



City Research Online

City St George's, University of London

Citation: Jiahui, G., Garrido, L., Liu, R., Susilo, T., Barton, J. & Duchaine, B. (2017). Normal voice processing after posterior superior temporal sulcus lesion. *Neuropsychologia*, 105, pp. 215-222. doi: 10.1016/j.neuropsychologia.2017.03.008

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/23453/>

Link to published version:

<https://doi.org/10.1016/j.neuropsychologia.2017.03.008>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

1
2
3
4
5 **Normal voice processing after posterior superior**
6
7
8 **temporal sulcus lesion**
9

10
11
12
13
14
15
16 Guo Jiahui^{a*}, Lúcia Garrido^b, Ran R. Liu^c, Tirta Susilo^d, Jason J. S.
17
18 Barton^e and Bradley Duchaine^a
19
20
21
22
23

24 *^aDepartment of Psychological and Brain Sciences, Dartmouth College, Hanover, NH,*
25 *USA; ^bDivision of Psychology, Department of Life Sciences, Brunel University, Uxbridge*
26 *UB8 3PH, United Kingdom; ^cDepartment of Medicine, division of neurology, Kingston*
27 *General Hospital, Queen's University, Ontario, Canada; ^dSchool of Psychology, Victoria*
28 *University of Wellington, Wellington, New Zealand; ^eDepartments of Medicine (Neurology)*
29 *& Ophthalmology and Visual Sciences, University of British Columbia, Canada*
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56

57
58 * Corresponding author at: Dartmouth College, Hanover, Department of Psychological and Brain
59 Sciences, NH, USA, 03755. Tel.: +1 603 277 0856.
60 E-mail address: Jiahui.Guo.GR@dartmouth.edu
61
62
63
64
65

1
2
3
4 Abstract

5
6 The right posterior superior temporal sulcus (pSTS) shows a strong response to voices,
7
8 but the cognitive processes generating this response are unclear. One possibility is that this
9
10 activity reflects basic voice processing. However, several fMRI and
11
12 magnetoencephalography findings suggest instead that pSTS serves as an integrative hub that
13
14 combines voice and face information. Here we investigate whether right pSTS contributes to
15
16 basic voice processing by testing Faith, a patient whose right pSTS was resected, with eight
17
18 behavioral tasks assessing voice identity perception and recognition, voice sex perception,
19
20 and voice expression perception. Faith performed normally on all the tasks. Her normal
21
22 performance indicates right pSTS is not necessary for intact voice recognition and suggests
23
24 that pSTS activations to voices reflect higher-level processes.

25
26 **Keywords:** *voice perception, pSTS, patient study*
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6 1. Introduction
7

8 The superior temporal sulcus (STS) extends anteriorly from the inferior parietal lobe
9 along the entire temporal lobe and is one of the longest sulci in the brain. The STS plays a
10 central role in processing social information, including the perception of faces (Haxby,
11 Hoffman, & Gobbini, 2000), voices (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), and
12 biological motion (Yovel & O’Toole, 2016). In addition to representing social perceptual
13 information, the STS integrates different social percepts to generate higher-level
14 representations (Campanella & Belin, 2007; Frith & Frith, 2003; Yovel & O’Toole, 2016). A
15 meta-analysis of more than 100 fMRI studies of the STS found motion processing, face
16 processing, and audiovisual integration reliably activated the posterior STS (pSTS), speech
17 processing activated the anterior STS, and theory-of-mind tasks led to activity along the
18 entire STS (Hein & Knight, 2008). A recent study used localizers to identify areas in the STS
19 showing selective responses to a variety of social stimuli (Deen, Koldewyn, Kanwisher, &
20 Saxe, 2015). The results showed selectivity to theory-of-mind reasoning in the angular gyrus
21 and surrounding sulci as well as the middle-to-anterior STS, biological motion in the most
22 posterior region of STS, face processing in a more anterior region of pSTS with weaker face
23 responses in middle-to-anterior STS, and a broad response to voices that peaked in middle
24 STS.
25

26 The STS response to voices extended into pSTS (Deen et al., 2015), and a number of
27 other studies have also suggested that the pSTS processes information about voices. In the
28 first paper to identify voice-selective areas, Belin et al. (2000) reported three clusters that
29 responded selectively to voices including one in pSTS (see also Watson, Latinus, Noguchi, et
30 al., 2014). The pSTS, along with anterior STS, showed an elevated response when
31 participants attended to vocal identity rather than the meaning of a sentence (von Kriegstein
32 & Giraud, 2004). In an adaptation study, repetition suppression to vocal identity was found in
33 pSTS and middle STS bilaterally (Andics et al., 2010). Notably, in a fair proportion of voice
34 studies, effects are more pronounced in right STS than left STS (Belin, Zatorre, & Ahad,
35 2002; Belin et al., 2000; Gainotti, 2013), and pSTS’s response to voices may also be stronger
36 on the right than the left (Schall, Kiebel, Maess, & von Kriegstein, 2014).
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 What sort of cognitive computations do pSTS activations to voices reflect? One
5
6 possibility is that pSTS carries out fundamental voice processing by representing the auditory
7
8 properties of voices and categorizing vocal identity and characteristics like sex, expression,
9
10 and age. These processes depend on voice-selective regions in more anterior regions of STS
11
12 (Belin, Bestelmeyer, Latinus, & Watson, 2011; Bestelmeyer, Belin, & Grosbras, 2011), but
13
14 pSTS may also play a role in them. Another possibility is that right pSTS voice activations
15
16 are driven by higher-level voice processing such as computations integrating voice
17
18 representations with information from faces and other types of social information (Belin et
19
20 al., 2011; Campanella & Belin, 2007; Thurman, van Boxtel, Monti, Chiang, & Lu, 2016;
21
22 Yovel & O’Toole, 2016). Consistent with this last account, pSTS shows cross-modal fMRI
23
24 adaptation effects between faces and voices when facial expressions are similar to the
25
26 preceding vocal expression (Watson, Latinus, Noguchi, et al., 2014). Such integration may be
27
28 specific to people: a conjunction analysis of fMRI data indicated that right pSTS integrates
29
30 face and voice information but not visual and auditory information about objects (Watson,
31
32 Latinus, Charest, Crabbe, & Belin, 2014). **A recent study found strong correlations between
33
34 the strength of the preference of a voxel in pSTS for visual mouth movements and the
35
36 magnitude of its auditory speech response, as well as its preference for vocal sounds (Zhu &
37
38 Beauchamp, 2017).** Magnetoencephalography (MEG) suggests that the right pSTS shows a
39
40 stronger response to combined face-voice stimuli than the sum of unimodal face and voice
41
42 components (Hagan et al., 2009). Finally, Deen et al. (2015) found that a pSTS region of
43
44 interest identified with a dynamic face localizer showed comparable activation to voices and
45
46 faces.

