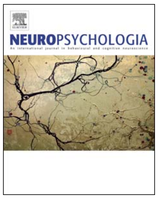
Neuropsychologia xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Neuropsychologia

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



Normal voice processing after posterior superior temporal sulcus lesion

Guo Jiahui[a,*], Lúcia Garrido[b], Ran R. Liu[c], Tirta Susilo[d], Jason J.S. Barton[e], Bradley Duchaine[a]

[a] Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA
[b] Division of Psychology, Department of Life Sciences, Brunel University, Uxbridge UB8 3PH, United Kingdom
[c] Department of Medicine, division of neurology, Kingston General Hospital, Queen's University, Ontario, Canada
[d] School of Psychology, Victoria University of Wellington, Wellington, New Zealand
[e] Departments of Medicine (Neurology) & Opthalmology and Visual Sciences, University of British Columbia, Canada

ARTICLE INFO

Keywords:
Voice perception
pSTS
Patient study

ABSTRACT

The right posterior superior temporal sulcus (pSTS) shows a strong response to voices, but the cognitive processes generating this response are unclear. One possibility is that this activity reflects basic voice processing. However, several fMRI and magnetoencephalography findings suggest instead that pSTS serves as an integrative hub that combines voice and face information. Here we investigate whether right pSTS contributes to basic voice processing by testing Faith, a patient whose right pSTS was resected, with eight behavioral tasks assessing voice identity perception and recognition, voice sex perception, and voice expression perception. Faith performed normally on all the tasks. Her normal performance indicates right pSTS is not necessary for intact voice recognition and suggests that pSTS activations to voices reflect higher-level processes.

1. Introduction

The superior temporal sulcus (STS) extends anteriorly from the inferior parietal lobe along the entire temporal lobe and is one of the longest sulci in the brain. The STS plays a central role in processing social information, including the perception of faces (Haxby et al., 2000), voices (Belin et al., 2000), and biological motion (Yovel and O'Toole, 2016). In addition to representing social perceptual information, the STS integrates different social percepts to generate higher-level representations (Campanella and Belin, 2007; Frith and Frith, 2003; Yovel and O'Toole, 2016). A meta-analysis of more than 100 fMRI studies of the STS found motion processing, face processing, and audiovisual integration reliably activated the posterior STS (pSTS), speech processing activated the anterior STS, and theory-of-mind tasks led to activity along the entire STS (Hein and Knight, 2008). A recent study used localizers to identify areas in the STS showing selective responses to a variety of social stimuli (Deen et al., 2015). The results showed selectivity to theory-of-mind reasoning in the angular gyrus and surrounding sulci as well as the middle-to-anterior STS, biological motion in the most posterior region of STS, face processing in a more anterior region of pSTS with weaker face responses in middle-to-anterior STS, and a broad response to voices that peaked in middle STS.

The STS response to voices extended into pSTS (Deen et al., 2015), and a number of other studies have also suggested that the pSTS processes information about voices. In the first paper to identify voice-selective areas, Belin et al. (2000) reported three clusters that responded selectively to voices including one in pSTS (see also Watson et al., 2014b). The pSTS, along with anterior STS, showed an elevated response when participants attended to vocal identity rather than the meaning of a sentence (von Kriegstein and Giraud, 2004). In an adaptation study, repetition suppression to vocal identity was found in pSTS and middle STS bilaterally (Andics et al., 2010). Notably, in a fair proportion of voice studies, effects are more pronounced in right STS than left STS (Belin et al., 2002, 2000; Gainotti, 2013), and pSTS's response to voices may also be stronger on the right than the left (Schall et al., 2014).

What sort of cognitive computations do pSTS activations to voices reflect? One possibility is that pSTS carries out fundamental voice processing by representing the auditory properties of voices and categorizing vocal identity and characteristics like sex, expression, and age. These processes depend on voice-selective regions in more anterior regions of STS (Belin et al., 2011; Bestelmeyer et al., 2011), but pSTS may also play a role in them. Another possibility is that right pSTS voice activations are driven by higher-level voice processing such as computations integrating voice representations with information from faces and other types of social information (Belin et al., 2011; Campanella and Belin, 2007; Thurman et al., 2016; Yovel and O'Toole, 2016). Consistent with this last account, pSTS shows cross-modal fMRI adaptation effects between faces and voices when facial expressions are similar to the preceding vocal expression (Watson et al.,

---

* Corresponding author.
E-mail address: Jiahui.Guo.GR@dartmouth.edu (G. Jiahui).

http://dx.doi.org/10.1016/j.neuropsychologia.2017.03.008
Received 13 September 2016; Received in revised form 4 March 2017; Accepted 5 March 2017

Please cite this article as: Jiahui, G., Neuropsychologia (2017), http://dx.doi.org/10.1016/j.neuropsychologia.2017.03.008

2014b). Such integration may be specific to people: a conjunction analysis of fMRI data indicated that right pSTS integrates face and voice information but not visual and auditory information about objects (Watson et al., 2014a). A recent study found strong correlations between the strength of the preference of a voxel in pSTS for visual mouth movements and the magnitude of its auditory speech response, as well as its preference for vocal sounds (Zhu and Beauchamp, 2017). Magnetoencephalography (MEG) suggests that the right pSTS shows a stronger response to combined face-voice stimuli than the sum of unimodal face and voice components (Hagan et al., 2009). Finally, Deen et al. (2015) found that a pSTS region of interest identified with a dynamic face localizer showed comparable activation to voices and faces.

