

Prosody production networks are modulated by sensory cues and social context

Martin Klasen,^{1,2} Clara von Marschall,^{1,2} Güldehen Isman,¹ Mikhail Zvyagintsev,^{1,2} Ruben C. Gur,³ and Klaus Mathiak^{1,2}

¹Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany, ²JARA - Translational Brain Medicine, 52074 Aachen, Germany, and ³Department of Psychiatry, University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104, USA

Correspondence should be addressed to Martin Klasen, Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany. E-mail: mklasen@ukaachen.de

Martin Klasen and Clara von Marschall contributed equally to this study.

Abstract

The neurobiology of emotional prosody production is not well investigated. In particular, the effects of cues and social context are not known. The present study sought to differentiate cued from free emotion generation and the effect of social feedback from a human listener. Online speech filtering enabled functional magnetic resonance imaging during prosodic communication in 30 participants. Emotional vocalizations were (i) free, (ii) auditorily cued, (iii) visually cued or (iv) with interactive feedback. In addition to distributed language networks, cued emotions increased activity in auditory and—in case of visual stimuli—visual cortex. Responses were larger in posterior superior temporal gyrus at the right hemisphere and the ventral striatum when participants were listened to and received feedback from the experimenter. Sensory, language and reward networks contributed to prosody production and were modulated by cues and social context. The right posterior superior temporal gyrus is a central hub for communication in social interactions—in particular for interpersonal evaluation of vocal emotions.

Key words: emotion; social interaction; fMRI; pSTG; prosody

Introduction

Emotional prosody is an essential part of human communication; its production and understanding are crucial for social interactions (Klasen *et al.*, 2012). Functional imaging and lesion studies have addressed the neurobiology underlying emotional prosody, suggesting specific neural networks for the comprehension and production of emotional prosodic cues (Schirmer and Kotz, 2006; Aziz-Zadeh *et al.*, 2010). Their embedding in social communication, however, is far less investigated. To an even larger extent than the pure linguistic component of speech, prosody processing depends on the context, i.e. speaker–speaker relationship, situational and social context

and the allower narrative. Therefore, the interactive component of emotional prosody needs to be considered, but so far little is known on its mechanisms and neurocognitive basis.

Synthesizing previous cognitive neuroscience research on prosody perception, Schirmer and Kotz (2006) introduced a neurobiological three-step model of emotional prosody comprehension. This model postulates a right-hemispheric specialization for prosody, based on its lower temporal resolution which supports the decoding of prosodic rather than semantic aspects of speech. In this model, first, acoustic information is analyzed by the auditory cortex. In this stage, spectral processing of lower frequencies is more proficient in the right hemisphere, which

Received: 30 May 2017; Revised: 23 January 2018; Accepted: 22 February 2018

© The Author(s) (2018). Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

encodes information over longer intervals (Mathiak et al., 2002). Second, emotional significance of prosody is decoded in terms of auditory object recognition by the right auditory 'what' processing stream (Altmann et al., 2007) in superior temporal gyrus (STG) and anterior superior temporal sulcus (STS). The third processing step also takes place in a lateralized fashion: The right inferior frontal gyrus (IFG) performs evaluative judgments, whereas semantic processing takes place in the left IFG (Schirmer and Kotz, 2006). Therefore, semantic processing should be minimized when studying pure prosody processing, e.g. by applying pseudowords. According to this model, right hemisphere areas play a major role in processing paralinguistic aspects of speech; nonetheless, the model also highlights left-hemisphere contributions. Indeed, functional magnetic resonance imaging (fMRI) studies showed primarily right-lateralized networks in frontotemporal (Aziz-Zadeh et al., 2010) and superior temporal regions (Witteman et al., 2012) for prosody comprehension, although left-hemisphere contributions have also been described (Kotz and Paulmann, 2011).

Compared with prosody perception, fewer studies addressed mechanisms of prosody production. Frontal and temporal networks were central in prosody production as in its comprehension (Aziz-Zadeh et al., 2010; Klaas et al., 2015). Moreover, affective components of prosody production have been related to basal ganglia and limbic areas (Pichon and Kell, 2013). Further, tasks for prosody production involved visual and auditory cortices (Aziz-Zadeh et al., 2010; Pichon and Kell, 2013; Belyk and Brown, 2016) but their functional significance was unclear; they may have been parts of the prosody production network or reflected just modality-specific cues of the experimental task. In general, prosody production yielded bilateral activation patterns but the contribution of posterior superior temporal gyrus (pSTG) seemed to be stronger in the right hemisphere (Pichon and Kell, 2013). Despite prosody being essentially social, neither comprehension nor production of emotional prosody has so far been investigated in the context of social communication.

The present study addresses two open research questions. First, in previous studies on prosody production, the emotion was always cued, i.e. the participants were told which emotion they were supposed to express. From an experimental perspective, this is a well-controlled approach but still lacks ecological validity and delivers limited insight into the mechanisms of stimulus selection and executive control. Aim of the present study was therefore to consider aspects of ecological validity that have not been investigated by previous studies on prosody production. To differentiate the neural underpinnings of free emotional prosody generation from cue-specific activation patterns, we therefore compared free generation of emotional prosody to production after visual and auditory cues. According to the definition by Schmuckler (2001), we accordingly put special emphasis on the aspect of the tasks (and, consequently, on a corresponding behavior of the participants) that considered aspects of real-life actions. Second (and as another aspect of task-related ecological validity), all studies so far focused on single aspects of prosody processing, i.e. either comprehension or production. However, natural communication is bidirectional, and passive reception or the production without expected response will be rather the exception than the rule. In face-to-face communication, prosody perception and production are closely intertwined and influenced by each other. Thus, social interactivity captures important ecologically valid aspects of prosodic communication and is processed across distributed neural networks (Gordon et al., 2013). However, interactive prosodic

communication is a major technical challenge in noisy high-field fMRI environment. An optical microphone and online digital signal processing (DSP) enabled bidirectional communication despite the scanner noise (Zvyagintsev et al., 2010).

The present study extended previous research by including two cued and an un-cued condition as well as a social interaction condition. During the entire experiment, participants listened and watched bi-syllabic pseudowords and subsequently uttered them with happy, angry, or neutral expression. The choice and processing of the emotion for each utterance was modified according to four conditions: (1) free production, where participants selected freely the emotion; (2) an auditory cue instructed on the emotion; (3) a visual cue was given and (4) the social interaction condition, where subjects selected freely the emotion and subsequently received a direct feedback on the emotion as perceived by the experimenter. Based on previous findings, we postulated the following hypotheses: (H1) In any condition, the processing and production of pseudowords with emotional prosody activates an extended right-lateralized network of inferior frontal and superior temporal areas, along with bilateral basal ganglia, motor, premotor and sensory regions; (H2) auditory and visual cues increase involvement of the respective sensory modality compared with the free (un-cued) production and (H3) the social interaction condition with interactive feedback leads to increased activity in right-hemispheric IFG and pSTG (Broca and Wernicke homotopes) compared with the free production condition without the feedback.

Materials and methods

Participants

A total of 30 German volunteers (age 20–33 years, median and s.d. 24 ± 3.2 ; 16 females) were recruited via advertisements at RWTH Aachen University. All participants had normal or corrected-to-normal vision, normal hearing, no contraindications against MR investigations and no history of neurological or psychiatric illness. The study protocol was approved by the local ethics committee and was designed in line with the Code of Ethics of the World Medical Association (Declaration of Helsinki, 1964). Written informed consent was obtained from all participants.