47 To clarify the role of the pSTS in voice processing suggested by fMRI studies, it would
48
49 be helpful to have complementary data from a lesion study. Here, we assessed the role of
50
51 right pSTS in voice processing with behavioral experiments in a patient, Faith, whose right
52
53 pSTS was lost due to a tumor resection. The surgery left the more anterior regions of Faith’s
54
55 STS intact, including those containing the temporal voice areas (TVAs) in the middle and
56
57 anterior STS. We tested Faith with eight behavioral tasks that tap a wide range of voice
58
59 processing abilities including identity discrimination, identity memory, sex categorization,
60
61 and expression categorization. Impaired performance with some or all of the tasks would
62
63 support the hypothesis that right pSTS is involved in basic aspects of voice processing, while
64
65

1
2
3
4 intact performance would be more consistent with the hypothesis that the voice activations
5 seen in the pSTS are reflections of higher-level voice processing.
6
7
8

9 10 2. Method & Results

11 2.1 Patient Case

12
13 Faith is a right-handed speech therapist, and English is her native language. In 2009 she
14 had a right occipitotemporal resection to remove a tumor, and in July 2015, she had a second
15 resection along the margin of the same location followed by proton radiation therapy.
16
17 Following her first surgery, she noted severe face processing deficits. Her impairments affect
18 many types of face processing, including perception of identity, expression, and gaze (Susilo,
19 Wright, Tree, & Duchaine, 2015). Faith believes her ability to process voices remains normal.
20 She completed the first eight tasks described below in April 2015 when she was 52-years-
21 old, and did a final task (three-alternative expression test) in February 2016 when she was 53.
22
23
24
25
26
27
28
29

30 2.2 Faith's lesion and its overlap with voice-selective activations in normal participants

31 2.2.1 Anatomical scan

32
33 Faith was scanned on a 3.0-T Phillips MR scanner (Philips Medical Systems, WA, USA)
34 with a SENSE (SENSitivity Encoding) 32-channel head coil. An anatomical volume was
35 acquired using a high-resolution 3D magnetization-prepared rapid gradient-echo sequence
36 (220 slices, field of view = 240 mm, acquisition matrix = 256×256 , voxel size = $1 \times 0.94 \times$
37 0.94 mm). This scan was skull stripped and then warped to Talairach space. The high-
38 resolution MR images of Faith's brain (Figure 1A) show a lesion extending from the
39 fusiform to the superior part of temporal lobe in the right hemisphere, encompassing a large
40 part of her posterior superior temporal sulcus (pSTS). The estimated lesion size on the axial,
41 coronal, and sagittal axes is 43 mm, 37 mm, and 37 mm, respectively.
42
43
44
45
46
47
48
49

50 2.2.2 Peak coordinates from five papers

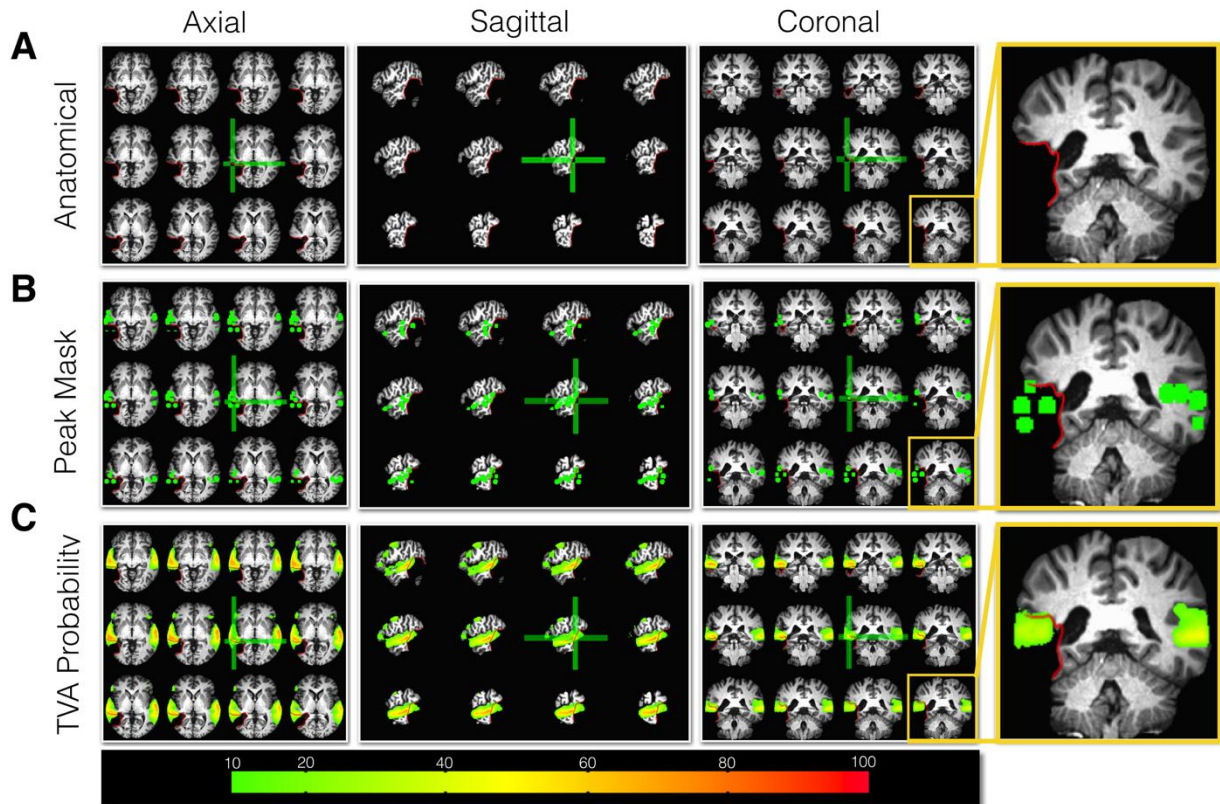
51
52 To demonstrate that Faith's lesion overlaps with voice-selective responses in the
53 literature, peak voxels implicated in voice processing from five papers (Belin et al., 2000;
54 Deen et al., 2015; von Kriegstein & Giraud, 2004; Watson, Latinus, Noguchi, et al., 2014;
55 Watson, Latinus, Charest, et al., 2014) are displayed on Faith's brain. Coordinates for 29
56 peak voxels in the temporal lobe that were listed in the tables in the five papers were
57
58
59
60
61
62
63
64
65

1
2
3
4 extracted manually and plotted on the standard brain in Talairach space with a 5mm radius
5 sphere centered at the peak voxel (Supplementary Table 1). The coordinates map was then
6 converted to a brain mask and overlaid on Faith's brain. Figure 1B shows the overlap of the
7 peak voxels with Faith's lesion.
8
9

10 2.2.3 Overlap with TVA probabilistic map

11 As another approach to determining which voice-selective regions may have been
12 affected by Faith's resection, we compared her lesion to a probability map of the temporal
13 voice areas (TVA) downloaded from <http://vnl.psy.gla.ac.uk/resources.php>. [Belin and](#)
14 [colleagues created](#) the map based on individually-thresholded data from 152 participants.
15 Each individual's T-map showing voice-selective voxels (voices > non-voice sounds) was
16 corrected for multiple comparisons based on the spatial extent at $q < 0.05$ (Chumbley &
17 Friston, 2009). They then applied a Gamma-Gaussian mixture model to separate the null
18 voxels from the active voxels (Gorgolewski, Storkey, Bastin, & Pernet, 2012). Data from
19 each participant was transformed into MNI space, binarized, summed, and normalized to 100
20 to create the group probability map. We converted this map from MNI space to Talairach
21 space so it could be overlaid on Faith's anatomical scan. Voxels that were voice-selective in
22 10% or more of the participants are displayed in Figure 1C.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 1. Faith’s anatomical scan, the peak coordinates mask, and the TVA probabilistic map.



Panel A displays Faith’s lesion on the axial (left), coronal (middle), and sagittal (right) planes with the spacing of one slice (around 1 mm). The right hemisphere lesion extended from the fusiform to the superior part of temporal lobe, encompassing a large part of her posterior superior temporal sulcus (pSTS). The estimated lesion size on the 3D axial, coronal, and sagittal axis is 43x37x37 (mm).

Panel B shows the peak coordinates mask of voices selective activations in the temporal lobe from the five studies. These peak coordinates mask is overlaid on Faith’s brain. Each dot is a sphere centered at the peak voxel with a radius of 5 mm. These slices show that Faith’s lesion has a large overlap with the voice selective activations.

Panel C shows a probabilistic map of temporal voice areas (TVA) overlaid on Faith’s brain. The TVA probabilistic map was downloaded from <http://vnl.psy.gla.ac.uk/resources.php> and normalized to Talairach space. For a voxel to be shown on this map, it had to be voice-selective in at least 10% of the participants. The color bar shows the scale for the TVA-probability map. Like the peak coordinates mask in Panel B, Faith’s lesion overlaps with the TVA probabilistic map.

