ELSEVIER

Is voice processing species-specific in human auditory cortex?
An fMRI study

Shirley Fecteau,[a,b,c,]* Jorge L. Armony,[c,d] Yves Joanette[a,b,c] and Pascal Belin[b,c]

[a]*Faculté de Médecine, Université de Montréal, Montreal, Canada*
[b]*Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM), Montreal, Canada*
[c]*Centre de Recherche en Neuropsychologie Expérimentale et Cognition (CERNEC), Université de Montréal, Montreal, Canada*
[d]*Douglas Hospital Research Center, McGill University, Montreal, Canada*

Recent studies suggested a sensitivity of regions of the human superior temporal sulcus (STS) to the sound of the human voice. However, the question of the species specificity of this response is still open. Healthy adult volunteers were scanned in an event-related fMRI design to compare responses in the STS to human and animal vocalizations, as well as to control nonvocal sounds (e.g., musical instruments). Bilateral activation of anterior STS was observed for human vocalizations, when contrasted with both nonvocal sounds and animal vocalizations. Animal vocalizations, compared to nonvocal sounds, elicited a more restricted left STS activation, although this region responded even more strongly to human vocalizations. This study provides the first evidence suggesting a species specificity in STS responses to vocalizations in humans.

Introduction

Species-specific vocalizations are social sounds upon which several species of vertebrates rely for their survival (e.g., Andrew, 1963; Altmann, 1967; Fossey, 1972; Gautier and Gautier, 1977; Green, 1975; Petersen, 1982; Seyfarth et al., 1980; Smith et al., 1982; Snowdon, 1982; Struhsaker, 1967; Tian and Rauschecker, 1998; van Lawick-Goodall, 1968; Winter et al., 1966). Evidence from neurophysiological studies on primary auditory cortex in nonhuman primates suggests that species-specific vocalizations are highly effective stimuli which can induce greater neural responses than other less behaviorally

relevant sounds (Wang and Kadia, 2001). More recently, a neuroimaging study (Poremba et al., 2004) on primary auditory cortex in rhesus monkeys showed that species-specific vocalizations induce greater activation in the left superior temporal gyrus (STG), relative to the right, whereas no such asymmetric responses were observed for human vocalizations and nonvocal sounds.

In humans, recent neuroimaging studies identified brain regions, located within the superior temporal sulcus (STS), which respond more strongly to human voices than to nonhuman sounds (Belin et al., 2000, 2002). However, the issue of species specificity of these responses is still unresolved. Given that some of the acoustic features of animal vocalizations are similar to human vocalizations (e.g., Rendall, 2003; Scherer and Kappas, 1988), it is possible that voice-responsive areas in STS could also respond, to a comparable degree, to nonhuman animal vocalizations. Alternatively, these regions could show a species-specific response and therefore respond more strongly to human than to animal vocalizations.

To address this issue, we conducted an event-related fMRI study to directly test the species specificity of STS responses to vocal sounds. To that effect, we compared brain activity elicited by human and animal vocalizations. Based on evidence from previous studies on the involvement of STS on human voice processing, we focused our analysis on this region.

Materials and methods

Participants

Fifteen healthy, right-handed adults ($n = 15$, mean age = 22.6 years, six females) with no history of neurological or psychiatric conditions participated in this study. Informed written consent was obtained, and the study was approved by the ethical committee from Centre Hospitalier de l'Université de Montréal.

* Corresponding author. Département de Psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montreal, Que., Canada, H3C 3J7.
E-mail address: shirley.fecteau@umontreal.ca (S. Fecteau).

Available online on ScienceDirect (www.sciencedirect.com).

Procedure

Auditory stimuli were presented binaurally through pneumatic headphones, sealed by foam ear inserts, and further shielded by plastic ear defenders (providing an attenuation of fMRI scanning noise of about 30 dB) using MCF (DigiVox, Montreal), with a sound-pressure level of 85–90 dB. They consisted of 144 sounds divided into five categories (Fig. 1): 24 speech sounds (e.g., vowels, words, and sentences in different languages), 24 non-linguistic vocalizations (e.g., laughs, screams, coughs), 24 cat vocalizations, 24 mixed-animal vocalizations, and 48 nonvocal sounds (e.g., mechanical sounds and musical instruments). The 24 mixed-animal vocalizations were produced by alligator, antelope, bird, camel, chimpanzee, chinchilla, cow, dog, donkey, eagle, elephant, frog, hippopotamus, horse, kangaroo, koala, moose, pig, raccoon, rooster, sea lion, sheep, turtle, and whale (sample stimuli from all categories can be obtained at http://www.mapageweb. umontreal.ca/belinp/). There was no significant difference in mean duration between human vocalizations (mean = 1.21; SD = 0.22) and animal vocalizations (mean = 1.34; SD = 0.32). The mean duration for nonvocal sounds was 1.56 (SD = 0.32). The reason for presenting a substantial number of cat vocalizations was to have a category of vocalizations which was comparable, as much as possible, to the human vocalizations category in terms of semantic familiarity and homogeneity. All human and animal vocalizations were produced by different individuals, and most stimuli used in this study were different from those used in Belin et al. (2000). Sounds were normalized to a common peak of intensity with Mitsyn (WLH) and CoolEdit Pro (Syntrillium Software Corp.). Stimuli were presented in a random order (mean SOA = 5.5 s, SD = 2.4 s). Null events (i.e., scanner noise) were included and were used as baseline in the main auditory contrast (described below).

