OXFORD

Loud and angry: sound intensity modulates amygdala activation to angry voices in social anxiety disorder

Doerte Simon[1],[*], Michael Becker[1],[*], Martin Mothes-Lasch[1], Wolfgang H.R. Miltner[2] and Thomas Straube[1]

[1]Institute of Medical Psychology and Systems Neuroscience, University of Muenster, Von-Esmarch-Str., 52, D-48149 Münster, Germany and [2]Department of Biological and Clinical Psychology, Friedrich Schiller University, Jena

[*]These authors contributed equally to this work
Correspondence should be addressed to Thomas Straube, Institute of Medical Psychology and Systems, Neuroscience, University Hospital Muenster, Von-Esmarch-Str., 52, D-48149 Münster, Germany. Email: thomas.straube@uni-muenster.de

Abstract

Angry expressions of both voices and faces represent disorder-relevant stimuli in social anxiety disorder (SAD). Although individuals with SAD show greater amygdala activation to angry faces, previous work has failed to find comparable effects for angry voices. Here, we investigated whether voice sound-intensity, a modulator of a voice's threat-relevance, affects brain responses to angry prosody in SAD. We used event-related functional magnetic resonance imaging to explore brain responses to voices varying in sound intensity and emotional prosody in SAD patients and healthy controls (HCs). Angry and neutral voices were presented either with normal or high sound amplitude, while participants had to decide upon the speaker's gender. Loud *vs* normal voices induced greater insula activation, and angry *vs* neutral prosody greater orbitofrontal cortex activation in SAD as compared with HC subjects. Importantly, an interaction of sound intensity, prosody and group was found in the insula and the amygdala. In particular, the amygdala showed greater activation to loud angry voices in SAD as compared with HC subjects. This finding demonstrates a modulating role of voice sound-intensity on amygdalar hyperresponsivity to angry prosody in SAD and suggests that abnormal processing of interpersonal threat signals in amygdala extends beyond facial expressions in SAD.

Key words: Social anxiety disorder; fMRI; prosody; amygdala; insula; sound intensity

Introduction

Social anxiety disorder (SAD) is characterized by persistent fear responses in social interactions or performance situations (APA, 2000). Cognitive-behavioral models of social anxiety propose dysfunctional allocation of processing resources to cues associated with negative evaluation (Rapee and Heimberg, 1997). In search for neural foundations of those processing biases, abnormal brain responses in different brain areas have been found in SAD individuals by means of functional brain imaging (Freitas-Ferrari *et al.*, 2010). In particular, greater activation of the amygdala has reliably been observed during processing of disorder-related stimuli such as angry facial expressions

(Schulz *et al.*, 2013). This is in line with neurobiological models proposing a central role of the amygdala for processing of threat-related signals (LeDoux, 1996; Ohman and Mineka, 2001; Straube *et al.*, 2011). Due to interconnections to brain stem areas, the hypothalamus, and cortical areas, the amygdala is critically involved in alerting responses and the modulation of autonomic, perceptual, and emotional processing of threatening and fear-related stimuli (LeDoux, 2000; Tamietto and de Gelder, 2010; Lipka *et al.*, 2011).

In daily life, angry faces are one of the most informative indicators of others' disapproval. Therefore, previous research has predominantly used this class of stimuli to investigate abnormal behavioral and neural responses to social threat signals

in SAD (Freitas-Ferrari *et al.*, 2010; Schulz *et al.*, 2013). However, faces are not the only stimulus class signaling the emotional state of others. Voices may also convey powerful emotional signals during social interaction. Detection and identification of emotional prosody are critical for the appreciation of the social environment, since prosodic information yields important cues to potential distress or safety characteristic of a social situation by directly indicating social approval and disapproval. Accordingly, angry emotional prosody represents a strong signal of social threat and rejection, which might be particularly relevant in SAD. However, a previous fMRI study investigating the processing of angry prosody in SAD (Quadflieg *et al.*, 2008) found similar responses in the amygdala, insula, and fronto-temporal regions to angry as compared with neutral prosody in both, SAD patients and healthy control (HC) participants. Greater activation in patients relative to controls was only observed in the orbitofrontal cortex (OFC), supporting the role of the OFC in enhanced processing of prosodic information (Sander *et al.*, 2005; Schirmer *et al.*, 2008).

According to these findings, one might expect a difference between processing of threatening faces and voices in SAD, with reliable amygdala hyper-responsivity only to faces but not to voices. However, it has to be considered that emotional states expressed in voices are reliably communicated by a combination of three perceptual dimensions—pitch, time and sound intensity (subjectively rated as loudness) (Pittam and Scherer, 1993). Most studies investigating neural correlates of emotional prosody have used normalized sound amplitudes across emotional stimuli (Quadflieg *et al.*, 2008; Ethofer *et al.*, 2009; Mothes-Lasch *et al.*, 2011) to control for differences between emotions not related to prosody. In these studies, sound intensity was treated as a confounding variable, thereby ignoring that sound intensity is a fundamental dimension in the construct of emotional prosody (Pittam and Scherer, 1993). Processing of angry prosody is modulated by the sound intensity of stimuli, which varies with the strength of the emotional response of the speaker and also with his/her physical distance (Schirmer *et al.*, 2007). As sound intensity of angry voices is associated with the physical distance as well as the emotional intensity of a potentially dangerous person, it signals the immediacy, closeness, and relevance of threat. Hence, this information should be prioritized in the information processing stream in SAD, in keeping with similar findings for other social threat-associated stimuli.

