

# Cortical Networks Representing Object Categories and High-level Attributes of Familiar Real-world Action Sounds

James W. Lewis<sup>1</sup>, William J. Talkington<sup>1</sup>, Aina Puce<sup>1,2</sup>,  
Lauren R. Engel<sup>1</sup>, and Chris Frum<sup>1</sup>

## Abstract

■ In contrast to visual object processing, relatively little is known about how the human brain processes everyday real-world sounds, transforming highly complex acoustic signals into representations of meaningful events or auditory objects. We recently reported a fourfold cortical dissociation for representing action (nonvocalization) sounds correctly categorized as having been produced by human, animal, mechanical, or environmental sources. However, it was unclear how consistent those network representations were across individuals, given potential differences between each participant's degree of familiarity with the studied sounds. Moreover, it was unclear what, if any, auditory *perceptual* attributes might further distinguish the four *conceptual* sound-source categories, potentially revealing what might drive the cortical network organization for representing acoustic knowledge. Here, we used functional

magnetic resonance imaging to test participants before and after extensive listening experience with action sounds, and tested for cortices that might be sensitive to each of three different high-level perceptual attributes relating to how a listener associates or interacts with the sound source. These included the sound's perceived concreteness, effectuality (ability to be affected by the listener), and spatial scale. Despite some variation of networks for environmental sounds, our results verified the stability of a fourfold dissociation of category-specific networks for real-world action sounds both before and after familiarity training. Additionally, we identified cortical regions parametrically modulated by each of the three high-level perceptual sound attributes. We propose that these attributes contribute to the network-level encoding of category-specific acoustic knowledge representations. ■

## INTRODUCTION

How is our knowledge of different types of everyday sounds and auditory objects represented in the brain? Through life experience, we learn about which sensory features or object attributes are most behaviorally relevant for successful object categorization and recognition (McClelland & Rogers, 2003; Miller, Nieder, Freedman, & Wallis, 2003; Rosch, 1973). This includes knowledge representations for the broad categories of “living” versus “nonliving” things (Silveri et al., 1997; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996; De Renzi & Lucchelli, 1994; Warrington & Shallice, 1984). Additionally, distinct processing pathways have been identified or proposed for various *visually defined* object categories, including faces, body parts, animals, tools, scenes and places, buildings, letters, and visual word forms (Farah, 2004; Hasson, Harel, Levy, & Malach, 2003; McCandliss, Cohen, & Dehaene, 2003; Polk et al., 2002; Downing, Jiang, Shuman, & Kanwisher, 2001; Haxby et al., 2001; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997; Perani et al., 1995; Allison, McCarthy, Nobre,

Puce, & Belger, 1994). The underlying mechanisms that establish semantic-level category-specificity remains an area of active research (Patterson, Nestor, & Rogers, 2007; Barsalou, Kyle Simmons, Barbey, & Wilson, 2003). At one extreme, theoretical models suggest that domain-specific networks may be innately predisposed to perform specific types of operations relatively independent of sensory experience, and consequently, develop to process certain types of knowledge representations (Burton, Snyder, & Raichle, 2004; Caramazza & Mahon, 2003; Pascual-Leone & Hamilton, 2001; Caramazza & Shelton, 1998). Conversely, sensory-motor property-based models posit that the brain largely self organizes based on an individual's sensory input experiences (Martin, 2007; Lissauer, 1890/1988). Regardless of mechanism, or combination therein, a multitude of visual studies indicate that various visual attributes, such as form, color, and biological motion, influence how certain conceptual categories of objects may become differentially represented in the brain (Roether, Omlor, Christensen, & Giese, 2009; Nielsen, Logothetis, & Rainer, 2008; New, Cosmides, & Tooby, 2007; Van Essen & Gallant, 1994). However, in hearing research, considerably less is known about what attributes might similarly influence how auditory “objects” or action events may eventually become represented

<sup>1</sup>West Virginia University, Morgantown, <sup>2</sup>Indiana University, Bloomington

as acoustic knowledge (Scott, 2005; Griffiths & Warren, 2004; Husain, Tagamets, Fromm, Braun, & Horwitz, 2004).

Much of hearing research has emphasized how acoustic signals from a single sound source, regardless of category, may be segregated and processed as a distinct event or chain of events, generally referred to as “auditory streaming” (Husain, Lozito, Ulloa, & Horwitz, 2005; Carlyon, 2004; Tougas & Bregman, 1990). For most species, this streaming of acoustic information must ultimately be transformed into neuronal representations that evaluate the behavioral relevance of the sound to the listener. Repeated sound exposures during development conditions a listener to generalize features characteristic of particular sound sources, such that subsequent exposure to similar acoustic events can lead to probabilistic representations of the identity, or at least category, of the sound source (Kumar, Stephan, Warren, Friston, & Griffiths, 2007; Körding & Wolpert, 2004). Akin to visual object and visual motion processing mechanisms in cortex, similar hierarchical processing mechanisms could exist for extracting sensory feature information from the sound waves exciting the auditory system, ultimately allowing for the rapid processing and recognition of everyday sounds. Thus, research in category-specific hearing perception may serve to uniquely inform cognitive models of knowledge representation, which have predominantly been based on visual and linguistic studies.

The strongest evidence to date for defining a distinct category of natural sound is for vocalizations. Pathways preferential or selective for processing vocalizations (including speech and other harmonically structured sounds) in humans implicate various intermediate processing stages beyond tonotopically organized primary auditory cortices (Leaver & Rauschecker, 2010; Lewis et al., 2009; Rauschecker & Scott, 2009; Staeren, Renvall, De Martino, Goebel, & Formisano, 2009; Fecteau, Armony, Joanette, & Belin, 2004; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). This is based on both top-down influences for extracting potential linguistic, communicative, or emotional content (Uppenkamp, Johnsrude, Norris, Marslen-Wilson, & Patterson, 2006; Thierry, Giraud, & Price, 2003; Belin, Zatorre, & Ahad, 2002; Binder et al., 2000), and on quantifiable bottom-up attributes including particular ranges of temporal cues and cadences (Wilden, Herzel, Peters, & Tembrock, 1998; Owren, Seyfarth, & Cheney, 1997; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) and harmonic content (Lewis et al., 2009; Riede & Zuberbühler, 2003; Riede, Herzel, Hammerschmidt, Brunnberg, & Tembrock, 2001). Neuropsychological lesion studies have documented selective deficits for some categories of nonverbal auditory material, including environmental sounds and animal vocalizations (Saygin, Leech, & Dick, 2010), wherein animal vocalizations could be regarded as a nonlinguistic subcategory of vocalizations. Animal vocalizations, relative to sounds produced by humans using hand tools, preferentially activate distinct cortical pathways (Altmann, Doehrmann, & Kaiser, 2007; Lewis, Phinney, Brefczynski-Lewis, & DeYoe, 2006; Lewis, Brefczynski, Phinney, Janik, & DeYoe, 2005), and this differential processing can

occur in the brain as early as 70 msec after sound onset (Murray, Camen, Gonzalez Andino, Bovet, & Clarke, 2006). Because vocalizations appear to represent a distinct category of sound, we retained action-sound stimuli explicitly devoid of any vocalization content for the present paradigm.

Human-produced (conspecific) action sounds are likely to represent another distinct category of real-world sound (De Lucia, Camen, Clarke, & Murray, 2009; Engel, Frum, Puce, Walker, & Lewis, 2009; Doehrmann, Naumer, Volz, Kaiser, & Altmann, 2008; Galati et al., 2008; Lahav, Saltzman, & Schlaug, 2007; Gazzola, Aziz-Zadeh, & Keysers, 2006; Lewis et al., 2005, 2006; Bidet-Caulet, Voisin, Bertrand, & Fonlupt, 2005; Pizzamiglio et al., 2005). In particular, the above studies indicate that human action sounds preferentially or selectively activate fronto-parietal networks, including presumed mirror-neuron systems, and bilateral posterior temporal regions. Processing in these regions is thought to reflect experience-based probabilistic matching to one’s own motor repertoire of sound-producing actions, such that a listener may effectively “embody” conspecific action sounds (or visual actions) as one means for generating a sense of meaning and behavioral relevance behind the sound, consistent with embodied or grounded cognition models (Barsalou, 2008).

With regard to the semantic categories of “living or animate things,” as identified in neuropsychological literature (e.g., Saygin et al., 2010), knowledge of *human* action sounds presumably represents just one acoustic subcategory. Thus, as a second category of action sound, we recently tested for and reported a dissociation of cortical networks for processing action sounds produced by non-human animals, thereby demonstrating distinct cortical representations for at least two subcategories of living action-sound sources (Engel et al., 2009). We also reported a dissociation of cortical networks for processing two acoustic subcategories of “nonliving things,” including mechanical and environmental sound sources (*ibid*). Although environmental sounds cannot be readily embodied by a listener, there is an ecological rationale for suspecting that these stimuli could be processed as a potentially distinct category of sound. However, automated machinery and their corresponding action sounds, represents a subcategory of nonliving things that have been around for less than 200 or so years. Thus, specific cortical networks for encoding knowledge representations for these two distinct categories of nonliving action sounds may, in part, be governed by different rules. Although the above-mentioned fourfold dissociation of cortical networks for different categories of sound sources was informative for models of knowledge representations, those results relied on measuring brain responses to novel sound samples presented out of context, and thus, one or some sound categories might have simply been more familiar to participants, thereby biasing the results. Consequently, it was unclear how “stable” this fourfold dissociation of network representations may have been, especially for the nonhuman action-sound categories.

Increasing one's familiarity with particular objects, through repeated exposure or training (perceptual learning), typically results in modulations in brain activation responses, which may result from stabilized and reweighted network representations that are less subject to interference (Doshier & Lu, 2009; Erickson et al., 2007). This has been demonstrated both at a neuronal level in monkeys (Peissig, Singer, Kawasaki, & Sheinberg, 2007; Miller et al., 2003; Desimone, 1996) and at a cortical network level in humans when assessed with hemodynamic neuroimaging techniques (Henson, Shallice, & Dolan, 2000; Grill-Spector et al., 1999; Buckner et al., 1995). For instance, studies on expertise training have reported *enhanced* activations to repeated exposure or practice with discriminating visual objects, such as when viewing faces (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999), highly familiar (visuo-) motor actions (Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005), and in response to cross-sensory associative learning of artificial "audiovisual" objects (Naumer et al., 2008). In contrast, other studies have reported *reduced* activations to repeated exposures with previously unfamiliar visual scenes (Yi & Chun, 2005), unfamiliar faces (Ryan et al., 2008; Gobbini & Haxby, 2006), and common everyday visual objects such as animals and cars (Grill-Spector et al., 1999), musical instruments, and sports equipment (Reber, Gitelman, Parrish, & Mesulam, 2005). Perceptual learning theories that account for reduced activation profiles include a *sharpening* model, wherein increasing familiarity leads to sparser and more refined neural encoding, and a *facilitation* model, wherein repeated exposures leads to faster neural processing (Wiggs & Martin, 1998; Desimone, 1996).

The effects of both sharpening and facilitative plastic changes, when measured using fMRI, can manifest as decreased signal amplitudes and/or decreased expanses in activated cortex (Mukai et al., 2007; Grill-Spector, Henson, & Martin, 2006; James & Gauthier, 2006). Regardless of whether enhanced or reduced activations prevail as a result of increased familiarity with recognizing particular objects or sensory events, those networks that remain significantly responsive should reflect neuronal encoding that is more highly selective or stable for representing that particular type of object or event (Grill-Spector et al., 2006; Hopfield & Brody, 2000; Hopfield & Tank, 1985). Consequently, a primary objective of the present study was to determine how consistent the cortical network activations were for representing human, animal, mechanical, and environmental action sounds after increasing a participant's familiarity with those sounds. Our hypothesis was that increasing one's familiarity with examples of readily categorized and recognized acoustic events that belonged to a given category of sound source would lead to reduced activation network profiles, consistent with sharpening and facilitation models.

