

Tuned to voices and faces: Cerebral responses linked to social anxiety

Benjamin Kreifelts^{a,*}, Kathrin N. Eckstein^a, Thomas Ethofer^{a,b}, Ariane Wiegand^a, Sarah Wächter^a, Carolin Brück^a, Michael Erb^b, Martin Lotze^c, Dirk Wildgruber^a

^a Department of Psychiatry and Psychotherapy, University of Tübingen, Germany

^b Department for Biomedical Magnetic Resonance, University of Tübingen, Germany

^c Functional Imaging Group, Department for Diagnostic Radiology and Neuroradiology, University of Greifswald, Germany



ARTICLE INFO

Keywords:

Amygdala
Temporal voice area
Fusiform face area
Superior temporal sulcus
Threat
Social anxiety

ABSTRACT

Voices and faces are the most common sources of threat in social anxiety (SA) where the fear of negative evaluation and social exclusion is the central element. SA itself is spectrally distributed among the general population and its clinical manifestation, termed social anxiety disorder, is one of the most common anxiety disorders. While heightened cerebral responses to angry or contemptuous facial or vocal expressions are well documented, it remains unclear if the brain of socially anxious individuals is generally more sensitive to voices and faces. Using functional magnetic resonance imaging, we investigated how SA affects the cerebral processing of voices and faces as compared to various other stimulus types in a study population with greatly varying SA ($N = 50$, 26 female). While cerebral voice-sensitivity correlated positively with SA in the left temporal voice area (TVA) and the left amygdala, an association of face-sensitivity and SA was observed in the right fusiform face area (FFA) and the face processing area of the right posterior superior temporal sulcus (pSTSFA). These results demonstrate that the increase of cerebral responses associated with social anxiety is not limited to facial or vocal expressions of social threat but that the respective sensory and emotion processing structures are also generally tuned to voices and faces.

1. Introduction

An angry tone of voice or a contemptuous face are typical threat signals dreaded by a socially anxious individual in face-to-face communication. They relate to central fears in social anxiety (SA) pertaining to humiliation, exclusion and negative social evaluation. SA can be characterized as ancient maladaptive mechanism which evolved parallel to a social hierarchy based on nonverbal and later verbal signals rather than on physical violence (Öhman, 1986; Trower and Gilbert, 1989). Accordingly, increased sensitivity to socially threatening nonverbal signals reflected in cognitive biases towards these signals is one of the key features in SA (Morrison and Heimberg, 2013). For facial expressions it is well documented that the cerebral sensitivity to social threat in SA is reflected most prominently in increased responses of the amygdala (Brühl et al., 2014; Etkin and Wager, 2007; Miskovic and Schmidt, 2012) but also in other parts of the cerebral emotion and salience processing systems (i.e., insula, mediofrontal and orbitofrontal cortex; Brühl et al., 2014; Etkin and Wager, 2007; Miskovic and Schmidt, 2012). Moreover, brain areas related to attention and emotion regulation (i.e., dorsolateral

prefrontal cortex, parietal cortex Brühl et al., 2014; Miskovic and Schmidt, 2012) and, finally, also regions implicated in the general processing of faces (i.e., occipital and occipito-temporal cortex; Brühl et al., 2014; Miskovic and Schmidt, 2012) exhibit increased responses to face cues.

For socially threatening vocal expressions the limited number of comparable studies suggests a similar pattern with altered responses to social threat in the orbitofrontal cortex (Quadflieg et al., 2008; Simon et al., 2017), amygdala (Simon et al., 2017) and dorsolateral as well as dorsomedial prefrontal cortex (Kreifelts et al., 2014).

However, it remains an open question if SA generally modulates the cerebral sensitivity to voices and faces irrespective of a socially threatening emotion, attitude or intention expressed in these signals. It is conceivable that not only social threat per se but also its evolutionary common carrier signals (i.e., voices and faces) are preferentially processed in SA. Primary candidates for such SA-related sensitivity to voices and faces would be the specialized sensory brain areas underpinning the general processing of human voices and faces (i.e., the TVA (Belin et al., 2000; Pernet et al., 2015), the FFA (Kanwisher et al., 1997; Kanwisher

* Corresponding author. Department of Psychiatry and Psychotherapy, University of Tübingen, Calwerstrasse 14, 72076, Tübingen, Germany, E-mail address: benjamin.kreifelts@med.uni-tuebingen.de (B. Kreifelts).

and Yovel, 2006), the pSTSFA (Bernstein and Yovel, 2015; Haxby et al., 2000) and the occipital face area (OFA; Bernstein and Yovel, 2015; Puce et al., 1996)) but also the amygdala which exhibits combined preferential responses to voices and faces (Mende-Siedlecki et al., 2013; Pernet et al., 2015).

Our study aimed to elucidate in a cohort of 50 individuals if and how SA is reflected in the neural activation of canonical voice and face perception structures. Convergent with the spectral distribution of SA severity in the population, the study cohort included a spectrum of individuals ranging from very low SA to severe social anxiety disorder (SAD), which represents a very common psychiatric disorder (Stein et al., 2010). The individual voice- and face-sensitivity were determined through functional magnetic resonance imaging (fMRI) experiments where the cerebral responses to voice and face cues were compared to various other types of acoustic and visual stimuli.

A linear relationship between SA and cerebral voice- and face-sensitivity was hypothesized within the following brain areas: 1) in the sensory voice processing areas (i.e. TVA) for voice-sensitivity and face processing areas (i.e., FFA, pSTSFA and OFA) for face-sensitivity 2) in the amygdala as central salience processing structure (Menon, 2015) which unites voice- and face-sensitivity.