Stimuli

Prosodic stimuli were disyllabic pseudowords uttered with happy, angry or neutral prosody. The pseudowords followed German phono-tactical rules, but had no semantic meaning (see Thönnessen et al., 2010). The emotional prosody of the pseudowords was validated in a pre-study by 25 participants, who did not take place in the fMRI study (see Klasen et al., 2011 for recognition rates and details of stimulus validation). Visual stimuli consisted of the pseudowords in written form on a white background. During the perception phase, the words were presented in black font. An exception was the visual cue condition, where the font color indicated the emotion that was to be produced. Color changes indicated the beginning of the production phase (see Experimental Design). Auditory and visual cues were presented by MR-compatible headphones and video goggles. The sound level was adjusted to comfortable listening levels.

Experimental design

The present experiment investigated four different conditions of emotional prosody production, which were employed during

different measurement sessions in a randomized order. Condition 1 investigated free prosody production, i.e. the participants were free to choose which one of the three emotions they wanted to express. Conditions 2 and 3 investigated prosody production following auditory or visual emotion cues, respectively. Condition 4 included social interaction: In addition to produce freely emotional prosody like in Condition 1, the participants received feedback from the experimenter about the expressed emotion (Figure 1). This feedback did not need to be incorporated in the further production trials. The participants' speech was recorded via an MRI compatible fiber-optic microphone attached to the head coil (prototype from Sennheiser electronic GmbH & Co KG, Wedemark-Wennebostel, Germany). Microphone recordings were filtered online in real time to minimize the scanner noise for the listener, using a template-based subtraction approach (Zvyagintsev et al., 2010). This setup allowed the experimenter to listen to the participant's speech and prosody in real time. Additionally, the speech signals from all sessions were saved to disk for offline analysis. In the social interaction condition (Condition 4), this signal was used for emotion recognition by the experimenter and online feedback.

During the free prosody production (Condition 1), participants pronounced a given pseudoword. For each trial, they were free to choose one of three emotions (happy, neutral and angry). Afterwards, a new pseudoword was presented which had to be expressed in a self-chosen emotion again. During Condition 2, participants listened to an auditory cue and repeated the recognized emotion before they listened to the next pseudoword. During the visual cue condition (Condition 3), participants were presented a color-coded pseudoword. In specific, red font color indicated that the pseudoword should be produced with angry prosody, blue font color indicated happy prosody and black color neutral prosody. Each emotion was assigned to one color and participants should utter the given pseudoword in the given emotion before they received a new cue. The social interaction Condition 4 represents an extended version of Condition 1 (free production). However, instead of hearing all pseudowords in neutral prosody, each subsequent pseudoword was presented with the individual emotion that was recognized by the experimenter in the previous trial. This feedback represented the social component of this condition since subjects were made aware of being listened to by another human and of their influence on this interaction.

Each fMRI session included eight blocks of prosody production and lasted 12 min. One block (65 s) included 12 trials of prosody production. Each trial began with the participant viewing a pseudoword in written form (2.7 s). In Conditions 2 and 3, the pseudoword was combined with an auditory or visual cue, respectively, for the instructed emotion. The auditory cue was randomly given by female or male voices. Afterwards, participants had to produce emotional prosody (2.7 s), before they received the next pseudoword. Blocks were separated by 17 s resting blocks with a fixation cross.

The four conditions were each presented in a separate measurement session. The order of experimental conditions was randomized across participants. Within each condition, the order of trials was randomized as well; accordingly, all blocks in the experiment comprised different stimuli and were also different between conditions.

Content analysis

To test reliability of prosody coding and task adherence, an offline content analysis of the speech recordings was performed

for two-thirds of the participants. For each trial in every condition, the coder (GI) listened to the participant's speech recording and coded the recognized emotion. For Conditions 2 and 3, the coding took place without knowledge of the stimulation protocol, i.e. the coder did not know which emotion the participant was instructed to express. Krippendorff's Alpha (Hayes and Krippendorff, 2007) assessed the reliabilities of task performance and coding. In the Conditions 1 (free production) and 4 (social), no cues were presented and thus accuracy could not be determined. Nevertheless, frequencies of the different emotions were assessed and compared with Chi-square tests.

fMRI data acquisition

Whole-brain fMRI was performed with echo-planar imaging (EPI) sequences (TE = 28 ms, TR = 2000 ms, flip angle = 77°, voxel size = 3 × 3 mm, matrix size = 64 × 64, 34 transverse slices, 3 mm slice thickness, 0.75 mm gap) on a 3 Tesla Siemens Trio scanner (Siemens Medical, Erlangen, Germany) with a standard 12-channel head coil. For each of the four sessions, 360 images were acquired, corresponding to an acquisition time of 12 min per session. After the functional measurements, high-resolution T1-weighted anatomical images were performed using a magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence (TE = 2.52 ms; TI = 900 ms; TR = 1900 ms; flip angle = 9°; FOV = 256 × 256 mm²; 1 mm isotropic voxels; 176 sagittal slices). Total scanning time including preparation was ~55 min.

fMRI data analysis

Due to technical problems, data from seven participants were lost or incomplete. Data from the remaining 23 participants (12 females) were included in the group analysis.

Whole-brain fMRI data analysis was performed using standard procedures in SPM12 (Wellcome Department of Imaging Neuroscience, London, UK; <http://www.fil.ion.ucl.ac.uk/spm>). Standard preprocessing included motion correction, Gaussian spatial smoothing (8 mm FWHM) and high-pass filtering with linear-trend removal. The first 8 images of each session were discarded to avoid T1 saturation effects. All functional images were co-registered to anatomical data and normalized into MNI space.

Whole-brain mapping of brain activation was calculated in a block design. Separate predictors were defined for each of the four conditions. First-level contrast maps were calculated, contrasting the prosody blocks to the fixation cross phases. Second-level results were obtained via one-sample t-tests for the four conditions and via paired t-tests, comparing each of the other Conditions 2, 3 and 4 (auditory cue, visual cue and social interaction) with the free production condition (Condition 1) to extract the specific contribution of the additional cognitive processes. The resulting t maps were thresholded at a voxel-wise $P < 0.001$ and a $P < 0.05$ on the cluster level with family-wise error (FWE) correction for multiple comparisons.

To further investigate condition effects and lateralization, we calculated a region of interest (ROI) analysis in anatomically defined sensory cortex areas and central nodes of the language and prosody networks. Based on the anatomical labels of the WFU Pickatlas (Maldjian et al., 2003), we defined spherical ROIs (radius: 10 mm) in bilateral Heschl's gyrus (HG, auditory cortex; $x = \pm 52$, $y = -16$, $z = 6$), primary visual area (V1; $x = \pm 10$, $y = -96$, $z = -2$), IFG ($x = \pm 48$, $y = 26$, $z = 6$) and pSTG ($x = \pm 58$, $y = -44$, $z = 12$). To test for condition-specific lateralization effects, we

