Author: Witteman, Jurriaan
Title: Towards a cognitive neuroscience of prosody perception and its modulation by alexithymia
Issue Date: 2014-12-18

Chapter 2.3

A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception

Witteman, J., Van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia, 50,* 2752-2763.

Abstract

With the advent of neuroimaging considerable progress has been made in uncovering the neural network involved in the perception of emotional prosody. However, the exact neuroanatomical underpinnings of the emotional prosody perception process remain unclear. Furthermore, it is unclear what the intrahemispheric basis might be of the relative right-hemispheric specialization for emotional prosody perception that has been found previously in the lesion literature. In an attempt to shed light on these issues, quantitative meta-analyses of the neuroimaging literature were performed to investigate which brain areas are robustly associated with stimulus-driven and task-dependent perception of emotional prosody. Also, lateralization analyses were performed to investigate whether statistically reliable hemispheric specialization across studies can be found in these networks. A bilateral temporofrontal network was found to be implicated in emotional prosody perception, generally supporting previously proposed models of emotional prosody perception. Right-lateralized convergence across studies was found in (early) auditory processing areas, suggesting that the right hemispheric specialization for emotional prosody perception reported previously in the lesion literature might be driven by hemispheric specialization for non-prosody-specific fundamental acoustic dimensions of the speech signal.

1. Introduction

For group-dwelling primates including humans the ability to recognize and strategically react to the emotions of con-specifics is essential as it promotes reproductive success (Silk, 2007). The 'social brain hypothesis' even proposes that selective pressure on such social processing capacities drove the relative expansion of the neocortex in the course of primate evolution (Dunbar, 1998). The importance of the ability to read the (emotional) mind of others in modern human society is exemplified by the devastating effects that disorders of 'theory of mind' processing such as autism spectrum disorders can have on daily functioning of affected individuals (for a review on the evolution and disorders of theory of mind see e.g. Brüne & Brüne-Cohrs, 2006).

How something is being said is an important source of information regarding the emotional state of the speaker. This emotional prosodic (or suprasegmental, i.e., pertaining to a phonological element whose domain is something larger than a single segment, Trask, 1996, p. 343) layer of speech uses a variety of acoustic cues such as intensity, speaking rate, F_0 and spectral balance to signal emotional states (Scherer, 2003). Cross-cultural studies have found better-than-chance decoding accuracy between cultures for basic vocal emotions such as anger and sadness (Scherer, Banse, & Wallbott, 2001; Van Bezooijen, Otto, & Heenan, 1983), suggesting that the human brain might indeed harbor a dedicated emotional voice recognition system that is rooted in our evolutionary history.

Our knowledge of how the brain processes emotional prosody has been greatly advanced with the advent of neuroimaging techniques. Using this technique there generally have been two approaches to delineating the neural network involved in emotional prosody perception. On the one hand, it has been tested which regions in the brain show a stronger response to emotional prosodic categories (such as anger) than to neutral prosody: this network includes areas involved in the extraction of acoustic cues that are important for emotional prosody perception and therefore has been called 'stimulus-driven' (Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009). On the other hand, it has been investigated which brain areas are more active when attention is directed to emotional prosody as compared to direction of attention to a non-emotional prosodic dimension of the same stimulus material (such as the segmental structure or semantics). As this network of regions is sensitive to emotional prosodic task demands independently of acoustics, it has been called 'task-dependent' (Wildgruber et al., 2009). By comparing these stimulus-driven and task-dependent networks, insight can be gained in the regional sensitivity of the brain to earlier and later stages of the emotional prosodic perception process respectively despite the relatively low temporal resolution of functional neuroimaging.

Imaging studies employing stimulus-driven contrasts have generally found a network of areas comprising the superior temporal cortex (Beaucousin et al., 2007; Brück, Kreifelts, Kaza, Lotze, & Wildgruber, 2011; Dietrich, Hertrich, Alter, Ischebeck, & Ackermann, 2008; Ethofer et al., 2007; Ethofer, Van de Ville, Scherer, & Vuilleumier, 2009; Ethofer et al., 2011; Frühholz, Ceravolo, & Grandjean, 2011; Grandjean et al., 2005; Imaizumi et al., 1997; Kotz et al., 2003; Leitman et al., 2010; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Morris, Scott, & Dolan, 1999; Mothes-Lasch, Mentzel, Miltner, & Straube, 2011; Sander et al., 2005; Wiethoff et al., 2007; Wittfoth et al., 2009) and frontal cortical areas (Brück et al., 2011b; Ethofer et al., 2009; Frühholz et al., 2011; Grandjean et al., 2005; Kotz et al., 2003; Leitman et al., 2010; Mitchell et al., 2003; Morris et al., 1999). Additionally, many of these studies find involvement of subcortical structures such as the insula (Ethofer et al., 2009; Leitman et al., 2010; Morris et al., 1999; Mothes-Lasch et al., 2011), basal ganglia (Ethofer et al., 2009; Kotz et al., 2003; Leitman et al., 2010; Morris et al., 1999) and to a lesser extent the amygdala (Ethofer et al., 2009; Mothes-Lasch et al., 2011; Sander et al., 2005; Wiethoff, Wildgruber, Grodd, & Ethofer, 2009).

Studies using task-dependent contrasts tend to find frontal activations (Bach et al., 2008; Brück, Kreifelts, Kaza, Lotze, & Wildgruber, 2011; Buchanan et al., 2000; Ethofer et al., 2006; Ethofer et al., 2009; George et al., 1996; Imaizumi et al., 1997; Mitchell et al., 2003; Wildgruber et al., 2004) in addition to superior temporal activation (Bach et al., 2008; Ethofer et al., 2006; Ethofer et al., 2009; Frühholz, Ceravolo, & Grandjean, 2011; Gandour et al., 2003; Wildgruber et al., 2005). This pattern of results seems to suggest that superior temporal regions are primarily involved in the initial stages of the emotional prosodic perception process, while frontal areas are involved in later stages.

This literature together with electrophysiological and clinical data has inspired two rather similar multi-stage models (Schirmer & Kotz, 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009) of emotional prosody perception. Both models propose that in an initial phase there is extraction of basic acoustic properties from the speech signal in the primary and secondary auditory cortex. In the model of Wildgruber et al. (2009) the mid-superior temporal gyrus (mid-STG) is explicitly mentioned as an important brain region participating in this phase. Furthermore, both models suggest that, although processing is bilateral in this stage, there might be relative superiority of the right auditory cortex. Both models postulate a second phase in which there is integration of acoustic information into an emotional percept or 'gestalt' (or where meaningful suprasegmental representations of acoustic sequences are realized; Wildgruber et al., 2009) but the models differ in the exact neuroanatomical underpinnings. While the model of Schirmer and Kotz (2006) and its further elaboration by Kotz and Paulmann (2011) suggest that this integration of acoustic cues is achieved by moving anteroventrally along the auditory 'what'-pathway to the anterior superior temporal sulcus (STS), the model by Wildgruber et al. (2009) and its further expansion by Brück et al. (2011a) proposes that this stage is supported by the posterior superior temporal cortex. Both models suggest right-hemispheric specialization in this second phase. Finally, both models propose a third phase in which emotional prosody is explicitly evaluated and integrated with other cognitive processes. While the model by Brück et al. (2011a) assumes that both (left and right) inferior frontal gyri (IFG) have a similar evaluative function in this phase, the Kotz and Paulmann (2011) model further specifies that the right IFG and orbitofrontal cortex (OFC) are involved in the explicit evaluation of emotional prosody and that the left IFG is responsible for integrating emotional prosodic information with other layers of the speech signal (such as semantics). Concerning the role of subcortical structures, the model of Kotz and Paulmann (2011) proposes that the amygdala and basal ganglia (BG) are important for the detection of emotional salience and evaluation (stage 3) of emotion prosody respectively. The model by Brück et al. (2011a) proposes that the amygdala plays a role in implicit emotional prosodic perception but leaves the role of the BG unspecified.