2. Material and methods

2.1. Participants

50 participants (24 male, mean age 25.1 years, SD 3.7 years, range: 19–39 years) were recruited at the Universities of Tübingen (study site 1) and Greifswald (study site 2) through public announcements inviting socially anxious and very outgoing individuals. All participants underwent the Structured Clinical Interview for DSM-IV (SCID; Wittchen et al., 1997). SA severity was assessed using the Liebowitz Social Anxiety Scale (LSAS; self-report, Stangier and Heidenreich, 2003). Eighteen participants fulfilled the DSM-IV criteria of SAD while the other participants showed varying levels of SA below the clinical threshold. Three participants (two of them with SAD) had a mild depression, one (with SAD) suffered from generalized anxiety disorder and six suffered from specific phobia (three of them with SAD). The remaining participants did not suffer from any psychiatric disorder. Seven of the participants (five of them with SAD) had a history of major depression and all of these had been in remission for more than six months before study inclusion. None of the participants had a history of substance abuse, or neurological illness, or was taking any regular medication. General anxiety was evaluated applying the State-Trait-Anxiety-Inventory (STAI; Laux et al., 1981) which differentiates between state anxiety as a state of anxiety characterized by e.g., feelings of tension, apprehension, nervousness, and worry at the present moment (subscale STAI X1) and trait anxiety as individual anxiety proneness characterized as the tendency to experience stressful situations as threatening or dangerous and to react to such situations with increased state anxiety (subscale STAI X2). The “Mehrfachwahl-Wortschatz-Intelligenz-Test” (MWT-B; Lehrl, 2005) was applied to determine verbal intelligence. All participants reported normal hearing and normal or corrected to normal visual acuity and were right-handed as assessed with the Edinburgh Inventory (Oldfield, 1971).

The psychometric and socio-demographic data of the study population are given in Table 1. The study was performed according to the Code of Ethics of the World Medical Association (Declaration of Helsinki) and the protocol of human investigation was approved by the local ethics committees where the study was performed. All individuals gave their written informed consent prior to their participation in the study. All participants received a monetary expense allowance.

2.2. Stimuli and experimental design

Two fMRI experiments were performed to localize face-sensitive (Kanwisher et al., 1997) and voice-sensitive (Belin et al., 2000) brain

Table 1
Socio-demographic and psychometric data.

	mean (SD)	study site 1 (n = 28)	study site 2 (n = 22)
age (years)	25.1 (3.7)	24.8 (4.0)	25.4 (3.5)
gender	24m/26f	14m/14f	10m/12f
MWT-B score	31.1 (2.7)	31.1 (2.3)	31.0 (3.2)
LSAS	38.1 (31.5)	42.6 (28.1)	32.4 (35.3)
SAD	18/50	13/28	5/22
STAI state (X1)	37.7 (8.5)	39.0 (9.0)	36.0 (7.7)
STAI trait (X2)	42.0 (9.3)	39.9 (12.0)	44.6 (2.3)

MWT-B = “Mehrfachwahl-Wortschatz-Intelligenz-Test”, a short test of verbal intelligence; LSAS = Liebowitz Social Anxiety Scale; STAI = State Trait Anxiety Inventory; study site 1 = Tübingen, study site 2 = Greifswald.

areas. For reasons of comparability to the existing literature on general face and voice processing, the experimental designs were closely adapted from seminal studies in this area (Belin et al., 2000; Epstein et al., 1999; Kanwisher et al., 1997).

The face sensitivity experiment included pictures from four different categories (faces, houses, objects and natural scenes) using a block-design. All stimuli were black-and-white pictures and unknown to the participants. The face stimuli showed neutral to slightly friendly expressions. The house stimuli depicted different types of buildings (e.g., brick, wooden, concrete). The object stimuli comprised household objects and items of clothing, while the natural scenes included different types of panoramas (e.g., mountains, coast, river). Each block and category contained 20 stimuli. Within blocks the stimuli were presented in random order for 300 ms interleaved with 500 ms of fixation (1 block = 20 stimuli × (300 ms picture + 500 ms fixation) = 16 s). Eight blocks of each category pseudorandomized within the experiment were shown separated by short ~1 s rest periods. To ascertain constant attention, a one-back task was employed in which the participants had to press a button on a fiber optic system (LumiTouch, Photon Control, Burnaby, Canada) with their right index finger when a picture was directly repeated. Positions of repeated stimuli were randomized within blocks.

The voice sensitivity experiment consisted of a passive-listening block design experiment with 24 stimulation blocks and 12 silent periods (each 8 s). Participants were instructed to listen with their eyes closed. The stimuli included 12 blocks of human vocal sounds (e.g., onomatopoeia, speech [e.g., finish language, single syllables], sighs), 6 blocks with animal sounds (e.g., various cries, gallops) and 6 blocks with environmental sounds (e.g., telephones, cars, planes). Stimuli were normalized with respect to mean acoustic energy. The blocks were separated by 2 s of silence. Sound and silence blocks were randomized across the experiment.

2.3. Image acquisition

MRI was performed using a PRISMA (Tübingen) or VERIO 3T (Greifswald) scanner (Siemens, Erlangen, Germany). Structural T1-weighted images (PRISMA/VERIO): 176 slices, TR = 2300/(1900) ms, TE = 2.96/(2.52) ms, voxel size: 1 × 1 × 1 mm³) and functional images (30/(34) axial slices acquired in sequential descending order, 3 mm thickness + 1 mm gap, TR = 1.7/(2.0) s, TE = 30 ms, voxel size: 3 × 3 × 4 mm³, field of view 192 × 192 mm², 64 × 64 matrix, flip angle 90°) were acquired. The time series consisted of 336/(303) images for the face experiment and 231/(195) images for the voice experiment. A field map (36/(34) slices, slice thickness 3 mm, TR = 400/(488) ms, TE(1) = 5.19/(4.92) ms, TE(2) = 7.65/(7.38) ms) was acquired.

