Neuropsychologia 49 (2011) 1807–1815

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Shape from sound: Evidence for a shape operator in the lateral occipital cortex

Thomas W. James[a,b,c,*], Ryan A. Stevenson[a,b,d], Sunah Kim[b,c,e], Ross M. VanDerKlok[a], Karin Harman James[a,b,c]

[a] Department of Psychological and Brain Sciences, Indiana University, United States
[b] Program in Neuroscience, Indiana University, United States
[c] Cognitive Science Program, Indiana University, United States
[d] Department of Speech and Hearing Sciences, Vanderbilt School of Medicine, United States
[e] Vision Science Program, School of Optometry, University of California, Berkeley, United States

ARTICLE INFO

Article history:
Received 30 November 2010
Received in revised form 21 February 2011
Accepted 3 March 2011
Available online 11 March 2011

Keywords:
Multisensory
fMRI
Visual cortex
Metamodal
Object recognition

ABSTRACT

A recent view of cortical functional specialization suggests that the primary organizing principle of the cortex is based on task requirements, rather than sensory modality. Consistent with this view, recent evidence suggests that a region of the lateral occipitotemporal cortex (LO) may process object shape information regardless of the modality of sensory input. There is considerable evidence that area LO is involved in processing visual and haptic shape information. However, sound can also carry acoustic cues to an object's shape, for example, when a sound is produced by an object's impact with a surface. Thus, the current study used auditory stimuli that were created from recordings of objects impacting a hard surface to test the hypothesis that area LO is also involved in auditory shape processing. The objects were of two shapes, rods and balls, and of two materials, metal and wood. Subjects were required to categorize the impact sounds in one of three tasks, (1) by the shape of the object while ignoring material, (2) by the material of the object while ignoring shape, or (3) by using all the information available. Area LO was more strongly recruited when subjects discriminated impact sounds based on the shape of the object that made them, compared to when subjects discriminated those same sounds based on material. The current findings suggest that activation in area LO is shape selective regardless of sensory input modality, and are consistent with an emerging theory of perceptual functional specialization of the brain that is task-based rather than sensory modality-based.

For decades, the principal organizational theory for the functions of the occipital, temporal, and parietal cortices was based on the modality of sensory input. The posterior cortex was grossly separated into visual, auditory, and somatosensory systems (Frackowiak et al., 2004; Kolb & Whishaw, 2003), and it was usually only within those systems that the cortex was further separated based on more specific perceptual and cognitive functioning (for example, see Van Essen, Casagrande, Guillery, & Sherman, 2005). More recently, new evidence has made it clear that sensory processing occurs in isolated systems only at the very lowest levels (Foxe & Schroeder, 2005). An alternative theory to the parallel processing of discrete sensory inputs is the "metamodal" brain, for which the primary organizing principle is task requirements, rather than sensory modality (James, VanDerKlok, Stevenson, & James, 2011; Lacey, Tal, Amedi, & Sathian, 2009; Pascual-Leone & Hamilton, 2001). According to this view, regions of cortex instantiate "operators" that perform a specific calculation or implement a specific cognitive operation. Operators have the capacity to process input from multiple sensory modalities. One condition for a multisensory operator to develop is that the sensory inputs must all contain the type of information necessary for successful calculation. Also, operators develop preferences or weightings for the specific input modalities that provide the most reliable information. With typical development, operators in different individuals will show very similar patterns of preference across sensory modalities, giving the impression that the brain is organized based on sensory modalities, rather than cognitive operations. It is cases of atypical development—especially atypical development of sensory systems—that demonstrate the capacity of operators to complete the same calculations using non-preferred sensory inputs and that provide the most compelling evidence for the metamodal brain hypothesis (Pascual-Leone & Hamilton, 2001). In sum, the metamodal brain hypothesis has two tenets. First, the brain is by nature multisensory, and second, the multisensory nature of operators may be latent. The latent multisensory nature of operators may give the impression that the brain is organized based on sensory-specific functioning. The current work uses the first tenet, that the

* Corresponding author at: 1101E Tenth St., Bloomington, IN 47405, United States. Tel.: +1 812 856 0841; fax: +1 812 855 4691.
E-mail address: thwjames@indiana.edu (T.W. James).