Imaging details

Scanning was performed on a 1.5-T MRI system (Magneton Vision, Siemens Electric, Erlangen, Germany) at the Centre Hospitalier de l'Université de Montréal. Functional scans were acquired with a single-shot echo planar gradient-echo (EPI) pulse sequence (TR = 2.6 s, TE = 40 ms, flip angle = 90°, FOV = 215 mm, matrix = 128 × 128). The 28 axial slices (resolution 3.75 × 3.75 mm in-plane, 5-mm thickness) in each volume were aligned with the AC–PC line, covering the whole brain. A total of 320 volumes were acquired (the first four volumes of each series were
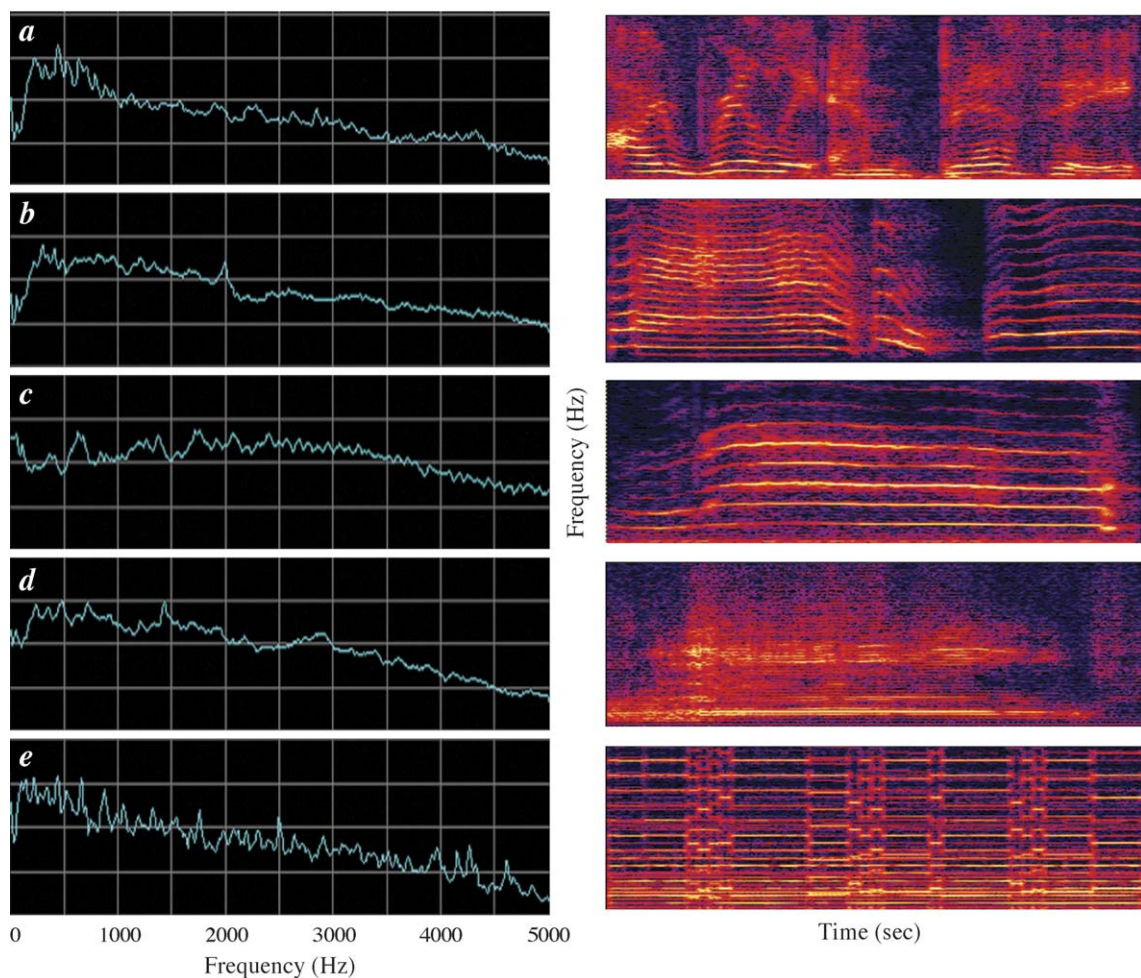

Fig. 1. Stimulus categories. Left panels: average frequency distribution. *X*-axis: frequency (0–5000 Hz). *Y*-axis: amplitude (84 to 36 dB). Right panels: spectrogram of a representative sample in each category. *X*-axis: time (variable). *Y*-axis: frequency (0–5000 Hz). Color indicates amplitude. (a) Human speech stimuli. (b) Human nonlinguistic vocalizations. (c) Cat vocalizations. (d) Mixed-animal vocalizations. (e) Nonvocal stimuli. See Materials and methods for description of stimuli. Sample stimuli can be heard at http://www.mapageweb.umontreal.ca/belinp/.

later discarded to allow for T1 saturation). After the functional scanning, T1-weighted anatomical images were obtained for each participant ($1 \times 1 \times 1$ mm resolution). Scanner noise was continuous throughout the experiment providing a constant background (baseline).

fMRI analysis

Image processing and statistical analysis were performed using SPM99 (Wellcome Department of Cognitive Neurology; Friston et al., 1995; Worsley and Friston, 1995), as done in previous studies (Armony and Dolan, 2001, 2002). Briefly, the imaging time series was realigned to the first volume to correct for interscan movement. Time series were shifted using sinc interpolation to correct for differences in slice acquisition times. The functional images were then spatially normalized to a standard stereotaxic space (Talairach and Tournoux, 1988) based on a template provided by the Montreal Neurological Institute (Evans et al., 1994) to allow for group analysis. Finally, functional data were spatially smoothed with an 8-mm isotropic Gaussian kernel to compensate for residual inter-participant variability and to allow for the application of Gaussian random field theory in the statistical analysis (Friston et al., 1995). Each subject's structural scan was coregistered with the mean realigned functional image and normalized using the parameters determined for the functional images. A mean anatomical image was created from the participants' individual scans, onto which activation was overlaid for anatomical localization.

