Cortical Organization of Environmental Sounds by Attribute

Julia Hocking,^{1*} Katie L. McMahon,¹ and Greig I. de Zubicaray²

¹The University of Queensland, Centre for Advanced Imaging, Brisbane, Australia

²University of Queensland, School of Psychology, Brisbane, Australia

Abstract: Semantic knowledge is supported by a widely distributed neuronal network, with differential patterns of activation depending upon experimental stimulus or task demands. Despite a wide body of knowledge on semantic object processing from the visual modality, the response of this semantic network to environmental sounds remains relatively unknown. Here, we used fMRI to investigate how access to different conceptual attributes from environmental sound input modulates this semantic network. Using a range of living and manmade sounds, we scanned participants whilst they carried out an object attribute verification task. Specifically, we tested visual perceptual, encyclopedic, and categorical attributes about living and manmade objects relative to a high-level auditory perceptual baseline to investigate the differential patterns of response to these contrasting types of object-related attributes, whilst keeping stimulus input constant across conditions. Within the bilateral distributed network engaged for processing environmental sounds across all conditions, we report here a highly significant dissociation within the left hemisphere between the processing of visual perceptual and encyclopedic attributes of objects. *Hum Brain Mapp* 32:688–698, 2011. © 2011 Wiley-Liss, Inc.

Key words: fMRI; auditory; object; semantic; perceptual; encyclopedic

INTRODUCTION

Functional neuroimaging of the organization of object knowledge has provided rich data on the multiple ways in which different perceptual or semantic attributes of objects modulate patterns of brain activation. Much of this data has been derived from studies of visual or spoken words and pictures. By contrast, far less is known about the processing and organization of attributes related to environmental sounds. Because object knowledge is derived from

the interaction of multiple senses, characteristic object sounds will make a significant contribution to object knowledge and representation in the brain. Our study is designed to investigate neuronal responses when attention is directed to different attributes of real object sounds to further understand the neuroanatomical organization of semantic knowledge.

To date, the investigation of environmental sound processing at a conceptual level has focused predominantly on either the difference between meaningful and meaningless sounds or on categorical distinctions between knowledge types, e.g., living vs. nonliving categories: [Doehrmann et al., 2008; Engel et al., 2009; Kraut et al., 2006; Lewis et al., 2004]. Contrasting meaningful relative to meaningless sounds is important, because it allows the identification of regions preferentially engaged by auditory object conceptual processing relative to lower-level auditory sensory input processing. For example, Lewis et al. [2004] contrasted recognized environmental sounds with unrecognizable reversed sounds to identify the network involved in the process of environmental sound

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Julia Hocking, The University of Queensland, Centre for Advanced Imaging, St Lucia, Queensland 4072, Australia. E-mail: julia.hocking@cmr.uq.edu.au

Received for publication 22 October 2009; Revised 1 February 2010; Accepted 8 February 2010

DOI: 10.1002/hbm.21040

Published online 9 March 2011 in Wiley Online Library (wileyonlinelibrary.com).

recognition. They reported activation in a network of regions, including bilateral posterior middle temporal gyri and posterior superior temporal sulcus, as well as left lateralized activation in inferior frontal cortex, angular gyrus, posterior cingulate, anterior fusiform, and the supramarginal gyrus. Similarly, Engelien et al. [2006] reported greater activation in bilateral superior temporal, left inferior frontal gyrus, left parahippocampal gyrus, and right superior mid-orbital frontal regions when participants listened to meaningful relative to meaningless sounds. The predominant left > right hemisphere pattern of activation is consistent with neuroimaging data that has used verbal (spoken word) input to access meaning, suggesting access from both verbal and nonverbal auditory inputs to a shared semantic system subserved by the left hemisphere. However, this type of manipulation does not allow us to differentiate between the different types of semantic or perceptual information that can be accessed from the same meaningful input.

A further set of studies have manipulated categorical differences. This is in line with the previous work from both functional neuroimaging and neuropsychological data that have distinguished between differential processing of manmade versus natural items or tool versus animal knowledge, using pictorial or lexical input stimuli. For example, Engel et al. [2009] reported distinct activation patterns associated with living versus nonliving sounds, as well as within-category differences between human, animal, mechanical, and environmentally produced sounds. In addition to looking at categorical differences between animals and nonliving objects, Kraut et al. [2006] investigated the threatening versus nonthreatening valence of these different categories of sounds. Two right superior temporal regions responded differentially to threatening animal stimuli, suggesting sensitivity to semantic object features in addition to categorical processing differences between living and nonliving items in this region.