2.4. Analysis of sample data

Potential associations between study site/MRI scanner, age and gender which were used as covariates of no interest in the MRI analysis and SA severity (i.e. LSAS scores), general state (i.e., STAI X1) and trait (i.e., STAI X2) anxiety and verbal intelligence (i.e., MWT-B) were systematically investigated applying bivariate correlation analyses.

2.5. Analysis of fMRI data

Data were analyzed with statistical parametric mapping software (SPM8, Wellcome Department of Imaging Neuroscience, London, UK, <http://www.fil.ion.ucl.ac.uk/spm/>). Preprocessing comprised the removal of the first five EPI images from each run, realignment, unwarping based on a static field map, normalization into MNI space (Montreal Neurological Institute; Collins et al., 1994; resampled voxel size: $3 \times 3 \times 3 \text{ mm}^3$) and spatial smoothing using a Gaussian filter with 8-mm full width half maximum. For the voice experiment, three regressors were defined (vocal sounds [V], animal sounds [A], and environmental sounds [E]) using a box car function convolved with the hemodynamic response function (HRF) corresponding to the duration of the blocks of stimuli. Four regressors (faces [F], houses [H], objects [O], and scenes [S]) were defined for the face experiment. Low frequency components were removed through a high-pass filter with a cutoff frequency of 1/128 Hz. The error term was modeled as a first order autoregressive process to account for serial autocorrelations (Friston et al., 2002). The motion parameters (i.e., translation and rotation on the x-, y- and z-axes) were included in the models at single subject level as covariates. Voice-sensitivity was defined by the contrast vocal > non-vocal sounds (i.e., $V > (A, E)$); contrast weights [1 -1/2 -1/2]) and face-sensitivity by the contrast faces > non-face stimuli (i.e., $F > (H, O, S)$); contrast weights [1 -1/3 -1/3 -1/3]). The individual contrast images were calculated and statistically evaluated at the group level in a random-effects analysis using one-sample t-tests to define the FFA, the OFA, the pSTSFA and the TVA as functional regions of interest (ROI). Statistical significance was assessed at $p < 0.05$ with FWE correction at voxel level for multiple comparisons across the whole brain. When functional ROIs (i.e., FFA, OFA, pSTSFA) were not separable at the voxel-wise threshold of $p < 0.05$, FWE-corrected, the FFA was defined as comprising all supra-threshold voxels within the fusiform gyrus, the OFA as comprising all supra-threshold voxels in the inferior occipital gyrus and the pSTSFA as comprising all supra-threshold voxels in the gyri adjacent to the pSTS (i.e., middle and superior temporal gyri and the angular gyrus) at a statistical threshold of $p < 0.05$ with FWE correction at voxel level for multiple comparisons across the whole brain. For the TVA the temporal gyri and the temporal pole were defined as a priori anatomical ROI. Only voice-sensitive voxels within this ROI were included in the TVA. The amygdala served as anatomical ROI and was defined with the Automated Anatomic Labeling (AAL) toolbox (Tzourio-Mazoyer et al., 2002).

The relationships between SA severity (LSAS scores) and cerebral voice- and face-sensitivity were investigated using linear regression analyses. Age, gender and the study site/MRI scanner were included as covariates in all group analyses. Hypothesis-based ROI analyses focused on TVA, FFA, OFA, pSTSFA and amygdala ($p < 0.05$ with FWE correction for multiple comparisons across the respective ROI at voxel level) complemented by an explorative whole-brain analysis ($p < 0.05$ with FWE correction for multiple comparisons at voxel level). Mean contrast estimates were extracted from significant clusters for the purpose of visualization, post-hoc characterization and validation of observed effects. For the post-hoc decomposition of voice- and face-sensitivity effects, a positive correlation between SA and the response to voices, or respectively faces, was expected and tested one-tailed while no predictions were made on the relationships between SA and the responses to the other stimulus classes and two-tailed testing was applied. The validation analyses included the replication of the linear regression analyses to ensure that SA-associations were based on voice- or face-preferential modulations of cerebral responses as defined by minimum difference contrasts (i.e., $V - \max(A, E)$ and $F - \max(H, O, S)$) and if associations remained significant when general anxiety estimates (STAI) included as covariates in partial correlations. The minimum difference contrasts were calculated at the individual level as the minimum of the contrast images $V - A$ and $V - E$ for voice-sensitivity and $F - H$, $F - O$ and $F - S$ for face-sensitivity. Finally, it was tested if the association between SA and voice-

sensitivity and between SA and face-sensitivity differed in significant clusters using the matrix approach described by A. Paul Beaulne ([http://www.spsstools.net/Syntax/Regression RepeatedMeasure/CompareRegressionCoefficients.txt](http://www.spsstools.net/Syntax/Regression%20RepeatedMeasure/CompareRegressionCoefficients.txt); for details see the Supplementary material). For the validation analyses, the direction of association between SA and cerebral responses remain the same as in the initial analysis and one-tailed testing was applied.

As the TVA includes distinct peaks of voice-sensitivity assumed to exhibit distinct functional profiles, significant effects in the TVA were referenced to the TVA voice-sensitivity peaks in the present study as well as to the maxima of voice-sensitivity clusters observed in the seminal study by Belin et al., (2000) and the large scale study by Pernet et al., (2015) based on the Euclidian distance.

2.6. Data availability

Group level statistical maps and SPM code underlying the findings of the study are available upon request from the corresponding author.

3. Results

3.1. Participant sample data

50 individuals with greatly varying SA, 18 of them with a diagnosis of SAD took part in the study. Details of the sample characteristics can be found in Table 1 including the characteristics of the subsamples included at the two study sites. The covariates gender and study site/MRI scanner were not substantially correlated with SA severity (both $p > 0.05$, both $\text{abs}(r) \leq 0.21$), while age was negatively correlated with SA severity ($r = 0.37$, $p = 0.008$). SA severity positively correlated with general state ($r = 0.61$, $p < 0.001$) and trait anxiety ($r = 0.46$, $p = 0.001$). The covariates study site/MRI scanner, age and gender were not associated with general state or trait anxiety or verbal intelligence (all $p > 0.05$, all $\text{abs}(r) \leq 0.25$). Also verbal intelligence was not significantly correlated with any of the remaining sample variables (all $p > 0.05$, all $\text{abs}(r) \leq 0.27$).