0028-3932/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2011.03.004

brain is inherently multisensory, as a framework for investigating the shape processing operations involved in multisensory object recognition.

In the field of object recognition, it has been suggested that a region of the lateral occipitotemporal cortex (LO) may be the site of an operator that is dedicated to processing volumetric shape (Amedi et al., 2007; Lacey et al., 2009). Several research groups have established that area LO is involved in visual and tactile/haptic recognition of objects. These studies (Amedi, Jacobson, Hendler, Malach, & Zohary, 2002; Amedi, Malach, Hendler, Peled, & Zohary, 2001; James et al., 2002; Kim & James, 2010; Sathian & Zangaladze, 2002; Stilla & Sathian, 2008) report evidence for a sub-region of area LO called the lateral occipital tactile-visual area (LOtv) that is object selective for both visually presented and haptically presented object stimuli compared to texture stimuli. Although objects and textures differ along many dimensions (e.g., curvature, roughness, weight, color, etc.), it is clear from comparing across the studies that the most important dimension influencing selective activation in LOtv is that the object stimuli are discriminated mainly based on their volumetric shape, whereas the textures are not (James et al., 2002; James, Kim, & Fisher, 2007; Stilla & Sathian, 2008; Tal & Amedi, 2009). For instance, in the study by James et al. (2002), novel objects were used and the objects were constructed such that they all had the same texture, hardness, etc. and only differed based on their volumetric shape properties. Using a priming paradigm, the results showed that brain activation in area LO was suppressed when objects were repeated, regardless of whether the objects were presented within or across sensory modalities. Activation showed recovery from suppression when non-repeated objects were presented. The results were taken as evidence that neurons in area LO are tuned to specific shape features of objects, but that the tuning was invariant to the input sensory modality.

For vision and haptics, shape characteristics of objects are salient and shape information is important for successful recognition. Thus, the existence of a common neural substrate, such as area LOtv, for processing shape information across the two sensory systems is not surprising. A shape *operator*, however, should process signals from any sensory system that produces signals that contain shape information, not just the sensory systems for which shape information is the most salient. Recently, it has been suggested that brain regions exist that are selective for objects presented through the auditory modality (Amedi et al., 2007; Beauchamp, Lee, Argall, & Martin, 2004; James et al., 2011; Lewis, Brefczynski, Phinney, Janik, & Deyoe, 2005; Lewis et al., 2004). In most of these studies, sounds of manual tools (e.g., hammer, saw, etc.) were used as stimuli, and subjects were required to recognize the tools based on the sound (Beauchamp et al., 2004; James et al., 2011; Lewis et al., 2005; Lewis et al., 2004). These studies found greater activation with tool sounds than with other sounds in the posterior middle temporal gyrus (pMTG). Of particular interest is that the coordinates of area pMTG and area LO are very similar – both are at the junction of the temporal and occipital lobes – and it is clear that they show considerable overlap. Further evidence that area LO is object selective for sounds comes from a study that used sounds produced by a visual-to-audio sensory substitution device. Subjects listened to audio waveforms that had been transformed from pictures of objects using the sensory substitution device. These "substitution sounds" produced greater activation in area LOtv than control sounds (Amedi et al., 2007).

The results of the studies described above converge to suggest that activation in area LO/pMTG (and perhaps specifically area LOtv) is object selective across three sensory input modalities, vision, touch, and hearing. There is evidence that object selectivity in area LO for vision and haptics is driven by shape, rather than other object characteristics (James et al., 2002). What is missing is evidence that object selectivity for sounds in area LO is also based on the shape characteristics of the objects that made the sounds. The hypothesis that area LO is the site of a shape operator would be strongly supported by results indicating that activation was driven by changes in sounds that were based on manipulations of the shape characteristics of the objects that produced them.