Under laboratory conditions, normal sound amplitude of angry voices might not suffice to induce differential amygdala responses to angry *vs* neutral prosody in SAD because angry voices per se might not represent sufficiently intense threats. Rather, it is conceivable that only an angry and loud voice abnormally activates the amygdala in SAD. This study used a factorial design with prosody and sound intensity as independent factors in order to investigate the role of sound intensity on brain responses to angry *vs* neutral prosodic stimuli in SAD. In particular, we predicted that greater sound intensity of angry utterances ('loud and angry') provokes greater activation in the amygdala and in other emotion-related brain areas in SAD as compared with healthy participants.

Materials and methods

Participants

Twenty subjects with SAD (12 women; mean, 29.20 years ± 7.44 years) and 20 control subjects (11 women; mean, 26.85 ± 6.29 years) took part in this study. All were right-handed with normal or corrected to normal vision. Groups were matched for age [$t(38) = 1.08$, $P = 0.287$], gender ($\chi^2 = 0.10$, $P = 0.749$) and education level ($U = 1.28$, $P = 0.355$) (Table 1). Participants were recruited via public announcement and gave written informed consent prior to participation. The study was approved by the Ethics Committee of the University of Jena. For SAD patients, diagnosis was confirmed by a trained clinical psychologist administering the Structured Clinical Interview for DSM-IV Axis I and II disorders (SCID I and II, Wittchen *et al.*, 1996; Fydrich *et al.*, 1997). Participants with current psychotropic medication, ongoing major-depressive episode, psychotic disorder, and history of seizures or head injury with impairment of consciousness were excluded. In the SAD sample, comorbidities included specific phobia ($n = 1$), panic disorder ($n = 1$), obsessive compulsive disorder ($n = 1$), generalized anxiety disorder (partial remission, $n = 1$) and depressive episodes in the past ($n = 5$). Five Patients also met the criteria of an Axis II avoidant personality disorder. HC subjects were free of any psychopathology and medication. After scanning, all participants completed the LSAS (Liebowitz Social Anxiety Scale, German version, Stangier and Heidenreich, 2005) and the BDI (Beck Depression Inventory, German version, Hautzinger *et al.*, 1995) questionnaire. SAD patients scored significantly higher on LSAS [$t(38) = 9.76$, $P < 0.05$] and BDI [$t(38) = 3.10$, $P < 0.05$] than HC participants.

Stimuli

All prosodic stimuli were evaluated in previous studies (Quadflieg *et al.*, 2008; Mothes-Lasch *et al.*, 2011) and consisted of a set of 20 semantically neutral bisyllabic nouns (five letters) spoken in either angry or neutral prosody by two women and two men. Stimuli were recorded and digitized through an audio interface at a sampling rate of 44 kHz with 16 bit resolution. Utterances were normalized in amplitude (70%) and edited to a common length of 550 ms using Adobe Audition (v1.5, Adobe Systems, San Jose, CA). Loudness of the stimuli was varied by presenting the stimuli at two different sound intensity levels according to individually adjusted hearing levels in the scanner environment. Relative to a 'loud but not painful' sound intensity level, which was determined in a stepwise manner for every participant individually using an independent set of neutral and angry voice stimuli, an attenuation factor of 0.2 was applied on both channels to voice stimuli during the 'normal' intensity condition. At the headphone speakers, maximal sound pressure level was 67 dB (A) (SEM: 1.55) for the normal sound intensity condition and 84 dB (A) (SEM: 1.94) for the loud condition. Individual sound intensity level calibrations did not differ between groups [$F(1,38) = 0.850$, $p = 0.362$].

Experimental design

Auditory stimuli were presented binaurally via scanner-compatible headphones (Commander XG MRI audio system, Resonance Technology, Northridge, CA). Participants were instructed to listen to the words and to decide whether they were spoken by a male or a female speaker, and then to press a button of an optic fiber response box with their right index or middle finger. Gender-button assignment was counterbalanced across subjects. Prior to scanning participants performed several practice trials to become acquainted with the task. The actual experimental task consisted of 160 trials, with each stimulus presented twice, yielding 40 trials per condition. During the task, a central white fixation cross was projected onto a screen inside the scanner bore. Acoustic stimuli were

Table 1. Overview of demographic and mood characteristics for patients with SAD and HCs concerning age, gender, education level, symptom severity (LSAS) and depression (BDI)

	SAD	HC	group-statistic
Age (SD)	29.20 (7.4)	26.85 (6.2)	$t(38) = 1.08$, $P = 0.287$
Gender (female: male)	12: 8	11: 9	$\chi^2 = 0.10$, $P = 0.749$
education (10 years: 12 years: > 12 years)	3: 3: 14	0: 3: 17	$U = 1.28$, $P = 0.355$
LSAS (SD)	64.0 (16.7)	18.4(12.6)	$t(38) = 9.76$, $P < 0.001$
BDI (SD)	13.8 (12.3)	4.45 (5.5)	$t(38) = 3.10$, $P = 0.004$

LSAS, Liebowitz Social Anxiety Scales, german version (20); BDI, Beck Depression Inventory, german version (21).

presented in four pseudorandomized orders. Inter-stimulus-intervals were jittered within a range of 1800–5500 ms. Accuracy and latency of responses were recorded. After scanning, participants rated all presented acoustic stimuli on a nine-point Likert scale to assess pleasantness (1 = very unpleasant, 9 = very pleasant) and arousal (1 = not arousing, 9 = very arousing). Behavioral data were analyzed using Analysis of Variance (ANOVA) and independent sample as well as pairwise *t*-tests using SPSS software (Version 22, IBM Corp., Armonk, NY). 2 × 2 × 2 level repeated measures ANOVA with group (SAD *vs* HC) as between group factor and sound intensity level (loud *vs* normal) and emotion (angry *vs* neutral prosody) as within-groups factors were used to analyze hit rates, reaction times, and ratings of pleasantness and arousal. *Post-hoc*, *t*-tests were calculated.

