Neuroscience and Biobehavioral Reviews 37 (2013) 24–35

Contents lists available at SciVerse ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: A quantitative meta-analysis

Sascha Frühholz[a,b,*], Didier Grandjean[a,b]

[a] Neuroscience of Emotion and Affective Dynamics (NEAD) Laboratory, Department of Psychology, University of Geneva, 40 bd du Pont d'Arve, CH-1205 Geneva, Switzerland
[b] Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

ARTICLE INFO

Article history:
Received 6 August 2012
Received in revised form 8 October 2012
Accepted 4 November 2012

Keywords:
Vocal expressions
Voice
Prosody
Paraverbal/nonverbal expressions
Valence
Attentional focus

ABSTRACT

Vocal expressions of emotions consistently activate regions in the superior temporal cortex (STC), including regions in the primary and secondary auditory cortex (AC). Studies usually report broadly extended functional activations in response to vocal expressions, with considerable variation in peak locations across several auditory subregions. This might suggest different and distributed functional roles across these subregions instead of a uniform role for the decoding of vocal emotions. We reviewed recent studies and conducted an activation likelihood estimation meta-analysis summarizing recent fMRI and PET studies dealing with the processing of vocal expressions in the STC and AC. We included two stimulus-specific factors (paraverbal/nonverbal expression, stimulus valence) and one task-specific factor (attentional focus) in the analysis. These factors considerably influenced whether functional activity was located in the AC or STC (influence of valence and attentional focus), the laterality of activations (influence of paraverbal/nonverbal expressions), and the anterior-posterior location of STC activity (influence of valence). These data suggest distributed functional roles and a differentiated network of auditory subregions in response to vocal expressions.

Contents

* Corresponding author at: Neuroscience of Emotion and Affective Dynamics (NEAD) Laboratory, Department of Psychology, University of Geneva, 40 bd du Pont d'Arve, CH-1205 Geneve, Switzerland. Tel.: +41 22 379 9455; fax: +41 22 379 9219.
E-mail addresses: fruehholz@gmail.com, sascha.fruehholz@unige.ch (S. Frühholz).

0149-7634/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neubiorev.2012.11.002

1. Introduction

Emotional states can considerably influence the tone of a voice by modifying the different components of the vocal tract during vocal expressions, from which individuals infer emotional states in other persons (Banse and Scherer, 1996). Therefore, emotional features of vocal intonations are important for social interactions. A prioritized decoding of these emotional features by the human brain should allow fast and adaptive behavior in accordance with the social context. Regions in the bilateral superior temporal cortex (STC) of the human brain comprising the superior temporal gyrus (STG) and the middle temporal gyrus (MTG), as well as the superior temporal sulcus (STS), are generally sensitive to human voices compared with other environmental sounds (Belin et al., 2000; Schirmer et al., 2012). Furthermore, subregions within this voice-sensitive cortex are also more responsive to the emotional than to the neutral tone of a voice (i.e. Bach et al., 2008; Ethofer et al., 2009b; 2012; Grandjean et al., 2005; Sander et al., 2005). These emotion-sensitive areas are typically located in the mid-posterior STC. STC activity in response to vocal expressions is a consistent finding across many studies (Bach et al., 2008; Frühholz et al., 2012; Grandjean et al., 2005; Sander et al., 2005), leading to the proposal of a general "emotional voice area" (EVA) located in the STC (Ethofer et al., 2012). The EVA might have a general functional role for the decoding of emotional cues from vocal expressions.

Thought to be located in the posterior STC (pSTC), the EVA not only comprises areas in the higher-level auditory cortex (AC), but also areas in the primary and secondary AC, especially in the right hemisphere (Ethofer et al., 2012). Several studies have demonstrated this broad extension of emotion-sensitive cortical regions (Beaucousin et al., 2007; Leitman et al., 2010b; Sander et al., 2005; Wildgruber et al., 2004), which covers regions across different levels of auditory processing. This finding might not suggest a unique and general role of the emotion-sensitive STC and AC, but rather a distribution of functional roles across these different areas. Lower-level auditory cortices might decode acoustic properties of vocal expression, whereas higher-level auditory regions might subserve a sensory-integrative role, such as building up percepts of vocal expression based on pre-processed auditory features and representing an acoustic object (Schirmer and Kotz, 2006). This assumption of a functional distribution is also indicated across studies that show a considerable variation of peak activations in response to vocal emotional expressions. These studies report peak activations for emotional compared with neutral vocal expressions that range from the bilateral pSTC (Beaucousin et al., 2007; Ethofer et al., 2006a, 2009a; Mitchell et al., 2003; Sander et al., 2005; Wildgruber et al., 2005) and bilateral mid-STC (mSTC; Bach et al., 2008; Ethofer et al., 2006b, 2007, 2009a; Grandjean et al., 2005; Leitman et al., 2010b; Sander et al., 2005; Wiethoff et al., 2008) to the right anterior STG (aSTC; Bach et al., 2008). Studies also report peak activations in the primary and secondary AC in response to vocal expressions (Fecteau et al., 2007; Leitman et al., 2010b; Meyer et al., 2005; Phillips et al., 1998; Sander et al., 2005; Warren et al., 2006a; Wildgruber et al., 2005).

Together, the broad extension of emotion-sensitive areas in the STC and AC and the variation in peak activity might suggest different and distributed functional roles rather than a unique functional role for all of these regions. Specifically, the activity of these different regions and their functional roles might depend on several stimulus- as well as task-specific modalities, which figure as bottom-up and top-down factors, respectively, during the processing of vocal expression. With respect to top-down factors, one frequently investigated factor during the processing of vocal expressions is attentional focus, sometimes referred to as the level of processing of vocal expressions (Bach et al., 2008; Grandjean et al., 2005; Sander et al., 2005). The emotional cue of a voice can be presented inside the focus of attention (explicit processing of emotion) when participants are asked to explicitly discriminate, classify, or evaluate the emotional tone of a voice (Alba-Ferrara et al., 2011; Beaucousin et al., 2007; Ethofer et al., 2006a, 2009a; Frühholz et al., 2012; Grandjean et al., 2005; Johnstone et al., 2006; Kotz et al., 2003; Leitman et al., 2010b; Mitchell, 2006; Sander et al., 2005; Szameitat et al., 2010; Wildgruber et al., 2005; Wittfoth et al., 2010). This emotional cue can also be presented outside the focus of attention (implicit processing of emotion) when participants are, for example, asked to discriminate the gender of one voice from another (Alba-Ferrara et al., 2011; Bach et al., 2008; Ethofer et al., 2012; Fecteau et al., 2007; Frühholz et al., 2012; Morris et al., 1999; Mothes-Lasch et al., 2011) or to make a linguistic decision about a nonemotional linguistic feature (Buchanan et al., 2000; Ethofer et al., 2006a; Mitchell et al., 2003), or when emotional voices are presented outside the spatial focus of attention (Grandjean et al., 2005; Sander et al., 2005). For the latter, Grandjean and colleagues (2005), for example, have shown that a region in the right pSTC is sensitive to vocal expressions independent of attentional focus, whereas an adjacent region in the right mSTC is responsive to vocal expression only when the emotional cue is inside the focus of attention. We recently replicated this finding by showing that the right mSTC was active only when the emotional cue was inside the focus of attention, whereas regions in the right pSTC were generally active during the processing of vocal emotional expressions (Frühholz et al., 2012). Although these studies show a dependency of STC activity in response to vocal expressions, the dependency of primary or secondary AC activity is not yet known or has not been investigated systematically. There are, however, reasons to believe that even low-level AC activity might depend on attentional focus toward or away from vocal expressions. Studies usually report primary and secondary AC activity specifically during implicit attention toward vocal expressions (Fecteau et al., 2007; Phillips et al., 1998), but sometimes also during explicit attention (Leitman et al., 2010b).