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 minutes and contains one run in total. It contains 40 eight-seconds blocks of sounds (16 bit, mono, 22050 Hz sampling rate). Half of the blocks consists of vocal sounds (speech and nonspeech), and the other half consists of nonvocal sounds (industrial sounds, environmental sounds, and a few animal vocalizations). All sounds have been normalized and a 1kHz 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 = 2000ms, 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, Lee, Kay, & Tank, 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 anatomical data and transformed to a standard space in the Talairach template (Talairach & Tournoux, 1988). Each volume was blurred with a 4-mm FWHM Gaussian kernel. Time series of each run were scaled by the mean of the baseline before passing into the

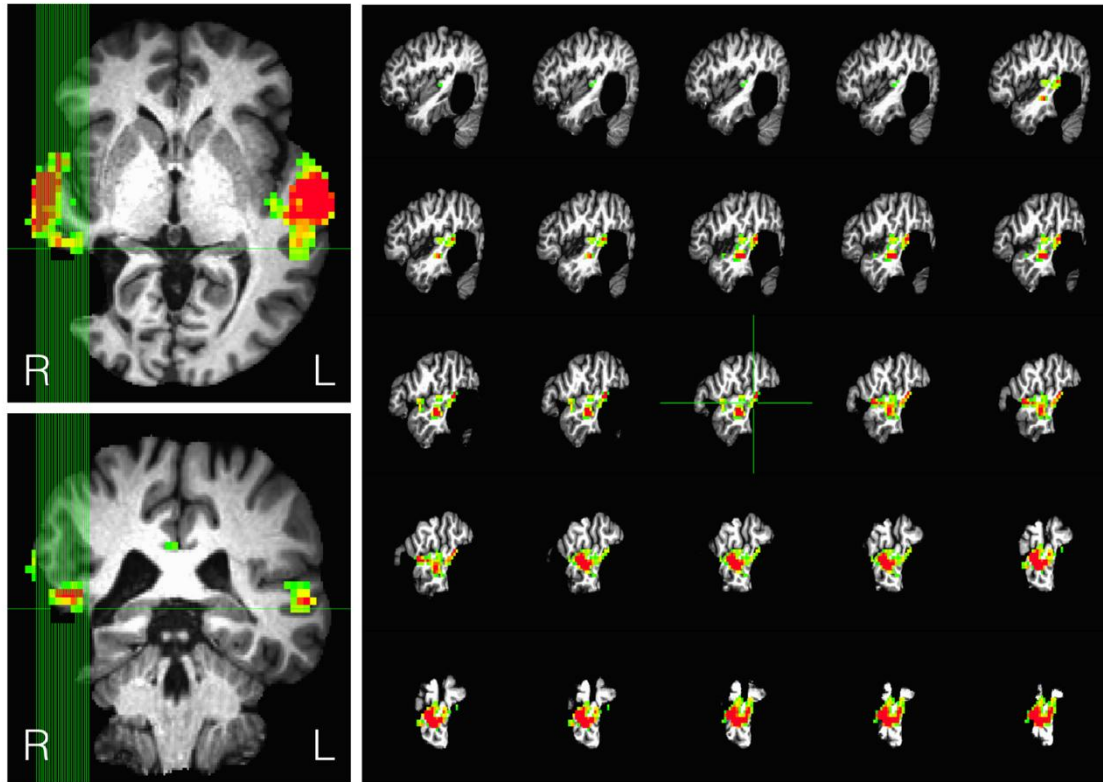
1
2
3
4 deconvolution analysis. Detrending and motion correction were carried out by including
5 trends and head motion as regressors in the regression model. Repetition times with
6 excessive motion (>0.3 mm) were removed. A general linear model procedure was used for
7 ROI analysis. Voice-selective regions were identified with a vocal $>$ non-vocal contrast.
8

9 10 11 12 2.3.4 Results

13 Two large clusters were localized at $p < 0.0001$ uncorrected on both temporal lobes
14 (Figure 2). Although Faith's lesion impacted part of her right temporal lobe, a cluster of 484
15 voxels was still found on the right side (peak: 62, -10, 2). The other cluster was in the left
16 temporal lobe and consisted of 461 voxels (peak: -64, -22, 5). Peak voxels in both
17 hemispheres were on the superior temporal gyrus (STG) in the vicinity of the peak
18 coordinates reported in previous experiments (Belin et al., 2000; Deen et al., 2015; von
19 Kriegstein & Giraud, 2004; Watson, Latinus, Noguchi, et al., 2014; Watson, Latinus,
20 Charest, et al., 2014). These voice localizer results indicate that, despite Faith's extensive
21 right hemisphere lesion, her remaining right STS still shows typical voice-selectivity.
22
23
24
25
26
27
28
29

30
31 To summarize Faith's imaging results, Faith's lesion disrupted her right pSTS but left
32 more anterior sections of STS intact. The lesioned part of Faith's right pSTS overlapped with
33 voice-selective activations in normal participants but the unlesioned part of Faith's STS
34 showed a typical voice-selective response.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 **Figure 2. Voice-selective areas in Faith.**
5



33 This figure shows voxels in Faith that responded significantly more strongly to vocal sounds than to non-vocal
34 sounds ($p < 0.0001$) in a standard functional TVA localizer (Belin et al., 2000).
35

36 2.4 Behavioral Testing

37 2.4.1 Control participants

38
39
40
41 Three groups of participants provided control data for the behavioral tasks. All were
42 compensated with course credit or reimbursement.
43

44 The first group of controls (Vancouver group) were a subset of the 73 controls aged 19-
45 70 years old reported in Liu et al. (2014). We selected 19 participants to create an age and
46 sex-matched group for Faith (age range 36-70; mean = 51.7, S.D. = 11.9). All participants
47 had no history of neurological or psychiatric diseases, and no visual or auditory complaints.
48 Participants were required to be native English speakers that had lived in North America for
49 five years or more.
50

51 The second group of controls (London group) were drawn from Garrido et al. (2009).
52 This group consisted of eight women between 46 and 64 years old (mean = 56.6; S.D. = 5.5).
53 They averaged 15.9 (S.D. = 2.0) years of education. All were native British English speakers
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 and reported normal hearing. One control (control 6) was taking fluoxetine, as well as
5 ropinirole for treatment of restless leg syndrome. The other controls reported no neurological
6 or psychiatric history.
7
8

9
10 The third control group (Dartmouth group) consisted of 18 undergraduate students tested
11 at Dartmouth College. Their ages ranged from 18 to 22 (mean = 18.94; S.D. = 1.26), and
12 eight were female. All reported normal hearing, and none reported neurological disorders.
13
14

15 16 17 2.4.2 General Material and Procedure

18
19 Faith wore headphones for all of the behavioral tasks. She did the voice identity
20 discrimination task and the three-alternative vocal expressions task on a 13-inch
21 MacBookPro and did the other tests were executed on a Dell laptop.
22
23

24 Faith's behavioral scores were compared to the controls' results using the modified t-test
25 for single case studies developed by Crawford & Howell (1998). Differences between Faith
26 and the controls were considered significant when the one-tailed probability was equal to or
27 below 0.05 because our expectation was that Faith's performance on the voice processing
28 tests would either be normal or impaired but not superior to the controls.
29
30
31
32
33

34 35 2.4.3 Voice Identity Discrimination

36 37 2.4.3.1 Material and Procedure

38
39 The auditory stimuli were created from 20 male and 20 female volunteers between the
40 ages of 20–31 (Liu et al., 2014). Each stimulus was used only once as a target or as a
41 distractor. Controls wore headphones and were tested with an IBM Lenovo laptop running
42 SuperLab software.
43
44