Data analysis was performed in a two-stage mixed-effects analysis (equivalent to a random effects analysis) in which BOLD responses for each subject were first modeled using a synthetic hemodynamic function in the context of the fixed-effects general linear model. Subject-specific linear contrasts on the parameter estimates were then entered into a second-level analysis to perform between-subjects analyses, resulting in a t statistic for each voxel. These t statistics (transformed to Z statistics) constitute a statistical parametric map (SPM). SPMs were thresholded at $P = 0.05$, corrected for multiple comparisons across the brain, except for the STS, where a $P = 0.001$ (uncorrected) was used, based on our a priori hypothesis on the involvement of this region in voice processing (Belin et al., 2000). All statistical comparisons involved comparisons between some of the five conditions (speech, nonlinguistic vocalizations, cat vocalizations, mixed-animal vocalizations, and nonvocal sounds), except for the main auditory contrast in which all experimental conditions were compared to null events.

After the scanning session, participants were asked to rate the familiarity of the cat and mixed-animal vocalizations using a visual analog scale (1–100). The order of presentation of the stimuli was different from that of the scanning session and was counterbalanced across participants.

The effects of stimulus familiarity on brain activation were analyzed by multiplying the HRF to the mean-corrected familiarity score for each stimulus. This was done on an individual basis, that is, using the familiarity scores from each participant.

Results

Auditory activation

When all categories of auditory stimuli were compared to baseline (i.e., scanner noise), a large part of the superior temporal

plane corresponding to auditory cortex was activated bilaterally $P < 0.001$ corrected; see Table 1a and Fig. 2 for the time course of the two highest peaks). These maxima were located within the superior temporal gyrus (STG) and were activated by the five categories of stimuli to comparable extent.

Human versus nonhuman sounds

Human vocalizations (speech and nonlinguistic vocalizations) elicited greater bilateral responses in STS, compared to nonhuman sounds (animal vocalizations and nonvocal sounds), as shown in Table 1b and Fig. 3. At the individual level, 11 out of 15 participants showed significant bilateral STS responses; one showed only left STS activation, one showed only right STS activation, and two participants did not show significant STS activation ($P > 0.01$).

Human and animal responses

The contrast of human versus nonvocal sounds yielded significant bilateral activations in STS (see Table 1i and Fig. 4). This STS activation was very similar to the one observed in the previous contrast (human versus nonhuman). Although cat vocalizations, contrasted with the nonvocal sounds, activated a restricted left anterior STS region (see Table 1g and Figs. 4 and 5), this region responded significantly more to human than to animal vocalizations (Table 1c and Figs. 4 and 5).

We further explored the left STS activation for animal vocalizations compared to the nonvocal sounds. We investigated whether this left STS response was modulated by the familiarity ratings of cat and mixed-animal stimuli. No correlation between the familiarity ratings and the left STS response to animal vocalizations was observed (z score < 1).

Discussion

STS responses to human voices

In this study, we investigated neural responses to human and animal vocalizations, focusing on the region of the superior temporal sulcus (STS). Previously, Belin et al. (2000, 2002) showed, using an fMRI block design, that circumscribed regions along the upper bank of the STS responded selectively to human voices, whether they contained linguistic information. This study provides further support for the existence of voice-selective areas in the human brain, using different stimuli and different participants. In addition, by using an event-related design, we can exclude potential confounding effects due to habituation and/or stimulus expectation in these responses.

The areas of voice-selective activation observed here are fairly similar to those obtained by Belin et al. (2000) (see Fig. 3a). Although we observed extended activation of STS/middle temporal gyrus, our peaks of activation (Table 1b) correspond fairly well to the maxima reported in Belin et al. (2000), especially in the right hemisphere, with three maxima in the middle and anterior parts of STS (see coordinates from Belin et al., 2000, in Fig. 3a). When overlaid onto the participants' mean MR image, the activations are clearly located within the STS region, although Talairach coordinates suggest more inferior activation (compared to Belin et al., 2000). These

Table 1
Coordinates of significant STS responses for the different contrasts of interest: coordinates are in Talairach space; coordinates have been chosen according to their z scores (highest peaks in each hemisphere; $P < 0.001$), their location (if two local maxima were closer than 5 mm, only the greater maximum was reported), and their voxel size (minimum of three connected voxels)