A longstanding question in the neuropsychological literature is whether there is hemispheric specialization within this emotional prosodic perception network. An important source of evidence regarding hemispheric specialization is provided by the lesion literature as it gives information on the necessity of cerebral structures for emotional prosody perception. A recent meta-analysis of the lesion literature (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011) indeed showed that while damage to both cerebral hemispheres degrades emotional prosody perception performance, damage to the right hemisphere compromised performance more, suggesting that there is *relative* right-hemispheric specialization for emotional prosody perception. However, it remains unclear through which mechanism such right hemispheric specialization for emotional prosodic perception might emerge: hemispheric specialization might emerge in any of the three stages proposed by the two emotional prosody perception models.

Indeed, on the one hand the *cue-dependent lateralization hypotheses* propose that the right-hemisphere specialization for emotional prosody perception can be traced back to a (non-prosody specific) advantage of the right hemisphere for early acoustic processing such as spectral processing (Van Lancker & Sidtis, 1992) as spectral parameters are important for decoding of emotional prosody (Scherer, 2003). The idea that there is hemispheric specialization for processing of elementary dimensions of speech such as temporal vs. spectral information (Zatorre & Belin, 2001) or integration of the speech signal over smaller vs. longer time windows (Poeppel, 2003) in the left and right hemisphere respectively has recently indeed been gaining support (see e.g. Boemio, Fromm, Braun, & Poeppel, 2005; Jamison, Watkins, Bishop, & Matthews, 2006). Both proposed models of emotional prosody perception (Brück, Kreifelts, & Wildgruber, 2011a; Kotz and Paulmann, 2011) indeed assume relative right hemispheric specialization for the first (acoustic) stage of emotional prosody perception in the auditory cortex.

On the other hand, the *functional lateralization hypothesis* proposes that the right hemisphere is specialized for the processing of the emotional communicative *function* of emotional prosody (Van Lancker, 1980). This proposal might correspond to the more abstract processing of emotional prosody in stage 2 (integration of acoustic cues into an emotional percept) or stage 3 (explicit evaluation of emotional prosody) in the models of emotional prosody perception described above. A closer look at the proposed models of emotional prosody perception reveals that while both models assume right-hemispheric specialization for stage 2 emotional prosody perception, the Kotz and Paulmann (2011) model also allows for right-hemispheric specialization in the last evaluative stage while the Brück et al. (2011a) model assumes bilateral processing. As a consequence of the multi-stage nature of emotional prosodic perception, right-hemispheric specialization for the whole *process* of emotional prosody perception (as frequently assessed through behavioral tasks in lesion studies) might thus depend on right-hemispheric specialization for just one stage or could be additive or even interactive (for instance, there might be top-down modulated right-hemispheric specialization for acoustic processing based on task demands) across multiple stages.

In sum, although considerable progress has been made in delineating the neural network involved in emotional prosody perception some questions remain unresolved. First, recently proposed models disagree on the exact neuroanatomical substrate of the proposed second stage (integration of acoustic cues into an emotional percept or representation of meaningful suprasegmental acoustic sequences) of emotional prosody perception. Second, is it unclear where in the process of emotional prosody perception the right-hemispheric specialization arises that has been found in the lesion literature.

The exact pattern of results found in fMRI studies is dependent on many variables including the scanner hardware, scanning parameters, exact task instructions and analysis methods and hence results may vary considerably between studies (Bennett & Miller, 2010). Therefore, there seems to be merit in determining which cerebral structures are reliably activated across studies that employ a variety of slightly different tasks to measure the same hypothesized underlying cognitive process. This will provide information about which neural network is generally involved (without study-specific idiosyncratic effects) in the hypothesized cognitive process. To the best of our knowledge, no such quantitative meta-analysis of the neuroimaging literature of emotional prosodic perception has as yet been undertaken. Therefore, in the present investigation activation likelihood estimation (ALE) meta-analysis, a commonly used and valid coordinate based voxel-wise quantitative meta-analysis method for neuroimaging data (Eickhoff et al., 2009; Laird et al., 2005; Turkeltaub et al., 2011; Turkeltaub, Eden, Jones, & Zeffiro, 2002), was used to address following questions:

(1) Which neural networks are reliably involved in stimulus-driven and task-dependent perception of emotional prosody?
(2) Can statistically robust lateralization of activation probability be demonstrated for these networks?

Two ALE meta-analyses were performed in an attempt to answer these questions: one on stimulus-driven processing of emotional prosody and another on task-driven processing of emotional prosody. It was hypothesized that these meta-analyses would yield the bilateral temporofrontal network with subcortical involvement that has been proposed by recent models of emotional prosody perception. Additionally, for each of the two meta-analyses subtraction analyses were performed to investigate whether there was statistically reliable lateralization of ALE clusters. It was hypothesized that there would be multiple lateralized clusters, with clusters in early auditory processing areas (reflecting cue-dependent lateralization) and in multimodal temporal or frontal areas (possibly corresponding to functional lateralization).

2. Methods

2.1. Study and experiment selection

The PubMed database (www.ncbi.nlm.nih.gov/pubmed/) was searched for relevant publications until September 2011 using the following search string for the title or the abstract (where the asterisk denotes a wildcard): (brain OR neural) AND emotion* AND (prosod* OR vocal OR voice). Additionally, the reference lists of articles and recent reviews (Brück et al., 2011a; Kotz & Paulmann, 2011) were checked for potentially relevant publications. The following inclusion criteria were applied: (1) an emotional prosodic perception task was employed (i.e. studies using nonverbal vocal emotional sounds such as laughter were excluded) (2), emotional prosody was the only relevant independent variable that was varied (i.e. studies contrasting emotional prosody that was congruent with emotional lexical semantics to neutral prosody in words with neutral lexical semantics were excluded, since emotional semantics and emotional prosody were confounded in that case), (3) subjects were healthy adult participants, (4) whole brain analyses were reported (ROI analyses were excluded), (5) peak coordinates of significantly activated areas were reported in the MNI (Evans et al.,

1993) or Talairach (Talairach & Tournoux, 1988) standard stereotactic space, and (6) the article was published in a peer-reviewed international journal.

Experiments (contrasts) were selected from each paper as follows. For the meta-analysis on the stimulus-driven network of prosodic perception, the active condition had to be emotional prosody (i.e. emotional expression superimposed on an utterance containing speech segments such as voiced nasals, vowels, words or pseudowords) and the control condition either neutral prosody or speech material devoid of emotional prosody (synthesized speech). For the meta-analysis on the task-dependent network of emotional prosodic perception, attention had to be directed to the emotional prosody in the active condition, while in the control condition the same material had to be presented without attention directed at the emotional prosody. If a study reported multiple experiments that were eligible for inclusion in one of the two meta-analyses, the contrast was chosen that best answered the question "which network in the brain is involved in the processing of emotional prosody?" (e.g. when foci were presented for individual emotions separately and for all emotions combined, the latter experiment was included as it is more likely that it engages the full emotional prosodic perception network) to prevent one study from driving the meta-analysis.

All included studies used fMRI except George et al. (1996) and Imaizumi et al. (1997), which used PET. As can be observed in Table 1, a total of 16 studies (N = 296 subjects; 93 foci) was included in the meta-analysis on the stimulus-driven network of emotional prosodic perception. All of the included studies compared speech with emotional prosody to speech with neutral prosody or to synthesized speech without prosody.

Table 2 lists the 11 studies (N = 153 subjects; 50 foci) that were included in the meta-analysis on the task-driven network of emotional prosodic perception. All the included studies compared emotional prosody processing to the processing of a different dimension (such as speaker gender or lexical semantics) of the same speech stimuli. Note that Ethofer et al. (2006) was excluded from the task-driven analysis, as a predefined ROI was used in the analyses (i.e. no whole-brain corrected coordinates were reported).

