doi:10.1093/scan/nst180 SCAN (2014) 9, 1819–1827

'Inner voices': the cerebral representation of emotional voice cues described in literary texts

Carolin Brück,[1,2] Benjamin Kreifelts,[1] Christina Gößling-Arnold,[2,3] Jürgen Wertheimer,[2,3] and Dirk Wildgruber[1,2]
[1]Department of Psychiatry and Psychotherapy, Eberhard Karls University, Tübingen 72076, Germany, [2]Werner Reichardt Centre for Integrative Neuroscience (CIN), Tübingen 72076, Germany and [3]Department of Comparative Literature, Eberhard Karls University, Tübingen 72074, Germany

While non-verbal affective voice cues are generally recognized as a crucial behavioral guide in any day-to-day conversation their role as a powerful source of information may extend well beyond close-up personal interactions and include other modes of communication such as written discourse or literature as well. Building on the assumption that similarities between the different 'modes' of voice cues may not only be limited to their functional role but may also include cerebral mechanisms engaged in the decoding process, the present functional magnetic resonance imaging study aimed at exploring brain responses associated with processing emotional voice signals described in literary texts. Emphasis was placed on evaluating 'voice' sensitive as well as task- and emotion-related modulations of brain activation frequently associated with the decoding of acoustic vocal cues. Obtained findings suggest that several similarities emerge with respect to the perception of acoustic voice signals: results identify the superior temporal, lateral and medial frontal cortex as well as the posterior cingulate cortex and cerebellum to contribute to the decoding process, with similarities to acoustic voice perception reflected in a 'voice'-cue preference of temporal voice areas as well as an emotion-related modulation of the medial frontal cortex and a task-modulated response of the lateral frontal cortex.

Keywords: emotion; fMRI; literature; non-verbal communication; voice

INTRODUCTION

With only a slight change in the sound of their voices human beings are able to communicate a wealth of information: modulations of voice characteristics such as pitch, loudness, voice quality or tempo allow listeners to uncover attitudes, intentions or emotions behind spoken words (Banse and Scherer, 1996; Szameitat *et al.*, 2009; Sauter *et al.*, 2010)—crucial knowledge that aids 'survival' in our social environment.

For decades now research has aimed at understanding how voice signals are used to decipher affective meaning encoded in the sound of a voice. Particularly studies regarding the processing of affective voice cues such as speech prosody or laughter have contributed greatly to our current understanding of how the brain analyses, integrates and evaluates vocal expressions of emotions (Ackermann *et al.*, 2004; Schirmer and Kotz, 2006; Wildgruber *et al.*, 2006; Meyer *et al.*, 2007; Wildgruber *et al.*, 2009; Brück *et al.*, 2011b).

However, vocal signals may not only serve as a valuable guide in any day-to-day conversation, rather their role may also translate to other forms of communication as well. Considering narrative literature, for example, written *descriptions* of affective voice cues—just as their acoustic counterparts—may frequently be used to convey emotions and may similarly lead the readers to a better understanding of the characters described to send these signals.

Given the suggested similarities in the functional roles of vocal emotional cues both in literature and day-to-day interactions one might ask whether such similarities emerge with respect to cerebral mechanisms employed in the decoding process as well.

Received 13 December 2012; Revised 12 September 2013; Accepted 28 December 2013
Advance Access publication 5 January 2014

The authors would like to thank Cyril Belica, Holger Keibel, Marc Kupietz and Rainer Perkuhn of the Institute for the German Language (IDS) Mannheim for their support in determining measures of lexical complexity and syntactic complexity employed in the stimulus selection process.

This work was supported by the Werner Reichardt Centre for Integrative Neuroscience (CIN), Tübingen (CIN 2009-17).

Correspondence should be addressed to Carolin Brück, Department of Psychiatry and Psychotherapy, Eberhard Karls University, Calwerstraße 14, 72076 Tübingen, Germany. E-mail: Carolin.Brueck@med.uni-tuebingen.de

To address this question, this study aimed at exploring brain responses associated with the decoding of voice signals described in literary texts. Building on the hypothesis that similarities may emerge with respect to task- or emotion-driven as well as voice-sensitive brain responses, frequently reported for the decoding of affective voice cues (Wildgruber *et al.*, 2006, 2009; Brück *et al.*, 2011b), particular emphasis was placed on the evaluation of task- and emotion-related as well as voice-sensitive modulations of brain activation.

Analyses of voice-sensitive effects focused on the temporal voice area (TVA), a brain region located in the superior temporal cortex suggested to exhibit a preferential responding to human voices (Belin *et al.*, 2000; Belin *et al.*, 2002; Bestelmeyer *et al.*, 2011) and to play a role in a broad range of voice-related abilities including the perception of affective voice cues (Ethofer *et al.*, 2009b).

Considering emotion-related effects on the other hand, published data suggest emotion-driven modulations of activation for several structures implicated in emotional voice decoding such as the amygdalae (Wiethoff *et al.*, 2009), the anterior rostral medial frontal cortex (arMFC) (Brück *et al.*, 2011a) and the TVA (Ethofer *et al.*, 2012). While, particularly with respect to TVA activation, such reported emotion-driven increases in responding may reflect effects unique to the decoding of voice-based acoustic information, modulations of the amygdalae and arMFC resemble results documented for a variety of emotion perception tasks (e.g. facial emotion processing; Kesler-West *et al.*, 2001; Fusar-Poli *et al.*, 2009; Sabatinelli *et al.*, 2011). Latter findings, in turn, may outline a cue-independent contribution of both structures to perceptual mechanisms more commonly involved in deciphering other people's states of the mind (Zald, 2003; Amodio and Frith, 2006; Peelen *et al.*, 2010).

Aside effects of emotion, however, brain responses to affective voice cues have also been described to differ depending on the task instructions they are presented with. Compared to a more implicit processing of emotions encoded in a voice (e.g. via task instructions that distract attention away from expressed emotions), instructions to focus on the explicit evaluation of emotional information have been documented to increases activation within the lateral frontal lobe (Wildgruber *et al.*, 2004, 2005; Ethofer *et al.*, 2006, 2009a), a brain

region assumed to contribute to meaning analysis across a variety of emotion-related tasks (Kober *et al.*, 2008; Wager *et al.*, 2010; Lindquist *et al.*, 2012).

METHODS

Participants

Twenty-two volunteers (11 female, all right-handed, all native speakers of German, $mean_{age} = 24.95$, s.d. = 3.70) consented to participate. Participants were screened to exclude hearing or vision impairments as well as past or present psychiatric or neurological disorders, or current medical treatment that might affect brain function.