fMRI data acquisition and analysis

Scanning was performed in a 3-Tesla magnetic resonance scanner (Magnetom Trio, Tim System 3T; Siemens Medical Systems, Erlangen, Germany). After acquisition of a T1-weighted anatomical scan, one run of T2*-weighted echo planar images consisting of 335 volumes was recorded (TE, 30 ms; TR = 2080 ms, flip angle, 90°; matrix, 64 × 64; field of view, 192 mm^2, voxel size = 3 × 3 × 3 mm^3). Each volume comprised 35 axial slices (slice thickness 3 mm; interslice gap 0.5 mm; in-plane resolution 3 × 3 mm) which were acquired with caudally tilted orientation relative to the anterior–posterior commissure line in order to reduce susceptibility artifacts (Deichmann *et al.*, 2003). Prior to that, a shimming procedure was performed. To ensure steady-state tissue magnetization, the first 10 volumes were discarded from analysis.

Functional MRI-data preprocessing and analysis were performed using Brain Voyager QX software (Version 2.4; Brain Innovation, Maastricht, Netherlands). First, all volumes were realigned to the first volume to minimize artifacts of head movements. Further data pre-processing comprised correction for slice time errors and temporal (high-pass filter: 3 cycles per run; low-pass filter: 2.8 s; linear trend removal) as well as spatial (8 mm full-width half-maximum isotropic Gaussian kernel) smoothing. The anatomical and functional images were coregistered and transformed to normalized Talairach-space (Talairach and Tournoux, 1988) resulting in voxel sizes of 3 × 3 × 3 mm^3 on which the statistical inferences were based.

Statistical analyses were performed by multiple linear regression of the signal time course at each voxel. We modeled predictors for (i) loud and angry voices, (ii) loud and normal voices, (iii) loud and neutral voices, (iv) normal and neutral voices. For supplementary analyses, we modeled the same four predictors separately for male and female speakers in order to investigate interactions between voice gender and participants' gender. Expected blood oxygenation level dependent signal change for each predictor was modeled by a 2-gamma hemodynamic response function. On the first level, predictor estimates based on z-standardized time course data were generated for each subject using a random-effects model. On the second level, a 2 ('Group': SAD, HC) × 2 ('Emotion': angry, neutral) × 2 ('Sound Intensity': normal, loud) repeated measures analysis of variance was conducted. Statistical maps of group effects ('Group × Emotion interaction, 'Group × Sound Intensity' interaction, and three-way interaction of Group, Emotion and 'Sound Intensity') were assessed. We tested for interactions between voice gender and participant gender under consideration of sound intensity and emotional prosody by conducting a 2 ('Voice gender': male speaker, female speaker) × 2 ('Emotion': angry, neutral) × 2 ('Sound Intensity': normal, loud) × 2 (Participants' gender: male, female) repeated measures analysis of variance. Analysis focused on a priori defined regions of interest (ROIs). Search regions were defined according to WFU Pickatlas software (version 3.0.4; Maldjian *et al.*, 2003). Key structures involved in threat and prosody processing were selected, i.e. insula (volumes: left: 16 797 mm^3; right: 16 025 mm^3), OFC (volumes: left: 37 597 mm^3; right: 40 243 mm^3), superior temporal gyrus (volumes: left: 20 489 mm^3; right: 27 651 mm^3) and the amygdala (volumes: left: 1523 mm^3; right: 1498 mm^3) (Witteman *et al.*, 2012). A cluster-size threshold estimation procedure was used (Goebel *et al.*, 2006) to correct for multiple comparisons within the search regions. Significant clusters of contiguously activated voxels within the ROIs were determined by a Monte Carlo simulation based on 1000 iterations. After setting the voxel-level threshold to $P < 0.005$ (uncorrected) and specifying the FWHM of the spatial filter based on an estimate of the maps' smoothness, the simulation resulted in a minimum cluster size k of contiguously activated voxels within the search regions corresponding to a false positive rate of 5% (corrected). The cluster size threshold k was separately assessed for each contrast computed in amygdala [k_{amy}] and all other regions combined [k_{other}], respectively. Please note that k can only be computed if there is a minimum number of active voxels in a masked map.

Results

Behavioral data

For accuracy of gender classification, main effects of 'Emotion' [$F(1,38) = 23.59$, $P < 0.05$] and 'Sound Intensity' [$F(1,38) = 10.09$, $P < 0.05$] were found, indicating reduced accuracy for normal sound intensity and angry prosody (see Table 2). Furthermore, a significant interaction of 'Emotion' and 'Sound Intensity' [$F(1,38) = 8.26$, $P < 0.05$] reflected that participants accuracy was highest in detecting loud neutral voices [all $t(39) > 1.99$, all $P < 0.06$] and lowest in detecting normal angry voices [all $t(39) > 3.22$, all $P < 0.05$]. Analysis of reaction times revealed main effects of 'Emotion' [$F(1,38) = 19.93$, $P < 0.05$] and 'Sound Intensity'

Table 2. Performance data and post-scanning ratings of pleasantness and arousal for patients with SAD and HCs

Group	Auditorystimuli	Hit Rate[a] [%(SD)]	Reaction Times[a] [mean (SD)]	Pleasantness[b] [mean (SD)]	Arousal[b] [mean (SD)]
HC	loud angry	97.40% (3.19)	853.0 ms (158.6)	2.86 (1.19)	4.59 (2.33)
	loud neutral	98.90% (1.25)	820.0 ms (171.8)	5.60 (1.39)	2.36 (1.38)
	normal angry	93.13% (9.28)	959.3 ms (188.4)	3.35 (1.07)	4.30 (2.07)
	normal neutral	97.88% (3.27)	949.5 ms (262.8)	5.79 (1.41)	2.34 (1.39)
SAD	loud angry	97.88% (2,84)	821.0 ms (121.9)	3.28 (1.14)	4.18 (1.84)
	loud neutral	99.25% (1.43)	779.2 ms (122.0)	5.60 (0.88)	2.92 (1.42)
	normal angry	93.63% (7.32)	961.9 ms (150.5)	3.45 (1.05)	3.59 (1.71)
	normal neutral	98.00% (3.50)	859.4 ms (145.7)	5.32 (0.92)	2.63 (1.24)

[a]gender classification task,
[b]post-scanning rating, SD, standard deviation.