Anatomical location	Side	Size	x	y	z	z score
(a) Main auditory activation > baseline (null events)						
Superior temporal gyrus	R	3958	62	−28	2	6.19
			62	−28	−4	6.15
			64	−26	8	6.03
Superior temporal gyrus	L	3509	−50	−30	4	6.34
			−54	−18	−2	5.59
			−46	−20	0	5.39
(b) Human vocalizations (speech and nonlinguistic vocalizations) > nonhuman sounds (animal vocalizations and nonvocal sounds)						
Superior temporal sulcus	R	1037	60	−24	−8	5.23
			52	8	−24	5.04
			60	−14	−10	5.00
Superior temporal sulcus	L	545	−58	−10	−12	4.69
			−64	−24	−10	4.41
			−66	−32	0	4.18
(c) Human vocalizations > animal vocalizations (cat and mixed-animal vocalizations)						
Superior temporal sulcus	R	841	50	10	−26	4.63
			60	−24	−8	4.60
			60	−14	−10	4.34
Superior temporal sulcus	L	289	−58	−10	−12	4.82
			−64	−18	−4	3.78
			−60	−20	−6	3.55
(d) Animal vocalizations (cat and mixed-animal vocalizations) > human vocalizations						
No significant STS response						
(e) Human vocalizations > cat vocalizations						
Superior temporal sulcus	R	75	60	−20	−10	3.81
			62	−24	−8	3.56
			52	−20	−16	3.41
Superior temporal sulcus	L	93	−66	−36	0	3.68
			−62	−42	4	3.46
			−66	−44	6	3.35
(f) Human vocalizations > mixed-animal vocalizations						
Superior temporal sulcus	R	755	56	−34	−4	4.68
			60	−22	−6	4.35
			48	12	−26	4.21
Superior temporal sulcus	L	286	−58	−12	−12	4.49
			−62	−14	−14	4.42
			−52	−28	−12	3.76
(g) Cat vocalizations > nonvocal sounds						
Superior temporal sulcus	L	6	−62	−16	−12	3.49
(h) Mixed-animal vocalizations > nonvocal sounds						
No significant STS response						
(i) Human vocalizations > nonvocal sounds						
Superior temporal sulcus	R	794	54	−24	−10	5.16
			58	−30	−8	5.12
			54	6	−20	5.10
Superior temporal sulcus	L	525	−64	−26	−8	4.76
			−62	−16	−14	4.68
			−50	−38	−12	4.01

differences, of the order of 1 cm, which are well within the variability typically observed in neuroimaging studies, could reflect individual differences between groups (e.g., Belin et al., 2002), and/or they could be due to differences in design (event-related fMRI vs. block design) and in analysis (random vs. fixed effects).
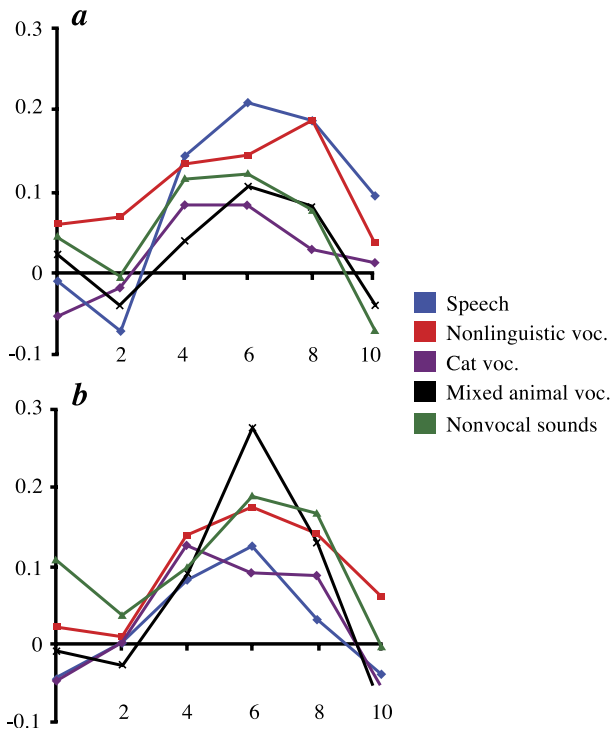
Fig. 2. Group-averaged peristimulus time response in the main auditory > baseline contrast (a) in the right STG voxel [62, −28, 2], the maximal peak in the right hemisphere, and (b) in the left STG voxel [−50, −30, 4], the global maximum. *X*-axis: time (seconds).

We hypothesize that the anterior activations observed here could correspond to a processing stream specialized in source identification: the 'what' stream, that is, source recognition. Indeed, animal studies have suggested an organization of auditory cortex in processing streams, similar to that evidenced in visual cortex. Rauschecker (1998) and Rauschecker and Tian (2000) proposed the existence of an anterolaterally directed stream for the processing of auditory objects, including conspecific vocalizations. More recently, Tian et al. (2001) showed that responses of anterolateral neurons of auditory belt cortex, relative to caudolateral neurons, seemed to be more specific for stimuli such as monkey calls. In the human brain, recent neuroimaging studies support the existence of an anterior stream for sound recognition (Alain et al., 2000; Maeder et al., 2001). Our results, as well as those from Belin et al. (2000), are consistent with these studies; that is, we observed more anterior activation in the STS than in the posterior part of the STS, and these could indeed reflect processes associated with identification of the source of the voice (e.g., speaker identification, Belin and Zatorre, 2003; Belin et al., 2004).