Although rare, environmental sound processing can be selectively impaired in patients with nonverbal auditory agnosia [Fujii et al., 1990; Habib et al., 1995; Spreen et al., 1965; Taniwaki et al., 2000], with further evidence for auditory agnosia selective for verbal input (pure word deafness: [Auerbach et al., 1982; Coslett et al. 1984; Di Giovanni et al., 1992; Metz-Lutz and Dahl, 1984; Takahashi et al., 1992; Tanaka et al., 1987; Yaqub et al., 1988] providing a double dissociation between these verbal and nonverbal recognition abilities. This suggests distinct processing mechanisms for these different types of auditory input (verbal versus nonverbal). Auditory agnosia for verbal material may follow a pattern of either bilateral or left superior temporal lesions [Auerbach et al., 1982; Coslett et al., 1984; Hickok and Poeppel, 2007], whereas auditory agnosia for nonverbal material generally follows right hemisphere damage [Fujii et al., 1990; Spreen et al., 1965]. However, it is likely that both hemispheres contribute to environmental sound processing, because environmental sounds require not only nonverbal analysis of tone

sequences but also the attribution of meaning. For example, Schnider et al. [1994] reported that patients with right hemisphere lesions were impaired on an acoustic recognition task (e.g., hearing a sound of a baby crying and incorrectly matching this with a picture of a cat), whereas patients with left hemisphere damage were impaired on a semantic recognition task (hearing a baby crying and incorrectly matching this with a picture of a baby laughing). This left lateralized environmental sound semantic processing deficit was comorbid with a more generalized semantic deficit in language comprehension and picture matching. Thus, they concluded that the right hemisphere is necessary for perceptual sound discrimination, whereas the left is involved when attributing meaning to those sounds.

An inherent difficulty in examining categorical differences between auditory stimuli is that activation observed for one category relative to another could be attributable to low-level acoustic properties, in particular, when processing living versus manmade sounds. Indeed, Lewis et al. [2005] explored these potential differences using measures of harmonic content and temporal dynamics of stimulus sounds. They concluded that activation in the middle superior temporal gyri observed for animal (right lateralized) and tool (left lateralized) sounds may have been modulated by differences between the low-level acoustic properties of these sounds. Multiple different studies have consistently shown that the anatomicofunctional organization of object knowledge along the dimension of category is inconsistent and incomplete in that it is modulated by task as well as stimulus differences [Devlin et al., 2002; Gauthier, 2000; Gorno-Tempini et al., 2000; Moore and Price, 1999; Moss et al., 2005; Mummery et al., 1998; Noppeney and Price, 2002; Price et al., 2003; Rogers et al., 2005; Tyler et al., 2000, 2004]. However, a right (environmental sound) versus left (spoken words) hemispheric dissociation has been reported that did control for low-level acoustic properties of sounds [Thierry et al., 2003], suggesting fundamental differences in the way that meaning derived via environmental sound input accesses semantic knowledge compared with spoken word input.

The novel question we ask here is how environmental, nonverbal sound stimuli modulate this shared semantic network whilst keeping task (attribute verification) and stimuli (living and manmade sounds) constant across conditions. This ensures control of potential confounds between low-level acoustic and perceptual properties of the stimuli and allows identification of regional responses to different information types about the same concepts. We used fMRI to scan participants while they carried out a stimulus attribute verification task. In line with object processing models supporting a neuronal network organizationally weighted by properties most salient to the type of knowledge being accessed and not constrained by category, we hypothesized that access to different types of attributes pertaining to a range of environmental sounds would place differential demands on the object processing

network. Specifically, we probed visual perceptual, encyclopedic, and categorical attribute knowledge about sounds relative to an auditory perceptual verification task to determine the differential patterns of response to these contrasting types of object-related attributes.

MATERIALS AND METHODS

Participants

Thirteen native English speakers (6 females; mean age, 26.5 years; range, 20–34 years) gave written informed consent to take part. All had normal or corrected to normal vision, normal hearing, were right-handed and were free of any history of neurological problems. The study was approved by The University of Queensland's Medical Research Ethics Committee, and all participants were paid a gratuity to cover their expenses.

Experimental Design and Stimuli

Stimuli

Auditory stimuli were 40 environmental sounds of 20 animals and 20 manmade objects (1,500 ms duration, 16 bit mono, 44,100 Hz; see Supporting Information for a complete list). As a measure of low-level acoustic properties, Praat software (www.praat.org) was used to extract the harmonic-to-noise ratio for each sound. A subsequent *t*-test between living and manmade sounds revealed no difference between the two groups ($t = 0.499$, $P = 0.623$). These stimuli were then divided into equal groups: animals with or without fur and manmade items predominantly made of metal or not. Each of these four groups was further subdivided into equal groups of animals that did or did not live in Australia and manmade items that were or were not a tool. Within these groups, objects did not differ on the measure of frequency, as determined by the log-transformed hyperspace analogue to language corpus norms ($t = -2.68$, $P = 0.79$) [Lund and Burgess, 1996]. An additional set of 16 sounds (eight living and eight manmade) was used for a practice task before scanning. Environmental sounds were downloaded from the Internet, with the majority coming from www.sounddogs.com. For the auditory perceptual condition, sounds were manipulated using Audacity (version 1.2.6; <http://audacity.sourceforge.net/>) to either fade in (get louder) or fade out (get quieter) over the first (fade in) or final (fade out) 500 ms.

Design

In each experimental block, participants viewed a written question projected onto a screen at the foot of the scanner bed via a mirror mounted on the head-coil and heard a set of five sounds presented through MR-compatible piezoelectric headphones. There were four conditions in total

(visual perceptual, encyclopedic, categorical, and auditory perceptual). Within each of these four conditions, there were two possible question types designed to reflect the attributes relating to each condition. This resulted in a total of eight different trial types:

- Visual perceptual: (a) have fur? (b) made of metal?
- Encyclopedic: (a) live in Australia? (b) a tool?
- Categorical: (a) living? (b) manmade?
- Auditory perceptual: (a) louder? (b) quieter?

