Valence-specific conflict moderation in the dorso-medial PFC and the caudate head in emotional speech

Abbreviated title: Conflict in emotional speech

Sonja A. Kotz a, b*, Reinhard Dengler c, Matthias Wittfoth d, e

a Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
b School of Psychological Sciences, The University of Manchester, Manchester, UK
c Department of Neurology, Hannover Medical School, Hannover, Germany
d Department of Clinical Psychiatry, Social Psychiatry, and Psychotherapy, Hannover Medical School, Hannover, Germany
e NICA – NeuroImaging and Clinical Applications, Hannover, Germany

* Corresponding author.
Sonja A. Kotz, PhD
Max Planck Institute for Human Cognitive and Brain Sciences
Dept. of Neuropsychology
Stephanstrasse 1a
04103 Leipzig
Germany
E-mail: kotz@cbs.mpg.de
Keywords: fMRI; conflict; prediction error; dmPFC; caudate head; emotional speech

Funding

This work was supported by the German Science Foundation [KO2268/3-2 to Sonja A. Kotz and FOR 499-KO2267/3-2 to Christine Schroeder nee Kohlmetz].

Acknowledgements

The authors wish to thank Denise Scheermann for recruiting participants and conducting the measurements, as well as all the participants of this study for their motivation.
Abstract

Emotional speech comprises of complex multimodal verbal and non-verbal information that allows deducting others’ emotional states or thoughts in social interactions. While the neural correlates of verbal and non-verbal aspects and their interaction in emotional speech have been identified, there is very little evidence on how we perceive and resolve incongruity in emotional speech, and whether such incongruity extends to current concepts of task-specific prediction errors as a consequence of unexpected action outcomes (“negative surprise”). Here we explored this possibility while participants listened to congruent and incongruent angry, happy, or neutral utterances and categorized the expressed emotions by their verbal (semantic) content. Results reveal valence-specific incongruity effects: negative verbal content expressed in a happy tone of voice increased activation in the dorso-medial prefrontal cortex (dmPFC) extending its role from conflict moderation to appraisal of valence-specific conflict in emotional speech. Conversely, the caudate head bilaterally responded selectively to positive verbal content expressed in an angry tone of voice broadening previous accounts of the caudate head in linguistic control to moderating valence-specific control in emotional speech. Together these results suggest that control structures of the human brain (dmPFC and sub-compartments of the basal ganglia) impact emotional speech differentially when conflict arises.
Introduction

In social communication we often rely on verbal (e.g. semantic content) and non-verbal emotion expressions (e.g. tone of voice, face, gesture, and body) to infer how someone feels and thinks. Most of these social exchanges employ congruent verbal and non-verbal emotion expressions that are known to facilitate information processing (e.g. de Gelder et al., 2004; Paulmann and Pell, 2009; Jessen and Kotz, 2011; Jessen and Kotz, 2013). However, communicating emotions is a complex process and, at times, marked by incongruity that leads to problems in decoding what a speaker intends to say (e.g. Schirmer, 2004; Mitchell, 2006a, b; Wittfoth et al., 2010). For example, in a professional conversation you generally do not expect someone to raise their voice in anger. However, a speaker, who is unhappy about the course of a conversation may smile at you, while her/his tone of voice may signal the emotion that the speaker really wants to express at that moment. This leads to mixed communicative signals and thus conflict in how to interpret the speaker’s communicative intent.

Consequently, it is not only important to understand how we (i) perceive and integrate different forms of emotion expressions, but also (ii) perceive and resolve mixed communicative signals. We therefore need to identify the neural correlates that underlie the perception of verbal and non-verbal emotion expression when they do and do not match our expectations.