Species specificity of STS responses

The main objective of this study was to explicitly test whether STS responses were selective for human vocalizations. To test the species specificity of the voice-selective STS response, we included a category of animal vocalizations from several different nonhuman species. Bilateral STS regions were significantly more responsive to human vocalizations than to either animal vocalizations (see Tables 1c, e, f) or nonvocal sounds (see Table 1i). Furthermore,
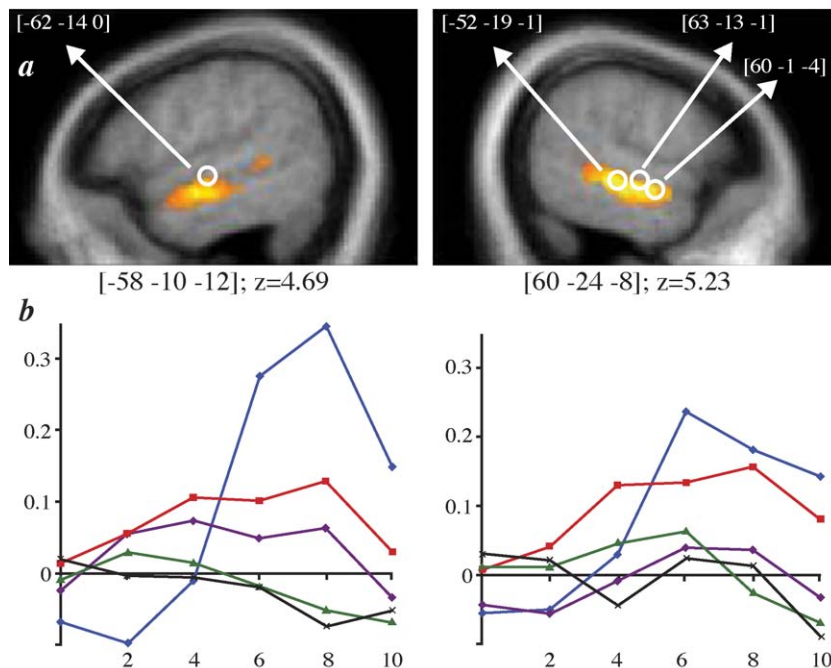

Fig. 3. (a) Human vocalizations (speech and nonlinguistic vocalizations) > nonhuman sounds [animal vocalizations (cat and mixed-animal vocalizations)] and nonvocal sounds (e.g., musical instruments, environmental sounds) contrast *P* < 0.001. (b) Group-averaged peristimulus time response in the human > nonhuman contrast in the global maximum, the right STS voxel [60, −24, −8], and in the left STS voxel [−58, −10, −12], the maximal peak on the left hemisphere. *X*-axis: time (seconds). The four white circles are the maximum peaks of the contrast human vocalizations > nonhuman vocalizations observed in Belin et al. (2000).
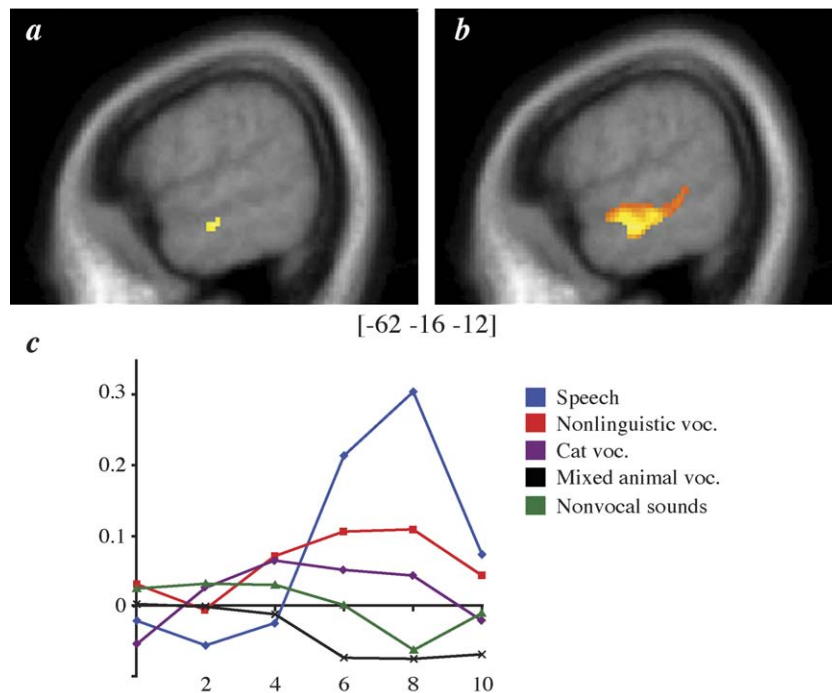
Fig. 4. (a) Animal vocalizations (cat and mixed-animal vocalizations) > nonvocal sounds (e.g., musical instruments, environmental sounds) contrast, (b) human vocalizations (speech and nonlinguistic vocalizations) > nonvocal sounds contrast ($P < 0.001$), and (c) group-averaged peristimulus time response in the left STS voxel $[-62, -16, -12]$, the global maximum in the animal vocalizations (cat and mixed-animal vocalizations) > nonvocal sounds contrast. X-axis: time (seconds).

compared to a common baseline (i.e., nonvocal sounds), stronger STS responses were found for human vocalizations than for animal vocalizations (see Figs. 4 and 5). Thus, with the present comparison set of animal vocalizations, there seems to be species specificity in the human STS response to vocalizations. It is important to emphasize that our criterion for species selectivity is a significantly stronger response to human vocalizations than to other types of sounds, including animal vocalizations and nonvocal sounds. Thus, we cannot rule out the possibility that these regions may also respond to other types of stimuli not included in this study. In addition, in the present study, we cannot demonstrate that STS does not respond to animal vocalizations, as this would be akin to attempting to prove the null hypothesis. However, our data also indicate that some areas in the STS show a significant response to animal vocalizations as compared to the nonvocal sounds, although of much lesser magnitude than for the human sounds.

