

# Human Brain Regions Involved in Recognizing Environmental Sounds

James W. Lewis<sup>1</sup>, Frederic L. Wightman<sup>4</sup>, Julie A. Brefczynski<sup>2</sup>, Raymond E. Phinney<sup>1</sup>, Jeffrey R. Binder<sup>2,3</sup> and Edgar A. DeYoe<sup>1,2</sup>

<sup>1</sup>Department of Radiology, Medical College of Wisconsin, Milwaukee, WI 53226, USA, <sup>2</sup>Department of Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, WI 53226, USA, <sup>3</sup>Department of Neurology, Medical College of Wisconsin, Milwaukee, WI 53226, USA and <sup>4</sup>Heuser Hearing Institute, University of Louisville, KY 40203, USA

**To identify the brain regions preferentially involved in environmental sound recognition (comprising portions of a putative auditory 'what' pathway), we collected functional imaging data while listeners attended to a wide range of sounds, including those produced by tools, animals, liquids and dropped objects. These recognizable sounds, in contrast to unrecognizable, temporally reversed control sounds, evoked activity in a distributed network of brain regions previously associated with semantic processing, located predominantly in the left hemisphere, but also included strong bilateral activity in posterior portions of the middle temporal gyri (pMTG). Comparisons with earlier studies suggest that these bilateral pMTG foci partially overlap cortex implicated in high-level visual processing of complex biological motion and recognition of tools and other artifacts. We propose that the pMTG foci process multimodal (or supramodal) information about objects and object-associated motion, and that this may represent 'action' knowledge that can be recruited for purposes of recognition of familiar environmental sound-sources. These data also provide a functional and anatomical explanation for the symptoms of pure auditory agnosia for environmental sounds reported in human lesion studies.**

**Keywords:** auditory cortex, fMRI, multimodal processing, object motion, temporal lobe

## Introduction

Different aspects of the sounds we hear are thought to be processed along different regions of the brain. Similar to the visual system, a growing body of evidence would suggest that the human and non-human primate auditory systems are at least roughly organized along two major cortical streams or networks (Mishkin *et al.*, 1983; Rauschecker, 1998; Romanski *et al.*, 1999; Rauschecker and Tian, 2000; Clarke *et al.*, 2002). One includes a dorsally directed network in both hemispheres that is involved in processing spatial information about sound (a 'where is it?' pathway), such as for sound-source localization, motion perception and spatial attention (Griffiths *et al.*, 1996; Baumgart *et al.*, 1999; Bushara *et al.*, 1999; Weeks *et al.*, 1999; Lewis *et al.*, 2000; Maeder *et al.*, 2001; Zatorre and Penhune, 2001; Warren *et al.*, 2002). The other stream involves a relatively more ventrally located network that is involved in aspects of sound recognition (a 'what is it?' pathway), which generally includes the processing of speech sounds or species specific vocalizations, and natural or environmental sounds. In humans, spoken language recognition involves a widespread cortical system, much of which is lateralized to the left hemisphere (Binder *et al.*, 1997; Belin *et al.*, 2000; Price, 2000). As for the processing of environmental (non-verbal) sounds, candidate regions have been reported for both hemispheres

(Engelien *et al.*, 1995; Giraud and Price, 2001; Humphries *et al.*, 2001; Maeder *et al.*, 2001). However, we still have only a fragmentary understanding of the precise brain regions and processing pathways that constitute a system for non-verbal sound recognition.

Although rare, lesions to portions of temporal or temporo-parietal cortex can lead to a 'pure' auditory agnosia for environmental sounds, which is defined as an impaired capacity to recognize auditory information (such as a doorbell ring or typewriter sounds) despite adequate hearing and speech comprehension (Spreeen *et al.*, 1965; Albert *et al.*, 1972; Vignolo, 1982; Fujii *et al.*, 1990; Schnider *et al.*, 1994; Engelien *et al.*, 1995; Clarke *et al.*, 2000; Clarke *et al.*, 2002; Saygin and Moineau, 2002). Such auditory agnosia has been observed after right, left, and bilateral lesions, though left hemisphere (and bilateral) lesions tend to produce additional deficits in verbal comprehension. Thus far, no precise anatomical locations have been correlated with producing auditory agnosia. Nonetheless, these lesion studies do suggest that the cortical pathways for processing environmental sounds and spoken language are at least partially separable at some level, though closely linked with one another, especially in the language dominant hemisphere (Vignolo, 1982; Schnider *et al.*, 1994; Saygin *et al.*, 2003).

Current neurological and cognitive models for how spoken language is processed include input, intermediate and output processing stages (Grabowski and Damasio, 2000; Price, 2000; Binder and Price, 2001; Wise *et al.*, 2001; Binder, 2002). This model serves as a starting point for assessing how non-verbal, environmental sounds might be processed and subsequently recognized. Primary auditory cortex and some of the surrounding cortex represent input stages, which are thought to be involved in processing physical features of the spoken word sounds. Intermediate processing stages include lexical-semantic and other associative processes, which involve a wide range of cortices predominantly in the left hemisphere (of right-handed subjects), including the classically defined Wernicke's area among other cortical regions. Output stages involve phonological access and articulatory planning (whether vocalizations are produced or not), for which the left inferior frontal cortex is widely implicated, including the cortex of and surrounding the classically defined Broca's area.

In contrast to spoken words, whose sounds tend to have an arbitrary relationship with the concepts they represent, most environmental sounds bear a natural and physical correspondence to the visible and sometimes tangible object movements that produce the sound. When environmental sounds are first experienced and learned, information regarding the identity of the sound-source is typically obtained in the context of other

sensory information (e.g. hearing the sound of a basketball bouncing while viewing the bouncing motion and/or making arm and hand movements to dribble the ball). Presumably, these separate streams of relevant sensory information (sound, sight, touch) can merge and integrate in the cortex to help provide a unified percept of an object and its functional dynamics. An important step in understanding the complexities of such multimodal integration would be to identify which brain regions participate preferentially in environmental sound-source recognition. Such regions, in contrast to some of the lateralized structures that appear to be specialized for spoken language processing in humans, may represent part of a more general or rudimentary system for sound recognition.

We used functional magnetic resonance imaging (fMRI) to reveal brain regions involved in the recognition of common environmental sounds. In an effort to maximally activate a sound recognition system, we included a diverse range of non-verbal, environmental sound categories (see Appendix). To reveal which regions were more sensitive to the process of 'recognition' *per se*, we also presented temporally reversed (backwards) versions of the same sounds, which served as control stimuli that were comparably complex and matched on many acoustical features, but were generally judged as unrecognizable. Some of these sound pairs can be heard in the online Supplementary Material. The results indicate that the posterior portions of the middle temporal gyrus (pMTG) in both hemispheres are primary loci involved with the retrieval of knowledge ('recognition') associated with a variety of environmental sounds. Some of these fMRI data can be viewed at <http://brainmap.wustl.edu/vanessen.html>, which contains a database of surface-related data from various brain mapping studies. Portions of these data have been reported previously (Lewis *et al.*, 2001).

## Materials and Methods

### Subjects and Task

Twenty-four right-handed adults (aged 21–47 years, 13 women) with no history of neurological, psychiatric or auditory symptoms participated in the imaging study. Informed consent was obtained following guidelines approved by the Medical College of Wisconsin Human Research Review Committee.

Environmental (non-verbal) sound samples were compiled from a CD collection (General 6000 series, Sound Ideas) and from various websites. Samples (44.1 kHz, 16-bit, monophonic) were trimmed to ~2 s duration (1.1–2.5 s range) and temporally reversed for the backward presentations (Cool Edit Pro, Syntrillium Software Co.). The temporally reversed sounds were chosen as a baseline control as they were typically judged to be unrecognizable, yet matched the physical features of the original sounds in five important aspects, including overall intensity, duration, spectral shape or content, spectral variation or motion (Thivard *et al.*, 2000), and acoustic complexity.

Six subjects, not included in the fMRI experiments, screened numerous pairs of backward- and then forward-played sound samples, retaining 105 sounds that could generally be verbally identified when played forward but typically not identified when played backward (see Appendix for complete list, and in the online Supplementary Material to hear some of these stimuli). Across seven separate fMRI scans, subjects (with eyes closed) were presented with 350 sound trials (105 forward, 105 backward, 140 silent) in a pseudo-random order: since recognition of a backward sound would be facilitated by previous experience with the corresponding forward sound, a given backward sound always preceded its forward presentation by at least two non-silent trials.

During each fMRI scan, subjects indicated by right hand button press whether they (1) could recognize or identify the sound (i.e.

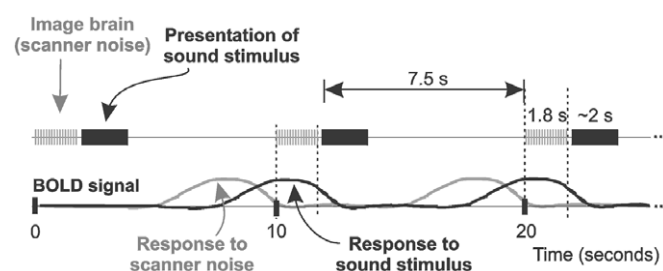
verbalize, describe, visualize or have a general sense of familiarity about the likely sound-source); (2) were uncertain; or (3) could not recognize the sound. Here we define 'recognition' as a sense of familiarity or implicit knowledge about a sound, whereas 'identification' additionally involves a verbal or semantic labeling of the sound-source. Button responses (and reaction times relative to sound onset) were collected during scanning both to help engage the listener to the sounds and to subsequently model the resulting fMRI data relative to each individual's judgment as to whether each sound was recognizable or not, which varied across subjects. Twelve subjects used their right index finger to button press for recognizable sounds, middle for uncertain, and ring finger for unrecognizable sounds. For the other 12 subjects, the fingering order was reversed in order to control for possible differences in response output.