Table 1. Overview of studies included in the meta analysis on stimulus-driven effects.

Study	**N (males)**	**Active condition**	**Control condition**	**Emotions**	**Task(s)**	**N foci**
Bach 2008	16 (8)	Emotional	Neutral	Anger, fear, neutral	• Categorize gender • Categorize emotional prosody	9
Beacousin 2007	23 (11)	Emotional	No prosody (text to Speech)	Anger, sadness, joy, no emotional intonation	• Categorize emotional semantics	10
Brück 2011b	24 (12)	Emotional	Neutral	Anger, joy, neutral	• Identify emotional prosody • Identify vowel • Identify word	4
Ethofer 2007	24 (12)	Emotional	Neutral	Anger, fear, joy, eroticism, neutral	• Passive listening	6
Ethofer 2009	24 (12)	Anger	Neutral	Anger, neutral	• Categorize emotional prosody • Categorize word class	7
Ethofer 2011	22 (13)	Emotional	Neutral	Anger, joy, sadness, relief, neutral	• Categorize gender	2
Frühholz 2011	17 (3)	Anger	Neutral	Anger, neutral	• Discriminate emotional prosody • Discriminate gender	13
Grandjean 2005	15 (8)	Anger	Neutral	Anger, neutral	• Categorize gender	8
Kotz 2003	12 (4)	Anger	Neutral	Joy, anger, neutral	• Categorize emotional prosody	12
Leitman 2010	19 (19)	Emotional	Neutral	Joy, anger, fear, neutral	• Categorize emotional prosody	4
Mitchell 2003	13 (13)	Emotional	Neutral	Joy, sadness, neutral	• Passive listening	6
Mothes-Lasch 2011	24 (16)	Anger	Neutral	Anger, neutral	• Categorize gender	1
Rota 2008	10 (10)	Emotional	Neutral	Joy, sadness, anger, neutral	• Categorize emotional prosody	0
Sander 2005	15 (8)	Anger	Neutral	Anger, neutral	• Categorize gender	8
Schirmer 2008	14 (14)	Anger	Neutral	Anger, neutral	• Watch movie and listen passively	0
Wiethoff 2007	24 (12)	Emotional	Neutral	Joy, eroticism, anger, fear, sadess, disgust, neutral	• Passive listening	3
Total	**296**					**93**

Table 2. Overview of studies included in the meta analysis on task driven effects.

Study	**N** (male)	**Active condition**	**Control condition**	**Emotion**	**Tasks**	**N foci**
Bach 2008	16 (8)	Categorize prosody	Categorize gender	Anger, fear, neutral.	• Categorize gender • Categorize emotional prosody	11
Brück 2011	24 (12)	Identify prosody	Identify vowel + Identify words	Anger, joy, neutral	• Identify emotional prosody • Identify vowel • Identify word	1
Buchanan 2000	10 (10)	Detect emotional prosody	Detect phoneme	Anger, joy, sad, neutral	• Detect joy • Detect sadness • Detect phoneme	3
Ethofer 2009	24 (12)	Categorize emotional prosody	Categorize word class	Anger, neutral	• Categorize emotional prosody • Categorize word class	4
Frühholz 2011	17 (3)	Discriminate emotional prosody	Discriminate gender	Neutral, Anger	• Discriminate emotional prosody • Discriminate gender	3
Gandour 2003	10 (5)	Discriminate emotional prosody	Passive listening	Anger, joy, sadness	• Discriminate emotional prosody • Passive listening	7
George 1996	13 (8)	Categorize emotional prosody of sentence	Categorize word	Joy, sadness, anger, neutral.	• Categorize emotional prosody • Categorize word	1
Imaizumi 1997	6 (6)	Categorize emotional prosody of words	Categorize speaker identity	Surprise, disgust, joy anger	• Categorize emotional prosody • Categorize speaker identity	13
Mitchell 2003	13 (13)	Attend prosody	Attend semantics	Joy, sadness, neutral	• Attend prosody • Attend semantics	3
Wildgruber 2004	10 (6)	Discriminate emotional prosody	Discriminate linguistic prosody	Emotional expressiveness	• Discriminate emotional prosody • Discriminate linguistic prosody	2
Wildgruber 2005	10 (5)	Categorize emotional prosody	Categorize vowels	Joy, anger, fear, sadness, disgust	• Categorize emotional prosody • Categorize vowels	2
Total	**153**					**50**

2.2. ALE meta-analyses

Activation Likelihood Estimation (ALE) is a widely used and valid coordinate-based meta-analysis procedure that tests for consistent brain activation across studies for a given set of (similar) tasks (Laird et al., 2005; Turkeltaub, Eden, Jones, & Zeffiro, 2002). In brief, ALE meta-analysis uses the peak-coordinates of activation clusters reported in the previous literature and models the location uncertainty of these coordinates as 3D-Gaussian probability density distributions. The voxelwise convergence of these distributions is then tested against a null-distribution and corrected for multiple comparisons to test which voxels show above-chance convergence across studies. As a last step cluster analysis is performed with a user-specified minimum cluster volume. Recently, the ALE-algorithm has been further optimized by adopting a random-effects approach (allowing for generalization of the results beyond the experiments analyzed) and more precisely modeling location uncertainty (Eickhoff et al., 2009) and by minimizing the cumulative effects of multiple neighboring within experiment foci (Turkeltaub et al., 2011). All ALE-analyses were performed using the latest version of the ALE-algorithm (Turkeltaub et al., 2011) at the time of writing, as implemented in GingerALE 2.1.1 (www.brainmap.org/ale).

As discussed in the Introduction, an important unresolved question in the literature is whether there is hemispheric specialization for emotional prosodic perception. Therefore, we additionally applied a lateralization analysis as reported previously by Turkeltaub and Coslett (2010) for each of the two meta-analyses. To formally test for significant lateralization of ALE-values, we multiplied x-coordinates of the input foci by −1 (effectively left-right flipping each coordinate) and performed a subtraction analysis where we subtracted the flipped foci from the original (unflipped) foci.

All meta-analyses were performed in Talairach space. When foci were reported in MNI space, they were converted to Talairach space using the Lancaster transform (Lancaster et al., 2007) as implemented in the GingerALE software. For the two main meta-analyses on stimulus and task-driven emotional prosody perception, above-chance voxel wise convergence was tested corrected for multiple comparisons using the false discovery rate (FDR) approach (Genovese, Lazar, & Nichols, 2002) as implemented in GingerALE with $q = 0.05$ and a minimum cluster extent of 100 mm^3. These FDR corrected images were subsequently used as input images for the subtraction analyses to test for lateralization effects. At present there is no established method to correct ALE subtraction maps for multiple comparisons (Eickhoff et al., 2011, p. 941). Further, for smaller data-sets with 10-20 experiments as the present, an uncorrected threshold with a small extent threshold has been suggested to provide sufficient protection against false positive ALE-clusters (Gobras, Beaton, & Eickhoff, 2012, p. 439). Therefore subtraction results are reported with a 100 mm^3 extent threshold in combination with an uncorrected threshold at a more lenient $p = 0.05$ and a more conservative $p = 0.001$ level for ALE-values. Statistical maps were overlaid onto the Talairach anatomical template and visualized using Mango software (http://ric.uthscsa.edu/mango/). Additionally, similarly to previous studies (Turkeltaub & Coslett, 2010; Van der Laan, De Ridder, Viergever, & Smeets, 2011) we applied one further conservativeness criterion. To avoid discussing clusters to which only one or two studies directly contributed, at least three of the included studies had to contribute directly to a significant ALE-cluster in order to be discussed in the discussion-section (all significant clusters are reported in the results section).