Ethics statement

The experiment was conducted in accordance with the ethical principles expressed in the Declaration of Helsinki, and the study protocol was reviewed and approved by the local ethics committee. All participants received detailed information about the purpose and procedure of the study, and gave written consent prior to involvement in this research.

MRI data acquisition

Magnetic resonance imaging (MRI) data were acquired on a 3 T MRI scanner (Tim Trio, Siemens, Erlangen, Germany) equipped with a 12-channel head coil. All functional images were obtained using a BOLD-sensitive echo planar imaging sequence covering the whole brain with 30 slices (slice thickness: 4 mm thickness + 1 mm gap, FoV = 192 mm × 192 mm, 64 × 64 matrix, voxel size 3 × 3 × 4 mm^3, TR = 1700 ms, TE = 30 ms and flip angle = 90°). In addition to functional data, high-resolution structural images were collected from each participant as anatomical reference (magnetization prepared rapid gradient echo: TR = 2300 ms, TE = 2.96 ms, 176 slices, slice thickness: 1 mm, FoV = 256 mm × 256 mm).

Stimulus material, tasks and procedure

Main experiment

Stimulus material comprised a set of 78 text samples. Text samples were selected from an original pool of 212 face and voice descriptions gathered from novels and narratives published in German. The selection was based on a series of behavioral experiments and text analyses conducted to evaluate key features of each text sample such as the emotional valence, emotional arousal, aesthetic value, text length, lexical complexity or syntactic complexity. The selection procedure aimed at composing two subsets of text stimuli, one set of voice descriptions, one set of face descriptions, each matched with respect to the key characteristics reported above. Of the 78 texts chosen in the process 39 described vocal expressions and 39 facial expressions. Within each sub-set one-third of the text samples conveyed emotionally neutral expressions, whereas the remaining two-thirds communicated either a positive ($n = 13$) or a negative ($n = 13$) emotional state. Examples of texts included in the sub-sets are provided in Table 1. Ratings of the respective text characteristics are summarized in Table 2.

The selected text samples were used to devise two different tasks: Task 1 targeted explicit emotion processing by asking participants to focus on emotions expressed in the texts and rate each of the 78 text samples with respect to the valence of the communicated emotions (from here on referred to as emotion judgment task or emo). Participants were offered a choice of four different response alternatives: ++ (highly positive), + (positive), − (negative) and −− (highly negative). Task 2 targeted a more implicit processing of emotional information by diverting attention away from expressed emotions through instructions to focus on the evaluation of a text's aesthetic value (from here on referred to as aesthetics judgment task or aes). Participants were asked to give their opinion on how well they thought each of the presented 78 text samples was written choosing one of four answers alternatives: ++ (very well), + (well), − (poorly) and −− (very poorly).

Text samples were back-projected onto a translucent screen placed in the back of the scanner bore and viewed by the participants via a

Table 1 Examples of text stimuli employed in the main experiment

Cue	Emotional state	Sample text	English translation
Facial	Positive	Sie zeigte, für die Dauer eines Herzschlags, ihr Lachen von einer Wange zur anderen. *Kirchhoff, Bodo (2002) Die Weihnachtsfrau. Frankfurt a. M.: Fischer Taschenbuch, p. 32.*	For the duration of a heartbeat, she smiled from cheek to cheek. *English translation provided by the authors.*
	Neutral	[Er] bedeckte die Oberlippe mit der Unterlippe und behielt diese Stellung [. . .]. *Gogol, Nikolai Vasilevich (n.d., 2nd ed., approx. 1960). Tschitschikows Abenteuer oder Tote Seelen. German translation by Elisabeth and Wladimir Wonsiatsky. München: Simon-Herold-Verlag, p. 20.*	[H]e pursed his lips, and retained this attitude unchanged [. . .]. *Gogol, Nikolai Vasilevich (2008) Dead Souls. translated by D. J. Hogarth, p. 15.*
	Negative	Krespel schnitt ein Gesicht, als wenn jemand in eine bittere Pomeranze beißt, und dabei aussehen will, als wenn er süßes genossen; aber bald verzog sich dies Gesicht zur graulichen Maske [. . .]. *Hoffmann, Ernst Theodor Amadeus (2001) Rat Krespel. In: Steinecke, H. & Segebrecht, W. (Eds.) Sämtliche Werke. Frankfurt a. M.: Deutscher Klassiker Verlag. Bd. 4, p. 44.*	Krespel made a face like someone biting into a sour orange who wants to look as if it were a sweet one; but soon his expression changed into a horrifying mask [. . .]. *Hoffmann, E.T.A. (1972) Councillor Krespel. In: Tales by E.T.A. Hoffmann. Edited and translated by Leonard J. Kent and Elizabeth C. Knight. University of Chicago Press, p. 129f.*
Vocal	Positive	Als sie sprach, klang ihre Stimme sanft und kehlig und mit einem italienischen Akzent behaftet. *Brown, Dan (2003) Illuminati. German translation by Axel Merz. Bergisch Gladbach: Bastei Lübbe, p. 75.*	When she spoke, her voice was smooth – a throaty, accented English. *Brown, Dan (2001) Angels and Demons. London: Corgi, p. 70.*
	Neutral	Die Frau [. . .] sprach langsam und leise in einer Sprache, die Julie sich nicht erinnern konnte, jemals gehört zu haben. *Hoffmann, Ernst Theodor Amadeus (1992) Lebens-Ansichten des Katers Murr. In: Steinecke, H. & Segebrecht, W. (Eds.) Sämtliche Werke. Frankfurt a. M.: Deutscher Klassiker Verlag. Bd. 5, p. 218f.*	The women [. . .] spoke slowly and quietly in a language Julie couldn't remember to have ever heard before. *English translation provided by the authors.*
	Negative	Er sprach sehr langsam, und die Worte schienen ihm gegen seinen Willen entpreßt zu werden. *Wilde, Oscar (1985). Das Bildnis des Dorian Gray. German translation by Hedwig Lachmann and Gustav Landauer. Frankfurt a. M.: Insel, p. 25.*	He spoke very slowly, and the words seemed wrung out of him almost against his will. *Wilde, Oscar (2011). The Picture of Dorian Gray. London: Harvard University Press, p. 87.*

mirror system mounted on the head coil. Texts were displayed centered in the middle of the screen in a 20 pt black Arial font against a light gray background. Text lengths ranged between one and five lines. Participants were asked to refrain from reading aloud and instructed to indicate their answers by pressing one of four buttons on a fiber optic response pad (LumiTouch, Photon Control, Burnaby, Canada) placed in their right hand (−− index finger, − middle finger, + ring finger, ++ little finder, reversed key arrangement for half of the participants). Stimulus presentation was controlled using the software package *Presentation 14.2* (Neurobehavioral Systems Inc., Albany, CA, USA) installed on a standard personal computer. Trial onset was synchronized with scan onset with each trial starting with a fixation cross displayed for either 1700, 2125, 2550, 2975 or 3400 ms (i.e. TR + ¼ steps of the TR) allowing to jitter stimulus onset relative to scan onset. The fixation interval was followed by the presentation of the respective text sample to read. As far as the reading period is concerned, no time limitations were imposed. Trials continued only after the reader had indicated an answer. Each trial was concluded by a second fixation interval with a fixed timeframe of 6800 ms (= 4 scans) separating consecutive trials. Moreover, fixation periods ranging from 10 200 to 11 900 ms were included as null events and randomly interspersed between stimulus presentations (= 8 null trials per task). Measurements for each task were obtained in separate runs, and the corresponding task instructions were provided immediately before starting each run. Text order within each task was fully randomized, and task order was balanced among participants.