There are many natural classes of auditory stimuli that contain useable information for determining not only an object's shape, but also its size, length, or material composition. It has also been shown that human listeners are capable of using acoustic information to recognize objects (Freed, 1990; Grassi, 2005; Warren & Verbrugge, 1984). Producing sounds that are diagnostic of these characteristics usually requires that the object be involved in an environmental event (Gaver, 1993b), such as when it is struck against a surface or dropped from a height (Gaver, 1993a). In the current study, auditory stimuli were created from recordings of objects impacting a hard surface. The objects were of two shapes, rods and balls, and of two materials, metal and wood. Subjects were required to categorize the impact sounds in one of three tasks, (1) by the shape of the object while ignoring material (i.e., as rods or balls), (2) by the material of the object while ignoring shape (i.e., as metal or wood), or (3) by the four combinations of shape and material (i.e., as a metal rod, wood rod, metal ball, or wood ball). Previous work on visual recognition suggests that shifting subjects' attention from one object property to another (i.e., between shape and material), is sufficient to preferentially activate brain regions involved in processing that specific object property (Cant & Goodale, 2007; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991). Thus, consistent with Amedi et al. (2007), we hypothesized that categorizing the sounds by the shape of the object involved in the impact would preferentially activate area LO and in particular the LOtv.

1. Materials and methods

1.1. Subjects

Subjects included 12 right-handed native English speakers (6 female, mean age = 21.7). All subjects reported normal or corrected-to-normal visual acuity and no history of hearing impairment. The experimental protocol was approved by the Indiana University Institutional Review Board and Human Subjects Committee.

1.2. General procedures

Subjects lay supine in the bore of the MRI with their head in the radio frequency coil and a response pad placed on their right thigh. Stimuli for visual and auditory presentations and timing cues for haptic presentations were delivered using Matlab 5.2 and Psychophysics Toolbox 2.53 (Brainard, 1997; Pelli, 1997) on an Apple Powerbook G4 (Titanium) running Mac OS 9.2. Visual stimuli were projected with a Mitsubishi XL30U LCD projector onto a rear-projection screen located inside the scanner bore behind the subject. Subjects viewed the screen through a mirror located above the head coil. Auditory stimuli were heard through pneumatic headphones. Foam was placed around the headphones inside the headcoil to reduce subject head movement. Haptic stimuli were placed on a "table" by an experimenter who stood in the MRI room. The table rested on the subject's abdomen/thighs and was angled toward the subject to make the stimuli easy to reach. The table had a non-skid surface to prevent the objects from sliding off or moving during manual exploration.

Subjects were tested on two or three different days to complete all of the data collection. Data from the audiovisual action-selective functional localizer and the visuohaptic object-selectivity functional localizer were collected as part of another study, which has been published elsewhere (James et al., 2011).

1.3. Impact sound procedures

Examples of the impact stimuli are shown in Fig. 1. Impact stimuli consisted of audio recordings of objects dropped onto the floor from a height of approximately 1 m. Four objects were used to create the impact sounds. Two of the objects were rods, each 1 cm in diameter and 30 cm long, one made of metal and one of wood. The metal rod was a section of rebar and the wood rod was a section of hardwood dowel. The other two objects were balls, each 3 cm in diameter, one made of metal and one of wood. The metal ball was a large stainless steel marble and the wood ball was made of hardwood. Recordings were made with a handheld digital recorder. Recordings of impacts with each of the four objects were made in three different rooms in the Psychology building and each object was recorded being dropped several times in each room. From this large set of recordings, 24 recordings were selected and