The task was to indicate via a two-choice (“yes” or “no”) key-press response whether each sound was associated with the attribute that corresponded to the question. Before scanning, participants were familiarized with all 40 sound stimuli by hearing them through headphones, in random order, three times. Participants confirmed when questioned that they recognized all sounds and were familiar with the items before the commencement of scanning. They then completed a practice run to familiarize them with the in-scanner task, using the additional set of 16 stimuli. Each question type in this practice session was allocated two environmental sound stimuli, so that the participants could familiarize themselves with the “yes” or “no” key press response for each question type. Participants were instructed to respond to the question as quickly and as accurately as possible during scanning. Because manmade items are generally manufactured from a range of materials, participants were instructed to consider whether the predominant material was metal or not, which, in turn, would depend upon their own experience with that item. For example, a doorbell could be considered predominantly metal or plastic depending upon a person's own experience. Importantly, this difference between participants' knowledge does not confound the difference between the conditions of interest, as the question retains the attribute quality “visual perceptual” independent of the individual's experience with that specific item.

The eight trial types were blocked, with five trials per block. Trial duration was 3.5 s [1.5 s for each sound, followed by 2 s to allow for a response (see Fig. 1)]. The question was presented for 3 s at the beginning of each block, giving a total block time of 20.5 s. Blocks were followed by a blank screen, the duration of which randomly alternated between equal multiples of 3, 4, 5, or 6 × TR (resolution time, 2.1 s). Stimulus presentation within a block was pseudorandomized, so that there were two *yes* responses and two *no* responses randomized for the first four trials in every block and either a *yes* or *no* response for the fifth trial to avoid participants predicting the final response. In each of the four scanning sessions, there were two blocks of trials for each question type (i.e., the same question was presented twice in each session), for each of the four conditions. This gave a total of 80 trials (16 blocks) per session and 320 trials (64 blocks) over the four sessions. Every environmental sound was presented twice

Figure 1.

Example of one experimental block for the visual perceptual (“have fur?”) condition. Following initial presentation of a 3-s fixation cross, sounds of 1.5-s duration were presented every 3.5 s (five sounds per block). The question remained on the screen throughout the block. In this condition, participants gave a “yes” (the animal has fur) or a “no” (the animal does not have fur) response with a key press.

in every condition for each of the four sessions. No two blocks comprised the same set of five items. For conditions (1) and (2), blocks comprised either animals (have fur? live in Australia?) or manmade objects (made of metal? a tool?). For conditions (3) and (4), animals and manmade objects were intermixed in each block. The order of the conditions (eight question types) was counterbalanced within and across participants. Left and right key—press for “yes” and “no” response—was likewise counterbalanced across participants.

Image Acquisition

Participants were scanned on a Bruker Medspec 4T system equipped with a transverse electromagnetic head coil [Vaughan et al., 2002]. A point-spread function mapping sequence was acquired before the echo planar imaging (EPI) acquisitions to correct geometric distortions in the

EPI data [Zeng and Constable, 2002]. Functional images used a T2*-weighted EPI sequence for blood oxygen level dependent contrast. The imaging parameters were as follows; TE = 30 ms, TR = 2,100 ms, and FOV (field of view) = 230 × 230. In a single acquisition, 36 slices were acquired, each 3-mm thick with a 0.6-mm gap between the slices. The first five volumes were discarded to allow for T1 equilibration effects. Between the four sessions (i.e., after two sessions were completed), a high-resolution 3D T1-weighted high-resolution image was acquired using an MP-RAGE sequence (TI = 900 ms, TR = 2,300 ms, TE = 2.94 ms, 256 × 256 × 176 matrix, and 0.9-mm³ isotropic voxels).

Data Analysis

Rigid-body motion correction was carried out on all EPI series using INRIAlign [Freire et al., 2002], and a mean

realigned image was created for each participant. This mean image was then coregistered to the corresponding structural (T1) image, using SPM5 (Wellcome Trust Centre for Neuroimaging, London). Each individual's T1 image was then normalized to the SPM5 MNI T1 template using the unified segmentation procedure [Ashburner and Friston, 2005]. The resulting spatial normalization parameters were applied to the EPI time series data and resliced to $3 \times 3 \times 3$ mm voxels. Images were then spatially smoothed with a full width half maximum Gaussian kernel of 9 mm. A mean T1 image was created from all individuals, using the Imcalc function in SPM5, for displaying the resulting statistical parametric maps.

First level (fixed effects) statistical analyses modeled each question type independently by convolving the onset times for each response with the hemodynamic response function [Mechelli et al., 2003] and high-pass filtered using a set of discrete cosine basis functions with a cut-off period of 128 s. Trials with omitted responses or responses meeting the exclusion criteria were similarly modeled as nuisance covariates (see Behavioral Results below). Parameter estimates were calculated for all voxels using the general linear model by computing a contrast image for each of the eight questions (two for each of the four conditions). The parameter estimates for each participant were then entered into a second-level random effects ANOVA that enabled the identification of the main effect of all conditions relative to the implicit baseline, the effect of all conditions relative to the perceptual baseline task, and the effect of each condition relative to all others. Any difference between each question type within each condition was also tested. Using a whole-brain height threshold of $P < 0.001$, uncorrected, we report only those regions that reached a level of significance family-wise error (FWE) corrected for voxel-wise comparisons across the whole brain at $P < 0.05$, with a minimum expected cluster size (spatial extent) of 10.155 voxels in this dataset.

