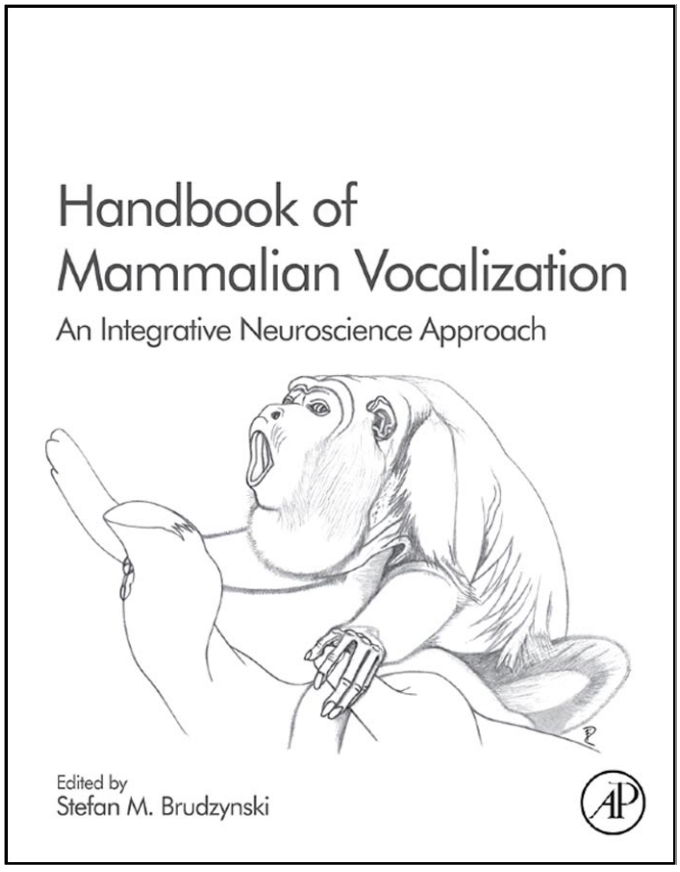
Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.

This chapter was originally published in the book *Handbook of Mammalian Vocalization*, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who know you, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:
http://www.elsevier.com/locate/permissionusematerial

From Sophie K. Scott, Brain mechanisms for processing perceived emotional vocalizations in humans. In: Stefan M. Brudzynski, editors, Handbook of Mammalian Vocalization. Oxford: Academic Press, 2009, pp. 187-198.
ISBN: 978-0-12-374593-4
© Copyright 2010 Elsevier BV.
Academic Press.

Author's personal copy

CHAPTER 5.5

Brain mechanisms for processing perceived emotional vocalizations in humans

Sophie K. Scott[1*], Disa Sauter[2] and Carolyn McGettigan[1]

[1]*Institute of Cognitive Neuroscience, University College, London, UK*
[2]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Abstract: Humans express emotional information in their facial expressions and body movements, as well as in their voice. In this chapter we consider the neural processing of a specific kind of vocal expressions, non-verbal emotional vocalizations e.g., laughs and sobs. We outline evidence, from patient studies and functional imaging studies, for both emotion-specific and more general processing of emotional information in the voice. We relate these findings to evidence for both basic and dimensional accounts of the representations of emotion. We describe in detail an fMRI study of positive and negative non-verbal expressions of emotion, which revealed that prefrontal areas involved in the control of orofacial movements were also sensitive to different kinds of vocal emotional information.

Keywords: voice; emotion; non-verbal expressions of emotion; functional magnetic resonance imaging (fMRI); superior temporal sulcus; premotor cortex; supplementary motor area; insula; amygdala

I. What are emotional vocalizations?

Humans are the talking apes, and our use of spoken language is considered to be one of our unique skills. Indeed, there is arguably no other sound in nature which has the complex spectrotemporal structure seen in speech. However, our voices also simultaneously express a multitude of other cues, about our mood, age, sex, health, social class and origins (Karpf, 2006). In this chapter we will outline the neural systems recruited to perceive a subset of human vocalizations: non-verbal vocal expressions of emotion. By non-verbal expressions of emotion, we specifically mean non-linguistic vocalizations, such as laughter, screams and sobbing. While speech itself is obviously frequently inflected with emotional information (e.g., we can hear whether a speaker is smiling or not (Tartter and Braun, 1994)), these non-verbal tokens are ideal for investigating the expression of emotional information when there is no concurrent verbal information,[1] and they thus closely mirror the kind of emotional information available via other channels, such as facial emotional expressions. In this chapter, we will address only some aspects of emotional speech, but will mainly emphasize non-verbal emotional vocalizations.

Speech production requires extremely fast and accurate movements of the articulators (the tongue, jaw, lips and soft palate), which shape the sounds produced at the larynx. Some of the complexity in speech can be seen in Fig. 1, which shows the acoustic structure of emotional speech and of non-verbal emotional vocalizations. Compared to the articulations of speech, non-verbal emotional expressions typically use less precise supraglottal articulations. The sound produced at the larynx is shaped by roughly positioned pharyngeal/oral/labial constrictions, which may in turn correlate with aspects of the associated facial expression.

*Corresponding author. E-mail: sophie.scott@ucl.ac.uk

[1]These kinds of sounds have been previously designated as "affect bursts" (Scherer, 1994): we avoid this term since the term "burst" connotes several acoustic features (rapid onsets, brief durations) which are not necessarily seen in the sounds themselves, which can start slowly and be relatively lengthy.

Stefan M. Brudzynski (Ed.)
Handbook of Mammalian Vocalization
ISBN 978-0-12-374593-4

DOI: 10.1016/B978-0-12-374593-4.00019-X
Copyright 2010 Elsevier B.V. All rights reserved.

