The Superior Temporal Sulcus Differentiates Communicative and Noncommunicative Auditory Signals

Sarah Shultz[1], Athena Vouloumanos[2], and Kevin Pelphrey[1]

Abstract

■ Processing the vocalizations of conspecifics is critical for adaptive social interaction. A species-specific voice-selective region has been identified in the right STS that responds more strongly to human vocal sounds compared with a variety of nonvocal sounds. However, the STS also activates in response to a wide range of signals used in communication, such as eye gaze, biological motion, and speech. These findings raise the possibility that the voice-selective region of the STS may be especially sensitive to vocal sounds that are communicative, rather than to all human vocal sounds. Using fMRI, we demonstrate that the voice-selective region of the STS responds more strongly to communicative vocal sounds (such as speech and laughter) compared with noncommunicative vocal sounds (such as coughing and sneezing). The implications of these results for understanding the role of the STS in voice processing and in disorders of social communication, such as autism spectrum disorder, are discussed. ■

INTRODUCTION

The human voice conveys important social information, including emotional state and speaker characteristics (Campanella & Belin, 2007; Belin, Fecteau, & Bedard, 2004). As such, the ability to recognize and process the communicative information in conspecific vocalizations is important for successful social interaction. In typical human adults, a region in the STS responds preferentially to human vocal sounds compared with naturally occurring nonhuman vocalizations, other nonvocal sounds, and well-matched acoustic controls (Leaver & Rauschecker, 2010; Fecteau, Armony, Joanette, & Belin, 2004; Kriegstein & Giraud, 2004; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Binder et al., 2000). Furthermore, this voice-selective region is hypoactive in response to vocal sounds in adults with autism spectrum disorder (ASD), a disorder characterized by deficits in language and social communication (Gervais et al., 2004). However, an alternative explanation may be that this region of cortex is especially sensitive to communicative signals rather than to vocal sounds per se. Although in much of the animal communication literature, any aspect of an organism's phenotype that influences the behavior of others (including, e.g., size, coloring, etc.; Maynard Smith & Harper, 2003) can be considered communicative, here we consider communicative signals to be voluntarily and flexibly produced with the intention of sharing information with another person (as in Tomasello, 2008; Grice, 1968). Because the vocal stimuli used in pre-

vious studies contained both communicative (e.g., speech, laughter) and noncommunicative (e.g., coughs) signals, it is unknown whether this region is indeed specialized for processing all human vocal sounds or whether this region is especially sensitive to a subset of conspecific vocalizations: vocal signals with communicative significance. This distinction has important implications for understanding the neural infrastructure of voice processing and the reported neural deficits in ASD. If this region of the STS is sensitive to vocal communicative signals, then this suggests that processing of the communicative aspect of vocal sounds is separable from the vocal quality of sounds. This would further suggest that the reported hypoactivation in ASD may reflect a difficulty in recognizing and extracting the communicative significance of vocal sounds rather than a deficit in voice processing per se. The current study seeks to clarify the role of the STS in processing vocal sounds by examining whether this voice-selective region is especially sensitive to communicative vocal sounds compared with noncommunicative vocal sounds.

Several lines of research suggest that the STS may be especially sensitive to communicative vocal sounds. The STS is engaged in a wide range of social tasks, including theory of mind and mentalizing (Saxe, 2006; Zilbovicius et al., 2006; Gallagher & Frith, 2003), biological motion perception (Puce & Perrett, 2003; Allison, Puce, & McCarthy, 2000), face perception (Haxby, Hoffman, & Gobbini, 2000), and speech processing (Vouloumanos, Kiehl, Werker, & Liddle, 2001; Price, 2000), suggesting that the STS supports a variety of social functions (Hein & Knight, 2008). What might account for the role of STS in such a wide range of

[1]Yale University, [2]New York University

social tasks? Redcay (2008) has argued that the STS is engaged by a process that is common to all these functional domains: interpreting the communicative significance of both auditory and visual inputs. This hypothesis was supported by several studies demonstrating that the STS is activated by stimuli that convey social communicative significance (Redcay, 2008), such as body movements, eyegaze, head orientation, lip reading, facial expressions, vocal sounds, and speech.