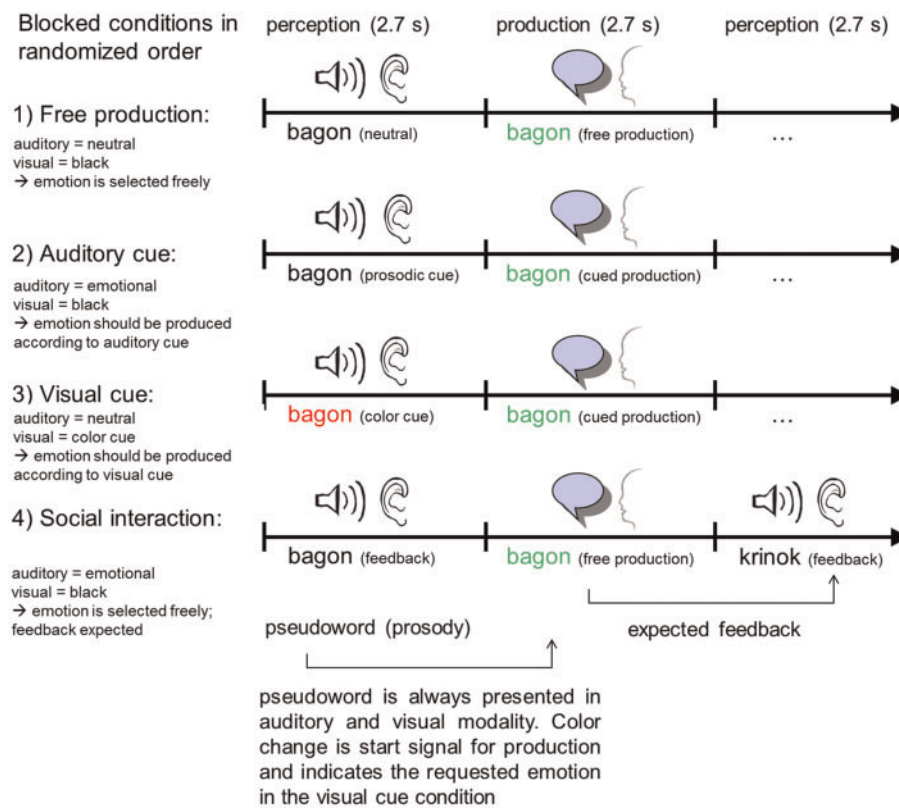


Fig. 1. Study design. Production of emotional prosody was investigated in four different conditions: (1) free production, where participants selected freely the emotion; (2) an auditory cue indicated the emotion; (3) a visual cue indicated the emotion and (4) a social interaction condition, where subjects selected freely the emotion and subsequently received a direct feedback on the emotion as perceived by the experimenter.

calculated hemisphere \times condition ANOVAs for all anatomical regions (Heschl's gyrus, V1, IFG and pSTG). For a detailed investigation of potential hemisphere \times condition interactions, we calculated additional *post-hoc* t-tests. Specifically, we calculated the laterality coefficient (right-left) for each anatomical region. For each of the four laterality coefficients, experimental conditions were then compared using paired t tests. Bonferroni correction for multiple comparisons was applied and significance level was considered according to a $P < 0.05$.

Results

Behavioral data

During the auditory or visual cue conditions (Condition 2 and 3), participants were instructed to utter the written pseudoword in a given emotion (neutral, happy or angry). To determine whether the participants correctly translated the cues into vocal emotions and whether these were correctly recognized, offline coding of the recorded speech assessed their reliabilities. Due to technical problems, sound recordings from eight participants of the final fMRI data sample were incomplete; offline codings and reliability assessment were thus performed in two-thirds of the participants ($n = 15$). Therefore, a trained coder categorized the perceived emotion of each of the recorded utterances. During this procedure, the coder was blind to the cued emotion. Krippendorff's alpha was used to evaluate the reliability of emotion production with respect to the instructed cue (Hayes and Krippendorff, 2007). For auditory cues, cued and recognized emotions were identical in 82.0% of all trials, resulting in a Krippendorff's alpha of 0.73. For visually cued emotions, cued

and recognized emotions were identical in 83.5% of the trials, corresponding to a Krippendorff's alpha of 0.75. Thus, the fidelity of the emotion production and recognition admit the scanner noise yielded a satisfactory reliability.

Utterances in the free production condition and the interactive condition were coded as well. In contrast to the cued conditions, any emotion was correct but we were interested in a potential behavioral bias. Therefore, the distributions were investigated in a 2 (condition) \times 3 (emotion) chi-square test but no significant difference emerged between the two conditions [free production: angry = 443; happy = 508; neutral = 484; interactive production: angry = 448; happy = 526; neutral = 467; $\chi^2(2) = 0.63$, $P = 0.73$]. Also, there was a trend toward more happy stimuli in both conditions; however, this was not significant after Bonferroni correction [social interaction: $\chi^2(2) = 6.89$, $P = 0.06$; free: $\chi^2(2) = 4.52$, $P = 0.22$]. Moreover, we investigated if there was a tendency to 'sustain emotions' in the social interaction and free conditions, i.e. if the participants more frequently repeated the emotion from the previous trial. Therefore, we calculated two additional coefficients: an 'interaction coefficient' for the social interaction condition, which quantified the percentage of trials in which the participant followed the emotion of the experimenter, and a 'repetition coefficient' for the social interaction and free conditions, which quantified the percentage of trials where the produced emotion was identical to the one in the previous trial (please note that the emotion in the perception phase was always neutral for the free condition). On an average, the interaction coefficient in the social interaction condition was 24.80%. The repetition coefficient, in turn, was 30.24% in the free and 26.32% in the social interaction condition. All coefficients did not reach chance level

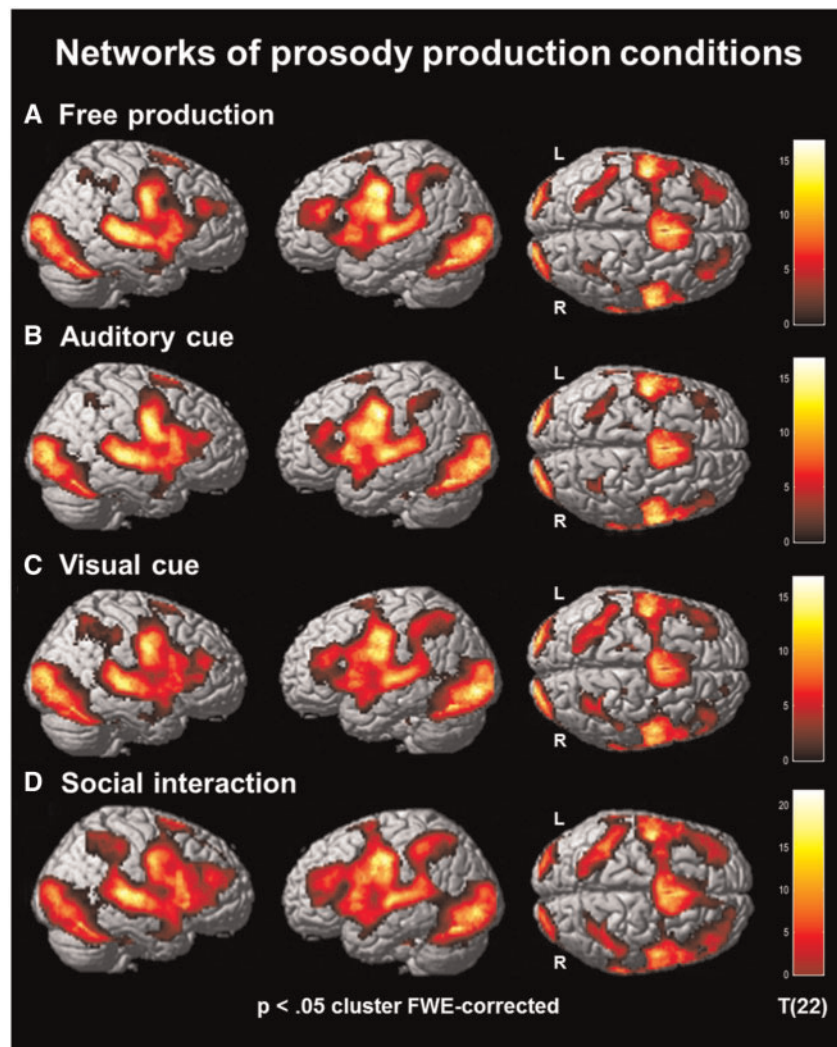


Fig. 2. Networks of prosody production conditions. Brain activation patterns for free production (A), auditory cue (B), visual cue (C) and social communication (D). Sensory and motor areas emerged for all conditions. Activity in parietal, inferior frontal and posterior temporal regions was most pronounced in the social interaction condition (cluster-wise $p_{FWE} < 0.05$).