3. Results

3.1. Main analyses

3.1.1. Stimulus-driven network

The ALE-analysis of experiments contrasting emotional prosody to neutral prosody or speech without prosody revealed 15 clusters of above chance convergence (see Table 3 and Figure 1). Five clusters met our additional criterion of having at least three directly contributing studies. The most concurrent ALE-cluster lay in the right mid superior temporal gyrus and continued medially into the transverse temporal gyrus (peak Talairach coordinates (44, −24, 8), ALE-value = 2.67•10^{-3}, volume = 2696 mm^3). In the left temporal cortex two significant ALE-clusters were found. The cluster with the highest convergence was located in the mid-STG (peak Talairach coordinates (−58, −22, 2), ALE-value = 1.57•10^{-3}, volume = 2696 mm^3), while the other ALE-cluster was located in the posterior STG (peak Talairach coordinates (−42, −34, 10), ALE-value = 1.52•10^{-3}, volume = 848 mm^3). Two ALE-clusters were located in the inferior frontal gyri. One was located in the lateral left IFG (peak Talairach coordinates (−54, 20, 2), ALE-value = 1.23•10^{-3}, volume = 680 mm^3). The other cluster was located in the right IFG ALE-cluster (peak Talairach coordinates (48, 28, 6), ALE-value = 1.23•10^{-3}, volume = 976 mm^3) and extended into the dorsal anterior insula.

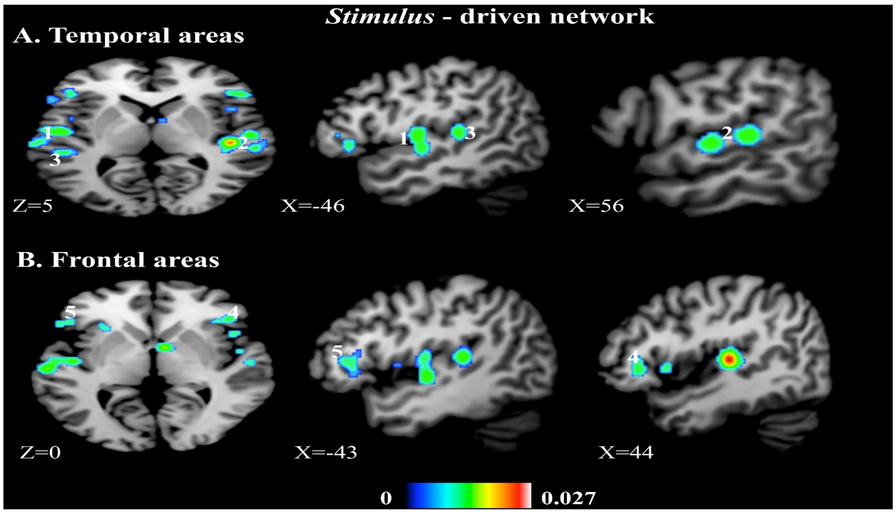

Figure 1. Results of the meta-analysis of studies contrasting emotional prosody to neutral prosody or synthesized speech. All ALE-clusters were corrected for multiple comparisons at an FDR of 0.05 and had a minimum cluster extent of 100 mm^3. For display purposes the cluster-uncorrected statistical maps are shown. Clusters to which at least three studies have directly contributed are marked with a number that corresponds to the cluster number in Table 3. In the top-panel A, 1 = left STG, 2 = right STG and 3 = left STG. In the bottom-panel B, 4 = right IFG, 5 = left IFG.

Table 3. Locations (Talairach space) of significant ALE-clusters.

Clu ster	**Cluster size (mm^3)**	**Anatomical Label**	**Peak voxel coordinates**			**ALE-value (•10^{-3})**	**No. stu dies**
			x	**y**	**z**		
***Stimulus*-driven analysis:**							
1	2696	L Superior Temporal Gyrus	−58	−22	2	1.57	7
		L Superior Temporal Gyrus	−48	−12	6	1.56	
		L Insula	−42	−14	−4	1.48	
2	2696	R Superior Temporal Gyrus	44	−24	8	2.67	7
		R Superior Temporal Gyrus	56	−30	8	1.55	
		R Superior Temporal Gyrus	54	−16	4	1.54	
3	848	L Superior Temporal Gyrus	−42	−34	10	1.52	3
4	808	R Inferior Frontal Gyrus	48	28	6	1.23	3
		R Inferior Frontal Gyrus	44	28	2	1.18	
		R Inferior Frontal Gyrus	36	24	0	0.91	
5	680	L Inferior Frontal Gyrus	−54	20	2	1.23	3
		L Inferior Frontal Gyrus	−42	28	6	1.08	
		L Inferior Frontal Gyrus	−46	24	0	1.02	
6	528	L Inferior Frontal Gyrus	−34	6	−12	1.74	2
7	504	R Medial Globus Pallidus	8	−2	−2	1.66	2
8	400	R Medial Frontal Gyrus	16	40	14	1.52	2
9	400	L Superior Occipital Gyrus	−32	−72	30	1.52	2
10	392	R Middle Occipital Gyrus	24	−86	18	1.48	2
11	392	L Inferior Frontal Gyrus	−52	20	24	1.47	2
12	224	R Insula	46	12	2	1.05	2
13	192	R Superior Temporal Gyrus	50	−4	−4	1.02	2
14	152	R Subcallosal Gyrus	26	4	−10	0.91	2
		R Putamen	26	−2	−6	0.89	2
15	152	L Claustrum	−26	18	−2	0.99	2
***Task*-driven analysis:**							
1	880	R Superior Temporal Gyrus	46	−34	6	1.35	3
2	512	R Inferior Frontal Gyrus	40	34	2	0.93	3
		R Inferior Frontal Gyrus	36	34	−6	0.81	
3	184	L Middle Frontal Gyrus	−42	34	0	0.90	1
4	168	R Superior Frontal Gyrus	10	20	50	0.86	1
5	128	R Middle Frontal Gyrus	48	26	24	0.85	1

L = Left hemisphere; *R* = Right hemisphere. All reported ALE-clusters were thresholded at $q < 0.05$ corrected and an extent threshold of 100 mm^3

3.1.2. Task-driven network

The ALE-analysis of experiments contrasting processing of emotional prosody to processing a different dimension of the same material revealed 5 clusters of above chance convergence (see Table 3 and Figure 2). Two of these clusters fulfilled our additional criterion of having at least three directly contributing studies. The cluster of highest convergence lay in the right posterior STG (peak Talairach coordinates (46, −34, 6), ALE-value = $1.35 \bullet 10^{-3}$, volume = 880 mm^3). The second cluster was located in the right lateral IFG (peak Talairach coordinates (40, 34, 2), ALE-value = $.93 \bullet 10^{-3}$, volume = 416 mm^3). Although significant ALE-voxels were also noted in the left IFG, these voxels did not survive the additional criterion of a minimum cluster volume of 100 mm^3.

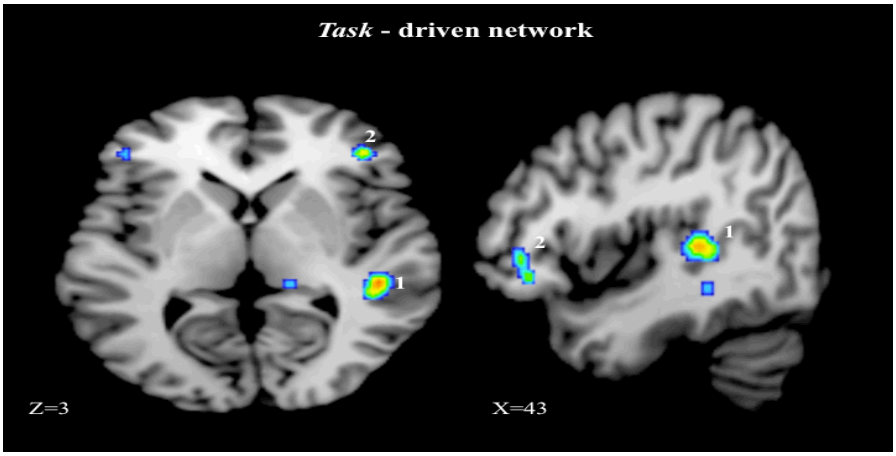

Figure. 2. Results of the meta-analysis of studies contrasting processing of emotional prosody to processing of a different dimension of the same stimulus material. All ALE-clusters were corrected for multiple comparisons at a FDR of 0.05 and had a minimum cluster extent of 100 mm^3. For display purposes the cluster-uncorrected statistical maps are shown. Clusters to which at least three studies have directly contributed are marked with a number that corresponds to the cluster number in Table 4. 1 = right posterior STG, 2 = right IFG.