Furthermore, studies suggest activation in the STS increases as a function of the communicative significance of stimuli, with the STS consistently showing the greatest response to meaningful stimuli of communicative significance (Redcay, 2008). For instance, within the auditory domain, left STS activation increased in response to more communicative sounds: Activation was strongest to words, less strong to naturalistic sounds including animal vocalizations and instruments, and least to tones (Specht & Reul, 2003). Similarly, narratives elicited greater activation in the STS compared with sentences (Xu, Kemeny, Park, Frattali, & Braun, 2005), and the response to sentences was greater than to pseudoword sentences (Roder, Stock, Neville, Bien, & Rosler, 2002), suggesting a sensitivity to communicative content. Similar findings have been reported within the visual domain: Activity in the STS is greater in response to communicative signals such as eye gaze (which may direct the attention of another person toward a source or communicative affective information, flexibly and intentionally) and facial expressions of emotion compared with less communicative signals such as person identity, a phenotypic aspect that is not intentional (LaBar, Crupain, Voyvodic, & McCarthy, 2003; Hoffman & Haxby, 2000). Finally, the STS responds more strongly to goal-directed actions compared with simple biological motion and nonbiological motion (Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004; Pelphrey et al., 2003). Remarkably, the same physical stimulus elicits greater neural responses in the STS based on when the listener interprets it as speech rather than nonspeech (Möttönen et al., 2006; Dehaene-Lambertz et al., 2005). Similarly, the pSTS responds differentially to perceptually similar actions based on whether they are perceived as intentional (a character lifting their arm upward with their own volition) or not (a piston pushing the character's arm upward in the same motion; Morris, Pelphrey, & McCarthy, 2008). Together, these studies suggest that interpreting the communicative significance of stimuli may be a common process underlying the many functions of the STS.

In the current study, we investigated whether the voice-selective region of the STS may be especially sensitive to the communicative significance of vocal sounds by examining neural activity in response to eight sound categories that varied in terms of whether they are vocally produced and whether they typically serve a communicative function (see Table 1 for stimulus properties). We first identified a voice-selective region by contrasting the neural response to human vocal sounds (both communicative—adult-directed-speech, infant-directed speech, communicative

Table 1. Properties of Sounds Used in the Experiment

Sound	Communicative	Vocal
Infant-directed speech	√	√
Adult-directed speech	√	√
Communicative vocal nonspeech	√	√
Noncommunicative vocal nonspeech	X	√
Rhesus macaque calls	X	X
Clapping	X	X
Walking	X	X
Water	X	X

vocal nonspeech (e.g., laughter, sounds of agreement)—and noncommunicative—physiological vocalizations like sneezing) with nonvocal sounds (walking, clapping, rhesus monkey calls, and sounds of water). After identifying voice-selective regions, we next compared the response to each of the four human vocal sound categories for all voxels within this region. If this region responds nonspecifically to all human vocal sounds, then we would expect all human vocal conditions to elicit similar levels of activation. However, if this region responds selectively to human communicative vocal sounds, then we would expect to see increased activation in response to communicative vocal sounds (e.g., laughter) but not to noncommunicative vocal sounds (e.g., sneezing).

METHODS

Participants and Stimuli

Twenty healthy adults participated in the study (seven men, mean age = 22 years, all right-handed). All of the participants were fluent in English, and none were fluent in Japanese. All participants gave written, informed consent and the Yale Human Investigations Committee approved the protocol.

Eight types of auditory stimuli were presented: infant-directed speech, adult-directed speech, human communicative vocal nonspeech, human noncommunicative vocal nonspeech, sounds of walking, sounds of hand clapping, rhesus macaque vocalizations, and sounds of water. All sounds were sampled at 44,100 Hz, and equalized for mean intensity using PRAAT 5.1.07 (Boersma & Weenink, 2009). The sounds were concatenated into five 20-sec sound files per sound category, each consisting of 11–15 tokens separated by 600–1000 msec of silence. The selection and ordering of tokens comprising the 20-sec sound files was pseudorandom such that the same token was never played twice in a row.

Infant-directed speech consisted of 15 tokens of Japanese words spoken by three female native Japanese speakers. Words were spoken in infant-directed speech, with slightly higher pitch and exaggerated pitch contour, to

allow us to compare affective speech to speech produced in a neutral tone (see adult-directed speech). All words were spoken in Japanese to ensure that brain regions responsive to speech were not simply sensitive to semantic content.

Adult-directed speech consisted of the same 15 tokens of Japanese words used in the infant-directed speech condition spoken by the same three female native Japanese speakers. Words were spoken in a normal, neutral tone.

Human communicative vocal nonspeech consisted of 15 tokens produced by three women: agreement (3), disagreement (3), disgust (3), inquiry (3), and laughter (3). These vocalizations were chosen because they can carry affective and semantic meaning and are often produced with the intent of communicating this meaning to another listener.

Human noncommunicative vocal nonspeech consisted of 15 tokens produced by three women: coughs (3), throat clearings (3), yawns (4), hiccups (3), and sneezes (2). These sounds were selected because humans produce them primarily for physiological reasons.