Previous fMRI research has identified predominantly right-lateralized fronto-temporal neural correlates for verbal and non-verbal congruent vocal emotion expressions (e.g. (Paulmann et al., 2010; Bruck et al., 2011; Kotz and Paulmann, 2011). Furthermore, behavioral and EEG patient studies have implicated the basal ganglia in the processing of non-verbal vocal emotion expressions such as tone of voice (e.g. (Cancelliere and Kertesz, 1990; Pell, 1996; Pell and Leonard, 2003; Sidtis and Van Lancker Sidtis, 2003; Dara et al., 2008; Kotz et al., 2009; Paulmann and Pell, 2009; Paulmann et al., 2011), verbal emotion expressions (Castner
et al., 2007; Hillier et al., 2007; Kotz et al., 2009; Paulmann and Pell, 2009), and the integration of emotion expressions (Paulmann and Pell, 2009; Paulmann et al., 2011). To date, only a few fMRI studies have reported enhanced activation of the basal ganglia in response to verbal vocal (Kotz et al., 2005) and non-verbal vocal emotion expressions (Morris et al., 1999; Wildgruber et al., 2002; Kotz et al., 2003, 2006; Grandjean et al., 2006; Bach et al., 2008; Fruhholz et al., 2012).

In an attempt to specify how verbal and non-verbal vocal emotion expressions interact, recent imaging studies have started to utilized incongruity paradigms (Schirmer et al., 2004; Mitchell, 2006a, b; Wittfoth et al., 2010). Results confirm the involvement of the above described fronto-striato-temporal network, but extends it to areas responding to incongruity or conflict: (i) the dorsal anterior cingulate cortex (dACC), more generally ascribed to conflict resolution (e.g. Kanske and Kotz, 2011a, b), and (ii) valence-specific responses in the caudate nucleus/thalamus (Wittfoth et al., 2010). In light of recent seminal work on cognitive and affective conflict that has implicated the role of the dACC/medial prefrontal cortex (mPFC) in monitoring predictions of action outcomes in specific task contexts (Alexander and Brown, 2011; Egner, 2011), the question arises whether a mixed communicative signal (e.g. an incongruent verbal and non-verbal emotion expression) reflects an unexpected action outcome (e.g. a speaker’s communicative intent) that leads to a task-specific prediction error and increased activation in dACC, mPFC, or both (e.g. Egner, 2011). Further, it is of specific interest whether task-specific prediction errors, termed “negative surprise” (Egner, 2011) are valence-specific or not (Alexander and Brown, 2011). Taken together, such evidence would indeed confirm (i) a domain-general role of the dACC/mPFC in responding to prediction errors independent of the stimulus type used to induce conflict in task specific action.
outcome, and (ii) would specify whether “negative surprise” is a valence-specific phenomenon or not.

Therefore, we set out to test and extend the following research questions: (1) Do dACC, mPFC, or both respond to a task-specific prediction error (e.g. incongruent verbal and non-verbal emotion expression)? (2) Is the response in these areas valence-specific or not?

We instructed participants to categorize the emotional content of an utterance (task specificity: emotional semantic content). We expected to find activation in dACC, mPFC or both in response to incongruent verbal and non-verbal emotion expressions. Should the unexpected outcome of mismatching verbal and non-verbal expressions be valence-specific, we predicted activation in dACC, mPFC, or both to increase when negative verbal content is spoken in a happy tone of voice (task-specific “negative surprise”). On the other hand, a mismatch between positive verbal content and an angry tone of voice may engage subcomponents of the basal ganglia (e.g. caudate; Wittfoth et al., 2010) in support of a valence-specific response to an unexpected action outcome in emotional speech.

**Materials and Methods**

**Participants**

Twenty right-handed, normal-hearing native speakers of German were recruited from the University of Magdeburg student population with an age range of 24 to 37 years (mean age 26.9; SD = 3.4, 10 females). None of them reported any current or past medical or neurological condition. They received a brief MR safety screening and gave informed consent prior to the experiment. The protocol was approved by the local ethics committee of the University of Magdeburg and of the Hannover Medical School in accordance with the Declaration of Helsinki.
Paradigm

The stimulus material consisted of 96 German sentences that were spoken by a semi-professional actress. Stimuli were recorded with a DAT recorder, digitized at a 16-bit/44.1 kHz sampling rate, and normalized to 75 dB. The material consisted of an equal number (32) of three verbal (v) emotion expressions: positive (P) (e.g. ‘Sie hat ihre Prüfung bestanden.’; ‘She has passed her exam.’), negative (N) (e.g. ‘Sie hat ihm das Herz gebrochen.’; ‘She has broken his heart.’), or neutral (X) (‘Sie hat das Buch gelesen.’; ‘She has read the book.’).

