NeuroImage 76 (2013) 45–56

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus

Thomas Ethofer [a,b,*], Johannes Bretscher [a], Sarah Wiethoff [c], Jeanne Bisch [a], Sarah Schlipf [a], Dirk Wildgruber [a], Benjamin Kreifelts [a]

[a] *Department of General Psychiatry, University of Tübingen, Tübingen, Germany*
[b] *Department of Biomedical Resonance, University of Tübingen, Tübingen, Germany*
[c] *Department of Neurology, University of Tübingen, Tübingen, Germany*

ARTICLE INFO

Article history:
Accepted 26 February 2013
Available online 16 March 2013

Keywords:
Superior temporal sulcus
Orbitofrontal cortex
Voice
Face
Audiovisual integration
Social signals

ABSTRACT

It was the aim of this study to delineate the areas along the right superior temporal sulcus (STS) for processing of faces, voices, and face–voice integration using established functional magnetic resonance imaging (fMRI) localizers and to assess their structural connectivity profile with diffusion tensor imaging (DTI). We combined this approach with an fMRI adaptation design during which the participants judged emotions in facial expressions and prosody and demonstrated response habituation in the orbitofrontal cortex (OFC) which occurred irrespective of the sensory modality. These functional data were in line with DTI findings showing separable fiber projections of the three different STS modules converging in the OFC which run through the external capsule for the voice area, through the dorsal superior longitudinal fasciculus (SLF) for the face area and through the ventral SLF for the audiovisual integration area. The OFC was structurally connected with the supplementary motor area (SMA) and activation in these two areas was correlated with faster stimulus evaluation during repetition priming. Based on these structural and functional properties, we propose that the OFC is part of the extended system for perception of emotional information in faces and voices and constitutes a neural interface linking sensory areas with brain regions implicated in generation of behavioral responses.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Faces and voices carry a wealth of socially relevant cues including age, gender, identity, attractiveness, as well as intentions, attitudes and the current affective state of the protagonists. Individuals living in complex social systems need to extract this information in a fast and efficient manner. Consequently, the majority of humans can be regarded as face and voice experts. To predict the emotional state, intentions, attitudes, and future actions of other people, correct interpretation of facial and vocal features that can dynamically change during natural communication, such as eye gaze position, facial expression, and speech melody (prosody), is required.

Current models on face perception (Haxby et al., 2000; Ishai, 2008; Tsao and Livingstone, 2008) propose a functional specialization of the face-sensitive cortices along the posterior superior temporal sulcus (pSTS, Puce et al., 1998) for processing of such dynamic facial cues while the occipital face area (e.g. Halgren et al., 1999) and the fusiform face area (Kanwisher et al., 1997) have been proposed as critical sites for assessment of invariant facial features (e.g., identity, gender). This model is supported by neurophysiological data obtained in non-human primates (Hasselmo et al., 1989; Perrett et al., 1985, 1992), observations made in patients with brain lesions (Akiyama et al., 2006; Campbell et al., 1990; Grüsser and Landis, 1991), and neuroimaging experiments comparing top-down (judgment of identity versus eye gaze, Hoffman and Haxby, 2000) or bottom-up effects (presentation of dynamic versus static stimuli, Pitcher et al., 2011) during face processing.

The neural correlates for voice perception are less well understood than those for processing of faces (Latinus and Belin, 2011). However, voice-sensitive areas have been described along the superior temporal sulcus (STS, Belin et al., 2000). In analogy to the face processing system, separate pathways for assessment of invariant and changeable vocal features have been identified by functional magnetic resonance imaging (fMRI) studies (for a detailed review see Campanella and Belin, 2007). While invariant information expressed in the voice (e.g., gender, identity) has been shown to be represented in the anterior STS adjacent to the temporal pole (aSTS, Belin and Zatorre, 2003; Charest et al., 2013; Latinus et al., 2011; von Kriegstein et al., 2003; von Kriegstein and Giraud, 2006), a modulation of the response amplitude by emotional prosody has been demonstrated for the middle part of the STS (mSTS, Ethofer et al., 2006b; Grandjean et al., 2005) which has been shown to be particularly sensitive to specific acoustic properties (Andics et al., 2010; Kriegstein and Giraud, 2004; Wiethoff