Author's personal copy

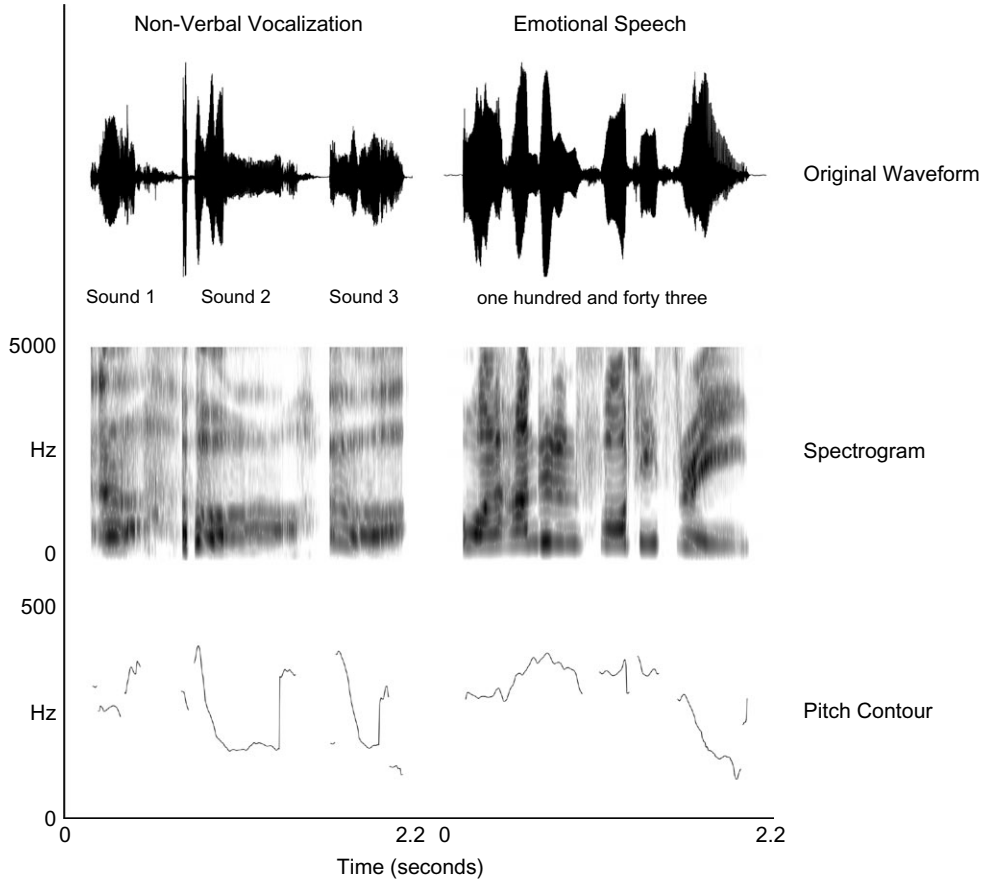


Fig. 1. A comparison of emotional speech and human emotional vocalizations. The top panel shows oscillograms of three examples of non-verbal disgust vocalizations on the left, and an oscillogram of a disgusted emotional speech example from the same (female) speaker. The middle and lower panel show the spectrograms and pitch profiles of the exemplars, respectively. The non-verbal vocalizations show less detailed spectral change than the speech, reflecting the cruder, relatively static positioning of the articulators in these emotional vocalizations. In contrast, the pitch profiles of the non-verbal vocalizations show big pitch changes throughout each utterance, while the pitch changes in the emotional speech are confined to the end of the utterance.

The spectral detail and changes in speech are much coarser in non-verbal vocalizations. The quality and control of the sound produced at the larynx has a big affect on the sounds found in emotional vocalizations. Thus, a sound like laughter, for example, has been described as more like a way of modified breathing than it is like a way of speaking (Abercrombie, 1967; cited in Kohler, 2008). These differences in the production of non-verbal emotional vocalizations are reflected in their acoustic structure. Unlike speech, which can be understood with only a certain degree of coarse spectrotemporal structure (Shannon et al., 1995), the recognition of emotional vocalizations is dependent on the preservation of information such as pitch and fine spectral structure, as well as cues such as those from the amplitude envelope (Sauter, 2006). This observation suggests that, unlike meaning in spoken language, the recognition of emotion in non-verbal emotional expressions is dominated by effects created at the glottis and postural properties (e.g., tension) in the articulators. These non-verbal emotional expressions are arguably more like animal vocalizations than they are like spoken language, since they do not require the dynamic spectral shaping caused by the rapid movements of the articulators (Jürgens, 1986).

Voice quality and pitch, as well as loudness and rate, are also important in the emotional inflection of spoken language. The recognition of emotional states in speech is strongly influenced by the pitch and pitch changes in the speech (Banse and Scherer, 1996;

Author's personal copy

Bänziger and Scherer, 2005) and by fine spectral properties (Murray and Arnott, 1993). However, the recognition rates for emotional information in speech are somewhat lower (around 50%) (Banse and Scherer, 1996) than they are for non-verbal, emotional, vocalizations (around 80%) (Sauter, 2006). While these differences may be due to the quality of the emotional stimuli (poor exemplars may not be representative of the potential emotional information in speech), this difference between emotional speech and non-verbal emotional vocalizations may also reflect the fact that emotion in speech is overlaid on the speech signal, and is thus somewhat more constrained in its expression than it is in non-verbal emotional vocalizations. Thus, there could be conflicts between the prosodic cues in sentence-level speech which denote the emotional information and those that cue linguistic information, e.g., the rising pitch of a question, the falling pitch of normally produced words, or the changes in pitch used to express linguistic emphasis and stress. Fig. 1 shows the pitch contours in both emotional speech and non-verbal emotional vocalizations. The pitch changes in the latter occur over most of the duration of the sounds, in contrast to that of the speech, where there are big changes in pitch only at the end of the utterance. These conflicts need not have a long-term effect on the communicative sense of the uttered speech, but could be enough to create difficulty in specifically making stimuli that unambiguously express a single emotion. In contrast, non-verbal emotional expressions are not produced in tandem with any other constraints other than those that affect the social and cultural ease with which these can be produced.