3.2. fMRI analysis

3.2.1. Voice- and face-sensitive areas

All expected voice- and face-sensitive areas could be localized with the exception of the left OFA (see Table S1). In the left inferior occipital cortex no suprathreshold face-sensitive voxels could be determined. In the right hemisphere, the pSTSFA and the OFA were encompassed in one large cluster at the voxel-wise threshold of $p < 0.05$, FWE-corrected. Thus, they were disambiguated combining the voxel-wise threshold of $p < 0.05$, FWE-corrected with appropriate anatomical ROIs for the pSTSFA and the OFA (see Material and methods).

3.2.2. Association of SA and cerebral voice- and face-sensitivity

3.2.2.1. Voice-sensitivity. Significant positive linear relationships between SA severity and voice-sensitivity were found in the left TVA (Fig. 1 A, Fig. S1 A; peak coordinate: 66 -33 0, $t = 3.8$, $p_{\text{FWEcorr}} = 0.041$) and the bilateral amygdalae (Fig. 1 B; left: peak coordinate: 18 -6 -15, $t = 2.9$, $p_{\text{FWEcorr}} = 0.044$; right: peak coordinate: 21 0 -18, $t = 2.9$, $p_{\text{FWEcorr}} = 0.049$).

In the left TVA, the correlation between voice-sensitivity and SA was driven by increased responses to voices in individuals with greater SA ($r = 0.48$, $p = 0.0002$, one-tailed) while the correlation between SA and responses to animal sounds was considerably weaker ($r = 0.31$, $p = 0.03$) and non-significant between SA and environmental sounds ($r = 0.18$, $p > 0.05$).

The peak of the SA and voice-sensitivity correlation was located between the middle and posterior absolute voice-sensitivity maxima in the

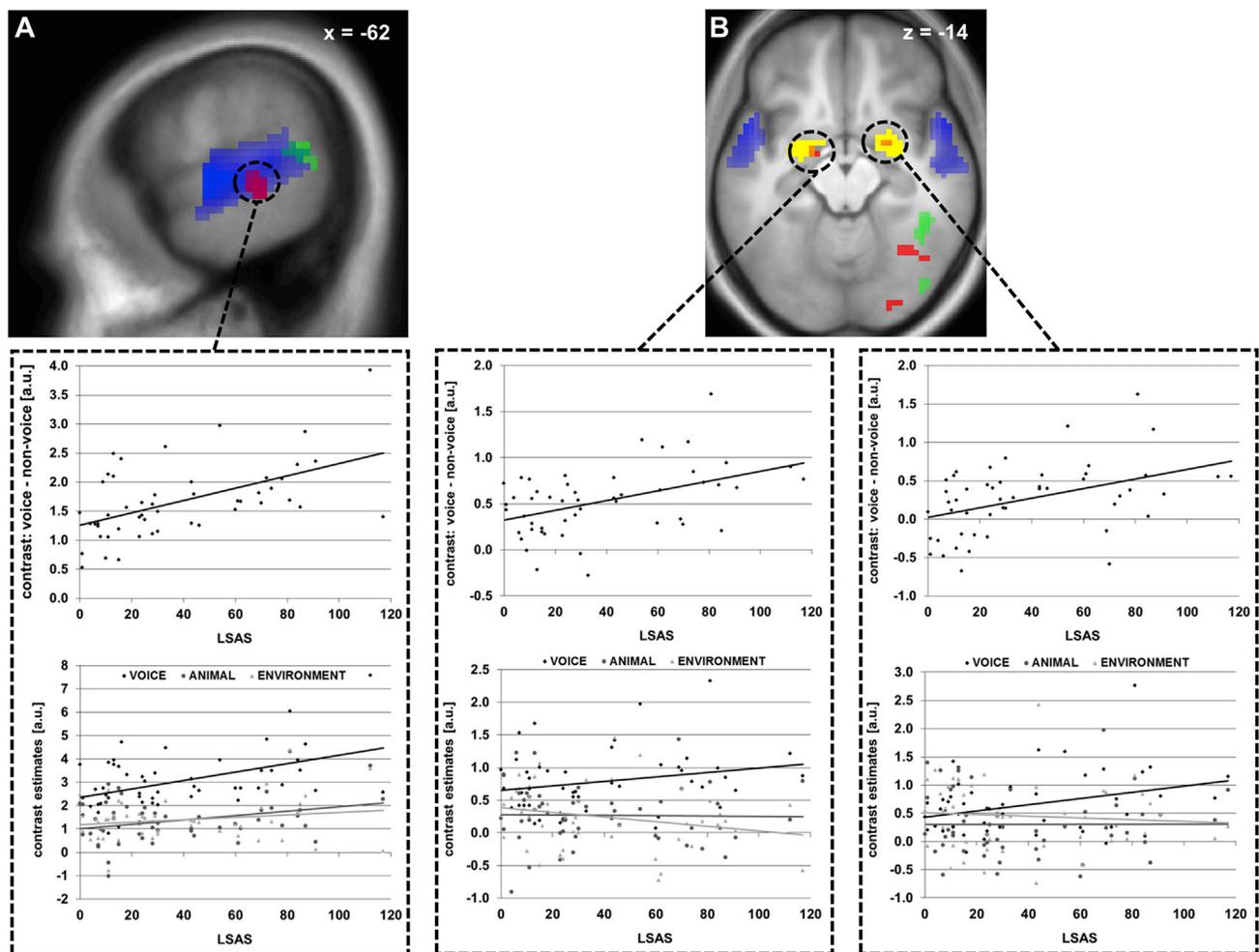


Fig. 1. Linear associations between SA severity and cerebral voice-sensitivity.