3.2. Lateralization analyses

An important issue in the literature is whether there is hemispheric specialization for emotional prosodic perception. For the stimulus-driven analysis, convergence seemed to be higher in the right temporal lobe than the left. For the task-driven analysis, both significant ALE-clusters were right lateralized. These qualitative impressions were followed up with a formal test of lateralization of ALE-clusters by subtracting ALE-maps that were flipped around the x-axis (i.e. right/left flipping the ALE-map) from the original ALE-maps. One weakness of the ALE meta-analysis is that activation magnitude (effect size) and spatial extent of clusters are not considered. Therefore, similarly to Turkeltaub and Coslett (2010), we checked all included studies for whether the effect size (Z or T-value) and spatial extent of peak (maximally active) clusters in the vicinity of lateralized ALE clusters was larger in the left or right hemisphere. To test whether lateralization of cluster activation magnitude and extent was statistically significant between the hemispheres, a binomial test was performed.

3.2.1. Stimulus-driven network

Table 4 lists the two ALE-clusters that showed significant lateralization in the subtraction analysis at $p = 0.05$ uncorrected for the stimulus-driven contrasts. As can be seen in Figure 3, the only ALE cluster that fulfilled our additional criterion of having at least three directly contributing studies was located in the right transverse temporal gyrus (peak Talairach coordinates (50,−24,10), $Z = 1.95$, volume = 624 mm^3). No ALE-cluster survived the more conservative $p = 0.001$ threshold. Although peak activation of temporal lobe clusters tended to be greater in the right than left hemisphere, this effect was not statistically reliable (P(binomial) = 0.27). However, there was a trend for peak cluster extent to be larger in the right than left temporal lobe (P(binomial) = 0.065).

3.2.2. Task-driven network

As can be observed in Table 4, the lateralization analysis on the task-driven network of emotional prosodic perception revealed only one significantly lateralized ALE-cluster that was located in the posterior right STG (peak Talairach coordinates (44.25, −38.25, 8.75), $Z = 2.14$, volume = 880 mm^3) at $p = 0.05$ uncorrected. No ALE-cluster survived the more conservative $p = 0.001$ threshold. For all included studies reporting clusters in the temporal lobe, peak activation and extent were greater in the right than the left hemisphere (P(binomial) < 0.05).

Table 4. Locations of significant clusters for the *lateralization* analyses.

Cluster	Cluster size (mm³)	Anatomical Label	Peak voxel coordinates x	y	z	*Z*-value	No. studies
***Stimulus*-driven lateralization analysis:**							
1	624	R Transverse Temporal Gyrus	50	−24	10	1.95	3
		R STG	42	−26	4	1.87	
		R Transverse Temporal Gyrus	40	−24	12	1.82	
		R STG	38	−26	6	1.80	
2	192	L Insula	−42	−14	4	2.01	0
***Task*-driven lateralization analysis:**							
1	880	R STG	44.25	−38.25	8.75	2.13	3

L = left hemisphere; *R*=Right hemisphere. All reported ALE-clusters were thresholded at $p < 0.05$ uncorrected and a cluster extent of 100 mm³

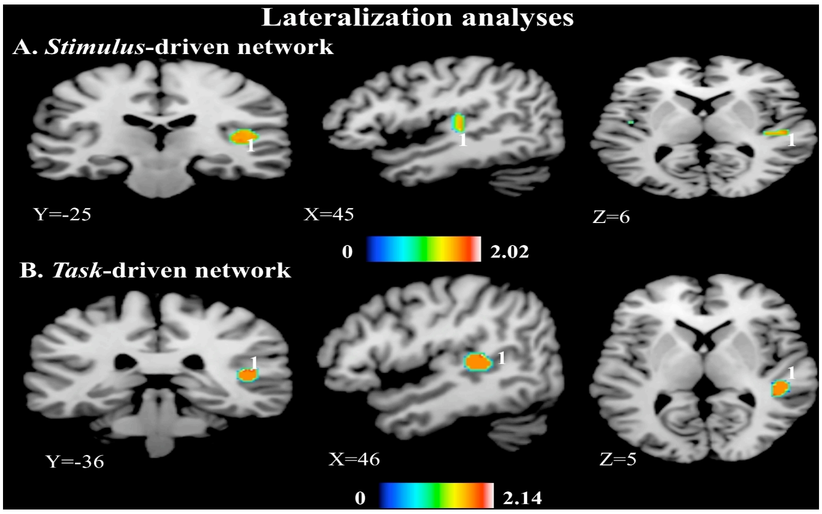

Figure 3. Results of the lateralization analyses. All ALE-clusters were significant at p = 0.05 uncorrected and had a minimum cluster size of 100 mm³. Clusters to which at least three studies have directly contributed are marked with a number which corresponds to the cluster number in Table 4.

4. Discussion

Using activation-likelihood estimation (ALE) meta-analysis, we investigated which neural networks are reliably engaged in stimulus-driven and task-driven emotional prosody perception and to which extent these networks show hemispheric asymmetry. Although all reported clusters in the main analyses passed the criterion of a less than 5% family-wise probability that the convergence found could have occurred by chance, the maximum percentage of experiments directly contributing to a cluster was 41. Exactly the same maximum percentage of contributing experiments was found in a recent ALE-meta analysis of visual food-cue processing, which the authors qualified as a 'moderate' level of convergence as compared to other ALE-meta analyses (Van der Laan, De Ridder, Viergever, & Smeets, 2011). As these authors pointed out, a recent review reported that the test-retest reliability of fMRI results as expressed by the intraclass correlation coefficient is 0.5 (Bennett & Miller, 2010) which implies an even lower correspondence of results across *different* experiments (tasks) that intend to measure the same neurocognitive process (as has been done in the present investigation). Hence, a moderate degree of convergence across experiments as found in the present meta-analysis could have been expected.

In addition to heterogeneity induced by non-prosody specific variables that differed between included studies such as scanning parameters (e.g. whether measures were taken to reduce echo planar imaging noise) and analysis details, there was variation in factors specific to the subject at hand that might have reduced convergence. For instance, in the case of stimulus-driven processing it can be observed in Table 1 that there is variation in the exact task instruction, varying from passive listening to discrimination and categorization. These different task instructions might engage the different components of the stimulus-driven network, lowering overall convergence. This also holds for the task-driven analysis: as can bee seen in Table 2 there was variation in the exact task instructions for both the active and control conditions, likely inducing variation in the exact locations of activation clusters found across studies. Another source of variation is whether stimulus-driven processing was implicit or explicit: the few studies that have directly compared implicit to explicit stimulus-driven emotional prosody processing, suggest that these two modes of processing might tax different components of the emotional prosody perception network (Bach et al., 2008; Frühholz, Ceravolo, & Grandjean, 2011). A third source of variation might be the number and kind of emotions included in the study, which varied from only angry and neutral to seven different emotional categories, as can been seen in Table 1. As there is evidence for emotion specific processing in the auditory cortex (Ethofer, Van de Ville, Scherer, & Vuilleumier, 2009) and since variation in the number of emotions in categorization tasks might induce variation in the degree to which the working memory system is taxed (Hoekert, Kahn, Pijnenborg, & Aleman, 2007) this factor might have compromised convergence of activation clusters across studies.