(33.33% in a paradigm with three emotions) and were in a very similar range. The data thus showed no tendency to sustain the same emotion, neither in a social context nor during free production.

fMRI data

To investigate the neural networks associated with each of the conditions (Hypothesis H1), t-tests contrasted the production phases to the modeled baseline condition (fixation cross) for all four conditions separately. As expected, all four contrasts revealed strong activations within various functional networks each (Figure 2A–D). Visual areas including the occipital lobe and extra-striate areas of the ventral pathway were found throughout all conditions with the highest T-values in the visually cued condition (Figure 2C). The auditory cortex, in particular the STG, was activated as well. During the interactive condition, the temporal cortex pattern seemed more extended compared to the other conditions (Figure 2D). Other activated brain regions, such as the dorsal anterior cingulate cortex (ACC), prefrontal cortex (PFC; in particular middle frontal gyrus, MFG) and inferior parietal lobe (IPL), indicated contributions from executive control

and working memory networks. Like the temporal cluster, these activation patterns were enlarged in the interactive condition. Further activation clusters were located in prosody-related areas including the IFG and pSTG (Figure 2A–D).

In addition to the general activation patterns, we determined the neural correlates of cued and interactive production compared to the free production (Hypotheses H2 and H3). Therefore, paired t-tests contrasted the free production condition with production after auditory and visual cue. No significant difference between the auditory cue and the free production condition emerged at a FWE-corrected threshold. For the contrast visual cue vs free production, the mapping revealed significant activations in the visual cortex, in particular, extra-striate areas of bilateral ventral and right dorsal pathways (cluster-wise $p_{FWE} < 0.05$; Figure 3A; Table 1). The contrast social interaction vs free production resulted in increased activation in several networks (Figure 3B; Table 1): first, higher activation emerged in the prosody processing network with Broca's (IFG) and Wernicke's (pSTG) areas and their right-hemispheric homotopes. Second, the reward system, including ventral tegmental area (VTA) and nucleus accumbens (NAcc), was more active during interactive task. Finally, the

Table 1. Clusters from mapping in Figure 3A–C

Figure	Cluster	Anatomical regions	Brodman areas	Peak voxel (MNI)			Cluster size (voxels)	T values (peak voxel)						
1. Visual cued vs free production														
1.	1	Middle occipital gyrus	17, 18, 19	40	−72	10	3295	7.00						
		Middle temporal gyrus	21, 36, 37											
		Fusiform gyrus	36, 37											
		Inferior occipital gyrus	18, 19											
		Lingual gyrus	18, 19											
		Calcarine gyrus	17											
		Cuneus	18, 19											
		Parahippocampal gyrus												
		Inferior temporal gyrus	20											
	Cerebellum R													
	Superior occipital gyrus R	18, 19												
	2	Lingual gyrus	18, 19	−24	−86	−10	1295	6.72						
		Inferior occipital gyrus	18, 19											
		Fusiform gyrus	36, 37											
		Middle occipital gyrus	17, 18, 19											
		Parahippocampal gyrus												
		Cuneus	18											
		Middle temporal gyrus	36, 37											
		Cerebellum												
		Calcarine gyrus L	17											
	3	Precuneus	7	28	−62	36	724	4.73						
		Superior occipital gyrus R	19											
		Cuneus R	19											
		Angular gyrus R	39											
		Inferior parietal gyrus	39, 40											
		Middle occipital gyrus R	19											
		Superior parietal gyrus	7											
2. Interactive vs free production														
2.		1	Cerebellum L							−14	−78	−32	604	5.80
	Pyramis													
	Lingual gyrus		18, 19											
	2	Lentiform nucleus		36	−4	−14	2142	5.93						
		Putamen												
		Thalamus												
		Pons												
		Parahippocampal gyrus												
		Pallidum												
		Amygdala												
		Hippocampus R	28, 35											
		Caudate												
		Superior temporal gyrus	28, 35, 38											
	3	Insula		42	8	26	3396	7.13						
		Middle temporal gyrus	21, 38											
		Inferior frontal gyrus R	44, 45, 47											
		Middle frontal gyrus R	6, 8, 9, 46											
		Insula	13											
		Precentral gyrus	6											
		Putamen R												
		Superior frontal gyrus	6, 8, 9, 10											
		Clastrum												
		Olfactory gyrus R	45, 47											
	4	Superior temporal gyrus	22, 39, 41,	44	−32	4	4036	9.28						
		Middle temporal gyrus	42											
		Inferior parietal gyrus	21, 37											
		Supramarginal gyrus	39, 40											
		Angular gyrus R	40											
		Insula	39											
		Postcentral gyrus	13											
		Precuneus	5, 40											
		Inferior temporal gyrus R	7											
		Middle occipital gyrus R	37											
			19											

(continued)

Table 1. Continued

Figure	Cluster	Anatomical regions	Brodman areas	Peak voxel (MNI)			Cluster size (voxels)	T values (peak voxel)
3. Laterality coefficient	5	Middle temporal gyrus	21, 37	−44	−68	−10	1070	5.15
		Superior temporal gyrus	22, 39					
		Middle occipital gyrus	19					
		Inferior occipital gyrus	19					
		Inferior temporal gyrus	37					
		Fusiform gyrus	37					
	6	Inferior frontal gyrus	44, 45, 46,	−44	4	22	1520	6.24
		Middle frontal gyrus	47					
		Precentral gyrus	6, 8, 9, 46					
		Superior temporal gyrus	6					
		Insula L	22, 38					
	7		13	−40	−44	56	1280	6.05
		Inferior parietal gyrus	40					
		Superior parietal gyrus	5, 7					
		Postcentral gyrus	2, 5, 40					
		Precuneus	7					
		Supramarginal gyrus	40					
	Angular gyrus L							
1	Superior temporal gyrus	22	48	−34	4	82	4.97	
	Middle temporal gyrus	21						
2	Superior temporal gyrus	22	54	−42	12	103	5.18	
	Middle temporal gyrus R	21						

frontoparietal working memory network, including MFG and IPL, contributed more to the interactive condition as well.

The contrast of social interaction vs free production suggested a right-hemispheric dominance, specifically for areas in the IFG and pSTG. To test this possibility, we calculated a laterality coefficient for this contrast, quantifying regions where activation was stronger in the right as compared with the left hemisphere. For each voxel, the laterality coefficient was calculated by subtracting the value of the corresponding voxel in the other hemisphere. This was performed by subtraction of a right-left flipped version of the same contrast map, i.e. (social interaction vs free) – (social interaction vs free [right-left flipped along the y-axis]) which then entered a t-test on the second level. To assure that only areas of the main contrast (social interaction vs free) were included, an inclusive masking with this contrast was performed. A dominance emerged toward the right-hemispheric homotope of Wernicke's area specifically for the social interaction condition (cluster-wise pFWE < 0.05; Figure 3C; Table 1).