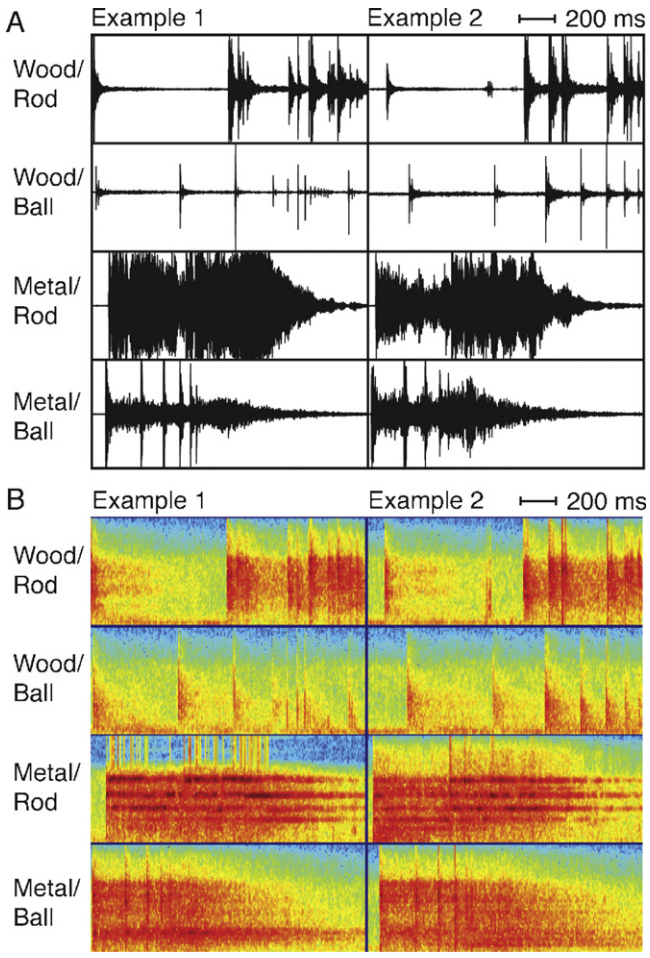

Fig. 1. Waveforms and spectrograms of impact sounds. Waveforms of two examples of each of the four stimulus categories are shown in (a), with time on the horizontal axis (0–1500 ms) and amplitude on the vertical axis. The same eight sounds are shown as spectrograms in (b), with time on the horizontal axis, frequency band on the vertical axis (0–11.6 KHz), and power indicated by the color scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

used in the study as impact sound stimuli. The 24 stimuli were chosen such that there were six sounds for each of the four objects and such that two of those six sounds were recorded in each of the three rooms. The different acoustics and flooring surfaces of the rooms provided variability in examples of the impact sounds, such that subjects could not perform the matching task based on idiosyncratic features of specific recordings. Scrambled nonsense versions of each of the 24 impact sound stimuli were also created. Audio waveforms were partitioned into 10 ms intervals and the bits in half of the intervals (determined randomly) were exchanged with the bits from the other half of the intervals. Intervals were exchanged with the interval that matched it most closely in amplitude. Scrambling the waveforms made them unrecognizable and, subjectively, they sounded similar to noise.