Each of the three verbal emotion expressions was spoken either with a positive (happy), negative (angry), or neutral tone of voice (nv) resulting in a total number of 288 experimental stimuli. Of the resulting nine experimental conditions, three were congruent (congruent positive (P,Pnv)); congruent negative (N,Nnv); congruent neutral (X,Xnv)), while two were incongruent (incongruent positive (P,Nnv); incongruent negative (N,Pnv)). The remaining conditions were paired with neutral verbal (X,Pnv, X,Nnv) or non-verbal vocal expressions (P,Xnv, N,Xnv) and served as filler conditions to counterbalance the design. The average duration of an expression was about 1.6 seconds. Of further note is that the experimental stimuli were constructed in a way that verbal emotion expression could only be recognized at the end of an utterance (the noun), while the impact of non-verbal vocal information started with the onset of an utterance (Paulmann and Kotz, 2008). Thus, we aimed at studying the interaction of non-verbal vocal and verbal emotion expressions in a semi-controlled fashion.

Participants categorized the verbal emotional content of each expression by pressing one of three buttons on a custom-built response box with their right hand as fast and accurately as possible. Half of the participants responded with the right-hand index finger to positive verbal content and with their right-hand ring finger to negative verbal content; the response assignment was reversed for the other half of the participants. All participants pressed with
the right-hand middle finger in response to neutral verbal content. Auditory stimuli were delivered using a high-frequency shielded transducer system. This transmission system includes a piezoelectric loudspeaker enabling the transmission of strong sound pressure levels (~105 dB) with excellent attenuation characteristics. The loudspeakers were embedded in tightly occlusive MR-compatible headphones allowing unimpeded conduction of the stimuli. Participants wore noise protection ear plugs to provide additional noise attenuation. Auditory stimulation was adjusted during a short training session to a comfortable listening level, which was the same for all participants. After the scanning, all participants reported that the acoustic quality and comprehension of the presented stimuli was optimal.

**MRI data acquisition**

Images were acquired employing a standard head coil with a Siemens Trio 3-T scanner (Erlangen, Germany) at the ZENIT (Zentrum für neurowissenschaftliche Innovation und Technologie, University Clinic of Magdeburg, Germany). Functional images were recorded axially along the AC-PC plane with a T2*-weighted gradient-echo echoplanar imaging (EPI) sequence (repetition time TR = 3500 ms, echo time TE = 30 ms, flip angle = 90°, FoV 192 x 192 mm², matrix 64 x 64, 3 x 3 x 5 mm in-plane resolution) with 18 slices of 4 mm thickness interspersed with 1-mm in-between gaps (seven slices extended ventrally from the AC-PC plane). Slices were acquired in an interleaved manner. One session was recorded with two short breaks of stimulation (2 x 20 seconds, no scanner interruption), resulting in a total of 658 volumes. All participants were instructed to lie still during the stimulation and pauses and to shut their eyes throughout the scanning session. The first three volumes were discarded to allow for magnetic saturation effects. For the structural scans, we used a T1-weighted sequence in the same orientation as the functional sequence to provide detailed anatomical images aligned to the functional scans [192 1-mm slices, TR 2500 ms, TE 4.77 ms, FoV of
256 mm]. High-resolution structural images were also acquired for the purpose of cross-subject registration. Scanning consisted of an event-related design with jittered inter-trial-intervals randomly chosen from a distribution ranging from zero to two seconds. Forty-eight null events were included in the design. All stimuli were presented in pseudo-randomized order with the constraint that no expression type was immediately repeated in the same condition.