To get further insight into the influence of modality-specific cues on neural patterns of prosody production and its lateralization, ROI analyses addressed bilateral sensory cortices and pSTG-IFG networks. The defined ROIs included bilateral Heschl's gyrus and primary visual area V1 as well as IFG (Broca's area) and pSTG (Wernicke's area and their respective right-hemispheric homotopes). To test hemispheric specialization in the different prosodic conditions, we calculated separate two-factorial ANOVAs (hemisphere × condition) for the anatomical structures. A main effect of hemisphere was observed only in V1 [$F(1, 22) = 14.31, P < 0.01$]. Main effects of conditions were observed in pSTG [$F(3, 20) = 3.08, P < 0.05$] and, on a trend level, in V1 [$F(3, 20) = 2.74, P = 0.05$] and IFG [$F(3, 20) = 2.84, P = 0.06$]. Notably, a significant hemisphere × condition interaction emerged for pSTG only [$F(3, 20) = 4.74, P < 0.01$] but none of the other regions [Heschl's gyrus: $F(3, 20) = 1.57, P = 0.21$;

V1: $F(3, 20) = 1.10, P = 0.35$; IFG: $F(3, 20) = 0.56, P = 0.59$]. Post-hoc t-tests on the laterality coefficient (right-left) compared all experimental conditions for the four anatomical regions. After Bonferroni correction for multiple comparisons, the contrast interactive vs. free in right pSTG yielded the only significant result of all comparisons [$t(22) = 4.54, P < 0.001$; Figure 4].

Discussion

Our study investigated neural networks during emotional prosody production in unrestricted, cued and social contexts. During all conditions, distributed network activity emerged in line with the three-step model of prosody perception (Schirmer and Kotz, 2006) and, in particular, right-lateralized IFG and pSTG activity, i.e. in Broca's and Wernicke's areas and their right-hemispheric homotopes. Compared with uncued conditions, sensory responses were larger in the dorsal and the ventral visual stream after visual cues and in auditory cortices after auditory and visual cues. The latter finding is in line with encoding and mapping of color code to the vocal emotion. For the first time, we investigated social interaction using emotional prosody; the participants received an emotion feedback from the experimenter on their prosody production. The expectation and processing of the social feedback led to higher activity in bilateral language network and basal ganglia exhibiting right-lateralized pSTG activity. In context of vocal emotions, the right pSTG appeared as a central hub for social communication in concert with reward processing.

In line with Hypothesis H1, production of emotional prosody activated a right-lateralized frontotemporal network, along with bilateral basal ganglia, motor, premotor and sensory regions in all four conditions. Previous neuroimaging studies established a central role of temporal and inferior frontal regions in supporting emotional prosody (e.g. Schirmer and Kotz, 2006).

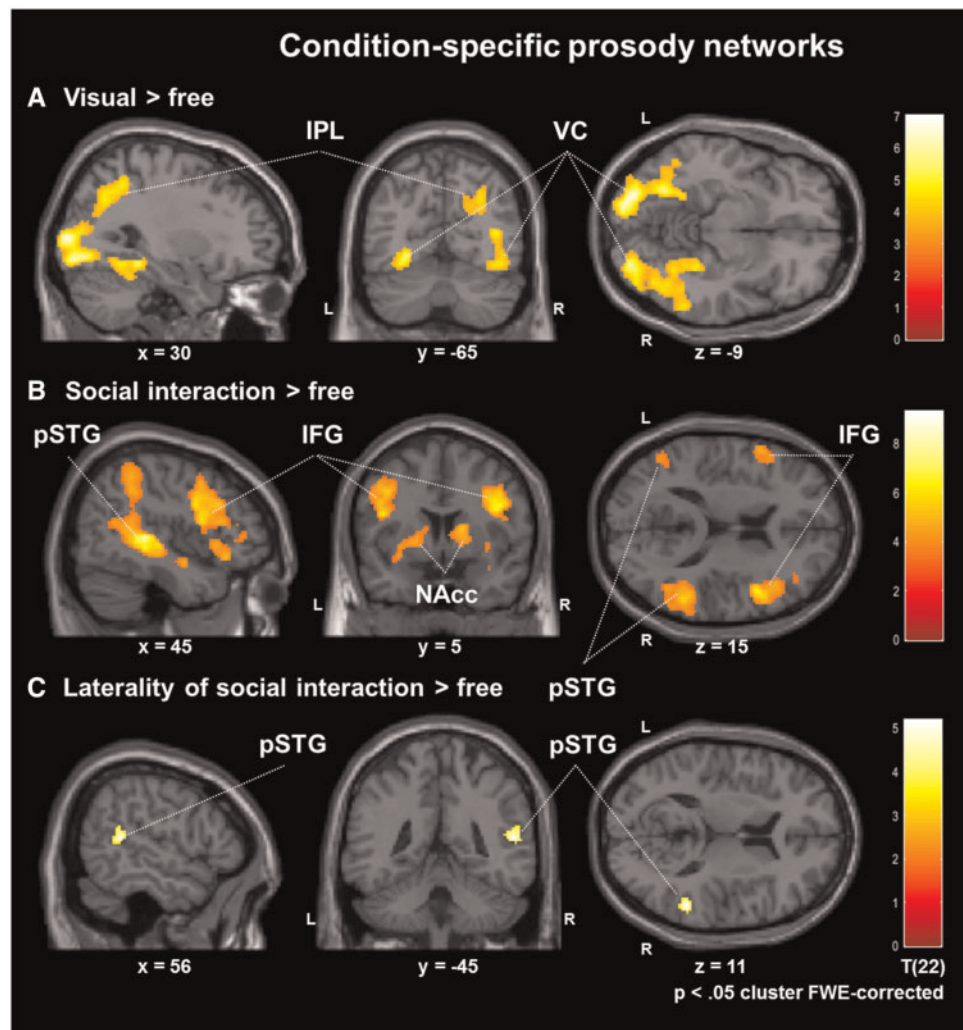


Fig. 3. Condition-specific prosody networks. Compared with free production, visual cues revealed activations in the ventral and dorsal pathways (A). Social interaction vs free production increased activation in the prosody processing network with Broca's (IFG) and Wernicke's (pSTG) areas and their right-hemispheric homotopes. Moreover, the reward system and the frontoparietal working memory network emerged specifically for the social interaction condition (B). Lateralization for networks of social interaction was determined by the laterality coefficient (social interaction vs free) – [social interaction vs free (right-left flipped)]. After masking with the main contrast, dominance emerged in the right pSTG for social interaction (C; cluster-wise pFWE < 0.05).

In specific, the IFG and the superior temporal cortex have traditionally been described as supporting general aspects of prosody processing (Ethofer et al., 2009; Park et al., 2010). Recent research has shed more light on the specific functions of these areas: Brück et al. (2011) observed that during emotional prosody processing, activation in right IFG and pSTG was driven by task demands, whereas activity in the middle portions of the STG were associated with stimulus properties. Thus, for prosody comprehension, activity in IFG and pSTG seems to reflect cognitive processing rather than mere stimulus encoding (Brück et al., 2011).

During prosody production, IFG and STG (anterior and posterior portions) have been associated with motor components, with a bilateral IFG and a right-lateralized STG pattern (Pichon and Kell, 2013). However, the specific contribution of the right pSTG seems to extend beyond motor execution. Brück et al. (2011) describe the role of the pSTG as integrating the prosodic cue with internal representations of social emotions. The pSTG thus seems to support the processing of social aspects. Kreifelts et al. (2013) accordingly highlighted the role of this region for

the skill acquisition for non-verbal emotional communication. In line with this notion, our study also found enhanced right pSTG activation specifically in the interactive condition (cf. Hypothesis H3). Krach et al. (2008) emphasize that pSTG activity is a specific signature of human-to-human (not human-to-machine) interactions. In a similar vein, real-life social interactions have also consistently been associated with the right pSTG (Redcay et al., 2010). Accordingly, this region has recently been described as being involved in both social interaction and attentional reorientation (Krall et al., 2015). These two functions may be closely intertwined; social interaction requires understanding the partner's message and expressing an adequate response. The right pSTG may thus support rapid switches between external stimulation and internal cognitive processes in social communication. In summary, there is strong evidence that the right pSTG supports functions specific to social communication.