Each subject performed eight runs. The protocol for these runs is shown in Fig. 3a. Each run contained eight 16-s stimulation blocks. These stimulation blocks were interleaved with seven 16-s rest intervals, plus a rest interval at the beginning and at the end of each run. There were eight trials per block. During a trial, a sound stimulus was presented for 1.5 s, followed by .5-s inter-stimulus interval. Subjects performed one of four one-back matching tasks on the last seven trials of each block. An instruction cue was presented during the rest interval preceding the stimulation block. The instruction was one of "shape", "material", "both", or "scrambled". It is worthwhile noting that the same intact impact sound stimuli were presented for shape, material, and intact ("both") blocks. Scrambled sounds, rather than intact sounds, were presented for the scrambled blocks. For shape blocks, subjects performed the matching task based on the shape of the object that made the impact sound. In other words, they matched based on whether the stimulus represented a rod or ball (two-alternative forced choice—2AFC) and ignored whether it was made of metal or wood. For material blocks, subjects performed the opposite task, basing their match judgments on the material of the object that produced the impact sound and ignoring the shape. That is, a 2AFC matching task for whether it was made of metal or wood. For intact ("both") blocks, subjects did not match the impact sounds based on a specific characteristic of the object. Instead, used all of the acoustic information available to them to make a 4AFC matching judgment. That is, they were required to match the intact stimuli based on the four specific alternatives, metal rod, metal ball, wood rod, and wood ball. For the scrambled blocks, the scrambled impact sound stimuli were presented, rather than the intact impact sound stimuli. For the scrambled blocks, the subjects performed the same 4AFC task as for the intact blocks. That is, they were required to match the four specific sounds, but in this case, the sounds were scrambled, rather than intact. The order of the blocks was randomized for each run and subject.

1.4. Visuohaptic object-selectivity procedures

The purpose of these runs was to functionally localize the LOtv part of the LO. The stimuli and procedures for this part of the study have been described previously (Kim & James, 2010). Examples of visual stimuli are shown in Fig. 2c and d and the protocol is shown in Fig. 3b. Briefly, the visual runs used grayscale images of 40 objects and 40 textures. Each stimulus subtended 12° of visual angle. The haptic runs used 20 three-dimensional familiar objects (e.g., cup, book, etc.) and 20 two-dimensional textures (e.g., fabric, sandpaper, etc.), all MR-compatible and sized to be easily explored with the hands. Each subject performed two visual runs and two haptic runs. Each run contained five 16-s blocks of object presentation and five 16-s blocks of texture presentation. These stimulation blocks were interleaved with nine 16-s rest intervals, plus a rest interval at the beginning and at the end of each run. Object and texture stimulation blocks had four trials per block. During a trial, a stimulus was presented for 3 s and followed by 1-s inter-stimulus interval. For haptic trials, subjects received auditory cues to begin and end manual two-handed exploration of the objects. The auditory cues were not necessary for the visual trials – the subjects were cued by the onset and offset of the visual stimuli – but they were included in the visual trials to match the haptic trials. The order of the blocks was randomized.

1.5. Audiovisual action-selectivity procedures

The purpose of these runs was to functionally localize the pMTG part of the LO. The stimuli and procedures for this part of the study have been described previously (James et al., 2011). Examples of stimuli are shown in Fig. 2a and b and the protocol is shown in Fig. 3c. Briefly, stimuli consisted of audio and video recordings of manual actions involving a moveable implement (e.g., hammer, paper cutter, paper towel dispenser, etc.). Separate video and audio files were extracted from the raw recordings, such that they could be presented separately as visual and auditory stimuli. Scrambled nonsense versions of the video and audio signals were also created. Video sequences were scrambled on a frame-by-frame basis. For each frame, the locations of half of the pixels in the image were exchanged with the locations of the other half of the pixels. Each pixel exchanged locations with the pixel that was closest to it in intensity. Audio waveforms were partitioned into 10 ms intervals and the bits in half of the intervals (determined randomly) were exchanged with the bits from the other half of the intervals. Intervals were exchanged with the interval that matched it most closely in amplitude. Each subject performed two visual runs and two auditory runs. Each run contained three 12-s blocks of action presentation and three 12-s blocks of scrambled presentation. These stimulation blocks were interleaved with five 12-s rest intervals, plus a rest interval at the beginning and at the end of each run. Action and scrambled stimulation blocks had eight trials per block. During a trial, a stimulus was presented for 2 s with no inter-stimulus interval. The order of the blocks was randomized. Subjects performed a one-back matching judgment on the last seven stimuli in each block.