Also unclear from our earlier study was whether there were any behaviorally relevant *high-level* perceptual attributes that might be associated with the fourfold dissocia-

tion of category-specific cortical networks. In vision studies, tools typically represent one distinct object category, which is, in part, due to their motor affordance properties, such as how graspable they appear to be (Grèzes, Tucker, Armony, Ellis, & Passingham, 2003; Ellis & Tucker, 2000). Additionally, knowledge representations of humans (and animals) are distinguished as living things by the visual system, in part, due to their biological motion attributes (Pelphrey, Morris, & McCarthy, 2004; Frith & Frith, 1999; Johansson, 1973). Hence, various high-level perceptual attributes can be important in contributing to representations of distinct categories of visual objects or agents. Thus, another objective of the present study was to identify cortical regions that were sensitive to different high-level acoustic perceptual attributes or dimensions, with the expectation that how a listener interacts with a sound source will influence how the associated category-specific acoustic knowledge becomes represented in the brain.

## METHODS

### Participants

We tested 14 right-handed participants (age = 20–36 years, 8 women). All participants were native English speakers with no previous history of neurological, psychiatric disorders, or auditory impairment, and had a self-reported normal range of hearing. Informed consent was obtained for all participants following guidelines approved by the West Virginia University Institutional Review Board.

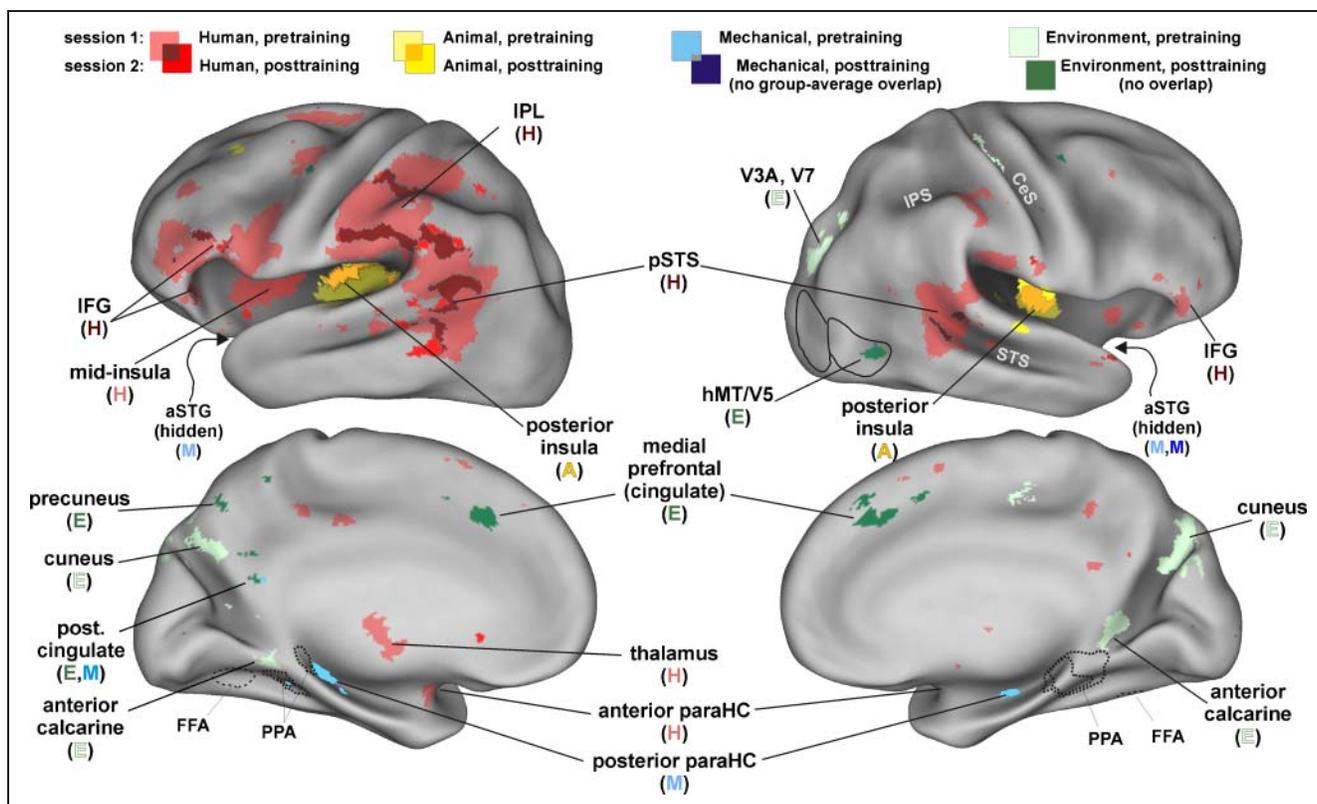
### Sound Stimulus Creation and Presentation

Sound stimuli consisted of 256 sound stimuli reported previously (Engel et al., 2009). These included professional compilations of action sounds (Sound Ideas, Richmond Hill, Ontario, Canada) from four conceptual categories: human, animal, mechanical, and environmental (see Appendix 1 for list). All sound stimuli were edited to  $3.0 \pm 0.5$  sec duration, matched for total root mean squared (RMS) power, and onset/offset ramped to 25 msec (Cool Edit Pro, Syntrillium Software, owned by Adobe). Sound stimuli were converted to one channel (mono, 44.1 kHz, 16 bits) but presented to both ears. During fMRI scanning, high-fidelity sound stimuli were delivered using a Windows PC computer, with Presentation software (version 11.1, Neurobehavioral Systems) via a sound mixer and MR-compatible electrostatic ear buds (STAX SRS-005 Earspeaker System; Stax, Gardena, CA), worn under sound-attenuating ear muffs.

### Scanning Paradigms

Each participant underwent two fMRI scans. The initial Scanning Session 1 consisted of eight separate functional imaging runs, across which the 256 sound stimuli and 64





**Figure 1.** A fourfold dissociation of preferential activation for different categories of action sounds pre- and postfamiliarity training. Group-averaged ( $n = 14$ ) data show preferential activation to the perception of sounds produced by humans (red), animals (yellow), mechanical sources (blue), and environmental sources (green), resulting from pretraining scanning (Session 1) and after familiarity training (Session 2). Data are illustrated on three-dimensional inflated renderings of the PALS cortical surface model ( $\alpha < 0.05$ , corrected; refer to Methods). CeS = central sulcus; FFA = fusiform face area; paraHC = parahippocampus; IPS = intraparietal sulcus; PPA = parahippocampal place area; aSTG = anterior superior temporal gyrus; pSTS = posterior superior temporal gyrus. Refer to text for other details.

### High-level Perceptual Attributes of the Category-specific Sounds

The 213 sound stimuli retained for analyses were presented in random order to a subset of the participants ( $n = 9$ ), who each rated the sounds using Likert scales (1 to 5), assessing three different high-level semantic attributes: concreteness, effectuation, and spatial scale. The *concreteness* dimension (CC) was defined as the degree to which the sound constituted an actual instance, event, or thing (“thingness”): being definite, not vague, or elusive (1 = definite/concrete, 5 = vague, elusive). Our concreteness characterization was similar to that described in previous work regarding word recognition (Fliessbach, Weis, Klaver, Elger, & Weber, 2006). However, concrete words, which refer to things that can be sensually experienced, are usually contrasted with words that depict abstract concepts, which did not apply well to the action sounds. In the present study, the sound of rain, for example, was typically rated as being less concrete of a sound source than that of a person producing sound by typing on a keyboard. The *effectuality* dimension (EF) was defined by the sense that the listener

could conceivably have caused, affected, or influenced the sound production in some way: This included assessing the degree to which the heard sound (as opposed to the written description of the sound) provided a sense that the source was manipulable, palpable, graspable, or tangible (1 = more effectual, 5 = not tangible/effectual). This scale was designed to have homology to motor affordance properties associated with visual objects, ostensibly including graspable objects such as tools (Goldenberg, Hentze, & Hermsdorfer, 2004; Grèzes, Tucker, et al., 2003; Grèzes & Decety, 2002). The *spatial scale* dimension (SS) referred to the extent to which the heard sound provided a sense of depicting a small localized, discrete event, such as a stopwatch ticking, or if it manifested on a scale that was large relative to the listeners body size, such as a strong wind or a freight train (1 = small, 5 = grand). For each perceptual dimension, the ratings for each sound were averaged across the group. Group-averaged ratings were then assessed by a *K*-means analysis to identify clusters of rated sounds that corresponded to any of the four conceptual categories (human, animal, mechanical, or environmental), and separately assessed for clustering significance using a MANOVA analysis.

## Correlations between Semantic Attributes and Brain Activation

Using the above described Likert-scale ratings, we also conducted a second multiple linear regression analysis of the fMRI data. For each individual ( $n = 9$  of 14), his or her Likert-scale responses (1 to 5) were used as multipliers in each of three regression terms. The data analyzed were further restricted to only those sounds correctly categorized after Session 1 and those sounds for which all three ratings were collected (some participants did not rate all sounds within the allotted time). Using the same fMRI analyses described earlier, we performed a conjunction analysis to identify brain regions sensitive to one or another semantic attribute [e.g., (EF > CC)  $\cap$  (EF > SS)].

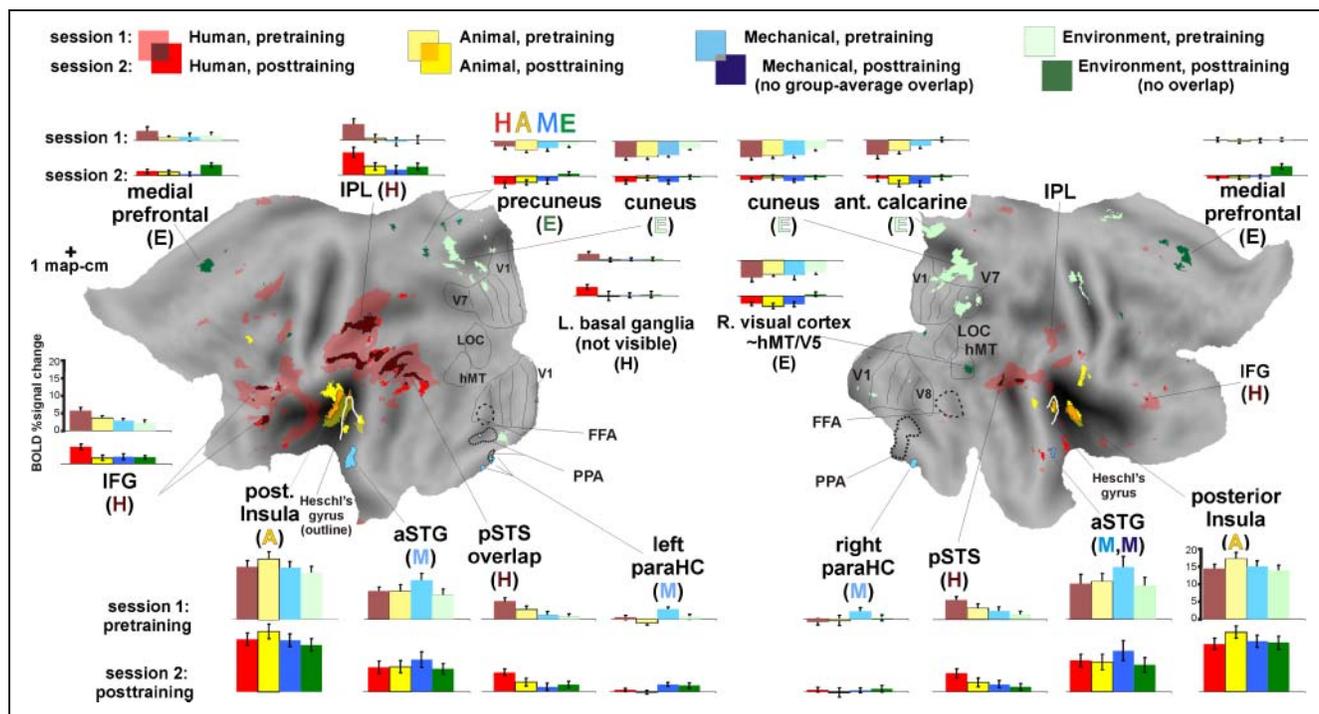
## RESULTS

This study was comprised of two fMRI scanning sessions. In Session 1, which was reported in our earlier study (Engel et al., 2009), participants listened to 256 examples of common real-world sounds for the first time, and silently determined if each sound was produced by a human or not. Participants were instructed to not provide any overt motor responses during scanning in order to avoid potentially confounding activation in motor-related cortices. However, immediately after scanning, participants heard all of the sounds again and provided overt keyboard responses to indicate which category he or she thought each

sound belonged to: human, animal, mechanical, or environmental. For the present study, a subset of the participants ( $n = 14$ ) listened to all the sounds four to five additional times over the course of several days or weeks, practicing to the point where they could readily recognize and verbally identify each sound (refer to Methods). A second fMRI scan (Session 2) ensued using the same task and identical scanning parameters to Session 1. To maintain a precisely balanced statistical comparison between the pre- and post-training conditions, only those sounds that were correctly categorized after Session 1 were retained for analyses in both Sessions 1 and 2. There were two main findings from this study: one related to the persistence of a fourfold dissociation of networks for representing different categories of real-world sound sources, and the other related to the identification of high-level perceptual attributes that may influence or determine how acoustic knowledge representations and category-specificity may become organized in the human brain.