Besides this task dependency of STC activity in response to vocal expressions as a top-down factor, STC responses to vocal expressions might also depend on stimulus-specific features as bottom-up factors. We recently demonstrated a dependency of STC activity on acoustic features such as the pitch and intensity of a voice during vocal emotional expressions (Frühholz et al., 2012). We found less sensitivity to acoustic features in mSTC compared with pSTC. Besides basic acoustic features, two more important stimulus-specific features of emotional voices might be present. The first feature concerns the emotional dimension of vocal expressions, represented by the factors *valence* and *arousal*, which determine the pleasantness and the excitement of perceived vocal emotions, respectively. The effects of arousal have seldom been investigated, but the few existing results point to a bilateral STC sensitivity to the level of arousal in vocal emotions (Warren et al., 2006b; Wiethoff et al., 2008). With respect to valence, vocal expression can be neutral or of positive (i.e. happy) or negative valence (i.e. anger, fear, disgust). Like the effects of arousal, the question of how vocal expressions of different valence elicit activity in the STC has been rarely investigated. Ethofer et al. (2009b) report different voxel patterns in the bilateral voice-sensitive cortex in the STC, including higher-level auditory regions in the STG and lower-level auditory regions in the transverse temporal gyrus in response to different vocal expressions, without providing a clear picture about the spatial distribution of activation maps across different vocal expressions. Other studies report bilateral STC activity elicited by both positive and negative vocal expressions (Fecteau et al., 2007; Frühholz et al., 2012; Grandjean et al., 2005; Johnstone et al., 2006; Kotz et al., 2003; Leitman et al., 2010b; Meyer et al., 2005; Mothes-Lasch et al., 2011; Phillips et al., 1998; Sander et al., 2005; Warren et al., 2006a), but only in the primary and secondary AC in response to a negative expression (Fecteau et al., 2007; Kotz et al., 2003; Leitman et al., 2010b; Phillips et al., 1998; Sander et al., 2005). However, positive expressions can also sometimes elicit primary and secondary AC activity, although surprisingly only for positive nonverbal expressions rather than for vocal expressions superimposed in speech utterances (Meyer et al., 2005; Szameitat et al., 2010).

The last finding is related to the second important stimulus-specific factor, the *medium* of vocal expressions, which covers both nonverbal expressions and emotional intonations superimposed on speech and speech-like verbal utterances. For the latter, the emotional tone of voice can modulate verbal speech utterances, including both neutral and emotional semantics (Beaucousin et al., 2007; Ethofer et al., 2006a; Leitman et al., 2010b), but it can also modulate vocal utterances based on pseudo language as well as on interjections that sound like speech but are devoid of any semantic meaning (Bach et al., 2008; Dietrich et al., 2008; Frühholz et al., 2012; Grandjean et al., 2005; Sander et al., 2005). These speech-based intonations are usually referred to as emotional speech prosody and generally represent a paraverbal emotional cue. We therefore subsume these cases under the category of paraverbal expressions. Unlike these speech or speech-like paraverbal expression of emotions, vocal expression can also be the nonverbal expression of emotional states (Dietrich et al., 2007; Fecteau et al., 2007; Meyer et al., 2005; Morris et al., 1999; Phillips et al., 1998; Szameitat et al., 2010; Warren et al., 2006a). This issue of the medium as represented by the distinction between paraverbal and nonverbal expressions seems important in terms of lower- and higher-level activity in the AC. The type of medium can influence the acoustic richness and diversity of acoustic and emotional cues and can change the way we extract the emotional value from these different types of vocal emotional expressions. Nonverbal expressions seem to be decoded according to a limited range of acoustic features, such as the pitch or the fundamental frequency (f0), the spectral center of gravity, and the number of amplitude onsets (Sauter et al., 2010). Paraverbal expressions are classified and discriminated according to several features of the f0, the intensity, and the spectral and temporal properties (Banse and Scherer, 1996; Juslin and Laukka, 2003; Leitman et al., 2010a; Patel et al., 2011). The diversity of acoustic features from which the emotional tone of a paraverbal expression is decoded and discriminated might involve stronger primary and secondary AC activity. Another factor concerning paraverbal and nonverbal expressions might be the laterality of STC activity, with the assumption that paraverbal expressions might more strongly activate the left hemisphere, while nonverbal mediums might more strongly activate the right hemisphere. The only recent study provides mixed and inconsistent evidence for this assumption, specifically in relation to STC activity (Kotz et al., 2003).

Taken together, these findings suggest that, instead of a unique emotion-sensitive area in the STC and AC with a general role during the processing of vocal expressions, several subregions might be present with different functional roles during the decoding of vocal expressions. The location of functional activities, as well as the functional role of these subregions, might depend on top-down and bottom-up factors during the processing of vocal expressions. Although recent studies on vocal expressions provide evidence for a functional distribution, this issue has not yet been investigated systematically. The current meta-analysis therefore aimed at determining the dependency and distribution of functional activity in the STC and AC in response to vocal expressions based on the *attentional focus* as a top-down factor, as well as on stimulus *valence* (negative, positive) and stimulus *medium* (paraverbal expressions, nonverbal expressions) as bottom-up factors, during the processing of vocal emotions. We did not include stimulus *arousal* as another factor in this meta-analysis because of a very small amount of activation foci reported in previous experiments (see Section 2.3). We collected peak activations across 29 functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) studies, including 128 peak activations that dealt with the processing of vocal emotional expressions and their influence on bilateral STC activity.

2. Materials and methods

2.1. Literature search and data collection

To find both fMRI and PET studies dealing with vocal expressions of emotions, we thoroughly searched the BrainMap database using Sleuth (http://brainmap.org/sleuth/; see Laird et al., 2005), as well as PubMed (www.pubmed.org), using the following search strings: “vocal AND (emotions OR expressions)”, “(emotional OR affective OR emotion) AND (prosody OR prosodic OR speech OR voices OR intonation OR vocalization)”, “speech AND (melody OR prosody OR intonation)”, “(anger OR fear OR disgust OR sad OR happy OR arousal) AND prosody”, “auditory emotions.” We included only those studies that were published in English before April 2012, included only healthy young and middle-aged human adults, and used auditory stimulation and a scanning space covering at minimum the whole bilateral STC and AC in a full posterior-to-anterior direction. Studies were excluded that were based solely on predefined regions of interest. Peak locations had to be reported in standard stereotactic coordinate space (Talairach or Montreal Neurological Institute [MNI] space). We were specifically interested in activation foci in the STC, including the lateral STG and the STS and the superior portion of the MTG, as well as in the entire plane of the STC hidden in the lateral sulcus. The latter regions included the transverse temporal gyrus, as well as the planum temporale (PTe; mostly covered by Brodmann area [BA] 42) and the planum polare (PPo; mostly covered by BA 52). The latter regions represent primary and secondary AC (see Fig. 1 for a definition of all of these regions).

2.2. Data set

This meta-analysis dealt with the processing of vocal expression in the human STC and AC. We therefore excluded all studies that reported the processing of vocal expressions in primate and nonprimate animals, as well as studies that used animal vocalizations in human participants. We also excluded studies or peak activations in studies that compared vocal expressions only against baseline activity instead of comparing them against other emotional or neutral vocal expressions. This was done to ensure that functional activity was specific for the emotional dimension in vocal expression. The final data set included 27 independent studies with a total of 489 participants reporting 128 activation foci (see Table 1). Please note that Table 1 reports 29 studies; however, in two cases, we combined the activation foci of two studies into a single study because the activation foci were acquired with the same sample of participants. All coordinates were taken in or to MNI space. Peak activations reported in Talairach space were transformed to MNI space using a linear transformation (Lancaster et al., 2007).

2.3. Data analysis

First, we entered all 128 activation foci derived from all studies included in this meta-analysis in a general analysis to find activation likelihood estimation (ALE) maps that provided an estimation of the spatial extent of emotion-sensitive regions in the STC and AC, independent of stimulus- and task-specific factors. In a second approach, peak activations were categorized according to the three main factors of interest in this meta-analysis (Table 2). First, all peak