1.6. Imaging parameters and analysis

Imaging was carried out using a Siemens Magnetom TRIO 3-T whole-body MRI with eight-channel phased-array head coil. The field of view was 22 cm × 22 cm × 11.2 cm, with an in-plane resolution of 64 × 64 pixels and 33 axial slices per volume (whole brain), creating a voxel size of 3.44 mm × 3.44 mm × 3.4 mm. Voxels were re-sampled to 3 mm × 3 mm × 3 mm during pre-processing. Images were collected using a gradient echo EPI sequence for BOLD imaging (TE = 30 ms, TR = 2000 ms, flip angle = 70°). High-resolution T1-weighted anatomical volumes were acquired using a turbo-flash 3-D sequence (TI = 1100 ms, TE = 3.93 ms, TR = 14.375 ms, flip angle = 12°) with 160 sagittal slices with a thickness of 1 mm and field of view of 256 × 256 (voxel size = 1 mm × 1 mm × 1 mm).

Functional volumes were pre-processed using BrainVoyager™ 2.2.0. Pre-processing steps included linear trend removal, 3-D spatial Gaussian filtering (FWHM 6 mm), slice scan-time correction, and 3-D motion correction. Anatomical volumes were transformed into the common stereotactic space of Talairach and Tournoux using an 8-parameter affine transformation. The eight parameters were the AC and PC points, and six points representing the bounding box of the cortex, which were manually selected. Functional volumes were coregistered to the anatomical volume, thus transforming them into the common stereotactic space.

Data were analyzed using separate random-effects general linear models for the audio impact sounds, the visuohaptic objects and textures, and the audiovisual actions. Multiple runs for each experiment were appended, rather than averaged.