Walking sounds consisted of 15 tokens of three women walking on two surfaces: tile (7) and wood (8).

Clapping sounds consisted of 15 tokens of three women clapping. Clapping sounds were initially selected to convey communicative intent (e.g., praise as in applause) through nonvocal means. However, recordings of clapping were not recognized as applause and were instead reported as sounding like "snapping twigs." As such, the clapping sounds were considered to be noncommunicative human sounds.

Rhesus macaque vocalizations consisted of 15 tokens produced by three free-ranging adult female rhesus macaques recorded in Cayo Santiago, Puerto Rico: grunts (2), coos (2), girneys (3), noisy screams (4), and arched screams (4). These calls differ from one another on valence and referential function (Hauser, 2000) and thus represent a wide range of rhesus vocalizations.

Water sounds consisted of 15 tokens downloaded from www.findsounds.com: running water (3), boiling water (3), water being poured (3), splashing water (3), and lake water (3).

Stimuli were presented in a block design with one sound category played per block. Each of the eight sound categories was played five times for a total of 40 blocks. Blocks were separated by a 12-sec intertrial interval and presented in pseudorandom order such that the same sound category was never repeated more than twice in a row. Participants were instructed to listen to the stimuli at all times.

Sound Properties

Behavioral Ratings

An additional 15 adult participants rated each sound token along three dimensions: communicativeness, emotional content, and valence. Ratings of communicativeness were collected to verify that sounds categorized as communicative or noncommunicative were in fact perceived as such. Ratings of emotional content and valence were also obtained to statistically control for the influence of emotional content and valence on the neural responses to communicative and noncommunicative sounds.

Participants were presented with each sound one at a time and asked to rate sounds on all three dimensions. Sounds were played to the participants in random order, and ratings were recorded on a laptop computer. Instructions to participants were given as follows:

> You will hear a series of sounds presented one at time. You will be asked three questions after hearing each sound. The first question is *How communicative is this sound?* Sounds are considered to be communicative when they are voluntarily or flexibly produced with the goal of sharing or communicating information to another person. In other words, a communicative sound is 1) intentionally produced with the goal of sharing information with another person and 2) can be flexibly adjusted to be appropriate for a given situation. State your answer by clicking anywhere along the bar. One end of the bar will be labeled *Least*. The other end of the bar will be labeled *Most*. After you make your rating a second question will appear. The second question is: *How emotional is this sound?* We are not concerned with the valence of the emotion (in other words we are not asking if the sound evokes a positive or negative emotion). We just want to know how much emotion the sound conveys. Make your rating by clicking anywhere along the bar. One end of the bar will be labeled *Least*. The other end of the bar will be labeled *Most*. Remember, *Most* can mean positive or negative. After you make your rating a third question will appear: *Does this sound convey a positive or negative emotion?* Make your rating by clicking anywhere along the bar. One end of the bar will be labeled *Most negative*. The other end of the bar will be labeled *Most positive*. The midpoint of the bar corresponds to neutral. Each sound will be presented one at time. You can take as much time as you like to answer each question. Feel free to use the full range of the bar when answering the three questions.

The values on the bar scaled arbitrarily from 0 (corresponding to *Least* or *Most negative*) and 660 (corresponding to *Most* or *Most positive*). Reliability between raters was high (intraclass correlation coefficient = .73, $p <$ .0001). MANOVAs with Sound Category (two levels: communicative and noncommunicative) as the fixed factor revealed significant main effects of Sound Category for communicativeness, $F(1,118) = 268.0$, emotional content, $F(1,118) = 141.9$, and valence, $F(1,118) = 7.75$; $p < .05$ for all. Communicative sounds were significantly higher on measures of communicativeness, emotional content, and positive valence compared with noncommunicative sounds. Results from these ratings confirmed that sounds

categorized as communicative were in fact perceived as such (see Table 2 for means and standard deviations of ratings for each sound category).

Acoustic Analyses

Acoustic analyses were conducted to allow us to measure and statistically control for the influences of low-level acoustic features on neural responses to communicative versus noncommunicative sounds. Several acoustic properties were assessed, including duration, intensity, harmonicity (the degree of acoustic periodicity, also referred to as the harmonic-to-noise ratio), and pitch. All acoustic analyses were performed using Praat software 5.1.07 (Boersma & Weenink, 2009). MANOVA with Sound Category (two levels: communicative and noncommunicative) as the fixed factor revealed significant main effects of Sound Category for intensity, $F(1,118) = 11.77$, harmonicity, $F(1,118) = 68.69$, and pitch, $F(1,89) = 6.19$, all $p < .05$. There was no main effect of Sound Category for duration, $F(1,118) = .24$, $p = .62$. Communicative sounds were significantly higher on measures of intensity and harmonicity compared with noncommunicative sounds. Noncommunicative sounds were significantly higher in pitch than communicative sounds. Means and standard deviations of each acoustic property for each sound category are provided in Table 2.