### Imaging and Data Analysis

Scanning was conducted at 1.5 Tesla on a General Electric (GE Medical Systems, Milwaukee, WI) Signa scanner, equipped with a commercial head coil (Medical Advances Inc., Milwaukee, WI) suited for whole-head, echo-planar imaging of blood-oxygenation level dependent (BOLD) signals (Bandettini *et al.*, 1993; Ogawa *et al.*, 1993). Subjects wore earplugs and were presented with binaural sound stimuli, which could easily be heard via custom electrostatic headphones (Koss Inc., Milwaukee, WI). To compensate for frequency specific attenuations by the earplugs, subjects listened to several cycles of six sine wave tones (128, 256, 1024, 2048, 4096 and 8192 Hz) just prior to scanning. The sound intensity for each tone was adjusted using a nine-band equalizer (CFX-12, Mackie Co.) until all tones were perceived to be at roughly the same loudness (typically 70–85 dB, L-weighted through ear plugs).

We used a 'silent' clustered-acquisition fMRI design that allowed stimulus events to be presented during scanner silence. The scanning cycle, schematized in Figure 1, was repeated every 10 s, and consisted of presentation of a sound or silent event (~2 s), followed by silence during which time the subjects responded. The collection of BOLD signals (brain images, 1.8 s slice package) started 7.5 s after onset of each sound stimulus (or silence), and was the only time that the scanner made noise. In each scanning run, we acquired 52 gradient-recalled image volumes ( $T_E = 40$  ms,  $T_R = 10$  s), which included 16 axial slices of 6 mm thickness, with in-plane voxel dimensions of  $3.75 \times 3.75$  mm. For most subjects, this volume covered nearly the entire brain, originating at the temporal pole (Talairach coordinate  $z \approx -41$ ) (Talairach and Tournoux, 1988), and extending up to or within ~1 cm of the dorsal-most portions of the brain (range  $z = +55$  to  $+65$ ).  $T_1$ -weighted anatomical MR images were collected using a spoiled GRASS pulse sequence (1.1 mm slices, with  $0.9375 \times 0.9375$  mm in-plane resolution).

Data were viewed and analyzed using the AFNI software package (Cox, 1996) and related plug-in software (available at <http://afni.nimh.nih.gov/afni/index.shtml>). For each subject, the seven scans were concatenated into one time series, with the exception that for two subjects we retained six of seven scans and for one subject we retained five of seven scans (due to technical difficulties or excessive head motion during data acquisition). The first acquired brain volume in each scan (always a response to silence) was discarded, and the



**Figure 1.** Schematic of the imaging paradigm. Upper panel shows cycles of stimulus presentation and MRI clustered-acquisitions. Lower panel shows the ideal BOLD response signals for a responsive voxel of brain.

remaining 51 brain volume images were motion corrected by re-registering them to the 20th brain volume of the last scan (closest to the time of anatomical image acquisition). This 3D motion registration accounted for global head translations ( $x$ ,  $y$ ,  $z$ ) and rotations (yaw, pitch, roll) using a least-squares fit algorithm (AFNI plug-in). We then performed multiple linear regression analyses based on the button responses modeling the sound stimuli relative to silent events. With the clustered acquisition design, the whole-brain BOLD response to each sound stimulus could be treated as an independent event. Consequently, we could disregard or censor particular events from the regression model in accordance with each subject's button responses (e.g. including only those sound stimulus pairs for which the forward-played version was judged as recognizable and the corresponding backward-played version was not recognizable).

For the group results, individual anatomical and functional brain maps were transformed, using AFNI, into standardized Talairach coordinate space (Talairach and Tournoux, 1988). Functional data (multiple regression coefficients) were spatially low-pass filtered (4 mm rms Gaussian filter), then merged by combining coefficient values for each interpolated voxel across all subjects. For the main data in Figure 2, the combination of individual voxel probability threshold ( $t$ -test,  $P < 0.001$ ) and the cluster size threshold (3 voxel minimum) yielded the equivalent of a whole-brain corrected significance level of  $\alpha < 0.05$ . A split-half (or 'cross-validation') correlation test (Binder *et al.*, 1997) was used to estimate how well the environmental sound recognition pattern of activation would generalize to other subject samples. Using the same threshold setting as in the full data set, a voxel-by-voxel correlation between two subgroups (roughly matched for age, gender and button response fingering order) was 0.60, yielding a Spearman-Brown estimated reliability coefficient of  $\rho_{XY} = 0.75$  for the entire sample of 24 subjects. This indicates the level of correlation that would be expected between the activation pattern of our sample of 24 subjects and activation patterns from other random samples of 24 subjects matched for age and gender. Public domain software packages SureFit and Caret (<http://brainmap.wustl.edu:8081/sums>) were used to project data onto the Colin Brain atlas model in Talairach coordinate space, and were used to display the data (Van Essen *et al.*, 2001; Van Essen, 2004).

## Results

For each subject, we initially analyzed only those sound stimuli that were judged as both recognizable when played forwards and unrecognizable when played backwards (on average 56 of 105 sound pairs were retained, see Appendix numbers in parentheses). Our first analysis, illustrated in Figure 2*a*, examined the group-averaged ( $n = 24$ ) pattern of activation (yellow hues) evoked by the recognizable, forward-played sound stimuli relative to silence. Similarly, Figure 2*b* shows the pattern for the corresponding backward-played, unrecognizable sound stimuli relative to silence. As expected, in both comparisons the strongest and most extensive activation (75% or greater of maximum intensity MR response) included primary auditory cortex (PAC) and the immediately surrounding cortex on the superior temporal plane (collectively termed PAC+) (Engelien *et al.*, 1995; Giraud and Price, 2001; Maeder *et al.*, 2001; Wessinger *et al.*, 2001), located bilaterally within the lateral sulcus (LaS) and superior temporal gyrus (STG). Moderate activation (~25–75% of maximum intensity) was present bilaterally in inferior frontal cortex, and in anterior cingulate cortex (not visible in lateral views). Additionally, there was activity within a large swath of cortex including the left central sulcus (CeS) and left post-central sulcus (PoCeS), which was most likely related to planning and/or production of the right hand button presses (Burton *et al.*, 1999; Burton, 2002). Finally, in both conditions, a few regions (dark blue) showed either a depression below baseline in response to sound presentation or relatively greater activation during the

silent periods. This included bilateral portions of extrastriate visual cortex, the right precentral cortex, and left orbital frontal cortex (not visible in lateral views).

To reveal brain regions preferentially involved in the process of recognizing environmental sounds, we effectively subtracted (via multiple linear regression analysis) the activation pattern for the unrecognized, backward-played sounds (Fig. 2*b*) from that for the recognized, forward-played (Fig. 2*a*) sounds (prior to thresholding). The resulting group-averaged pattern of cortical activity is illustrated on a three-dimensional model in Figure 2*c*, on the corresponding flat map representation in Figure 2*d*, and on select axial slices from the brain of one subject in Figure 2*e*. The Talairach coordinates of specific cortical foci are indicated in Table 1. Since the forward- and backward-played sounds were matched on many physical attributes (refer to Materials and Methods), they comparably activated auditory input stages, including the PAC+ and the STG bilaterally. A predominant difference between the forward and backward sounds in this analysis was that only the forward-played sounds were judged as recognizable. Thus, the activation foci revealed after the subtraction (in Fig. 2*c–e*, yellow) should largely reflect high-level processing associated with the recognition (and/or identification) of environmental sounds.