Our results confirmed the assumption of a distributed cortico-striatal network of IFG, STG and reward system for the free production of prosody. These findings are well in line with

ROI activations to different prosody production conditions

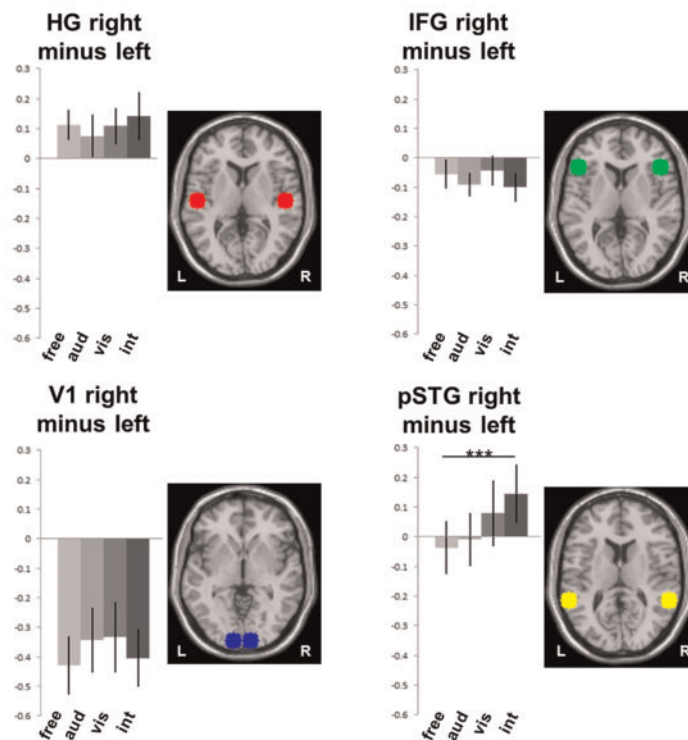


Fig. 4. ROI activations to different prosody production conditions. To investigate the modulating effects of cue modality and social context, ROI analyses addressed specifically bilateral sensory cortices V1 and Heschl's gyrus and pSTG-IFG networks. t-Tests on the laterality coefficient (right-left) compared all experimental conditions for the four anatomical regions. After Bonferroni correction for multiple comparisons, the contrast social interaction vs free in right pSTG yielded the only significant result of all comparisons [$t(22) = 4.54$, $P < 0.001$].

previous studies (Aziz-Zadeh et al., 2010; Pichon and Kell, 2013; Belyk and Brown, 2016). This is an interesting, yet a somewhat unspecific finding; the complexity of the task makes it difficult to disentangle functional contributions of single brain regions (Mathiak and Weber, 2006). To solve this issue, we compared free production to cued production in auditory and visual domains. During the cued conditions, sensory activation patterns corresponded to the respective domains (cf. Hypothesis H2). Effects of auditory cues in bilateral Heschl's gyrus were only observed at the ROI level but not at the level of whole-brain contrasts. Neural responses were pronounced after visual cues, first, in primary visual areas (V1) and in higher association cortices of the ventral and dorsal pathways. It is well established that attention to a sensory modality increases cortical responses in sensory areas of this modality (Mathiak et al., 2007). Second, visual cues yielded similar effect sizes as auditory cues in bilateral Heschl's gyrus. Previous research has shown that visual processing pathways reach far down into the STG, making it a central hub for the multimodal integration of auditory and visual signals (Robins et al., 2009; Klasen et al., 2011). Visual cues can influence responses in auditory cortex (Harrison and Woodhouse, 2016) and facilitate processing in the auditory domain (Wolf et al., 2014). Moreover, the STG supports auditory but not visual sensory imagination (Zvyagintsev et al., 2013). Taken together, this suggests that visual cues are first translated into an auditory prosodic imagination before verbalization. In contrast to the cues, the interactive condition did not increase responses in the sensory cortices.

Separate memory systems support visual and auditory processing (Burton et al., 2005). In our study, visual processing was

only relevant for the prosody selection after visual cues. Moreover, this represented a cross-modality task. The visual color cue was transformed into an auditory-vocal category. Therefore, in addition to the visual cue processing, a recoding at the level of auditory-vocal representation needs to be achieved. Such process may lead to the observed activation increase in the auditory domain in addition to the visual encoding (see the cross-modality model in Nakajima et al., 2015). An alternative explanation may follow the line of dominant modality. For emotional prosody, the auditory modality is dominant to the visual (Repp and Penel, 2002). The auditory activity increase after visual cues may be an increase of selective attention to extract the missing prosodic information from the auditory stimuli (Woodruff et al., 1996).

The role of the reward system during social interactions is well established (Walter et al., 2005). Similar to monetary reward, social reward is elicited for instance by a smile, a good reputation or sharing a positive outcome with somebody and activates the striatum and the PFC (Izuma et al., 2008; Fareri et al., 2012; Mathiak et al., 2015). Accordingly, the present study revealed higher activation of the striatum and the PFC in the interactive than in the free production condition. Indeed, in video games, interactivity was required for the elicitation of reward responses (Kätsyri et al., 2013a). Reward responses to social interactions seem to be specifically impaired in substance abuse (Yarosh et al., 2014) and thus may represent a specific component. Notably, the produced emotions in the social interaction condition did not differ in their frequency from the other conditions. Therefore, enhanced activation in the reward system cannot be attributed to more positive (i.e. happy) emotional

cues. Even if not paired with rewarding valence, emotional interaction may thus elicit rewarding responses in striatal and PFC structures (Krach et al., 2010). Despite a considerable number of studies on socially rewarding stimuli, it remains unclear whether social interactions are rewarding per se. Recent evidence, however, suggests that the latter is the case (Fareri and Delgado, 2014). Interaction with a human partner exerts stronger social influence on an individual than computer-generated interaction (Fox et al., 2015). This distinction can be observed at the neural level as well (Fareri and Delgado, 2013). In specific, responses of the reward system seem to depend strongly on the social context. For instance, compared with a computer opponent, game interactions with a human partner led to enhanced effects in the reward system (Kätysyri et al., 2013b). We conclude that the social component adds an additional rewarding value to interactivity.

MFG and IPL activation suggest an enhanced involvement of working memory during the interactive condition. Functional imaging studies point out a specialized social working memory for maintenance and manipulation of social cognitive information, which allows humans to understand complex multifaceted social interactions (Meyer et al., 2012; Meyer et al., 2015). Specifically, social working memory involves processing in a neural network encompassing MFG, supplementary motor area (SMA), and superior and inferior parietal regions (Meyer et al., 2012). However, this study and previous literature on the social working memory employ only social rating or n-back tasks without any interactive component (Thornton and Conway, 2013; Meyer et al., 2015; Xin and Lei, 2015). According to our findings, interactivity may increase the involvement of social working memory supporting online adjustment of ongoing social behavior to the dynamic context (compare Schmidt and Cohn, 2001).