II. Two theories of emotion

There are two broad approaches to the study of emotion: those which identify different emotions as separate categories, which exist without reference to each other, and those in which different emotions are represented as points in emotional space, a space which is defined along two or three general dimensions (e.g., arousal, valence, approach–avoid). A dominant approach in the former class of theories is the theory of basic emotions. According to the theory of basic emotions, there is a subset of emotions that are recognized across different cultures (i.e., are universal), that are innate, that have distinctly different expressions, that can be processed in a rapid and automatic way, and that are implemented in different neural systems (Ekman et al., 1969; Ekman, 1992a,b). In contrast, a dominant dimensional account of emotions is the circumplex model (Russell, 1980). In this model, emotions are points in emotional space along the dimensions of arousal and valence.

III. Positive and negative emotions in the voice

Emotional vocalizations have been identified that mirror the categories of facial expressions of the basic emotions: fear; sadness; disgust; anger; happiness; and surprise (Scott et al., 1997). Notably, the majority of these six basic emotions – fear, sadness, anger and disgust – are negative in their valence. It has been suggested that further basic emotions, reflecting more positive moods, might be preferentially expressed with the voice, while being generally expressed facially with a smile (Ekman, 1992b). We specifically investigated this claim, developing a set of non-verbal emotional expressions of the hypothesized positive emotions of amusement, relief, sensual pleasure, contentment and triumph. We found good recognition rates for these stimuli (Sauter and Scott, 2007), with the exception of contentment, which seemed to be processed as a weaker version of sensual pleasure. While we have not tested whether it would be possible for people to distinguish between facial expressions of these emotions, this is good evidence that positive non-verbal expressions of emotion are well-recognized for the categories of amusement, triumph, relief and sensual pleasure, which goes some way to support Ekman's original argument. There is now evidence that, across cultures, people can infer pride from postural and facial cues (Tracy and Robins, 2008), which suggests that at least one more positive emotion can be read in the face and body. It is also, however, possible that faces and voices vary in how well they convey different kinds of emotion.

IV. Emotional vocalizations: general neural systems

Functional imaging (fMRI and PET) has been used to identify the cortical systems important in the perceptual processing of human vocalizations. The bilateral superior temporal sulcus (STS) has been designated a voice processing area (Belin et al., 2000), since it shows a robust response to human vocalizations across a wide range, relative to an acoustic control

Author's personal copy

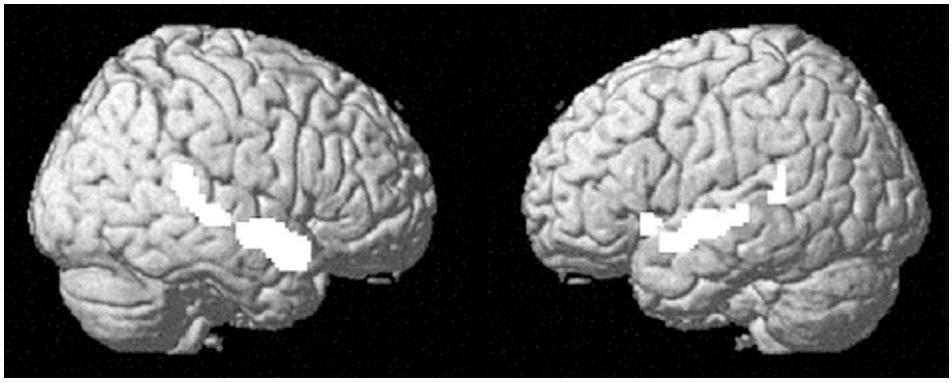


Fig. 2. Projection fields (white) of the contrast of emotional vocalizations over an acoustic control condition (data from Warren et al., 2006b) (data thresholded at false discovery rate = 0.05). The activity runs along the left and right superior temporal sulcus.

condition.[2] This STS system can be fractionated to some degree. The left mid-STS shows an enhanced response to sublexical phonetic structure in speech (Liebenthal et al., 2005), and the left anterior STS has been linked to the processing of intelligible speech (Scott et al., 2000; 2006) and to the formation of auditory word forms (Cohen et al., 2004). In contrast, the right STS has been argued to be important for processing pitch variation, since it shows a sensitivity to the dynamic pitch information in speech (Scott et al., 2000) and music (Patterson et al., 2002). In an adaptation study, the right anterior STS showed a specific response to changes in speakers, but not to changes in the words the different speakers said, suggesting a role in the processing of vocal identity (Belin et al., 2003). The right anterior STS has also been shown to be sensitive to signals with sufficient spectral detail for speaker identity to be available, over an acoustic control condition (Scott et al., 2006; Warren et al., 2006a).