The main novel finding of this study was that environmental sound recognition evoked activity bilaterally in and surrounding posterior portions of the middle temporal gyri (pMTG) and superior temporal sulci (pSTS), including a single robust but isolated site in the right hemisphere, and a larger, more ventrally directed focus in the left hemisphere. Several of the other cortical foci revealed by this contrast were strongly left lateralized (inferior frontal cortex, anterior fusiform gyrus, angular gyrus, and to some extent the posterior cingulate cortex), and in locations consistent with their involvement in retrieving and selecting semantic and verbal information (Binder *et al.*, 1999; Price, 2000; Martin and Chao, 2001;

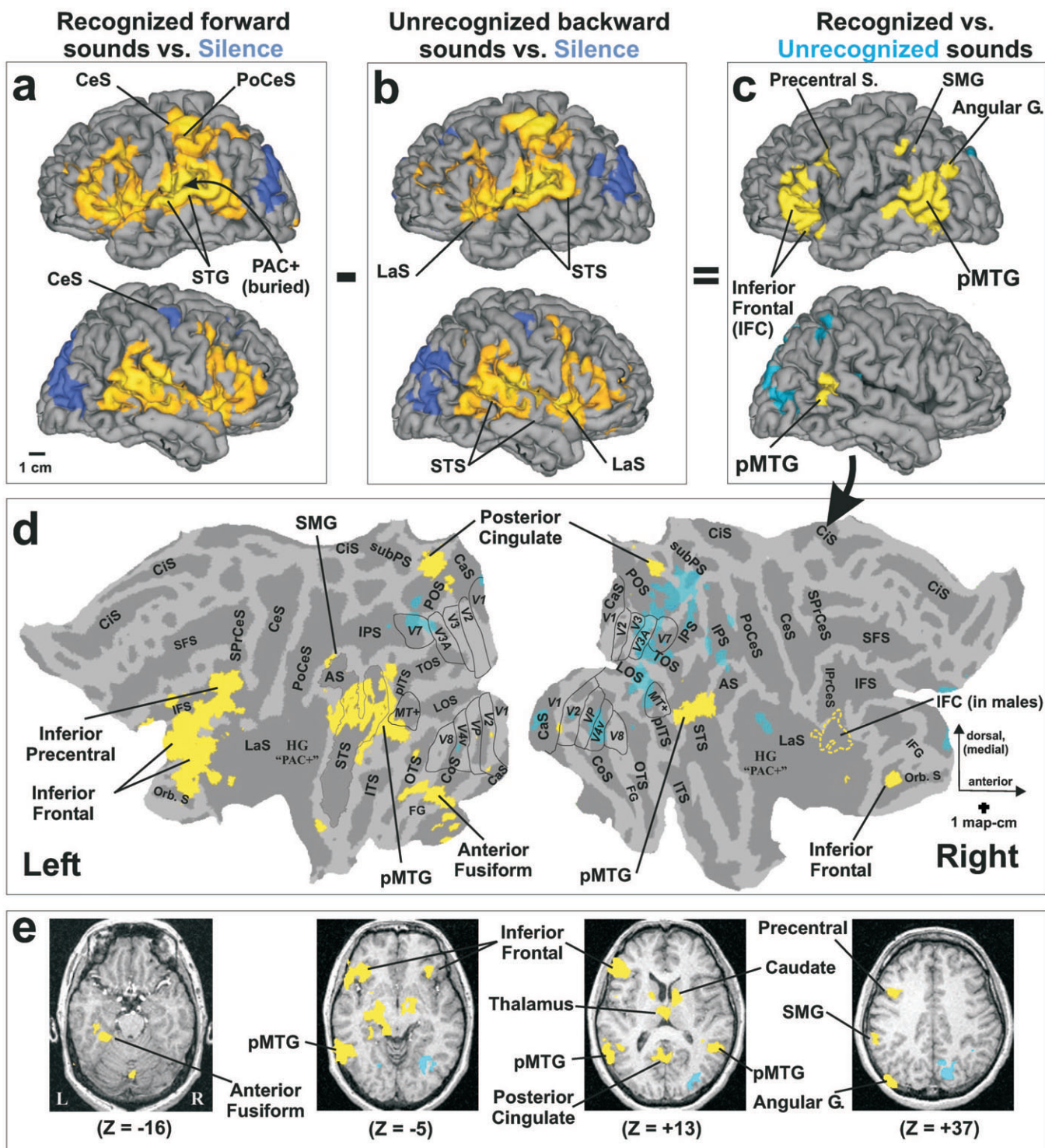
**Table 1**

Coordinates of environmental sound recognition foci reported in stereotaxic space (Talairach and Tournoux, 1988)

Anatomical location	Talairach coordinates			Volume (mm <sup>3</sup> )
	<i>x</i>	<i>y</i>	<i>z</i>	
Right hemisphere				
pMTG (and pSTS)	50	−49	13	1892
Inferior frontal g. (orbital s.)	31	32	−3	533
Left hemisphere				
pMTG (and pSTS)	−55	−52	5	9925
Inferior frontal cortex (IFC)				
Inferior precentral sulcus	−41	8	28	5235
‘Inferior frontal gyrus, dorsal’	−48	32	13	6217
‘Inferior frontal gyrus, ventral’	−42	30	−2	4908
Angular gyrus	−45	−76	32	2041
Posterior cingulate	−2	−56	16	1899
Anterior fusiform	−26	−36	−13	1400
Supramarginal gyrus (SMG)	−57	−38	38	549

The inferior frontal cortex was comprised of three major foci, evident at higher threshold settings ( $P < 0.00001$ , uncorrected). The reported volumes, however, are all based on the data shown in Figure 2 ( $P < 0.001$ , corrected to  $\alpha < 0.05$ ). IT, inferotemporal cortex; pMTG, posterior middle temporal gyrus; pSTS, posterior portions of superior temporal sulcus.





**Figure 2.** Brain regions involved in environmental sound recognition. Yellow hues show group-averaged ( $n = 24$ ) activated regions (relative increases in BOLD signal) and dark blue shows relative decreases in BOLD signal projected onto the Colin brain atlas, evoked by (a) recognizable, forward sounds relative to silence or (b) the corresponding unrecognized, backward sounds relative to silence. (c) Data from panel b subtracted from a, revealing regions preferentially involved in recognizing sounds (yellow) versus not recognizing the corresponding backward-played sounds (light blue), both relative to silence. Brain surfaces in a–c depict an approximation of layer 2. (d) Flat maps showing data from panel c. Identified visual areas (V1, V2, V3, MT+, etc.) are from the Colin atlas database and are indicated by black outline. The left STS is outlined in gray for clarity. See text for other details. (e) Axial sections of data from panel c displayed on the brain of one subject. All colored foci are statistically significant at a corrected  $\alpha < 0.05$ . Data sets are available for visualizing or downloading at: [http://brainmap.wustl.edu:8081/sums/directory.do?dir\\_id=707085](http://brainmap.wustl.edu:8081/sums/directory.do?dir_id=707085) (human colin atlas). AS, angular sulcus; CaS, calcarine sulcus; CeS, central sulcus; CiS, cingulate sulcus; CoS, collateral sulcus; FG, fusiform gyrus; HG, Heschl's gyrus; IFG, inferior frontal gyrus; IFS, inferior frontal sulcus; IPrCeS, inferior precentral sulcus; IPS, intraparietal sulcus; ITS, inferotemporal sulcus; LaS, lateral sulcus; LOS, lateral occipital sulcus; Orb. S, orbital sulcus; OTS, occipito-temporal sulcus; pITS, posterior inferotemporal sulcus; PoCeS, postcentral sulcus; POS, parieto-occipital sulcus; SMG, supramarginal gyrus; SPiCeS, superior precentral sulcus; STS, superior temporal sulcus; subPS, subparietal sulcus; TOS, transverse occipital sulcus.

Binder *et al.*, 2003). These activation sites are specifically addressed further in the Discussion.

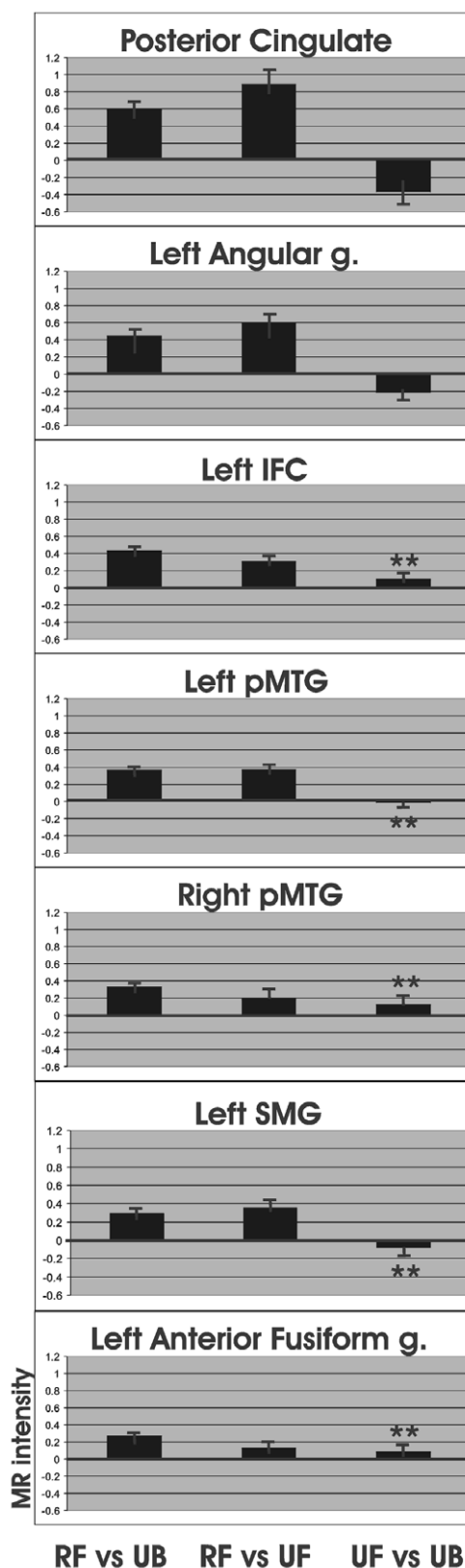
Regardless of the fingering scheme used for button responses, the group-averaged reaction time to the sounds judged as recognizable was 2.6 s (after sound onset), compared to 3.0 s for the corresponding backward-played, unrecognizable sounds. The significantly longer reaction times to the backward sounds [ $F(1, 22) = 16.1$ ,  $P < 0.0006$ , cross-nested three-factor ANOVA] may imply that there were greater processing demands required to judge a backward-played sound as unrecognizable. Consequently, this would suggest that those regions preferentially engaged by the recognizable environmental sounds (yellow in Fig. 2c–e) were not simply modulated by overall greater task difficulty or task demands (Barch *et al.*, 1997).