In order to test the possibility that STS responses to human voices were simply due to stimulus familiarity or homogeneity (i.e., same species), we included in our design a stimulus category entirely comprised of cat vocalizations. A comparison between human and cat vocalizations also revealed bilateral significant STS activation, similar to that between human and mixed-animal vocalizations (see Tables 1e, f), providing further support for human specificity of STS responses to vocal stimuli. Moreover, no correlation was observed between familiarity ratings and neural activity, suggesting that this measure does not account for the STS activation observed here.

Animal studies of species specificity

Although direct comparison between results from animal and human studies is difficult, it is nonetheless informative to examine

our findings in the context of studies of species specificity in the nonhuman primate brain. Several neurophysiological studies reported that processing species-specific vocalizations is different from artificial sounds (Newman and Wollberg, 1973; Tian et al., 2001; Wang et al., 1995; Wollberg and Newman, 1973). Fewer studies compared species-specific to heterospecific vocalizations. Wang and Kadia (2001) showed that responses of subpopulation of neurons in A1 marmoset were selective to species-specific vocalizations over backwards vocalizations, which was not observed in the cat A1 (Wang and Kadia, 2001). More recently, Poremba et al. (2004), using PET in rhesus monkeys, showed asymmetric processing of species-specific vocalizations (left > right hemisphere) within the left superior temporal gyrus, close to regions homologous to our STS peaks, which was not observed for human vocalizations and nonvocal sounds (e.g., glass breaking, tones). Thus, although cerebral responses associated with the rhesus vocalizations were not directly compared to those observed for other categories of sounds, species-specific calls appeared to be processed differently than other sounds, including heterospecific vocalizations. Thus, there is converging evidence from studies involving nonhuman primates, suggesting that species-specific calls evoke stronger responses in some regions of auditory cortex than other types of stimuli. Given the relatively few physiological studies addressing the question of species-specific responses to vocalizations, less is currently known about the extent of such specificity in other regions of auditory cortex, particularly STS.

Parallel with the brain responses observed for human faces

We observed significant STS activation for animal vocalizations when contrasted with nonvocal sounds, yet this activation was smaller than that elicited by human voices (see Figs. 4 and 5).
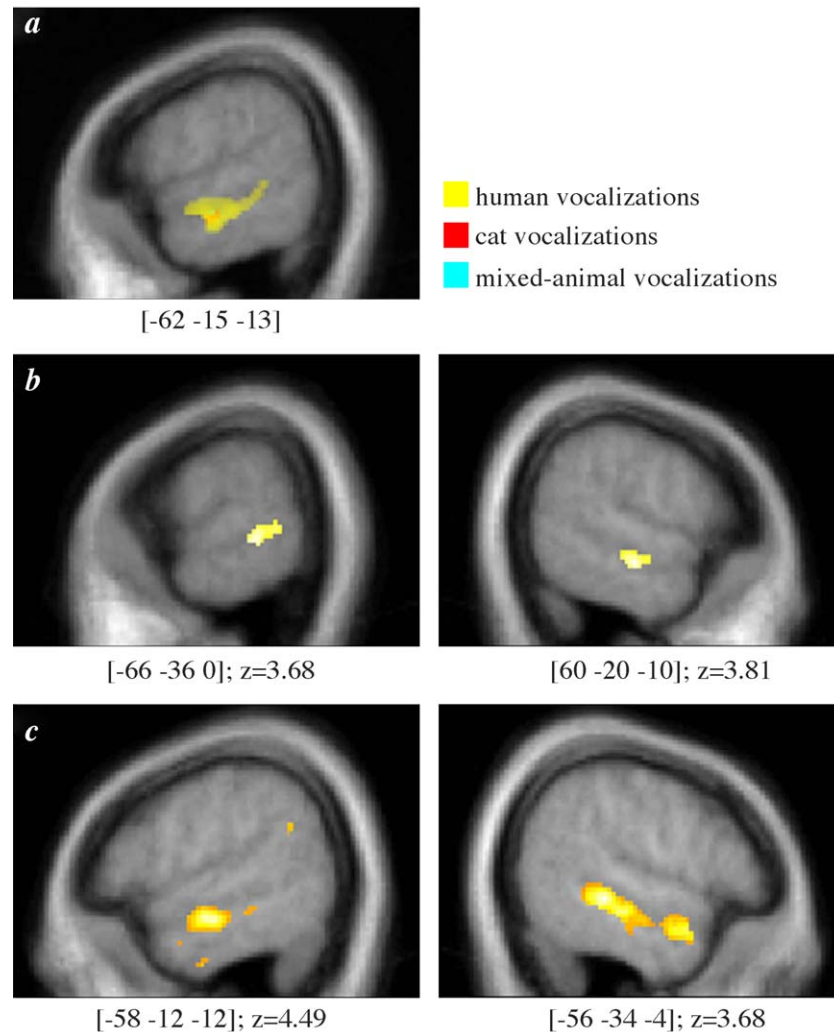
Fig. 5. (a) Superposition of the three main contrasts of interest [human vocalizations (speech and nonlinguistic vocalizations) > nonvocal sounds, cat vocalizations > nonvocal sounds, and mixed-animal vocalizations > nonvocal sounds contrasts] in the global maximum for the cat vocalizations > nonvocal sounds contrast, the left STS voxel $[-62, -15, -13]$ ($P < 0.001$), (b) human vocalizations (speech and nonlinguistic vocalizations) > cat vocalizations in the right STS $[60, -20, -10]$ ($P < 0.001$) and in the left STS $[-66, -36, 0]$ ($P < 0.001$) and human vocalizations (speech and nonlinguistic vocalizations) > mixed-animal vocalizations in the right STS $[56, -34, -4]$ ($P < 0.001$) and in the left STS $[-58, -12, -12]$ ($P < 0.001$).