## Effects of Familiarity Training on Cortical Networks Representing Real-world Action Sounds

After familiarity training, there still existed a fourfold dissociation of brain networks preferentially activated by each sound category—human, animal, mechanical, and environmental—although with a few significant network activation differences (Figures 1 and 2; Table 1). These results are addressed below by sound category.



**Figure 2.** Flat map cortical surface renderings and BOLD percent signal change histograms from the data shown in Figure 1. BOLD percent signal changes (mean  $\pm$  SE relative to silent events) are shown for selected regions of interest, including those defined by complete overlap between the pre- and posttraining conditions (refer to Figure 1 for other details). V1–V8 = visual areas from PALS database; LOC = lateral occipital cortex.

**Table 1.** Talairach Coordinates of Activation Foci Preferential for One Category of Sound Postfamiliarity Training or Showing Overlap across the Two Scanning Sessions

	Condition	Anatomical Location	Talairach Coordinates			Volume (mm <sup>3</sup> )
			x	y	z	
<i>Right Hemisphere</i>						
H > AME	overlap	IFG	47	32	3	58
H > AME	overlap	pSTS/pMTG	52	-43	12	8147
A > HME	overlap	posterior insula	38	-19	11	728
M > HAE	pretraining	aSTG	46	4	-4	206
M > HAE	pretraining	parahippocampus	27	-19	-14	84
E > HAM	pretraining	anterior calcarine	13	-52	4	796
E > HAM	pretraining	cuneus	18	-74	28	3972
E > HAM	posttraining	hMT/V5	41	-67	0	237
E > HAM	posttraining	medial prefrontal	5	23	39	535
<i>Left Hemisphere</i>						
H > AME	overlap	IFG	-39	34	27	569
H > AME	overlap	IPL	-49	-36	40	19,386
H > AME	overlap	mid-insula	-41	1.8	9.8	10,896
H > AME	overlap	pSTS/pMTG	-47	26	4.5	9063
A > HME	overlap	posterior insula	-35	-26	16	948
M > HAE	pretraining	aSTG	-51	4	-1	536
M > HAE	pretraining	parahippocampus	-28	-36	-7	649
E > HAM	pretraining	cuneus	-15	-65	11	120
E > HAM	posttraining	precuneus	-11	-76	28	1456
E > HAM	posttraining	medial prefrontal	-6	17	39	320

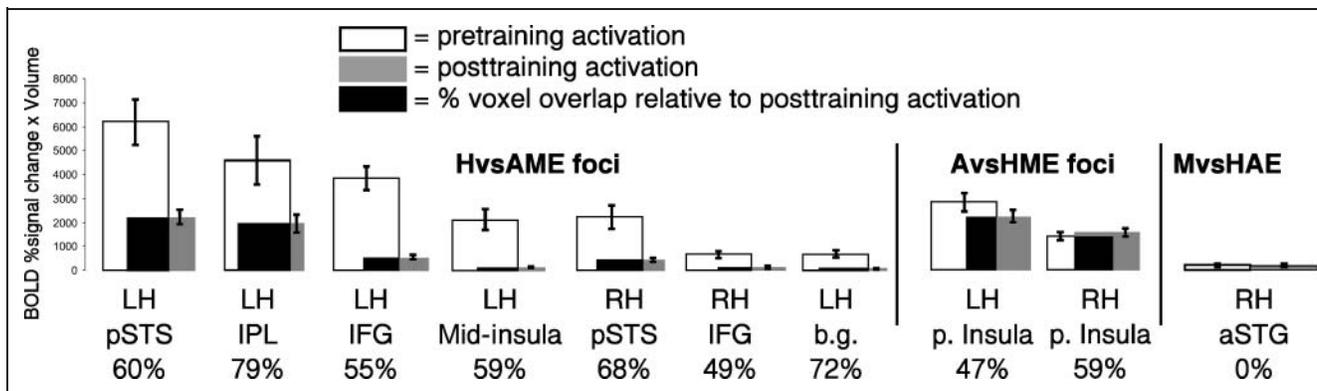
Refer to text for abbreviations.

Increased familiarity with the human-produced action sounds (e.g., walking in high-heeled shoes, typing on a keyboard, biting and chewing; all explicitly devoid of vocalizations) generally demonstrated a reduced extent of activation (Figure 1, Session 2, saturated red), confined to the same general regions showing preferential activation prior to training (Session 1; transparent red and overlap denoted by dark red). The global changes in activation extent were more evident in the corresponding cortical flat map renderings (Figure 2, red hues). However, within the regions of pre- and posttraining overlap (dark red), the relative activation *strength* (% BOLD signal changes relative to silent events) did not significantly differ [Figure 2, red histograms in Session 1 vs. 2;  $t(13)$ ,  $p > .1$  for the red regions illustrated].

To quantitatively assess the differences in the degree of volumetric activation in the pre- versus posttraining conditions, we performed a regional “activation-expansion” analysis (Figure 3). For this analysis, anatomical regions were blocked

in Talairach coordinate space (e.g., lateral temporal cortex region), and voxels within a blocked region that showed significant preferential activation to one category of sound either before *or* after training (combined regions of interest) were retained to generate a measure of activation expansion within that more inclusive functionally defined region (e.g., Figures 1 and 2; right pSTS region, all red hues). Different combined regions of interest (ROIs) showed varying degrees of overlap, resulting from the pre- versus posttraining scanning sessions. Some ROIs showed no significant activation after familiarity training, hence, no overlap is depicted (e.g., right hemisphere IPL for human actions, and foci preferential for environmental sounds). Overall, this analysis quantified the reductions in activation extent observed after familiarization training for several regions preferentially activated by different categories of action sounds (Figures 1 and 2), especially for human-produced action sounds.

Correctly categorized animal action sounds (e.g., galloping, flight sounds, a dog panting; all devoid of vocalization



**Figure 3.** Quantification of change in activation expande as a function of pre- versus postfamiliarity training. Histograms depict the group-averaged volumetric BOLD responses of several combined ROIs from Figures 1 and 2 from the pretraining scan (white; mean  $\pm$  SE relative to responses to silent events) relative to the partially overlapping activation patterns resulting from the posttraining scan (gray; mean  $\pm$  SE). The degree of direct volumetric overlap (depicted in Figures 1 and 2 as overlap regions) is indicated on the histograms (black) and in the figure text as percentages. HvsAME = foci from Figure 1 preferential for human-produced action sounds; AvsHME = preferential for animal action sounds; MvsHAE = preferential for mechanical action sounds; RH = right hemisphere; b.g. = basal ganglia.

content) led to significant preferential activation largely restricted to the bilateral posterior insula in both Session 1 (Figures 1 and 2, transparent yellow) and Session 2 (saturated yellow and overlap denoted by yellow-orange). The ROI analyses revealed similar activation strengths before and after training (Figure 2, yellow hues), with 47% volumetric overlap in activation expanses in the left hemisphere and 59% in the right (Figure 3). Unlike the foci preferential for human action sounds, foci preferential for animal action sounds—the bilateral posterior insulae—did not show a significant change in overall activation expande, but rather appeared to shift in location. The anatomical alignment of a participant’s brain volume between Sessions 1 and 2 was on the order of 1–2 mm<sup>3</sup>, indicating that these apparent ROI displacements were not simply due to alignment or registration errors per se. However, as addressed further in the following section, a voxelwise ANOVA (Figure 4, yellow hues) indicated that there were some significant effects of within-subject variability in the posterior insula relative to training effects for this category of sound. Regardless of the cause of the apparent displacement in foci, the bilateral posterior insulae, in general, showed significant preferential activation to animal action sounds relative to all three of the other categories of sound both before and after familiarity training.

Correctly categorized mechanical action sounds (e.g., a ceiling fan, a printer—all judged to be independent of an animate agent directly instigating the action) in Session 1 yielded preferential activation along bilateral parahippocampal cortex and anterior STG (aSTG) (Figures 1 and 2, light blue). The locations of the pretraining parahippocampal foci overlapped with, or were juxtaposed to, the PPA (dotted outlines). After training, the left parahippocampal activation remained comparable in strength of activation, but the relative increase in activation by the familiarity training with environmental sounds precluded a “preferential” status (Figure 2, Session 1 vs. Session 2, blue and green

histograms). Rather, this portion of the left parahippocampus was preferential for representing nonliving sound sources in general [mechanical and environmental vs. human and animal;  $t(27) = 4.62, p < .00001$ ].

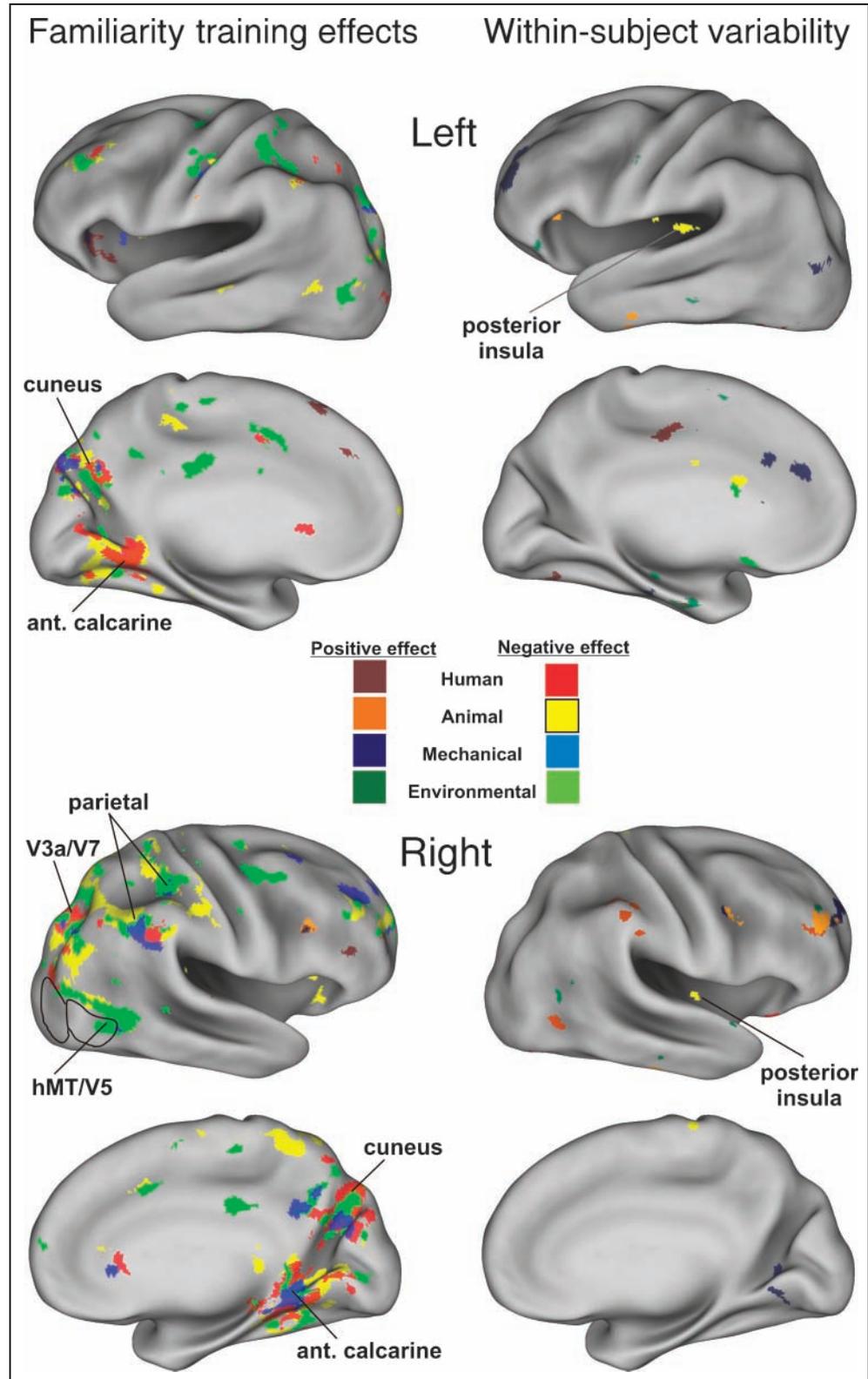
After familiarity training, significant preferential group-averaged activation by the mechanical action sounds was restricted to a small portion of the right aSTG (Figure 2, dark blue; also see Figure 5B). Most participants showed preferential bilateral aSTG activation foci for mechanical sounds after training (not illustrated). Thus, individual variability in the location of these foci, together with limitations of volumetric averaging across subjects, appeared to account for the relatively small extent of the right aSTG focus and only a trend toward significant preferential activation of the left aSTG focus after training. Nonetheless, as with the networks representing human action-sound sources, networks representing mechanical action sounds similarly showed reduced activation expanses after familiarity training.