Correlations of LSAS scores with voice-sensitivity (red) rendered onto sagittal (A) and transversal slices of the study population mean anatomical scan (B). Functional and anatomical ROIs are superimposed in different colors (voice-sensitive areas (TVA) = blue; face-sensitive (FFA, pSTSFA and OFA) = green; amygdala = yellow). Results are shown at a threshold of $p < 0.005$ ($t(46) \geq 2.7$), uncorrected, at voxel-level for visualization purposes only. Areas with significant associations of SA and voice-sensitivity ($P_{FWE} < 0.05$ at voxel level corrected for multiple comparisons across the a priori ROIs) are circled in broken lines. The x- and z-coordinates refer to the MNI system. The top row of diagrams illustrates the relationship between the SA severity and voice-sensitivity in the statistically significant areas while the bottom row of diagrams details the association of SA severity and cerebral responses to voices as compared to the other types of stimuli.

left TVA with the minimal distance to the posterior voice-sensitivity maxima of the present study as well as previous reference studies (see Table S2).

In the left amygdala, a borderline-significant positive relationship between SA and responses to voices ($r = 0.23$, $p = 0.05$, one-tailed) was observed while this relationship was significant in the right amygdala ($r = 0.33$, $p = 0.01$, one-tailed). This was contrasted by non-significant mostly negative relationships between SA and responses to the other types of sounds (all $\text{abs}(r) \leq 0.24$, all $p > 0.05$). Within the left TVA and the bilateral amygdalae, the validation analyses demonstrated that SA was correlated with voice-preferential responses also using the minimum difference contrast ($V - \max(A,E)$); all $r \geq 0.41$, all $p \leq 0.002$, one-tailed) and that the observed effects could not be solely attributed to general anxiety (all partial $r \geq 0.28$, all $p \leq 0.02$, one-tailed). In the left TVA the association between SA and voice-sensitive (i.e., $V > (A, E)$) as well as voice-preferential (i.e., $V - \max(A,E)$) responses was stronger than between SA and face-sensitive as well as face-preferential responses, respectively (both $t \geq 2.7$, both $p \leq 0.004$, one-tailed). In the amygdalae, the results were comparable for the comparisons of voice- and face-preferential responses (both $t \geq 2.4$ both $p \leq 0.02$) but non-significant for the comparisons of

voice- and face-sensitive responses (both $t \leq 1.9$ both $p \geq 0.06$).

3.2.2.2. Face-sensitivity. Face-sensitivity correlated with SA severity in the right FFA (Fig. 2, Fig. S1 B; peak coordinate: 45–54 -18, $t = 3.3$, $p_{FWE\text{corr}} = 0.012$) and the right pSTSFA (Fig. 2, Fig. S1 B; peak coordinate: 42–57 12, $t = 3.7$, $p_{FWE\text{corr}} = 0.033$).

The correlation between SA and face-sensitivity in both areas was based on increased responses to faces in individuals with greater SA (right FFA: $r = 0.36$, $p = 0.005$; right pSTSFA: $r = 0.36$, $p = 0.005$, both one-tailed) whereas the correlations between SA and the responses to all other classes of stimuli were non-significant (all $\text{abs}(r) \leq 0.18$, all $p > 0.05$).

Again, the validation analyses showed for both areas a correlation between SA and face-preferential responses using the minimum difference contrast (both $r \geq 0.38$, both $p \leq 0.003$, one-tailed) and the stability of the effects when accounting for general anxiety (all partial $r \geq 0.32$, all $p \leq 0.01$, one-tailed). The relationship between SA and face-sensitive as well as face-preferential responses was stronger than between SA and voice-sensitive as well as voice-preferential responses, respectively, in both areas (all $t \geq 1.7$, all $p \leq 0.047$, one-tailed).

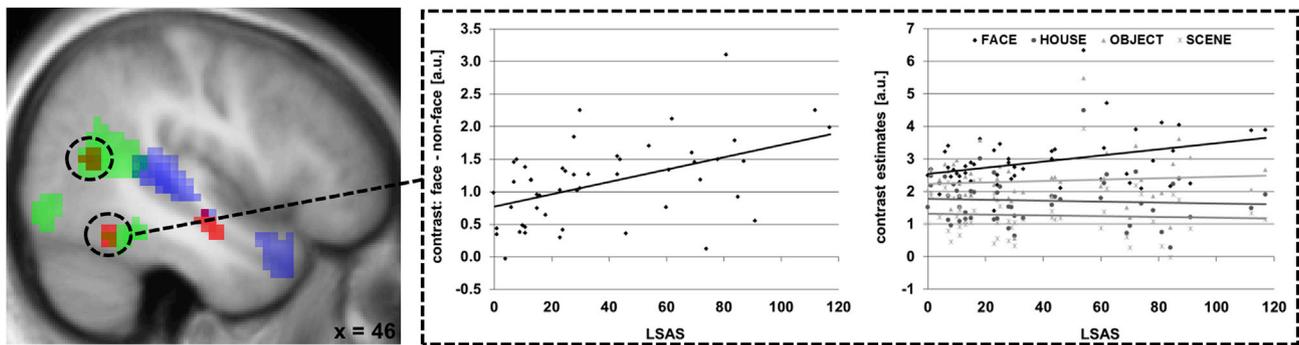


Fig. 2. Linear associations between SA severity and cerebral face-sensitivity.