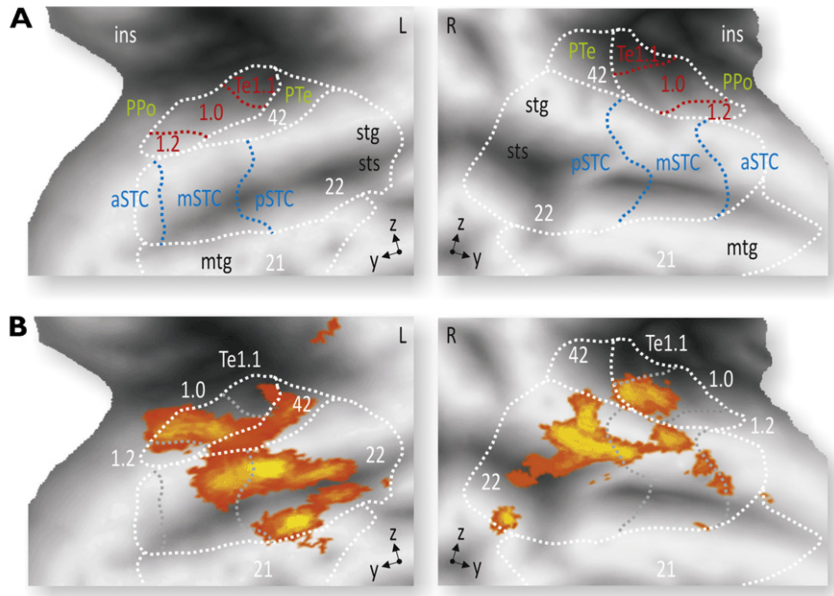

Fig. 1. (A) The STC and the AC as the main regions of interest. The AC is distinguished into primary AC (red) and secondary AC (green). The primary AC is represented by the subregions Te1.0, Te1.1, and Te1.2, which are predominantly located on the transverse temporal gyrus. These AC subregions were defined according to Morosan et al. (2001), as implemented in the Anatomy toolbox (Eickhoff et al., 2005) for SPM8. The secondary AC is located anterior and posterior to the primary AC and is represented by the planum polare (PPo) and the planum temporale (PTe; BA 42), respectively. BA 42 was defined according to its location relative to the primary auditory cortex and higher-level auditory regions in BA 22 (see Hackett, 2011). The STC is mainly represented by BA 21 and 22, which were defined according to the CARET atlas (Van Essen et al., 2001). The white dotted lines denote BA areas. Along the anterior-to-posterior direction, we defined the subregions of aSTC ($y > 10$), mSTC ($10 < y < -15$), and pSTC ($y < -15$). (B) Emotional vocal expressions generally elicited extended activities in the STC independent of stimulus modality, task, and emotional valence. Activity is rendered on the left (L) and right (R) cortical flat surface of the human Colin atlas implemented in CARET software (Van Essen et al., 2001). Abbreviations: *ins* insula, *mtg* middle temporal gyrus, *stg* superior temporal gyrus, *sts* superior temporal sulcus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

activations were divided according to their activity in response to "paraverbal" or "nonverbal" vocal expressions (factor *stimulus medium*). Paraverbal expressions comprised vocal expression that used speech based on pseudo language, as well as semantic neutral or emotional speech. All other types of vocal expressions were classified as nonverbal vocal expressions. Second, peak activations were classified as being elicited by "positive" or "negative" expressions (factor *valence*). Positive expressions comprised expressions of happiness or laughter. Negative expressions comprised expressions of anger, fear, or disgust. Only those peak activations were included that could be uniquely categorized in one of these categories. We did not include peak activations that were reported to be a combined response to both positive and negative expressions.

Unlike the studies that reported foci classification according to valence as one emotional feature of vocal expressions, only two studies (Warren et al., 2006b; Wiethoff et al., 2008) reported three different activation foci for arousal as the other emotional feature. Only one left and one right STC focus were located within a proper brain mask. These two foci did not allow a valid ALE map estimation of the arousal effects of vocal emotions, and thus the factor *arousal* was not included in this meta-analysis.

Finally, we separated peak activations according to the attentional focus (factor *attention*). Peak activations in the category "explicit attention" were taken from studies that asked participants to directly focus on the emotional cue of a voice and to explicitly categorize vocal expression according to their valence. Peak activations during "implicit attention" were taken from studies in which participants were asked to focus on a nonemotional stimulus feature, such as the gender of a voice or a nonemotional linguistic feature of spoken speech, or when emotional expression was presented outside the spatial focus of attention. Studies that reported peak activation during passive listening or in which activations were reported to be combined for explicit and implicit attention tasks were not taken into account here. ALE maps were calculated for each of these six categories (paraverbal/nonverbal, negative/positive, explicit/implicit attention). ALE maps were also calculated for all three subcategories of verbal expression (pseudo-language-based speech, semantic neutral speech, semantic emotional speech), of negative expressions (anger, fear, disgust), and of the implicit attention condition (focus on gender, focus on linguistic feature, spatial manipulation). However, the resulting ALE maps for fear and disgust have to be considered with caution because only one study (Phillips et al., 1998) reported a few peak activations specifically for the expressions of fear ($n = 5$) and disgust ($n = 10$).

ALE maps were estimated using the algorithm implemented in the GingerALE toolbox (http://brainmap.org), allowing a coordinate-based meta-analysis of neuroimaging results reported as activation foci (Eickhoff et al., 2009). These activation foci are modeled as three-dimensional Gaussian distributions represented by modeled activation (MA) maps of activation likelihoods. The number of subjects in each study is taken into account to estimate the spatial uncertainty of each activation focus. Subsequently, ALE maps were generated by combining the MA maps of all studies into one category. ALE maps represent the likelihood of each voxel to be activated across the studies entered into the ALE algorithm. These ALE maps are compared against a null distribution of the random spatial association of the MA maps of each study, as revealed by a permutation procedure. The original ALE maps were then tested against the ALE maps resulting from the null distribution. The resulting statistical maps were each thresholded at a false discovery rate (FDR) corrected p-value of <0.05 and a cluster volume of 20 mm^3. This relatively low cluster volume threshold was chosen so that we would be especially able to identify subregions in the STC and AC that might be small compared with the broad extent of the general emotion-sensitive cortex.

Table 1
Studies included in this meta-analysis with the number of subjects (*N*), stimulus material, imaging method used, number of activation foci, and contrasts from which foci were derived.

Authors	*N*	Stimulus	Method	Foci	Contrasts
Alba-Ferrara et al. (2011)	19	Neutral semantic speech	fMRI	4	Emotional vs. neutral voices[c]; complex emotions vs. simple emotions
Bach et al. (2008)	16	Pseudo-language-based speech	fMRI	5	Emotional vs. neutral voices[c]; implicit (gender[d]) vs. explicit (valence[e])
Beaucousin et al. (2007)	23	Emotional semantic speech	fMRI	3	Explicit (valence[e]) vs. implicit (linguistic[f]); human emotional voices vs. synthetic emotional voices
Buchanan et al. (2000)	10	Neutral semantic speech	fMRI	1	Implicit (linguistic[f]) vs. explicit (valence[e])
Dietrich et al. (2007)	16	Nonverbal expressions	fMRI	1	Affective bursts vs. vocal gestures
Dietrich et al. (2008)	16	Pseudo-language-based speech	fMRI	2	Emotional vs. neutral voices[c]
Ethofer et al. (2006a)[a]	24	Emotional semantic speech	fMRI	2	Implicit (linguistic[f]) vs. explicit (valence[e]); explicit (valence[e]) vs. implicit (linguistic[f])
Ethofer et al. (2006b)[a]	24	Emotional semantic speech	fMRI	2	Increasing emotional prosodic intensity
Ethofer et al. (2007)	20	Neutral semantic speech	fMRI	4	Emotional vs. neutral voices[c]
Ethofer et al. (2009a)	24	Neutral semantic speech	fMRI	3	Emotional vs. neutral voices[c]; explicit (valence[e]) vs. implicit (linguistic[f])
Ethofer et al. (2012)	22	Pseudo-language-based speech	fMRI	2	Emotional vs. neutral voices[c]
Fecteau et al. (2007)	14	Nonverbal expressions	fMRI	7	Emotional vs. neutral voices[c]; positive vs. neutral voices; negative vs. neutral voices; negative vs. positive voices
Frühholz et al. (2012)	17	Pseudo-language-based speech	fMRI	6	Anger vs. neutral voices; explicit (valence[e]): anger vs. neutral; implicit (gender[d]): anger vs. neutral
Grandjean et al. (2005)[b]	15	Pseudo-language-based speech	fMRI	3	Anger vs. neutral voices; explicit (spatial[g]) vs. implicit (no spatial)
Johnstone et al. (2006)	40	Neutral semantic speech	fMRI	3	Happy vs. angry voices
Kotz et al. (2003)	12	Neutral semantic speech	fMRI	5	Normal speech: positive vs. negative voices; normal speech: negative vs. neutral voices; prosodic (delexicalized) speech: positive vs. neutral voices; prosodic (delexicalized) speech: negative vs. neutral voices
Leitman et al. (2010b)	20	Neutral semantic speech	fMRI	17	Emotional vs. neutral voices[c]; emotional voices and acoustic cue interaction; anger voice and acoustic cue interaction; happy + fear voices and cue interaction
Meyer et al. (2005)	12	Nonverbal expressions	fMRI	2	Laughter vs. sound
Mitchell et al. (2003)	13	Emotional semantic speech	fMRI	2	Emotional vs. neutral voices[c]; implicit (linguistic[f]) vs. explicit (valence[e])
Mitchell (2006)	28	Emotional semantic speech	fMRI	6	Prosody with congruent; semantic vs. prosody-only; prosody with incongruent; semantic vs. prosody-only
Morris et al. (1999)	6	Nonverbal expressions	PET	1	Emotional vs. neutral voices[c]
Mothes-Lasch et al. (2011)	24	Neutral semantic speech	fMRI	2	Anger vs. neutral voices
Phillips et al. (1998)	6	Nonverbal expressions	fMRI	15	Fearful vs. mild happy voices; disgust vs. mild happy voices; fearful vs. disgust voices; disgust vs. fearful voices
Sander et al. (2005)[b]	15	Pseudo-language-based speech	fMRI	8	Anger vs. neutral voices; anger (spatial[g]) vs. neutral voices; anger (no spatial) vs. neutral voices; anger (spatial[g]) vs. anger (no spatial)
Szameitat et al. (2010)	18	Nonverbal expressions	fMRI	6	Tickle vs. joy; tickle vs. taunt; tickle vs. (taunt + joy)
Warren et al. (2006a,b)	20	Nonverbal expressions	fMRI	4	F-contrast between emotions; correlation with positive valence; correlation with arousal
Wiethoff et al. (2008)	24	Neutral semantic speech	fMRI	9	Emotional vs. neutral voices[c]; correlation with intensity, f0, duration, arousal
Wildgruber et al. (2005)	10	Neutral semantic speech	fMRI	1	Explicit (valence[e]) vs. implicit (linguistic[f])
Wittfoth et al. (2010)	20	Emotional semantic speech	fMRI	2	Emotional vs. neutral voices[c]