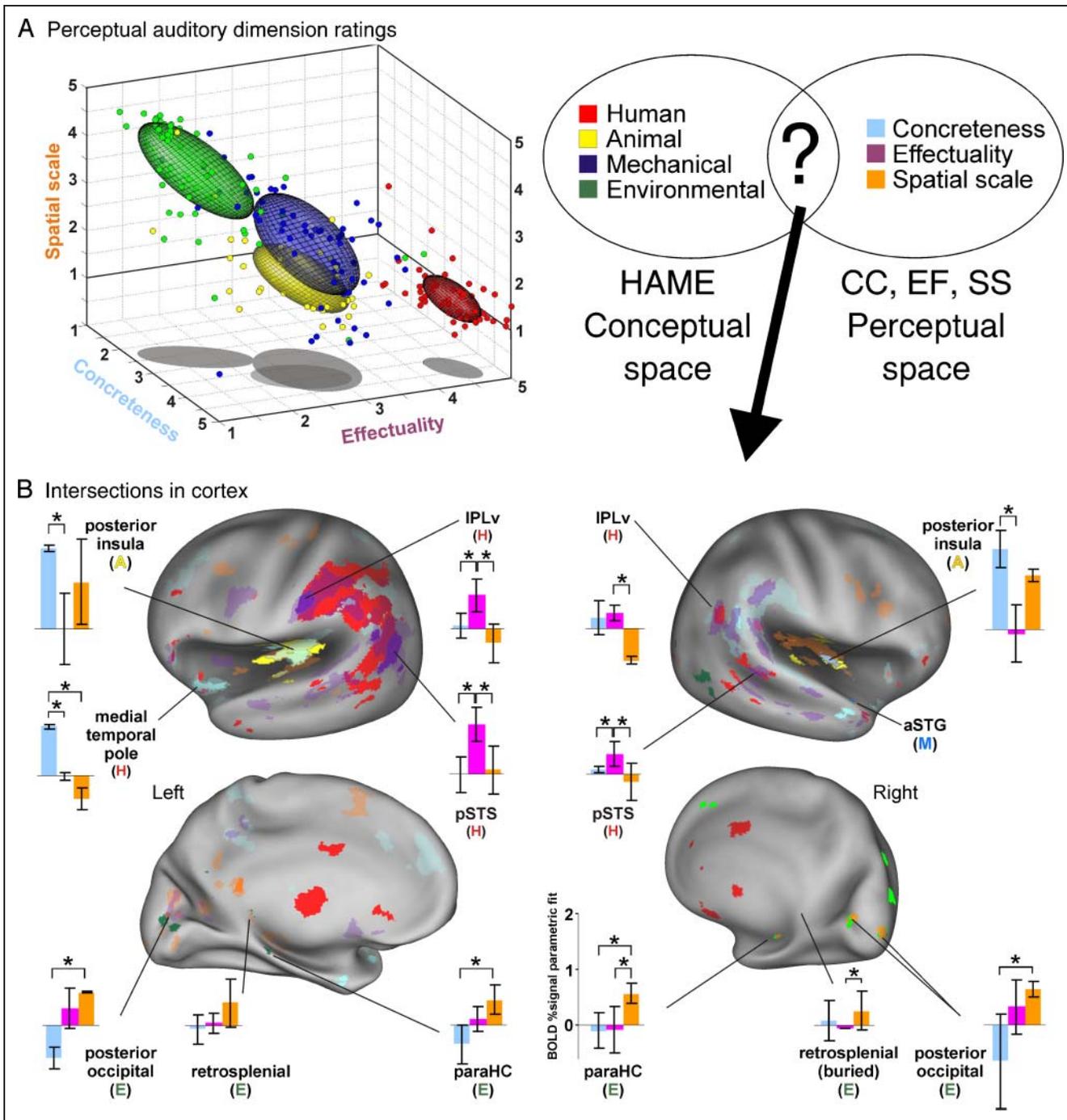
Somewhat unexpectedly, cortical responses to hearing environmental sounds after listening practice (e.g., wind, forest fires, rain, ocean waves) led to striking changes in the global cortical pattern of differential network activation (Figures 1 and 2, green hues). In the pretraining condition, the correctly categorized environmental sounds led to significant differential activation that was largely characterized by differences of “negative” differential BOLD signal changes, relative to the other sound categories (Figure 2, light green histograms). This included the cuneus, anterior calcarine cortex, and various visual-related areas such as V7. Interestingly, these were regions that showed the greatest degree of familiarity training effects in the ANOVA for all four categories of sound (cf. Figure 1, light green; Figure 4, all colors). However, after familiarity training, those same correctly categorized environmental sounds led to positive differential BOLD signal activation in different regions (dark green). This included the bilateral medial prefrontal

cortices (dorso-anterior cingulate), precuneus regions, retrosplenial cortex, and the right hemisphere visual motion processing area hMT/V5 (Van Essen, 2005; Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998). Note that both before and after

training, hearing the environmental sounds relative to silent events did yield robust, positive BOLD signals (activation) in many of the other ROIs illustrated (Figure 2, green present in most histograms). Rather, the above-mentioned regions were

**Figure 4.** The effects of training are greater than within-subject variability. Left panel brain images depict significant effects due to training separately for each category of sound [ANOVA;  $F(1, 12) = 9.33, p < .01$ , corrected]. Right panels show significant effects for within-subject variability as a function of stimulus category. For all brain images, dark hues depict positive effects and primary colors depict negative effects. Refer to Figure 1 and text for other details.





**Figure 5.** High-level perceptual attribute analysis of the four conceptual categories of real-world sounds. (A) Likert-scale ratings of 213 (of the 256) sound stimuli color coded by category as in Figure 1. Each dot depicts the group-averaged coordinates for that particular sound stimulus (not all dots are visible from this vantage point). To facilitate visualization of the clusters corresponding colored ellipsoids were charted, which represent 1.5 standard deviation density plots normalized to represent the same relative degree of variance by category. The Spatial Scale dimension ranges from 0 to 5 to facilitate visualization of the projected shadows of the ellipsoids. (B) Group-averaged ( $n = 9$ ) cortical maps showing regions preferential for category membership (same color key as in Figure 1) relative to an overlay of regions sensitive to any one of the high-level perceptual attribute ratings of the same sounds: concreteness (cyan), effectuality (purple), and spatial scale (orange). Histograms depict response profiles (mean  $\pm$  SE, relative to silent events) only to those foci showing direct overlap between regions preferentially activated by both the conceptual- and perceptual-based regression analyses. \*denotes  $t$ -test significance at  $p < .05$ , corrected. Refer to text for other details.

the only ones showing *preferential* activation for environmental sounds relative to the other three categories of sound-source actions. Together, the above results demonstrated that after familiarizing participants with all of the sound stimuli, there was still a robust dissociation of networks preferential for each of the four categories of action sounds: human, animal, mechanical, and environmental.

### Training Effects versus Within-subject Variability

Some of the differences in activation patterns or activation expanses before versus after familiarity training could conceivably have been related to factors other than training, such as within-subject variability in BOLD responses. To assess this possibility, we conducted a two-factor mixed effects ANOVA comparing pre- versus posttraining scanning runs with the first versus second half of the scanning runs within each scanning session. This was performed on a voxelwise basis for each of the four separate categories of sound (refer to Methods). For all categories of sound, the results revealed regions showing either significant negative effects [Figure 4, primary colors;  $F(1, 12) = 9.33$ ,  $p < .01$ ,  $\alpha < .05$ , corrected for whole-brain multiple comparisons] or positive effects (dark hues). Overall, the effect of training showed substantially greater volumetric expanses of BOLD signal variation (178,690 voxels, 3.85% of total imaging volume showing a significant effect to at least one category of sound) relative to within-subject variability (59,197 voxels, 1.27%). The bilateral cuneus, bilateral anterior calcarine, and right occipital–parietal regions showed the greatest effects (mostly negative effects) with familiarity training. Regions showing significant within-subject variability did not show substantial overlap with regions preferentially activated by the familiar human-produced action sounds (cf. Figures 1 and 4, red hues), thereby indicating that within-subject variability did not account for those results. As mentioned earlier, there was significant within-subject variability along bilateral posterior insular cortex that overlapped with portions of the foci preferential for animal action sounds. Thus, the perceptual learning effects with animal action sounds appeared to have the least degree of change in preferential cortical network representations.

### High-level Perceptual Attributes of Different Categories of Action Sounds

Apart from category membership, what are some of the perceptual or semantic attributes that could distinguish the four categories of action sound and potentially drive the cortical organization for how we encode meaning behind every day sounds? Intuitively, the real-world sound stimuli we presented could at least be qualitatively described along perceptual dimensions that related to how, or if, a person might interact with the sound source. For

the present study, we had a subset of the participants rate each of the sounds along three perceptual dimensions (refer to Methods), including its perceived concreteness (Did it evoke the sense of a distinct visual or palpable form?), its effectuality (Could it be produced, affected, or controlled in some way by the listener?), and its spatial scale (Size relative to the observer?). Note, that in our earlier study, participants had rated the sounds for overall life-long familiarity, which showed no significant differences across categories, and for pleasantness, which showed only a slight preference for environmental sounds—we had attempted to balance those perceptual attributes across the four categories (Engel et al., 2009).

Together, the three sets of group-averaged Likert-scale ratings along dimensions of concreteness, effectuality, and spatial scale were plotted as functions of experimentally defined category membership (Figure 5A, colored data points). Density ellipsoids were also plotted, which were constructed based on the means and covariance matrices of the respective category ratings. These three perceptual attribute dimensions were sufficient to distinguish all four conceptual categories of sound from one another, as revealed through two types of analyses. We performed a *K*-means cluster analysis assuming four a priori clusters. This analysis yielded two highly uniform clusters, one predominantly including the human action sounds (98% human, 2% mechanical) and one cluster predominantly including environmental sounds (4% animal, 8% mechanical, 88% environmental). The two other clusters were preferential for the remaining two categories, one for animal sounds (65% animal, 35% mechanical) and one for mechanical sounds (21% animal, 52% mechanical, 27% environmental). The four conceptual categories of sound were also found to be separable along the three dimensions using a MANOVA [whole model difference of vector means, Wilks' Lambda  $F(9, 501) = 91.5$ ,  $p < .0001$ ; and on all individual contrast levels,  $p < .0001$ ]. Both the human (red) and environmental (green) action-sound categories were most readily separated when using these perceptual dimensions. The animal (yellow) and mechanical (blue) action sounds could be distinguished from one another along the dimension of spatial scale (Tukey's HSD;  $p < .0001$ ), but not on concreteness ( $p = .1737$ ) or effectuality ( $p = .2345$ ) alone.