Data Analysis

Behavioral data acquired during scanning were assessed with respect to response times and error rates. A two-factorial repeated-measures ANOVA with the factors verbal emotion expression (positive, negative, and neutral) and non-verbal vocal emotion expression (happy, angry, and neutral tone of voice) was conducted for reaction times and error rates. The statistical threshold was set to p<0.05 and Greenhouse-Geisser corrections were applied when necessary.

Functional MRI data were pre-processed with SPM5 (Welcome Department of Cognitive Neurology, London). As the slices of each volume were not acquired simultaneously, a timing correction procedure was applied. All volumes were realigned to the 10th volume, ‘unwarped’ to remove variance caused by movement-by-field-in homogeneity interactions, normalized to a standard EPI template, and smoothed with a Gaussian kernel of 8 mm full-width at half-maximum to account for anatomical differences between participants and to allow for statistical inference using the Gaussian Random Field theory. At the first level, data were high-pass filtered (128 sec) and an autoregressive function (AR-1) was employed to estimate temporal autocorrelation in the data and to correct the degrees of freedom, accordingly.
The first-level model consisted of ten regressors, which comprised all experimental conditions and an additional regressor, in which incorrect or delayed response trials were entered. Trial onsets in SPM were time points at the end of each expression motivated by the expression characteristics (early and on-going vocal emotion expression and late verbal emotion expression). As mentioned above, expressions were composed in a way that participants were not able to identify verbal emotional content before hearing the last word of the utterance. This ensured that participants fully processed the spoken utterance before categorizing the verbal content of an utterance.

Nine regressors (all experimental conditions) were analysed in a repeated-measure ANOVA. In order to examine incongruity effects, we restricted our analyses to those conditions, which did not contain neutral information (see explanation above). As the task-specific feature was always the verbal content of an expression, the task-irrelevant information (non-verbal vocal emotion expressions) was either congruent (PvPnv, NvNnv) or incongruent (PvNnv, NvPnv).

In order to address our specific research questions we report (i) an overall incongruity effect, (ii) two task-specific incongruity effects by comparing incongruent and congruent expressions when the verbal emotion expressions (task-specific) changed valence (PvNnv > NvNnv and NvPnv > PvPnv), but non-verbal vocal emotion expressions (task-irrelevant) stayed constant, and (iii) valence-specific effects (congruent).

We applied a statistical threshold of p < 0.001 with an additional cluster extend threshold corrected for multiple comparisons in order to protect against false positive results (Forman et al., 1995). This threshold (k = 13 voxels) was calculated by SPM using an empirically determined extent threshold. This method is implemented in SPM and refers to the estimated smoothness of the images. After determining of the number of resels, the expected Euler characteristic is calculated. This is used to give the correct threshold (number of voxels) that
is required to control for false positive results. Coordinates of cluster maxima are reported in MNI space and labeled using the aal database within the WFU-Pickatlas (Tzourio-Mazoyer et al., 2002; Maldjian et al., 2003; Maldjian et al., 2004).

Results

Behavioral data

Main effects for both verbal and non-verbal information were observed when analysing reaction times (RT) and error rates (ER).

Participants responded slowest to positive verbal expressions and fastest to negative verbal expressions (main effect verbal content for RT: \( F(2,38)=44.7, p<0.001 \)). However, participants displayed lower ERs for positive verbal expressions and increased ERs for neutral verbal expressions (main effect verbal for ER: \( F(2,38)=7.04, p=0.003 \)).

When sentences were neutrally intoned, participants responded slower compared to sentences spoken in an emotional tone of voice (main effect nonverbal for RT: \( F(2,38)=22.63, p<0.001 \)). Further, participants made more errors responding to sentences spoken in a neutral tone of voice than to sentences in a happy tone of voice (main effect nonverbal for ER: \( F(2,38)=4.61, p=0.038 \)).