Image Acquisition and Preprocessing

Data were acquired using a 3.0-T Siemens TIM TRIO scanner using a 12-channel head coil. Functional images were collected using a standard echo-planar pulse sequence (parameters: repetition time = 2 sec, echo time = 25 msec, flip angle = 60°, field of view = 220 mm, matrix = 64^2, voxel size = 3.4 × 3.4 × 4 mm, 34 slices). High-resolution T1-weighted anatomical images of the whole brain were acquired for registration using a 3-D MPRAGE sequence (repetition time = 1900 msec, echo time = 2.96 msec, flip angle = 9°, field of view = 256 mm, matrix = 256^2, voxel size = 1 × 1 × 1 mm, 160 slices).

Data were preprocessed and analyzed using the Brain-Voyager QX 2.0 software package (Brain Innovation, Maastricht, The Netherlands). The first nine volumes of the functional data set were discarded to allow for T1 equilibrium. Preprocessing of the functional data included slice time correction (using sinc interpolation), 3-D rigid-body motion correction (using trilinear-sinc interpolation), spatial smoothing with a FWHM 4-mm Gaussian kernel, linear trend removal, and temporal high-pass filtering (fast Fourier transform based with a cutoff of 3 cycles/time course). Functional images were registered to anatomical images, which were in turn normalized to Talairach space. Estimated motion plots and cine loops revealed that no participant had head motion greater than 3 mm of translation in any direction or 3 degrees of rotation about any axis.

fMRI Data Analysis

A random effects multiparticipant statistical analysis was performed by multiple linear regression of the time course of the BOLD response in each voxel. The model included explanatory variables for each of the eight sound categories: infant-directed speech, adult-directed speech, human communicative vocal nonspeech, human noncommunicative vocal nonspeech, walking sounds, clapping sounds, rhesus macaque vocalizations, and water sounds. Each sound category was modeled as a boxcar function peaking during each sound trial, convolved with a double-gamma hemodynamic response function. Linear contrasts using t statistics were performed to compare activations among experimental conditions. Given our a priori hypotheses, we were interested in first localizing the voice-selective region identified by Belin and colleagues (2000) by performing a vocal > nonvocal contrast. We then wanted to test our specific hypothesis by examining whether this region is especially sensitive to communicative vocal sounds (infant-directed speech, adult-directed speech, and human communicative vocal nonspeech) compared with noncommunicative vocal sounds (human noncommunicative vocal nonspeech). Areas of activation from the vocal > nonvocal contrast were thresholded using a false discovery rate of $q < .05$ to correct for multiple comparisons (Genovese, Lazar, & Nichols, 2002).

A second random effects multisubject analysis was performed to statistically control for the influence of sound properties (duration, pitch, intensity, harmonicity, emotional content, and valence) on the observed fMRI signal. The mean values per block of each sound property were included as additional regressors in the analysis. Importantly, sound property values were z-normalized before being entered into the model (Leaver & Rauschecker, 2010).

RESULTS

The vocal > nonvocal contrast revealed greater activation in the left STS and superior temporal gyrus (STG) and the right STS and STG. The peak coordinates of these regions correspond closely to those reported in previous studies examining voice-selective brain regions (see Table 3) and fell within the middle STG.

To examine our main question of interest, whether this voice-selective region responds equally to all vocal sounds or more specifically to communicative compared with noncommunicative vocal sounds, we examined the beta values from the voxels comprising the regions identified in the vocal > nonvocal contrast. Note that this is a rather conservative approach, biased against finding a stronger response to communicative compared with noncommunicative vocal sounds: By first identifying a voice-selective region, we are more likely to obtain high beta values for all vocal conditions. A repeated-measures ANOVA conducted on beta values from voxels within the voice-sensitive