Correlations of LSAS scores with face-sensitivity (red) rendered onto a sagittal slice of the study population mean anatomical scan. Functional ROIs are superimposed in different colors (voice-sensitive areas (TVA) = blue; face-sensitive (FFA, pSTSFA and OFA) = green). Results are shown at a threshold of $p < 0.005$ ($t(46) \geq 2.7$), uncorrected, at voxel-level for visualization purposes only. Areas with significant associations of SA and face-sensitivity ($p_{FWE} < 0.05$ at voxel level corrected for multiple comparisons across the a priori ROIs) are circled in broken lines. The x-coordinate refers to the MNI system. The diagrams illustrate the relationship between the SA severity and face-sensitivity in the right FFA (left) and the association of SA severity and cerebral responses to faces as compared to the other types of stimuli (right).

In the other ROIs and outside the ROIs no significant associations between SA severity and voice- or face-sensitivity were observed.

4. Discussion

To our knowledge, the present study is the first to directly link the neuroscience of SA with that of general voice- and face processing. Our data demonstrate that individual SA severity correlates with the cerebral sensitivity to voices and faces in central nodes of the voice (Belin et al., 2000; Pernet et al., 2015), face (Bernstein and Yovel, 2015; Haxby et al., 2000; Kanwisher and Yovel, 2006) and salience (Menon, 2015) processing systems and that this effect even occurs during presentation of neutral or slightly positive social cues and cannot be explained by concomitantly increased general anxiety. In other words, cerebral responses in socially anxious subjects appear to be tuned to voices and faces per se.

This effect might be explained by the fact, that social threat is most often communicated by facial and vocal cues. Also, the specific location of the SA-dependent cerebral responses are in line with the assumption that not only the sensory systems underlying voice (i.e., TVA) and face perception (i.e., FFA and pSTSFA) but also that part of the salience processing network where voice- and face-sensitivity converge (i.e., amygdala) are more sensitive to human voices and faces as compared to other types of complex and meaningful visual and acoustic stimuli. Within the auditory and visual association cortex all observed response associations with SA were modality-specific, i.e., involved the preferred stimulus of the respective voice or face processing region, with uniformly increased responses to voices or faces compared to all other stimulus classes in higher socially anxious individuals. Within the bilateral amygdala such a relationship could be observed in a statistically significant manner for voice-preferential responses but not for face-preferential responses and with a stronger association of SA and voice-preferential responses than with face-preferential responses. This points to an increased salience of vocal cues in SA. Possibly, the lack of a comparable finding for the visual modality may be explained by the processing bias towards visual cues known in adults (Robinson and Sloutsky, 2004; Santorelli, 2006). If face cues are generally more salient than voice cues in humans, this may lead to a decreased sensitivity for detecting associations of SA and face-sensitivity through a potential ceiling effect.

Within the TVA, the strongest association between SA severity and voice-sensitivity was observed closest to the posterior voice-sensitivity peak which is thought to be related to the extraction of meaningful patterns from vocal cues (Brück et al., 2011; Schirmer and Adolphs, 2017; Wildgruber et al., 2006). Additionally, the SA association maximum converges well with the section of the TVA preferentially processing emotional cues (Ethofer et al., 2012). This may indicate that,

within the voice processing network, SA primarily modulates the sensitivity of the TVA for the extraction of emotionally relevant information from voices. For faces an increased sensitivity within the pSTSFA is well in line with the assumption that this face area preferentially processes dynamic aspects of faces (Pitcher et al., 2011) and may be an indicator of an automatically enhanced effort to decipher facial expressions. The increased face-sensitivity within the FFA which exhibits a preference for time-constant aspects of faces (e.g., identity; Gobbini and Haxby, 2007; Haxby et al., 2000), in contrast, might rather be explained through an increased automatic parsing of faces for socially problematic or unproblematic acquaintances, or possibly reflect increased effort related to decreased facial identity recognition ability in SA (Davis et al., 2011). However, increased face-preferential responses in the FFA might also be less specifically but more parsimoniously explained by increased general attention to faces (Wojciulik et al., 1998) or increased social relevance of faces (Van Bavel et al., 2011) in SA.

Then again, the question arises if the increased cerebral voice- and face-sensitivities in SA represent cerebral characteristics which exist without causal relations to SA, or if they are the consequence of dysfunctional learning or potentially even one of the causes. Although it may seem more convincing to think of them as a consequence convergent with present models of SA (Morrison and Heimberg, 2013; Spence and Rapee, 2016; Wong and Rapee, 2016), an opposite mechanism should be considered. Increased sensitivity of the voice and face processing systems might be a genetically determined resource with the potential to develop social and emotional competencies (Grossmann et al., 2013; Neuhaus et al., 2016; Verhallen et al., 2017) but which in the context of other variables (e.g., an effective fear learning network, see also Fox and Kalin, 2014; or fear-related overgeneralization of social cues, see Ahrens et al., 2016) may turn into a risk factor or even pathogenetic factor.

Interestingly, also neutral or even positive faces (Bell et al., 2011; Gutiérrez-García and Calvo, 2017; Kreifelts et al., 2014; Machado-de-Sousa et al., 2010; Peschard and Philippot, 2017) and voices (Kreifelts et al., 2014; Peschard and Philippot, 2017; Ritter et al., 2015) tend to be misinterpreted as socially threatening in SA. These findings would be in line with the hypothesis that the stimuli in the present study could have been perceived as social threat and the activation patterns might reflect social threat processing rather than general face processing. But when not only socially threatening stimuli are evaluated as more threatening but also neutral and positive or even facial and vocal expressions which do not carry a specific meaning outside the context of direct bodily contact (i.e., tickling laughter; Kreifelts et al., 2017; Kreifelts et al., 2014; Ritter et al., 2015) then obviously these effects pertain to overall voice and face processing and might better be explained as the consequence of a negative social expectation bias (Ly and Roelofs, 2009; Wiemer and Pauli,

2016) when confronted with the carrier signals of human social communication, potentially as a consequence of fear-related over-generalization during social learning.