$[F(1,38) = 29.60, P < 0.05]$ and a significant interaction of 'Emotion' and 'Group' $[F(1,38) = 5.87, P < 0.05]$. The latter effect was due to faster reaction times to neutral relative to angry prosody in SAD compared with HC subjects $[t_{anger-neutral}(38) = 2.42, P < 0.05; t_{loud-normal}(38) = 0.17, P > 0.05]$ (see Table 2).

Analyses of valence and arousal ratings revealed significant main effects of 'Emotion' and 'Sound Intensity' for valence $[F(1,38) = 105.59, P < 0.05; F(1,38) = 6.85, P < 0.05]$ and arousal $[F(1,38) = 12.63, P < 0.05; F(1,38) = 17.76, P < 0.05]$, and an interaction of *'Emotion'* and *'Sound Intensity'* [valence: $F(1,38) = 9.72, P < 0.05$; arousal: $F(1,38) = 17.21, P < 0.05$]. Participants rated angry prosody as less pleasant and more arousing than neutral prosody [valence: $t(39) = -10.26, P < 0.05$; arousal: $t(39) = 5.83, P < 0.05$]. Furthermore, loud voices were rated as more unpleasant and arousing than normal voices [loud: $t(39) = -2.31, P < 0.05$; normal: $t(39) = 4.10, P < 0.05$] and, in particular, loud and angry prosody was perceived as more unpleasant and arousing than the other categories [$t(39) = -10.51, P < 0.05$] (Table 2). Finally, a significant interaction of 'Group' and 'Sound Intensity' [$F(1,38) = 12.15, P < 0.05$] was found for valence ratings, indicating that SAD patients rated loud stimuli more negatively than HCs (Table 2).This interaction was marginally significant for arousal ratings [$F(1,38) = 4.02, P = 0.052$]. The interaction effect of 'Group' and 'Emotion' was non-significant [$F(1,38) = 2.44, P = 0.13$] as was the three-way interaction term of 'Group', 'Emotion' and 'Sound Intensity' [$F(1,38) = 0.10, P = 0.76$].

fMRI data

Effects across groups. Although not of main interest in this study, the effects across groups are reported for the sake of completeness. The ANOVA revealed main effects of 'Emotion' (cluster size thresholds, $k_{amy} = 189$, $k_{other} = 324$) and 'Sound Intensity' (cluster size thresholds, $k_{amy} = 216$, $k_{other} = 513$) as well as an interaction of 'Emotion' and 'Sound Intensity' (cluster size thresholds, $k_{amy} = 81$, $k_{other} = 216$) in bilateral superior temporal regions, insula, OFC and amygdala (see Table 3). In general, participants showed higher activation to angry as compared with neutral prosody and higher activation to loud as compared with normal acoustic stimuli. Furthermore, the prosody effect was more pronounced for the loud as compared with the normal condition.

Group effects. Group × Emotion. The interaction of *Group* and *Emotion* revealed two significant clusters in the left OFC (Cluster 1: peak x, y, z: −21, 23, −12; $F(1,38) = 11.51, P < 0.05$, corrected, size 126 mm³; Cluster 2: peak x, y, z: −17, 63, −9; $F(1,38) = 15.33$, $P < 0.05$, corrected, size 171 mm³, cluster size thresholds, $k_{amy} = 81$, $k_{other} = 216$). SAD patients showed greater differences of

Table 3. Significant clusters for the main effect of 'Emotion', main effect of 'Sound Intensity' and the interaction of 'Emotion' and 'Sound Intensity' in the regions of interest (amygdala, superior temporal region, insula, OFC)

Area	La	mm³	x	y	z	F score
'main effect Emotion'						
amygdala	R	216	18	−6	−13	31.56
amygdala	L	270	−19	−9	−10	22.56
superior temporal region	R	17874	48	−30	2	81.69
superior temporal region	L	13500	−63	−21	8	57.02
insula	R	7020	39	21	−9	67.76
insula	L	5049	−36	21	−5	76.96
OFC	R	9342	39	21	−10	67.76
OFC	L	7020	−36	24	−7	78.81
'main effect Sound Intensity'						
amygdala	R	1026	27	0	−16	19.91
amygdala	L	1377	−21	−3	−19	18.07
superior temporal region	R	23382	48	−18	5	213.68
superior temporal region	L	19737	−51	−15	5	172.76
insula	R	7533	40	−20	5	246.49
insula	L	6831	−45	−13	3	133.03
OFC lateral	R	8019	43	28	−10	28.28
OFC lateral	L	10233	−32	20	−24	26.15
OFC medial	L	2700	−2	51	−15	22.80
'interaction of Emotion and Sound Intensity'						
amygdala	R	14391	18	−7	−13	15.75
amygdala	L	378	−24	0	−16	17.18
superior temporal region	R	13203	51	−24	−4	31.40
superior temporal region	L	3051	−47	−6	−10	23.73
insula	L	945	−45	−6	−5	18.64
OFC	R	5481	48	30	−7	29.87
OFC	L	2268	−45	21	−7	17.23

La, Lateralization (R, right; L, left); mm³, cluster size in mm³; x,y,z, coordinates of peak voxel in Talairach space; F score, F-statistic at the voxel of maximal activity within significant clusters, IFG: inferior frontal gyrus; OFC, orbitofrontal cortex.

activation between angry and neutral prosody as compared with the HC subjects (Figure 1).