Functional localizer of temporal voice areas

To allow comparisons between activation patterns obtained in the main experiment and the TVA implicated in the direct perception of voice signals, all participants completed a functional localizer scan (adapted from Belin *et al.*, 2000) aimed at defining voice-sensitive brain areas: Participants were instructed to close their eyes and listen carefully to a series of sound stimuli presented to them. Acoustic stimulations included 12 blocks of human vocal sounds (VS), 6 blocks of environmental sounds (ES), 6 blocks of animal sounds (AS) and 12 blocks of silence. Each block measured 10 s in duration (i.e. 8 s of auditory stimulation plus 2 s of silence). Sound stimuli within the respective blocks were normalized to the same mean acoustic intensity, and block order was randomized among participants. Stimulus presentation was controlled using the software package *Presentation 14.2* (Neurobehavioral Systems Inc., Albany, CA, USA), and sounds stimuli were delivered via MRI compatible headphones (Sennheiser Electronic GmbH & Co. KG, Wedemark-Wennebostel, Germany; in-house modified).

Data analysis: behavioral data

Ratings (i.e. emotional valence/aesthetic value) and reading durations (i.e. time between stimulus onset and button press) were analyzed as behavioral data. To this end, obtained ratings were re-coded to numeric values (emo: ++ = 1, + = 2, − = 3, −− = 4; aes: ++ = 4, + = 3, − = 2, −− = 1) and averaged among text samples pertaining to the same valence categories and type of cue, resulting in six single measures obtained for each participant within each task condition: $mean_{face\_pos}$, $mean_{face\_neu}$, $mean_{face\_neg}$, $mean_{voice\_pos}$, $mean_{voice\_neu}$ and $mean_{voice\_neg}$. Reading durations were averaged in a similar fashion.

Data analysis: Imaging data

All images were processed and analysed using the software package SPM8 (http://www.fil.ion.ucl.ac.uk/spm/).

Preprocessing

EPI raw data were realigned to correct for head motion, unwarped using a static field map, co-registered with obtained anatomical images, normalized to MNI space and smooth with an isotropic Gaussian kernel of 8 mm full-width at half maximum. The first five images of each run were discarded from further analyses to exclude measurements preceding T1 equilibrium.

Statistical analysis: main experiment

Based on the research questions outlined earlier, statistical analyses aimed at evaluating cue-independent task and emotion effects as well as 'voice'-related effects on brain activation associated with the processing of the presented text samples. The respective analyses were based on a general linear model with each event modeled as a separate regressor convolved with the canonical HRF. Events were time-locked to the onset of each stimulus and modeled durations corresponded to the individual reading durations obtained for the respective text samples. Time series were high-pass filtered (cut-off frequency: 1/128 Hz) to remove low-frequency noise. Serial autocorrelations within the data were accounted for by modeling the error term as an autoregressive process. Estimated beta values were used to define *t*-contrasts for each subject corresponding to the main effect of each of the 12 different experimental factor level combinations (i.e. combinations of task, cue and valence). Computed contrasts then were subjected to a second-level group analysis of variance employing a full-factorial design with task (emo/aes), emotional valence (positive/negative/neutral) and type of cue (facial/vocal) specified as within-subject factors and unequal variances assumed for measurements in each level. Resulting main effects and interactions were assessed for significance at cluster level using a cluster-defining threshold of $P < 0.001$ uncorrected, and a

Table 2 Summary of key characteristics of text samples describing facial and vocal cues

Cue	Valence[a]				Arousal[b]	Aesthetic value[c]	Text length (no. of characters)	Word frequency[d] (a.u.)	Syntactic complexity[d] (a.u.)
	All	Pos	Neu	Neg					
Facial									
Mean	5.01	3.10	4.96	6.97	4.17	4.44	104.82	8.17	5.31
s.d.	1.65	0.50	0.26	0.41	1.09	0.54	57.34	1.37	0.77
Vocal									
Mean	5.06	3.35	5.09	6.75	4.42	4.51	101.79	8.25	5.27
s.d.	1.46	0.42	0.47	0.32	1.04	0.43	59.01	1.52	0.77

[a]Valence ratings measured on a scale ranging from 1—very positive to 9—very negative; categorization: neutral = 4.5–5.5, positive <4.5, negative >5.5.
[b]Arousal ratings measured on a scale ranging from 1—very low to 9—very high.
[c]Aesthetic value measured on a scale ranging from 1—very poorly written to 9—very well written.
[d]Determined for each word using the German Reference Corpus (DeReKo, Kupietz *et al.*, 2010) and averaged among words within each text sample.

cluster-wise significance levels of $P < 0.05$ corrected for multiple comparisons (across the whole brain) as criterion. Corrected cluster-level *P*-values were determined using the NS Toolbox (http://fmri.wfubmc.edu/cms/software#NS). Additionally, analyses were conducted to explore relationships between regional brain activation and behavioral responses given by the participants (see Supplementary Data).

Statistical analysis: functional localizer

Analyses of localizer data relied on a general linear model with each of the 3 stimulation blocks VS, AS and ES modeled as a separate regressor using a boxcar function of 8 s in duration convolved with the HRF. Voice-sensitive brain activation was evaluated by contrasting brain responses to VS with activation elicited by both ES and AS (*t*-contrast: VS > AS, ES). The respective contrasts then were subjected to a second-level random effects analysis. Results were assessed for cluster-wise significance using a cluster-defining threshold of $P < 0.001$ uncorrected, and a cluster-wise significance levels of $P < 0.05$ corrected for multiple comparisons (across the whole brain) as criterion.