To clarify the role of the pSTS in voice processing suggested by fMRI studies, it would be helpful to have complementary data from a lesion study. Here, we assessed the role of right pSTS in voice processing with behavioral experiments in a patient, Faith, whose right pSTS was lost due to a tumor resection. The surgery left the more anterior regions of Faith's STS intact, including those containing the temporal voice areas (TVAs) in the middle and anterior STS. We tested Faith with eight behavioral tasks that tap a wide range of voice processing abilities including identity discrimination, identity memory, sex categorization, and expression categorization. Impaired performance with some or all of the tasks would support the hypothesis that right pSTS is involved in basic aspects of voice processing, while intact performance would be more consistent with the hypothesis that the voice activations seen in the pSTS are reflections of higher-level voice processing.

2. Method and results

2.1. Patient case

Faith is a right-handed speech therapist, and English is her native language. In 2009 she had a right occipitotemporal resection to remove a tumor, and in July 2015, she had a second resection along the margin of the same location followed by proton radiation therapy. Following her first surgery, she noted severe face processing deficits. Her impairments affect many types of face processing, including perception of identity, expression, and gaze (Susilo et al., 2015). Faith believes her ability to process voices remains normal. She completed the first eight tasks described below in April 2015 when she was 52-years-old, and did a final task (three-alternative expression test) in February 2016 when she was 53.

2.2. Faith's lesion and its overlap with voice-selective activations in normal participants

2.2.1. Anatomical scan

Faith was scanned on a 3.0-T Phillips MR scanner (Philips Medical Systems, WA, USA) with a SENSE (SENSitivity Encoding) 32-channel head coil. An anatomical volume was acquired using a high-resolution 3D magnetization-prepared rapid gradient-echo sequence (220 slices, field of view = 240 mm, acquisition matrix = 256 × 256, voxel size = 1 × 0.94 × 0.94 mm). This scan was skull stripped and then warped to Talairach space. The high-resolution MR images of Faith's brain (Fig. 1A) show a lesion extending from the fusiform to the superior part of temporal lobe in the right hemisphere, encompassing a large part of her posterior superior temporal sulcus (pSTS). The estimated lesion size on the axial, coronal, and sagittal axes is 43 mm, 37 mm, and 37 mm, respectively.

2.2.2. Peak coordinates from five papers

To demonstrate that Faith's lesion overlaps with voice-selective responses in the literature, peak voxels implicated in voice processing from five papers (Belin et al., 2000; Deen et al., 2015; von Kriegstein and Giraud, 2004; Watson et al., 2014b, 2014a) are displayed on Faith's brain. Coordinates for 29 peak voxels in the temporal lobe that were listed in the tables in the five papers were extracted manually and plotted on the standard brain in Talairach space with a 5 mm radius sphere centered at the peak voxel (Supplementary Table 1). The coordinates map was then converted to a brain mask and overlaid on Faith's brain. Fig. 1B shows the overlap of the peak voxels with Faith's lesion.

2.2.3. Overlap with TVA probabilistic map

As another approach to determining which voice-selective regions may have been affected by Faith's resection, we compared her lesion to a probability map of the temporal voice areas (TVA) downloaded from http://vnl.psy.gla.ac.uk/resources.php. Belin and colleagues created the map based on individually-thresholded data from 152 participants. Each individual's T-map showing voice-selective voxels (voices > non-voice sounds) was corrected for multiple comparisons based on the spatial extent at $q < 0.05$ (Chumbley and Friston, 2009). They then applied a Gamma-Gaussian mixture model to separate the null voxels from the active voxels (Gorgolewski et al., 2012). Data from each participant was transformed into MNI space, binarized, summed, and normalized to 100 to create the group probability map. We converted this map from MNI space to Talairach space so it could be overlaid on Faith's anatomical scan. Voxels that were voice-selective in 10% or more of the participants are displayed in Fig. 1C.

2.3. Voice localizer to assess Faith's temporal voice areas

2.3.1. Stimuli and experiment procedures

To exam whether Faith shows voice-selective areas in her intact cortex we conducted a standard TVA localizer (Belin et al., 2000). Voice stimuli were designed by Belin and his colleagues and were downloaded from http://vnl.psy.gla.ac.uk/resources.php. This functional localizer lasts 10 min and contains one run in total. It contains 40 eight-seconds blocks of sounds (16 bit, mono, 22,050 Hz sampling rate). Half of the blocks consists of vocal sounds (speech and non-speech), and the other half consists of nonvocal sounds (industrial sounds, environmental sounds, and a few animal vocalizations). All sounds have been normalized and a 1 kHz tone of similar energy was provided for calibration. The order of the sound blocks was provided on the website and optimized for the vocal vs. nonvocal contrast.

2.3.2. MRI acquisition

Faith was scanned on the same 3.0-T Phillips MR scanner as the anatomical scan (Philips Medical Systems, WA, USA) with a SENSE (SENSitivity Encoding) 32-channel head coil. Functional images were collected using echo-planar functional images (time to repeat = 2000 ms, time echo = 35 ms, flip angle = 90°, voxel size = 3 × 3 × 3 mm). Each volume consisted of 36 interleaved 3 mm thick slices with 0 mm interslice gap. The slice volume was adjusted to cover most of the brain including the entire temporal lobe. Previous studies found that the location and extent of susceptibility effects are influenced by the slice orientation and phase-encoding direction (Ogawa et al., 1990; Ojemann et al., 1997). In our study, we adopted oblique slice orientation aligned with each participant's anterior commissure–posterior commissure (AC–PC) line, because it produces fewer susceptibility artifacts than the commonly used transverse orientation (Ojemann et al., 1997) and at the same time provides better coverage of the brain. The phase-encoding direction (anterior–posterior) was chosen to move the signal loss away from the more anterior part of the brain.

2.3.3. Data analysis

Imaging data were analyzed using the AFNI software package (Cox, 1996). Before statistical analysis, the first volume was discarded to allow for magnetic saturation effects, and each volume was registered to the third volume. The EPI data were warped to align with the