Group × Sound Intensity. The interaction of 'Group' and 'Sound Intensity' revealed a significant cluster in the left insula (peak x, y, z: −42, −7, 2; $F(1,38) = 18.06, P < 0.05$, corrected, size 891 mm³, cluster size thresholds, $k_{amy} = 81$, $k_{other} = 108$). Individuals with SAD showed greater activation to loud acoustic stimuli in the left insula relative to HC individuals (Figure 2).

Group × Sound Intensity × Emotion. The three-way interaction of 'Group', 'Emotion' and 'Sound Intensity' was significant in the right amygdala (peak x, y, z: 21, −10, −10; $F(1,38) = 10.46, P < 0.05$, corrected, size 108 mm³, cluster size threshold, $k_{amy} = 81$,).
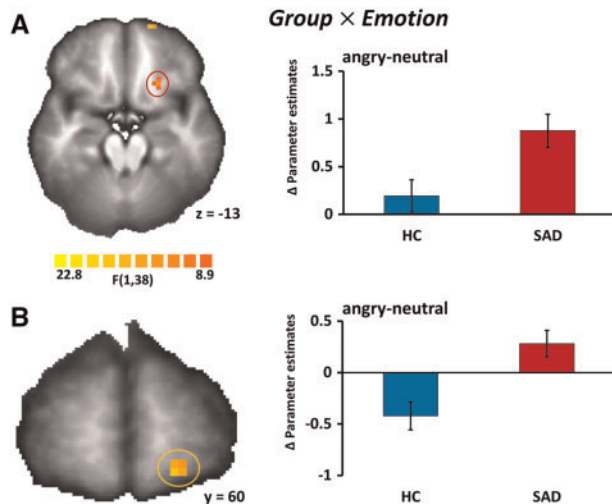
Fig. 1. Brain activation to angry and neutral prosody differs between HCs and patients with a diagnosis of SAD. (A) Statistical parametric map (left) showing a significant interaction of group (SAD vs HC) and emotion (angry prosody vs neutral prosody) in the left OFC (peak x, y, z: −21, 23, −12); $P_{corrected} < 0.05$. Bar graphs (right) of differences in parameter estimates (mean$_{angry-neutral}$ ± SEM) of the OFC cluster marked in red shown separately for HC and SAD. (B) Statistical parametric map (left) showing a significant interaction of group (SAD vs HC) and emotion (angry prosody vs neutral prosody) in a second, more anterior cluster in the left OFC (peak x, y, z: −17, 63, −9); $P_{corrected} < 0.05$. Bar graphs (right) of differences in parameter estimates (mean$_{angry-neutral}$ ± SEM) of the OFC cluster marked in orange shown separately for HC and SAD. Images are in radiological convention.
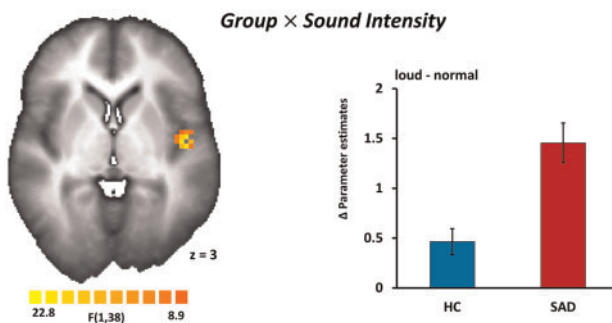

Fig. 2. Sound intensity of voice stimuli is coded differentially in the insulae of patients with SAD and HCs. A statistical parametric map (left) reveals a significant interaction of group (SAD vs HC) and sound intensity (loud voices vs normal voices) in the left insula (peak x, y, z: −42, −7, 2); $P_{corrected} < 0.05$. Bar graphs (right) of differences in parameter estimates (mean$_{loud-normal}$ ± SEM) of the insula cluster separately for HC and SAD. Image is in radiological convention.

Post hoc t-tests revealed that loud and angry prosody in comparison to loud but neutral prosody led to greater activity in SAD patients ($t = 3.18$, $P < 0.05$, corrected), while there was no significant group effect for the normal sound intensity condition. We also found a three-way interaction in the insula (peak x, y, z: −37, −3, 11; $F(1,38) = 11.26$, $P < 0.05$, corrected, size 162 mm^3, cluster size threshold, $k_{other} = 108$). However, this effect was mainly driven by differential responses during the normal sound intensity condition, that is, relatively less activation to angry vs. neutral prosody in SAD patients as compared with HC subjects (Figure 3).

Assessing covariate contributions. In order to test for differences in brain activation relative to the individual sound level, we calculated the Pearson correlation between a difference measure for the sound intensities from the loud and normal conditions and beta estimates from the insula, amygdala and OFC clusters. Even at the most liberal threshold without any control for multiple correlations, no significant correlation between sound intensity and brain activation was detectable (all r's < 0.28; all P's > 0.08). Furthermore, in order to account for differences in accuracy or reaction times between groups (see Behavioral data) we used hit rates and reaction times as covariates in the between-groups analyses of fMRI data. We re-assessed all $2 \times 2 \times 2$ repeated-measures ANOVAs as reported before including the hit rates or the reaction times from the four conditions (see Table 2) as covariates. These analyses confirmed without exception that neither accuracy measures nor reaction times had any effect on the significance of the reported effects for any of the above clusters.