The STS is also very sensitive to non-verbal emotional expressions; Fig. 2 is from an fMRI study comparing non-verbal expressions of emotion to a well-matched acoustic control condition (Warren et al., 2006b), and shows the comparison of all emotional vocalizations over spectrally rotated equivalents (the control condition). This shows extensive bilateral STS activation, extending further in the anterior STS on the right than on the left side of the brain. This result suggests that, as for speech and speaker information, the STS is an important stage in the perception and representation of non-verbal emotional stimuli, following acoustic processing in the superior temporal gyrus. Within the STS there may be a left/right distinction, with the left anterior STS fields being more sensitive to linguistic information, and the right anterior STS fields being more sensitive to talker identity (Belin and Zatorre, 2003) and emotion (Grandjean et al., 2005; Schirmer and Kotz, 2006) information in the voice. It has been suggested that, in face processing, visual speech is processed very distinctly from visual identity and emotional expression information, which are seen to share many properties at perceptual levels (Calder and Young, 2005). This broad left/right distinction between speech information and talker/emotion in the voice may thus mirror functional aspects of the organization of facial processing.

V. Emotional vocalizations: specific neural systems

In the theory of basic emotions (Ekman et al., 1969; Ekman, 1992a,b), it is hypothesized that there are distinct neural systems for different basic emotion categories. Neuropsychological studies of emotion recognition impairments following brain damage, and functional imaging studies of the neural systems recruited by different emotions in participants with undamaged brains provide good ways of testing this hypothesis, although typically small numbers of patients are studied in the former case. These techniques allow us to identify different neural systems

[2]Control conditions are important in functional imaging studies, since the observed activation is always determined relative to some other condition. In studies with communication sounds, such as speech or other vocalizations, this allows us to distinguish between neural activation associated with low-level acoustic stimulation, and activation associated with the specific properties of the sounds of interest.

Author's personal copy

that are recruited in the perception of different emotions, and that are distinct from general auditory processing in the superior temporal gyrus (STG) and STS. The "basic" emotions, comprising anger, disgust, fear, sadness, surprise and happiness, have been investigated mainly using facial expressions of emotion; however, there are some neuropsychological and neuroscientific investigations using non-verbal vocal expressions of emotion.

VI. Neuropsychological studies of emotion recognition

VI.A. Fear and anger

Patients with amygdala damage have difficulty in processing identity from faces, and also in processing facial signals communicating certain emotions, particularly fear and anger. Work with a patient who had bilateral amygdala resections (patient DR) found that she also showed a constellation of problems in dealing with vocal information. She showed deficits in perceiving non-emotional prosodic information in speech (e.g., the difference between a question and a command), while performing normally on tests of speech perception, sentence comprehension and recognition of environmental noises (Scott et al., 1997). Consistent with her problems in identifying people from photographs, she performed poorly on voice recognition tasks, both matching unfamiliar voices and recognizing familiar voices. In emotional processing, she was impaired on the recognition of sadness, fear and anger when tested with emotionally inflected single words: she did not perform well on the other emotions tested (happiness and disgust); however, her performance was not significantly different from that of age-matched control participants (suggesting that the emotional speech stimuli were not necessarily well-recognized). When tested with non-verbal emotional expressions, however, the results were clearer: DR was impaired on the recognition of fear and anger, and performed at ceiling for the recognition of the other emotions tested (disgust, sadness, surprise and happiness), while the control subjects recognized all the different emotional noises well. Thus, as in processing facial emotional information, DR had a specific problem with the processing of anger and fear from the voice, and this problem was clearer when tested with non-verbal emotional vocalizations. This finding has been replicated in other patients, and extended to emotional bodily postures of fear (Sprenglemeyer et al., 1999). These results have been taken as evidence that the amygdala (while clearly also involved in a variety of processes) has an important role in the processing of threat. Thus, the amygdala has been hypothesized to process angry cues from others because the angry person could be a threat, and to process fearful cues from others because whatever is threatening them could also threaten you. Some studies have failed to replicate this effect with vocal expressions of fear (e.g., Anderson and Phelps, 1998). However, in this case, the patient did show poorer recognition of the non-verbal recognition of fearful sounds than of any other non-verbal emotional expression; the control subjects also performed very variably on the recognition of these fear sounds, suggesting that the fear stimuli used in this study were generally not well-recognized.

VI.B. Disgust

There is some evidence that disorders affecting the basal ganglia can lead to specific deficits in processing emotions. Thus, patients with Huntingdon's chorea, a heritable dementia whose effects are initially mainly in the basal ganglia, have been shown to have deficits in the perception of disgust from the face and in emotional speech (Sprengelmeyer et al., 1996). A case study of a patient with specific damage to the putamen and the anterior insula (Calder et al., 2000) also showed deficits in the recognition of disgust from emotional facial expressions, and from both verbal and non-verbal vocal emotional expressions.

VI.C. Sadness and anger

Studies of patients with frontal lesions, grouped by lesion location (Hornak et al., 2003), have revealed that specific problems in processing non-verbal emotional vocalizations of sadness can be seen following damage to the ventromedial frontal lobe. In a subset of these patients, problems with vocal expressions of anger were also reported.

VI.D. Happiness

The presupplementary motor area (pSMA) has been implicated in the production of laughter (Fried et al., 1998). A depth electrode investigation of an epileptic patient showed that stimulation of the pSMA resulted

Author's personal copy

in first smiling, then laughter and a sensation of mirth (Krolak-Salmon et al., 2006). Increasing the stimulating current led to signs of distress, such as moaning and tears (Krolak-Salmon et al., 2006). Recording from this area also showed an enhanced response to facial expressions of happiness (i.e., smiling faces) (Krolak-Salmon et al., 2006). The authors suggested that these effects might reflect an involvement of a brain mirror system in the recognition, experience and production of happiness, since the pSMA was involved in the production of smiles and laughter, in the experience of mirth and the perception of happiness in the face (Krolak-Salmon et al., 2006).