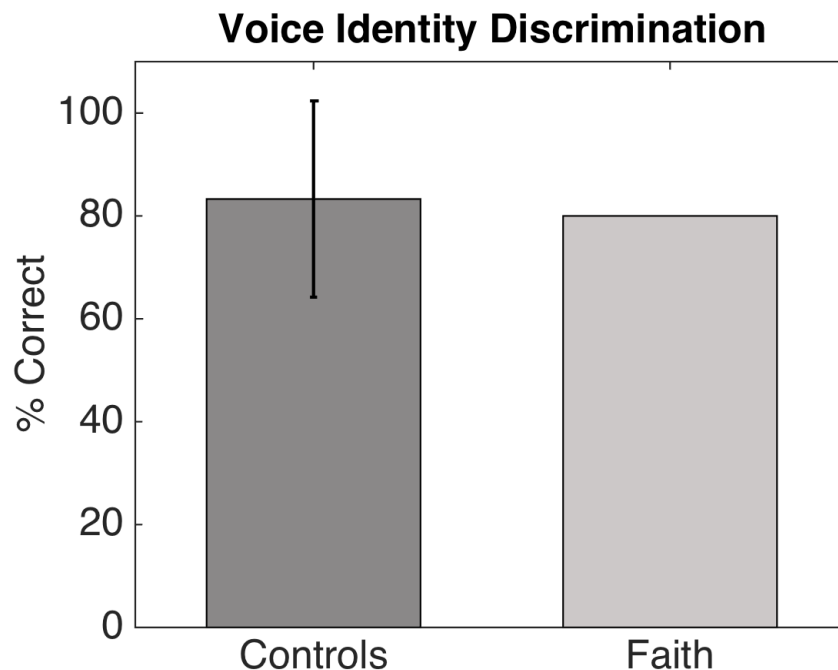
45
46 Each trial of the task consisted of a target voice and two choice voices. Audio stimuli
47 consisted of two different texts that the volunteers read. For the initial target voice, the
48 subjects read the phrase: "This is by far one of the most amazing books I have ever read, it
49 tells the story of a Colombian family across generations." One choice voice was the target
50 voice while the other was a distractor. Both choice voices read the phrase: "After a hearty
51 breakfast, we decided to go for a walk on the beach. It was a lovely morning with the crisp
52 smell of the ocean in the air." Volunteers were asked to read both texts at the same speed. All
53 recordings were 10 s in duration.
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 Participants attempted to select the choice voice that matched the target voice. No
5 feedback on performance was provided. In each trial, the participant heard the target voice
6 first. After a 1.5-s pause, a ring tone sounded for 875 ms, which served as an auditory mask
7 and separated the target from the test choices. Next, the participant heard the choice voices,
8 sequentially. Choice voices were preceded by 1 s of silence and were followed by 1.5 s of
9 silence. The participant was asked to press the number button “1” or “2” to indicate which
10 choice voice matched the target voice. The 40 trials were divided into two equal blocks; male
11 voices were presented in one block and females in the other.
12
13
14
15
16
17
18

19 2.4.3.2 Results

20 Faith’s accuracy in the voice discrimination task was 80.0% (Figure 3). The average
21 performance of the controls was 83.3% (S.D. = 9.5%). Faith’s performance was comparable
22 to the controls’ mean ($p = 0.37 > 0.05$), and her z score was -0.35.
23
24
25
26
27

28 **Figure 3.** Faith’s performance on voice identity discrimination task.
29



55 Faith’s accuracy on the voice identity discrimination task compared with the Vancouver controls. The error bar
56 shows ± 2 standard deviations (S.D.).
57
58
59
60
61
62
63
64
65

2.4.4 Voice Identity Recognition

2.4.4.1 Learning six speakers

2.4.4.1.1 Material and procedure

At the start of the test, participants learned the voices of six unfamiliar young female speakers (Garrido et al., 2009). All voices were native British English speakers and had similar accents. Speakers read sentences with three key words taken from the BKB Sentence List (Bench, Kowal, & Bamford, 1979). All samples were recorded in an anechoic chamber using Cool Edit 96 (<http://www.syntrillium.com>) and were normalized for peak amplitude using the program PRAAT (Boersma & Weenink, 2005).

Participants attempted to learn name-voice pairings, and they were told that the names would also be necessary for a later task. They were first presented with the name of a speaker on the screen, each of which started with a different letter from A to F, and then they heard a sentence said by that speaker. After that, they heard a number of sentences and for each one, they decided if it was said by the same speaker or not. Half the sentences were said by the target speaker, while the other half were said by one of the other five speakers. This procedure was repeated for each of the six speakers.

In the practice block, participants were presented with two test sentences per speaker. In the first test block, six test sentences followed the sentence presenting each speaker's voice. In the second and third test blocks, there were ten test sentences. No sentences were repeated.

2.4.4.1.2 Results

Over the three learning blocks, the control participants' mean accuracy was 75.2% (S.D. = 2.9). Faith's overall performance was 75.6%, which corresponds to a z-score of 0.13. (Figure 4A). Taking a closer look at each learning block, controls responded correctly on 71.5% (S.D. = 4.1) of trials in the first block, 74.4% (S.D. = 5.3) in the second block, and 78.1% (S.D. = 5.5) in the third block. Faith scored 77.8%, 63.3% and 86.7% in these three blocks respectively. The z-score of the three blocks was 1.52, -2.09, and 1.57. Only the result of the second block was slightly lower than control performance ($t(7) = -1.98$, $p = 0.04$).

2.4.4.2 Naming the six speakers

2.4.4.2.1 Material and Procedure

1
2
3
4 Immediately after the learning task finished, participants were presented with 60 new
5 sentences (6 speakers x 10 sentences per speaker) in a random order and were asked to select
6 the name associated with that speaker in the previous task. The six names were shown on the
7 computer screen. Feedback (one beep) was given for incorrect responses.
8
9

10 11 12 2.4.4.2.2 Results

13 With six choices, chance performance on this task was 16.7%. The mean percent correct
14 for the London controls was 35.6% (S.D. = 10.9). Faith correctly identified 38.3% of the test
15 items, which corresponds to a z-score of 0.25 (Figure 4B).
16
17
18
19

20 21 2.4.4.3 Old-new discrimination with the six speakers

22 23 2.4.4.3.1 Material and Procedure

24 Immediately after the name identification task, participants' voice memory was tested
25 with an old-new discrimination task. For each sentence, participants decided whether it was
26 said by one of the six speakers used in the two previous tasks or a new speaker. New
27 speakers were young females with accents similar to the six target speakers. New recordings
28 were done for all speakers (targets and distracters) in a silent room. Peak amplitude for all
29 stimuli was matched using PRAAT. The test trials included six sentences by each of the old
30 speakers (36 'old' trials) and four sentences said by each of the nine new speakers (36 'new'
31 trials).
32
33
34
35
36
37
38