Testing for interactions between voice gender and participant gender. Next, we tested if the regions reported above show an interaction of participant gender and speaker gender. At the chosen a-priori voxel-level threshold none of the clusters showed an interaction between voice gender and participant gender as indicated by $2 \times 2 \times 2 \times 2$ repeated-measures ANOVAs across HC and SAD [all $F(1,38) < 1.55$; all $P > 0.05$, corr]. Furthermore, neither the factor 'Emotion' nor the factor 'Sound Intensity' showed a significant interaction with voice gender or participant gender [all $F(1,38) < 4.31$; all $P > 0.05$, corr.].

Discussion

This study investigated differences in brain activation between patients with SAD and HC subjects in response to angry and neutral prosody, presented in either normal or loud intensity. Results show that brain activation is modulated by emotional prosody, sound intensity and group. In SAD, we found generally greater responses to loud voices in the insula, and to angry prosody in the OFC. Furthermore, we observed an interaction of sound intensity, emotion and group in insula and amygdala. In particular, the amygdala showed greater activation to angry vs neutral prosody specifically in response to loud voices in SAD as compared with HC subjects.

Reaction times revealed that participants regardless of diagnosis responded slower to emotional sounds than to neutral sounds but also slower to normal sound intensity than to loud sound intensity. These main effects imply that overall processing of angry prosody was more elaborate, possibly as a result of increased attention to threatening voices, and that processing of loud stimuli might have been facilitated by stimuli's increased salience. Furthermore, the reaction time data also show an interaction of group and prosody which results from SAD reacting slower to angry stimuli. This might be related to an interference between processing of angry prosody and the main task of gender classification and at least partially associated with OFC activation and distracted attention to threat, even though our covariance analysis does not support a significant association between behavioral and imaging data.

Brain activation during loud and angry prosody processing in SAD individuals revealed that sound intensity modulates the effects of angry prosody on amygdala activation in SAD individuals. The finding of elevated amygdala activation in patients complements previous studies reporting greater amygdala responses during threat processing in SAD using other stimuli than emotional voices (Freitas-Ferrari *et al.*, 2010; Schulz *et al.*, 2013). The amygdala has been suggested to be of essential relevance for mediation of automatic, bottom-up processing of
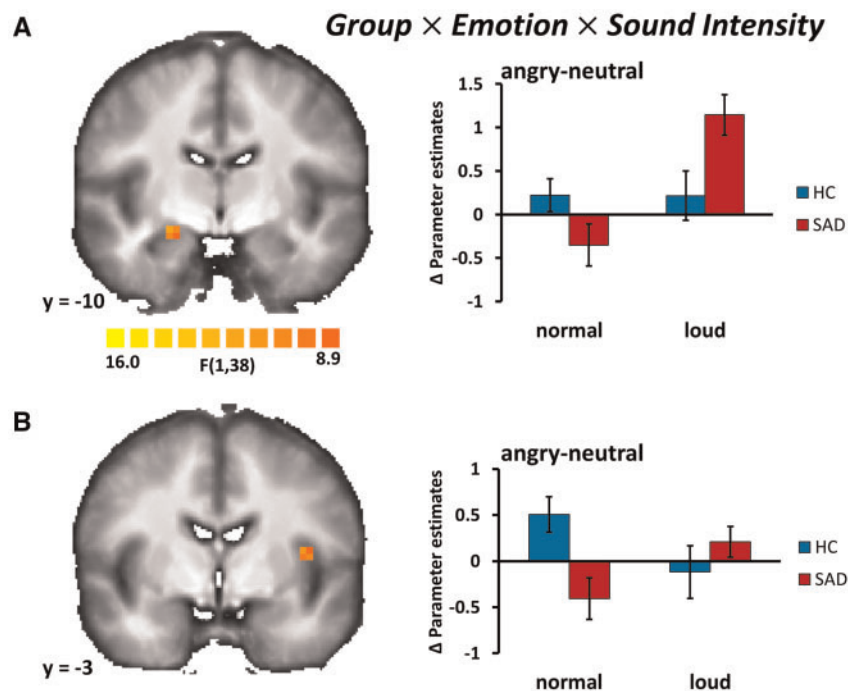
Fig. 3. In patients with SAD, the amygdala responds to loud and angry voices to a higher degree than in HCs. **(A)** Statistical parametric map (left) showing a significant three-way interaction of group (SAD *vs* HC), sound intensity (loud voices *vs* normal voices), and emotion (angry prosody *vs* neutral prosody) in the right amygdala (peak x, y, z: 21, −10, −10); $P_{corrected} < 0.05$. Bar graphs (right) of differences in parameter estimates (mean$_{angry-neutral}$ ± SEM) of the amygdala cluster shown separately for HC and SAD. **(B)** Statistical parametric map (left) showing a significant three-way interaction of group (SAD *vs* HC), sound intensity (loud voices *vs* normal voices) and emotion (angry prosody *vs* neutral prosody) in the left insula (peak x, y, z: −37, −3, 11); $P_{corrected} < 0.05$. Bar graphs (right) of differences in parameter estimates (mean$_{angry-neutral}$ ± SEM) of the insula cluster shown separately for HC and SAD. Images are in radiological convention.

emotional and particularly threatening stimuli (LeDoux, 1996; Ohman and Mineka, 2001), and this response is enhanced in anxiety disorders (Shin and Liberzon, 2010). Aside from triggering defensive autonomic responses and behavior (LeDoux, 1996, 2000; Shin and Liberzon, 2010), the amygdala guides attention to behaviorally relevant signals (LeDoux, 1996, 2000; Adolphs *et al.*, 2005; Gamer and Büchel, 2009; Shin and Liberzon, 2010) including angry prosody (Frühholz and Grandjean, 2013; Frühholz *et al.*, 2015).