Aiming to evaluate the contribution of the TVA to the processing of literary voice description, TVA masks were generated based on the results of the functional localizer scans and used to explore brain activation within these regions during reading. To this end, beta values (estimated for each event) were extracted from all voxel within the left or right TVA and subsequently averaged among voxels within the same hemisphere. Aiming to further explore activation differences related to task, emotional valence or type of cue, the respective mean beta values were subjected to separate repeated-measures analyses of variance (i.e. one for the right and one for the left TVA). Moreover, to evaluate the role of writing style, a second exploratory analysis was conducted to explore effects of the use of direct speech on TVA activation during reading. The motivation to test for effects of direct speech was derived from recent research findings suggesting differences in reporting style modulate reading-related TVA responses (Yao *et al.*, 2011). Of the 39 voice descriptions employed in the current experiment, 13 utilized direct speech quotations (e.g. 'Das ist nicht zu ertragen', sprach die Fürstin leise mit zitterender Stimme[1]). To infer differences between the two different reporting styles, beta estimates corresponding to 'direct-speech' or 'no-direct speech' text samples were extracted from the TVA and compared by means of paired-samples *t*-test.

RESULTS

Behavioral data

On average, judgments of emotional valence replicated valence categories assigned to the texts employed in the study (Figure 1): On a four-point scale ranging from 1—highly positive to 4—highly negative, text samples selected to represent positive states of the mind received average ratings of $\text{mean}_{\text{pos}} = 1.73$ (±0.06 s.e.m.), while mean ratings obtained for neutral and negative text samples averaged to values of $\text{mean}_{\text{neu}} = 2.50$ (±0.04 s.e.m.) and $\text{mean}_{\text{neg}} = 3.28$ (±0.05 s.e.m.), respectively.

Considering judgments of aesthetic value (Figure 1), ratings obtained on a four-point scale ranging from 1—very poorly written to 4—very well written indicated that overall the highest aesthetic value was assigned to text samples expressing positive emotions ($\text{mean}_{\text{pos}} = 2.83 \pm 0.08$ s.e.m.), followed by texts expressing neutral ($\text{mean}_{\text{neu}} = 2.70 \pm 0.06$ s.e.m.) and negative states of the mind ($\text{mean}_{\text{neg}} = 2.59 \pm 0.08$ s.e.m.).

Reading durations obtained during the emotion judgement and aesthetics judgment task (Figure 1), revealed that participants took longest to read text samples expressing a neutral as compared to an emotional state of the mind ($\text{mean}_{\text{neu\_emo}} = 6455$ ms ± 376 ms s.e.m.; $\text{mean}_{\text{neu\_aes}} = 6518$ ms ± 412 ms s.e.m.; $\text{mean}_{\text{pos\_emo}} = 5753$ ms ± 326 ms s.e.m.; $\text{mean}_{\text{pos\_aes}} = 6143$ ms ± 412 ms s.e.m.; $\text{mean}_{\text{neg\_emo}} = 5764$ ms ± 317 ms s.e.m.; $\text{mean}_{\text{neg\_aes}} = 6270$ ms ± 426 ms s.e.m.).

fMRI data: analysis of variance

Significant results are summarized in Table 3 and Figure 2.

Main effect of task

Analyses of fMRI data indicated a significant main effect of task on cerebral responses within the left lateral frontal cortex (i.e. left middle and inferior frontal cortex). *Post hoc* comparisons computed on beta values extracted from this activation cluster revealed that this main effect was driven by an increased activation of the lateral frontal cortex during the emotion judgment task as compared to aesthetics judgment task.

Main effect of valence category

A significant main effect of valence category on brain activation was observed within the arMFC (including the anterior cingulate cortex), the cerebellum and the posterior cingulate cortex (PCC). *Post hoc* comparisons computed for each activation cluster indicated that valence-related effects observed within the medial frontal cortex were driven by increasing responses of this region to text samples expressing positive (as relative to negative or neutral) emotions, while effects observed for the cerebellum and PCC were explained by stronger responses of these regions to text samples conveying both positive and negative (as compared to neutral) emotional states.

Main effect of cue type

Moreover, analyses indicated cue-related activation differences in the left posterior and mid and superior temporal cortex as well as the right superior temporal cortex and right parietal cortex extending into the superior occipital cortex. *Post hoc* inspections of the observed main effect of cue type evidenced that all the reported regions responded more strongly to descriptions of vocal as compared to facial cues.

Interactions

As far as the modeled interaction terms are concerned, no significant findings emerged at the chosen statistical thresholds.

fMRI data: TVA activation

Comparisons between cue-related activation patterns obtained in the reading experiment and voice-sensitive brain activation (as determined by the functional localizer) indicated a substantial overlap between temporal brain structures implicated in the processing of voice *descriptions* and the TVA: 83% (= 165 of 199) of the voxels activated within the right superior temporal cortex as well as 13% (= 25 of 194) of the voxels activated within the left mid and superior temporal cortex, and 22% (= 13 of 59) of the voxels activated within the left posterior superior temporal cortex proved to overlap with voice-sensitive brain structures located within the right and left hemisphere (right TVA: activation peak: 60, −18, −3, k_e: 510, $P_{\text{corr}} = 0.000$; left TVA: activation peak: −60, −9, −0, k_e: 289, $P_{\text{corr}} = 0.000$; Figure 3).

[1] English translation: 'I cannot take this any longer' the baroness said quietly with a quivering voice. Hoffmann, E.T.H. (1912). *Lebensansichten des Katers Murr*. Hamburg: Verlag Alfred Janssen, p. 194, English translation provided by the authors.