RESULTS

Behavioural Results

Reaction times were analyzed using a repeated measures ANOVA modeling the four conditions. Trials with omitted responses were removed from the behavioral analysis, as were responses with latencies faster than 300 ms or slower than 3,500 ms, resulting in 1.3% (56 responses from a total of 4,160) being discarded. Although responses are to some degree subjective (see Methods section), we also removed erroneous responses. To deal with the subjective nature of certain questions, we first determined whether a person responded consistently for a particular object (e.g., "made of metal" for a camera was always a "Yes"-response) and retained those response latencies. If, however, they responded inconsistently to a particular object within a question type across the Experiment (i.e., the four scanning sessions), we considered that event to be

TABLE I. Means and standard deviations for response times and errors by condition

	Response times		% Errors	
	Mean	SD	Mean	SD
Perceptual	1,611	278	8.6	3.6
Encyclopedic	1,591	257	4.3	3.1
Categorical	1,307	284	0.9	0.2
Baseline	1,683	249	0.5	0.3

erroneous. Because any one question for an individual item was shown only twice throughout the experiment, one response would thus be considered an error and removed from the analysis and one would be retained (i.e., experimenter bias on judgment of what constitutes a correct or incorrect response does not impact the error results). We then analyzed the response latencies and percentage errors for each condition with two repeated measures ANOVAs. Means and standard deviations are shown in Table I. The response latency ANOVA identified a significant main effect of condition ($F[36,3] = 43.649$, $P < 0.0005$). Post hoc analysis revealed significantly faster responses for the categorical condition relative to all others (visual perceptual: $t = 11.082$; encyclopedic: $t = 10.553$; auditory perceptual: $t = -11.587$, all P 's < 0.0005). There was no significant difference between auditory perceptual relative to visual perceptual ($t = -1.893$, $P = 0.083$) or encyclopedic ($t = -1.908$, $P = 0.081$) conditions, nor did perceptual and encyclopedic response latencies differ ($t = 0.549$, $P = 0.593$). An analysis of percent errors identified a main effect of condition ($F[3,36] = 43.29$; $P < 0.0005$). Post hoc t -tests revealed that all conditions were significantly different from each other, with the visual perceptual condition containing the most errors and the auditory perceptual the least (see Table I).

Imaging Results

All four conditions combined increased activation in a bilateral network of distributed regions (Fig. 2a). A predominantly left lateralized subset of these regions were activated at an FWE corrected level for the three-object attribute questions relative to the auditory-perceptual task. These regions included the calcarine gyrus extending into the precuneus, the fusiform gyrus, angular gyrus, middle temporal gyrus, inferior, and superior frontal gyrus. The only right lateralized activation was observed in the inferior and anterior cerebellum. Co-ordinates and Z-scores of all regions are listed in Table II.

To detect responses specific to each attribute type, we contrasted activation for each semantic attribute condition relative to all other conditions [e.g., visual perceptual $>$ (encyclopedic + categorical + auditory perceptual)]. Visual perceptual questions increased activation in the left inferior temporal lobe, with two distinct cluster peaks in the

Figure 2.

fMRI results and effect sizes. (a) Activation for all four conditions, rendered on the SPM standard surface model. (b) Increased activation for visual perceptual verification relative to all other conditions, rendered at $P < 0.001$ uncorrected on the mean T1-weighted image from all participants with plot of mean centered effect sizes for the lateral fusiform ($-48, -54, -12$) and medial fusiform ($-30, -36, -18$). (c) Regions significantly activated for the encyclopedic verification condition relative to all others in

precuneus ($-12, -60, 21$), medial superior frontal gyrus ($-15, 33, 45$), and angular gyrus ($-45, -72, 39$), rendered at $P < 0.001$ uncorrected on the mean subject T1-weighted image. Key: V, visual perceptual condition; F, fur question; M, metal question; E, encyclopedic condition; A, Australia; T, tool; C, categorical condition; L, living; N, nonliving/manmade; A, auditory perceptual conditions; U, up/louder; D, down/quieter.

lateral and medial fusiform gyrus (Fig. 2b). We also tested the simple effect of question type within this condition (have fur? Vs. made of metal?), which revealed no difference between conditions in either direction.

Encyclopedic attribute questions relative to all others increased activation in the left medial superior frontal

gyrus, left precuneus, and left angular gyrus (Fig. 2c). As shown in the plot of effect size for each of these regions (Fig. 2c), the effect appeared to be driven by responses to living things (i.e., the question “live in Australia?” relative to “a tool?”), with relative decreases for all other conditions. However, on interrogation, the simple t -contrast

TABLE II. Anatomical regions, co-ordinates, and Z-scores for all conditions relative to the implicit baseline and for each condition relative to all others

	<i>x</i>	<i>y</i>	<i>z</i>	Z-score
All > auditory perceptual baseline				
L calcarine gyrus/precuneus	-6	-51	12	6.7
L fusiform gyrus	-27	-33	-18	6.5
L inf frontal gyrus	-36	24	-15	6.3
L sup frontal gyrus	-15	33	45	6.2
L angular gyrus	-45	-72	39	5.8
R inf cerebellum	39	-69	-39	5.0
R ant cerebellum	6	-51	-42	4.7
L mid temporal gyrus	-60	-15	-15	4.7
Visual perceptual > all others				
L inf temp/fusiform	-48	-54	-12	5.9
L ant fusiform	-30	-36	-18	5.6
Encyclopedic > all others				
L sup frontal gyrus	-15	33	45	5.6
L precuneus	-12	-60	21	5.4
L angular gyrus	-45	-72	39	4.9