Right-hemispheric IFG and STG have been associated with the recognition (STG) and the evaluation (IFG) of prosody processing in previous studies (Schirmer and Kotz, 2006; Ethofer et al., 2009; Park et al., 2010; Frühholz et al., 2015). However, our results suggest further contributions of the IFG which can be ascribed to the interactive task. Notably, the IFG's role in the mirror neuron system (Aziz-Zadeh et al., 2006; Molnar-Szakacs et al., 2009) may be of relevance. Previous studies already underline the function of the mirror neuron system during language development (Rizzolatti and Arbib, 1998) and speech perception (Aziz-Zadeh and Ivry, 2009). In specific, mirror neurons may underlie the ability to map acoustic representations onto their corresponding motor gestures (Aziz-Zadeh and Ivry, 2009). Furthermore, mirror neuron activity correlates with social cognition and emotion processing (Enticott et al., 2008). Since these functions were not necessary during free production, interactive communication may require higher involvement of the mirror neuron system. The missing activation of the mirror neuron system during the auditorily cued condition may be attributed to the involvement of the phonological loop, maintaining phonemic information for several seconds (Salamé and Baddeley, 1982).

The right hemisphere exhibits higher proficiency in the processing of pitch, frequency and intensity (Mathiak et al., 2002; Liebenthal et al., 2003; Schirmer and Kotz, 2006). These cues are linked to prosody perception. Indeed, right hemisphere lateralization was found during prosody comprehension (Kotz et al., 2006). Our study confirmed the right-hemispheric specialization of the pSTG (homotope of Wernicke's area). Furthermore, this structure responded specifically to social prosodic interaction. Notably, such lateralization did not emerge at the IFG. The extraction of specific acoustic cues from complex speech signals

(i.e. supra-segmental features such as pitch contours and rhythmic structures) has been associated with the posterior temporal cortex (Wildgruber et al., 2005). In contrast, prosody production involves the right dorsolateral PFC (Brück et al., 2011). Conceivably, the evaluation of the feedback takes precedence over the production during interactive prosodic communication. Specifically, enhanced activation in pSTG may reflect a prosodic pattern matching process (Wildgruber et al., 2006): In social communication, responses from the interaction partner will be compared with the expected pattern (Burgoon, 1993). Conceivably, the right pSTG is a processing unit for prosody semantics—in analogy to its left-hemispheric counterpart (Wernicke's area), which codes the semantics of language (Démonet et al., 1992).

Limitations

The present study has some limitations that have to be taken into account when interpreting the results. First, sound recordings from eight participants were incomplete and thus offline codings were performed on two-thirds of the participants only. We cannot exclude the possibility that this may have influenced the results; however, we have no reason to assume that the remaining 15 participants were not representative for the entire sample. The losses concerned data from four female and four male participants, leaving the gender proportions of the sample intact. Second, for the free and social interaction conditions, we do not know for sure what the intended emotion of the participant was. A separate assessment in each trial might have been a solution; however, this would have disrupted both the timing and the psychological processes and therefore also the comparability with the other conditions.

Another limitation concerns the aspect of ecological validity. Although the task (and, as a consequence, the behavior of the participants) was designed to overcome some of the limitations of other studies, the setting remains the one of a neuroscientific experiment, with a scanner environment and sensory stimulation via headphones and video goggles. Also, the subjects were aware that the interactive communication in the social condition took place via pre-recorded stimuli (pseudowords) and not via the actual voice of the communication partner. The setting thus differs considerably from those where prosody production and communication take place in everyday life.

In a social interaction, any type of feedback by the interaction partner will also be inherently social. This, however, makes it difficult to distinguish two core aspects of social interaction: the social aspect (i.e. an interaction with a human partner) and the interactive aspect (i.e. receiving feedback in the interaction). Also, communications via emotional prosody will usually involve a human interaction partner. The present study thus cannot distinguish between those aspects, which must be considered a limitation of the results from Condition 4.

Conclusion

The present study separated for the first time cue-specific activations from prosodic networks. Compared with auditory signals, visual cues seemed to undergo an additional processing step, transforming them to auditory representations before verbalization. The visual cues increased activation in sensory cortices, IFG, and pSTG relative to uncued prosody production. Thus, cues for prosody production did not only affect sensory areas but higher levels of cortical processing as well. Moreover, social interaction influenced the prosody production network. In

addition to reward responses, the social feedback enhanced activity in the prosody network of IFG and pSTG. Extending findings from previous research, right-lateralization for emotional prosodic communication was only observed in the pSTG. We thus established this region as a central unit for the social aspects of emotional prosody.

Funding

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft (DFG), IRTG 1328, IRTG 2150, and MA 2631/6-1), the Federal Ministry of Education and Research (APIC: 01EE1405A, APIC: 01EE1405B, and APIC: 01EE1405C), and the Interdisciplinary Center for Clinical Research (ICCR) Aachen.

Conflict of interest. None declared.

References

- Altmann, C.F., Bledowski, C., Wibrall, M., Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *Neuroimage*, *35*(3), 1192–200.
- Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., Iacoboni, M. (2006). Lateralization of the human mirror neuron system. *The Journal of Neuroscience*, *26*(11), 2964–70.
- Aziz-Zadeh, L., Ivry, R.B. (2009). The human mirror neuron system and embodied representations. *Advances in Experimental Medicine and Biology*, *629*, 355–76.
- Aziz-Zadeh, L., Sheng, T., Gheyntanchi, A. (2010). 'Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS One*, *5*(1), e8759.
- Belyk, M., Brown, S. (2016). Pitch underlies activation of the vocal system during affective vocalization. *Social Cognitive and Affective Neuroscience*, *11*(7), 1078–88.
- Brück, C., Kreifelts, B., Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, *8*(4), 383–403.
- Burgoon, J.K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology*, *12*(1–2), 30–48.
- Burton, M.W., LoCasto, P.C., Krebs-Noble, D., Gullapalli, R.P. (2005). A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage*, *26*(3), 647–61.
- Démonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., Rascol, A., Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, *115*(Pt 6), 1753–68.
- Enticott, P.G., Johnston, P.J., Herring, S.E., Hoy, K.E., Fitzgerald, P.B. (2008). Mirror neuron activation is associated with facial emotion processing. *Neuropsychologia*, *46*(11), 2851–4.
- Ethofer, T., Kreifelts, B., Wiethoff, S., et al. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, *21*(7), 1255–68.
- Fareri, D.S., Niznikiewicz, M.A., Lee, V.K., Delgado, M.R. (2012). Social network modulation of reward-related signals. *The Journal of Neuroscience*, *32*(26), 9045–52.
- Fareri, D.S., Delgado, M.R. (2013). Differential reward responses during competition against in- and out-of-network others. *Social Cognitive and Affective Neuroscience*, *9*(4), 412–20.
- Fareri, D.S., Delgado, M.R. (2014). Social rewards and social networks in the human brain. *The Neuroscientist*, *20*(4), 387–402.
- Fox, J., Ahn, S.J., Janssen, J.H., Yeykelis, L., Segovia, K.Y., Bailenson, J.N. (2015). Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Human-Computer Interaction*, *30*(5), 401–32.
- Frühholz, S., Klaas, H.S., Patel, S., Grandjean, D. (2015). Talking in fury: the cortico-subcortical network underlying angry vocalizations. *Cerebral Cortex*, *25*(9), 2752–62.
- Gordon, I., Eilbott, J.A., Feldman, R., Pelphrey, K.A., Vander Wyk, B.C. (2013). Social, reward, and attention brain networks are involved when online bids for joint attention are met with congruent versus incongruent responses. *Society for Neuroscience*, *8*(6), 544–54.
- Harrison, N.R., Woodhouse, R. (2016). Modulation of auditory spatial attention by visual emotional cues: differential effects of attentional engagement and disengagement for pleasant and unpleasant cues. *Cognitive Processing*, *17*(2), 205–11.
- Hayes, A.F., Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89.
- Izuma, K., Saito, D.N., Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*(2), 284–94.
- Kätsyri, J., Hari, R., Ravaja, N., Nummenmaa, L. (2013a). Just watching the game ain't enough: striatal fMRI reward responses to successes and failures in a video game during active and vicarious playing. *Frontiers in Human Neuroscience*, *7*, 278.
- Kätsyri, J., Hari, R., Ravaja, N., Nummenmaa, L. (2013b). 'The Opponent matters: elevated fMRI reward responses to winning against a human versus a computer opponent during interactive video game playing. *Cerebral Cortex*, *23*(12), 2829–39.
- Klaas, H.S., Frühholz, S., Grandjean, D. (2015). Aggressive vocal expressions—an investigation of their underlying neural network. *Frontiers in Behavioral Neuroscience*, *9*, 121.
- Klasen, M., Kenworthy, C.A., Kircher, T.T.J., Mathiak, K.A., Mathiak, K. (2011). Multimodal representation of emotions. *The Journal of Neuroscience*, *31*(38), 13635–43.
- Klasen, M., Chen, Y.-H., Mathiak, K. (2012). Multisensory emotions: perception, combination and underlying neural processes. *Reviews in the Neurosciences*, *23*(4), 381–392.
- Kotz, S.A., Meyer, M., Paulmann, S. (2006). Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. *Progress in Brain Research*, *156*, 285–94.
- Kotz, S.A., Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, *5*(3), 108–25.
- Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS One*, *3*(7), e2597.
- Krach, S., Paulus, F., Bodden, M., Kircher, T. (2010). The rewarding nature of social interactions. *Frontiers in Behavioral Neuroscience*, *4*, 22.
- Krall, S.C., Rottschy, C., Oberwelland, E., et al. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, *220*(2), 587–604.
- Kreifelts, B., Jacob, H., Brück, C., Erb, M., Ethofer, T., Wildgruber, D. (2013). Non-verbal emotion communication training induces specific changes in brain function and structure. *Frontiers in Human Neuroscience*, *7*, 648.
- Liebethal, E., Binder, J.R., Piorowski, R.L., Remez, R.E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience*, *15*(4), 549–58.
- Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fmri data sets. *Neuroimage*, *19*(3), 1233–9.