Table 2. Acoustic Properties and Behavioral Ratings by Sound Category

Sound Category	Duration (msec)	Intensity (db)	Pitch (Hz)	Harmonicity (db)	Communicativeness	Emotion	Valence
Communicative Signals							
Infant-directed speech	895 (240)	70.2 (.3)	289 (55)	15.0 (3.8)	553 (19)	417 (43)	450 (41)
Adult-directed speech	623 (91)	70.0 (.6)	228 (31)	9.8 (4.1)	552 (19)	304 (32)	331 (27)
Communicative vocal nonspeech	729 (180)	70.5 (.1)	267 (59)	13.1 (5.9)	591 (25)	482 (95)	308 (194)
Mean for communicative signals	749 (210)	70.2 (.4)	261 (13)	12.6 (5.1)	565 (26)	401 (97)	363 (129)
Noncommunicative Signals							
Noncommunicative vocal nonspeech	838 (452)	70.2 (.5)	303 (67)	8.7 (7.7)	280 (151)	208 (93)	278 (38)
Rhesus macaque calls	724 (328)	70.6 (.3)	335 (116)	7.9 (8.2)	281 (132)	258 (119)	274 (54)
Clapping	737 (182)	67.9 (1.8)		−1.4 (1.1)	347 (23)	239 (20)	391 (11)
Walking	831 (92)	68.3 (1.1)	217 (125)	0.7 (1.4)	123 (19)	88 (11)	314 (11)
Water	747 (473)	70.1 (.5)	356 (86)	−0.4 (2.2)	58 (15)	54 (11)	324 (15)
Mean for noncommunicative signals	775 (324)	69.4 (1.5)	306 (13)	3.1 (6.6)	218 (141)	170 (107)	317 (52)

Values are given as mean (standard deviation). Communicativeness and emotion are rated on a scale from 0 (*least*) to 660 (*most*). Valence is rated on a scale from 0 (*most negative*) to 660 (*most positive*).

Table 3. Peak Coordinates of STS and STG Activation for Vocal > Nonvocal Contrasts Reported in Previous Studies and in the Current Study

| | Talairach Coordinates | | |
	x	y	z
Right STS/STG			
Belin et al., 2000	58	6	−10
Belin et al., 2000	60	−1	−4
Belin et al., 2000	63	−13	−1
Belin et al., 2002	63	−13	−1
Gervais et al., 2004	60	−12	4
Gervais et al., 2004	56	−20	−4
Average of prior studies	60	−9	−3
Current study	54	−10	−2
Left STS/STG			
Belin et al., 2002	−62	−14	0
Belin et al., 2000	−62	−14	0
Gervais et al., 2004	−64	−40	12
Gervais et al., 2004	−64	−12	4
Average of prior studies	−63	−20	4
Current study	−54	−16	4

region revealed a main effect of Sound Category in both the left and right hemispheres ($F(1,19) = 36.36, p < .001$; $F(1,19) = 19.60, p < .001$, respectively). Follow-up paired samples t tests revealed that voice-sensitive regions in the left and right hemispheres responded more strongly to each of the three sound types comprising the communicative vocal sound category (infant-directed speech, adult-directed speech, and human communicative vocal nonspeech) compared with the noncommunicative vocal sounds (e.g., coughs, yawns; all $ps < .01$; see Figure 1). This finding was further supported by a whole-brain communicative > noncommunicative contrast which revealed significant activation in both the left and right STS and STG. The peak coordinates of these clusters fell within the middle STG (left: −57, −16, 4; right: 57, −11, 1) and corresponded closely to the peak coordinates reported in the vocal > nonvocal contrast.

Furthermore, hemispheric differences were observed in the sensitivity to vocal communicative nonspeech sounds. There was no difference in the right hemisphere in response to all three vocal communicative conditions (all $ps > .17$). However, the left hemisphere responded more strongly to speech sounds compared with communicative vocal nonspeech sounds (all $ps < .05$), suggesting that the voice-selective region of the left hemisphere is especially sensitive to linguistic communicative signals.

Including sound properties (duration, intensity, pitch, harmonicity, emotional content, and valence) as additional regressors in our model resulted in no significant voxels responding more strongly to communicative compared with noncommunicative sounds. Importantly, however, one of the sound properties, emotional content, was highly correlated with ratings of how communicative the sounds were ($r = .92, p < .0001$), raising the possibility that adding emotional content as a regressor was accounting for much of the variance in the observed fMRI signal. We therefore performed the analysis without emotional content as an additional regressor. In this case, we observed significant activation in the right STG and STS and in the left STG in response to communicative compared with noncommunicative sounds. Although the clusters in both hemispheres extended continuously from anterior to middle regions, local peak foci were observed in anterior and middle portions of the STS and STG (see Figure 2 for areas of activation and Table 4 for peak coordinates).

DISCUSSION

Vocal sounds elicited bilateral activation in the STS and STG, replicating previous studies of voice perception (Belin et al., 2000) using a new corpus of vocal and nonvocal stimuli. Importantly, the current results extend previous findings by demonstrating that, while this region is selective for human voices, it is especially sensitive to communicative vocal sounds compared with noncommunicative vocal sounds. Communicative vocal sounds (infant-directed speech, adult-directed speech, and communicative nonspeech sounds, such as laughter) elicited greater activation than vocal noncommunicative sounds (e.g., coughs and yawns), in the left and right anterior to middle STG and the right middle STS, even when controlling for sound properties. Moreover, we found a hemispheric asymmetry in activation within the communicative sounds, such that human speech sounds elicited more activation than communicative nonspeech vocalizations in the left but not the right hemisphere.