VII. Separate neural systems for different vocal expressions of emotion?

There is thus some evidence from patient studies that damage to different brain areas can lead to specific deficits in the recognition of different vocal expressions of emotion. Amygdala damage can lead to deficits in the recognition of fear and anger, insula and putamen damage can detrimentally affect disgust recognition, orbito-frontal lesions can affect sadness, anger and disgust recognition, and the expression of happiness (laughter) recognition is associated with stimulation of the pSMA. In many cases, these recognition deficits are seen across different modalities (e.g., face as well as voice), which would be congruent with the notion of basic expressions of emotion (Ekman et al., 1969; Ekman, 1992a,b). As might be expected, however, there is no clear evidence of simple one-to-one mapping between brain regions and recognition of specific emotions – the amygdala does not exclusively support fear recognition and likewise single emotions can be affected by a variety of brain regions – for example, disgust deficits can be seen following damage to the insula and putamen (Calder et al., 2000), or orbitofrontal lesions (Hornak et al., 2003). This pattern is borne out by functional imaging studies, which have tended to show more distributed patterns of activation to emotional vocalizations.

VIII. Functional imaging studies of emotion recognition

In line with some of the patient literature, some functional imaging studies of non-verbal emotional vocalizations have found a specific amygdala response to fearful vocalizations (e.g., Phillips et al., 1998). In this study, a range of different brain areas was also activated to the vocal expressions of fear, in addition to the amygdala activation. This study showed insula responses to facial expressions of disgust, but not to vocal expressions of disgust. In other studies, amygdala activation has been seen to a range of emotional vocalizations, both positive and negative (Sander et al., 2001, 2003; Fecteau et al., 2007). It is known from functional imaging studies that brain regions other than the amygdala are activated by fearful sounds, and not only fearful sounds activate the amygdala. This variation may reflect, in part, the range of emotional vocalizations used in the study (e.g., Phillips et al., 1998 only investigated fear and disgust). Functional imaging methods see all the brain areas recruited by a task or perceptual processes, and it can be difficult to distinguish between brain areas that are somewhat peripheral to the task or process of interest from those that are critical. Insula activation has been seen to both positive and negative non-verbal emotional expressions, as well as to disgust, e.g., laughter and crying vocalizations (Morris et al., 1999; Sander et al., 2003), which may reflect its role in specific emotions, or may reflect the more general role of the anterior insula in the representation of vocalizations (Wise et al., 1999).

IX. Dimensional accounts of emotion and functional imaging studies

As mentioned in the introduction, in contrast to the basic emotions approach to the nature of emotional processing, there are theories which posit that emotional states are represented as points in emotional space, with two (or more) coordinates. Thus, in the circumplex model of emotion (Russell, 1980), different emotions are plotted along the dimensions of valence and arousal. In vocal emotion processing, it has been suggested that, rather than specific emotions, "speech acoustics provide an external cue to the level of non-specific arousal associated with emotional processes" (Bachorowski, 1999, p. 55). Consistent with this, patients with Parkinson's disease, a degenerative disease of the basal ganglia, have been reported to show a general problem with prosody (Lloyd, 1999). Recent work has shown, however, that the Parkinson's patients remain sensitive to the intensity in emotional speech, even though they have deficits in recognizing the emotion conveyed, or in rating the valence of the

Author's personal copy

emotional speech (Dara et al., 2007). This finding suggests that a single dimension in Russell's circumplex model, arousal, might be preserved while another (valence) could be selectively damaged. An alternative interpretation could be that quite simple properties of emotional speech, such as intensity, map onto the perceived emotional intensity – louder sounds are more aroused, across all emotion categories (Banse and Scherer, 1996). This interpretation might mean that it is easier to establish arousal than valence from a simple appraisal of the stimuli, in the presence of brain damage, allowing the patients to make accurate judgments of emotional intensity.

In a recent functional imaging study of non-verbal vocal expressions of emotion, we used fMRI to present people with non-verbal vocal expressions of two negative emotion categories (disgust and fear) and two positive emotion categories (amusement and triumph) (Warren et al., 2006b). These four emotional categories were chosen since they were all highly recognizable and also mapped clearly along the dimension of valence: the amusement and triumph sounds were rated as highly positive, and the fear and disgust sounds as highly negative (Sauter, 2006). We also included a condition in which participants were prompted to make orofacial movements, by instructing them to smile. This permitted an analysis of brain areas involved in both the perception of emotional vocalizations and the brain areas involved in orofacial movement. We also included a control condition of spectrally-rotated versions of the emotional stimuli, which we used to control for the brain response to low-level acoustic properties of the emotional sounds.

Comparing all the emotional categories with the acoustic control condition revealed activity in the left and right STS, the left insula and pSMA (Sauter, 2006). Comparing individual emotion categories with the control condition revealed activation in the bilateral anterior STS for every emotion. In addition to this, amusement showed activation in the inferior frontal gyrus, fear showed extensive activation of the bilateral anterior insula, and triumph showed left insula activation, pSMA and left and right premotor activation.

Further to the simple contrast of all the different emotion categories with the acoustic baseline described above, a contrast of brain areas that varied significantly across emotion conditions (relative to the acoustic control condition) was run, to identify any specific cortical responses to the kinds of emotional information in the sounds. This contrast identified significant patterns in activation across the superior and inferior temporal lobes, the precentral gyrus and prefrontal cortex, limbic and mesial temporal cortex, including the amygdala, hippocampus and basal ganglia. This contrast was used to mask a contrast of the areas which were activated by orofacial movement (smiling to command), revealing activation in the precentral gyrus, pSMA and insula, and in the temporal and occipital lobes, which was affected both by the emotional sounds subjects heard and when they (silently) smiled.