The clustering we revealed was ellipsoidal rather than spherical (Figure 5A). Thus, the three perceptual dimensions had some interdependence and did not represent truly orthogonal dimensions. Nonetheless, the presence of distinct clusters indicated that these dimensions could provide utility in understanding how high-level perceptual attributes of real-world sounds might relate to representations in neural networks, and how a listener may learn to conceptually categorize them. Thus, we next directly tested whether any brain regions showed parametric sensitivity to any of these perceptual dimensions.

We reanalyzed the posttraining fMRI data ( $n = 9$  of 14) using three terms in a multiple linear regression model representing the three perceptual dimensions: concreteness

(CC), effectuality (EF), and spatial scale (SS). Unlike the analysis of category membership describe earlier (Figure 1), here we used the Likert-scale ratings (1 to 5) as multipliers to parametrically model for sensitivity along each perceptual dimension. For direct comparison, a separate analysis of the same fMRI data was modeled using the four conceptual categories of sound (as described earlier for Figure 1 data, but for this subset of 9 participants). The perceptual dimension and category membership regression models were precisely matched in that we only analyzed BOLD signal responses to those sounds correctly categorized and that had successfully been rated along all three dimensions. Our *null* hypothesis was that there would be no relationship between regions activated using category membership as regressors (H, A, M, and E) versus using parametric ratings along perceptual dimensions as regressors (CC, EF, SS).

Results from the above regression analysis revealed robust activation preferential for each of the three perceptual dimensions. Because of the lack of complete orthogonality between dimensions, results from the conjunction analysis represent a conservative approach for identifying brain regions that are significantly preferential for processing information along any one perceptual dimension relative to *both* of the other two (Figure 5B; solid hues cyan, purple, and orange;  $\alpha < .05$ , corrected). Voxels (brain regions) showing significant preferential sensitivity to only one dimension relative to either of the other two dimensions are indicated by transparent hues on the cortical surface models (reflecting slightly less constrained criteria; same color scheme;  $\alpha < .05$ , corrected). These activation foci were overlaid on top of the foci preferentially responsive to category membership.

Cortical regions showing sensitivity to the concreteness quality of real-world sounds (cyan hues) included the bilateral posterior insular regions, anterior insular cortices, and anterior STG regions. Concreteness-sensitive cortices directly overlapped both with a subset of foci preferential for human action sounds (left medial temporal pole and ventral portions of the anterior insulae; cyan overlapping red) and with foci preferential for animal action sounds (bilateral posterior insulae; cyan overlapping yellow). Brain regions showing sensitivity to effectuality ratings (purple) included bilateral parietal and posterior STS regions, which directly overlapped with several foci that were preferential for human-produced action sounds (purple hues overlapping red). Cortices sensitive to the spatial scale ratings of real-world sounds (orange) most strongly involved parahippocampal regions, midline occipital cortices and, to some extent, the bilateral posterior insulae. The parahippocampal and midline occipital regions showed significant direct overlap with regions preferential for environmental sounds (orange hues overlapping green). Thus, the results from the above analyses indicated that there was a significant relationship between cortical representations of category membership and parametric sensitivity to ratings along high-level acoustic perceptual dimensions that are

related to how, or the extent to which, the listeners physically associated with those sound sources.

## DISCUSSION

The present study of hearing perception using real-world sounds yielded two major findings. The first was a verification of the existence of distinct cortical networks preferentially representing four conceptual categories of action sounds—human, animal, mechanical, and environmental—although with some differences based on listener familiarity. The second was that three high-level acoustic perceptual attributes—concreteness, effectuality, and spatial scale—could also distinguish those four conceptual categories of sound psychophysically, and that there was direct overlap between brain regions preferential for category membership and those showing parametric sensitivity to at least one of these three perceptual dimensions. Although a causal relationship remains to be verified, these results suggest that how a listener physically interacts with sound sources (in a multisensory and sensory–motor manner) is related to the cortical networks that are ultimately utilized to represent those acoustic knowledge representations. The four auditory object and event categories, plus their respective cortical network representations and sensitivity to perceptual dimensions, are addressed in turn below.

### Human Action-sound Network

The network of activated regions preferential or selective for human-produced action sounds was greater in overall volume than for any of the other categories of action sounds both before and after the training sessions. As addressed in our earlier study, we interpreted this as support for grounded (embodied) cognition models, wherein the brain effectively encodes sounds relative to associations with the listener’s motor or body representations, which will be greater for human (conspecific) actions in general (Engel et al., 2009). The regions that were most “stable” (present before and after familiarity training) predominantly included those implicated in representing motor actions and/or in relating observed behaviors to one’s own bodily representations or motor schemas (Norman & Shallice, 1980). In particular, even after extensive listening practice and verbal identification, the left-lateralized IFG and IPL regions remained preferentially activated, representing regions that at least roughly corresponded with mirror-neuron systems (Rizzolatti & Craighero, 2004; Gallese & Goldman, 1998; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) and auditory mirror-neuron systems (Ricciardi et al., 2009; Kohler et al., 2002). These systems are thought to play a role in representing a sense of meaning behind observed goal-oriented actions of others as they interact with inanimate objects, whether heard or viewed. Cortical networks may achieve this by probabilistically matching the

heard sound to one's own repertoire of meaningful sound-producing motor actions (Engel et al., 2009; Aglioti, Cesari, Romani, & Urgesi, 2008; Galati et al., 2008; McNamara et al., 2008; Pazzaglia, Pizzamiglio, Pes, & Aglioti, 2008; Lahav et al., 2007; Gazzola et al., 2006; Lewis et al., 2005, 2006; Bidet-Caulet et al., 2005; Iacoboni et al., 2005; Pizzamiglio et al., 2005). Many of the human action sounds we used included implements and tool use. Networks involved with hearing (or viewing) human tool use actions typically involve the same basic networks that represent complex hand and arm action events (for a review, see Lewis, 2006). For instance, the presence of a nonautomated hand tool (e.g., a hammer) may effectively be represented in cortical networks as an extension of one's hand. In support of this contention, right- versus left-handed listeners have mirror-opposite lateralization differences in their organization of cortical networks for tool sound knowledge (Lewis et al., 2006). Thus, although the presence of tools and other hand-held objects that were used in the production of some of our human-produced sound stimuli might have led to some differences in cortical activation profiles, such effects were unlikely to have had a significant impact on the existence of the observed fourfold dissociation of networks for representing the sound categories demonstrated in the present study.

Interestingly, the posttraining left IFG and IPL foci preferential for human action sounds roughly overlapped, respectively, with classically defined Broca's area (*pars opercularis* and *pars orbitalis*) for speech production (MacNeilage, 1998; Hinke et al., 1993; Geschwind, 1965) and Wernicke's area for speech perception (Shuster & Lemieux, 2005; Binder & Price, 2001; Wise et al., 2001). The extent to which symbolic representations for spoken or signed language may be interrelated to those of embodied motor action sequences represented in mirror-neuron systems remains an active area of research (Lingnau, Gesierich, & Caramazza, 2009; Lotto, Hickok, & Holt, 2009; Lewis, 2006; Arbib, 2005; Emmorey et al., 2004; Rizzolatti & Craighero, 2004; Grèzes, Armony, Rowe, & Passingham, 2003; Corballis, 1999). On this note, several of our participants indicated that they would subvocally name some of the sounds (applicable to all four categories) during the scanning sessions, especially after the familiarity training. Some earlier studies have distinguished networks for spoken words from those for semantic processing of environmental sounds (Engelien et al., 2006; Lewis et al., 2004; Thierry et al., 2003). However, other studies indicate that verbs or sentences describing particular categories of actions generally activate much of the same network(s) as the sensory stimuli themselves (Bedny, Caramazza, Grossman, Pascual-Leone, & Saxe, 2008; Kiefer, Sim, Hermberger, Grothe, & Hoenig, 2008; Dick et al., 2007; Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006; Buccino et al., 2005; Tettamanti et al., 2005; Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004). Thus, if some of the fourfold dissociation of networks preferential for processing each respective sound-source category were reflecting linguistic representations (a subvocal

inner dialog strategy) or other conceptual-level representations, then this would have been a category-specific semantic phenomenon in and of itself. This interpretation would also be generally consistent with grounded cognition and cognitive mediation models for how linguistic or other abstracted symbolic representations of sensory events may ultimately become organized in cortical networks (Barsalou, 2008; Vygotsky, 1978).

The bilateral pSTS regions also showed prominent pre- and posttraining activation that was preferential for human action sounds. Activation in these locations was consistent with a role in biological motion processing and, more generally, in the abstraction of semantically meaningful complex action dynamics, whether heard, viewed, or both (Lewis, 2010; Doehrmann et al., 2008; Martin, 2007; Kable, Kan, Wilson, Thompson-Schill, & Chatterjee, 2005; Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Lewis et al., 2004; Kable, Lease-Spellmeyer, & Chatterjee, 2002; Calvert, Campbell, & Brammer, 2000). One possibility is that the pSTS/pMTG complexes may develop to process highly familiar, complex motion attributes in general, however with a preference for human actions because they are typically more behaviorally relevant to a human listener during development. As with the left IPL focus, the bilateral pSTS/pMTG foci are reported to be sensitive to human action sounds in individuals who have never had visual experience (Lewis et al., *in press*; Ricciardi et al., 2009), indicating that these regions may represent domain-specific hubs that are optimal for biological motion processing whether heard, viewed, or audiovisually associated. Our finding that the bilateral IPL and pSTS regions showed parametric sensitivity to the perceptual dimension of effectuality indicates that these regions are associated not only with visual motor affordances (Grèzes & Decety, 2002) but also with one's acoustically perceived ability to manipulate or affect the external object or sound source.

Other regions preferential for human-produced action sounds that showed reduced activation with familiarity training included the basal ganglia, anterior parahippocampal cortex, cingulate cortices, and insular cortices, which are regions that are related to various aspects of motor representations and/or limbic sensory systems (Heimer & Van Hoesen, 2006; Augustine, 1996). For instance, middle and anterior insular cortices are thought to play a crucial role in acquiring internal models of one's own behaviors, including meta-representations of one's self, as well as modeling the behaviors of others (Craig, 2009; Mutschler et al., 2007). Taken together, the above findings support the idea that embodied representations largely mediate a sense of meaning or recognition behind heard human-produced actions, again supporting grounded cognition theories for knowledge representations (Barsalou, 2008; Barsalou et al., 2003).

Why were activation expanses generally reduced after familiarity training? One possibility was that after training the participants may have become less interested in attending to the highly familiar action-sound stimuli. A more plausible possibility was that extensive familiarization with the

sounds may have led to Hebbian-like reweighting of the network representations (Doshier & Lu, 2009; Ma, 1999; Hopfield, 1995). This would then allow the representations to settle more quickly and/or to rely on reduced expanses of interconnected cortical regions, consistent with facilitation and/or sharpening models of perceptual learning (Mukai et al., 2007; Grill-Spector et al., 2006; James & Gauthier, 2006). Most ROIs that showed overlap between the pre- and posttraining sessions (for human and also animal and mechanical categories) revealed comparable activation response magnitudes, yet a decrease in overall cortical activation expanses (cf. Figure 2 vs. Figure 3 histograms). In this regard, results of the present study appeared to be more consistent with a sharpening model of perceptual learning (Grill-Spector et al., 2006), but here germane to networks mediating hearing perception.

Another account of the present results regarding the post-training stabilized network preferential for human-produced action sounds is in the context of auditory “what–where” models (Recanzone & Cohen, 2009; Tardif, Spierer, Clarke, & Murray, 2008; Ahveninen et al., 2006; Barrett & Hall, 2006; Clarke et al., 2002; Maeder et al., 2001; Rauschecker, 1998a). In the simplified form of this model, acoustic signal processing pathways more specialized for encoding patterns of “what” the sound is versus spatial information regarding “where” the sound is located relative to the listener are thought to more definitively diverge shortly after primary auditory cortices. In particular, the processing of vocalizations (i.e., harmonically structured acoustic events) are typically routed laterally and anteriorly relative to the primary auditory cortices (Lewis et al., 2009; Belin et al., 2000; Rauschecker, 1998b). However, sounds produced during speech repetition paradigms are reported to be routed more posteriorly along the superior temporal plane, wherein audiomotor interfaces may link those acoustic representations with facial articulatory gestures (Rauschecker & Scott, 2009; Warren, Wise, & Warren, 2005). Although speculative, one possibility that was consistent with the present results is that acoustic representations routed along medial, posterior, and dorsal pathways (relative to primary auditory cortices) are more amenable to being probabilistically correlated with intermodal invariant sensory attributes (Lewkowicz, 2000); listening experience may enable the strengthening of associations between acoustic cues with coincidentally timed and/or spatially localized cues that are derived separately from tactile–motor systems and, if available, from the visual system (for a review, see Lewis, 2010). This could lead to representations not only of “where” the sound source is located relative to one’s body, but also encode “what” the sound attributes depict in terms of sensory–motor properties relative to one’s motor system (especially human-produced actions). This basic multisensory cortical mechanism may apply not only to speech articulation functions related to the mouth and vocal tract (Liberman & Whalen, 2000; Liberman & Mattingly, 1985), but more generally to all sounds that can be produced by the observer.