Key: L, left; R, right; ant, anterior; inf, inferior; mid, middle; sup, superior.

between these two question types (Australia vs. tool) revealed no significant difference in either direction, at a corrected level. In the co-ordinates reported for (Encyclopedic > all others), we observed the following Z-scores for the simple within condition contrast of (Australia > tool): left medial superior frontal, $Z = 1.2$; left precuneus, $Z = 4.4$; left angular gyrus, $Z = 1.9$.

No regions were detected at a corrected level threshold for categorical-attribute questions relative to all others, and there was no difference between these question types (living? vs. manmade?). For co-ordinates and Z-scores of activated regions for each contrast, see Table II.

DISCUSSION

These data revealed distinct and highly significant patterns of neuronal activation associated with access to different types of attribute information in response to the same environmental sound input. Together, all conditions activated a bilateral network of activation in ventral occipitotemporal, superior temporal, parietal, and frontal regions. These regions represent the distributed auditory object processing network from early sensory processes through access to stored conceptual representations and task-related executive processes. Conditions requiring access to stored knowledge about these objects, relative to a perceptual baseline task, were predominantly left lateralized with right cerebellar activation. Most interesting was the clear dissociation occurring with access to different types of attributes. Visual perceptual questions increased activation in the ventral temporal lobe. In contrast, encyclopedic questions were associated with left lateralized

activation in the medial superior frontal gyrus, angular gyrus, and the precuneus. No regions were more activated for the categorical verification task. We will first discuss each of these information types and then discuss how this relates to neuropsychological data from patients with non-verbal auditory agnosia.

Questions tapping visual perceptual knowledge from auditory concepts increased activation in the left fusiform gyrus—a region that has been consistently associated with visual object and amodal semantic processing (for a review, see Martin [2007]). Unimodal, meaningful auditory information is thought to quickly converge onto a shared, modality-independent semantic system. This is illustrated by crossmodal priming tasks, where environmental sound primes elicit faster response times to subsequently presented visual targets, compared with incongruent primes [Schneider et al., 2008]. The data here suggest that the fusiform region either stores information about conceptual (visual) attributes accessible via unimodal sound input or is engaged by explicit visual mental imagery processes required for verifying the presence or absence of the visual attribute (i.e., fur or metal). The design of this study does not permit us to adjudicate between these possible differences nor are they mutually exclusive explanations (cf. [Martin, 2007] for a discussion of this issue). Moreover, these visual attributes (fur and metal) also have a tactile component, which could be used as an additional strategy for this task. Supporting evidence for a role of the inferior occipitotemporal cortex in nonvisual discrimination comes from studies of congenitally blind individuals [Buchel et al., 1998] and is also implicated in the convergence of tactile and visual object information in sighted individuals ([Pietrini et al., 2004; for a review, see Amedi et al. [2005]). Here, we conclude that the left fusiform is engaged when access to stored visual knowledge about objects is required from environmental sound inputs, as predicted by models proposing that visual information is stored in the sensory regions that first acquired that knowledge. Importantly, we demonstrate that this is also true for nonverbal auditory input and that the effect was specific to the visual perceptual task, because significant activation was not observed in this region for the other attribute verification conditions.

It is perhaps surprising that no differential activation was observed for the distinction between living and man-made concepts within the visual perceptual condition. This is particularly relevant for the ventral occipitotemporal regions as living and manmade items have been strongly associated with activation in the lateral and medial fusiform gyrus, respectively, in a variety of tasks [Chao et al., 2002; Devlin et al., 2005; Noppeney et al., 2006; Okada et al., 2000; Price et al., 2003; Rogers et al., 2005; Whatmough et al., 2002]. Category effects for auditory word stimuli have also been observed in ventral occipitotemporal regions bilaterally for both sighted and blind participants [Mahon et al., 2009]. Previous imaging data have suggested differences between categorical processing

of auditory objects and related these differences to the differential weighting of functional versus sensory features [Doehrmann et al., 2008; Engel et al. 2009; Kraut et al., 2006; Lewis et al., 2005; Murray et al., 2006], with man-made items increasing activation in regions involved in processing action and motor properties and living items increasing activation in visual association regions. For manmade items, it is possible that there were no significant effects within the categorical question, because the set of objects were made up of 50% tool and 50% other man-made items. Because the nontool manmade items are not strongly linked with functional features (e.g. musical instruments, which have been associated with perceptually based semantic deficits in patients; [Warrington and Shallice, 1984] this may have attenuated the weighting of functional properties and hence explain why we did not detect differences between these conceptual categories. However, this does not explain why there was no increased activation for attribute verification of living things, blocks of which consisted of only animals. We therefore presume that the verification of visual object attributes from environmental sound input does not tap into the same underlying function reported in previous studies of category effects and might thus be both a task—(verification) and modality—(nonverbal sound) specific effect.