Within these areas co-activated by production and perception of emotional vocalizations, we identified regions that correlated with the valence and arousal ratings for the stimuli (shown in Fig. 3) which had been collected with a different set of subjects in a previous behavioral experiment. In terms of valence, the triumph sounds were rated most positive, followed by amusement, then fear and disgust (Sauter, 2006). The correlation with valence alone was seen in the right inferior frontal gyrus. In terms of arousal, the categories of triumph and fear were rated as the most aroused, followed by amusement and disgust, and this

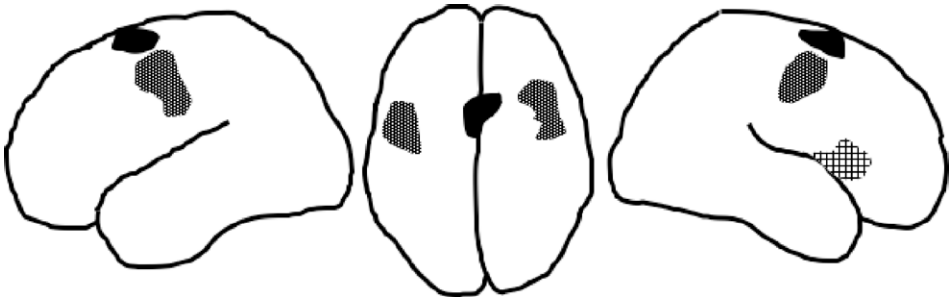


Fig. 3. Brain regions co-activated by orofacial movements and by emotional vocalizations (adapted from Warren et al., 2006b). Key: pre-supplementary motor areas (solid black); lateral premotor (black dots); inferior frontal gyrus (hatching). Activity in regions pre-supplementary motor areas correlate with the perceived arousal of the non-verbal emotional vocalizations: Activity in right inferior frontal gyrus correlates with perceived valence, and activity in bilateral premotor cortex correlates with both valence and arousal scores.

Author's personal copy

led to activation of the pSMA. This result provides evidence that emotional vocalizations of high arousal therefore engage a region involved in higher-order aspects of complex motor control, which may link to a specific preparation for responsive action.

The left and right precentral gyri were co-activated by increases in both valence and arousal. The precentral gyri are somatotopically organized, and the peaks of activation were greatest in the face motor area (Buccino et al., 2001; Carr et al., 2003; Leslie et al., 2004). The activation also extended into more ventral regions associated with motor control of the articulators (Murphy et al., 1997; Blank et al., 2002; Wilson et al., 2004). This activation extended into primary motor regions bilaterally, even though the subjects did not actually move their faces during the scanning session.

With the proviso that only four different emotional categories were tested, this finding is a striking demonstration that motor regions associated with producing facial expressions are activated by hearing emotional vocalizations, and that different regions within this system correlate with the arousal and valence of the stimuli. The perception of vocal expressions of positive or arousing emotions automatically engages preparation for responsive orofacial gestures. We tend to express emotions like cheering and laughing in groups, and laughter has been shown to be a very contagious behavior (Provine, 2000) – when we hear someone laughing, we are primed to join in, or at very least to smile. The finding of increased premotor cortex activation to the sounds of laughter (and cheering) would provide a mechanism for this link.

The finding that the dimensions of valence and arousal have specific neural bases within frontal areas associated with the perception of emotional vocalizations and the production of orofacial movements may provide evidence in support of the circumplex model of emotion (Russell, 1980). The circumplex model of emotion posits that different emotional states are processed and represented as points in an emotional space, along the dimensions of valence and arousal. The current data do not provide full support for this model, however, since no negative correlations were found. There were no increases in cortical activity with lower ratings of valence, despite the fact that both the fear and disgust stimuli were rated as highly negative in valence. Likewise, there were no increases in cortical activity with lower ratings of arousal. It seems highly unlikely that the unpleasantness of disgust and fear are encoded cortically simply as a lack of activation. Further work with a wider palate of emotional states may be the way to address this point and identify whether convincing correlations with more negative valence or decreased arousal can be determined, as would be predicted by dimensional accounts of emotion representations. Different imaging techniques will also facilitate the development of this work – e.g., Sauter and Eimer (in press) have shown that the timing of electrophysiological responses to non-verbal emotional vocalizations varies with the perceived arousal of the expressions.

X. Role of mirror systems in perception of emotional vocalizations

The finding that brain areas associated with motor control are activated by both emotion perception and orofacial gestures has some congruency with theories about the involvement of mirror systems in human communication (Warren et al., 2006b). Several recent studies have suggested that the premotor cortex is co-activated by produced and perceived speech (Wilson et al., 2004), and the results of Warren et al. (2006a,b) extend this into non-verbal vocalizations and orofacial gestures. Indeed, these results go somewhat beyond the results from speech perception studies, since the current investigation revealed activation relative to complex acoustic control conditions, unlike the speech studies (Scott et al., 2009).

Does the involvement of the "mirror" system in the perception of emotional vocalizations and production of orofacial movements suggest that the recognition of emotional vocalizations entails the imitation of the emotions in question? We would argue not, for several reasons. First, the mirror system activations seen in Warren and colleagues study (Warren et al., 2006b) are not likely to reflect the *emotional* contagion of the stimuli *per se*, since the disgust vocalizations led to the weakest activations, despite disgust being an exceedingly contagious emotion (Nemeroff and Rozin, 1994). The results are also unlikely to reflect a mirror system involvement in the recognition of the sounds, since the fear and disgust sounds were very highly recognizable (Sauter, 2006), yet led to lower activation in "mirror" motor areas. Instead of recognition of emotions, or emotional recognition, we argued that these results are related to behavioral responses to the stimuli – either preparation for general action to respond in pSMA (such as readiness to run away for the fear sounds) or preparation for responsive emotional orofacial expressions or vocalizations

Author's personal copy

(such as smiling or laughing) in right IFG and bilateral precentral gyri. It is striking that behavioral mirroring in human communication tends to be in positive contexts – if we like someone, we tend to adopt the same position and gestures when talking to them, and will do so more if we are trying to express affiliation towards them (e.g., Chartrand and Bargh, 1999). A recent review (Scott et al., 2009) has linked motor cortex activation during speech perception to the employment of motor areas in the timing and coordination of communicative cooperative actions, such as turn taking in conversation. In this context, the lateral premotor system responses to more positive and more arousing emotional stimuli reflect the more social nature of the emotions used in this experiment. Laughter and cheering tend to be vocal expressions we produce in groups, and which play an important part in reinforcing positive social bonds between people (Provine, 2000).