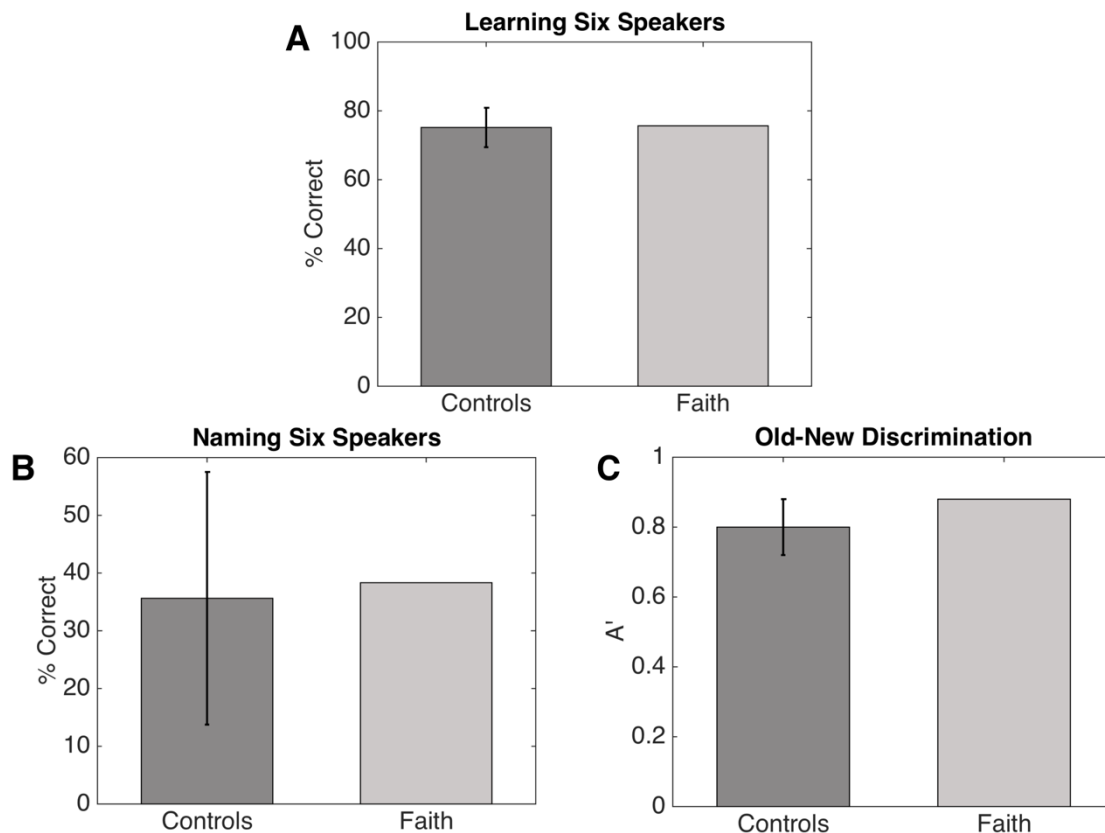
39 40 2.4.4.3.2 Results

41 A' for the controls was 0.80 (S.D. = 0.04). Faith's A' of 0.88 was better than all the
42 controls ($z = 2.0$) (Figure 4C).
43
44
45

46 In summary, the three voice identity recognition tasks demonstrated Faith had normal
47 discrimination, recognition and familiarity for voice identity.
48
49
50
51
52
53
54

55 **Figure 4.** Faith's performance on voice identity recognition tasks
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



Performance of Faith and the London control group on the three voice tasks involving the six speakers. (A) Faith's accuracy compared with the control group on the three learning blocks. Scores from the three blocks were combined to calculate overall accuracy. (B) Faith's accuracy compared with the controls on the Naming the Six Speakers task. (C) Comparison of Faith and the controls on the Old-New Discrimination task. All error bars display the ± 2 standard deviations (S.D.).

2.4.5 Vocal Sex Perception

2.4.5.1 Material and Procedure

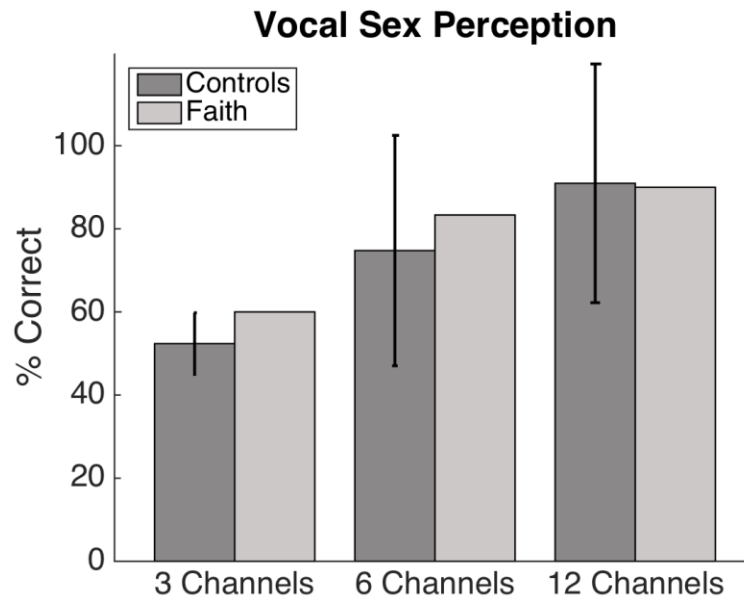
Twenty-six native English speakers read sentences aloud from the BKB Sentence List (stimuli were collected using a microphone and the program Cool Edit 96). Different speakers read different sets of sentences. Sentences were matched for peak amplitude and noise-vocoded using PRAAT to increase the difficulty of the task to avoid ceiling effects. Three, six, and twelve frequency channels were used.

1
2
3
4 On each trial, participants listened to one sentence and were asked to decide whether the
5 speaker was a man or a woman. Half the sentences were spoken by males and half by
6 females. Thirty sentences for each frequency channel were presented, making a total of 90
7 trials.
8
9

10 2.4.5.2 Results

11
12 Seven out of the eight London control participants performed this task. For the three
13 noise-vocoded levels (3,6,12), participants correctly perceived the sex of the speaker on 52.4%
14 (S.D. = 3.7), 74.8% (S.D. = 13.9) and 91.0% (S.D. = 14.4) of trials. Faith scored 60.0%, 83.3%
15 and 90.0% (z = 2.05, 0.62, -0.07), indicating she categorizes vocal sex normally (Figure 5).
16
17
18
19
20
21

22 **Figure 5.** Faith's performance on vocal sex perception task



45
46 Faith's accuracy on the Vocal Sex Perception task compared with seven London controls. Error bars show ± 2
47 standard deviations (S.D.).
48
49

50 2.4.6 Vocal Expression Perception

51 2.4.6.1 Recognition of vocal expression of emotion

52 2.4.6.1.1 Material and Procedure

53
54 Stimuli were 90 non-verbal sounds expressing one of the following emotions:
55 achievement/triumph, amusement, anger, disgust, fear, pleasure, relief, sadness and surprise
56
57
58
59
60
61
62
63
64
65

1
2
3
4 (10 stimuli for each emotion) (Sauter, 2006; Sauter & Scott, 2007). After each stimulus, the
5 list of nine possibilities was presented on the screen and participants selected the one that
6 best described the expression of the voice.
7
8

9 10 2.4.6.1.2 Results

11 London controls selected the correct adjective on 82.2% (S.D. = 6.1) of trials. Faith's
12 accuracy was 73.3% ($p = 0.16 > 0.05$), which placed her 1.47 standard deviations below the
13 mean.
14
15
16
17

18 19 2.4.6.2 Recognition of vocal expression via paralinguistic cues in speech

20 21 2.4.6.2.1 Material and Procedure

22 Stimuli consisted of emotionally inflected spoken three-digit numbers. Like the previous
23 task, there were ten stimuli for each of the nine emotion categories. There were another ten
24 stimuli expressing contentment, to make a total of 100 trials in this task. All ten adjectives
25 were presented on the computer screen in each trial, and participants were also asked to
26 select the adjective that best described the emotion in the voice.
27
28
29
30

31 32 2.4.6.2.2 Results

33 The controls correctly identified the emotions in 72.1% (S.D. = 5.5) of trials. Faith's
34 score of 66.0% ($z = -1.12$) did not differ significantly ($p = 0.10 > 0.05$) from the controls'
35 mean score.
36
37
38
39
40