### Regional Analyses Involving ‘Miss-trials’

The backward-played sounds were chosen as a baseline condition primarily to address the issue of ‘recognition’ while controlling for numerous physical features of sound. Although the two sound datasets were ideally matched for overall duration, intensity, spectral variation, spectral content and acoustic complexity, they did differ in that the backward sounds were distorted in the temporal domain, thereby altering the temporal phase and onset of events. Thus, the activation in Figure 2 could conceivably reflect differences in the temporal acoustic properties of the forward- versus backward-played sounds, rather than differences in recognition. To investigate this issue, we focused specifically on the process of recognition itself by comparing the ‘miss-trials’ (i.e. the forward-played sounds judged as unrecognizable).

Figure 3 illustrates a comparison of BOLD signal differences within seven different regions of interest (ROIs). These ROIs were derived from the data shown in Figure 2c–e (yellow), and column 1 depicts the relative MR signals for the Recognized Forward (RF) versus Unrecognized Backward (UB) sounds, all being significantly positive in sign. Column 2 shows a similar Recognized versus Unrecognized comparison, but only includes sounds that were Forward-played (RF versus UF). If any of the ROIs were not sensitive to recognition *per se*, then the sign of the differential MR activity in column 2 should ideally drop to zero or be negative. However, in this comparison all of the cortical regions of interest were significantly positive in sign, suggesting that responses in these ROIs did indeed reflect recognition rather than stimulus differences. This was further supported by another analysis whose results are shown in column 3. This analysis directly compared responses among only the Unrecognized sounds: Forward versus Backward (UF versus UB). For such trials, no judgment of recognition was reported, so any potential differences should be due to differences in the stimulus properties of

forward versus backward sounds. In this case, five of the seven ROIs showed no significant difference from zero (*t*-test,  $P < 0.05$ ). In the remaining two ROIs (posterior cingulate and left angular gyrus) the small differences were negative in sign and



**Figure 3.** Analysis of ‘miss-trial’ data within regions of interest (ROIs) associated with sound recognition. Column 1 (RF vs. UB) shows the group-averaged MR BOLD response differences (arbitrary MR intensity units) within seven ROIs derived from Figure 2c–e, comparing Recognized, Forward-played sounds (RF; 1344 trials across all subjects) versus the corresponding Unrecognized Backward-played sounds (UB; 1344 trials). Column 2 (RF versus UF) compares the Recognized versus Unrecognized Forward-played sounds (UF; 373 trials). Column 3 (UF versus UB) illustrates Forward versus Backward sounds that were all deemed Unrecognizable. Standard errors were derived from the average of mean unthresholded MR intensities across the 24 subjects. Values not statistically different from zero are indicated by asterisks (\*\* $P < 0.05$ ).



are likely to reflect task-unrelated factors (Binder *et al.*, 1999, 2003). Together, these data indicate that the acoustical differences between forward and backward sounds did not account for our main results. [We also compared the ‘false positive’ trials, examining the Backward-played sounds judged as Recognizable (RB). The results from a full complement of miss-trial cross-comparisons did qualitatively support the findings illustrated in Figure 3, in that the ROIs were primarily involved in the process of ‘perceived recognition’. However, establishing on what grounds a subject ‘recognized’ a backward-played sound was more difficult to interpret than with ‘not recognizing’ a forward-played sound. Thus, we opted to focus on the Unrecognized Forward (UF) type of miss-trial.]

### Gender Differences

We compared the whole-brain activation patterns evoked by the men relative to the women for the main paradigm. Overall, both of these group-averaged data sets showed activation patterns comparable to that shown in Figure 2, with the notable exception of a focus in the right inferior frontal cortex (IFC) indicated by dotted yellow outline in Figure 2*d*. In response to recognizable sounds, this right IFC focus showed a significant increased BOLD signal in males, and a decreased signal in females (peak activation at Talairach  $x = 44$ ,  $y = 18$ ,  $z = 11$ ;  $\alpha < 0.05$  in a two-sample *t*-test for means). However, this region was no longer statistically significant after merging data across genders.

### Discussion

We used fMRI to measure brain responses to a wide range of recognizable environmental (non-verbal) sounds in contrast to comparably complex, yet unrecognizable, backward versions of the same sounds. The results revealed a network of brain regions that were preferentially involved in the process of recognizing environmental sounds. In contrast to earlier studies that examined environmental sounds, the present study revealed strong bilateral activation in posterior portions of the middle temporal gyri (pMTG, left > right). Moreover, the pMTG foci, in addition to the other regions of this network, were shown to represent a hierarchical stage beyond the early processing of the physical features of spectrally and temporally complex sounds, since earlier sound processing stages (e.g. primary and nearby surrounding auditory cortex, and the STG), were comparably activated by both the recognizable sounds and unrecognizable control sounds.

To interpret the environmental sound recognition data further, Figure 4 shows a direct comparison of the present findings with three other datasets illustrating cortical processing networks described in previous publications from our institution (colored cortex), all superimposed onto one brain model. These studies used the same scanning equipment and similar processing techniques, thereby providing a relatively accurate and direct comparison. Together, these data support distinctions between cortical regions engaged in (i) input stages of acoustic processing (blue hues); (ii) phonetic and semantic aspects of spoken language (red); and (iii) recognizing non-verbal, environmental sounds (yellow, from Fig. 2*c,d*). A system involved in visual motion processing is also illustrated (green), which may share some later stage processing mechanisms in common with environmental sound recognition. Furthermore, we have included the Talairach coordinate foci

reported in several earlier studies germane to present findings (Fig. 4*a* symbols). Together, these data are discussed in the context of (i) current sound processing models; (ii) semantic knowledge; (iii) visual object recognition and multimodal pathways; and (iv) human lesion studies.

### A Cortical Model for Environmental Sound and Spoken Language Processing

Several of the cortical regions activated by recognizable environmental sounds in the present study are largely consistent with, and appear to parallel, current neurological and cognitive models for how spoken language is processed. This includes input, intermediate and output processing stages (Grabowski and Damasio, 2000; Price, 2000; Binder and Price, 2001), which largely apply to words depicting a wide range of categories (Démonet *et al.*, 1992; Vandenberghe *et al.*, 1996; Price *et al.*, 1997; Bookheimer *et al.*, 1998; Lebrun *et al.*, 1998; Mummery *et al.*, 1998; Binder *et al.*, 1999; Pulvermüller, 2001; Roskies *et al.*, 2001; Grossman *et al.*, 2002; Noppeney and Price, 2002).

#### Input Stages

The blue hues in Figure 4 illustrate some of the input stages of acoustical processing reported in a study by Binder *et al.* (2000). Light blue shows cortex more responsive to tones than white noise, and dark blue represents cortex more responsive to passively heard words than tones. These data ( $n = 28$ ) show a dorsal to ventral progression along the bilateral STG related to the processing of increasingly complex acoustic structure. Much of this bilateral cortex was also activated by both the recognized and unrecognized environmental sounds of the present study (compare Fig. 2*a,b* with Fig. 4*a* blue).

Previous imaging studies involving environmental sounds, which were contrasted with a variety of different baseline conditions, have also reported activity along what might be considered input cortical sites in or near the STG. For instance, Engelien *et al.* (1995) compared passive listening to environmental sounds to a silent rest condition and similarly revealed strong bilateral PAC+ and STG activity. Giraud and Price (2001) found that middle portions of the left and right STG were more responsive to environmental sounds produced by living and nonliving things (including speech sounds) relative to white noise. Similarly, Maeder *et al.* (2001) observed middle portions of the left and right STG regions, among other sites, that responded more to environmental sound recognition than to localization of white noise stimuli. Consistent with these earlier studies, the present study showed that large extents of the bilateral PAC+ and STG were strongly activated in response to hearing and attending to a wide range of recognizable environmental sounds. Our data suggest that these various STG foci were activated in the previously mentioned studies because of differences in the acoustic complexity of the sounds (or differences in the degree of attention paid to the stimuli) relative to the control sounds (e.g. white noise). However, these input cortical STG sites appear to be relatively insensitive to the *recognition* of the sounds we presented, as they were comparably activated by the corresponding unrecognizable backward control sounds.

#### Intermediate Stages

Intermediate stages of speech processing include lexical-semantic and other associative processes. Red regions in Figure





4, from a study by Binder *et al.* (1997), depict cortex at intermediate and output processing stages, being more responsive to the comprehension of spoken words (recalling knowledge pertaining to animal names) than to processing simple tone patterns ( $n = 30$ ). Many of the brain regions involved in recognizing the environmental sounds of the present study, outside the bilateral pMTG and left SMG, showed a large degree of overlap with those involved in comprehending spoken words (Fig. 4, orange). This also included portions of subcortical structures such as the medial thalamus and caudate nuclei (not shown). Common to both studies, subjects were required to recognize the sounds and, to varying degrees, their meaning, thereby placing demands on semantic knowledge retrieval. Additionally, subjects in the present study indicated that they would 'internally' name (subvocalize in their head) many of the environmental sounds, perhaps giving them a better sense that they had accurately recognized the sounds. Together, these common task demands are likely to account for much of the overlap in terms of activation of lexical retrieval and phonological planning or short-term verbal memory processes (Price *et al.*, 1994; Paus *et al.*, 1996; Hickok *et al.*, 2000; Wise *et al.*, 2001; Binder, 2002).