[*] Corresponding author at: Department of Biomedical Magnetic Resonance University of Tübingen Otfried-Müller-Str. 51 72076 Tübingen, Germany. Fax: 49 7071 294371.
E-mail address: Thomas.Ethofer@med.uni-tuebingen.de (T. Ethofer).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.02.064

et al., 2008). In agreement with these results, the spatial activation pattern of voice-sensitive areas along the STS can be employed to successfully predict which emotion (Ethofer et al., 2009b) or which speaker (Formisano et al., 2008) was perceived by the listener.

While integration of auditory and visual speech-related signals is thought to rely on the left STS (Arnal et al., 2009; Blank and von Kriegstein, 2012), integration of socially relevant signals from emotional prosody and facial expressions has been shown to occur at the overlap of right hemispheric face- and voice-sensitive STS cortices (Kreifelts et al., 2009; Szycik et al., 2008; Wright et al., 2003). Neuroimaging studies investigating effective connectivity (Friston et al., 1997, 2003) of these STS regions revealed interactions with the FFA (Kreifelts et al., 2007; Muller et al., 2012) and the amygdala (Muller et al., 2012) during audiovisual integration of emotional signals. A recent neuroimaging study demonstrated direct connections between STS areas responsive to voice identity and the FFA (Blank et al., 2011). Apart from that, however, it is unknown which other brain areas are structurally interconnected with the different parts of the STS. Previous lesion studies (Hornak et al., 1996, 2003) demonstrated impaired recognition of both emotional facial expressions and prosody in patients with damaged orbitofrontal cortex (OFC), an area which is long known from animal studies to receive input from multiple senses (Jones and Powell, 1970) and thus fulfills the criteria of a multisensory convergence zone (Driver and Noesselt, 2008; Mesulam, 1998). In line with these findings, neuroimaging data indicated enhanced activation during active judgment of emotions in faces and voices (e.g. Ethofer et al., 2006a; Sabatinelli et al., 2011). Based on these convergent results, we predicted structural fiber connections towards and functional activation within the orbital part of the inferior frontal cortex during processing of social signals in faces and voices. To directly test this hypothesis, we combined diffusion tensor imaging (DTI) with a factorial adaptation fMRI paradigm which was specifically designed to test for regional habituation during repeated exposure to faces, voices, and face–voice combinations. Such attenuation of brain responses has been termed repetition suppression — a robust phenomenon which occurs consistently after repeated presentation of identical stimuli and can be exploited to examine which brain regions participate in processing of a certain stimulus type. It has been proposed that repetition suppression reflects top-down mediated perceptual expectations (Summerfield et al., 2008) as well as bottom-up sharpening of neural responses (Larsson and Smith, 2012) which typically results in more accurate and faster behavioral responses (Grill-Spector et al., 2006). Therefore, adaptation designs additionally offer the opportunity to reveal the neural structures which mediate repetition priming effects (Schacter and Buckner, 1998). It should be noted, however, that repetition suppression has also been found outside of neural systems engaged in processing of a particular stimulus type presumably via carry-over effects from other brain areas (Mur et al., 2010) questioning the specificity of effects based on response habituation alone.