### Animal Action-sound Network

The left and right posterior insular regions showed robust differential activation to nonvocal animal action sounds, being preferentially activated both before and after familiarity training, and under different listening task conditions, as reported in our previous study (Engel et al., 2009). Because the posterior insular foci are roughly situated between primary auditory and secondary tactile–motor cortices (Burton, 2002; Foxe et al., 2002), this activation may reflect some form of audiomotor transformation or representation. Animal actions are arguably more difficult to “embody” in terms of our own motor repertoire of sound-producing actions (e.g., flapping our arms to fly or shaking our head and body to dry off). However, observations of such actions may, nonetheless, instill some form of meaningful, goal-oriented motor behaviors that can be roughly modeled through motor emulation (Mutschler et al., 2009; Buccino et al., 2004).

Interestingly, the posterior and anterior insular regions showed prominent parametric sensitivity to the perception of concreteness of the sound source (especially in contrast to effectuality). Additionally, the bilateral posterior insulae appeared to partially overlap regions reported to be activated in response to viewing impossible finger movements (Costantini et al., 2005). Thus, activation of the posterior insulae may relate less to the observer’s motor schemas per se, and more to abstracted representations reflecting interoceptive assessments of behavior associated with the motor actions of animate agents (Craig, 2009). Although the functions of the posterior insulae remain enigmatic, the present data do provide novel evidence for semantic knowledge models, indicating that the category of “living things” can be subdivided by acoustic representations for human versus nonhuman biological actions. To our knowledge, these insular regions have not been reported to be responsive or preferential to visual depictions of nonhuman animals. Thus, the semantic system for representing “living things” appears to be more widely distributed, dynamic, and multimodal than previously recognized.

### Mechanical Action-sound Network

The mechanical action sounds, which were judged as not being directly associated with a biological agent instigating the action, represented a unique category of sound source from an evolutionary perspective, because neural systems will not have had time to develop any domain-specific regions for this sound-source category. Nonetheless, mechanical sounds led to significant preferential activation along the bilateral aSTG and parahippocampal regions. The bilateral aSTG were preferentially activated in all individuals prior to familiarity training, and in most individuals after training. We propose that the aSTG foci, which are located closer to auditory cortex *proper*, are representing “bottom–up” acoustic signal features that contribute to the categorization of this type of sound. This could include,

for instance, processing of the more regular temporal cadences and/or specific frequency content of motors and automated machinery (e.g., the rhythm of a ticking stopwatch or helicopter in flight).

Left parahippocampal cortex remained sensitive to mechanical action sounds after familiarity training, but also became more sensitive to environmental sounds after training. Thus, this region was generally preferential for sound sources belonging to the semantic category of “nonliving things.” Additionally, the parahippocampal regions were parametrically sensitive to the perceived spatial scale of action sounds, but were not significantly sensitive to the concreteness or effectuality dimensions. These activated foci were in close proximity to parahippocampal regions involved in processing visual scenes (Haxby et al., 2001; Aguirre, Zarahn, & D’Esposito, 1998; Epstein & Kanwisher, 1998; Penfield & Milner, 1958), for which analogous structures in the macaque are reported to subserve scale-invariant visual object processing (Hung, Kreiman, Poggio, & DiCarlo, 2005). Thus, one possibility is that the parahippocampal regions preferentially function to represent visual and acoustic sensory inputs that cannot be embodied by the listener—the actions or content of which may reside on a wide range of spatial scales relative to the observer’s body. The non-embodiable nature and variable spatial scale attributes of our mechanical sound stimuli applied to several examples of our environmental sounds as well, yet environmental sounds preferentially activated other cortical regions, as addressed next.

### Environmental Action-sound Networks

Surprisingly, increased familiarity with the environmental sounds led to some dramatic shifts in the location of regions showing preferential cortical network activation to this category of sound. Prior to training, this included posterior midline regions such as the cuneus and anterior calcarine cortices. After training, the network preferentially activated bilateral medial prefrontal cortices, left precuneus and retrosplenial cortex, the right hMT/V5 region, and various posterior occipital regions (cf. Figures 1 and 5, dark green hues). These changes in network representations may reflect one or multiple factors, which are discussed below in the contexts of acoustic crowding, reverse hierarchy theories of perceptual learning, and default-mode networks.

The medial prefrontal (anterior cingulate) activation preferential for environmental sounds may have reflected a form of “acoustic crowding,” similar to the visual crowding effects associated with visual object discrimination difficulty (Gerlach, 2009; Gale, Done, & Frank, 2001; Snodgrass & McCullough, 1986). Acoustically, the environmental sounds had relatively fewer changes in signal energy across frequency bands (not shown), and thus, may be better described in terms of acoustic textures (McDermott, Oxenham, & Simoncelli, 2009; Reddy, Ramachandra, Kumar, & Singh,

2009; Gygi, Kidd, & Watson, 2004), such as the relatively slow spectro-temporal modulations that occur with tree branches blowing in the wind or the flow of a waterfall. Although not explicitly required by our task, the medial prefrontal activation might have effectively taken on a greater role in discriminating the specific identity of these sounds after familiarity training. Thus, greater attentional demand or conflict monitoring (Sohn, Albert, Jung, Carter, & Anderson, 2007; Jansma, Ramsey, Slagter, & Kahn, 2001) may have been evoked by individuals who were attempting to more precisely distinguish among our set of environmental sound stimuli.

Environmental sounds also differentially activated portions of the cuneus, retrosplenial, and anterior calcarine regions before and/or after familiarity training, although with regional dependence on the participant’s degree of familiarity. Interestingly, these regions showed some of the greatest overall effects of training across all four categories of action sounds (Figure 4). One possibility is that these changes are related to perceptual learning in the context of reverse hierarchy theories. In vision literature, practice-induced improvements in learning are proposed to begin at high-level areas of the visual system, and the formation of specific representations entails encoding backward to the input levels, which have better signal-to-noise ratios (Ahissar, Nahum, Nelken, & Hochstein, 2009; Ahissar & Hochstein, 2004). Despite being presented in the auditory modality, these posterior midline foci overlapped with regions that have been associated with high-level visual perception and visual imagery (Cavanna & Trimble, 2006; Kosslyn, Thompson, Sukel, & Alpert, 2005; Ishai, Ungerleider, & Haxby, 2000; Fletcher et al., 1995) and with processing related to retrieving visual spatial contexts from long-term memory (Burgess, Maguire, Spiers, & O’Keefe, 2001). In addition, the right hMT/V5 visual motion area showed increased preferential activation to environmental sounds, which may have reflected associations with visual motion properties (e.g., visualizing tree branches blowing in the wind, or a waterfall). An intriguing possibility is that some of the presumed high-level “visual areas” along the anterior occipital lobe may be regarded more as regions that perform supramodal or metamodal operations (Pascual-Leone & Hamilton, 2001). These regions may enable the encoding and/or monitoring of dynamic situational relationships with objects (or agents), regardless of the sensory input modality, consistent with their reported roles in mental imagery and episodic memory retrieval (Wakusawa et al., 2009; Walther, Caddigan, Fei-Fei, & Beck, 2009; Hassabis, Kumaran, & Maguire, 2007; Cavanna & Trimble, 2006).

Prior to familiarity training, the environmental sounds differentially activated the bilateral precuneus and anterior calcarine foci, which appeared to at least partially overlap with reported “default-mode” networks (Mason et al., 2007; Burton et al., 2004; Raichle et al., 2001; Shulman et al., 1997). One possibility was that, prior to familiarity training, the environmental sounds might have been more easily or

quickly determined to be “nonhuman” (our button response instructions emphasized accuracy over speed, and thus, the reaction time data did not properly address this issue). If so, then the simple task of judging whether the environmental sounds were produced by a human or not may have been easier to perform, such that the default-mode networks were not “turned off” as strongly when attending to environmental sounds as when attending to the other three more engaging (concrete and effectual) categories of sound-source events.

Another component of default-mode networks, involving the medial prefrontal regions together with precuneus and retrosplenial regions, includes processing related to self-referential and introspective-oriented mental activity (Gusnard, Akbudak, Shulman, & Raichle, 2001). The environmental sounds we used tended to be less “embodiable,” which may represent a key high-level perceptual attribute that the central nervous system utilizes in order to encode acoustic representations and, consequently, convey a sense of meaning or purpose to the listener. In other words, the actions of wind, forest fire, ocean waves, and thunder are typically well out of a person’s ability to directly physically control or influence. Thus, they are less readily represented in cortical networks relating to motor representations of one’s self, or to repertoires of sound-producing motor actions. They are also less amenable to being encoded in terms of purposeful intentions, properties, or behaviors; in this regard, machines, as well as humans and animals, may be considered as producing actions that have a purpose. Thus, default-mode networks may incorporate as-

pects of how embodiable and/or purposeful an acoustic event may be, and consequently, engage processes that assess whether the observer might need to interact with the sound source. Although speculative, this category-specific difference in sound-source encoding may explain why environmental sounds, especially subdued “sounds of nature,” are so often used as an aid for relaxation and stress reduction—because they can help take one’s mind off of one’s self.

In summary, we presented an identical set of sounds to participants before and after familiarity training, and in both conditions, revealed a fourfold dissociation of networks representing category-specific knowledge of human, animal, mechanical, and environmental sound sources. These global network dissociations were at least qualitatively similar to reported category-specific knowledge representations of visual objects. Thus, the existence of distinct networks for acoustic subcategories of living and nonliving things indicates that the semantic system for conceptual knowledge is more multisensory and widespread than previously recognized. In particular, these results suggest that how a listener interacts with a particular sound source—its concreteness as a distinct object or sound source, its effectuality or the extent to which the listener could affect, produce, or in some way influence the sound source, and the sound source’s range in spatial scale relative to the listener’s body—influences how acoustic knowledge representations become encoded, and why particular cortical networks show, or appear to show, category-specificity at a conceptual level.