We expected verification of encyclopedic attribute knowledge to engage regions involved in retrieval of associative, modality-independent knowledge of objects. Activation was left lateralized and reflected a dorsal posterior-anterior pattern, as opposed to the ventral posterior activation seen for visual perceptual questions. The three clusters of activation in angular gyrus, precuneus, and medial superior frontal gyrus have been previously implicated in episodic, semantic, and autobiographical memory retrieval tasks [Burianova and Grady, 2007; Gardini et al., 2006; Nyberg, 2002]. We speculate that accessing the more general or broad nature of encyclopedic object knowledge, which does not specifically relate to perceptual or functional properties, results in this wider network of regions common to multiple tasks. These regions would thus appear to reflect a network involved in visualizing and retrieving stored encyclopedic information about the heard-object concept.

The precuneus is part of the dorsal visual pathway; thus, its involvement in this network suggests a visual component to retrieving encyclopedic knowledge from environmental sounds. Bringing to mind a visualization (i.e., a mental image) of the object creating the heard sound may, as for the visual perceptual condition, be an implicit or explicit process. Indeed, the precuneus has been implicated in the explicit generation of visual imagery from concrete nouns [Gardini et al., 2006]. The response pattern in this region (see plot, Fig. 2) was weighted toward the question involving living things, although this difference did not reach a corrected level of significance. One might not expect information about the

encyclopedic properties of these objects to be subsumed by visual processing regions, but rather by amodal semantic regions (such as the angular gyrus). However, this region could be responding to visual-imagery processes, driven either by top-down semantic (e.g., in angular gyrus) to sensory association regions as an implicit by-product of conceptual processing or bottom-up mediating between environmental sound input before (or in parallel with) access to stored semantics. The differential weighting toward the question involving living things may be due to the increased demands previously associated with processing the more visually similar living items, compared to nonliving things, which are weighted more to functional or action features.

The angular gyrus has been consistently associated with retrieval of semantic information, from both visual and auditory words and pictures (for a review, see Vigneau et al. [2006]). Indeed, priming from incongruent visual inputs has been found to increase activation in the angular gyrus for environmental sounds but not the corresponding auditory words. This effect was associated with increased demands on conceptual (semantic) processing [Noppeney et al., 2008], and contrasted with a crossmodal (incongruent) priming effect in the medial superior frontal gyrus for both sounds and spoken words. We would concur with a conceptual role for the angular gyrus as part of the network involved in accessing and retrieving encyclopedic knowledge about objects. However, Noppeney et al. [2008] associated the effect in the medial superior frontal activation, which they observed for incongruent priming of both sounds and words, with higher conceptual decision processes and cognitive control mechanisms [cf. Botvinick et al., 2001; Brown and Braver, 2005; Duncan and Owen, 2000; Kerns et al., 2004; Paus, 2001]. The similarity in reaction times between the encyclopedic and visual perceptual verification conditions reported here suggest that medial frontal activation was specific to the task of retrieving encyclopedic information from environmental sounds that was no more demanding in cognitive terms (as measured by response latency) than retrieving visual perceptual information.

We did not observe differential activation for the categorical verification condition relative to the other question types. It could be argued that categorization (living/non-living) can be achieved either without recourse to the exact source of the sound or may require access to only the single concept without the additional associative semantic information required to access more complex conceptual attributes such as encyclopedic- or perceptually related features of that concept. Alternatively, information about category may have activated a subset of neuronal regions also engaged by the more difficult (as suggested by response times) encyclopedic and visual perceptual questions, which would be undetectable in this design.

Neuropsychological data have suggested that auditory nonverbal agnosia is the result of damage to the right hemisphere [Fujii et al., 1990; Spreen et al., 1965].

Although all tasks activated a bilateral network of regions, we did not observe activation in the right cerebral cortex for any of the three verification tasks that required access to meaning relative to the auditory perceptual baseline. It is important to note that we controlled for the low-level acoustic and perceptual properties of stimuli by using the same sounds across all conditions as well as utilizing an auditory perceptual baseline task to remove activation related to perceptual discrimination of the nonverbal sounds. Consistent with work by Schnider et al. [1994] and Vignolo [1982], these data therefore suggest that the right hemisphere is selectively involved in the acoustic and perceptual discrimination of nonverbal environmental sounds. When this region is damaged, the failure to discriminate sounds results in the inability to carry out high-level semantic analysis of those sounds, a function carried out by left hemisphere semantic regions. Consistent with this account, EEG evidence has shown that, although environmental sound processing involves a bilateral network of regions, differential processing of categories of sounds occurs in the right posterior superior and middle temporal cortices early in the time-course of environmental sound discrimination (70-ms poststimulus onset: [Murray et al., 2006].

As mentioned in the Introduction section, a previous functional imaging study [Thierry et al., 2003] has demonstrated a double dissociation between processing environmental sounds in the right hemisphere and spoken words in the left hemisphere, in the context of controlling for low-level perceptual properties. At first glance, our results appear to contradict these findings; however, there are two important points to note. First, the baseline task used by Thierry et al. [2003] used scrambled sounds with any reference to meaning removed. Here, we retained meaning in the sounds used for the baseline task against which other conditions were contrasted. We were thus able to control for both early acoustic as well as perceptual properties of meaningful sounds before high-level semantic analysis. Second, their results were not modulated by task difficulty, as demonstrated by the absence of an interaction between task (logical sequence and categorization) and stimulus type. They therefore concluded that the level at which this dissociation occurred was at an intermediary stage in the processing of meaning that interfaced between low-level acoustic properties and high-level semantic analysis. Here, we did not observe any significant activation for the categorical verification task when a high-level control for perceptual properties of meaningful sounds was used in a task very similar to their categorization task (which their results suggest would activate the same right hemisphere regions as sequence interpretation). We therefore conclude that the categorization verification task used here only requires access to intermediary stages in processing, which is not distinct from any intermediate semantic processing that may implicitly occur in response to a high-level baseline.