- Mathiak, K., Hertrich, I., Lutzenberger, W., Ackermann, H. (2002). Functional cerebral asymmetries of pitch processing during dichotic stimulus application: a whole-head magnetoencephalography study. *Neuropsychologia*, **40**(6), 585–93.
- Mathiak, K., Weber, R. (2006). Toward brain correlates of natural behavior: fMRI during violent video games. *Human Brain Mapping*, **27**(12), 948–56.
- Mathiak, K., Menning, H., Hertrich, I., Mathiak, K.A., Zvyagintsev, M., Ackermann, H. (2007). Who is telling what from where? A functional magnetic resonance imaging study. *NeuroReport*, **18**(5), 405–9.
- Mathiak, K.A., Alawi, E.M., Koush, Y., et al. (2015). Social reward improves the voluntary control over localized brain activity in fMRI-based neurofeedback training. *Frontiers in Behavioral Neuroscience*, **9**, 136.
- Meyer, M.L., Spunt, R.P., Berkman, E.T., Taylor, S.E., Lieberman, M.D. (2012). Evidence for social working memory from a parametric functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(6), 1883–8.
- Meyer, M.L., Taylor, S.E., Lieberman, M.D. (2015). Social working memory and its distinctive link to social cognitive ability: an fMRI study. *Social Cognitive and Affective Neuroscience*, **10**(10), 1338–47.
- Molnar-Szakacs, I., Wang, M.J., Laugeson, E.A., Overy, K., Wu, W.-L., Piggot, J. (2009). Autism, emotion recognition and the mirror euron system: the case of music., *McGill Journal of Medicine*, **12**(2), 87–98.
- Nakajima, J., Kimura, A., Sugimoto, A., Kashino, K. (2015). Visual attention driven by auditory cues—selecting visual features in synchronization with attracting auditory events. In: He, X., Luo, S., Tao, D., Xu, C., Yang, J., Hasan, M.A., editors. *mmm 2015*, Cham, Switzerland: Springer International Publishing.
- Park, J.-Y., Gu, B.-M., Kang, D.-H., et al. (2010). Integration of cross-modal emotional information in the human brain: an fMRI study. *Cortex*, **46**(2), 161–9.
- Pichon, S., Kell, C.A. (2013). Affective and sensorimotor components of emotional prosody generation. *The Journal of Neuroscience*, **33**(4), 1640–50.
- Redcay, E., Dodel-Feder, D., Pearrow, M.J., et al. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *Neuroimage*, **50**(4), 1639–47.
- Repp, B.H., Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, **28**(5), 1085–99.
- Rizzolatti, G., Arbib, M.A. (1998). Language within our grasp. *Trends in Neurosciences*, **21**(5), 188–94.
- Robins, D.L., Hunyadi, E., Schultz, R.T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain and Cognition*, **69**(2), 269–78.
- Salamé, P., Baddeley, A. (1982). Disruption of short-term memory by unattended speech: implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, **21**(2), 150–64.
- Schirmer, A., Kotz, S.A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, **10**(1), 24–30.
- Schmidt, K.L., Cohn, J.F. (2001). Human facial expressions as adaptations: evolutionary questions in facial expression research. *American Journal of Physical Anthropology*, **116**(S33), 3–24.
- Schmuckler, M.A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, **2**(4), 419–36.
- Thönnessen, H., Boers, F., Dammers, J., Chen, Y.-H., Norra, C., Mathiak, K. (2010). Early sensory encoding of affective prosody: neuromagnetic tomography of emotional category changes. *Neuroimage*, **50**(1), 250–9.
- Thornton, M.A., Conway, A.R.A. (2013). Working memory for social information: chunking or domain-specific buffer? *Neuroimage*, **70**, 233–9.
- Walter, H., Abler, B., Ciaramidaro, A., Erk, S. (2005). Motivating forces of human actions: neuroimaging reward and social interaction. *Brain Research Bulletin*, **67**(5), 368–81.
- Wildgruber, D., Riecker, A., Hertrich, I., et al. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage*, **24**(4), 1233–41.
- 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.
- Witteman, J., Van Heuven, V.J., Schiller, N.O. (2012). Hearing feelings: a quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, **50**(12), 2752–63.
- Wolf, D., Schock, L., Bhavsar, S., Demenescu, L.R., Sturm, W., Mathiak, K. (2014). Emotional valence and spatial congruency differentially modulate crossmodal processing: an fMRI study. *Frontiers in Human Neuroscience*, **8**, 659.
- Woodruff, P.W.R., Benson, R.R., Bandettini, P.A., et al. (1996). Modulation of auditory and visual cortex by selective attention is modality-dependent. *NeuroReport*, **7**(12), 1909–13.
- Xin, F., Lei, X. (2015). Competition between frontoparietal control and default networks supports social working memory and empathy. *Social Cognitive and Affective Neuroscience*, **10**(8), 1144–52.
- Yarosh, H.L., Hyatt, C.J., Meda, S.A., et al. (2014). Relationships between reward sensitivity, risk-taking and family history of alcoholism during an interactive competitive fMRI task. *PLoS One*, **9**(2), e88188.
- Zvyagintsev, M., Klasen, M., Mathiak, K.A., Weber, R., Edgar, J.C., Mathiak, K. (2010). Real-time noise cancellation for speech acquired in interactive functional magnetic resonance imaging studies. *Journal of Magnetic Resonance Imaging*, **32**(3), 705–13.
- Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K.A., Sack, A.T., Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience*, **37**(9), 1421–34.