Curiously, we did not observe significant differential activity for recognized environmental sounds along the middle portions of the MTG (mMTG) in either hemisphere (near or overlapping the progression of blue to purple to red cortex in Fig. 4a). Previous auditory studies of both human and non-human primates suggest that the cortex in this vicinity constitutes a major part of the ventrally directed 'what' stream (Binder *et al.*, 2000; Rauschecker and Tian, 2000; Maeder *et al.*, 2001). One possibility for this lack of differential activation was that subjects were not required to explicitly 'identify' each sound, such that there were insufficient task demands to modulate these regions. Alternatively, the lack of mMTG differential activity may reflect the stimulus properties or the category of sounds we ultimately retained in our analysis, which were biased by sound-sources that depicted manipulated objects and objects that typically have strong visual motion associations (see Semantic knowledge section below). Preliminary data pertaining to the processing of tool versus animal sounds support this latter hypothesis (Lewis and DeYoe, 2003), indicating that animal vocalizations (which were only sparsely represented in the present study) do preferentially activate the bilateral mMTG foci. Thus, animal sounds and speech sounds may be more effective stimuli for evoking the ventral temporal processing stream.

The present data appear to support the placement of the pMTG foci at intermediate, as opposed to output, processing stages. This is based in part on the location of the left pMTG, being situated between other intermediate regions for spoken language processing (Grabowski and Damasio, 2000; Price, 2000; Binder and Price, 2001). Additionally, the pMTG foci overlap cortex previously implicated in other aspects of semantic processing and in complex visual motion processing, both supporting an associative role with visual information, which is discussed in greater detail in the sections below.

In contrast to earlier studies involving environmental sounds, the bilateral activation of the pMTG foci appears to be much more pronounced in, if not unique to, the present study. Engelien *et al.* (1995) observed greater activity in middle portions of the left MTG region when subjects categorized, as opposed to passively listened to, a variety of environmental

sounds. However, their left MTG focus was located well anterior to the left pMTG focus of the present study. Their focus may relate more to the semantic processes of explicitly categorizing the sounds, which was not required of the subjects in the present study. The sound recognition study by Maeder *et al.* (2001), wherein subjects attended to animal cries amidst complex auditory scenes, revealed foci in the left and right angular gyri that appear to have included extensions of weaker activation into the vicinity of the left and right pMTG. The wider range of environmental sounds and actions that were specifically 'attended to' in the present study (e.g. manipulated tools, fluid movement and dropped objects) may explain this greater degree and extent of activation in the bilateral pMTG regions.

The left supramarginal gyrus (SMG) together with neighboring parietal cortex was also prominently activated in the environmental sound recognition paradigm (Fig. 4a), and may also represent an intermediate processing stage. Ventral portions of the left SMG have been implicated in tasks that require maintenance and manipulation of phonological information (Paulesu *et al.*, 1993; Mummery *et al.*, 1998; Binder and Price, 2001). However, in the present study the left SMG activation was situated more dorsally than in the above studies. Rather, this focus was contiguous with the large swath of activation along the left pre- and post-central cortex (cf. Fig. 2a versus 2c), which is perhaps more closely associated with the production of button responses (Burton *et al.*, 1999; Burton, 2002). Alternatively, preliminary data suggest that the left SMG activation (Brodmann area 44, or a possible homologue to monkey area 7b) may be related to audio-tactile or audio-motor associations (Lewis and DeYoe, 2003), being evoked by sounds produced by objects that are typically manipulated with the dominant (right) hand. Thus, it remains unclear whether the dorsal SMG activity of the present study was related to (i) covert phonological processing; (ii) slight differences in tactile attention, planning, or production of right hand button presses; or (iii) audio-tactile and audio-motor associations with some of the recognizable environmental sounds.

### Output Stages

The large expanse of activity in the left IFC evoked by recognizable environmental sounds (including the pars opercularis and triangularis of the inferior frontal gyrus) was consistent with representing output stages of processing. This stage involves phonological access and articulatory planning (whether vocalizations are subsequently produced or not), which can engage the left frontal operculum, left anterior insula and parietal operculum (Thompson-Schill *et al.*, 1997; Grabowski and Damasio, 2000; Price, 2000; Binder and Price, 2001). Portions of the left IFC were also preferentially activated in earlier environmental sound studies that explicitly or implicitly involved sound recognition. This includes studies where subjects would actively categorize as opposed to passively listen to environmental sounds (Engelien *et al.*, 1995), or attend to environmental sounds in contrast to white noise (Giraud and Price, 2001; Maeder *et al.*, 2001). Additionally, portions of the left IFC have also been activated with successful recognition of visual objects (Bar *et al.*, 2001). This process of identifying or naming environmental sounds or visual objects was potentially common to all the above recognition studies, and is consistent with the language processing 'output' role for the left IFC.



The right IFC may also have a role in sound recognition. In the present study, a small activation focus was present in the ventral-most portion of the right IFC (Fig. 2*d*; along the orbital sulcus), though its function remains unclear. In a study by Humphries *et al.* (2001), a larger expanse of the right IFC (and portions of the left dorsal prefrontal cortex) showed a greater response to sequences of environmental sound stimuli (e.g. a gunshot and then the sound of footsteps quickly fading into the distance) in contrast to hearing spoken sentences that described the same event. Their relatively strong activation in the right IFC may be explained by the nature of the 'nonverbal versus verbal' contrast, which is qualitatively different from the 'recognized versus unrecognized' contrast we performed. Interestingly, a separate analysis by gender of the present data did reveal significant activity along a larger extent of dorsal portions of the right IFC region, but only for males (Fig. 2*d*, yellow dotted outline). This finding appears to be at odds with earlier gender studies that suggest more bilateral processing in females than males (Shaywitz *et al.*, 1995; Jaeger *et al.*, 1998). Although issues of laterality and gender remain to be resolved, the present data do support a role for the left, and possibly right, IFC regions in environmental sound recognition and/or identification.

The posterior cingulate was activated in both the environmental sound recognition and spoken word semantics paradigm (Fig. 4*b,c*, orange). This region has been proposed to function in the retrieval of information from long-term memory (Valenstein *et al.*, 1987; Vogt *et al.*, 1992; Binder *et al.*, 1999), which may be part of a mechanism for judging whether or not a sound is recognizable or 'familiar'. Others have proposed a role for the posterior cingulate cortex in the spatial distribution of attention (Shulman *et al.*, 1997; Raichle *et al.*, 2001; Corbetta *et al.*, 2002) and processing of emotional state (Maddock, 1999). Presently, the actual role(s) and placement of the posterior cingulate within a cognitive model of sound recognition remains unclear.

### Semantic Knowledge

Portions of the left pMTG focus of the present study overlapped cortex implicated in storing knowledge associated with processing and identifying different object categories, notably including tools or artifacts. Spoken words, in addition to written words, photographs, drawings, and videos depicting different object categories have been used extensively to address issues pertaining to categorical knowledge and whether or not they are preferentially processed along different cortical pathways (Warrington and Shallice, 1984; Hillis and Caramazza, 1991; Perani *et al.*, 1995; Damasio *et al.*, 1996; Martin *et al.*, 1996; Tranel *et al.*, 1997; Mummery *et al.*, 1998; Spitzer *et al.*, 1998; Chao *et al.*, 1999; Moore and Price, 1999; Perani *et al.*, 1999; Martin, 2001; Martin and Chao, 2001; Beauchamp *et al.*, 2002; Devlin *et al.*, 2002; Grossman *et al.*, 2002). Several of these studies suggest that depictions of tools and artifacts preferentially activate a network including cortex near the left pMTG, left premotor (inferior precentral), and fusiform gyrus (for review, see Martin, 2001). The open triangles in Figure 4*a* show the Talairach coordinates for several of the reported foci (projected laterally to the outer surface for visibility) implicated in tool-related knowledge (Martin *et al.*, 1996; Mummery *et al.*, 1998; Chao *et al.*, 1999; Moore and Price, 1999; Perani *et al.*, 1999; Beauchamp *et al.*, 2002; Grèzes and Decety, 2002; Grossman *et al.*, 2002). Most of the environ-

mental sound stimuli retained in our analysis depicted objects and events that were associated with or produced by implements and tools. Though our data did not fully address whether different categories of sound-sources (e.g. tools versus animals) were processed by different networks or sub-networks, they do demonstrate that at least portions of the 'tool-related network' defined in visual- and word-related studies can also be activated by recognizable environmental sounds.

Portions of the left pMTG have also been implicated in the retrieval of 'action' knowledge. In a study by Phillips *et al.* (2002) subjects were instructed to indicate if a visually presented object (such as a tool) could, for instance, be manipulated by a twisting motion. This was in contrast to making a judgment as to the relative size of the same stimulus ('perceptual' knowledge). They reported activity in the left pMTG region (the 'X' in Fig. 4*a*) that was specific to action knowledge retrieval. Thus, activity in the left pMTG may be not be related so much to the category or categories of objects being depicted *per se*, but rather to the type of semantic knowledge (the task) that is being engaged. This finding is consistent with the present results, in that most of the environmental sounds that we presented were associated with complex movements or manipulations (ostensibly including tools and implements in action). This may account for why the left pMTG did not show much overlap with the spoken word paradigm in Figure 4 (red). The process of recognizing the wide variety of environmental sounds, in contrast to the spoken animal names, may have been placing greater demands on knowledge pertaining to how the sound was likely to have been produced (e.g. the visual or motor actions associated with the sound production).