Motor cortex in humans has been shown to respond to a variety of different sounds – including music, animal vocalizations and environmental sounds. The evidence for consistent activation of motor cortex by speech (relative to an acoustic control condition) is somewhat weaker (Scott et al., 2009). Thus, it may suggest that unlike the STS, motor areas are less sensitive to the particular acoustic structure of speech – i.e., its reliance on the processing of dynamic, broad spectral structure. In this context, the sensitivity of motor cortex to different emotional vocalizations may reflect the fact that they, like other non-speech sound, rely on the use of envelope, pitch and fine spectral cues for recognition (e.g., environmental noises, Gygi et al., 2004; Sauter, 2006).

XI. Conclusions

A wide range of different neural areas are recruited when people process non-verbal emotional vocalizations. Some are very generally involved in processing social communication, thus hearing emotional sounds activates the bilateral STS, with an emphasis on anterior regions in the right hemisphere. Some other brain regions appear to be specifically involved in the perception of certain emotions across different perceptual channels; damage to brain regions, such as the amygdala and insula, result in deficits in the recognition of some emotions from the voice (e.g., fear and disgust), mirroring deficits in the perception of the same emotions in the face. Functional imaging studies tend to reveal that a network of activity is seen to different vocal emotional expressions, and that some areas (such as the amygdala) are also activated by a range of different vocally-expressed emotions, in addition to those to which they have been linked by the neuropsychological literature. Some brain regions appear to be processing some of the dimensions hypothesized to be important in the conceptual representation of emotions; in frontal areas, brain regions can be identified that are correlated in activity with the perceived arousal (pSMA) or valence (right IFG) of emotional vocalizations, or be sensitive to both arousal and valence (right and left premotor cortex) (Warren et al., 2006a,b). This frontal activity seems to be a specific function of the behavioral correlates of the emotions used in this study, with the positive sounds of laughter and cheering being those commonly performed in groups of other people, and which have a specific role of reinforcing bonds between members of a group. We argue that, generally, mirroring behavior has strong positive affect in human communication – if we like someone, we copy their gestures and movements in conversations with them, and even start to use the same words. The motor system may thus underlie the coordination of behaviors such as conversation and contagious laughter, which are central to the affiliative aspects of human communication.

References

Abercrombie, D., 1967. Elements of General Phonetics. Edinburgh University Press, Edinburgh, UK.

Anderson, A.K., Phelps, E.A., 1998. Intact recognition of vocal expressions of fear following bilateral lesions of the human amygdala. Neuroreport 9 (16), 3607–3613.

Bachorowski, J.A., 1999. Vocal expression and perception of emotion. Curr. Dir. Psychol. Sci. 8, 53–57.

Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. J. Pers. Soc. Psychol. 70, 614–636.

Bänziger, T., Scherer, K.R., 2005. The role of intonation in emotional expressions. Speech Commun. 46, 252–267.

Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14 (16), 2105–2109.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403 (6767), 309–312.

Blank, S.C., Scott, S.K., Murphy, K.M., Warburton, E., Wise, R.J.S., 2002. Speech production: Wernicke, Broca and beyond. Brain 125, 1829–1838.

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., Freund, H.J., 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur. J. Neurosci. 13, 400–404.

Calder, A.J., Young, A.W., 2005. Understanding the recognition of facial identity and facial expression. Nat. Rev. Neurosci. 6 (8), 641–651.

Author's personal copy

Calder, A.J., Keane, J., Manes, F., Antoun, N., Young, A.W., 2000. Impaired recognition and experience of disgust following brain injury. Nat. Neurosci. 3 (11), 1077–1078.

Carr, L., Iacoboni, M., Dubeau, M.C., Mazziotta, J.C., Lenzi, G.L., 2003. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. Proc. Natl. Acad. Sci. USA, 100, 5497–5502.

Chartrand, T.L., Bargh, J.A., 1999. The chameleon effect: the perception-behavior link and social interaction. J. Pers. Soc. Psychol. 76, 893–910.

Cohen, L., Jobert, A., Le Bihan, D., Dehaene, S., 2004. Distinct unimodal and multimodal regions for word processing in the left temporal cortex. Neuroimage 23 (4), 1256–1270.

Dara, C., Monetta, L., Pell, M.D., 2007. Vocal emotion processing in Parkinson's disease: reduced sensitivity to negative emotions. Brain Res. 1188, 100–111.

Ekman, P., 1992a. Are there basic emotions? Psychol. Rev. 99, 550–553.

Ekman, P., 1992b. An argument for basic emotions. Cogn. Emot. 6, 169–200.

Ekman, P., Sorenson, E.R., Friesen, W.V., 1969. Pan-cultural elements in facial displays of emotion. Science 164, 86–88.

Fecteau, S., Belin, P., Joanette, Y., Armony, J.L., 2007. Amygdala responses to nonlinguistic emotional vocalizations. Neuroimage 36 (2), 480–487.