Activation in the STS and STG in response to communicative compared with noncommunicative sounds was observed even when statistically controlling for the influence of duration, intensity, harmonicity, pitch, and valence. Although these results suggest that communicative value plays a strong role in modulating activity in the STS and STG, two other acoustic properties, emotional and phonetic content, may also be contributing to the neural response in voice-selective areas. Ratings of emotional content were highly correlated with communicativeness and when emotion was included as a regressor, no regions showed selectivity for communicative versus noncommunicative sounds. In addition, two of the three communicative sound categories contained phonetic content (adult-directed speech and infant-directed speech). However, we believe that it is unlikely that emotional or phonetic content, rather than communicativeness, are
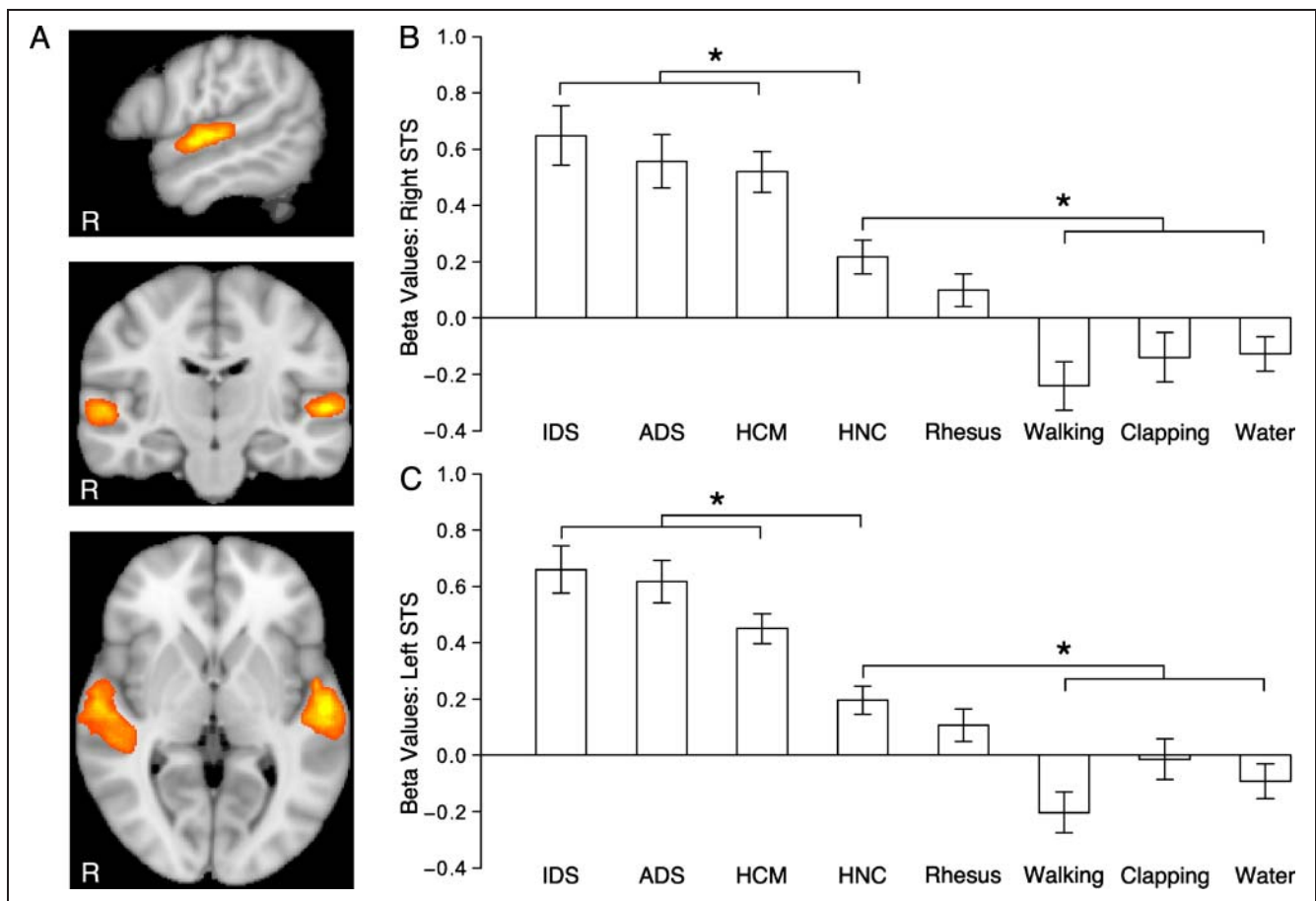
Figure 1. Experimental results. (A) The location of the left and right STG and STS regions in which activation was significantly greater in response to vocal compared with nonvocal sounds ($q < .05$). "R" indicates the right hemisphere. Bar graphs show beta values from regions within the right (B) and left (C) hemispheres. Error bars indicate standard errors of the means. IDS = infant-directed speech; ADS = adult-directed speech; HCM = human communicative vocal nonspeech; HNC = human noncommunicative vocal nonspeech.