### Relation to the Visual System

In contrast to many of the semantic and language processing studies mentioned above, the present study revealed significant activity evoked by environmental sound recognition in both the left and right pMTG regions. The bilateral representation of these activation foci together with their functional characteristics show some parallels to the visual recognition system. Additionally, their physical location, situated between auditory and visual cortex proper, was highly suggestive of a possible role in audio-visual or multimodal processing.

### Parallels between Sound and Visual Recognition Pathways

The sound recognition network revealed in the present study appears to parallel an object recognition or 'what is it?' pathway in the visual system. In humans, visual object processing is thought to follow a hierarchical progression bilaterally, from low level cues in early visual areas (i.e. V1, V2, hMT, etc.), to general object shape processing in the lateral occipital complex (near the LOS) and occipito-temporal sulcus (OTS) region, to more category-specific processing in the ventral temporal cortex, such as for faces or common objects (Malach *et al.*, 1995; Bar *et al.*, 2001; Grill-Spector *et al.*, 2001; Haxby *et al.*, 2001), and possibly to other structures along the temporal lobe (Murray and Richmond, 2001). In a similar manner, and largely consistent with earlier auditory studies (Engelien *et al.*, 1995; Clarke *et al.*, 2000; Rauschecker and Tian, 2000; Giraud and Price, 2001; Maeder *et al.*, 2001), a pathway for environmental sound recognition appears to follow a progression starting from low level input cues in early auditory cortex bilaterally (PAC, PAC+ and STG). This presumably leads to the higher-level sound recognition processing in

the bilateral pMTG plus a variety of mostly left lateralized structures implicated in semantic and/or linguistic processing. However, the actual processing hierarchy of the pMTG foci relative to these other semantic-related structures remains to be established.

The pMTG foci and OTS/LO complex may be at comparable processing levels in the auditory and visual systems, respectively, both being involved in the process of recognition. Analogous to the pMTG foci in sound recognition, the LO complex (e.g. Fig. 4a, asterisks), is preferentially activated by a wide range of identifiable visual objects and palpated objects as opposed to unidentifiable scrambled objects or palpated textures (Malach *et al.*, 1995; Amedi *et al.*, 2001; Grill-Spector *et al.*, 2001). Additionally, portions of the OTS are known to be modulated more by visual recognition success than by simple stimulus parameters (Bar *et al.*, 2001). Based on their close cortical proximity, the pMTG and OTS/LO complex in both hemispheres may be among the first cortical sites where the auditory and visual (and possibly tactile) recognition pathways can interact, and is addressed below. However, specifying the true extent of overlap and verifying multimodal response properties awaits further study.

#### *The pMTG Foci and Multimodal Processing*

The location and activation characteristics of the bilateral pMTG foci are consistent with their involvement in processing audio-visual or multimodal (or possibly supramodal) motion information for purposes of sound-source ('object') recognition. For instance, upon hearing and seeing an audio-visual event (such as a ping pong ball bouncing to rest) the temporal dynamics of the sound and motion attributes of the sight covary in time between the two sensory pathways. Based on non-human primate studies, this sensory information would, at least initially, be processed in the respective primary sensory cortices in *both* hemispheres, and then propagate along higher-level modality-specific regions (Van Essen *et al.*, 1990; Rauschecker, 1998; Kaas *et al.*, 1999). In the macaque monkey, the information from both modalities may then converge in 'association' cortices, such as cortex near the left and right posterior STS (Leinonen *et al.*, 1980; Bruce *et al.*, 1981). Homologous cortical regions in humans, possibly including the bilateral pMTG foci, may be involved in processing such stimulus-driven ('bottom up') covariance, and these cortical representations might *a priori* be expected to be present in both hemispheres.

Further illustrating a possible audio-visual link, the green regions in Figure 4 show cortex involved in visual motion discrimination during a task in which subjects assessed the speed of coherently moving dot arrays in contrast to randomly moving dots ( $n = 7$ ) (Lewis *et al.*, 2000). This paradigm activated the hMT+ (or V5+) visual motion complex, which in both hemispheres showed only a small degree of overlap with the pMTG sound recognition foci (Fig. 4c, chartreuse). However, cortical regions activated by more complex visual motion stimuli are known to be located anterior to the hMT+ complex, and thus partially overlap the pMTG foci in both hemispheres. For instance, the open circles in Figure 4a depict the Talairach coordinates reported in several studies for cortex involved in processing biological motion. This included the viewing of point-light displays defining movements of the human body or hand (Bonda *et al.*, 1996; Grossman *et al.*, 2000; Grèzes *et al.*, 2001; Grossman and Blake, 2002), and video clips showing

eye, mouth, or whole body movements (Puce *et al.*, 1998; Beauchamp *et al.*, 2002). Furthermore, portions of the left and right pMTG foci overlapped cortex activated by visual lip-reading (Fig. 4a, open squares) that was greatly enhanced when the corresponding speech sounds were present (Calvert *et al.*, 2000), suggesting that at least portions of the bilateral pMTG are involved in audio-visual *integration*.

With regard to multimodal (or supramodal) integration, normally sighted and hearing people typically become familiar with environmental sounds while simultaneously viewing and/or manipulating the sound-source (object) itself, such as pounding with a hammer or removing the cork from a bottle of wine. When subjects were asked to identify some of the forward-played sounds (replayed to them after the scan session), in some instances they would physically gesture and/or indicate that they could visualize how the sound was produced before they could provide an accurate verbal label (akin to the tip-of-the-tongue phenomenon) (Brown, 2000). One possibility is that the pMTG foci are involved in learning or mediating the stimulus-driven multimodal correlations between the sight and sound (and possibly touch and motor actions) associated with object movements, consistent with a 'sensory-functional' hypothesis for how feature information is encoded (Ettlinger and Wilson, 1990; Grossman *et al.*, 2002). Consequently, upon hearing environmental sounds in isolation the pMTG foci may be recruited, evoking a sense of recognition of the sound-source and, or by way of, its multimodal associations.

A related possibility is that the activity in the bilateral pMTG reflects processing pertaining more to multimodal associations or visual imagery subsequent to the 'initial recognition' of the sound. A comparison of the incoming sounds with stored representations may be taking place at earlier processing stages, such as in the bilateral STG regions. Any lack of differential BOLD activation (i.e. in response to the recognized versus unrecognized sounds) within such early recognition stages could have been due to processes beyond the resolution of our fMRI paradigm. For instance, differential processing may be taking place at a local neural circuit level (e.g. interspersed subpopulations of neurons responding differentially to sounds). Nonetheless, our results do suggest that the bilateral pMTG foci are involved in associating an auditory percept with stored knowledge pertaining to the likely sound-source. Thus, by this broader definition, the pMTG foci should be considered as having a role in the process of sound recognition.

#### *Human Lesions and Sound Recognition*

Lesions to the right, left, or bilateral temporal or temporal-parietal cortex can lead to severe impairments in the recognition of non-verbal, environmental sounds, while largely sparing verbal comprehension (Spreen *et al.*, 1965; Albert *et al.*, 1972; Vignolo, 1982; Fujii *et al.*, 1990; Schnider *et al.*, 1994; Engelen *et al.*, 1995; Clarke *et al.*, 2000; Saygin and Moineau, 2002). The presence of distinct bilateral pMTG foci involved in environmental sound recognition may, in part, provide a specific anatomical and functional explanation for the reported symptoms. Namely, that right hemisphere damage more commonly leads to a 'pure' auditory agnosia associated with discriminative or acoustic errors, while left hemisphere damage tends to produce more semantic-associative errors and is more likely to produce additional deficits in spoken language comprehension (Vignolo, 1982; Schnider *et al.*, 1994). The

right pMTG focus for sound recognition (Fig. 4, yellow) is sufficiently well isolated from most spoken language systems (red) that a lesion of widely varying size could disrupt environmental sound processing without also disrupting language functions. In the left hemisphere, however, only a focal lesion to portions of the pMTG cortex might selectively disrupt environmental sound processing without also disrupting the greater expanse of immediately surrounding language-related structures.

## Appendix

Environmental sounds listed in the order presented. Parentheses indicate the number of subjects for which that particular sound (forward plus backward presentation) was retained for analysis.

