulus onset (Kawasaki et al., 2001). Furthermore, a differentiation occurs at a similar time window during the processing of emotional speech (Paulmann & Kotz, 2008; Schirmer et al., 2005, 2007). The similarity of the dif-

ferential ERP responses found here in response to emotional versus neutral vocalizations and the ERP effects found before with emotional faces are particularly striking given that the visual stimuli used in previous studies have been static, whereas auditory stimuli inherently extend over time. Although the subject numbers in the current experiments were small, such similarities across studies suggest that the detection of emotionally salient visual and auditory information may be based on shared neural mechanisms. In addition to the fast detection of visual and auditory affective stimuli, there is also evidence that emotional signals from different sensory modalities are rapidly integrated. de Gelder, Böcker, Tuomainen, Hensen, and Vroomen (1999) investigated the detection of incongruence between affective signals from the face and voice, using MMN. On trials where a specific face–voice pairing deviated from the standard established in a given block, an MMN with a latency of about 180 msec after auditory stimulus onset was observed, suggesting that emotional information from vision and audition had already been combined, less than 200 msec after stimulus onset. It is possible that the mechanisms involved in rapidly detecting affective signals from auditory and visual cues may also be involved in the integration of these signals.

Although caution must be exercised in the interpretation of the localization of ERP effects given the limited spatial resolution of these measurements, the correlation pattern between the emotion differentiation and the perceived arousal of the sounds may indicate that the effect is linked to activity in the pre-SMA. This brain region is involved in motor planning (Rizzolatti & Luppino, 2001) and has previously been found to be sensitive to the arousal levels of emotional signals (Warren et al., 2006). Thus, affective auditory signals that are perceived as more aroused may to a greater extent engage neural mechanisms involved in motor production. This could reflect the listener's preparation to engage in either similar (e.g., for fear or positive emotions) or different actions (e.g., a common response to an angry expression is fear). Given the physical matching of emotional and neutral stimuli achieved by the spectral rotation procedure used in the current study, it is unlikely that the observed differences in magnitude between emotion categories reflect low-level acoustic differences. The greatest differentiation between emotional and control sounds was found for achievement sounds; this type of vocalization also elicited the greatest activation in the pre-SMA (compared with amusement, disgust, fear, and control sounds) in the study by Warren et al. (2006). It may be that achievement sounds are a particularly contagious vocal signal. However, recent work has shown that a closely related emotion, pride, has a reliably recognized postural display (Tracy & Matsumoto, 2008). Future work could investigate whether this potentially analogous visual signal produces a similar effect to the vocal cues in the current study.

In sum, our data demonstrate that the frontal areas of the human brain differentiate between emotional and

nonemotional conspecific signals in the auditory modality as early as 150 msec after the onset of the sound. The similarity of this effect to that found in previous studies of emotional face processing suggests that the human brain rapidly distinguishes affective from nonaffective conspecific signals in several modalities.

Acknowledgments

This work was supported by ESRC grant PTA-026-27-1372 to D. A. S. The authors would like to thank Sue Nicholas for her help with running participants and Frank Eisner for useful comments on previous versions of this manuscript. M. E. holds a Royal Society-Wolfson Research Merit Award.

Reprint requests should be sent to Disa Anna Sauter, School of Psychology, Birkbeck College, University of London, Henry Wellcome Building, London WC1E 7HX, United Kingdom, or via e-mail: d.sauter@bbk.ac.uk.

REFERENCES

- Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*, 165–178.
- Ashley, V., Vuilleumier, P., & Swick, D. (2004). Time course and specificity of event-related potentials to emotional expressions. *NeuroReport*, *15*, 211–216.
- Blessner, B. (1972). Speech perception under conditions of spectral transformation: I. Phonetic characteristics. *Journal of Speech and Hearing Research*, *15*, 5–41.
- Boersma, P., & Weenink, D. (2005). PRAAT: Doing phonetics by computer [computer software]. Retrieved from <http://www.praat.org/>.
- Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, *41*, 259–268.
- Bryant, G., & Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science*, *18*, 746–751.
- Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology*, *52*, 95–111.
- de Gelder, B., Böcker, K. B. E., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses. *Neuroscience Letters*, *260*, 133–136.
- Eimer, M., & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *NeuroReport*, *13*, 427–431.
- Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, *45*, 15–31.
- Eimer, M., Holmes, A., & McGlone, F. (2003). The role of spatial attention in the processing of facial expression: An ERP study of rapid brain responses to six basic emotions. *Cognitive, Affective, and Behavioral Neuroscience*, *3*, 97–110.
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development*, *8*, 181–195.
- Holmes, A., Vuilleumier, P., & Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: Evidence from event-related brain potentials. *Cognitive Brain Research*, *16*, 174–184.
- Kawasaki, H., Adolph, A., Kaufman, O., Damasio, H., Damasio, A. R., Granner, M., et al. (2001). Single-neuron responses to emotional visual stimuli recorded in human ventral prefrontal cortex. *Nature Neuroscience*, *4*, 15–16.
- Kiss, M., & Eimer, M. (2008). ERPs reveal subliminal processing of fearful faces. *Psychophysiology*, *45*, 318–326.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.
- Liddell, B. J., Williams, L. M., Rathjen, J., Shevrin, H., & Gordon, E. (2004). A temporal dissociation of subliminal versus supraliminal fear perception: An event-related potential study. *Journal of Cognitive Neuroscience*, *16*, 479–486.
- Mehler, J., Bertoncini, J., & Barriere, M. (1978). Infant recognition of mother's voice. *Perception*, *7*, 491–497.
- Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology*, *77*, 247–265.
- Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport*, *19*, 209–213.
- Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron*, *31*, 889–901.
- Sauter, D. (2006). An investigation into vocal expressions of emotions: The roles of valence, culture, and acoustic factors. PhD thesis, University of London.
- Sauter, D. A., Calder, A. J., Eisner, F., & Scott, S. K. (n.d.). *Perceptual cues in non-verbal vocal expressions of emotion*. Manuscript submitted for publication.
- Sauter, D. A., & Scott, S. K. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion*, *31*, 192–199.
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*, 24–30.
- Schirmer, A., Simpson, E., & Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research*, *1176*, 103–112.
- Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *NeuroReport*, *16*, 635–639.
- Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication*, *40*, 99–116.
- 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*, *385*, 254–257.
- Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences, U.S.A.*, *105*, 11655–11660.
- Warren, J. E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J. S., et al. (2006). Positive emotions preferentially engage an auditory-motor “mirror” system. *Journal of Neuroscience*, *26*, 13067–13075.