[a] The activation foci of these two studies were combined so as to represent a single study because the results reported in these two studies were acquired with the same sample of participants.
[b] Same as above.
[c] The contrast "emotional vs. neutral voices" includes the comparison of emotional expression of different valence with neutral expressions. Studies differ in the amount and the kind of emotional expression, which were entered in this contrast.
[d] "Gender" indicates an implicit gender discrimination task.
[e] "Valence" indicates an explicit valence discrimination task.
[f] "Linguistic" indicates an implicit linguistic discrimination task.
[g] "Spatial" indicates an explicit processing of emotional voices by directing attention to the spatial location of these voices (operationalized as attentional focus to the left or right ear). "No spatial" means direction of attention away from the spatial location of emotional voices.

Resulting ALE maps were rendered on a cortical flat surface of the human Colin atlas implemented in CARET software (Van Essen et al., 2001) to allow the exact cortical representation in relation to the boundaries of auditory-related BA 21, 22, 41, and 42, as well as according to the primary AC subregions Te1.0, Te1.1, and Te1.2, as defined by (Economo and Koskinas, 1925). The latter three regions of interest were created according to (Morosan et al., 2001), as implemented in the Anatomy toolbox (Eickhoff et al., 2005) for the SPM software (Version 8; Welcome Department of Cognitive Neurology, London, UK). BA 42 was defined according to its relative location to primary AC and BA 22 (see Hackett, 2011). Finally, we subdivided STC activity in an anterior-to-posterior direction at approximate *y*-coordinate boundaries in the anterior STC (aSTC) ($y > 0$), mSTC ($-20 > y > 0$), and pSTC ($y > -20$) (see also Frühholz et al., 2012; Grandjean et al., 2005).

3. Results

In a first analysis, we were interested in general activation in the STC elicited by emotional vocal expressions irrespective of stimulus medium, stimulus valence, and attentional task condition. Emotional vocal expressions generally elicited extended activity in the bilateral temporal cortex (Fig. 1; Table 3). This activity included regions in the bilateral primary AC in BA 41, including subregions in the left Te1.1 and Te1.0 and in the right Te1.0. In the left hemisphere, this primary AC activity extended rostrally (BA 52) and caudally (BA 42) into the secondary AC. Additional general activity in response to vocal expressions was revealed in the bilateral STC (BAs 21, 22, 38). This indicates a widespread network of STC and AC regions that are sensitive to emotional vocal expressions. However, subregions in this network were differentially sensitive

Table 2
Details about the study and peak voxel classification of all studies included in the final meta-analysis according to (A) stimulus medium, (B) stimulus valence, and (C) attentional focus. The table reports number of experiments, number of participants, and number of activation foci.

		Experiments	Participants	Foci
(A) Stimulus medium				
Nonverbal		6	78	19
Paraverbal	Pseudo language	5	86	26
	Neutral semantic speech	10	207	49
	Emotional semantic speech	6	132	17
(B) Stimulus valence				
Positive		6	116	11
Negative	Anger	4	76	19
	Fear	1	6	5
	Disgust	1	6	10
(C) Attentional focus				
Explicit		14	286	50
Implicit	Gender	8	117	33
	Linguistic	3	47	3
	Spatial	1	15	5

to vocal expressions, depending on (a) the medium of vocal expression (paraverbal, nonverbal); (b) the valence of vocal expressions (positive, negative), and (c) the attentional focus toward or away from vocal expressions (explicit attention, implicit attention).

3.1. *Paraverbal and nonverbal emotional vocal expressions*

ALE maps revealed bilateral activity in the higher-level AC (BA 21, 22) for paraverbal expressions located in the mSTC and pSTC, as well as for nonverbal vocal expressions in the right pSTC (Fig. 2A; Table 4A and B). Nonverbal expressions additionally activated the bilateral primary AC (left Te1.1, right Te1.0), and paraverbal expressions additionally activated the left primary (Te1.0) and secondary AC (BA 42). Since paraverbal expressions include different types of verbal mediums, we generated ALE maps separately for vocal emotions based on pseudo language, for neutral semantic speech, and for emotional semantic speech (Fig. 2B, Table 4C and E). Whereas all of these paraverbal expressions revealed bilateral activity in the STG and the MTG in the mSTC and pSTC (BA 21, 22), only neutral (Te1.0) and pseudo-language-based speech (Te1.1) revealed activity in the left primary AC.

3.2. *Positive and negative vocal expressions*

ALE maps for negative vocal expressions revealed bilateral primary AC activity (Te1.0; extending in the left hemisphere into BA 52), secondary AC activity (BA 42), and activity in the right mSTC and left pSTC (BA 22), whereas positive expressions elicited activity only in the bilateral left mSTC and in the right aSTC and pSTC (BA 21, 22, 38) (Fig. 2C; Table 5A and B). Negative vocal expressions in the studies included in this meta-analysis mainly comprised expressions of anger, disgust, or fear, for which we generated ALE maps separately (Fig. 2D; Table 5C and E). Angry voices revealed bilateral activity in the mSTC and pSTC (BA 22) and in the left secondary AC (BA 22/52). Fearful expressions revealed activity in the left secondary AC (BA 42) and the right mSTC (BA 22). Expressions of disgust revealed activity in the right primary (Te1.0/Te1.1) and secondary AC (BA 42), as well as in the left mSTC (BA 22).

3.3. *Explicit and implicit processing of emotional vocal expressions*

Vocal expressions presented inside (explicit attention) or outside the focus of attention (implicit attention) revealed activity mainly in the bilateral pSTC (BA 22), but only vocal expressions presented outside the focus of attention revealed activity in the bilateral primary (bilateral Te1.0, left Te1.1) and secondary AC (BA 42) (Fig. 2E; Table 6A and B). During implicit attention, participants could focus their attention on the gender of a voice, or on a linguistic feature of spoken speech, or on a different spatial location (i.e. attention was directed away from vocal expressions) (Fig. 2F; Table 6C and E). During implicit attention toward vocal expressions, most of the activity in the bilateral primary AC (bilateral Te1.0, left Te1.1), in the right secondary AC (BA 42), and in the left mSTC and right pSTC (BA 22) was driven when participants focused on the gender of a voice instead of focusing on a linguistic feature (left pSTC, BA 22) or a different spatial location (right mSTC/pSTC, BA 22).

All results for the main analyses on paraverbal and nonverbal, negative and positive, and explicit and implicit attention toward vocal expressions are summarized in Fig. 3.

Table 3
Peak activations in the superior temporal cortex of regions, which were generally sensitive to vocal expressions.