## APPENDIX 1: LIST OF SOUND STIMULI AS A FUNCTION OF CATEGORY

<i>Human</i>	<i>Animal</i>	<i>Mechanical</i>	<i>Environmental</i>
applause	bat flapping wings 1	airplane fly by	avalanche
banging on door	bat flapping wings 2	airplane engine starting	bubbling mud
blowing nose 1	bee buzzing around	airplane, prop	bubbling water
blowing nose 2	bees buzzing	airplane, prop2	fire crackling 1
blowing up balloon	butterfly flapping wings	automated metal puncher	fire crackling 2
bongo drums	buzzing insect	bells chimes	fire crackling 3
<i>camera, taking picture</i>	<i>cicada chirp</i>	<i>bells moving</i>	fire crackling 4
clapping hands	cows moving	boat motor	fire in fire place 5
counting change	dog breathing 1	<i>cars passing by</i>	fire in fire place 6
cymbal crash	dog breathing 2	chopper	forest fire
deep breathing	dog breathing heavily	church bells ringing	glacier break 1
<i>dialing on touch tone phone</i>	dog eating biscuit	clock	glacier break 2
<i>dialing telephone, rotary</i>	dog footsteps	clock ticking 1	glacier break 3
door knocker	dog lapping & eating	clock ticking 2	glacier break 4
doorknob 1	dog lapping & licking	clock ticking 3	heavy rain
doorknob 2	dog lapping up water 1	clocks, multiple	heavy rainstorm

**APPENDIX 1** (*continued*)

<i>Human</i>	<i>Animal</i>	<i>Mechanical</i>	<i>Environmental</i>
dribbling/shooting basketball	dog lapping up water 2	<i>coin falling on the table</i>	heavy rain
dribbling basketball	dog licking 1	conveyor	lake water wave ashore
dribbling basketball, echoey	dog licking 2	<i>cuckoo clock</i>	large river
eating apple	dog panting & sniffing	door <i>creaking</i> closed	mud bubbling
eating celery	dog panting 1	egg timer	ocean waves
eating chips 1	dog panting 2	exhaust fan	ocean waves
eating chips 2	dog panting 3	fax arriving	rain fall 1
footstep on hard surface	dog panting heavily	fax machine	rain fall 2
footsteps 1	dog panting, heavy breathing	fax machine, paper coming out	rain fall 3
footsteps 2	dog sniffing	fax or copy machine adjusting	<i>rain running down spout</i>
footsteps on the wood	dog swimming, shakes collar	<i>fax warming up and beeps</i>	rain falling with thunder
footsteps on rough surface	dog trotting 1	film projector reel rolling	river
human gargling	dog trotting 2	fireworks going off	river medium
jumping rope	dog trotting 3	flywheel	<i>rock splash in water</i>
knocking on door	dog trotting 4	garage door opening 1	rocks falling
knocking on wooden door	dog walking	garage door opening 2	<i>rocks splashing in water</i>
<i>money in vending machine</i>	fly buzzing	heavy machine, quiet	rockslide
opening bottle of champagne	<i>ben, caught, flight &amp; vocalizes</i>	helicopter	small brush fire
opening can of beer	<i>ben chased 1</i>	helicopter passing	small waterfall
pouring cereal	<i>ben chased 2</i>	industrial engine running	thunder
pouring juice	<i>ben flap around, vocalizes</i>	industry	water
<i>putting coin in slot machine</i>	<i>ben flapping</i>	industry generator (compressor)	water bubbling 1
raking gravel	hoofed animal footsteps	machinery 1	water bubbling 2
raking something	hoofed animal stampede	machinery 2	water bubbling 3
<i>ringing doorbell</i>	horse drawing carriage	machinery 3	<i>water dripping 1</i>
ripping paper up	horse eating	money from slot machine	water dripping 2
scratching	horse galloping 1	office machine	water dripping 3
<i>setting microwave</i>	horse galloping 2	office printer, printing	water dripping 4
<i>shaving, electric razor</i>	horse trotting 1	pain can lid rolling on floor	water dripping in cave
shuffling cards	horse trotting 2	police car passing	water flow
<i>starting large power tool</i>	horse trotting 3	pressbook	<i>water leaking</i>
<i>start up power tool</i>	horse trotting 4	printer 1	water running
<i>taking picture with Polaroid</i>	insect flying	printer 2	waves
tearing paper off pad	insects buzzing 1	<i>printer warming up</i>	waves, ocean
tennis ball rally	insects buzzing 2	printer, dot matrix	wind
<i>turning on television</i>	large herd passing by	printer, feeding paper	wind blowing 1
typing computer keyboard 1	<i>large mutt</i>	printer, office	wind blowing 2
typing computer keyboard 2	<i>pidgeon flutter</i>	refrigerator motor turning on	wind blowing 3

Downloaded From <https://academic.oup.com/ajph/article-pdf/112/5/826/19123414301000> on 18 May 2021

**APPENDIX 1** (continued)

Human	Animal	Mechanical	Environmental
<i>typing cash register</i>	<i>pidgeon flutter fast</i>	scanner adjusting	wind blowing 4
<i>using a table saw</i>	pigeon flight	stopwatch ticking	wind blowing 5
<i>using hand tools</i>	pigs feeding	train squeal breaks to a stop	wind blowing 6
<i>vacuuming</i>	rattle snake rattling 1	train, freight passing	wind blowing 7
writing on chalkboard 1	rattle snake rattling 2	train, steam engine driving by	wind blowing 8
writing on chalkboard 2	rattle snake rattling 3	washing machine 1	wind blowing 9
writing on chalkboard 3	water fowl flapping wings	washing machine 2	wind blowing 10
writing pencil on paper	<i>woodpecker 1</i>	<i>water going down drain</i>	wind blowing, cold
zippering	woodpecker 2	<i>water going down toilet</i>	wind gusting
zippering up tent	zebra trotting	windshield wiper	wind, fast

Italicized entries denote sound stimuli censored post hoc from all analyses.

**Acknowledgments**

We thank Dr. David Van Essen, Donna Hanlon, and John Harwell for continual development of software for cortical data analyses and presentation with CARET and the SUMS database. We also thank Gerry Hobbs and Doug Ward for assistance with statistical analyses, Ms. Mary Pettit for assistance with our participants, as well as Amy Prostko and Alisa Elliot for editorial and data analysis assistance. This work was supported by the NCRN NIH COBRE grant E15524 (to the Sensory Neuroscience Research Center of West Virginia University).

Reprint requests should be sent to James W. Lewis, Department of Physiology and Pharmacology, PO Box 9229, West Virginia University, Morgantown, WV 26506, or via e-mail: jlewis@hsc.wvu.edu.

**REFERENCES**

- Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*, 1109–1116.
- Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to “building” stimuli: Evidence and implications. *Neuron*, *21*, 373–383.
- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*, 457–464.
- Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *364*, 285–299.
- Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., et al. (2006). Task-modulated “what” and “where” pathways in human auditory cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 14608–14613.
- Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex*, *5*, 544–554.
- Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex*, *17*, 2601–2608.
- Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*, 105–124; discussion 125–167.
- Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research, Brain Research Reviews*, *22*, 229–244.
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Jacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*, 1818–1823.
- Barrett, D. J., & Hall, D. A. (2006). Response preferences for “what” and “where” in human non-primary auditory cortex. *Neuroimage*, *32*, 968–977.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.
- Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84–91.
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience*, *7*, 1190–1192.
- Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts are more than percepts: The case of action verbs. *Journal of Neuroscience*, *28*, 11347–11353.
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Research, Cognitive Brain Research*, *13*, 17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.
- Bidet-Caulet, A., Voisin, J., Bertrand, O., & Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *Neuroimage*, *28*, 132–139.
- Binder, J., Frost, J., Hammeke, T., Bellgowan, P., Springer, J., Kaufman, J., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*, 512–528.
- Binder, J. R., & Price, C. J. (2001). Functional neuroimaging of language. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 187–251). Cambridge, MA: MIT Press.
- Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., et al. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. *Journal of Cognitive Neuroscience*, *16*, 114–126.

- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Brain Research, Cognitive Brain Research*, *24*, 355–363.
- Buckner, R. L., Petersen, S. E., Ojemann, J. G., Miezin, F. M., Squire, L. R., & Raichle, M. E. (1995). Functional anatomical studies of explicit and implicit memory retrieval tasks. *Journal of Neuroscience*, *15*, 12–29.
- Burgess, N., Maguire, E. A., Spiers, H. J., & O'Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *Neuroimage*, *14*, 439–453.
- Burton, H. (2002). Cerebral cortical regions devoted to the somatosensory system: Results from brain imaging studies in humans. In R. J. Nelson (Ed.), *The somatosensory system: Deciphering the brain's own body image* (pp. 27–72). New York: CRC Press.
- Burton, H., Snyder, A. Z., & Raichle, M. E. (2004). Default brain functionality in blind people. *Proceedings of the National Academy of Sciences, U.S.A.*, *101*, 15500–15505.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*, 649–657.
- Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, *15*, 1243–1249.
- Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, *7*, 354–361.
- Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: the animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*, 1–34.
- Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*, 465–471.
- Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*, 564–583.
- Clarke, S., Thiran, A. B., Maeder, P., Adriani, M., Vernet, O., Regli, L., et al. (2002). What and where in human audition: Selective deficits following focal hemispheric lesions. *Experimental Brain Research*, *147*, 8–15.
- Corballis, M. C. (1999). The gestural origins of language. *American Scientist*, *87*, 138–145.
- Costantini, M., Galati, G., Ferretti, A., Caulo, M., Tartaro, A., Romani, G. L., et al. (2005). Neural systems underlying observation of humanly impossible movements: An fMRI study. *Cerebral Cortex*, *15*, 1761–1767.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.
- Craig, A. D. (2009). How do you feel—Now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*, 59–70.
- Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, R. D. (1996). A neural basis for lexical retrieval. *Nature*, *380*, 499–505.
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*, 179–229.
- De Lucia, M., Camen, C., Clarke, S., & Murray, M. M. (2009). The role of actions in auditory object discrimination. *Neuroimage*, *48*, 475–485.
- De Renzi, E., & Lucchelli, F. (1994). Are semantic systems separately represented in the brain? The case of living category impairment. *Cortex*, *30*, 3–25.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *93*, 13494–13499.
- Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Benterovato, S., D'Amico, S., et al. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: Evidence from functional magnetic resonance imaging and lesion data. *Journal of Cognitive Neuroscience*, *19*, 799–816.
- Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J., & Altmann, C. F. (2008). Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia*, *46*, 2776–2786.
- Dosher, B. A., & Lu, Z. L. (2009). Hebbian reweighting on stable representations in perceptual learning. *Learning & Perception*, *1*, 37–58.
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473.
- Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, *7*, 89–97.
- Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, *91*, 451–471.
- Emmorey, K., Grabowski, T., McCullough, S., Damasio, H., Ponto, L., Hichwa, R., et al. (2004). Motor-iconicity of sign language does not alter the neural systems underlying tool and action naming. *Brain and Language*, *89*, 27–37.
- Engel, L. R., Frum, C., Puce, A., Walker, N. A., & Lewis, J. W. (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., et al. (2006). Functional neuroanatomy of non-verbal semantic sound processing in humans. *Journal of Neural Transmission*, *113*, 599–608.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Erickson, K. I., Colcombe, S. J., Wadhwa, R., Bherer, L., Peterson, M. S., Scalf, P. E., et al. (2007). Training-induced functional activation changes in dual-task processing: An fMRI study. *Cerebral Cortex*, *17*, 192–204.
- Farah, M. J. (2004). *Visual agnosia*. Cambridge, MA: MIT Press.
- Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage*, *23*, 840–848.
- Fletcher, P. C., Frith, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S., & Dolan, R. J. (1995). The mind's eye: Precuneus activation in memory-related imagery. *Neuroimage*, *2*, 195–200.
- Fließbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *Neuroimage*, *32*, 1413–1421.
- Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., et al. (2002). Auditory–somatosensory multisensory processing in auditory association cortex: An fMRI study. *Journal of Neurophysiology*, *88*, 540–543.
- Frith, C. D., & Frith, U. (1999). Interacting minds: A biological basis. *Science*, *286*, 1692–1695.
- Galati, G., Comitteri, G., Spitoni, G., Aprile, T., Di Russo, F., Pitzalis, S., et al. (2008). A selective representation of the meaning of actions in the auditory mirror system. *Neuroimage*, *40*, 1274–1286.
- Gale, T. M., Done, D. J., & Frank, R. J. (2001). Visual crowding and category specific deficits for pictorial stimuli: A neural network model. *Cognitive Neuropsychology*, *18*, 509–550.

- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2, 493–501.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2, 568–573.
- Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, 16, 1824–1829.
- Gerlach, C. (2009). Category-specificity in visual object recognition. *Cognition*, 111, 281–301.
- Geschwind, N. (1965). Disconnexion syndromes in animals and man: I. *Brain*, 88, 237–294.
- Glover, G. H., & Law, C. S. (2001). Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magnetic Resonance in Medicine*, 46, 515–522.
- Gobbini, M. I., & Haxby, J. V. (2006). Neural response to the visual familiarity of faces. *Brain Research Bulletin*, 71, 76–82.
- Goldenberg, G., Hentze, S., & Hermsdorfer, J. (2004). The effect of tactile feedback on pantomime of tool use in apraxia. *Neurology*, 63, 1863–1867.
- Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to “mirror” and “canonical” neurones in the human brain: An fMRI study. *Neuroimage*, 928–937.
- Grèzes, J., & Decety, J. (2002). Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia*, 40, 212–222.
- Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: An fMRI study of implicit processing. *European Journal of Neuroscience*, 17, 2735–2740.
- Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, 5, 887–892.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10, 14–23.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24, 187–203.
- Gron, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., & Riepe, M. W. (2000). Brain activation during human navigation: Gender-different neural networks as substrate of performance. *Nature Neuroscience*, 3, 404–408.
- Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences, U.S.A.*, 98, 4259–4264.
- Gygi, B., Kidd, G. R., & Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America*, 115, 1252–1265.
- Hadjikhani, N., Liu, A. K., Dale, A., Cavanagh, P., & Tootell, R. B. H. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nature Neuroscience*, 1, 235–241.
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping*, 7, 213–223.
- Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience*, 27, 14365–14374.
- Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, 37, 1027–1041.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.
- Heimer, L., & Van Hoesen, G. W. (2006). The limbic lobe and its output channels: Implications for emotional functions and adaptive behavior. *Neuroscience and Biobehavioral Reviews*, 30, 126–147.
- Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science*, 287, 1269–1272.
- Hinke, R. M., Hu, X., Stillman, A. E., Kim, S.-G., Merkle, H., Salmi, R., et al. (1993). Functional magnetic resonance imaging of Broca’s area during internal speech. *NeuroReport*, 4, 675–678.
- Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature*, 376, 33–36.
- Hopfield, J. J., & Brody, C. D. (2000). What is a moment? “Cortical” sensory integration over a brief interval. *Proceedings of the National Academy of Sciences, U.S.A.*, 97, 13919–13924.
- Hopfield, J. J., & Tank, D. W. (1985). “Neural” computation of decisions in optimization problems. *Biological Cybernetics*, 52, 141–152.
- Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310, 863–866.
- Husain, F. T., Lozito, T. P., Ulloa, A., & Horwitz, B. (2005). Investigating the neural basis of the auditory continuity illusion. *Journal of Cognitive Neuroscience*, 17, 1275–1292.
- Husain, F. T., Tagamets, M. A., Fromm, S. J., Braun, A. R., & Horwitz, B. (2004). Relating neuronal dynamics for auditory object processing to neuroimaging activity: A computational modeling and an fMRI study. *Neuroimage*, 21, 1701–1720.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one’s own mirror neuron system. *PLoS Biology*, 3, 529–535.
- Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron*, 28, 979–990.
- James, T. W., & Gauthier, I. (2006). Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Human Brain Mapping*, 27, 37–46.
- Jansma, J. M., Ramsey, N. F., Slagter, H. A., & Kahn, R. S. (2001). Functional anatomical correlates of controlled and automatic processing. *Journal of Cognitive Neuroscience*, 13, 730–743.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14, 201–211.
- Kable, J. W., Kan, I. P., Wilson, A., Thompson-Schill, S. L., & Chatterjee, A. (2005). Conceptual representations of action in the lateral temporal cortex. *Journal of Cognitive Neuroscience*, 17, 1855–1870.
- Kable, J. W., Lease-Spellmeyer, J., & Chatterjee, A. (2002). Neural substrates of action event knowledge. *Journal of Cognitive Neuroscience*, 14, 795–805.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311.
- Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, 28, 12224–12230.

- Kohler, E., Keysers, C., Umiltà, A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*, 846–848.
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–247.
- Kosslyn, S. M., Thompson, W. L., Sukel, K. E., & Alpert, N. M. (2005). Two types of image generation: Evidence from PET. *Cognitive Affective & Behavioral Neuroscience*, *5*, 41–53.
- Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., & Griffiths, T. D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Computational Biology*, *3*, e100.
- Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, *27*, 308–314.
- Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience*, *30*, 7604–7612.
- Lewis, J. W., Frum, C., Brefczynski-Lewis, J., Talkington, W., Walker, N., Rapuano, K., et al. (in press). Cortical network differences in the sighted versus early blind for recognition of human-produced action sounds. *Human Brain Mapping*.
- Lewis, J. W. (2006). Cortical networks related to human use of tools. *The Neuroscientist*, *12*, 211–231.
- Lewis, J. W. (2010). Audio-visual perception of everyday natural objects: Hemodynamic studies in humans. In M. J. Naumer & P. J. Kaiser (Eds.), *Multisensory object perception in the primate brain* (pp. 155–190). New York: Springer Science+Business Media, LLC.
- Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., & DeYoe, E. A. (2005). Distinct cortical pathways for processing tool versus animal sounds. *Journal of Neuroscience*, *25*, 5148–5158.
- Lewis, J. W., Phinney, R. E., Brefczynski-Lewis, J. A., & DeYoe, E. A. (2006). Lefties get it “right” when hearing tool sounds. *Journal of Cognitive Neuroscience*, *18*, 1314–1330.
- Lewis, J. W., Talkington, W. J., Walker, N. A., Spirou, G. A., Jajosky, A., Frum, C., et al. (2009). Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *Journal of Neuroscience*, *29*, 2283–2296.
- Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex*, *14*, 1008–1021.
- Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*, 281–308.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.
- Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*, 187–196.
- Lingnau, A., Gesierich, B., & Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences, U.S.A.*, *106*, 9925–9930.
- Lissauer, H. (1890/1988). A case of visual agnosia with a contribution to theory. *Cognitive Neuropsychology*, *5*, 157–192.
- Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*, 110–114.
- Ma, J. (1999). The asymptotic memory capacity of the generalized Hopfield network. *Neural Networks*, *12*, 1207–1212.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral & Brain Sciences*, *21*, 499–511; discussion 511–546.
- Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: A human fMRI study. *Neuroimage*, *14*, 802–816.
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45.
- Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science*, *315*, 393–395.
- McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*, 293–299.
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*, 605–610.
- McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*, 310–322.
- McDermott, J. H., Oxenham, A. J., & Simoncelli, E. P. (2009). *Sound texture synthesis via filter statistics*. In IEEE; Workshop on Applications of Signal Processing to Audio and Acoustics. Mohonk, NY, October 18–21, 2009.
- McNamara, A., Buccino, G., Menz, M. M., Glascher, J., Wolbers, T., Baumgartner, A., et al. (2008). Neural dynamics of learning sound–action associations. *PLoS ONE*, *3*, e3845.
- Miller, E. K., Nieder, A., Freedman, D. J., & Wallis, J. D. (2003). Neural correlates of categories and concepts. *Current Opinion in Neurobiology*, *13*, 198–203.
- Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., & Ungerleider, L. G. (2007). Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *Journal of Neuroscience*, *27*, 11401–11411.
- Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *Journal of Neuroscience*, *26*, 1293–1302.
- Mutschler, I., Schulze-Bonhage, A., Glauche, V., Demandt, E., Speck, O., & Ball, T. (2007). A rapid sound–action association effect in human insular cortex. *PLoS ONE*, *2*, e259.
- Mutschler, I., Wieckhorst, B., Kowalewski, S., Derix, J., Wentlandt, J., Schulze-Bonhage, A., et al. (2009). Functional organization of the human anterior insular cortex. *Neuroscience Letters*, *457*, 66–70.
- Naumer, M. J., Doehrmann, O., Muller, N. G., Muckli, L., Kaiser, J., & Hein, G. (2008). Cortical plasticity of audio-visual object representations. *Cerebral Cortex*, *19*, 1641–1653.
- New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences, U.S.A.*, *104*, 16598–16603.
- Nielsen, K. J., Logothetis, N. K., & Rainer, G. (2008). Object features used by humans and monkeys to identify rotated shapes. *Journal of Vision*, *8*, 1–15.
- Norman, D., & Shallice, T. (1980). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self regulation* (pp. 1–18). New York: Plenum.
- Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions. *Journal of the Acoustical Society of America*, *101*, 2951–2963.
- Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research*, *134*, 427–445.

- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987.
- Pazzaglia, M., Pizzamiglio, L., Pes, E., & Aglioti, S. M. (2008). The sound of actions in apraxia. *Current Biology*, *18*, 1766–1772.
- Peissig, J. J., Singer, J., Kawasaki, K., & Sheinberg, D. L. (2007). Effects of long-term object familiarity on event-related potentials in the monkey. *Cerebral Cortex*, *17*, 1323–1334.
- Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, *16*, 1706–1716.
- Penfield, W., & Milner, B. (1958). Memory deficit produced by bilateral lesions in the hippocampal zone. *Archives of Neurology and Psychiatry*, *79*, 475–497.
- Perani, D., Cappa, S. F., Bettinardi, V., Bressi, S., Gorno-Tempini, M., Matarrese, M., et al. (1995). Different neural systems for the recognition of animals and man-made tools. *NeuroReport*, *21*, 1637–1641.
- Pizzamiglio, L., Aprile, T., Spitoni, G., Pitzalis, S., Bates, E., D’Amico, S., et al. (2005). Separate neural systems for processing action- or non-action-related sounds. *Neuroimage*, *24*, 852–861.
- Polk, T. A., Stallcup, M., Aguirre, G. K., Alsop, D. C., D’Esposito, M., Detre, J. A., et al. (2002). Neural specialization for letter recognition. *Journal of Cognitive Neuroscience*, *14*, 145–159.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, U.S.A.*, *98*, 676–682.
- Rauschecker, J. P. (1998a). Parallel processing in the auditory cortex of primates. *Audiology and Neuro-otology*, *3*, 86–103.
- Rauschecker, J. P. (1998b). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, *8*, 516–521.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*, 718–724.
- Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2005). Priming effects in the fusiform gyrus: Changes in neural activity beyond the second presentation. *Cerebral Cortex*, *15*, 787–795.
- Recanzone, G. H., & Cohen, Y. E. (2009). Serial and parallel processing in the primate auditory cortex revisited. *Behavioural Brain Research*, *206*, 1–7.
- Reddy, R. K., Ramachandra, V., Kumar, N., & Singh, N. C. (2009). Categorization of environmental sounds. *Biological Cybernetics*, *100*, 299–306.
- Ricciardi, E., Bonino, D., Sani, L., Vecchi, T., Guazzelli, M., Haxby, J. V., et al. (2009). Do we really need vision? How blind people “see” the actions of others. *Journal of Neuroscience*, *29*, 9719–9724.
- Riede, T., Herzel, H., Hammerschmidt, K., Brunnberg, L., & Tembrock, G. (2001). The harmonic-to-noise ratio applied to dog barks. *Journal of the Acoustical Society of America*, *110*, 2191–2197.
- Riede, T., & Zuberbühler, K. (2003). The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. *Journal of the Acoustical Society of America*, *114*, 1132–1142.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Research, Cognitive Brain Research*, *3*, 131–141.
- Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*, *9*, 11–32.
- Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*, 328–350.
- Ryan, J. D., Moses, S. N., Ostreicher, M. L., Bardouille, T., Herdman, A. T., Riggs, L., et al. (2008). Seeing sounds and hearing sights: The influence of prior learning on current perception. *Journal of Cognitive Neuroscience*, *20*, 1030–1042.
- Saygin, A. P., Leech, R., & Dick, F. (2010). Nonverbal auditory agnosia with lesion to Wernicke’s area. *Neuropsychologia*, *48*, 107–113.
- Scott, S. K. (2005). Auditory processing: Speech, space and auditory objects. *Current Opinion in Neurobiology*, *15*, 197–201.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.
- Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Meizen, F. M., & Raichle, M. E. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience*, *9*, 648–663.
- Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, *93*, 20–31.
- Silveri, M. C., Gainotti, G., Perani, D., Cappelletti, J. Y., Carbone, G., & Fazio, F. (1997). Naming deficit for non-living items: Neuropsychological and PET study. *Neuropsychologia*, *35*, 359–367.
- Snodgrass, J. G., & McCullough, B. (1986). The role of visual similarity in picture categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 147–154.
- Sohn, M. H., Albert, M. V., Jung, K., Carter, C. S., & Anderson, J. R. (2007). Anticipation of conflict monitoring in the anterior cingulate cortex and the prefrontal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *104*, 10330–10334.
- Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, *19*, 498–502.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Tardif, E., Spierer, L., Clarke, S., & Murray, M. M. (2008). Interactions between auditory “what” and “where” pathways revealed by enhanced near-threshold discrimination of frequency and position. *Neuropsychologia*, *46*, 958–966.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–281.
- Thierry, G., Giraud, A. L., & Price, C. (2003). Hemispheric dissociation in access to the human semantic system. *Neuron*, *38*, 499–506.
- Tougas, Y., & Bregman, A. S. (1990). Auditory streaming and the continuity illusion. *Perception & Psychophysics*, *47*, 121–126.
- Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006). Locating the initial stages of speech–sound processing in human temporal cortex. *Neuroimage*, *31*, 1284–1296.
- Van Essen, D. C. (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage*, *28*, 635–662.