In SAD, 'neural' activity to interpersonal social threat has been extensively investigated using face stimuli, and here especially pictures of angry faces (Freitas-Ferrari *et al.*, 2010; Schulz *et al.*, 2013). Although angry voices reliably signal interpersonal threat and have a high nosological relevance for individuals with SAD, prosodic stimuli have rarely been used in studies of SAD. A seminal study using prosodic stimuli (Quadflieg *et al.*, 2008) found no differences between SAD patients and HC subjects in amygdala responses to angry *vs* neutral prosody. This result suggested that the amygdala shows reliable hyper-responses only to threatening faces but not to threatening voices in SAD patients. However, using only a normalized sound intensity presented in a comfortable range (Quadflieg *et al.*, 2008) might not be an optimal protocol for investigating the neural correlates of SAD. For angry prosody, sound intensity is a crucial characteristic of this emotional expression, and especially the combination of sound intensity and angry prosody should represent a clear signal of social threat and rejection. This study shows that the sound intensity dimension is a relevant feature that differentially influences neural responses to anger prosody in SAD as compared with HC individuals especially in the amygdala. Thus, our results suggest that angry voices, similar to angry faces, may elicit greater amygdala responses in SAD.

Furthermore, we found between-group effects of sound intensity in the insula. The insula has repeatedly been shown to be involved in processing of aversive emotional cues in SAD (Straube *et al.*, 2004; Amir *et al.*, 2005; Etkin and Wager, 2007; Shah *et al.*, 2009; Boehme *et al.*, 2014). Further findings indicate that the insula plays an important role in interoception (Critchley *et al.*, 2004) and in the integration of affective arousal responses into perceived feelings induced by the current situation (Craig, 2009). Greater activation of the posterior insula in SAD identified in this study indicates that sound intensity *per se* might amplify interoceptive processing in SAD. Remarkably, there was also an interaction of sound intensity, prosody and group. However, in contrast to activation patterns in the amygdala, insular activation was found to be greater for angry *vs* neutral prosody in the normal sound intensity condition in the HC group as compared with patients. Thus, patients were rather characterized by reduced activation to angry prosody under the normal sound intensity condition, an effect which disappeared with loud stimuli, at least under the given experimental conditions. This finding might indicate successful avoidance of attentional focus on bodily responses and thus successful regulation of insula responses during moderate prosodic threat in SAD individuals.

Finally, a group by prosody interaction was detected in the OFC. This is in accordance with a previous study (Quadflieg *et al.*, 2008) that also detected greater activity during presentation of angry prosody in comparison to neutral prosody in SAD. Furthermore, the OFC has been shown to play a major role in the comprehension of emotional prosody (Sander *et al.*, 2005; Wildgruber *et al.*, 2005). Models of emotional prosodic processing suggest that the OFC accomplishes higher cognitive processes during the comprehension of emotional prosody, such as

the evaluation of the incoming emotional stimulus (Wildgruber *et al.*, 2009; Kotz and Paulmann, 2011).

With regard to these findings it should be noted, however, that the task used in this study involves implicit processing of angry and neutral voice prosody but does not include a condition of explicit processing of voice prosody. It should be noted that in healthy subjects different networks are recruited during implicit and explicit processing of voice prosody (Frühholz *et al.*, 2012). Most notably, with regard to our regions of interest, subregions activated to angry relative to neutral prosody within STG during explicit processing of voice prosody have been shown to recruit different subregions during implicit processing of the same voices in the context of a gender discrimination task (Grandjean *et al.*, 2005; Sander *et al.*, 2005). Furthermore, activation of bilateral amygdala is only found during implicit processing but not during explicit processing (Frühholz *et al.*, 2012), whereas activation of OFC is usually only reported during explicit processing but not during implicit processing.

We would like to mention some limitations of this study. First, as we have no explicit rating data on the perceived loudness of the stimuli, we are not able to present any data that would allow for inferences on group differences in loudness sensitivity. However, individual sound intensity level calibrations did not differ between groups on average and this might be taken as an indication that neither group under investigation has been more sensitive to loudness than the other. Second, we only investigated two different sound intensities. To understand processing of emotional prosody in more detail, it would be helpful to include further gradations of sound intensity. Since we only investigated two sound levels, it remains open whether similar differences between patients and controls will be found when sounds become even more threatening. Furthermore, we only investigated responses to angry *vs* neutral prosody. Future studies might also include other emotional expressions to investigate the emotional specificity of the findings.

In conclusion, the present event-related fMRI study demonstrated that different brain regions are recruited in SAD patients and HC subjects if emotional prosody and sound intensity interact. We found generally greater responses in the insula to loud voices and in the OFC to angry prosody in individuals with SAD. Furthermore, we found an interaction of sound intensity, emotion and group in insula and amygdala. The most important finding was an interaction of sound intensity, prosody, and diagnosis in amygdala activation, reflecting greater responses specifically to loud angry voices in SAD as compared with HC individuals. This finding indicates that sound intensity represents a highly relevant feature to activate a central hub of the neural fear network in SAD. The results of this study show that processing of angry faces and voices might engage similar neural networks if sound intensity, in addition to prosody, is considered as a crucial characteristic of voices.

Funding

Conflict of interest. None declared.

References

Adolphs, R., Gosselin, F., Buchanan, T.W., Tranel, D., Schyns, P., Damasio, A.R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, **433**(7021), 68–72.

Amir, N., Klumpp, H., Elias, J., Bedwell, J.S., Yanasak, N., Miller, L.S. (2005). Increased activation of the anterior cingulate cortex during processing of disgust faces in individuals with social phobia. *Biological Psychiatry*, **57**(9), 975–81.