Fried, I., Wilson, C.L., MacDonald, K.A., Behnke, E.J., 1998. Electric current stimulates laughter. Nature 391, 650.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., et al., 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nature Neurosci. 8, 145–146.

Gygi, B., Kidd, G.R., Watson, C.S., 2004. Spectral-temporal factors in the identification of environmental sounds. J. Acoust. Soc. Am. 115 (3), 1252–1265.

Hornak, J., Bramham, J., Rolls, E.T., Morris, R.G., O'Doherty, J., Bullock, P.R., Polkey, C.E., 2003. Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. Brain 126 (7), 1691–1712.

Jürgens, U., 1986. The squirrel monkey as an experimental model in the study of cerebral organization of emotional vocal utterances. Eur. Arch. Psychiat. Neurol. Sci. 236 (1), 40–43.

Karpf, A., 2006. The Human Voice. Bloomsbury, London, UK.

Kohler, K.J., 2008. "Speech-smile," "speech-laugh," "laughter" and their sequencing in dialogic interaction. Phonetica 65, 1–18.

Krolak-Salmon, P., Hénaff, M-A., Vighetto, A., Bauchet, F., Bertrand, O., Mauguière, F., et al., 2006. Experiencing and detecting happiness in humans: the role of the supplementary motor area. Ann. Neurol. 59, 196–199.

Leslie, K.R., Johnson-Frey, S.H., Grafton, S.T., 2004. Functional imaging of face and hand imitation: towards a motor theory of empathy. Neuroimage 21, 601–607.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. Cereb. Cortex 15 (10), 1621–1631.

Lloyd, A.J., 1999. Comprehension of prosody in Parkinson's disease. Cortex 35 (3), 389–402.

Morris, J.S., Scott, S.K., Dolan, R.J., 1999. Saying it with feeling: neural responses to emotional vocalizations. Neuropsychologia 37 (10), 1155–1163.

Murphy, K., Corfield, D.R., Guz, A., Fink, G.R., Wise, R.J., Harrison, J., Adams, L., 1997. Cerebral areas associated with motor control of speech in humans. J. Appl. Physiol. 83, 1438–1447.

Murray, I.R., Arnott, J.L., 1993. Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. J. Acoust. Soc. Am. 93, 1097–1108.

Nemeroff, C., Rozin, P., 1994. The contagion concept in adult thinking in the United States: transmission of germs and of interpersonal influence. Ethos 22, 158–186.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36 (4), 767–776.

Phillips, M.L., Young, A.W., Scott, S.K., Calder, A.J., Andrew, C., Giampietro, V., Williams, S.C.R., Bullmore, E.T., Brammer, M., Gray, J.A., 1998. Neural responses to facial and vocal expressions of fear and disgust. Proc. Royal Soc. B. 265 (1408), 1809–1817.

Provine, R.R., 2000. Laughter: A Scientific Investigation. Faber and Faber, London, UK.

Russell, J.A., 1980. A circumplex model of affect. J. Person. Soc. Psychol. 39, 1161–1178.

Sander, K., Brechmann, A., Scheich, H., 2003. Audition of laughing and crying leads to right amygdala activation in a low-noise fMRI setting. Brain Res. Brain Res. Protoc. 11 (2), 81–91.

Sauter, D., 2006. An investigation into vocal expressions of emotions: the roles of valence, culture, and acoustic factors. PhD thesis, University of London, UK.

Sauter, D.A. and Eimer, M. Rapid detection of emotion from human vocalizations, J. Cogn. Neurosci. (in press).

Sauter, D., Scott, S.K., 2007. More than one kind of happiness: can we recognize vocal expressions of different positive states? Motiv. Emotion 31, 192–199.

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

Schirmer, A., Kotz, S.A., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cogn. Sci. 10, 24–30.

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.

Scott, S.K., Blank, S.C., Rosen, S., Wise, R.J.S., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406.

Scott, S.K., Rosen, S., Lang, H., Wise, R.J.S., 2006. Neural correlates of intelligibility in speech investigated with noise vocoded speech – a Positron Emission Tomography study. J. Acoust. Soc. Am. 120 (2), 1075–1083.

Scott, S.K., McGettigan, C., Eisner, F., 2009. A little more conversation, a little less action: candidate roles for motor cortex in speech perception. Nature Rev. Neurosci. 10 (4), 295–302.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.

Sprengelmeyer, R., Young, A.W., Calder, A.J., Karnat, A., Lange, H., Hömberg, V., Perrett, D.I., Rowland, D.,

Author's personal copy

1996. Loss of disgust. Perception of faces and emotions in Huntington's disease. Brain 119 (5), 1647–1665.

Tartter, V.C., Braun, D., 1994. Hearing smiles and frowns in normal and whisper registers. J. Acoust. Soc. Am. 96 (4), 2101–2107.

Tracy, J.L., Robins, R.W., 2008. The nonverbal expression of pride: evidence for cross-cultural recognition. J. Pers. Soc. Psychol. 94, 516–530.

Warren, J.D., Scott, S.K., Price, C.J., Griffiths, T.D., 2006a. Human brain mechanisms for the early analysis of voices. Neuroimage 31 (3), 1389–1397.

Warren, J.E., Sauter, D.A., Eisner, F., Wiland, J., Dresner, M.A., Wise, R.J., Rosen, S., Scott, S.K., 2006b. Positive emotions preferentially engage an auditory-motor "mirror" system. J. Neurosci. 26 (50), 13067–13075.

Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. Nat. Neurosci. 7, 701–702.

Wise, R.J.S., Greene, J., Büchel, C., Scott, S.K., 1999. Brain systems for word perception and articulation. The Lancet 353 (9158), 1057–1061.