Unfortunately, the number of neuroimaging studies on emotional prosody perception published to date was too low to permit a reliable formal meta-analytic contrast of these moderating variables. In order to only discuss the part of the emotional prosody perception network that converges robustly across studies (and therefore represents the general network relatively independent of differential engagement of specific neural structures associated with the mentioned moderators), in the following only ALE-clusters will be addressed to which at least three of the included experiments directly contributed.

4.1. Stimulus-driven network

Contrasting the perception of emotional prosodic categories to neutral prosody (or speech devoid of prosody) engages areas involved in the processing of acoustic properties that differ between emotional and non-emotional prosody but likely also engages systems involved in more abstract processing of emotional information.

The clusters of highest convergence for the stimulus-driven contrast lay in the bilateral mid-STG, extending posteriorly and laterally with respect to HG. In the right hemisphere the STG cluster also extended medially into the transverse temporal (Heschl's) gyrus proper, which likely houses the human primary auditory cortex (Galaburda & Sanides, 1980). Indeed, as the mid-STG is located within the zone of purportedly speech-specific superior temporal cortex (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) and shows sensitivity to the emotional quality of speech, the mid-STG (and its posterolateral extension) has recently been coined the 'emotional voice area (EVA)' (Ethofer et al., 2011). Previous studies have shown that the mid-STG is sensitive to multiple acoustic properties relevant for emotional prosody perception (Wiethoff et al., 2007) and is active independently of task demands (Ethofer et al., 2009). This suggests that the mid-STG (and its posterolateral extension) is involved in the automatic integration of emotional prosodic acoustic cues relevant for the perception of emotional prosody.

In both hemispheres significant convergence in the STG extended from the mid-STG to the posterior STG, which produced a separate ALE-cluster in the left hemisphere. According to the Brück, Kreifelts, and Wildgruber (2011) model of emotional prosody perception, the posterior superior temporal cortex (STC) is involved in more abstract representation of emotional prosody (stage 2). Moreover, as will be elaborated further in the section on the task-dependent emotional prosody perception network, it appears that activity in the posterior STC during auditory processing can be influenced by task demands (e.g. Brechmann & Scheich, 2005) making it likely that this area participates later in the prosodic perception process than the mid-STG in line with the model proposed by Brück et al. (2011a).

As could be expected from both models of emotional prosody perception, significant clusters of convergence were found in the bilateral IFG. Both models of emotional prosody perception implicate the bilateral IFG in a final multimodal stage where emotional prosody is explicitly evaluated and integrated with other layers of speech. Lesion studies (Breitenstein, Daum, & Ackermann, 1998; Hornak et al., 2003; Hornak, Rolls, & Wade, 1996) and a more a recent TMS-study (Hoekert, Vingerhoets, & Aleman, 2010) have furthermore indicated that the inferior frontal cortex is not only sufficient but also necessary for adequate emotional prosody perception. If the IFG are indeed involved in a final multimodal explicit evaluative phase, it would be expected that the bilateral IFG are involved in the explicit evaluation of emotion in other modalities such as emotional faces or emotional scenes as well, which has been confirmed by recent meta-analyses of the respective neuroimaging literature (for emotional faces see Fusar-Poli et al., 2009a; for emotional faces and scenes see Sabatinelli et al., 2011).

In the right inferior frontal gyrus the ALE-cluster extended into the dorsal anterior insula. Of the two proposed models of emotional prosody perception proposed by Kotz and Paulmann (2011) and Brück, Kreifelts and Wildgruber (2011) only the latter proposes a role for the anterior insula, namely in automatic (implicit) processing of emotion. However, insula activation is not consistently found in fMRI studies during implicit processing of emotional prosody (Bach et al., 2008; Ethofer et al., 2011; Frühholz, Ceravolo, & Grandjean, 2011; Grandjean et al., 2005) but see (Morris, Scott, & Dolan, 1999; Mothes-Lasch, Mentzel, Miltner, & Straube, 2011). Furthermore, the neuroimaging literature suggests that there is functional differentiation within the anterior insula. For instance, a recent large scale meta-analysis of the functional neuroimaging literature reports that the dorsal insula might be an integration zone important for general maintenance of the task-set while the ventral insula is specifically involved in the recognition of emotion and empathy (Kurth, Zilles, Fox, Laird, & Eickhoff, 2010). Although the exact role of insular clusters in the emotional prosody perception process is admittedly speculative at this time, the dorsal insula might control general prosodic task performance while the ventral insula might specifically support emotion recognition and empathy during emotional prosody perception.

Although two ALE clusters were found in the basal ganglia (BG), more specifically the right medial globus pallidus and caudate body and the left claustrum, these clusters did not survive our additional criterion that at least three studies should have directly contributed to a cluster. Given the relatively low number of experiments in the analysis, this might have been a result of power being just too low to detect these clusters above threshold. Notably, no significant convergence was found in the amygdala, a structure often claimed to be essential for emotional processing. Although both models of emotional prosody perception implicate the amygdala in emotional prosody perception, namely in the detection of emotional salience (Kotz & Paulmann, 2011) and in automatic (implicit) emotional prosody perception (Brück, Kreifelts, & Wildgruber, 2011) lesion studies have failed to find an effect of amygdala damage on emotional prosody perception task performance (Adolphs, 2001; Adolphs & Tranel, 1999), suggesting that the amygdala might not be crucial for emotional prosody perception. Possibly the amygdala is only engaged in the emotional prosody perception process under specific task conditions, preventing robust convergence of peak activity in the amygdala across studies as analyzed by the present meta-analysis. Another possibility is that the amygdala habituates quickly to emotional prosody stimulation (Wiethoff, Wildgruber, Grodd, & Ethofer, 2009) reducing the ability to detect the amygdala activity over prolonged periods of stimulation as is typical for fMRI experiments. Lastly, the amygdala response to emotional prosody may vary considerably between individuals depending on personality traits as has been shown in a recent study (Brück, Kreifelts, Kaza, Lotze, & Wildgruber, 2011).

4.2. Task-driven network

Contrasting active processing of emotional prosody to processing of the same acoustic material while directing attention to a different attribute (such as semantics or speaker identity) captures more abstract later stages of the emotional prosody perception process. Both proposed models of emotional prosody perception implicate the bilateral IFG in the evaluation of emotional prosody (stage 3) but differ in the exact neuroanatomical underpinnings of stage 2 (integration of acoustic information into an emotional percept or representation of meaningful suprasegmental acoustic sequences). While the Brück, Kreifelts and Wildgruber (2011) model proposes that stage 2 processing occurs in the posterior superior temporal cortex, the Kotz and Paulmann model (2011) postulates that this more abstract stage of processing is supported by the anterior STS.

The most concurrent ALE-cluster for the task-driven network lay in the posterior STG/STS immediately posterior to HG, a region that is sometimes referred to as the 'planum temporale' (Westbury et al., 1999). Contrarily, no significant convergence across studies was found in the anterior STS (not even sub-threshold) for task-driven emotional prosody processing. This difference in convergence does not seem attributable to methodological factors such as a difference in susceptibility artifact in the BOLD signal (Devlin et al., 2000) as both regions lay in the superior temporal region, clear from air filled cavities. Therefore, stage 2 emotional prosody processing is more likely to be supported by the posterior STG/STS as proposed by Brück, Kreifelts, and Wildgruber (2011) than the anterior STS as proposed by Kotz and Paulmann (2011). The posterior auditory association cortex in the posterior STG seems (given its associative nature) well equipped for the proposed integrative function of stage 2 emotional prosody processing. Furthermore, a recent meta-analysis has shown that the posterior auditory association cortex is specialized in the processing of spectral information (Samson, Zeffiro, Toussaint, & Belin, 2010). As spectral parameters are important for the recognition of emotional prosody (Scherer, 2003) the posterior auditory association cortex might contribute to emotional prosody perception through enhanced spectral processing. Furthermore, neuroimaging studies have shown that activity of the posterior STG can be (top-down) modulated by task demands (Ahveninen et al., 2006; Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005; Jäncke, Mirzazade, & Shah, 1999), which is clearly compatible with a role for the posterior STG in task-dependent emotional prosody perception as found in the present analysis and as proposed by Brück et al. (2011a).