APA (2000). *Diagnostic and Statistical Manual of Mental Disorders.* Washington, DC: American Psychological Association.

Boehme, S., Ritter, V., Tefikow, S., *et al.* (2014). Brain activation during anticipatory anxiety in social anxiety disorder. *Social Cognitive and Affective Neuroscience*, **9**(9), 1413–8.

Craig, A.D. (2009). How do you feel - now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, **10**(1), 59–70.

Critchley, H.D., Wiens, S., Rotshtein, P., Ohman, A., Dolan, R.J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, **7**(2), 189–95.

Deichmann, R., Gottfried, J.A., Hutton, C., Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage*, **19(2 Pt 1)**, 430–41.

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

Etkin, A., Wager, T.D. (2007). Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *The American Journal of Psychiatry*, **164**(10), 1476–88.

Freitas-Ferrari, M.C., Hallak, J.E., Trzesniak, C., *et al.* (2010). Neuroimaging in social anxiety disorder: a systematic review of the literature. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, **34**(4), 565–80.

Frühholz, S., Ceravolo, L., Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, **22**(5), 1107–17.

Frühholz, S., Grandjean, D. (2013). Amygdala subregions differentially respond and rapidly adapt to threatening voices. *Cortex*, **49**(5), 1394–403.

Frühholz, S., Hofstetter, C., Cristinzio, C., *et al.* (2015). Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(5), 1583–8.

Fydrich, T., Renneberg, B., Schmitz, B., Wittchen, H.U. (1997). *Strukturiertes Klinisches Interview Für DSM-IV, Achse II (Persönlichkeitsstörungen).* Göttingen: Hogrefe.

Gamer, M., Büchel, C. (2009). Amygdala activation predicts gaze toward fearful eyes. *Journal of Neuroscience*, **29**(28), 9123–6.

Goebel, R., Esposito, F., Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, **27**(5), 392–401.

Grandjean, D., Sander, D., Pourtois, G., *et al.* (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, **8**(2), 145–6.

Hautzinger, M., Bailer, M., Worall, H., Keller, F. (1995). *Beck-Depressions-Inventar (BDI).* Testhandbuch. Bern: Hans Huber.

Kotz, S.A., Paulmann, S. (2011). Emotion, Language, and the Brain. *Language and Linguistics Compass*, **5**(3), 108–25.

LeDoux, J.E. (1996). *The Emotional Brain.* New York: Simon & Schuster.

LeDoux, J.E. (2000). Emotion circuits in the brain. Annual Reviews in. *Neuroscience*, **23**, 155–84.

Lipka, J., Miltner, W.H., Straube, T. (2011). Vigilance for threat interacts with amygdala responses to subliminal threat cues in specific phobia. *Biological Psychiatry*, **70**(5), 472–8.

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.

Mothes-Lasch, M., Mentzel, H.J., Miltner, W.H., Straube, T. (2011). Visual attention modulates brain activation to angry voices. *Journal of Neuroscience*, **31**(26), 9594–8.

Ohman, A., Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, **108**(3), 483–522.

Pittam, J., Scherer, K. (1993). Vocal Expression and Communication of Emotion. In Lewis, J., Haviland, J. M., editors. *Handbook of Emotions*. New York: Guilford Press.

Quadflieg, S., Mohr, A., Mentzel, H.J., Miltner, W.H.R., Straube, T. (2008). Modulation of the neural network involved in the processing of anger prosody: The role of task-relevance and social phobia. *Biological Psychology*, **78**(2), 129–37.

Rapee, R.M., Heimberg, R.G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, **35**(8), 741–56.

Sander, D., Grandjean, D., Pourtois, G., *et al.* (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. **28**(4), 848–58.

Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., Friederici, A.D. (2008). When vocal processing gets emotional: On the role of social orientation in relevance detection by the human amygdala. *Neuroimage*, **40**(3), 1402–10.

Schirmer, A., Simpson, E., Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research*, **1176**, 103–12.

Schulz, C., Mothes-Lasch, M., Straube, T. (2013). Automatic neural processing of disorder-related stimuli in social anxiety disorder: faces and more. *Frontiers in Psychology*, **4**, 282.

Shah, S.G., Klumpp, H., Angstadt, M., Nathan, P.J., Phan, K.L. (2009). Amygdala and insula response to emotional images in patients with generalized social anxiety disorder. *Journal of Psychiatry & Neuroscience*, **34**(4), 296–302.

Shin, L.M., Liberzon, I. (2010). The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology*, **35**(1), 169–91.

Stangier, U., Heidenreich, T. (2005). *Liebowitz Soziale Angst Skala [Liebowitz social anxiety scale]*. In: Collegium Internationale Psychiatriae Scalarum, editors. Hogrefe: Internationale Skalen für Psychiatrie Göttingen.

Straube, T., Kolassa, I.T., Glauer, M., Mentzel, H.J., Miltner, W.H. (2004). Effect of task conditions on brain responses to threatening faces in social phobics: an event-related functional magnetic resonance imaging study. *Biological Psychiatry*, **56**(12), 921–30.

Straube, T., Mothes-Lasch, M., Miltner, W.H. (2011). Neural mechanisms of the automatic processing of emotional information from faces and voices. *British Journal of Psychology*, **102**(4), 830–48.

Talairach, J., Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. Stuttgart: Thieme.

Tamietto, M., de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, **11**(10), 697–709.

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

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

Wittchen, H. U., Wunderlich, U., Gruschwitz, S., Zaudig, M., Eds. (1996). *Strukturiertes Klinisches Interview Für DSM-IV (SKID)*. Göttingen (Germany): Beltz-Test.

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.