For both reaction times (\( F(4, 76) = 6.3; p<0.001 \)) and accuracy rates (\( F(4, 76) = 4.4; p=0.003 \)), we also found an interaction of verbal and non-verbal emotion expressions (see Figure 1).

Resolving this interaction with Bonferroni-corrected post-hoc t-tests (\( p=0.05/4 = 0.0125 \)) for RTs, we found that participants responded significantly faster to congruent negative sentences (\( N_vN_{nv} \)) compared to both incongruent conditions (vs. \( P_vN_{nv} \): \( t(19)=-10.8, p<0.001 \); vs. \( N_vP_{nv} \):...
Comparing positive congruent sentences with both incongruent conditions did only yield one significant result (vs. N, P<sub>n v</sub>: p=0.016, vs. P, N<sub>n v</sub>: p=0.171).

With respect to error rates, we found that participants made less errors in the congruent positive condition (P, P<sub>n v</sub>) compared with both incongruent conditions (vs. N, P<sub>n v</sub>: t(19)=5.12, p=0.00006, vs. P, N<sub>n v</sub>: t(19)=4.22, p=0.00047). Here, comparing congruent negative sentences (N, N<sub>n v</sub>) with the two incongruent conditions did not yield significant results (vs. P, N<sub>n v</sub>: p=1.0, vs. N, P<sub>n v</sub>: p=0.29154).

Please include Figure 1 about here

**Functional MRI data**

**Overall incongruity**

As listed in Table 1, attenuated activation in the right thalamus was found in the overall contrast between incongruent > congruent utterances.

Please include Table 1 about here

**Task- (verbal) and valence-specific incongruity effects**

When compared to congruent negative expressions positive verbal expressions spoken in an angry tone of voice (P, N<sub>n v</sub> > N, N<sub>n v</sub>) revealed increased activation in the caudate head
bilaterally, the left hippocampus, the right precuneus, and the right thalamus. Negative verbal expressions intoned happily compared to congruent positive expressions ($N_v P_{nv} > P_v P_{nv}$) increased activation in the bilateral posterior superior temporal gyrus (STG), the right mid STG, the right middle frontal gyrus (MFG; BA46), and in the left inferior frontal gyrus (IFG; BA47; Table 1).

Task-(verbal) and valence-specific effects
Contrasting congruent positive ($P_v P_{nv}$) and negative ($N_v, N_{nv}$) emotion expressions revealed increased activation in the caudate head bilaterally and the lingual gyri (BA19). The reversed contrast led to increased activation in the dorso-medial PFC bilaterally (BA10), the left inferior temporal gyrus (ITG; BA20), the right middle temporal gyrus (MTG; BA21), and left posterior STG (BA22/41/42) (see Table 2/Figure 2).

Please insert Table 2/Figure 2 about here

Discussion
In the current study we set out to investigate whether incongruity of verbal (positive, negative, neutral) and non-verbal vocal (happy, angry, neutral) emotion expressions leads to activation in dACC, mPFC, or both comparable to reported activation in response to task-specific prediction errors of expected action outcomes (Alexander and Brown, 2011; Egner, 2011) and whether such activation is valence-specific or not (Wittfoth et al., 2010; Alexander and Brown, 2011). In a fully crossed design participants categorized the verbal emotional content of utterances (task-specificity) they listened to.
Overall incongruity-related activation

To test how verbal and non-verbal vocal emotion expressions interact, we applied an incongruity paradigm. Unlike the Stroop task, in which a neutral stimulus consists of a color-word written in an ink color unrelated to it, emotional verbal expressions spoken in an unexpected tone of voice can be informative in social communicative interactions (i.e. stepping over social boundaries in a communicative context). It is important to note that the expressions used in the current experiment ensured that participants were only able to detect a possible incongruity of verbal and non-verbal vocal emotional expressions at the end of an utterance. Our finding that the thalamus is associated with an incongruity response, is in line with the view that this structure is part of a network, which comes into play during the anticipation of aversive events (Chua et al., 1999; Ploghaus et al., 1999) as well as subsequent emotion regulation with a possible impact on conflict resolution (Herwig et al., 2007; Goldin et al., 2008). Furthermore, the thalamus responded specifically to an angry tone of voice that did not match the verbal emotional content, thereby emphasizing its role as a threat detector (Choi et al., 2012).