4.1. Perspectives and limitations

In face of the paucity of previous data on the behavioral processing correlates of face and voice cues in contrast to other classes of meaningful visual stimuli in terms of interpretation, attention and expectation biases, this should be explicitly addressed in future studies to aid the interpretation of the cerebral effects of SA on general face and voice processing. Under the hypothesis that, irrespective of explicitly expressed social threat, faces and voices might always be construed to be socially threatening in SA, it might also be valuable to balance stimulus material across categories for its potential to be evaluated as social threat to further disentangle the effects of social threat processing from general face and voice processing. Such an approach would be limited, however, by the constraint of comparators to potentially threat-relevant stimulus classes. With respect to the potential role of fear-related cue generalization during social learning, it might be valuable for future research to investigate contributing factors (e.g., frequency of socially threatening encounters, age during first socially threatening encounters, but also protective factors like frequency of positive social encounters).

Irrespective of the need to further clarify the exact source of the increased reactivity of the face and voice processing systems in SA, another avenue for further research can be envisaged in an adapted attention training directing attention away from faces and voices as tool to clarify the involvement of attention processes but on a more speculative note also to evaluate effects on SA symptoms and cognitive biases in analogy to established attention bias modification paradigms (Heeren et al., 2015).

And finally, a longitudinal approach adapted to the highest incidence rates of SA onset in later childhood and early adolescence (Wittchen et al., 1999) may shed light on the question if the increased reactivity of the face and voice processing systems is rather the consequence or the cause of SA.

5. Conclusion

In summary, this study demonstrated an increased sensitivity of the cerebral voice and face processing systems to their preferred cues in SA during presentation of neutral and slightly positive cues, thus connecting the neuroscience of SA with that of general face and voice processing. It appears plausible that the observed activation patterns reflect increased salience of and attention to voices and faces as common carrier signals of social threat, potentially automatically triggering the expectation of social threat and associated perception biases. Nevertheless, the present findings also emphasize the need for further research to disentangle the cerebral processes of basic social processing and those of threat processing in SA and clarify their role either as consequence or as cause of SA.

Acknowledgments

This work was supported by grants of the Fortüne-Program of the University of Tübingen (fortüne 1997-0-0, and fortune 2140-0-0).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.05.018>.

References

Ahrens, L.M., Pauli, P., Reif, A., Muhlberger, A., Langs, G., Aalderink, T., Wieser, M.J., 2016. Fear conditioning and stimulus generalization in patients with social anxiety disorder. *J. Anxiety Disord.* 44, 36–46.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. *Nature* 403, 309–312.

Bell, C., Bourke, C., Colhoun, H., Carter, F., Frampton, C., Porter, R., 2011. The misclassification of facial expressions in generalised social phobia. *J. Anxiety Disord.* 25, 278–283.

Bernstein, M., Yovel, G., 2015. Two neural pathways of face processing: a critical evaluation of current models. *Neurosci. Biobehav. Rev.* 55, 536–546.

Brück, C., Kreifelts, B., Wildgruber, D., 2011. Emotional voices in context: a neurobiological model of multimodal affective information processing. *Phys. Life Rev.* 8, 383–403.

Brühl, A.B., Delsignore, A., Komossa, K., Weidt, S., 2014. Neuroimaging in social anxiety disorder—a meta-analytic review resulting in a new neurofunctional model. *Neurosci. Biobehav. Rev.* 47, 260–280.

Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18, 192–205.

Davis, J.M., McKone, E., Dennett, H., O'Connor, K.B., O'Keary, R., Palermo, R., 2011. Individual differences in the ability to recognise facial identity are associated with social anxiety. *PLoS One* 6, e28800.

Epstein, R., Harris, A., Stanley, D., Kanwisher, N., 1999. The parahippocampal place area: recognition, navigation, or encoding? *Neuron* 23, 115–125.

Ethofer, T., Bretschner, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P., 2012. Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebr. Cortex* 22, 191–200.

Etkin, A., Wager, T.D., 2007. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am. J. Psychiatry* 164, 1476–1488.

Fox, A.S., Kalin, N.H., 2014. A translational neuroscience approach to understanding the development of social anxiety disorder and its pathophysiology. *Am. J. Psychiatry* 171, 1162–1173.

Friston, K.J., Glaser, D.E., Henson, R.N., Kiebel, S., Phillips, C., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: applications. *Neuroimage* 16, 484–512.

Gobbini, M.I., Haxby, J.V., 2007. Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41.

Grossmann, T., Vaish, A., Franz, J., Schroeder, R., Stoneking, M., Friederici, A.D., 2013. Emotional voice processing: investigating the role of genetic variation in the serotonin transporter across development. *PLoS One* 8, e68377.

Gutiérrez-García, A., Calvo, M.G., 2017. Social anxiety and threat-related interpretation of dynamic facial expressions: sensitivity and response bias. *Pers. Individ. Differ.* 107, 10–16.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends Cognit. Sci.* 4, 223–233.

Heeren, A., Mogoase, C., Philippot, P., McNally, R.J., 2015. Attention bias modification for social anxiety: a systematic review and meta-analysis. *Clin. Psychol. Rev.* 40, 76–90.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.

Kanwisher, N., Yovel, G., 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 2109–2128.

Kreifelts, B., Brück, C., Ethofer, T., Ritter, J., Weigel, L., Erb, M., Wildgruber, D., 2017. Prefrontal mediation of emotion regulation in social anxiety disorder during laughter perception. *Neuropsychologia* 96, 175–183.

Kreifelts, B., Brück, C., Ritter, J., Ethofer, T., Domin, M., Lotze, M., Jacob, H., Schlipf, S., Wildgruber, D., 2014. They are laughing at me: cerebral mediation of cognitive biases in social anxiety. *PLoS One* 9, e99815.

Laux, L., Glanzmann, P., Schaffner, P., Spielberger, C.D., 1981. *Das State-Trait-Angstinventar*. Beltz, Weinheim.

Lehrl, S., 2005. *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B*, fifth ed. Spitta Verlag, Balingen.