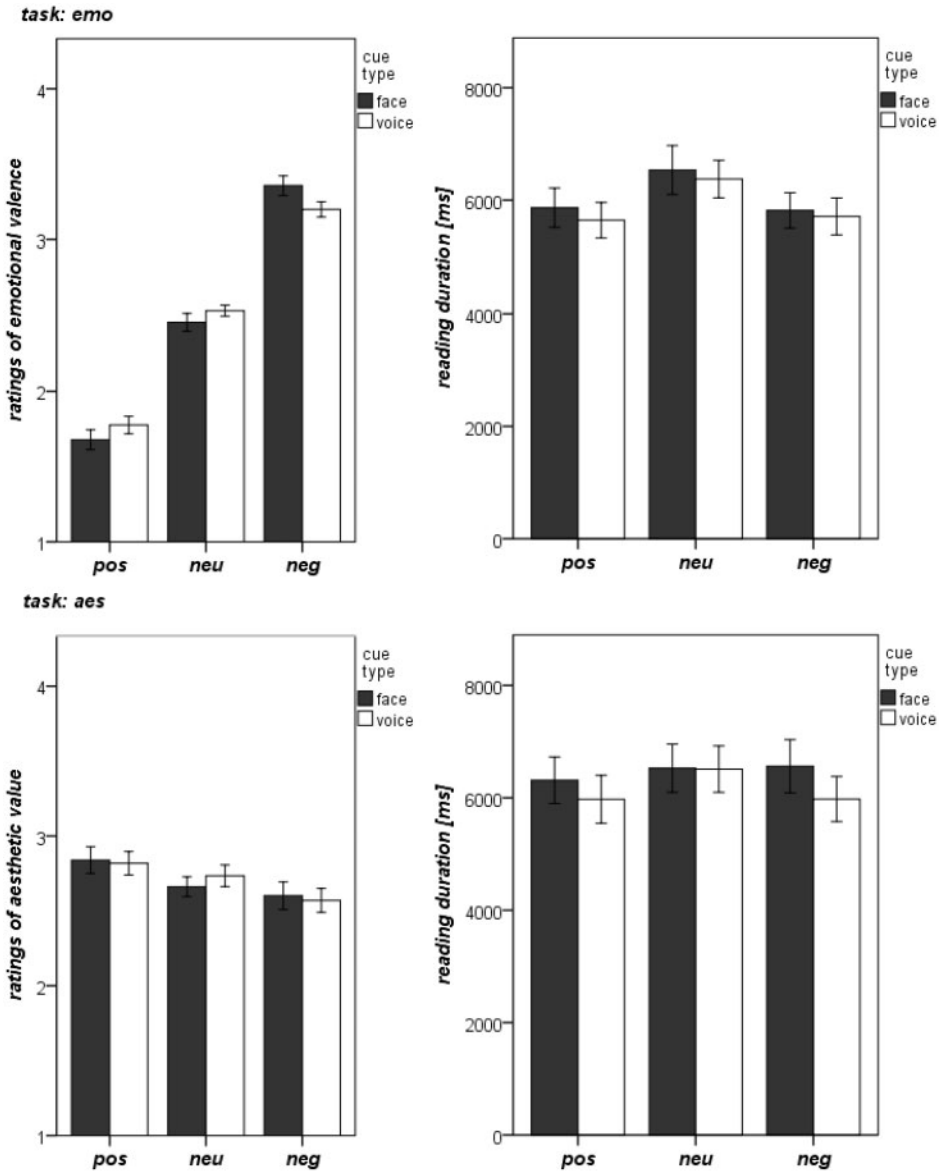

Fig. 1 Behavioral data: ratings of emotional valence and aesthetic values as well as corresponding mean reading durations observed for each valence category (positive = pos, neutral = neu, negative = neg) and type of cue (facial cues = dark gray bars, vocal cues = white bars). Results are shown as mean values ± 1 s.e.m.

Table 3 Significant results obtained from an analysis of variance computed on brain activation data

	Anatomical definition[a]		*x*	*y*	*z*	z_{max}	k_e	P_{corr}[b]
Main effect of task	L	Frontal mid/frontal inf	−48	15	36	4.38	168	0.000
Main effect of emotional valence	L/R	Cerebellum	−9	−42	−45	4.55	191	0.000
	L/R	Frontal medial/cingulum ant	0	48	−6	4.37	294	0.000
	L/R	Cingulum post	0	−30	24	4.37	141	0.001
Main effect of cue type	L	Temporal mid/temporal sup/temporal pole	−57	−9	−15	5.72	194	0.000
	L	Temporal mid	−60	−36	−3	4.43	59	0.048
	R	Temporal sup/temporal mid	63	−6	−15	4.35	199	0.000
	R	Parietal sup/occipital sup/occipital mid	30	−69	36	3.77	64	0.037

[a]Anatomical definitions are based on labels obtained using the cluster labeling tool provided by the SPM toolbox Automated Anatomical Labeling (AAL; Tzourio-Mazoyer *et al.*, 2002).
[b]Corrected for multiple comparisons across the whole brain at cluster level.

Analyses conducted on beta values extracted from both the right and left TVA indicated a significant main effect of cue type on brain activation observed within these regions explained by increased responses to descriptions of voices as relative to descriptions of faces [left TVA: $F(1,21) = 24.92$, $P < 0.001$; right TVA: $F(1,21) = 9.46$, $P = 0.006$]. Estimates of cue-type related activation differences obtained for each individual within the left and right TVA are displayed in Figure 3. Task instructions or emotional valence, on the other hand, did not influence

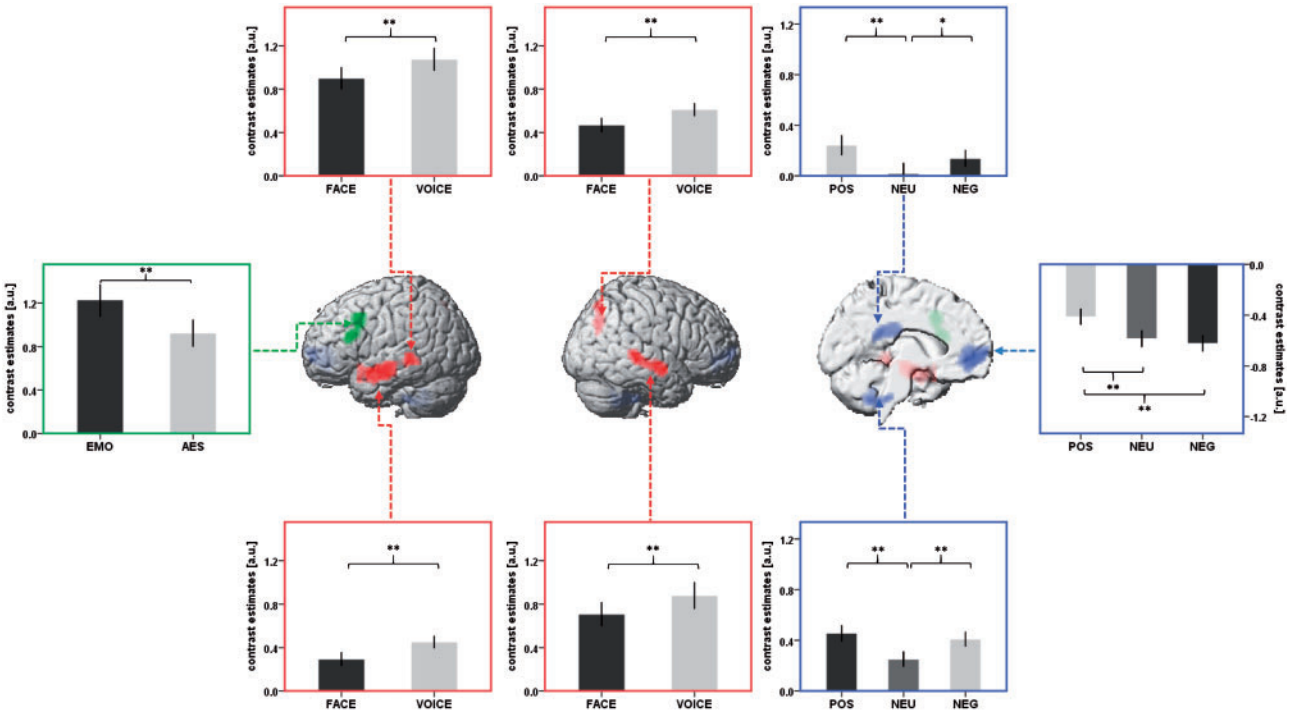

Fig. 2 Significant results obtained from the analysis of variance computed on brain activation data. Displayed are renderings of significant activation clusters (cluster-level P-value < 0.05 corrected for multiple comparisons across the whole brain) as well as beta estimates plotted as bar diagrams to further detail the findings (error bars: ±1 s.e.m., $^{**}P < 0.01$, $^{*}P < 0.05$). Green colors show activations corresponding to the main effect of task, whereas red colors depict activations corresponding to the main effect of cue type, and blue colors indicate activations reflecting the main effect of emotion.