This pattern is similar to what has been observed in the visual domain for the perception of human faces. Several studies have shown that faces elicit greater activation than objects in the fusiform gyrus (e.g., Kanwisher et al., 1997; McCarthy et al., 1997). Together, these results have been interpreted as evidence for the existence of a "face area" [fusiform face area (FFA); Kanwisher et al., 1997]. It has been suggested (Kanwisher et al., 1999) that FFA is selective for human faces but not for animals. However, Kanwisher et al. (1999) observed a significantly stronger response in this region for animal faces and whole animals than to object stimuli. Thus, the activity in the FFA was weakest for pictures of assorted objects, stronger for the animal stimuli, and strongest for human faces. Another study (Halgren et al., 2000), using MEG, showed an occipitotemporal peak of amplitude at 165 ms for animal faces, but at a level which was below its amplitude for human faces.

This notion of 'category-specific' areas in human brain has been challenged by Haxby et al. (2001; but also see Spiridon and Kanwisher, 2002). Haxby et al. (2001) proposed that the

representation of faces and different categories of objects is widely distributed and overlapping, arguing that any information, regardless of type, is processed by many different parts of the brain and that any brain region is likely to represent many different classes of information. These different notions from Haxby et al. (2001) and Spiridon and Kanwisher (2002), trying to account for the functional architecture underlying this capacity to generate a unique representation from an unlimited variety of stimuli, need to be further investigated within both the visual and auditory domains.

Moreover, as in the case of face processing, there is the question of the role of expertise: is this preference for human stimuli innate or the result of an overexposure to those stimuli? This question is still in debate within the visual domain (Kanwisher, 2000; Tarr and Gauthier, 2000). For example, Diamond and Carey (1986) hypothesized that the mechanisms involved in face recognition are also engaged when subjects make discriminations between structurally similar exemplars of a category for which they have gained substantial visual expertise. Indeed, Gauthier et al. (2000)

observed significant FFA activity with fMRI when bird experts viewed bird stimuli (but see Kanwisher, 2000). We are all expert in processing human voices, which is one of the earliest forms of social interaction in ontogeny and the sound category that we hear most often in our daily life. It is possible that the observed STS activation associated with human vocalizations could reflect an expertise in processing categorization at a subordinate level instead of being specific to human voices. In that case, significant STS responses to animal vocalizations should be observed in human subjects with expertise in animals.

In summary, this is the first study to present evidence suggesting selective responses to conspecific vocalizations in STS in the human brain. Bilateral STS responses were stronger to human voices (speech and nonlinguistic vocalizations) than to animal vocalizations and nonvocal sounds. The exact nature of this selectivity needs further exploration. For example, our design does not allow us to rule out the possibility that differences in cortical responses may be due to other factors, such as differences in the acoustic structures between the human and animal vocalizations acoustics, the familiarity and/or saliency of stimulus categories, the expertise for human vocalizations, or the vocal identity. Although the human vocalizations used here were produced by different speakers, these stimulus categories, unlike the other categories, might trigger a nearly automatic search for individual vocal identity. Future work will include participants who are experts in animal vocalizations to explore whether the possible search for the identity of the animal producing the sound will result in a different pattern of activation. This pattern may be similar to what we observed here: human participants, experts in human vocalizations, showed stronger activation to human vocalizations compared to categories of vocalizations for which they are not experts. The question of whether these responses reflect a hard-wired system for social communication or they are learned is still open.

References

Alain, C., Arnott, S.R., Hevenor, S., Graham, S., Grady, C.L., 2000. "What" and "where" in the human auditory system. Proc. Natl. Acad. Sci. 98, 12301–12306.

Altmann, S.A., 1967. The structure of primate communication. In: Altmann, S.A. (Ed.), Social Communication Among Primates. Chicago Univ. Press, Chicago, IL, pp. 325–362.

Andrew, R.J., 1963. The origins and evolution of the calls and facial expressions of the primates. Behaviour 20, 1–109.

Armony, J.L., Dolan, R.J., 2001. Modulation of auditory neural responses by a visual context in human fear conditioning. NeuroReport 12, 3407–3411.

Armony, J.L., Dolan, R.J., 2002. Modulation of spatial attention by fear-conditioned stimuli: an event-related fMRI study. Neuropsychologia 40, 817–826.

Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. NeuroReport 14, 2105–2109.

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. Brain Res., Cogn. Brain Res. 13, 17–26.

Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.

Diamond, R., Carey, S., 1986. Why faces are and are not special: an effect of expertise. J. Exp. Psychol. Gen. 115, 107–117.