Ly, V., Roelofs, K., 2009. Social anxiety and cognitive expectancy of aversive outcome in avoidance conditioning. *Behav. Res. Ther.* 47, 840–847.

Machado-de-Sousa, J.P., Arrais, K.C., Alves, N.T., Chagas, M.H., de Meneses-Gaya, C., Crippa, J.A., Hallak, J.E., 2010. Facial affect processing in social anxiety: tasks and stimuli. *J. Neurosci. Methods* 193, 1–6.

Mende-Siedlecki, P., Verosky, S.C., Turk-Browne, N.B., Todorov, A., 2013. Robust selectivity for faces in the human amygdala in the absence of expressions. *J. Cogn. Neurosci.* 25, 2086–2106.

Menon, V., 2015. Salience network. In: Toga, A.W. (Ed.), *Brain Mapping: an Encyclopedic Reference*. Academic Press: Elsevier, pp. 597–611.

Miskovic, V., Schmidt, L.A., 2012. Social fearfulness in the human brain. *Neurosci. Biobehav. Rev.* 36, 459–478.

Morrison, A.S., Heimberg, R.G., 2013. Social anxiety and social anxiety disorder. *Annu. Rev. Clin. Psychol.* 9, 249–274.

Neuhaus, E., Kresse, A., Faja, S., Bernier, R.A., Webb, S.J., 2016. Face processing among twins with and without autism: social correlates and twin concordance. *Soc. Cognit. Affect Neurosci.* 11, 44–54.

Öhman, A., 1986. Face the beast and fear the face: animal and social fears as prototypes for evolutionary analyses of emotion. *Psychophysiology* 23, 123–145.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.

Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage* 119, 164–174.

- Peschard, V., Philippot, P., 2017. Overestimation of threat from neutral faces and voices in social anxiety. *J. Behav. Ther. Exp. Psychiatry* 57, 206–211.
- Pitcher, D., Dilks, D.D., Saxe, R.R., Triantafyllou, C., Kanwisher, N., 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage* 56, 2356–2363.
- Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G., 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J. Neurosci.* 16, 5205–5215.
- Quadflieg, S., Mohr, A., Mentzel, H.J., Miltner, W.H., Straube, T., 2008. Modulation of the neural network involved in the processing of anger prosody: the role of task-relevance and social phobia. *Biol. Psychol.* 78, 129–137.
- Ritter, J., Brück, C., Jacob, H., Wildgruber, D., Kreifelts, B., 2015. Laughter perception in social anxiety. *J. Psychiatr. Res.* 60C, 178–184.
- Robinson, C.W., Sloutsky, V.M., 2004. Auditory dominance and its change in the course of development. *Child Dev.* 75, 1387–1401.
- Santorelli, N.T., 2006. Perception of Emotion from Facial Expression and Affective Prosody. Georgia State University.
- Schirmer, A., Adolphs, R., 2017. Emotion perception from face, voice, and touch: comparisons and convergence. *Trends Cognit. Sci.* 21, 216–228.
- Simon, D., Becker, M., Mothes-Lasch, M., Miltner, W.H., Straube, T., 2017. Loud and angry: sound intensity modulates amygdala activation to angry voices in social anxiety disorder. *Soc. Cognit. Affect Neurosci.* 12, 409–416.
- Spence, S.H., Rapee, R.M., 2016. The etiology of social anxiety disorder: an evidence-based model. *Behav. Res. Ther.* 86, 50–67.
- Stangier, U., Heidenreich, T., 2003. Die Liebowitz Soziale Angst-Skala (LSAS). Hogrefe, Göttingen.
- Stein, D.J., Ruscio, A.M., Lee, S., Petukhova, M., Alonso, J., Andrade, L.H., Benjet, C., Bromet, E., Demyttenaere, K., Florescu, S., de Girolamo, G., de Graaf, R., Gureje, O., He, Y., Hinkov, H., Hu, C., Iwata, N., Karam, E.G., Lepine, J.P., Matschinger, H., Oakley Browne, M., Posada-Villa, J., Sagar, R., Williams, D.R., Kessler, R.C., 2010. Subtyping social anxiety disorder in developed and developing countries. *Depress. Anxiety* 27, 390–403.
- Trower, P., Gilbert, P., 1989. New theoretical conceptions of social anxiety and social phobia. *Clin. Psychol. Rev.* 9, 19–35.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.
- Van Bavel, J.J., Packer, D.J., Cunningham, W.A., 2011. Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: evidence of in-group enhancement (not out-group disregard). *J. Cogn. Neurosci.* 23, 3343–3354.
- Verhallen, R.J., Bosten, J.M., Goodbourn, P.T., Lawrance-Owen, A.J., Bargary, G., Mollon, J.D., 2017. General and specific factors in the processing of faces. *Vis. Res.* 141, 217–227.
- Wiemer, J., Pauli, P., 2016. Fear-relevant illusory correlations in different fears and anxiety disorders: a review of the literature. *J. Anxiety Disord.* 42, 113–128.
- Wildgruber, D., Ackermann, H., Kreifelts, B., Ethofer, T., 2006. Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog. Brain Res.* 156, 249–268.
- Wittchen, H.U., Stein, M.B., Kessler, R.C., 1999. Social fears and social phobia in a community sample of adolescents and young adults: prevalence, risk factors and comorbidity. *Psychol. Med.* 29, 309–323.
- Wittchen, H.U., Zaudig, M., Fydrich, T., 1997. Strukturiertes Klinisches Interview für DSM-IV. Hogrefe, Göttingen.
- Wojculik, E., Kanwisher, N., Driver, J., 1998. Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *J. Neurophysiol.* 79, 1574–1578.
- Wong, Q.J.J., Rapee, R.M., 2016. The aetiology and maintenance of social anxiety disorder: a synthesis of complimentary theoretical models and formulation of a new integrated model. *J. Affect. Disord.* 203, 84–100.