driving the observed results. The influence of emotion and communicativeness can be dissociated by considering the response to adult-directed speech. Whereas adult-directed speech was rated as being significantly lower in emotional content than both infant-directed speech and nonspeech human communicative sounds, beta values for adult-directed speech were just as large as infant-directed speech and human communicative

sounds. The influence of phonetic content and communicativeness can be dissociated by considering the response to communicative nonspeech sounds (e.g., laughter) and noncommunicative sounds. Whereas communicative nonspeech sounds are devoid of phonetic content, beta values for communicative nonspeech sounds were larger than for all noncommunicative sounds, including vocal noncommunicative sounds. As

Figure 2. Experimental results depicting the location of the right STG and STS and left STG region in which activation was significantly greater in response to communicative compared with noncommunicative sounds ($q < .05$), even when statistically controlling for the influence of acoustic features (pitch, duration, intensity, harmonicity, and valence).
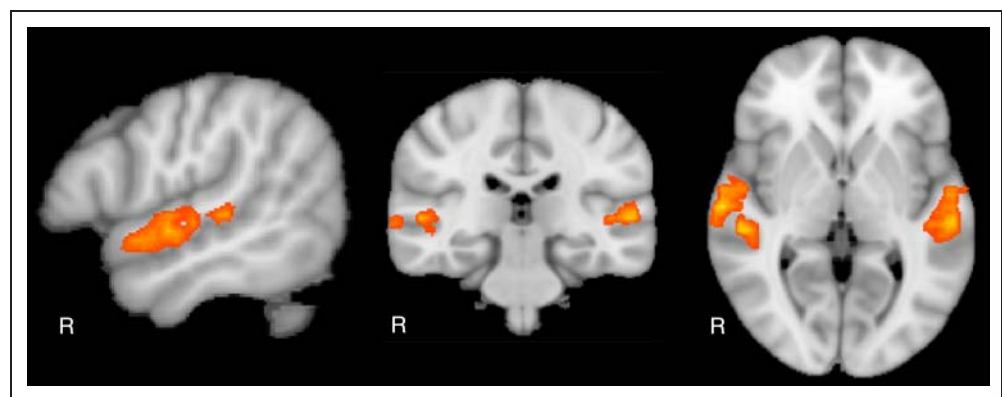
Table 4. Peak Coordinates of Activation for the Communicative > Noncommunicative Contrast, When Statistically Controlling for the Influence of Low-level and Perceptual Acoustic Features

| | Talairach Coordinates | | |
	x	y	z
Right anterior STG	53	−7	1
Right middle STS	48	−28	4
Left anterior STG	−51	−7	−2
Left middle STG	−57	−22	4

such, we propose that the most parsimonious explanation is that activation in this region reflects sensitivity to communicativeness.

Implications for the Neural Infrastructure of Voice Processing

The greater activation in the STS for communicative sounds is consistent with a key role for the STS in social perception and cognition in analyzing the communicative significance of auditory and visual inputs (Redcay, 2008). An important question for future research is whether this region is in fact voice selective or whether human non-vocal communicative signals would activate this region of the STS to the same degree as human vocal communicative signals. Although this has not been directly tested, there is reason to believe that this region may be sensitive to both vocal and nonvocal communicative signals. Although anterior portions of the STS are generally associated with speech processing, multiple fMRI studies revealed activation in the anterior STS for theory of mind tasks and in the middle STS for motion processing, face processing, and audiovisual integration (Hein & Knight, 2008). Given the multimodal nature of processing in the STS, it remains plausible that this voice-selective region is sensitive to communicative signals in both auditory and visual domains.

Finally, even within human vocal sounds, nonnative speech sounds preferentially engaged the voice-selective region in the left hemisphere, corroborating findings of left hemisphere dominance for language processing in right-handed adults (Möttönen et al., 2006; Vouloumanos et al., 2001). This further supports the dissociability of neural mechanisms processing linguistic and communicative aspects of language stimuli (Willems et al., 2010).