### Scan 1

ice dropped into glass (14)  
glass breaking 1 (19)  
rotary phone dialing (16)  
wood dropped on floor 1 (17)  
wood dropped on floor 2 (20)  
hammer hitting anvil (21)  
creaking door 1 (8)  
audience clapping (18)  
bubbles in water (13)  
footsteps (14)  
doorbell (15)  
hammer hitting nail 1 (20)  
door opening (19)  
boxing bell (12)  
glass breaking 2 (21)

### Scan 2

Polaroid camera (6)  
ping pong rally (20)  
card shuffle (14)  
striking match (4)  
air horn (17)  
scissors cutting hair (3)  
puppy barking (10)  
cannon fire (14)  
sonar ping (8)  
thunder (10)  
keys jingle & toss (7)  
machine gun (19)  
drum accent (17)  
pulling tape from dispenser (6)  
golf putt into cup 1 (16)

### Scan 3

tennis rally (19)  
typing on computer (16)  
cracking an egg (15)  
liquid filling vessel (9)  
bowling strike 1 (19)  
seagulls (9)  
chopping vegetables (14)  
toasting with glass (22)  
beating eggs (8)  
car crash (0)  
slot machine payout (8)  
forest fire (4)  
gasoline sloshing in can (19)  
door open & close (23)  
pouring beer (8)

### Scan 4

billiards break shot (22)  
revolver gun shot (19)  
draining bathtub (16)  
toaster (18)

removing a cork (21)  
manual typewriter 1 (10)  
whisky pouring into glass (19)  
hammering nail 2 (23)  
creaking door 2 (9)  
horse galloping (15)  
applying handcuffs (5)  
opening can of soda (8)  
knocking on door (23)  
locking door with key (11)  
grandfather clock (6)

### Scan 5

small gun fire (18)  
ping pong rally 2 (23)  
pool hall shot (21)  
manual typewriter 2 (13)  
glass breaking 3 (12)  
sneeze (14)  
explosion (16)  
ascending chimes (21)  
bongo drums (22)  
coins falling into drawer (11)  
water dripping in tub (22)  
chopping carrot (8)  
flipping through magazine (4)  
stapling paper (10)  
ping pong bounce to rest (21)

### Scan 6

push in coin slot (12)  
manual typewriter 3 (13)  
dribbling basketball (20)  
water dripping in metal dish (21)  
toilet flushing (5)  
paper cutter cutting (7)  
bird chirping (17)  
covering garbage can (16)  
bowling strike 2 (22)  
woman coughing (7)  
child laughter (12)  
ricocheting bullet (17)  
billiards shot into pocket (21)  
dive into swimming pool (15)  
golf putt into cup 2 (21)

### Scan 7

typewriter carriage return (7)  
filling bathtub (13)  
bite & chew chips (10)  
rooster call (3)  
tennis serve & return (14)  
shaking dice (1)  
tiger growl (3)  
basketball thru net & bounce (14)  
drink and swallow (13)  
metal file cabinet closing (14)  
racquetball rally (21)  
pig oinking (7)  
burp (14)  
unscrew & remove jar lid (12)  
parking meter (6)

## Supplementary Material

Supplementary material can be found at:  
<http://www.cercor.oupjournals.org/>.

## Notes

We thank Doug Ward for assistance with paradigm design and statistical analyses, Jennifer Junion-Dienger for assistance with acquiring sound samples, Jon Wieser, Dr David Van Essen, Donna Hanlon, and



John Harwell for assistance with cortical flattening and data presentation, and Wendy Huddleston for comments on the manuscript. This work was supported by grant R03 DC04642 to J.W.L., and grants EY10244 and MH51358 to E.A.D., and grant RR00058 to MCW.

Address correspondence to James W. Lewis, Department of Physiology and Pharmacology, PO Box 9229, West Virginia University, Morgantown, WV 26506, USA. Email: jlewis@hsc.wvu.edu.

## References

- Albert ML, Sparks R, Von Stockert T, Sax D (1972) A case study of auditory agnosia: linguistic and non-linguistic processing. *Cortex* 8:427–443.
- Amedi A, Malach R, Hendler T, Peled S, Zohary E (2001) Visuo-haptic object-related activation in the ventral visual pathway. *Nat Neurosci* 4:324–330.
- Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS (1993) Processing strategies for functional MRI of the human brain. *Magn Reson Med* 30:161–173.
- Bar M, Tootell RBH, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM (2001) Cortical mechanisms specific to explicit visual object recognition. *Neuron* 29:529–535.
- Barch DM, Braver TS, Nystrom LE, Forman SD, Noll DC, Cohen JD (1997) Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia* 35:1373–1380.
- Baumgart F, Gaschler-Markefski B, Woldorff MG, Heinze H-J, Scheich H (1999) A movement-sensitive area in auditory cortex. *Nature* 400:724–725.
- Beauchamp M, Lee K, Haxby J, Martin A (2002) Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34:149–159.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309–312.
- Binder J (2002) Wernicke's aphasia: a disorder of central language processing. In: *The neurological foundations of cognitive neuroscience* (D'Esposito ME, ed.), pp. 175–238. Cambridge, MA: MIT Press.
- Binder JR, Price CJ (2001) Functional neuroimaging of language. In: *Handbook of Functional Neuroimaging of Cognition* (Cabeza R, Kingstone A, eds), pp. 187–251. Cambridge, MA: MIT Press.
- Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17:353–362.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Rao SM, Cox RW (1999) Conceptual processing during the conscious resting state: a functional MRI study. *J Cogn Neurosci* 11:80–95.
- Binder J, Frost J, Hammeke T, Bellgowan P, Springer J, Kaufman J, Possing E (2000) Human temporal lobe activation by speech and non-speech sounds. *Cereb Cortex* 10:512–528.
- Binder J, McKiernan KA, Parsons ME, Westbury CF, Possing ET, Kaufman J, Buchanan L (2003) Neural correlates of lexical access during visual word recognition. *J Cogn Neurosci* 15:372–393.
- Bonda E, Petrides M, Ostry D, Evans A (1996) Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J Neurosci* 16:3737–3744.
- Bookheimer SY, Zeffiro TA, Blaxton TA, Gaillard WD, Malow B, Theodore WH (1998) Regional cerebral blood flow during auditory responsive naming: evidence for cross-modality neural activation. *Neuroreport* 9:2409–2413.
- Brown SR (2000) Tip-of-the-tongue phenomena: an introductory phenomenological analysis. *Conscious Cognit* 9:538–544.
- Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46:369–384.
- Burton H (2002) Cerebral cortical regions devoted to the somatosensory system: results from brain imaging studies in humans. In: *The somatosensory system: deciphering the brain's own body image* (Nelson RJ, ed.), pp. 27–72. New York: CRC Press.
- Burton H, Abend NS, MacLeod AM, Sinclair RJ, Snyder AZ, Raichle ME (1999) Tactile attention tasks enhance activation in somatosensory regions of parietal cortex: a positron emission tomography study. *Cereb Cortex* 9:662–674.
- Bushara KO, Weeks RA, Ishii K, Catalan MJ, Tian B, Rauschecker JP, Hallett M (1999) Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat Neurosci* 2:759–766.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
- Chao LL, Haxby JV, Martin A (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci* 2:913–919.
- Clarke S, Bellmann A, Meuli RA, Assal G, Steck AJ (2000) Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. *Neuropsychologia* 38:797–807.
- Clarke S, Thiran AB, Maeder P, Adriani M, Vernet O, Regli L, Cuisenaire O, Thiran J-P (2002) What and where in human audition: selective deficits following focal hemispheric lesions. *Exp Brain Res* 147:8–15.
- Corbetta M, Kincade JM, Shulman GL (2002) Neural systems for visual orienting and their relationships to spatial working memory. *J Cogn Neurosci* 14:508–523.
- Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio RD (1996) A neural basis for lexical retrieval. *Nature* 380:499–505.
- Démonet J-F, Chollet F, Ramsay S (1992) The anatomy of phonological and semantic processing in normal subjects. *Brain* 115:1752–1768.
- Devlin JT, Russell P, Davis MH, Price CJ, Moss HE, Fadili MJ, Tyler LK (2002) Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia* 40:54–75.
- Engelen A, Silbersweig D, Stern E, Huber W, Wolfgang D, Frith C, Frackowiak RSJ (1995) The functional anatomy of recovery from auditory agnosia: a PET study of sound categorization in a neurological patient and normal controls. *Brain* 118:1395–1409.
- Ettlinger G and Wilson, WA (1990) Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms. *Behav Brain Res* 40:169–192.
- Fujii T, Fukatsu R, Watabe S, Ohnuma A, Teramura K, Kimura I, Saso S, Korgure K (1990) Auditory sound agnosia without aphasia following a right temporal lobe lesion. *Cortex* 26:263–268.
- Giraud AL, Price CJ (2001) The constraints functional neuroimaging places on classical models of auditory word processing. *J Cogn Neurosci* 13:754–765.
- Grabowski TJ, Damasio AR (2000) Investigating language with functional neuroimaging. In: *Brain mapping: the systems* (Toga AW, Mazziotta JC, eds), pp. 425–462. New York: Academic Press.
- Grèzes J, Decety J (2002) Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia* 40:212–222.
- Grèzes J, Fonlupt P, Bertenthal B, Delon-Martin C, Segebarth C, Decety J (2001) Does perception of biological motion rely on specific brain regions? *Neuroimage* 13:775–785.
- Griffiths TD, Rees A, Witton C, Shakir RA, GB H, Green GGR (1996) Evidence for a sound movement area in the human cerebral cortex. *Nature* 383:425–427.
- Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. *Vision Res* 41:1409–1422.
- Grossman ED, Blake R (2002) Brain areas active during visual perception of biological motion. *Neuron* 35:1167–1175.
- Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R (2000) Brain areas involved in perception of biological motion. *J Cogn Neurosci* 12:711–720.
- Grossman M, Koenig P, DeVita C, Glosser G, Alsop D, Detre J, Gee J (2002) The neural basis for category-specific knowledge: an fMRI study. *Neuroimage* 15:936–948.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.