TVA responses (all main effects and interactions involving task or valence $P \geq 0.115$).

As far as effects of writing style are concerned, beta values extracted from the TVA indicated higher mean activation to text samples including direct speech statements for both the right ($mean_{no\_direct} = 0.40 \pm 0.05$ s.e.m.; $mean_{direct} = 0.53 \pm 0.07$ s.e.m.) and left ($mean_{no\_direct} = 0.59 \pm 0.09$ s.e.m.; $mean_{direct} = 0.69 \pm 0.11$ s.e.m.) TVA. However, only activation differences observed within the right TVA reached statistical significance at a conventional threshold of $P < 0.05$ [right TVA: $t(21) = -2.55$, $P = 0.019$; left TVA: $t(21) = -1.93$, $P = 0.068$].

DISCUSSION

Building on the assumption that similarities between written and acoustic representations of vocal cues may not only emerge with respect to their functional role in communication but may also extend to the cerebral mechanisms involved in the decoding process, this study sought to explore brain responses associated with processing emotional voice signals described in literary texts.

In line with the latter assumption obtained results suggest that the decoding of literary descriptions of non-verbal affective signals may indeed (partly) rely on a set of brain regions previously implicated in the auditory perception of emotional voices.

Similarities between the perception of acoustic and described emotional cues, for instance, emerged with respect to a task-dependent recruitment of lateral frontal brain structures, particularly the inferior frontal cortex, in situations that require the explicit evaluation of the emotional information expressed in a given cue.

Considering the frontal cortex's role in emotion processing, results of recent meta-analyses (Kober *et al.*, 2008; Wager *et al.*, 2010) implicate the lateral frontal cortex, particularly the inferior frontal cortex, in a functional group of brain regions essential to information selection and meaning analysis (Kober *et al.*, 2008; Wager *et al.*, 2010; Lindquist *et al.*, 2012) across a wide range of emotion-related tasks. Similar suggestions are also reflected in models of affective information processing which assume a key role of the lateral frontal cortex in high-order stages of the decoding process related to the appraisal, interpretation and conceptual categorization of expressed emotions (e.g. Schirmer and Kotz, 2006; Wildgruber *et al.*, 2009; Brück *et al.*, 2011b). Given research findings linking particularly the inferior frontal cortex to the mirror system of the human brain (Iacoboni and Dapretto, 2006; Van Overwalle and Baetens, 2009), increasing activation of this brain area in response to the emotional judgment task might be driven by increased efforts to mirror or simulate described facial or vocal expressions, a process assumed to aid or facilitate our understanding of others.

However, as far as task effects are concerned, it should be noted that the processing of the emotional connotations and the associated mirroring of described emotional expressions may have partially contributed to aesthetic judgments as well. In other words, valence judgments may actually be a part of the decision process that leads to judgments of aesthetic value with text samples expressing a positive emotion receiving slightly higher judgments of aesthetic values as compared to text samples expressing neutral or negative emotions (see 'Behavioral data' section; Figure 1). Analyses conducted to explore relationships between brain activation and aesthetic or valence ratings given by each participant revealed significant relationships with activation of the arMFC (see Supplementary Data) for both rating types extending suggestions of an overlap between both rating tasks to the level of brain activation as well.

Aside task instructions and the associated shift in the focus of attention, reading-related brain responses proved to be affected by the emotionality of described communication signals: Compared to neutral text samples, the processing of cue descriptions conveying emotional states more strongly engaged a set of midline structures