To localize potential convergence zones for processing of dynamic social information (i.e., emotional cues in facial expressions and prosody) irrespective of the sensory modality, we complemented the classical fMRI adaptation method with analysis strategies that enabled us to identify brain networks in which activity is correlated with behavioral effects (i.e., faster reaction times during classification of social cues) as well as structural approaches that reveal fiber connections towards these areas. Specifically, we hypothesized structural fiber projections between the areas for processing of faces, voices, and audiovisual integration along the right STS on the one hand and the OFC as a multisensory convergence zone for processing of dynamic facial and vocal cues on the other hand. Moreover, we predicted that modality specific cortices, such as the face-sensitive and voice-sensitive cortex in the STS, would show an enhanced response habituation for their respective preferred stimulus class (i.e., stronger habituation of face-sensitive STS to faces than voices and vice versa for the voice-sensitive STS), whereas multisensory convergence zones, such as the audiovisual integration area in the STS and the OFC, would elicit a consistent habituation pattern that does not depend on the sensory modality occurring similarly for faces, voices, and face–voice combinations.

Material and methods

Participants, stimulus material, and experimental design

Twenty-three healthy, right-handed German native speakers (13 females; 23.0 ± 4.2 years) participated in one DTI and three fMRI experiments. Right-handedness was assessed with the Edinburgh Inventory (Oldfield, 1971). None of the participants had a history of neurological or psychiatric illness, substance abuse, or impaired hearing. Vision was normal or corrected to normal. None of the participants was on any medication. The study was performed according to the Code of Ethics of the World Medical Association (Declaration of Helsinki). All subjects gave their written informed consent prior to inclusion in the study.

The fMRI experiments included a face localizer (Kanwisher et al., 1997), a voice localizer (Belin et al., 2000) and a bimodal face–voice integration localizer (Kreifelts et al., 2009) modified to additionally enable investigation of modality-dependent habituation effects (Ethofer et al., 2009a; Grill-Spector et al., 1999).

The face localizer was adapted from previous studies on face processing (Epstein et al., 1999; Haxby et al., 2000; Kanwisher et al., 1997) and included pictures from four different categories (faces, houses, objects, and natural scenes) presented using a block-design. Six blocks (duration: 16 s) of each category pseudorandomized within the experiment were presented separated by short rest periods of 1.5 s. Within each block, 20 stimuli of one category were presented in a random order for 300 ms interleaved with 500 ms of fixation. To keep the participants' attention fixed on the stimuli, they were instructed to press a button with their right index finger when they saw a picture directly repeated (one-back task). Positions of repeated stimuli were randomized within blocks with the restriction that one occurred during the first half of the block and one during the second half.

The voice localizer consisted of a passive-listening block design experiment with 32 stimulations and 16 silent epochs (each 8 s), as validated in previous research (Belin et al., 2000; Ethofer et al., 2009b). These stimuli included 16 blocks with human voices (HV; e.g., speech, sighs, laughs), eight blocks with animal sounds (AS; cries of various animals), and eight blocks with environmental sounds (ES; e.g., doors, telephones, cars).

The stimulus material used in the face–voice integration experiment consisted of short video clips (duration: 848 ms ± 295 ms, mean ± standard deviation) during which professional actors spoke words with emotionally neutral semantic content (the list of words is presented as supplemental material) in happy, neutral, or angry prosody with an emotionally congruent facial expression. These stimuli were evaluated in a prestudy outside the scanner including 20 subjects (10 females, 24.2 ± 3.8 years) to confirm that the emotional category intended by the actors was recognized with a high accuracy during auditory (A = sound clip without visual presentation; mean recognition rate: 84%), visual (V = mute video clip; mean recognition rate: 89%) and audiovisual presentation (AV = video clip with sound; mean recognition rate: 96%). To investigate modality-dependent habituation effects, each stimulus was presented three times during the course of the experiment (18 words × 3 modalities × 3 repetitions = 162 stimuli). These stimuli were presented within the framework of an event-related design with a varying inter stimulus interval (8.2–10.2 s) during three imaging runs (duration: about 10 min) each of which contained 18 auditory, 18 visual, and 18 audiovisual stimuli. On average, repetitions occurred with a temporal delay of 47.6 ± 36.9 s and 5.2 ± 4.0 intervening stimuli for all three modalities. To

Download English Version:

https://daneshyari.com/en/article/6029385

Download Persian Version:

https://daneshyari.com/article/6029385

Daneshyari.com