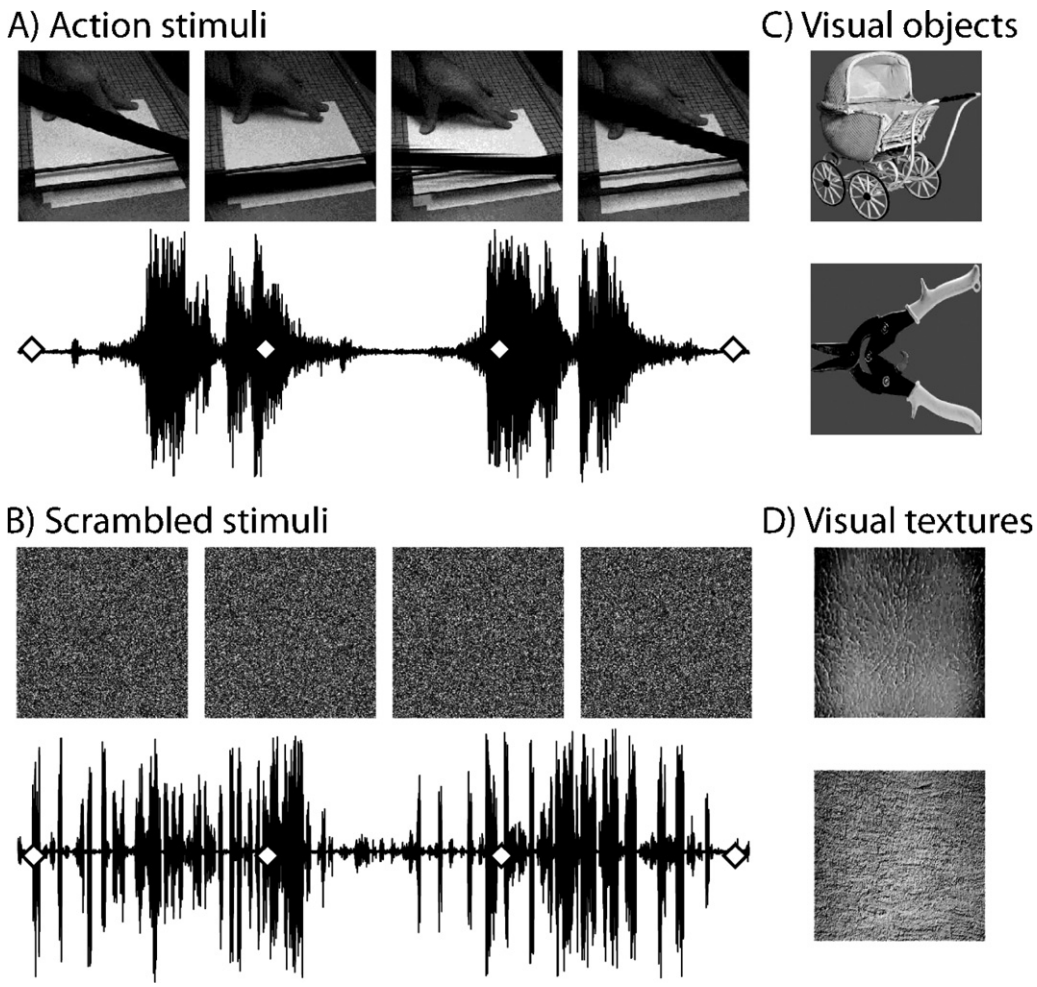

Fig. 2. Stimuli for functional localizer runs. An example of an intact stimulus used for testing action selectivity is shown in (a). Four frames of the video of the paper cutter are depicted with the sound waveform of the paper cutter. The white diamond symbols represent the time points when the video frames were extracted. A scrambled version of the paper cutter is shown in (b). Two examples of intact visual objects used to test for visuohaptic object selectivity are shown in (c). Two examples of visual textures are shown in (d). Haptic stimuli are not shown, but are described in Section 1.

Design matrices were constructed from predictors generated based on the timing of the blocked-design protocols for placement of canonical two-gamma hemodynamic response functions. For the impact sound runs, predictors representing the instruction cue were also included. All whole-brain contrasts were thresholded using a minimum voxel-wise p-value of 0.005 and corrected for multiple tests using a cluster threshold (Forman et al., 1995; Lazar, 2010; Thirion et al., 2007). The minimum number of contiguous voxels required to provide a false positive rate of 5% was estimated using the BrainVoyager QX Cluster-Level Statistical Threshold Estimator plugin ($p = 0.005$, alpha = 0.05; (Goebel, Esposito, & Formisano, 2006)). There were slight variations in the estimate across maps, but for consistency, we chose the most conservative estimate of a minimum of eight 3 mm × 3 mm × 3 mm voxels (216 mm^3). Whole-brain maps were re-sampled (using linear interpolation) from 3 mm × 3 mm × 3 mm to 1 mm × 1 mm × 1 mm to be shown at the same spatial resolution of the anatomical volumes. Labels for brain regions shown in the table were found with the Talairach Daemon (http://www.talairach.org/applet/) using the nearest coordinate located in grey matter.

2. Results

2.1. Behavioral results

Accuracy was measured for all of the functional runs. As expected, accuracy was at ceiling for the one-back matching judgments in the visuohaptic object-selectivity runs and the audiovisual action-selectivity runs. Accuracy results for the one-back matching judgments with the impact sounds in the auditory shape-selectivity runs are shown in Fig. 4. Accuracy was relatively poor for all conditions (<70%), but was significantly above chance as assessed by one-sample t-tests (all $t_{(11)} > 4.95$, $p < 0.001$). We attribute the moderate performance to the fact that the stimuli were highly similar to each other and that they were partially masked by the presence of the acoustic noise produced by the MRI. A one-way ANOVA showed that significant differences in accuracy existed among the four conditions ($F_{(3,33)} = 9.2$, $p = 0.001$, Greenhouse–Geisser corrected). Paired t-tests showed that the 4AFC matching task was more accurate with intact impact sounds than with scrambled sounds ($t_{(11)} = 2.54$, $p = 0.03$) and that the 2AFC task was more accurate when it was shape-based than material-based ($t_{(11)} = 2.42$, $p = 0.03$). The intact 4AFC task showed the best performance of the four conditions ($t_{(11)} = 2.60$, $p = 0.03$). We attribute the better performance with the 4AFC task to the fact that subjects could attend to any or all of the stimulus characteristics to make their judgment, whereas with the 2AFC tasks, the subjects were forced to attend to a specific set of characteristics (or possibly just a single characteristic) while actively ignoring a potentially orthogonal set of characteristics.