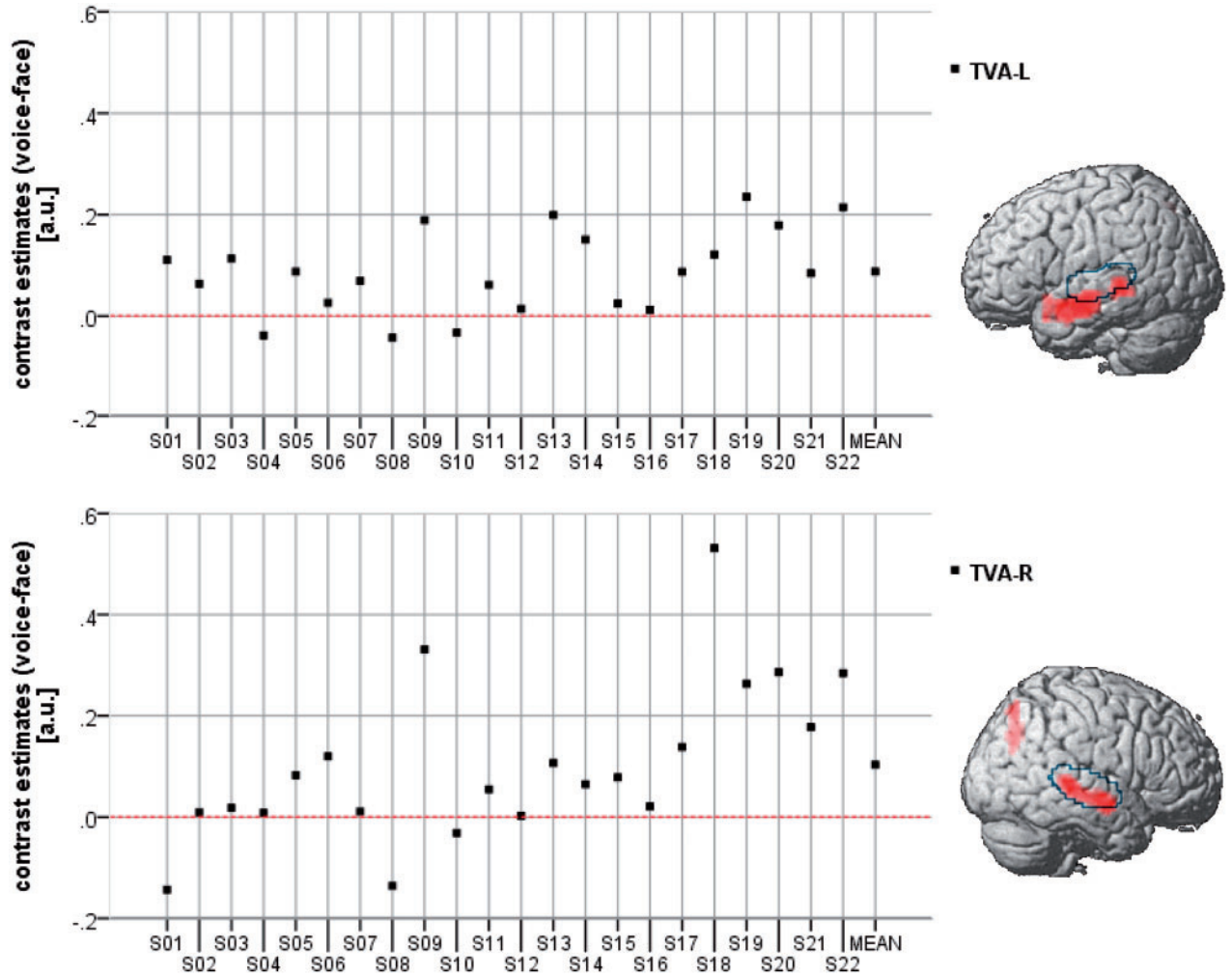

Fig. 3 Diagrams depicting contrast estimates of cue-type related activation differences obtained for each individual within the left (upper panel) and right (lower panel) TVA. Printed on the right are renderings showing the overlap between TVAs (outlined in blue) and 'voice-sensitive' activation clusters obtained in the reading experiment (main effect of cue type, voices > faces).

including aspects of the arMFC, PCC and cerebellum. While the role of the cerebellum remains elusive, hypotheses regarding the contribution of the PCC and arMFC may again be derived from several research reports tying the respective brain regions to different sub-functions and cognitive operations involved in emotion perception.

Activation of the medial frontal cortex has frequently been related to social cognitive processing with anterior rostral aspects linked to 'mentalizing' (Amodio and Frith, 2006)—processes by which inferences about the mental states of others are made (Frith and Frith, 2006). Enhanced arMFC activation observed in response to the presentation of emotional face and voice descriptions might thus be assumed to reflect mentalizing that further appeared to be modulated by the emotional salience of the respective signals (as reflected in increasing arMFC responses to emotional, particularly positive cues).

Considering the contribution of posterior cingulate brain structures, research findings linking the PCC to both the processing of stimuli carrying affective meaning (e.g. Maddock and Buonocore, 1997; Maddock *et al.*, 2003) as well as to episodic memory (e.g. Henson *et al.*, 1999; Maddock *et al.*, 2001) lead to assume a role of the PCC in the interaction between memory and emotions (Maddock, 1999). The term interaction in this case is used to describe a regulation of memory by emotion (Maddock, 1999) that perhaps may be most commonly expressed in a more efficient encoding, and thus an enhancement of memory, for emotional events. However, such interaction effects may also involve memory search and retrieval: One could assume that observed emotions cue the recall of similar emotional states or events (personally experienced in the past), and that the recall of such memories may in turn serve as a reference to interpret current observations (Lindquist *et al.*, 2012). In other words, literary descriptions of emotional expressions provided in this study could have cued in the reader an emotion-related memory search and retrieval mediated by the PCC.

While obtained responses of the PCC as well as of the lateral and medial frontal cortex may be considered to reflect brain responses reported across a wide range of emotion-related tasks and phenomena, temporal activation observed in this study appears to reveal a voice-sensitive modulation of activation. Latter assumptions are corroborated by observations of increased responding of this structure to voice descriptions that furthermore appeared to be enhanced by the use of direct speech quotations mimicking speech acts (Yao *et al.*, 2011). Considering the localization of observed 'voice'-sensitive modulations of brain activation, comparisons conducted between reading-related brain responses and functional localizer data revealed a substantial overlap between the identified 'voice'-sensitive activation clusters and areas specialized for the perception of human voices termed the TVA (Belin *et al.*, 2000, 2002).

In analogy to face-sensitive structures reported for the human visual system (i.e. fusiform face area; Kanwisher *et al.*, 1997), the TVA has

been suggested to represent a processing module that subserves the auditory analysis of voices (Campanella and Belin, 2007; Belin *et al.*, 2011) relevant to a rich set of voice cognition abilities including the extraction of emotional information encoded in a voice. However, the role of the TVA may not be limited to the auditory analysis and decoding of acoustic voice cues alone. Rather recent research reports as well as results obtained in this study demonstrate activation of voice-sensitive brain areas even in the absence of acoustic stimulation: Research published on the cerebral structures recruited during (non-clinical) auditory verbal hallucinations (Linden *et al.*, 2011) or the silent reading of text samples depicting different speech acts (Yao *et al.*, 2011) may serve as examples to illustrate activation of voice-sensitive brain structures that is not driven by acoustic stimulation.

However, a common denominator among experiments aimed at investigating verbal hallucinations and reading studies (including the current experiment) may be the shared experience of an 'inner voice' in the process. As far as reading is concerned, anecdotal reports as well as observations obtained in behavioral experiments identify occurrences of an inner voice, or the perceptual simulation of voice characteristic while reading, to be a commonplace phenomenon frequently observed among individuals (Alexander and Nygaard, 2008; Kurby *et al.*, 2009; Yao and Scheepers, 2011). Recent neuroimaging findings, moreover, link these perceptual simulations to a top-down activation of the TVA that occurs even when readers are not explicitly instructed to imagine the sound of a voice (Yao *et al.*, 2011). Latter findings connecting TVA activation with processes of auditory mental imagery, in turn, may be interpreted to suggest that the TVA may not only represent a processing site integral to the acoustic analysis of voice information but may also store acoustic information related to different vocal sounds that is re-activated during perceptual simulations. Recalling, recombining and modifying these stored sound information may not only give rise to auditory imagery, and thus the experience of hearing an inner voice (Kosslyn *et al.*, 2001) rather it may also facilitate the voice decoding process in the sense that mental representations formed on the basis of previous experiences may help in assigning meaning to presented voice descriptions. Activation of the TVA observed in the reading process may thus reflect the access of voice-related memories and the formation of mental images used in the process. Aside the more general observation of an increased TVA activation in response to voice descriptions, the link between TVA responses and auditory mental imagery is further substantiated by the observation that stylistic devices such as direct speech quotations, aimed at increasing mental imagery during reading, even further enhance reading-related TVA responses. As reasoned by Yao *et al.* (2011) direct speech quotations are 'assumed to entail a demonstration of the reported utterance' rather than a 'mere description' thus providing the reader with a more vivid and perceptually engaging exemplar of a speech act that raises the likelihood readers will 'activate "audible speech"-like representations' during reading (p. 3146). Yao *et al.* (2011) continue to link these direct-speech related simulation processes to increases in TVA activation strengthening suggestions that the TVA contributes to processes of auditory mental imagery.

