‘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 Wildgruber1,2

1Department of Psychiatry and Psychotherapy, Eberhard Karls University, Tübingen 72076, Germany, 2Werner Reichardt Centre for Integrative Neuroscience (CIN), Tübingen 72076, Germany and 3Department 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; Sramek 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.

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 amygdala (Wiethoff et al., 2009), the anterior rostral medial frontal cortex (aMFC) (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 amygdala and aMFC 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 (Zakl, 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 mm³, 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).

Table 1 Examples of text stimuli employed in the main experiment

<table>
<thead>
<tr>
<th>Cue</th>
<th>Emotional state</th>
<th>Sample text</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facial</td>
<td>Positive</td>
<td>Sie zeigte, für die Dauer eines Herzschlags, ihr Lachen von einer Wange zur anderen.</td>
<td>For the duration of a heartbeat, she smiled from cheek to cheek.</td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td>Hoffmann, Ernst Theodor Amadeus (2001) Rat Krespel. In: Steinecke, H. &amp; Segebrecht, W. (Eds.) Sämtliche Werke. Frankfurt a. M.: Deutscher Klassiker Verlag. Bd. 4, p. 44.</td>
<td>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 [...].</td>
</tr>
<tr>
<td>Vocal</td>
<td>Positive</td>
<td>Als sie sprach, klang ihre Stimme sanft und kehlig und mit einem italienischen Akzent behaftet.</td>
<td>When she spoke, her voice was smooth — a throaty, accented English.</td>
</tr>
</tbody>
</table>
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 finger, 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 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).

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

<table>
<thead>
<tr>
<th>Cue</th>
<th>Valence*</th>
<th>Arousal*</th>
<th>Aesthetic value†</th>
<th>Text length (no. of characters)</th>
<th>Word frequency‡</th>
<th>Syntactic complexity§</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Pos</td>
<td>Neu</td>
<td>Neg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial</td>
<td>Mean</td>
<td>5.01</td>
<td>3.10</td>
<td>4.96</td>
<td>6.97</td>
<td>4.17</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>1.65</td>
<td>0.50</td>
<td>0.26</td>
<td>0.41</td>
<td>1.09</td>
</tr>
<tr>
<td>Vocal</td>
<td>Mean</td>
<td>5.06</td>
<td>3.35</td>
<td>5.09</td>
<td>6.75</td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>1.46</td>
<td>0.42</td>
<td>0.47</td>
<td>0.32</td>
<td>1.04</td>
</tr>
</tbody>
</table>

*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.
†Arousal ratings measured on a scale ranging from 1—very low to 9—very high.
‡Aesthetic value measured on a scale ranging from 1—very poorly written to 9—very well written.
§Syntactic complexity determined for each word using the German Reference Corpus (DeReKo, Kupietz et al., 2010) and averaged among words within each text sample.

**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: meanface_pos, meanface_neu, meanface_neg, meanvoice_pos, meanvoice_neu and meanvoice_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
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 NIS Toolbox (http://fmri.wfubmc.edu/cms/software#NIS). 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 voxels 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 zitternder Stimme). 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 meanpos = 1.73 ($\pm 0.06$ s.e.m.), while mean ratings obtained for neutral and negative text samples averaged to values of meanneu = 2.50 ($\pm 0.04$ s.e.m.) and meanneg = 3.28 ($\pm 0.05$ 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 (meanpos = $2.83 \pm 0.08$ s.e.m.), followed by texts expressing neutral (meanneu = $2.70 \pm 0.06$ s.e.m.) and negative states of the mind (meanneg = $2.59 \pm 0.08$ 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 (meanneu_emo = 6455 ms $\pm 376$ s.e.m.; meanpos_emo = 5753 ms $\pm 326$ s.e.m.; meanneg_emo = 6143 ms $\pm 412$ s.e.m.; meanpos_emo = 5764 ms $\pm 317$ s.e.m.; meanneg_emo = 6270 ms $\pm 426$ 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, $-18$, $-3$, $k$: 510, $P_{corr} = 0.000$; left TVA: activation peak: $-59$, $-9$, $-0$, $k$: 289, $P_{corr} = 0.000$; Figure 3).

---

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

![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.](https://academic.oup.com/scan/article-abstract/9/11/1819/1685540/1823)

![Table 3 Significant results obtained from an analysis of variance computed on brain activation data](https://academic.oup.com/scan/article-abstract/9/11/1819/1685540/1823)

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

<table>
<thead>
<tr>
<th>Anatomical definition</th>
<th>$x$</th>
<th>$y$</th>
<th>$z$</th>
<th>$z_{\text{max}}$</th>
<th>$k_c$</th>
<th>$P_{\text{corr}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main effect of task</strong></td>
<td>L Frontal mid/frontal inf</td>
<td>-48</td>
<td>15</td>
<td>36</td>
<td>4.38</td>
<td>168</td>
</tr>
<tr>
<td><strong>Main effect of emotional valence</strong></td>
<td>L/R Cerebellum</td>
<td>-9</td>
<td>-42</td>
<td>-45</td>
<td>4.55</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td>L/R Frontal medial/cingulum ant</td>
<td>0</td>
<td>48</td>
<td>-6</td>
<td>4.37</td>
<td>294</td>
</tr>
<tr>
<td></td>
<td>L/R Cingulum post</td>
<td>0</td>
<td>-30</td>
<td>24</td>
<td>4.37</td>
<td>141</td>
</tr>
<tr>
<td><strong>Main effect of cue type</strong></td>
<td>L Temporal mid/temporal sup/temporal pole</td>
<td>-57</td>
<td>-9</td>
<td>-15</td>
<td>5.72</td>
<td>194</td>
</tr>
<tr>
<td></td>
<td>R Temporal mid</td>
<td>-60</td>
<td>-36</td>
<td>-3</td>
<td>4.43</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>R Parietal sup/occipital sup/occipital mid</td>
<td>63</td>
<td>-15</td>
<td>36</td>
<td>4.35</td>
<td>199</td>
</tr>
</tbody>
</table>

*Anatomical definitions are based on labels obtained using the cluster labelling tool provided by the SPM toolbox Automated Anatomical Labeling (AAL; Tzourio-Mazoyer et al., 2002).

*Corrected for multiple comparisons across the whole brain at cluster level.
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 \( \text{no} \_\text{direct} = 0.40 \pm 0.05 \text{ s.e.m.}; \text{mean} \_\text{direct} = 0.53 \pm 0.07 \text{ s.e.m.} \)) and left (mean \( \text{no} \_\text{direct} = 0.59 \pm 0.09 \text{ s.e.m.}; \text{mean} \_\text{direct} = 0.69 \pm 0.11 \text{ s.e.m.} \)) 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; 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.
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 Buonomore, 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 subsumes 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 investigated in verbal hallucinations and reading studies (including the clinical) auditory verbal hallucinations (Linden 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.

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