41 42 2.4.6.3 Three-alternative choices vocal expression of emotion

43 44 2.4.6.3.1 Material and Procedure

45 Stimuli consisted for speakers saying "Ah" so that it conveyed eight emotions (anger,
46 disgust, fear, neutral, pain, pleasure, sadness, and surprise) from the Montreal Affective
47 Voices (MAV) set (happiness was excluded due to ceiling effects). Each expression
48 contained ten clips of audios recorded by five actors and five actresses (Belin, Fillion-
49 Bilodeau, & Gosselin, 2008). Each trial consisted of three voices said by three different
50 actors and actresses. Two voices expressed the same emotion while the other one expressed a
51 different emotion. Participants were asked to pick the odd one out by pressing "1", "2", or
52 "3". There were 72 trials, with a 2-s interstimulus interval and a 2.5-s intertrial interval. The
53
54
55
56
57
58
59
60
61
62
63
64
65

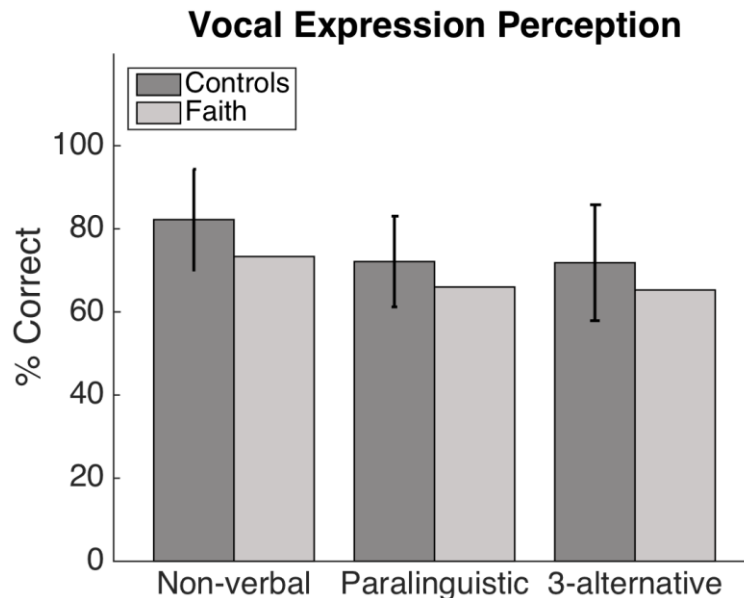
1
2
3
4 task was carried out using the online testing platform Testable (<http://www.testable.org>).
5
6 Participants wore headphones.

7 8 2.4.6.3.2 Results

9
10 The controls' mean accuracy was 71.8% (S.D. = 7.0%). Faith responded correctly on
11 65.3% of trials ($z = -0.94$), which was comparable with the Dartmouth control groups'
12 performance ($p = 0.19 > 0.05$) (Figure 6).
13
14
15
16

17 To exclude the possibility that Faith was impaired with particular vocal expressions, we
18 analyzed each expression category separately in all three vocal expression tasks
19 (supplementary material). None of these comparisons revealed a significant difference
20 between Faith and the controls, so our results provide no evidence that Faith has expression-
21 selective impairments (Supplementary Figure 1).
22
23
24
25
26

27
28 **Figure 6.** Faith's performance on vocal emotion perception task
29

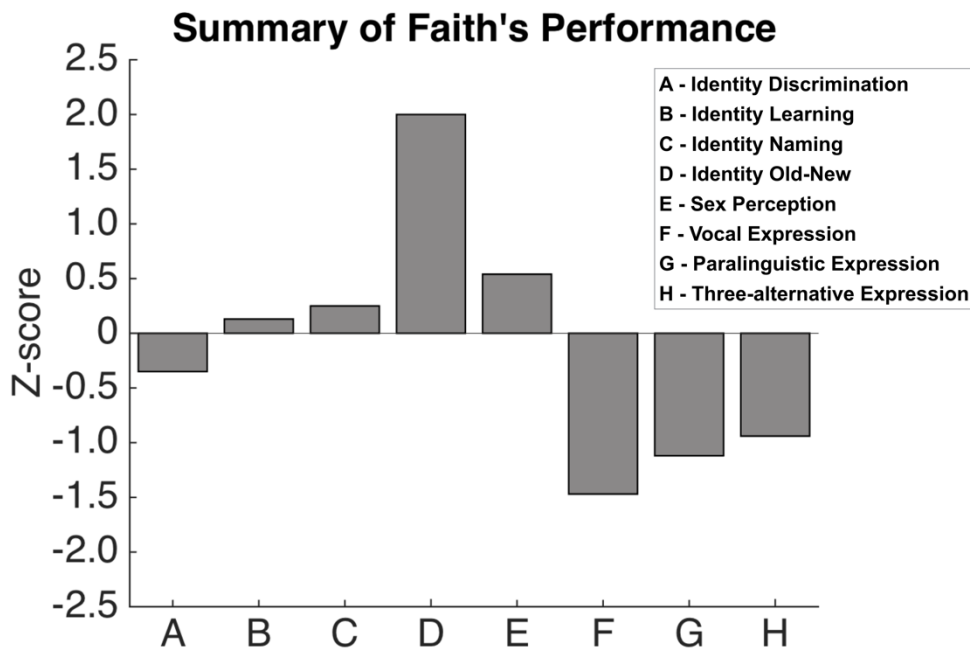


51
52 Faith's accuracy on the Vocal Expression Perception task. Her performance is compared with the London
53 controls in the non-verbal and paralinguistic tasks and with 18 Dartmouth controls on the three-alternative task.
54 Error bars display ± 2 standard deviations (S.D.).
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 3. Discussion
5

6 We tested voice processing in Faith, a woman whose right pSTS was resected after
7 surgery to remove a tumor. Eight voice tests measuring her abilities with identity
8 discrimination, familiarity and recognition, sex perception, and expression perception in
9 voices showed that her performance was comparable to that of control participants. Figure 7
10 presents a summary of Faith's behavioral performance relative to the controls. Her normal
11 performance on all the vocal tasks demonstrates that voice processing ability can remain
12 intact after a lesion to right pSTS.
13
14
15
16
17
18
19

20
21 **Figure 7.** Summary of Faith's Performance
22



48 Faith's z-scores on the behavioral voice tasks. Faith's performance was compared with the Vancouver controls
49 on task A, with the London controls on task B to G, and with the Dartmouth controls on task H. The three
50 blocks in the Learning the Six Speakers task were combined to calculate a single z-score, and the scores for
51 each of the three channels in the sex perception test were also combined. The z-scores ranged from -1.5 to 2,
52 and none of Faith's scores were significantly different from the controls.
53
54
55
56

57 Our investigation was motivated by fMRI results demonstrating increased signal in the
58 pSTS in response to voice stimuli (Andics et al., 2010; Belin et al., 2000; Deen et al., 2015;
59
60
61
62
63
64
65

1
2
3
4 von Kriegstein & Giraud, 2004; Watson, Latinus, Noguchi, et al., 2014; Watson, Latinus,
5 Charest, et al., 2014). We considered two explanations for these activations. According to the
6 first account, pSTS plays a role in fundamental voice processing by representing vocal
7 information and processing it to derive information about identity, sex, expression, and other
8 characteristics. The second possibility is that pSTS does not contribute to basic voice
9 processing but instead is involved in higher level voice processing (Belin et al., 2011;
10 Campanella & Belin, 2007; Thurman et al., 2016; Yovel & O’Toole, 2016). Because Faith’s
11 voice perception is normal, her findings indicate that the right pSTS is not necessary for
12 fundamental voice processing, and thus are not compatible with the first account.
13
14