- Hickok G, Erhard P, Kassubek J, Helms-Tillery A, Naeve-Velguth S, Strupp J, Strick P, Ugurbil K (2000) A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. *Neurosci Lett* 287:156–160.
- Hillis AE, Caramazza A (1991) Category-specific naming and comprehension impairment: a double dissociation. *Brain* 114:2081–2094.
- Humphries C, Willard K, Buchsbaum B, Hickok G (2001) Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Proc Natl Acad Sci U S A* 12:1749–1752.
- Jaeger J, Lockwood A, Van Valin R, Kemmerer D, Murphy B, Wack D (1998) Sex differences in brain regions activated by grammatical and reading tasks. *Neuroreport* 9:2803–2807.
- Kaas JH, Hackett TA, Tramo MJ (1999) Auditory processing in primate cerebral cortex. *Curr Opin Neurobiol* 9:164–170.
- Lebrun N, Clochon P, Etevenon P, Baron JC, Eustache F (1998) Effect of environmental sound familiarity on dynamic neural activation/inhibition patterns: an ERD mapping study. *Neuroimage* 8:79–92.
- Leinonen L, Hyvärinen J, Sovijärvi ARA (1980) Functional properties of neurons in the temporo-parietal association cortex of awake monkey. *Exp Brain Res* 39:203–215.
- Lewis JW, DeYoe EA (2003) Animal vs tool sounds evoke different activation patterns in right vs left handed subjects. *Human Brain Map Abstr*. 1488.
- Lewis JW, Beauchamp MS, DeYoe EA (2000) A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb Cortex* 10:873–888.
- Lewis JW, Wightman FL, Junion-Dienger JL, DeYoe EA (2001) FMRI Activation in response to the identification of natural sounds. *Soc Neurosci Abstr* 27.
- Maddock RJ (1999) The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends Neurosci* 22:310–316.
- Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S (2001) Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14:802–816.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RBH (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135–8139.
- Martin A (ed.) (2001) Functional neuroimaging of semantic memory. In: *Handbook of Functional Neuroimaging of Cognition* (Cabeza R, Kingstone A, eds), pp. 153–186. Cambridge, MA: MIT Press.
- Martin A, Chao LL (2001) Semantic memory and the brain: structure and processes. *Curr Opin Neurobiol* 11:194–201.
- Martin A, Wiggs CL, Ungerleider LG, Haxby JV (1996) Neural correlates of category-specific knowledge. *Nature* 379:649–652.
- Mishkin M, Ungerleider LG, Macko KA (1983) Object vision and spatial vision: two cortical pathways. *Trends Neurosci* 6:414–417.
- Moore CJ, Price CJ (1999) A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain* 122:943–962.
- Mummery CJ, Patterson K, Hodges JR, Price CJ (1998) Functional neuroanatomy of the semantic system: divisible by what? *J Cogn Neurosci* 10(6):766–777.
- Murray, EA and Richmond, BJ (2001) Role of perirhinal cortex in object perception, memory, and associations. *Curr Opin Neurobiol* 11:188–193.
- Noppeney U, Price C (2002) A PET study of stimulus- and task-induced semantic processing. *Neuroimage* 15:927–935.
- Ogawa S, Menon RS, Tank DW, Kim S-G, Merkle H, Ellerman JM, Ugurbil K (1993) Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. *Biophys J* 64:803–812.
- Paulesu E, Frith CD, Frackowiak RS (1993) The neural correlates of the verbal component of working memory. *Nature* 362:342–345.
- Paus T, Perry D, Zatorre R, Worsley K, Evans A (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur J Neurosci* 8:2236–2246.
- Perani D, Cappa SF, Bettinardi V, Bressi S, Gorno-Tempini M, Matarrese M, Fazio F (1995) Different neural systems for the recognition of animals and man-made tools. *Neuroreport* 21:1637–1641.
- Perani D, Schnur T, Tettamanti M, Gorno-Tempini M, Cappa SF, Fazio F (1999) Word and picture matching: a PET study of semantic category effects. *Neuropsychologia* 37:293–306.
- Phillips JA, Noppeney U, Humphreys GW, Price CJ (2002) Can segregation within the semantic system account for category-specific deficits? *Brain* 125:2067–2080.
- Price CJ (2000) The anatomy of language: contributions from functional neuroimaging. *J Anat* 197:335–359.
- Price CJ, Wise RJS, Watson JDG, Patterson K, Howard D, Frackowiak RSJ (1994) Brain activity during reading. The effects of exposure duration and task. *Brain* 117:1255–1269.
- Price CJ, Moore CJ, Humphreys GW, Wise RJS (1997) Segregating semantic from phonological processes during reading. *J Cogn Neurosci* 9:727–733.
- Puce A, Allison T, Bentin S, Gore JC, McCarthy G (1998) Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci* 18:2188–2199.
- Pulvermüller F (2001) Brain reflections of words and their meaning. *Trends Cogn Sci* 5:517–524.
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001) A default mode of brain function. *Proc Natl Acad Sci U S A* 98:676–682.
- Rauschecker JP (1998) Cortical processing of complex sounds. *Curr Opin Neurobiol* 8:516–521.
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of ‘what’ and ‘where’ in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800–11806.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131–1136.
- Roskies AL, Fiez JA, Balota DA, Raichle ME, Petersen SE (2001) Task-dependent modulation of regions in the left inferior frontal cortex during semantic processing. *J Cogn Neurosci* 13:829–843.
- Saygin AP, Moineau S (2002) Auditory agnosia with preserved verbal comprehension after unilateral left hemisphere lesion involving Wernicke’s area. *Soc Neurosci Abstr* 28.
- Saygin AP, Dick F, Wilson SW, Dronkers NF, Bates E (2003) Neural resources for processing language and environmental sounds. *Brain* 126:928–945.
- Schnider A, Benson DF, Alexander DN, Schnider-Klaus A (1994) Non-verbal environmental sound recognition after unilateral hemispheric stroke. *Brain* 117:281–287.
- Shaywitz B, Shaywitz S, Pugh K, Constable R, Skudlarski P, Fulbright R, Bronen R, Fletcher J, Shankweiler D, Katz L (1995) Sex differences in the functional organization of the brain for language. *Nature* 373:607–609.
- Shulman GL, Fiez JA, Corbetta M, Buckner RL, Meizen FM, Raichle ME (1997) Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *J Cogn Neurosci* 9:648–663.
- Spitzer M, Kischka U, Guckel F, Bellemann ME, Kammer T, Seyyedi S, Weisbrod M, Schwartz A, Brix G (1998) Functional magnetic resonance imaging of category-specific cortical activation: evidence for semantic maps. *Cogn Brain Res* 6:309–319.
- Spreen O, Benton AL, Fincham RW (1965) Auditory agnosia without aphasia. *Arch Neurol* 13:84–92.
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.
- Thivard L, Belin P, Zilbovicius M, Poline JB, Samson Y (2000) A cortical region sensitive to auditory spectral motion. *Neuroreport* 11:2969–2972.
- Thompson-Schill SL, D’Esposito M, Aguirre GK, Farah MJ (1997) Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc Natl Acad Sci U S A* 94:14792–14797.
- Tranel D, Logan CG, Frank RJ, Damasio AR (1997) Explaining category-related effects in the retrieval of conceptual and lexical knowledge for concrete entities: operationalization and analysis of factors. *Neuropsychologia* 35:1329–1339.

- Valenstein E, Bowers D, Verfaellie M, Heilman KM, Day A, Watson RT (1987) Retrosplenial amnesia. *Brain* 110:1631–1646.
- Van Essen DC (2004) Organization of visual areas in macaque and human cerebral cortex. In: *Visual neuroscience* (Chalupa L, Werner J, eds), pp. 507–521.
- Van Essen DC, Felleman DJ, DeYoe EA, Olavarria J, Knierim J (1990) Modular and hierarchical organization of extrastriate visual cortex in the macaque monkey. *Cold Spring Harb Symp Quant Biol* LV:679–696.
- Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. *J Am Med Informatics Assoc* 8:443–459.
- Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RS (1996) Functional anatomy of a common semantic system for words and pictures. *Nature* 383:254–256.
- Vignolo LA (1982) Auditory agnosia. *Phil Trans R Soc Lond B* 298:49–57.
- Vogt BA, Finch DM, Olson CR (1992) Functional heterogeneity in cingulate cortex: the anterior executive and posterior evaluative regions. *Cereb Cortex* 2:435–443.
- Warren J, Zielinski B, Green G, Rauschecker J, Griffiths T (2002) Perception of sound-source motion by the human brain. *Neuron* 34:139–148.
- Warrington EK, Shallice T (1984) Category specific semantic impairments. *Brain* 107:829–854.
- Weeks RA, Aziz-Sultan A, Bushara KO, Tian B, Wessinger CM, Dang N, Rauschecker JP, Hallet M (1999) A PET study of human auditory spatial processing. *Neurosci Lett* 262:155–158.
- Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 13:1–7.
- Wise R, Scott S, Blank S, Mummery C, Murphy K, Warburton E (2001) Separate neural subsystems within 'Wernicke's area'. *Brain* 124:83–95.
- Zatorre R, Penhune V (2001) Spatial localization after excision of human auditory cortex. *J Neurosci* 21:6321–6328.