In conclusion, we demonstrate a clear and highly significant dissociation in the processing of different semantic

object attributes from the same environmental sounds, consistent with the models of object processing proposing an underlying organization by salient object property. Knowledge about visual perceptual attributes engaged ventral temporal brain regions previously associated with visual (and amodal) object processing, whereas verifying encyclopedic attributes from environmental sounds engaged a dorsal temporal-frontal network. This dissociation was observed in the absence of differences between verification responses times, using the identical conceptual sound stimuli and the same attribute verification task. Future studies will need to contrast this verification task using different sensory modalities of conceptual input to determine whether these regions are selective for nonverbal sounds (i.e., unimodal) or if they are shared across modality (crossmodal or amodal) as well as investigating how connectivity between right hemisphere auditory association cortices and left hemisphere semantic regions modulates nonverbal auditory object processing.

ACKNOWLEDGMENTS

We thank Don Maillet, Alan Pringle, and Peter Hobden at the Centre for Magnetic Resonance for their assistance in setting up the auditory delivery system and all the people that participated in this study.

REFERENCES

- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005): Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559–571.
- Ashburner J, Friston KJ (2005): Unified segmentation. *NeuroImage* 26:839–851.
- Auerbach SH, Allard T, Naeser M, Alexander MP, Albert ML (1982): Pure word deafness. Analysis of a case with bilateral lesions and a defect at the prephonemic level. *Brain* 105:271–300.
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001): Conflict monitoring and cognitive control. *Psychol Rev* 108: 624–652.
- Brown JW, Braver TS (2005): Learned predictions of error likelihood in the anterior cingulate cortex. *Science* 307:1118–1121.
- Buchel C, Price CJ, Frackowiack RS, Friston K (1998): Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain* 121:409–419.
- Burianova H, Grady CL (2007): Common and unique neural activations in autobiographical, episodic, and semantic retrieval. *J Cogn Neurosci* 19:1520–1534.
- Chao LL, Weisberg J, Martin A (2002): Experience-dependent modulation of category-related cortical activity. *Cereb Cortex* 12:545–551.
- Coslett HB, Brashear HR, Heilman KM (1984): Pure word deafness after bilateral primary auditory cortex infarcts. *Neurology* 34:347–352.
- Devlin JT, Rushworth MF, Matthews PM (2005): Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia* 43:69–74.