Task- and valence-specific effects

When comparing positive verbal emotion expressions intoned angrily, we observed a task-specific (verbal content) activation increase in the caudate heads and in the lingual gyri. This pattern of activation is comparable to reported previous results (Wittfoth et al., 2010), which revealed increased activation of the caudate heads when positive verbal expressions were spoken in an angry tone of voice (task-specificity: tone of voice). So independent of task requirements we confirm a valence-specific activation pattern. Further, activation of the caudate heads not only occurred in response to positive verbal expressions, but even more so when verbal expressions mismatched non-verbal emotional vocal expression. We suggest that
the role of the caudate head(s) goes beyond the detection of positive social incentives, but also moderates the control of verbal and non-verbal vocal interaction (i.e. its congruity or incongruity) independent of task requirements (i.e. focus on verbal or non-verbal emotion expressions). On the one hand, this view is supported by observations that the caudate head is responsive to positive social signals in economic games (King-Casas et al., 2005; Delgado et al., 2008), or in the broader sense, whenever an outcome is perceived as desirable (Grahn et al., 2008). Our results may therefore extend monetary rewards to social communicative incentives. On the other hand, recent studies have emphasized the role of the caudate head in language control, i.e. during the processing of incongruent words (Ali et al., 2010), or language switching in bilinguals (Crinion et al., 2006). Furthermore, a recent meta-analytic connectivity study associated the head of the caudate nucleus with cognition and emotion-related structures (e.g. the IFG or the dACC), which seem to play a relevant role in the present investigation (Robinson et al., 2012). Crucially, independent of task demands (verbal here, non-verbal elsewhere (Witthoft et al., 2010)) or the incongruity of an emotion expression, the caudate head responds in a valence-specific manner. Consequently, the present results may encourage investigations of possible structural or neurodegenerative changes of the caudate head (Paulmann et al., 2011; Robotham et al., 2011). Although, there have been some attempts to link executive dysfunction to reduced activation in striatal regions in Parkinson’s or Huntington’s disease (Hasselbalch et al., 1992; Lawrence et al., 1998), the present finding may offer new insight into valence-specific incongruity effects in emotional speech and beyond in these patient groups.

Taking a closer look at the processing of negative compared to positive verbal emotion expressions, we find increased activation in the bilateral mid-to-posterior STG and in the dorso-medial PFC (dmPFC). In particular, the activation of the dmPFC is of interest as this
region has been attributed to inferring other people’s mental states (theory-of-mind) (Beaucousin et al., 2007; Sebastian et al., 2011). Furthermore, a recent imaging study using multi-voxel pattern analysis suggested not only a specific role of the dmPFC in mental state attribution, but found that this area together with the left STS responds modality independent (Peelen et al., 2010). Our findings corroborate this view of the dmPFC as a region associated with (1) appraisal as well as (2) conflict regulation in humans (Etkin et al., 2011) in a modality independent manner (Peelen et al., 2010).

However, on the basis of the current results, we propose that the role of the dmPFC extends to valence- and task-specific conflict regulation. Only negative verbal expression, but not positive verbal expression as well as the incongruity of negative verbal expression and happy tone of voice led to increased activation in the dmPFC. Additional activation was observed in the right MFG, an area within the lateral prefrontal cortex, which has frequently been associated with the processing of semantic constraints in sentences comprehension (for an overview see (Price, 2010), or with the processing of incongruent non-verbal vocal expressions (Wittfoth et al., 2010).