Region	BA	Volume (mm^3)	MNI *x*	MNI *y*	MNI *z*
L superior temporal gyrus	22	9024	−62	−24	0
	22		−52	−14	4
	41		−52	−32	10
	41		−40	−32	8
L middle temporal gyrus	22	24	−56	−44	8
R superior temporal gyrus	22	7640	64	−32	4
	22		62	−12	0
	41		62	−22	4
	42		46	−32	6
R transverse temporal gyrus	41		52	−22	8
R middle temporal gyrus	21		56	−4	−14
R superior temporal gyrus	38	128	52	8	−16

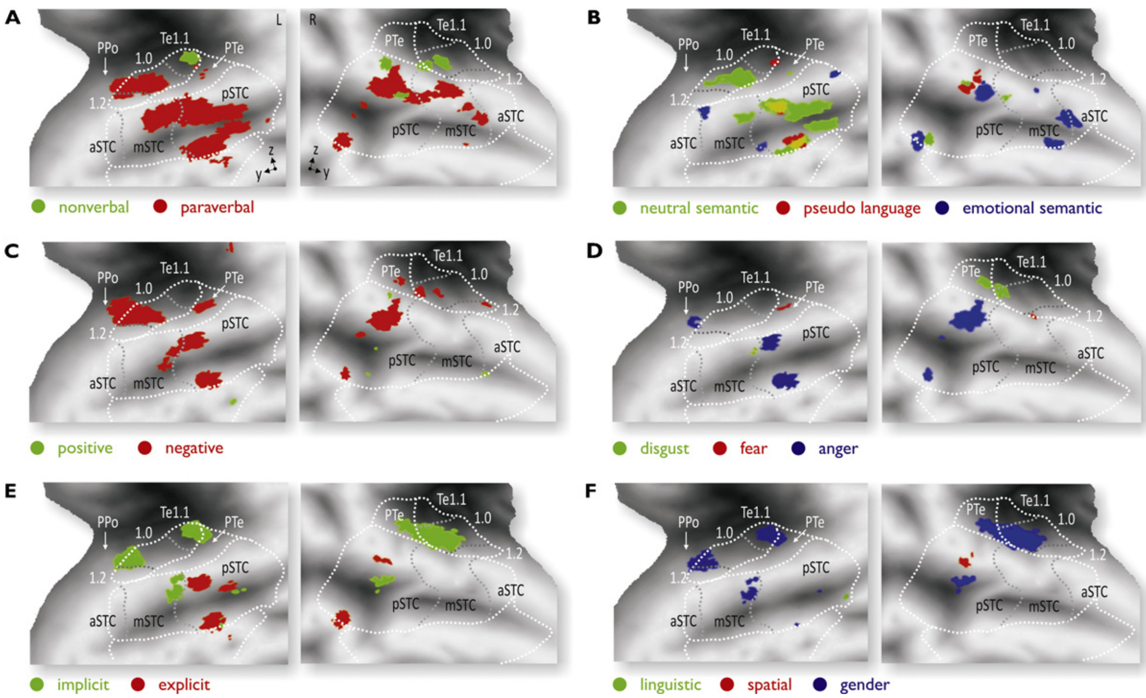

Fig. 2. Functional activation maps (A) for paraverbal and nonverbal vocal expressions and (B) for the different types of paraverbal expression (pseudo-language-based speech ["pseudo language"], semantic neutral speech ["neutral semantic"], and semantic emotional speech ["emotional semantic"]). (C) Functional activation maps for positive and negative vocal expressions. (D) Functional activation maps shown separately for the different types of negative vocal expressions (disgust, fear, anger). (E) Functional activation maps for vocal expressions presented inside (explicit) or outside the focus of attention (implicit). (F) During implicit attention, most of the activity was driven when participants focused on the gender of a voice, instead of focusing on a linguistic feature of speech (linguistic) or a different spatial location (spatial). For abbreviations, see Fig. 1.

Table 4
Peak activations for (A) nonverbal and (B) paraverbal vocal expressions. For verbal expressions, ALE maps were separately calculated for studies using (C) pseudo speech, (D) semantic neutral speech, or (E) semantic emotional speech.

Region	BA	Volume (mm^3)	MNI		
			x	*y*	*z*
(A) Nonverbal					
L superior temporal gyrus	41	304	−38	−32	8
R superior temporal gyrus	41	592	62	−20	4
	22		52	−16	8
	42	280	64	−28	14
(B) Paraverbal					
L superior temporal gyrus	22	6368	−62	−24	0
	22		−52	−32	6
	22		−52	−12	0
	41		−50	−14	6
R superior temporal gyrus	42/22	56	48	−36	4
R superior temporal gyrus	22	3440	64	−32	4
	22		62	−12	0
R middle temporal gyrus	21		56	−4	−14
(C) Pseudo speech					
L superior temporal gyrus	41	48	−46	−32	8
	22	672	−62	−26	0
R superior temporal gyrus	22	360	64	−30	6
(D) Neutral semantic speech					
L superior temporal gyrus	22	3168	−52	−34	6
	22		−64	−24	0
	41		−50	−14	8
R superior temporal gyrus	22	424	66	−34	4
	22	80	62	−12	0
(E) Emotional semantic speech					
L superior temporal gyrus	22	80	−54	−18	−8
	22	64	−54	−12	0
	22	56	−60	0	−4
	22	40	−62	−42	16
L middle temporal gyrus	21	32	−48	9	−39
R middle temporal gyrus	21	456	56	−2	−14
R superior temporal gyrus	38	32	39	15	−38
	22	336	64	−22	2
	22	280	66	−42	4
	22	32	60	−8	−4

Table 5
Peak activations for processing vocal expressions of (A) positive and (B) negative emotional valence. Peak activations for negative expressions are separately reported for expressions of (C) anger, (D) fear, or (E) disgust.

Region	BA	Volume (mm³)	MNI x	MNI y	MNI z
(A) Positive expressions					
L middle temporal gyrus	21	80	−58	−36	−9
R middle temporal gyrus	22	160	52	−38	−6
	21	144	50	−4	−24
R superior temporal gyrus	38	56	50	10	−18
(B) Negative expressions					
L superior temporal gyrus	22	2384	−62	−24	0
	22		−50	−12	4
	41	296	−52	−30	14
	22		−46	−16	−4
R superior temporal gyrus	22	1416	64	−30	4
R superior temporal gyrus	22	48	54	−6	−4
	42	304	52	−34	12
R transverse temporal gyrus	41	160	52	−20	8
(C) Angry expressions					
L superior temporal gyrus	22	488	−62	−26	0
R superior temporal gyrus	22	968	64	−30	4
(D) Fearful expressions					
L superior temporal gyrus	41/42	32	−40	−17	−6
L superior temporal gyrus	42	32	−52	−31	14
R superior temporal gyrus	22	80	54	−8	−2
(E) Expressions of disgust					
L superior temporal gyrus	22	80	−54	−20	4
R transverse temporal gyrus	41	336	52	−22	10
R superior temporal gyrus	42		54	−32	14

4. Discussion

The present meta-analytic data provide evidence, first, for a widespread and extended network of STC regions, including lower-level AC regions and higher-level auditory regions in the STG, STS, and MTG, which are responsive to vocal expressions of emotions. These emotion-sensitive regions cover areas in the primary and secondary AC, as well as in higher-level auditory regions in the STC. Interestingly, this network of STC and AC regions, which is sensitive to both paraverbal and nonverbal vocal emotions, is more extended compared with the suggested EVA (Ethofer et al., 2012), which was solely defined by paraverbally expressed vocal emotions. Second, and most important, across this widespread STC network, we found that different subregions seem to be differentially sensitive to stimulus-specific bottom-up features (medium, valence) and task-specific top-down conditions (attentional focus). These data indicate that different subregions in this emotion-sensitive STC network might represent different and distributed

Table 6
Peak activations for processing vocal expressions during (A) explicit or (B) implicit attention to the emotional cue of a voice. Peak activations for implicit processing are separately reported for (C) gender decision tasks, (D) linguistic tasks, and (E) tasks involving spatial manipulations.