Implications for Understanding Neurodevelopmental Disorders

The current findings may have important implications for elucidating the nature of voice processing deficits previously described in ASD, a neurodevelopmental disorder with impairments in social interaction and communication.

Unlike typically developing children, children with ASD fail to demonstrate preferential attention to speech (Kuhl, Coffey-Corina, Padden, & Dawson, 2005; Klin, 1991) and exhibit difficulty in extracting information about the mental states of others from voices (Rutherford, Baron-Cohen, & Wheelwright, 2002). Adults with ASD fail to activate voice-sensitive regions of the STS when listening to vocal compared with nonvocal sounds (Gervais et al., 2004), suggesting that individuals with ASD exhibit abnormal cortical voice processing. The results of the current study offer some insight into the nature of these reported cortical voice-processing deficits. Rather than having a deficit in voice processing per se, individuals with ASD may have a specific deficit in recognizing and extracting the communicative significance contained in vocal sounds (Redcay, 2008). This interpretation is parsimonious with reports indicating that children with ASD differ markedly from typically developing children on language tasks that require an understanding of the communicative intentions of others (reviewed by Redcay, 2008; Sabbagh, 1999). For instance, children with ASD perform better when responding to direct, as opposed to indirect, requests, suggesting a difficulty in understanding the communicative significance of linguistic input (Paul & Cohen, 1985). Characterizing the precise functional role of the STS in preferential processing of a subset of vocal stimuli, those with communicative functions, has important implications for the functional processing of socially relevant stimuli and for our conceptualization of the deficits present in ASD.

Reprint requests should be sent to Sarah Shultz, Department of Psychology, Yale University, 2 Hillhouse Ave., New Haven, CT 06511, or via e-mail: sarah.shultz@yale.edu.

REFERENCES

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 4,* 267–278.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8,* 129–135.

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

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13,* 17–26.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10,* 512–528.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer [Computer program]. (Version 5.1.07). Available from www.praat.org.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11,* 535–543.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage, 24,* 21–33.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage, 23,* 840–848.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Sciences, 7,* 77–83.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15,* 870–878.

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., et al. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience, 7,* 801–802.

Grice, H. P. (1968). Utterer's meaning, sentence meaning, and word meaning. *Foundations of Language, 4,* 1–18.

Hauser, M. D. (2000). A primate dictionary? Decoding the function and meaning of another species' vocalizations. *Cognitive Science, 24,* 445–475.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4,* 223–233.

Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—It's my area: Or is it? *Journal of Cognitive Neuroscience, 20,* 2125–2136.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3,* 80–84.

Klin, A. (1991). Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders, 21,* 29–42.

Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage, 22,* 948–955.

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. *Developmental Science, 8,* F1–F12.

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex, 13,* 1023–1033.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *The Journal of Neuroscience, 30,* 7604–7612.

Maynard Smith, J., & Harper, D. D. (2003). *Animal signals.* New York: Oxford University Press.

Morris, J. P., Pelphrey, K. A., & McCarthy, G. (2008). Perceived causality influences brain activity evoked by biological motion. *Social Neuroscience, 3,* 16–25.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage, 30,* 563–569.

Paul, R., & Cohen, D. J. (1985). Comprehension of indirect requests in adults with autistic disorders and mental retardation. *Journal of Speech and Hearing Research, 28,* 475–479.

Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *The Journal of Neuroscience, 23,* 6819–6825.

Price, C. J. (2000). The anatomy of language: Contributions from functional neuroimaging. *Journal of Anatomy, 197,* 335–359.

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 358,* 435–445.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. *Neuroscience & Biobehavioral Reviews, 32,* 123–142.

Roder, B., Stock, O., Neville, H., Bien, S., & Rosler, F. (2002). Brain activation modulated by the comprehension of normal and pseudoword sentences of different processing demands: A functional magnetic resonance imaging study. *Neuroimage, 15,* 1003–1014.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders, 32,* 189–194.

Sabbagh, M. A. (1999). Communicative intentions and language: Evidence from right-hemisphere damage and autism. *Brain and Language, 70,* 29–69.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16,* 235–239.

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia, 42,* 1435–1446.

Specht, K., & Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: An auditory rapid event-related fMRI-task. *Neuroimage, 20,* 1944–1954.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: MIT Press.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience, 13,* 994–1005.

Willems, R. M., de Boer, M., de Ruiter, J. P., Noordzij, M. L., Hagoort, P., & Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychological Science, 21,* 8–14.

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *Neuroimage, 25,* 1002–1015.

Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., & Boddaert, N. (2006). Autism, the superior temporal sulcus and social perception. *Trends in Neurosciences, 29,* 359–366.