15
16
17 Faith’s results indicate pSTS activations to voices reflect higher-level processes, and
18 several studies indicate that this region plays a role in integrating voice information with
19 other kinds of social information (Belin et al., 2011; Campanella & Belin, 2007; Yovel &
20 O’Toole, 2016). This evidence includes a meta-analysis that found that tasks involving
21 integration led to responses that clustered in pSTS (Hein & Knight, 2008) as well as fMRI
22 and MEG studies assessing the response to single (audio or visual) and multi-channel
23 (audiovisual) social stimuli (Deen et al., 2015; Watson, Latinus, Charest, et al., 2014). The
24 pSTS also shows a cross-modal fMRI adaptation effect when facial expressions were similar
25 to the preceding vocal expression (Watson, Latinus, Noguchi, et al., 2014). Also consistent
26 with an integrative role for pSTS are findings demonstrating that pSTS voxels selective for
27 visual mouth movements also respond strongly and selectively to voices whereas voxels
28 selective for moving eyes did not respond to voices (Zhu & Beauchamp, 2017).
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 What brain areas supported Faith’s normal voice processing? Previous studies have
44 demonstrated voice-selective regions in middle and anterior STS bilaterally, referred to as the
45 temporal voice areas (TVAs) (Belin et al., 2011; Latinus, Crabbe, & Belin, 2011; Yovel &
46 Belin, 2013). To see whether Faith’s TVAs are preserved, we carried out a standard fMRI
47 voice localizer in which blocks of voices and non-vocal sounds were presented (Belin et al.,
48 2000) (Figure 2). Faith showed a typical voice-selective response in regions anterior to her
49 lesion in both the right and left STS. Hence it is likely these more anterior STS regions as
50 well as frontal regions (Andics et al., 2010; Watson, Latinus, Noguchi, et al., 2014; Watson,
51 Latinus, Charest, et al., 2014) allowed her to perform normally in the lab and in daily life.
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 Furthermore, these results suggest that an intact right pSTS is not necessary to support voice-
5 selective activity in these anterior regions.
6

7
8 Faith's range of behavioural impairments make it difficult to provide evidence that the
9 pSTS integrates voice information with other social representations. Doing so would require
10 demonstrating that disruption of right pSTS impairs the ability to integrate social information
11 while not impairing the perceptual processes feeding an integrating mechanism. Because
12 Faith's face perception (Susilo et al., 2015) and body perception (unpublished data) are
13 severely disrupted by the lesions she has suffered, this makes it difficult to determine if there
14 is integration of voice with face or body information. Testing a patient with a lesion
15 selectively affecting pSTS or stimulating pSTS in neurologically intact participants may be
16 effective ways to address this issue. However, this may be more useful in examining
17 integration in identity perception rather than in expression perception, given Pitcher's (2014)
18 demonstration that transcranial magnetic stimulation to right pSTS disrupted facial
19 expression recognition.
20
21
22
23
24
25
26
27
28
29

30 **In conclusion, Faith's consistently normal performance on a range of voice processing**
31 **studies indicates that the right pSTS is not necessary for fundamental processing of voices.**
32

33 Her results are consistent with suggestions that right pSTS's responsiveness to voices reflects
34 higher-level voice processing effects, such as the integration of voices with other social
35 representations.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Acknowledgements

We are extremely grateful for Faith's participation in this project.

This project was supported by a Hitchcock Foundation grant to B.D. and CIHR grant MOP-102567, Canada Research Chair, and the Marianne Koerner Chair in Brain Diseases to J.B.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

- Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., & Vidnyánszky, Z. (2010). Neural mechanisms for voice recognition. *NeuroImage*, *52*(4), 1528–1540.
<https://doi.org/10.1016/j.neuroimage.2010.05.048>
- Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding Voice Perception. *British Journal of Psychology*, *102*(4), 711–725.
<https://doi.org/10.1111/j.2044-8295.2011.02041.x>
- Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. <https://doi.org/10.3758/BRM.40.2.531>
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26. [https://doi.org/10.1016/S0926-6410\(01\)00084-2](https://doi.org/10.1016/S0926-6410(01)00084-2)
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. <https://doi.org/10.1038/35002078>
- Bench, J., Kowal, Å., & Bamford, J. (1979). The Bkb (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology*, *13*(3), 108–112.
<https://doi.org/10.3109/03005367909078884>
- Bestelmeyer, P. E. G., Belin, P., & Grosbras, M.-H. (2011). Right temporal TMS impairs voice detection. *Current Biology*, *21*(20), R838–R839.
<https://doi.org/10.1016/j.cub.2011.08.046>
- Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer. [Computer program].
<<http://www.praat.org/>>.

- 1
2
3
4 Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in*
5
6 *Cognitive Sciences*, 11(12), 535–543. <https://doi.org/10.1016/j.tics.2007.10.001>
7
8
9 Chumbley, J. R., & Friston, K. J. (2009). False discovery rate revisited: FDR and topological
10
11 inference using Gaussian random fields. *NeuroImage*, 44(1), 62–70.
12
13 <https://doi.org/10.1016/j.neuroimage.2008.05.021>
14
15
16 Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic
17
18 resonance neuroimages. *Computers and Biomedical Research, an International Journal*,
19
20 29(3), 162–173.
21
22
23 Crawford, J. R., & Howell, D. C. (1998). Comparing an Individual's Test Score Against Norms
24
25 Derived from Small Samples. *The Clinical Neuropsychologist*, 12(4), 482–486.
26
27 <https://doi.org/10.1076/clin.12.4.482.7241>
28
29
30 Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional Organization of Social
31
32 Perception and Cognition in the Superior Temporal Sulcus. *Cerebral Cortex*, 25(11),
33
34 4596–4609. <https://doi.org/10.1093/cercor/bhv111>
35
36
37 Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing.
38
39 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431), 459–
40
41 473. <https://doi.org/10.1098/rstb.2002.1218>
42
43
44 Gainotti, G. (2013). Laterality effects in normal subjects' recognition of familiar faces, voices
45
46 and names. Perceptual and representational components. *Neuropsychologia*, 51(7), 1151–
47
48 1160. <https://doi.org/10.1016/j.neuropsychologia.2013.03.009>
49
50
51 Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., ... Duchaine, B.
52
53 (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition.
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 *Neuropsychologia*, 47(1), 123–131.

5
6
7 <https://doi.org/10.1016/j.neuropsychologia.2008.08.003>

8
9 Gorgolewski, K., Storkey, A. J., Bastin, M. E., & Pernet, C. R. (2012). Adaptive thresholding for
10 reliable topological inference in single subject fMRI analysis. *Frontiers in Human*
11 *Neuroscience*, 6, 245. <https://doi.org/10.3389/fnhum.2012.00245>

12
13
14
15
16 Hagan, C. C., Woods, W., Johnson, S., Calder, A. J., Green, G. G. R., & Young, A. W. (2009).
17 MEG demonstrates a supra-additive response to facial and vocal emotion in the right
18 superior temporal sulcus. *Proceedings of the National Academy of Sciences*, 106(47),
19 20010–20015. <https://doi.org/10.1073/pnas.0905792106>

20
21
22
23
24
25
26 Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for
27 face perception. *Trends in Cognitive Sciences*, 4(6), 223–233.
28
29 [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0)

30
31
32
33
34
35
36
37 Hein, G., & Knight, R. T. (2008). Superior Temporal Sulcus—It’s My Area: Or Is It? *Journal of*
38 *Cognitive Neuroscience*, 20(12), 2125–2136. <https://doi.org/10.1162/jocn.2008.20148>

39
40
41
42
43
44
45
46
47 Latinus, M., Crabbe, F., & Belin, P. (2011). Learning-Induced Changes in the Cerebral
48 Processing of Voice Identity. *Cerebral Cortex*, 21(12), 2820–2828.
49 <https://doi.org/10.1093/cercor/bhr077>