CONCLUSION

Whether an acoustic phenomenon in a day-to-day conversation or a vivid description in a book, emotional voice cues may share common characteristics that may not only relate to their functional role as valuable source of information but may also include cerebral mechanisms associated with the decoding processes. Similarities emerge with respect to the recruitment of both specialized voice perception areas as well as brain regions such as the posterior cingulate, or lateral and medial frontal cortex assumed to subserve functions relevant to emotion perception in a broader context. Observed similarities, in turn, may suggest a common perceptual mechanism that underlies the ability to decode emotional voice cues across a wide range of tasks or forms of presentation.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

REFERENCES

Ackermann, H., Hertrich, I., Grodd, W., Wildgruber, D. (2004). Das Hören von Gefühlen: Funktionell-neuroanatomische Grundlage der Verarbeitung affektiver Prosodie. *Aktuelle Neurologie*, *31*, 449–60.

Alexander, J.D., Nygaard, L.C. (2008). Reading voices and hearing text: talker-specific auditory imagery in reading. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(2), 446–59.

Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–77.

Banse, R., Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–36.

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

Belin, P., Zatorre, R.J., Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Research. Cognitive Brain Research*, *13*(1), 17–26.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–12.

Bestelmeyer, P.E., Belin, P., Grosbras, M.H. (2011). Right temporal TMS impairs voice detection. *Current Biology*, *21*(20), R838–9.

Brück, C., Kreifelts, B., Kaza, E., Lotze, M., Wildgruber, D. (2011a). Impact of personality on the cerebral processing of emotional prosody. *Neuroimage*, *58*(1), 259–68.

Brück, C., Kreifelts, B., Wildgruber, D. (2011b). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, *8*(4), 383–403.

Campanella, S., Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Science*, *11*(12), 535–43.

Ethofer, T., Anders, S., Erb, M., et al. (2006). Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage*, *30*(2), 580–7.

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*(1), 191–200.

Ethofer, T., Kreifelts, B., Wiethoff, S., et al. (2009a). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, *21*(7), 1255–68.

Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P. (2009b). Decoding of emotional information in voice-sensitive cortices. *Current Biology*, *19*(12), 1028–33.

Frith, C.D., Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–4.

Fusar-Poli, P., Placentino, A., Carletti, F., et al. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of Psychiatry and Neuroscience*, *34*(6), 418–32.

Henson, R.N., Rugg, M.D., Shallice, T., Josephs, O., Dolan, R.J. (1999). Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, *19*(10), 3962–72.

Iacoboni, M., Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*(12), 942–51.

Kanwisher, N., McDermott, J., Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–11.

Kesler-West, M.L., Andersen, A.H., Smith, C.D., et al. (2001). Neural substrates of facial emotion processing using fMRI. *Brain Research. Cognitive Brain Research*, *11*(2), 213–26.

Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K., Wager, T.D. (2008). Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage*, *42*(2), 998–1031.

Kosslyn, S.M., Ganis, G., Thompson, W.L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–42.

Kupietz, M., Belica, C., Keibel, H., Witt, A. (2010). The German Reference Corpus DEREKO: a primordial sample for linguistic research. In: Calzolari, N., Choukri, K., Maegaard, B., et al., editors. *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*, pp. 1848–54.

Kurby, C.A., Magliano, J.P., Rapp, D.N. (2009). Those voices in your head: activation of auditory images during reading. *Cognition*, *112*(3), 457–61.

Linden, D.E., Thornton, K., Kuswanto, C.N., Johnston, S.J., van de Ven, V., Jackson, M.C. (2011). The brain's voices: comparing nonclinical auditory hallucinations and imagery. *Cerebral Cortex*, *21*(2), 330–7.

Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences*, *35*(3), 121–43.

Maddock, R.J. (1999). The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends in Neuroscience*, *22*(7), 310–6.

Maddock, R.J., Buonocore, M.H. (1997). Activation of left posterior cingulate gyrus by the auditory presentation of threat-related words: an fMRI study. *Psychiatry Research*, *75*(1), 1–14.

Maddock, R.J., Garrett, A.S., Buonocore, M.H. (2001). Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience*, *104*(3), 667–76.

Maddock, R.J., Garrett, A.S., Buonocore, M.H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping*, *18*(1), 30–41.

Meyer, M., Baumann, S., Wildgruber, D., Alter, K. (2007). How the brain laughs. Comparative evidence from behavioral, electrophysiological and neuroimaging studies in human and monkey. *Behavioural Brain Research*, *182*(2), 245–60.

Peelen, M.V., Atkinson, A.P., Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, *30*(30), 10127–34.

Sabatinelli, D., Fortune, E.E., Li, Q., et al. (2011). Emotional perception: meta-analyses of face and natural scene processing. *Neuroimage*, *54*(3), 2524–33.

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*(11), 2251–72.

Schirmer, A., Kotz, S.A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Science*, *10*(1), 24–30.

Szameitat, D.P., Alter, K., Szameitat, A.J., Wildgruber, D., Sterr, A., Darwin, C.J. (2009). Acoustic profiles of distinct emotional expressions in laughter. *The Journal of the Acoustical Society of America*, *126*(1), 354–66.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273–89.

Van Overwalle, F., Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*(3), 564–84.

Wager, T.D., Feldman Barrett, L., Bliss-Moreau, E., et al. (2010). The neuroimaging of emotion. In: Lewis, M., Haviland-Jones, J.M., Feldman Barrett, L., editors. *Handbook of Emotions*, Vol. 3, New York: Guilfod Press, pp. 249–71.

Wiethoff, S., Wildgruber, D., Grodd, W., Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *Neuroreport*, *20*(15), 1356–60.

Wildgruber, D., Ackermann, H., Kreifelts, B., Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, *156*, 249–68.

Wildgruber, D., Ethofer, T., Grandjean, D., Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, *11*(4), 277–81.

Wildgruber, D., Hertrich, I., Riecker, A., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, *14*(12), 1384–9.

Wildgruber, D., Riecker, A., Hertrich, I., et al. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage*, *24*(4), 1233–41.

Yao, B., Belin, P., Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*(10), 3146–52.

Yao, B., Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition*, *121*(3), 447–53.

Zald, D.H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research. Brain Research Reviews*, *41*(1), 88–123.