Evans, A.C., Kamber, M., Collins, D.L., Macdonald, D., 1994. An MRI-based probabilistic atlas of neuroanatomy. In: Shorvon, S., Fish, D., Andermann, F., Bydder, G.M., Stefan, H. (Eds.), Magnetic Resonance Scanning and Epilepsy. Plenum, New York, pp. 263–274.

Fossey, D., 1972. Vocalizations of the mountain gorilla (Gorilla gorilla beringei). Anim. Behav. 20, 36–53.

Friston, K.J., Holmes, A.P., Poline, J.B., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Gautier, J.P., Gautier, A., 1977. Communication in old world monkeys. In: Sebeok, T.A. (Ed.), How Animals Communicate. Indiana Univ. Press, Bloomington, pp. 809–964.

Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. 3, 191–197.

Green, S., 1975. Variations of vocal pattern with social situation in the Japanese monkey (Macaca fuscata): a field study. In: Rosenblum, L.A. (Ed.), Developments in Field and Laboratory Research. Academic, New York, pp. 1–102.

Halgren, E., Raij, T., Marinkovic, K., Jousmaki, V., Hari, R., 2000. Cognitive response profile of the human fusiform face area as determined by MEG. Cereb. Cortex 10, 69–81.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.

Kanwisher, N., 2000. Domain specificity in face perception. Nature 3, 759–763.

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

Kanwisher, N., Stanley, D., Harris, A., 1999. The fusiform face area is selective for faces not animals. NeuroReport 10, 183–187.

Maeder, P.P., Meuli, R.A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J.-P., Pittet, A., Clarke, S., 2001. Distinct pathways involved in sound recognition and localization: a human fMRI study. NeuroImage 14, 802–816.

McCarthy, G., Puce, A., Gore, J.C., Allison, T., 1997. Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. 9, 605–610.

Newman, J.D., Wollberg, Z., 1973. Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. Brain Res. 54, 287–304.

Petersen, M.R., 1982. The perception of species-specific vocalizations by primates: a conceptual framework. In: Snowdon, C.T., Brown, C.H., Petersen, M.R. (Eds.), Primate Communication. Cambridge Univ. Press, Cambridge, UK, pp. 171–211.

Poremba, A., Malloy, M., Saunders, R.C., Carson, R.E., Herscovitch, P., Mishkin, M., 2004. Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature 427, 448–451.

Rauschecker, J.P., 1998. Parallel processing in the auditory cortex of primates. Audiol. Neuro-Otol. 3, 86–103.

Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. 97, 11800–11806.

Rendall, D., 2003. Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. J. Acoust. Soc. Am. 113, 3390–3402.

Scherer, K.R., Kappas, A., 1988. Primate vocal expression of affective states. In: Todt, D., Goedeking, P., Newman, E. (Eds.), Primate Vocal Communication. Springer, Heidelberg, pp. 171–194.

Seyfarth, R.M., Cheney, D.L., Marler, P., 1980. Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. Science (Washington, D.C.) 210, 801–803.

Smith, H., Newman, J.D., Symmes, D., 1982. Vocal concomitants of affiliative behavior in squirrel monkeys. In: Snowdon, C.T., Brown, C.H., Petersen, M.R. (Eds.), Primate Communication. Cambridge Univ. Press, Cambridge, pp. 30–49.

Snowdon, C.T., 1982. Linguistic and psycholinguistic approaches to primate communication. In: Snowdon, C.T., Brown, C.H., Petersen, M.R. (Eds.), Primate Communication. Cambridge Univ. Press, Cambridge, pp. 212–238.

Spiridon, M., Kanwisher, N., 2002. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron 35, 1157–1165.

Struhsaker, T.T., 1967. Auditory communication among vervet monkeys (*Cercopithecus aethiops*). In: Altmann, S.A. (Ed.), Social Communication Among Primates. Chicago Univ. Press, Chicago, IL, pp. 238–324.

Talairach, P., Tournoux, J., 1988. Stereotaxic Coplanar Atlas of the Human Brain. Thieme, Stuttgart, Gemany.

Tarr, M.J., Gauthier, I., 2000. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nature 3, 764–769.

Tian, B., Rauschecker, J.P., 1998. Processing of frequency-modulated sounds in the cat's posterior auditory field. J. Neurophysiol. 79, 2629–2642.

Tian, B., Reser, D., Durham, A., Kustov, A., Rauschecker, J.P., 2001. Functional specialization in rhesus monkey auditory cortex. Science 292, 290–293.

van Lawick-Goodall, J.A., 1968. A preliminary report on expressive movements and communication in the Gombe Stream chimpanzees. In: Jay, P.C. (Ed.), Primates: Studies in Adaptation and Variability. Rinehart and Winston, Holt, New York, pp. 313–374.

Wang, X., Kadia, S.C., 2001. Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. J. Neurophysiol. 86, 2616–2620.

Wang, X., Merzenich, M., Beitel, R., Schreiner, C., 1995. Representation of a species-specific vocalizations in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J. Neurophysiol. 74, 2685–2706.

Winter, P., Ploog, D., Latta, J., 1966. Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. Exp. Brain Res. 1, 359–384.

Wollberg, Z., Newman, J.D., 1973. Auditory cortex of squirrel monkey: response patterns of single cells to species-specific vocalizations. Science 175, 212–214.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—Again. NeuroImage 2, 173–181.