- Devlin JT, Russell RP, 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.
- Di Giovanni M, D'Alessandro G, Baldini S, Cantalupi D, Bottacchi E (1992): Clinical and neuroradiological findings in a case of pure word deafness. *Ital J Neurol Sci* 13:507–510.
- Doehrmann O, Naumer MJ, Volz S, Kaiser J, Altmann CF (2008): Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia* 46:2776–2786.
- Duncan J, Owen AM (2000): Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23:475–483.
- Engel LR, Frum C, Puce A, Walker NA, Lewis JW (2009): Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage* 47:1778–1791.
- Engelien A, Tuscher O, Hermans W, Isenberg N, Eidelberg D, Frith C, Stern E, Silbersweig D (2006): Functional neuroanatomy of non-verbal semantic sound processing in humans. *J Neural Transm* 113:599–608.
- Freire L, Roche A, Mangin JF (2002): What is the best similarity measure for motion correction in fMRI time series? *IEEE Trans Med Imaging* 21:470–484.
- Fujii T, Fukatsu R, Watabe S, Ohnuma A, Teramura K, Kimura I, Saso S, Kogure K (1990): Auditory sound agnosia without aphasia following a right temporal lobe lesion. *Cortex* 26:263–268.
- Gardini S, Cornoldi C, De Beni R, Venneri A (2006): Left medio-temporal structures mediate the retrieval of episodic autobiographical mental images. *Neuroimage* 30:645–655.
- Gauthier II (2000): What constrains the organization of the ventral temporal cortex? *Trends Cogn Sci* 4:1–2.
- Gorno-Tempini ML, Ciolotti L, Price CJ (2000): Category differences in brain activation studies: Where do they come from? *Proc Biol Sci* 267:1253–1258.
- Habib M, Daquin G, Milandre L, Royere ML, Rey M, Lanteri A, Salamon G, Khalil R (1995): Mutism and auditory agnosia due to bilateral insular damage—Role of the insula in human communication. *Neuropsychologia* 33:327–339.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Kerns JG, Cohen JD, MacDonald AW III, Cho RY, Stenger VA, Carter CS (2004): Anterior cingulate conflict monitoring and adjustments in control. *Science* 303:1023–1026.
- Kraut MA, Pitcock JA, Calhoun V, Li J, Freeman T, Hart J Jr (2006): Neuroanatomic organization of sound memory in humans. *J Cogn Neurosci* 18:1877–1888.
- Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA (2004): Human brain regions involved in recognizing environmental sounds. *Cereb Cortex* 14:1008–1021.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005): Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci* 25:5148–5158.
- Lund K, Burgess C (1996): Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav Res Methods* 28:203–208.
- Mahon BZ, Anzellotti S, Schwarzbach J, Zampini M, Caramazza A (2009): Category-specific organization in the human brain does not require visual experience. *Neuron* 63:397–405.
- Martin A (2007): The representation of object concepts in the brain. *Annu Rev Psychol* 58:25–45.
- Mechelli A, Henson RN, Price CJ, Friston KJ (2003): Comparing event-related and epoch analysis in blocked design fMRI. *Neuroimage* 18:806–810.
- Metz-Lutz MN, Dahl E (1984): Analysis of word comprehension in a case of pure word deafness. *Brain Lang* 23:13–25.
- Moore CJ, Price CJ (1999): A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain* 122 (Pt 5):943–962.
- Moss HE, Abdallah S, Fletcher P, Bright P, Pilgrim L, Acres K, Tyler LK (2005): Selecting among competing alternatives: Selection and retrieval in the left inferior frontal gyrus. *Cereb Cortex* 15:1723–1735.
- Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S (2006): Rapid brain discrimination of sounds of objects. *J Neurosci* 26:1293–1302.
- Noppeney U, Price CJ (2002): A PET study of stimulus- and task-induced semantic processing. *Neuroimage* 15:927–935.
- Noppeney U, Price CJ, Penny WD, Friston KJ (2006): Two distinct neural mechanisms for category-selective responses. *Cereb Cortex* 16:437–445.
- Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ (2008): The effect of prior visual information on recognition of speech and sounds. *Cereb Cortex* 18:598–609.
- Nyberg L (2002): Levels of processing: A view from functional brain imaging. *Memory* 10:345–348.
- Okada T, Tanaka S, Nakai T, Nishizawa S, Inui T, Sadato N, Yonekura Y, Konishi J (2000): Naming of animals and tools: A functional magnetic resonance imaging study of categorical differences in the human brain areas commonly used for naming visually presented objects. *Neurosci Lett* 296:33–36.
- Paus T (2001): Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nat Rev Neurosci* 2:417–424.
- Pietrini P, Furey ML, Ricciardi E, Gobbi ML, Wu WH, Cohen L, Fuazzelli M, Haxby JV (2004): Beyond sensory images: Object-based representation in the human ventral pathway. *Proc Natl Acad Sci USA* 101:5658–5663.
- Price CJ, Noppeney U, Phillips J, Devlin JT (2003): How is the fusiform gyrus related to category-specificity? *Cogn Neuropsychol* 20:561–574.
- Rogers TT, Hocking J, Mechelli A, Patterson K, Price C (2005): Fusiform activation to animals is driven by the process, not the stimulus. *J Cogn Neurosci* 17:434–445.
- Schneider TR, Engel AK, Debener S (2008): Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Exp Psychol* 55:121–132.
- Schnider A, Benson DF, Alexander DN, Schnider-Klaus A (1994): Non-verbal environmental sound recognition after unilateral hemispheric stroke. *Brain* 117 (Pt 2):281–287.
- Spree O, Benton AL, Fincham RW (1965): Auditory agnosia without aphasia. *Arch Neurol* 13:84–92.
- Takahashi N, Kawamura M, Shinotou H, Hirayama K, Kaga K, Shindo M (1992): Pure word deafness due to left hemisphere damage. *Cortex* 28:295–303.
- Tanaka Y, Yamadori A, Mori E (1987): Pure word deafness following bilateral lesions. A psychophysical analysis. *Brain* 110 (Pt 2):381–403.
- Taniwaki T, Tagawa K, Sato F, Iino K (2000): Auditory agnosia restricted to environmental sounds following cortical deafness and generalized auditory agnosia. *Clin Neurol Neurosurg* 102: 156–162.
- Thierry G, Giraud AL, Price C (2003): Hemispheric dissociation in access to the human semantic system. *Neuron* 38:499–506.

- Tyler LK, Stamatakis EA, Bright P, Acres K, Abdallah S, Rodd JM, Moss HE (2004): Processing objects at different levels of specificity. *J Cogn Neurosci* 16:351–362.
- Vaughan JT, Adriany G, Garwood M, Yacoub E, Duong T, Dela-Barre L, Andersen P, Ugurbil K (2002): Detunable transverse electromagnetic (TEM) volume coil for high-field NMR. *Magn Reson Med* 47:990–1000.
- Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N (2006): Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage* 30:1414–1432.
- Vignolo LA (1982): Auditory agnosia. *Philos Trans R Soc Lond B Biol Sci* 298:49–57.
- Warrington EK, Shallice T (1984): Category specific semantic impairments. *Brain* 107 (Pt 3):829–854.
- Whatmough C, Chertkow H, Murtha S, Hanratty K (2002): Dissociable brain regions process object meaning and object structure during picture naming. *Neuropsychologia* 40:174–186.
- Yaqub BA, Gascon GG, Al-Nosha M, Whitaker H (1988): Pure word deafness (acquired verbal auditory agnosia) in an Arabic speaking patient. *Brain* 111 (Pt 2):457–466.
- Zeng H, Constable RT (2002): Image distortion correction in EPI: Comparison of field mapping with point spread function mapping. *Magn Reson Med* 48:137–146.