Region	BA	Volume (mm³)	MNI x	MNI y	MNI z
(A) Explicit processing					
L superior temporal gyrus	22	1000	−64	−26	0
	22	264	−52	−32	12
R superior temporal gyrus	22	760	66	−36	4
(B) Implicit processing					
L superior temporal gyrus	22	824	−52	−10	−2
	22	32	−66	−28	4
	22		−46	−18	−2
	41	880	−40	−32	8
	42		−56	−20	4
	41		−62	−16	4
R transverse temporal gyrus	41	1992	52	−22	8
	22		50	−14	8
(C) Gender decision tasks					
L superior temporal gyrus	41	864	−40	−32	8
	22	808	−52	−10	−2
	41		−56	−20	4
	41		−62	−16	4
L transverse temporal gyrus	41		−46	−18	−2
L superior temporal gyrus	22	24	−66	−28	4
R transverse temporal gyrus	42	1976	52	−22	8
	41		50	−14	8
(D) Linguistic decision tasks					
L middle temporal gyrus	22	96	−48	9	−39
(E) Spatial manipulation					
R superior temporal gyrus	22	184	64	−30	4

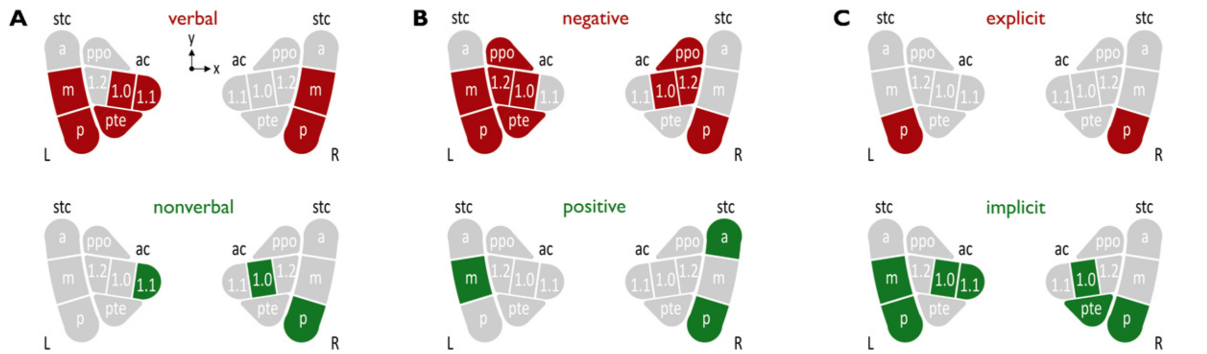

Fig. 3. Schematic summary for functional activity in response to vocal expressions in the superior temporal cortex (stc), comprising the superior and middle temporal gyrus and the superior temporal sulcus, and in the auditory cortex (ac), comprising the primary (Te1.0, Te1.1, Te1.2) and secondary auditory cortex (pte = planum temporale, ppo = planum polare). Summary for (A) paraverbal and nonverbal expressions, (B) negative and positive expressions, and (C) expressions presented inside (explicit) or outside the focus of attention (implicit).

functional roles during the decoding of emotional cues from voices.

4.1. The medium of emotional vocal expressions influences the laterality of activity

The first factor that we expected to differentially influence activity in the emotion-sensitive STC and AC was the distinction between paraverbal and nonverbal mediums, on which the emotional tone of a voice is superimposed. Paraverbal expression showed activity in the bilateral mSTC/pSTC that resembled activation generally found for human speech in the pSTC (Leaver and Rauschecker, 2010; Okada et al., 2010; Uppenkamp et al., 2006), but also bilateral responses to emotionally intonated speech in the mSTC and pSTC (Ethofer et al., 2012; Grandjean et al., 2005; Sander et al., 2005). Although these data indicate a similar role of both left and right STC for decoding emotional cues from vocal expressions, there might be a functional lateralization with stronger lexical processing in the left STC and stronger prosodic or intonation processing in the right STC (Gandour et al., 2004; Scott and Johnsrude, 2003). This assumption is corroborated by the fact that nonverbal expressions, which are not superimposed on a verbal medium and based only in voice intonation features, activated the right but not the left STC.

Besides activity in the STC, paraverbal and nonverbal expressions also revealed activity in the AC. This AC activity is thought to represent a more efficient and acoustic processing of vocal intonations (Ethofer et al., 2012). Nonverbal expressions revealed activity in the bilateral primary AC, whereas paraverbal expressions seem to be associated with activity in the left primary AC and the left PTe (BA 42). These results are suggestive of an in-depth processing of acoustic features of vocal emotions. The primary AC might specifically decode diverse acoustic features of vocal expressions, such as pitch, which is mainly based on the f0 (Bendor and Wang, 2005). The primary AC is also sensitive to the (temporal) amplitude (Giraud et al., 2000; Gourevitch et al., 2008) and the harmonics-to-noise ratio (Lewis et al., 2009). These important features (Banse and Scherer, 1996; Patel et al., 2011) might help to classify and discriminate vocal expressions (Banse and Scherer, 1996). After decoding of features in the primary AC, the PTe might figure as an integration site for these different types of spectral and temporal analyses (Griffiths and Warren, 2002), allowing the representation of a specific spectro-temporal acoustic pattern, especially for paraverbal expressions. This left lateralized activity in the primary and secondary AC for paraverbal expression might be related to the stronger linguistic-based representation of emotions in voices and the general left hemisphere dominance for spoken speech. This initial analysis in the left primary and secondary AC might separate verbal from acoustic features of vocal intonations, which are subsequently fed forward to the left and right higher-level STC, respectively, as suggested earlier.

Interestingly, neutral semantic and pseudo-language-based speech activated the left primary AC, but paraverbal expressions based on emotional semantic speech did not. Only neutral semantic and vocal expressions based on pseudo-language are superimposed on neutral mediums. This might indicate that only emotional intonations on neutral mediums receive low-level feature decoding, because these are the only cues, from which emotional information could be inferred. However, this in-depth decoding of vocal acoustic features might not only depend on the medium of vocal expressions, but also on the valence of vocal intonations, as discussed in the next section.

4.2. Valence influences lower-level and higher-level auditory cortical activity

Interestingly, negative expressions seem to activate the bilateral primary and secondary AC, but this activity was not found for positive expressions. This might indicate two important processing characteristics for negative expressions. First, negative rather than positive expressions might receive more in-depth and more efficient processing in the AC, perhaps because of their increased importance for immediate adaptive behavior (Ohman et al., 2001; Vuilleumier, 2005). Negative expressions, such as anger or fear, are characterized by specific vocal acoustic features that might receive prioritized decoding in the AC. This increase of activity for negative emotions might be related to the low road suggested by LeDoux (2012), who proposed an enhanced auditory discriminability in low-level auditory regions, as induced by subcortical regions such as the amygdala.

This is related to the second important processing characteristics of negative expressions. The negative spectrum of emotions (i.e. fear, anger, disgust, sadness) usually comprises more emotional valences compared with positive emotions (i.e. happiness). Different negative emotions involve different strategies for adaptive behavior, such as approach-related (i.e. anger) or withdrawal-related behavior (i.e. fear) (Marsh et al., 2005). Therefore, it is important to reliably differentiate between different emotions in the negative spectrum. This distinction might be based on distinguishing vocal acoustic features, such as spectral (i.e. mean and variation of pitch, mean and variation of the spectral center of gravity, harmonics-to-noise ratio) or amplitude features (mean and variation of intensity, amplitude onsets) (Banse and Scherer, 1996; Juslin and Laukka, 2003; Leitman et al., 2010a; Sauter et al., 2010). Negative expression elicited activity in the PPo (BA 52), a region that has been found to decode temporal amplitude dynamics (Giraud et al., 2000), spectral patterns (Hall et al., 2002; Warren et al., 2005), and the pitch of auditory sounds (Warren and Griffiths, 2003). This left PPo activity was observed with angry expression, but not with other vocal expressions. Angry expressions usually include a diversity of spectral and temporal features. This makes them a unique vocal expression of emotion that is able to drive brain responses considerably, enabling adaptive contextual behavior (Frühholz et al., 2012; Grandjean et al., 2005; Sander et al., 2005).

Besides this activity in the primary and secondary AC for negative expressions, negative and positive expressions revealed activity in the left and right pSTC and mSTC that has previously been shown to be consistently responsive to different types of vocal expression (Fecteau et al., 2007; Frühholz et al., 2012; Grandjean et al., 2005; Johnstone et al., 2006; Kotz et al., 2003; Leitman et al., 2010b; Meyer et al., 2005; Mothes-Lasch et al., 2011; Phillips et al., 1998; Sander et al., 2005; Szameitat et al., 2010; Warren et al., 2006a). Although these regions located in the lateral STC might be a type of higher-level perceptual and feature-integrative representation of vocal expressions, there is evidence that these areas also involve feature sensitivity for vocal expression. We have recently shown that these areas in the left mSTC and right pSTC are sensitive to the pitch and amplitude of angry voices (Frühholz et al., 2012). These areas might, however, also be sensitive to vocal features of other negative or positive expressions.