One crucial question remains to be discussed at this stage: Why may dmPFC activation be specific for “negative” conflict, and the caudate for “positive” conflict? More specifically, can these valence-specific activation patterns be considered as domain general effects, that is, not speech (verbal or non-verbal) specific? Upon elaborations on the valence-specific response of the caudate heads within the basal ganglia above, we will further consider the specific role of dmPFC in response to negative conflict.

Our finding of a distinction of a valence specific response pattern in dmPFC is supported by a number of studies. These results indicate that dmPFC activation links to a number of different processes, none of them mutually exclusive. These processes include the detection of change...
in the environment that signal the subsequent need for behavioral adjustment, especially negative change including unfavorable outcomes and response conflict (Holroyd and Coles 2002, Rushworth, Walton et al. 2004, Yeung, Botvinick et al. 2004, Botvinick 2007, Alexander and Brown 2011), social threats (Eisenberger and Lieberman 2004, Eisenberger 2013), or emotional reappraisal (Etkin, Egner et al. 2011, Ochsner, Silvers et al. 2012). In particular, recent investigations of the “social brain” have suggested that dmPFC may respond sensitively to cognitive imbalance and may therefore reflect a degree of cognitive dissonance (Izuma, Matsumoto et al. 2010, Izuma and Adolphs 2013). In other words, dmPFC encodes the discrepancy of current negative information and a desired outcome.

Accumulating data also suggest that dmPFC activation corresponds to the ability to reason about other peoples’ mental states, generally referred to as theory of mind (Ochsner, Knierim et al. 2004). Considering our experimental setup, in which participants listen to statements of one person about a third person, our findings fit well to the concept that dmPFC may come into play in triadic relations between two minds and an object (Saxe 2006). Our results combine and can extend these perspectives of the role of the dmPFC. We therefore suggest that dmPFC encodes the social context of triadic relations of negative flavor in emotional speech, thus inducing an emotional imbalance and signaling a need for a subsequent adjustment.

**Conclusions**

The current study set out to shed further light on the interaction of verbal and non-verbal vocal emotion expressions. Utilizing an incongruity paradigm it also allowed investigating whether a mismatch between verbal and non-verbal emotion expressions, as often to be found in social communicative interactions, relies on valence-specific activation patterns in the human brain that may or may not be domain specific. The current results clearly emphasize
that the appraisal and monitoring of emotional speech incongruities extend beyond classically
described networks involved in the perception and the categorization of verbal and non-verbal
vocal expressions. Both the dmPFC and the caudate head in their more general functions of
conflict regulation and cognitive/affective control, respectively, respond in a valence-specific
manner when encountering incongruity in emotional speech and confirm the importance of
these control structures in human social communication.
References


Figure captions

Figure 1: Behavioral data: Average reaction times and error rates

Reaction times in response to verbal emotion expressions were slowest responding to positive verbal information and fastest responding to negative verbal information. Furthermore, participants responded fastest when listening to congruent negative expressions ($N_v N_{nv}$).

Figure 2: Task- and valence specific effects

A: Displayed are significant clusters in the left dmPFC. Contrast estimates of the left dmPFC support the view that negative verbal expressions (cues) are processed by this area. Blue circles indicate the area, from which the contrast estimates were extracted. B: Significant clusters in bilateral caudate heads shown in an axial view ($z=2$). Contrast estimates of the left caudate head displays elevated signals for positive verbal emotion expressions (cues). Contrast estimates for the right caudate head were comparable (not shown here). Furthermore, activation of the left lingual gyrus can be seen.

Table 1: Activation clusters of task-specific incongruity effects

Order of presentation: overall incongruity, positive conflict, negative conflict. Abbreviations: MFG=middle frontal gyrus; IFG=inferior frontal gyrus; STG=superior temporal gyrus; SFG=superior frontal gyrus. Whole-brain analysis of all suprathreshold voxels; contrasts at $p<0.001$, $k=13$.