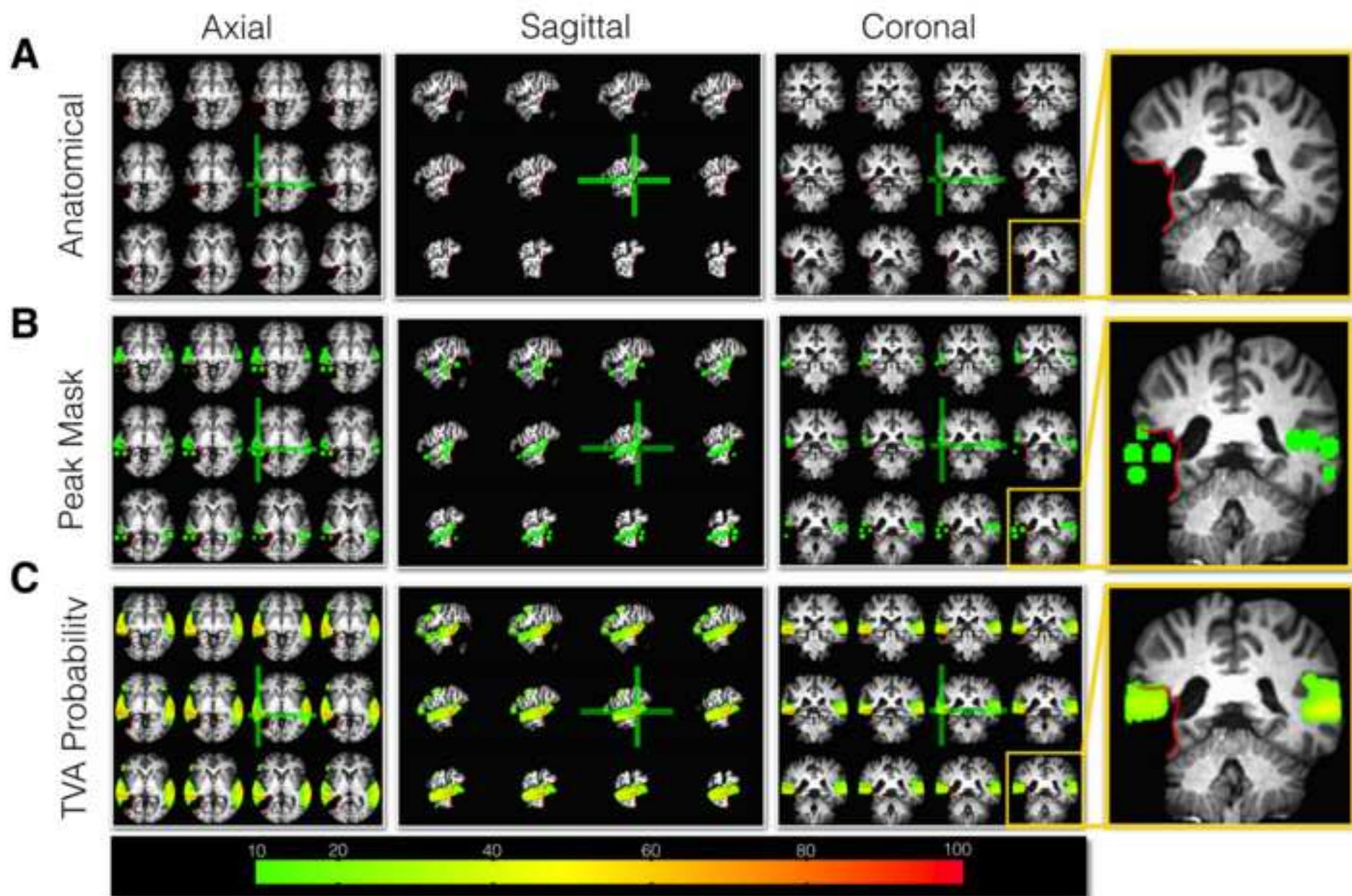
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65 Liu, R. R., Pancaroglu, R., Hills, C. S., Duchaine, B., & Barton, J. J. S. (2014). Voice
66 Recognition in Face-Blind Patients. *Cerebral Cortex*, bhu240.
67 <https://doi.org/10.1093/cercor/bhu240>

68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

- 1
2
3
4 Ojemann, J. G., Akbudak, E., Snyder, A. Z., McKinstry, R. C., Raichle, M. E., & Conturo, T. E.
5
6 (1997). Anatomic Localization and Quantitative Analysis of Gradient Refocused Echo-
7
8 Planar fMRI Susceptibility Artifacts. *NeuroImage*, 6(3), 156–167.
9
10 <https://doi.org/10.1006/nimg.1997.0289>
11
12
13
14 Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and fMRI Reveal Dissociable
15
16 Cortical Pathways for Dynamic and Static Face Perception. *Current Biology*, 24(17),
17
18 2066–2070. <https://doi.org/10.1016/j.cub.2014.07.060>
19
20
21 Sauter, D. A. (2006). *An investigation into vocal expressions of emotions: the roles of valence,*
22
23 *culture, and acoustic factors.* (Doctoral). University of London. Retrieved from
24
25 <http://discovery.ucl.ac.uk/1445045/>
26
27
28 Sauter, D. A., & Scott, S. K. (2007). More than one kind of happiness: Can we recognize vocal
29
30 expressions of different positive states? *Motivation and Emotion*, 31(3), 192–199.
31
32 <https://doi.org/10.1007/s11031-007-9065-x>
33
34
35
36 Schall, S., Kiebel, S. J., Maess, B., & von Kriegstein, K. (2014). Voice Identity Recognition:
37
38 Functional Division of the Right STS and Its Behavioral Relevance. *Journal of Cognitive*
39
40 *Neuroscience*, 27(2), 280–291. https://doi.org/10.1162/jocn_a_00707
41
42
43 Susilo, T., Wright, V., Tree, J. J., & Duchaine, B. (2015). Acquired prosopagnosia without word
44
45 recognition deficits. *Cognitive Neuropsychology*, 32(6), 321–339.
46
47 <https://doi.org/10.1080/02643294.2015.1081882>
48
49
50 Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-*
51
52 *Dimensional proportional system: an approach to cerebral imaging.* New York: Thieme
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

- 1
2
3
4 Thurman, S. M., van Boxtel, J. J. A., Monti, M. M., Chiang, J. N., & Lu, H. (2016). Neural
5
6 adaptation in pSTS correlates with perceptual aftereffects to biological motion and with
7
8 autistic traits. *NeuroImage*, *136*, 149–161.
9
10 <https://doi.org/10.1016/j.neuroimage.2016.05.015>
11
12
13
14 von Kriegstein, K. V., & Giraud, A.-L. (2004). Distinct functional substrates along the right
15
16 superior temporal sulcus for the processing of voices. *NeuroImage*, *22*(2), 948–955.
17
18 <https://doi.org/10.1016/j.neuroimage.2004.02.020>
19
20
21 Watson, R., Latinus, M., Charest, I., Crabbe, F., & Belin, P. (2014). People-selectivity,
22
23 audiovisual integration and heteromodality in the superior temporal sulcus. *Cortex*, *50*,
24
25 125–136. <https://doi.org/10.1016/j.cortex.2013.07.011>
26
27
28
29 Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2014). Crossmodal
30
31 Adaptation in Right Posterior Superior Temporal Sulcus during Face–Voice Emotional
32
33 Integration. *The Journal of Neuroscience*, *34*(20), 6813–6821.
34
35 <https://doi.org/10.1523/JNEUROSCI.4478-13.2014>
36
37
38 Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends*
39
40 *in Cognitive Sciences*, *17*(6), 263–271. <https://doi.org/10.1016/j.tics.2013.04.004>
41
42
43 Yovel, G., & O’Toole, A. J. (2016). Recognizing People in Motion. *Trends in Cognitive*
44
45 *Sciences*, *0*(0). <https://doi.org/10.1016/j.tics.2016.02.005>
46
47
48 Zhu, L. L., & Beauchamp, M. S. (2017). Mouth and voice: A relationship between visual and
49
50 auditory preference in the human superior temporal sulcus. *Journal of Neuroscience*,
51
52 2914–16. <https://doi.org/10.1523/JNEUROSCI.2914-16.2017>
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure
[Click here to download high resolution image](#)



Supplementary Material

[Click here to download Supplementary Material: Supplementary Jiahui, Garrido, Liu, Susilo, Barton, & Duchaine_revised2.docx](#)