In line with both models, significant convergence across studies was found in the right IFG for task-dependent processing of emotional prosody. Although significant ALE-values were also found in the left IFG homolog, these voxels did not survive cluster thresholding which might be a consequence of the relatively low power of the task-dependent analysis. As was already pointed out in the discussion of the IFG clusters in the stimulus-driven network, the IFG likely plays a role in explicit multimodal emotional evaluation and the integration of emotional prosody with other layers of speech. Interestingly, a recent imaging study suggests that there is functional connectivity between the right posterior temporal cortex and the right IFG during emotional prosody processing (Ethofer et al., 2006). Although in their dynamic causal model a forward projection from the posterior temporal cortex to the IFG was more probable than vice versa, another interesting possibility might be that activity in the posterior auditory association cortex can be selectively enhanced to flexibly augment processing of task-relevant (spectral) acoustic properties through top-down mediation by the IFG as has been suggested by Leitman et al. (2010).

4.3. Hemispheric specialization for emotional prosody

As outlined in the introduction, both proposed models of emotional prosody perception allow for hemispheric specialization in multiple stages of the emotional prosody perception process. More specifically, both models postulate (relative) right-hemispheric specialization for the first and second stage of emotional prosody perception while only the model by Kotz and Paulmann (2011) additionally leaves open the possibility of right-hemispheric specialization for emotional evaluation in the final stage. Indeed, a recent meta-analysis of the lesion literature found a statistically robust greater emotional prosodic performance degradation following right than left hemispheric damage (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011), suggesting that the right cerebral hemisphere is of relatively greater importance than the left in the decoding of emotional prosody. However, due to the typically limited spatial specificity of acquired lesions this meta-analysis could not address at which intrahemispheric loci this right-hemispheric specialization might arise, an issue the present meta-analysis is better equipped to address.

The lateralization analysis for the stimulus-driven network revealed right-lateralized convergence of ALE-clusters in the medial transverse temporal gyrus. However, the cluster only emerged at a more lenient threshold and therefore has to be interpreted with caution. Nevertheless, this result suggests that hemispheric specialization might emerge very early at the level of the (medial) HG (where the human PAC is likely to be located). Previous studies outside the prosody domain have indeed found functional lateralization at the level of HG. For instance, it has been found that the left and right HG may already show an advantage for temporal (Jamison, Watkins, Bishop, & Matthews, 2006; Warrier et al., 2009) and spectral (Mathys, Loui, Zheng, & Schlaug, 2010; Warrier et al., 2009; Zatorre, 1988) processing, respectively. As spectral variation is an important acoustic dimension for recognizing emotional prosody (Scherer, 2003), the right HG might make a relatively greater contribution to emotional prosody perception than its left hemispheric homolog through enhanced spectral processing.

The task-driven lateralization analysis revealed a right-lateralized ALE-cluster in the posterior STG, but again only at the more lenient threshold warranting caution in interpreting this result. Previous studies have shown that the right posterior STG might be specialized in pitch processing (Hyde, Peretz, & Zatorre, 2008; Johnsrude, Penhune, & Zatorre, 2000). Hence the right posterior STG, similarly to the right HG, might contribute disproportionally as compared to the left posterior STG to the perception of emotional prosody through its superior pitch-processing capability. Furthermore, in line with a task-dependent role of the posterior STG in emotional prosody perception as found in the present investigation, previous studies have shown that lateralization of posterior STG activity during auditory processing is dependent on task demands (Brechmann & Scheich, 2005; Geiser, Zaehle, Jancke, & Meyer, 2007). For instance, Brechmann and Scheich (2005) showed that when subjects had to categorize the pitch direction of frequency-modulated tones, the right posterior STG was more active than the left, but when the duration of the same material had to be categorized the hemispheric specialization shifted to the left STG. As these authors suggest, the left or right posterior STG might be recruited through top-down mediation to dynamically enhance task-relevant auditory processing. In the case of emotional prosodic processing, the right posterior STG might be similarly recruited to enhance spectral processing to support emotional prosody perception.

The present meta-analyses therefore suggest that the statistically robust relative right-hemispheric specialization for emotional prosody perception found previously in the lesion literature (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011) might be explained by multiple (relative) right-hemispheric asymmetries in the primary and associative auditory cortex, possibly corresponding to the proposed first and second stage of emotional prosody perception, respectively. In the latter case increased activity in the right posterior auditory association cortex might reflect increased effort to extract acoustic properties relevant to emotional prosody perception through top-down modulation by the right IFG (Leitman et al., 2010), which would render hemispheric specialization for emotional prosody perception an interactive dynamic process, although this hypothesis admittedly requires further explicit evaluation. These hemispheric asymmetries early in the emotional prosody perception process are in line with cue-dependent hypotheses (e.g. Van Lancker & Sidtis, 1992) of hemispheric specialization for emotional prosody perception. More generally, the results suggest that hemispheric specialization for emotional prosody can be explained by specialization of the left and right cerebral hemisphere for the processing of more basic (non-prosody-specific) acoustic dimensions such as spectral versus temporal processing (Zatorre & Belin, 2001) or integration over small versus large temporal windows (Poeppel, 2003), respectively. Such an early auditory bottom-up account of hemispheric specialization for emotional prosody has, in fact, been suggested previously (Ethofer et al., 2011; Obleser, Eisner, & Kotz, 2008; Robin, Tranel, & Damasio, 1990) and can also explain why no right hemispheric specialization has been found for visual emotional (i.e. facial) expression in a recent large-scale meta-analysis of the neuroimaging literature (Fusar-Poli et al., 2009b).

4.4. Strengths and limitations

To the best of the authors' knowledge, this is the first quantitative meta-analysis of the neuroimaging literature on emotional prosody perception. Individual imaging studies generally suffer from small sample size and the results reflect study specific-details in the experimental design. Quantitative meta-analysis alleviates these problems and enables one to study which neural network is reliably associated with a neurocognitive process with high statistical power. A limitation of ALE meta-analysis is that only the peak-coordinates of activation clusters are considered and not the effect size or extend of activation. However, when an attempt was made to compensate for this weakness by manually checking the extent and magnitude of lateralized clusters for the lateralization analyses, it was found that these measures largely confirmed the coordinate-based analyses. A further limitation was that due to the limited number of neuroimaging papers on emotional prosody perception published to date, it was not possible to formally test for the influence of moderating variables that have been suggested in the literature. Hence, the exact role of factors such as appraisal level and specific emotions in the neural processing of emotional prosody warrants further investigation.

4.5. Conclusion

In sum, the present meta-analyses implicate a bilateral temporofrontal network of areas including the PAC, mid-STG, post-STG and the IFG in the perception of emotional prosody, largely confirming recently proposed models of emotional prosody perception. Activation likelihood was larger in the posterior-STG than the anterior-STS, suggesting that the proposed second stage of emotional prosody perception is more likely to be supported by the posterior-STG, as suggested by Brück, Kreifelts and Wildgruber (2011) than the anterior-STS as suggested by Kotz and Paulmann (2011). Concerning hemispheric specialization, multiple right lateralized ALE-clusters were found in the auditory cortex but only at a more lenient threshold. These results suggest that the relative right-hemispheric specialization for emotional prosody perception found previously in the lesion literature (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011) might be a multi-stage (and possibly interactive) process that might be best explained by relatively early hemispheric specialization for prosody-relevant acoustic processing. However, future meta- (or mega-) analyses with a larger number of experiments are needed to conclusively test this hypothesis.