Table 2: Activation clusters of task- and valence-specific effects

Abbreviations: PFC=prefrontal cortex; MTG=middle temporal gyrus; ITG=inferior temporal gyrus; STG=superior temporal gyrus; SFG=superior frontal gyrus. Whole-brain analysis of all suprathreshold voxels; contrasts at $p<0.001$, $k=13$. 
Behavioral data: Average reaction times and error rates
84x101mm (220 x 220 DPI)
Task- and valence-specific effects

201x162mm (300 x 300 DPI)
<table>
<thead>
<tr>
<th>Region</th>
<th>Side</th>
<th>BA</th>
<th>z-value</th>
<th>Clustersize</th>
<th>MNI coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thalamus</td>
<td>R</td>
<td></td>
<td>4.29</td>
<td>59</td>
<td>X -18, Y 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incongruent (P_vN_v + N_vP_nv) &gt; Congruent (P_vP_nv + N_vN_nv)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caudate Head</td>
<td>L</td>
<td></td>
<td>4.30</td>
<td>102</td>
<td>X -10, Y 20, Z 2</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td></td>
<td>3.88</td>
<td>110</td>
<td>X 10, Y 20, Z 4</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>L</td>
<td></td>
<td>4.11</td>
<td>73</td>
<td>X -32, Y -44, Z 0</td>
</tr>
<tr>
<td>Thalamus</td>
<td>R</td>
<td></td>
<td>3.99</td>
<td>51</td>
<td>X 2, Y -18, Z 22</td>
</tr>
<tr>
<td>Precuneus</td>
<td>R</td>
<td></td>
<td>3.84</td>
<td>23</td>
<td>X 26, Y -48, Z 20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incongruent (N_vP_nv) &gt; Congruent (P_vP_nv)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid STG</td>
<td>R</td>
<td>22</td>
<td>3.77</td>
<td>113</td>
<td>X 70, Y -26, Z 4</td>
</tr>
<tr>
<td>Posterior STG</td>
<td>L</td>
<td>22</td>
<td>3.54</td>
<td>20</td>
<td>X -62, Y -32, Z 8</td>
</tr>
<tr>
<td>MFG</td>
<td>R</td>
<td>46</td>
<td>3.58</td>
<td>13</td>
<td>X 46, Y 30, Z 20</td>
</tr>
<tr>
<td>IFG</td>
<td>L</td>
<td>47</td>
<td>3.48</td>
<td>16</td>
<td>X -46, Y 22, Z 4</td>
</tr>
<tr>
<td>Region</td>
<td>Side</td>
<td>BA</td>
<td>z-value</td>
<td>Clustersize</td>
<td>MNI coordinates</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
<td>----</td>
<td>---------</td>
<td>-------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Positive &gt; Negative verbal expressions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lingual Gyrus</td>
<td>R</td>
<td>19</td>
<td>4.71</td>
<td>187</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>19</td>
<td>3.46</td>
<td>37</td>
<td>-28</td>
</tr>
<tr>
<td>Caudate Head</td>
<td>R</td>
<td>19</td>
<td>4.24</td>
<td>98</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>3.76</td>
<td>32</td>
<td>-10</td>
<td>20</td>
</tr>
<tr>
<td><strong>Negative &gt; Positive verbal expressions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dorso-medial PFC</td>
<td>L</td>
<td>10</td>
<td>4.70</td>
<td>301</td>
<td>-6</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>10</td>
<td>4.21</td>
<td>45</td>
<td>20</td>
</tr>
<tr>
<td>Mid MTG</td>
<td>R</td>
<td>21</td>
<td>4.49</td>
<td>283</td>
<td>66</td>
</tr>
<tr>
<td>ITG</td>
<td>L</td>
<td>20</td>
<td>4.03</td>
<td>34</td>
<td>-52</td>
</tr>
<tr>
<td>Posterior STG</td>
<td>L</td>
<td>22/42</td>
<td>3.82</td>
<td>34</td>
<td>-62</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>41</td>
<td>3.44</td>
<td>15</td>
<td>-40</td>
</tr>
<tr>
<td>SFG</td>
<td>R</td>
<td>10</td>
<td>3.36</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>