Positive expressions activated an area in the right aSTC, which was not significantly active in response to negative expressions. Although this area has been shown to be sensitive to voices (Belin et al., 2000) and to vocal expressions (Bach et al., 2008; Warren et al., 2006a), it is not consistently found for all types of vocal expressions. However, this right aSTC has recently been shown to be involved in the recognition of familiar voices (Kriegstein and Giraud, 2004). Although the studies that used positive expressions of emotion used vocal sounds that were unfamiliar to the participants, compared with negative emotions, positive emotions might elicit stronger feelings of familiarity, as was shown for facial expressions (Lander and Metcalfe, 2007).

4.3. The arousal level of vocal expressions

Not only are vocal expressions of positive emotions pleasant in terms of their affective valence, but they also usually signal and probably induce a medium to high level of arousal in the sender and the perceiver, respectively (Scherer, 1989). Although the factor *arousal* was not included in this meta-analysis because of the very small number of activation foci, the results of a study by Warren et al. (2006b) indicate that the right aSTC also encodes the arousal level of positive emotions, similar to the encoding of their valence in this region. The right aSTC might therefore figure as an integration site for a combined representation of the valence and the arousal of positive emotions. Besides the activity in the aSTC, Wiethoff et al. (2008) also report that activity in the left mSTC is associated with the arousal level of vocal emotions of both positive and negative valence. This region also overlaps with the valence effects for positive and negative emotions as revealed by this meta-analysis.

These results might indicate that activity in the bilateral STC in response to emotional compared with neutral expressions is strongly driven by the arousal level of emotional expressions. No study has yet reported the effects of arousal level in the primary and secondary AC, suggesting that activity in the AC, especially in response to negative vocal emotions as found in this meta-analysis, might be more driven by the emotional valence, but not strongly by the arousal level of vocal emotions. However, more empirical studies are needed to fully determine the influence of the arousal level on activity in the AC and STC.

4.4. Attentional foci elicit different activations in low-level auditory regions

Thus far we have discussed the distribution of functional roles across the STC and AC on the basis of stimulus-specific features as bottom-up factors during the processing of vocal expressions. However, we also expected that cognitive top-down factors might influence the activity and distribution of functional roles in these areas considerably. The most investigated top-down factor during the processing of vocal expressions is the manipulation of the attentional focus. Vocal expressions can be presented either inside (explicit attention) or outside the focus of attention (implicit attention). In the present meta-analysis, both explicit and implicit attention revealed activity in the bilateral STC that was located on the posterior parts of this area. Implicit attention also elicited activity in the left mSTC. These results seem to contradict the recent notion that the pSTC is usually involved only during explicit attention toward vocal expressions (Bruck et al., 2011). In the right STC, both types of processing similarly activated the pSTC, and thus the right STC indeed seems to respond to vocal emotions independent of the level of attentional processing. For the left STC, we found that implicit attention seems to elicit activity predominantly in the mSTC, while explicit attention to vocal expressions seems to elicit activity predominantly in the pSTC, confirming the notion of a pSTC that is sensitive to vocal emotions during explicit attention. Some activity during implicit attention was, however, also located in the left pSTC, which is in accordance with a recent observation of pSTC activity during both explicit and implicit attention toward vocal emotions (Frühholz et al., 2012).

One possible reason for this inconsistency concerning left STC activity might stem from different types of contrasts, which are used to compare explicit or implicit processing of vocal emotions. A predominance of pSTC activity during explicit attention is usually found when the explicit attention task is simply compared with the implicit task independent of the emotion effect (Bach et al., 2008; Beaucousin et al., 2007; Buchanan et al., 2000; Ethofer et al., 2009a; Mitchell et al., 2003; Wildgruber et al., 2005). Such a comparison makes it difficult, however, to tease apart task effects from emotion effects. Thus, pSTC activity might be strongly driven by the task, but less by the emotional value of voices. Instead, a pSTC activity for emotional voices during both the explicit and the implicit attention condition is usually found when the processing of emotional voices is directly compared with the processing of neutral voices within each attentional condition (Frühholz et al., 2012). Only the latter comparisons should enable the direct identification of vocal emotion effects, depending on the attentional focus, and these comparisons provide some evidence that pSTC activity is independent of the task.

Thus, left STC activity seems to be only partially separable according to the general distinction of explicit and implicit attention. However, when only the different tasks during implicit attention are considered, there may be some separate functional roles of the left STC. In the left STC, we found that focusing on the gender of a voice elicited activity mainly in the mSTC, whereas focusing on a linguistic feature elicited activity in the left pSTC. This indicates that the type of information that is inside the focus of attention while implicitly processing emotional cues from voices can influence the location of STC activity, especially in the left STC. However, we have to note that ALE maps for implicit processing using linguistic decision tasks were based on only three activation foci. Thus, more studies are required for a valid estimation of the left STC location during linguistic tasks before a clear distinction can be made in the left STC across different implicit tasks during the processing of vocal expression.

While STC activity did not reveal a large dissimilarity for the explicit or implicit processing of vocal expressions, activity in the AC seems to be considerably influenced by attentional focus. Implicit processing of vocal expressions activated the primary and secondary AC, but we did not find this activity for explicit processing of vocal expressions. In addition, this primary AC activity was mainly found when participants had to focus on the gender of a voice. In general, implicit processing of vocal expressions outside the focus of attention might challenge the decoding of emotional cues from voices, which might require more efficient and in-depth decoding of vocal features in the primary and secondary AC for reliable detection and categorization of emotional voices. This might especially be the case when participants have to focus on the gender of a voice as another important and salient feature of a voice. Again, however, we have to note that this priority of primary and secondary AC activity during implicit processing when focusing on the gender of a voice might be due to the majority of studies that used a gender decision task compared with linguistic tasks or spatial manipulations. More studies are needed to provide reliable activation maps in the comparison between each of the implicit processing conditions.

5. Conclusions

The data of this meta-analysis seem to provide evidence for a functional segregation and a functional distribution in subregions of the STC and AC in response to vocal expressions. Instead of a single and unique region that is sensitive to emotional cues of voices, multiple subregions appear to be active in response to vocally expressed emotions. Activity in these subregions depended on stimulus-specific features and on top-down conditions such as attentional focus. The medium and the valence of vocal expression, as well as attentional focus toward or away from vocal expressions, seem to influence several aspects of the spatial distribution of functional activations: whether activity was located in the AC or the STC, the laterality of activations, and the anterior–posterior distinction of STC activity. This suggests a differential weighting and thus distributed processing of acoustical and emotional information in vocal expressions of information, depending on the conditions set by the task and the available sensory acoustic information provided by vocal features of intonation. However, we have to note that all the activation maps reported here were not derived from direct statistical comparisons based on our conditions of interest. We did not compute statistical comparisons, because of the lack of overlapped activations of the resulting ALE maps. With accumulating evidence and additional activation foci reported in future studies, we might be able to provide an even more detailed picture of STC and AC subregional activity in response to vocal expressions on the basis of intraregional statistical comparisons.

Acknowledgments

This meta-analysis was supported by the Swiss National Science Foundation (SNSF 105314_124572/1) and by the NCCR in Affective Sciences at the University of Geneva (51NF40-104897).

References

Alba-Ferrara, L., Hausmann, M., Mitchell, R.L., Weis, S., 2011. The neural correlates of emotional prosody comprehension: disentangling simple from complex emotion. PLoS ONE 6, e28701.

Bach, D.R., Grandjean, D., Sander, D., Herdener, M., Strik, W.K., Seifritz, E., 2008. The effect of appraisal level on processing of emotional prosody in meaningless speech. Neuroimage 42, 919–927.

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

Beaucousin, V., Lacheret, A., Turbelin, M.R., Morel, M., Mazoyer, B., Tzourio-Mazoyer, N., 2007. FMRI study of emotional speech comprehension. Cerebral Cortex 17, 339–352.

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

Bendor, D., Wang, X., 2005. The neuronal representation of pitch in primate auditory cortex. Nature 436, 1161–1165.

Bruck, C., Kreifelts, B., Wildgruber, D., 2011. Emotional voices in context: a neurobiological model of multimodal affective information processing. Physics of Life Reviews 8, 383–403.

Buchanan, T.W., Lutz, K., Mirzazade, S., Specht, K., Shah, N.J., Zilles, K., Jancke, L., 2000. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. Brain Research Cognitive Brain Research 9, 227–238.

Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A., Ackermann, H., 2007. Semiotic aspects of human nonverbal vocalizations: a functional imaging study. Neuroreport 18, 1891–1894.

Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A., Ackermann, H., 2008. Understanding the emotional expression of verbal interjections: a functional MRI study. Neuroreport 19, 1751–1755.

Economo, C., Koskinas, G.N., 1925. Die Cytoarchitektonik der Hirnrinde des Erwachsenen Menschen. Springer-Verlag, Berlin.

Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Human Brain Mapping 30, 2907–2926.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325–1335.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., Wildgruber, D., 2006a. Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. Neuroimage 30, 580–587.

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., Grodd, W., Wildgruber, D., 2006b. Effects of prosodic emotional intensity on activation of associative auditory cortex. Neuroreport 17, 249–253.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P., 2012. Emotional Voice Areas: Anatomic Location, Functional Properties, and Structural Connections Revealed by Combined fMRI/DTI. Cerebral Cortex 22, 191–200.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., Wildgruber, D., 2009a. Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. Journal of Cognitive Neuroscience 21, 1255–1268.

Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P., 2009b. Decoding of emotional information in voice-sensitive cortices. Current Biology 19, 1028–1033.

Ethofer, T., Wiethoff, S., Anders, S., Kreifelts, B., Grodd, W., Wildgruber, D., 2007. The voices of seduction: cross-gender effects in processing of erotic prosody. Social Cognitive and Affective Neuroscience 2, 334–337.

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

Frühholz, S., Ceravolo, L., Grandjean, D., 2012. Specific brain networks during explicit and implicit decoding of emotional prosody. Cerebral Cortex 22, 1107–1117.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., Lowe, M., 2004. Hemispheric roles in the perception of speech prosody. Neuroimage 23, 344–357.

Giraud, A.L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., Kleinschmidt, A., 2000. Representation of the temporal envelope of sounds in the human brain. Journal of Neurophysiology 84, 1588–1598.

Gourevitch, B., Le Bouquin Jeannes, R., Faucon, G., Liegeois-Chauvel, C., 2008. Temporal envelope processing in the human auditory cortex: response and interconnections of auditory cortical areas. Hearing Research 237, 1–18.

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

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends in Neuroscience 25, 348–353.

Hackett, T.A., 2011. Information flow in the auditory cortical network. Hearing Research 271, 133–146.

Hall, D.A., Johnsrude, I.S., Haggard, M.P., Palmer, A.R., Akeroyd, M.A., Summerfield, A.Q., 2002. Spectral and temporal processing in human auditory cortex. Cerebral Cortex 12, 140–149.

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, 242–249.

Juslin, P.N., Laukka, P., 2003. Communication of emotions in vocal expressions and music performance: Different channels same code? Psychological Bulletin 129, 770–814.

Kotz, S.A., Meyer, M., Alter, K., Besson, M., von Cramon, D.Y., Friederici, A.D., 2003. On the lateralization of emotional prosody: an event-related functional MR investigation. Brain and Language 86, 366–376.

Kriegstein, K.V., Giraud, A.L., 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22, 948–955.

Laird, A.R., Lancaster, J.L., Fox, P.T., 2005. BrainMap: the social evolution of a human brain mapping database. Neuroinformatics 3, 65–78.

Lancaster, J.L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J.C., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Human Brain Mapping 28, 1194–1205.

Lander, K., Metcalfe, S., 2007. The influence of positive and negative facial expressions on face familiarity. Memory 15, 63–69.

Leaver, A.M., Rauschecker, J.P., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 30, 7604–7612.

LeDoux, J., 2012. Rethinking the emotional brain. Neuron 73, 653–676.

Leitman, D.I., Laukka, P., Juslin, P.N., Saccente, E., Butler, P., Javitt, D.C., 2010a. Getting the cue: sensory contributions to auditory emotion recognition impairments in schizophrenia. Schizophrenia Bulletin 36, 545–556.

Leitman, D.I., Wolf, D.H., Ragland, J.D., Laukka, P., Loughead, J., Valdez, J.N., Javitt, D.C., Turetsky, B.I., Gur, R.C., 2010b. It's not what you say, but how you say it: a reciprocal temporo-frontal network for affective prosody. Frontiers in Human Neuroscience 4, 1–13.

Lewis, J.W., Talkington, W.J., Walker, N.A., Spirou, G.A., Jajosky, A., Frum, C., Brefczynski-Lewis, J.A., 2009. Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. Journal of Neuroscience 29, 2283–2296.

Marsh, A.A., Ambady, N., Kleck, R.E., 2005. The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. Emotion 5, 119–124.

Meyer, M., Zysset, S., von Cramon, D.Y., Alter, K., 2005. Distinct fMRI responses to laughter speech, and sounds along the human peri-sylvian cortex. Brain Research Cognitive Brain Research 24, 291–306.

Mitchell, R.L., 2006. How does the brain mediate interpretation of incongruent auditory emotions? The neural response to prosody in the presence of conflicting lexico-semantic cues. The European Journal of Neuroscience 24, 3611–3618.

Mitchell, R.L., Elliott, R., Barry, M., Cruttenden, A., Woodruff, P.W., 2003. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. Neuropsychologia 41, 1410–1421.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., Zilles, K., 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13, 684–701.

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

Mothes-Lasch, M., Mentzel, H.J., Miltner, W.H., Straube, T., 2011. Visual attention modulates brain activation to angry voices. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 31, 9594–9598.

Ohman, A., Flykt, A., Esteves, F., 2001. Emotion drives attention: detecting the snake in the grass. Journal of Experimental Psychology: General 130, 466–478.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cerebral Cortex 20, 2486–2495.

Patel, S., Scherer, K.R., Bjorkner, E., Sundberg, J., 2011. Mapping emotions into acoustic space: the role of voice production. Biological Psychology 87, 93–98.

Phillips, M.L., Young, A.W., Scott, S.K., Calder, A.J., Andrew, C., Giampietro, V., Williams, S.C., Bullmore, E.T., Brammer, M., Gray, J.A., 1998. Neural responses to facial and vocal expressions of fear and disgust. Biological sciences/The Royal Society 265, 1809–1817.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P., 2005. Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. Neuroimage 28, 848–858.

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

Scherer, K.R., 1989. Vocal correlates of emotional arousal and affective disturbance. In: Wagner, H.L., Manstead, S.R. (Eds.), Handbook of Psychophysiology: Emotion and Social Behavior. Wiley, Chichester, pp. 165–197.

Schirmer, A., Fox, P.M., Grandjean, D., 2012. On the spatial organization of sound processing in the human temporal lobe: a meta-analysis. Neuroimage 63, 137–147.

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

Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends in Neuroscience 26, 100–107.

Szameitat, D.P., Kreifelts, B., Alter, K., Szameitat, A.J., Sterr, A., Grodd, W., Wildgruber, D., 2010. It is not always tickling: distinct cerebral responses during perception of different laughter types. Neuroimage 53, 1264–1271.

Uppenkamp, S., Johnsrude, I.S., Norris, D., Marslen-Wilson, W., Patterson, R.D., 2006. Locating the initial stages of speech-sound processing in human temporal cortex. Neuroimage 31, 1284–1296.

Van Essen, D.C., Drury, H.A., Dickson, J., Harwell, J., Hanlon, D., Anderson, C.H., 2001. An integrated software suite for surface-based analyses of cerebral cortex. Journal of the American Medical Informatics Association 8, 443–459.

Vuilleumier, P., 2005. How brains beware: neural mechanisms of emotional attention. Trends in Cognitive Sciences 9, 585–594.

Warren, J.D., Griffiths, T.D., 2003. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 23, 5799–5804.

Warren, J.D., Jennings, A.R., Griffiths, T.D., 2005. Analysis of the spectral envelope of sounds by the human brain. Neuroimage 24, 1052–1057.

Warren, J.D., Scott, S.K., Price, C.J., Griffiths, T.D., 2006a. Human brain mechanisms for the early analysis of voices. Neuroimage 31, 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. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 26, 13067–13075.

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., Ethofer, T., 2008. Cerebral processing of emotional prosody: influence of acoustic parameters and arousal. Neuroimage 39, 885–893.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., Ackermann, H., 2004. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. Cerebral Cortex 14, 1384–1389.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., Ackermann, H., 2005. Identification of emotional intonation evaluated by fMRI. Neuroimage 24, 1233–1241.

Wittfoth, M., Schroder, C., Schardt, D.M., Dengler, R., Heinze, H.J., Kotz, S.A., 2010. On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. Cerebral Cortex 20, 383–392.