6. References[1]

Adolphs, R., Tranel, D., & Damasio, H. (2001). Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology*, *15*, 396-404.

Adolphs, R., & Tranel, D. (1999). Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia*, 37, 1285-1292.

Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., . . . Belliveau, J.W. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Science of the United States of America*, *103,* 14608-14613.

*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.

*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.

Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences, 1191*, 133-155.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience, 8*, 389-395.

Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind--evolution ontogeny brain mechanisms and psychopathology. *Neuroscience & Biobehavioral Reviews, 30*, 437-455.

Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex, 15,* 578-587.

Breitenstein, C., Daum, I., & Ackermann, H. (1998). Emotional processing following cortical and subcortical brain damage, contribution of the fronto-striatal circuitry. *Behavioural Neurology, 11*, 29-42.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011a). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Phyics of Life Reviews, 8,* 383-403.

*Brück, C., Kreifelts, B., Kaza, E., Lotze, M., & Wildgruber, D. (2011b). Impact of personality on the cerebral processing of emotional prosody. *NeuroImage, 58,* 259-268.

*Buchanan, T.W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jankce, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research, 9*, 227-238.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jäncke, L. (2005). Scanning silence: Mental imagery of complex sounds. *NeuroImage, 26,* 1119-1127.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E.,. . . Tyler, L. K. (2000). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *Neuroimage, 11,* 589-600.

Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A., & Ackermann, H. (2008). Understanding the emotional expression of verbal interjections: A functional MRI study. *Neuroreport, 8,* 1751-1755.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology, 6,* 178-190.

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: Random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping, 30,* 2907-2926.

Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., & Fox, P. T. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *Neuroimage,* 57, 938-949.

*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.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., & Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody, A dynamic causal modeling study. *NeuroImage, 21,* 580-587.

*Ethofer, T., Bretscher, J., Geschwind, 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. (2009). 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. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology, 19,* 1028-1033.

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Petersen, S. E. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. *Nuclear Science Symposium and Medical Imaging Conference, 3*, 1813-1817.

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

Fusar-Poli P., Placentino A., Carletti, F., Allen, P., Landi, P., Abbamonte, M., . . . Politi, P. L. (2009b). Laterality effect on emotional faces processing, ALE meta-analysis of evidence. *Neuroscience Letters, 452,* 262-267.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., . . . & Politi, P. (2009a). Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of Psychiatry and Neuroscience, 34,* 418-432.

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology, 190,* 597-610.

*Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., & Li, X. (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Human Brain Mapping, 18,* 149-157.

Geiser, E., Zaehle, T., Jäncke, L., & Meyer, M. (2007). The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience, 20,* 541-552.

Genovese, C. R., Lazar, N. A., Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15,* 870-878.

*George, M. S., Parekh, P. I., Rosinsky, N., Ketter, T. A., Kimbrell, T. A., Heilman, K., . . . Post, R. M. (1996). Understanding emotional prosody activates right hemisphere regions. *Archives of neurology, 53,* 665-670.

Gobras, M., Beaton, S., & Eickhoff, S. B. (2012). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human Brain Mapping, 33,* 431-454.

*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.

Hoekert, M., Kahn, R. S., Pijnenborg, M., & Aleman, A. (2007). Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophrenia Research, 96,* 135-145.

Hoekert, M., Vingerhoets, G., & Aleman, A. (2010). Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception, an rTMS study. *BMC Neuroscience, 11,* 93.

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,* 1691-1712.

Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia, 34,* 247-261.

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia, 46,* 632-639.

*Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., Itoh, K., Kato, T., . . . Nakamura, K. (1997). Vocal identification of speaker and emotion activates different brain regions. *NeuroReport, 8,* 1809-2812.

Jäncke, L., Mirzazade, S., Shah, J. N. (1999). Attention modulates activity in the primary and the secondary auditory cortex: A functional magnetic resonance imaging study in human subjects. *Neuroscience Letters, 266,* 125-128.

Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex ,16,* 1266-1275.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain, 123,* 155-163.

*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.

Kotz, S. A., & Paulmann, S. (2011). Emotion language and the brain. *Language and Linguistics Compass, 5,* 108-125.

Kurth, F., Zilles, K., Fox, P., Laird, A., & Eickhoff, S. (2010). A link between the systems: Functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Structure and Function, 214,* 519-534.

Laird, A. R., Fox, P. M., Price, C. J., Glahn, D.C., Uecker, A. M., Lancaster, J. L., . . . Fox, P. T. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping, 25,* 155-164.

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

*Leitman, D. I., Wolfm D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., . . . Gur, R. C. (2010). "It's not what you say but how you say it": a reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience*, *19,* 1.

Mathys, C., Loui, P., Zheng, X., & Schlaug, G. (2010). Non-invasive brain stimulation applied to Heschl's gyrus modulates pitch discrimination. *Frontiers in Psychology, 1*, 193.

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

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. R, & Straube, T. (2011). Visual attention modulates brain activation to angry voices. *The Journal of Neuroscience, 31,* 9594-9598.

Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *The Journal of Neuroscience, 28,* 8116-8123.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41,* 245-255.

Robin, D. A., Tranel, D., & Damasio, H. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and Language, 39*, 539-555.

*Rota, G., Veit, R., Nardo, D., Weiskopf, N., Birbaumer, N., & Dogil, G. (2008). Processing of inconsistent emotional information: An fMRI study. *Experimental Brain Research, 186,* 401-407.

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., Beck, S., & Jeffries, J. (2011). Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage, 54,* 2524-2533.

Samson, F., Zeffiro, T. A., Toussaint, A., & Belin, P. (2010). Stimulus complexity and categorical effects in human auditory cortex: An activation likelihood estimation meta-analysis. *Frontiers in Psychology, 1*, 241.

*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.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227-256.

Scherer, K. R., Banse, R., & Wallbott, H.G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76-92.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10,* 24-30.

*Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., & Friederici, A. D. (2008). When vocal processing gets emotional: An the role of social orientation in relevance detection by the human amygdala *NeuroImage, 40,* 1402-1419.

Silk, J. B. (2007). Social components of fitness in primate groups. *Science, 317,* 1347-1351.

Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain: 3-d proportional system: An approach to cerebral imaging. Stuttgart: Thieme.

Turkeltaub, P. E., & Coslett, B. H. (2010). Localization of sublexical speech perception components. *Brain and Language, 114*, 1-15.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage, 16,* 765-780.

Turkeltaub, P.E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2011). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, *33,* 1-13.

Trask, R. L. (1996). *A dictionary of phonetics and phonology*. London: Routledge.

Van Bezooijen R., Otto, S.A., & Heenan, T.A. (1983). Recognition of vocal expressions of emotion. *Journal of Cross-Cultural Psychology, 14*, 387-406.

Van der Laan, L. N., De Ridder, D. T. D., Viergever, M. A., & Smeets, P. A. M. (2011). The first taste is always with the eyes: A meta-analysis on the neural correlates of processing visual food cues. *NeuroImage, 55*, 296-303.

Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. Papers in Linguistics: *International Journal of Human Communication, 13*, 201-277.

Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research, 35,* 963-970.

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., & Kraus, N. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *The Journal of Neuroscience, 29,* 61-69.

Westbury, C. F., Zatorre, R. J., & Evans, A.C. (1999). Quantifying variability in the planum temporale: A probability map. *Cerebral Cortex, 9,* 392-405.

*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.

Wiethoff, S., Wildgruber, D., Grodd, W., & Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *NeuroReport, 20,* 1356-1360.

*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.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology, 11,* 277-281.

Witteman J, Van IJzendoorn, M. H, Van de Velde D, Van Heuven, V. J., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia, 49,* 3722-3738.

Wittfoth, M., Schröder, 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.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *The Journal of the Acoustical Society of America, 84*, 566-572.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946-953.

[1] References marked with an asterisk (*) were included in the meta-analyses