A subset of subjects were given a verbal debriefing at the end of the session to determine if any explicit strategies were used to perform the different tasks with the impact sounds. Subjects had difficulty articulating any strategies used with the 2AFC material task and both of the 4AFC tasks. However, with the shape task, subjects consistently indicated using the pattern of impacts across time to differentiate balls from rods. During stimulus generation, when

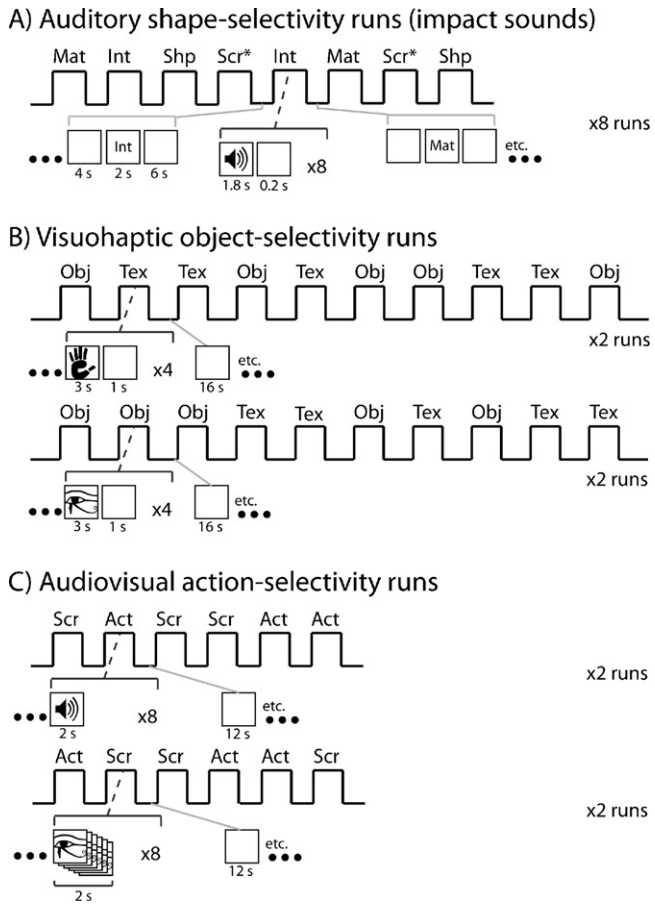

Fig. 3. Schematic of protocols for functional runs. The timing of protocols is depicted with boxcar functions that represent stimulation and rest intervals (blocks). Time is represented horizontally and the functions are drawn to scale. Above each stimulation interval is a label for task performed during that block. Below each protocol is a more detailed depiction of the trial structure within each block. Blank boxes indicate rest periods. Visual, haptic, and auditory stimuli are indicated by a box with an eye, hand, or speaker symbol, respectively. Below each box is a number representing the number of seconds that the stimulus in the box is presented for. If a stimulus cycle is repeated during a block, that is indicated by "x#" after the boxes. The number of runs of each protocol for each subject is shown to the right (i.e., "x# runs"). The protocols for runs with impact sounds are shown in (a). Shp indicates that subjects performed a 2AFC shape matching task, Mat indicates a 2AFC material matching task, Int indicates a 4AFC task on intact sounds, and Scr* indicates a 4AFC task on scrambled sounds. Note that the Scr* task was the only one of the four that used different stimuli. The protocols for runs testing visuohaptic object selectivity are shown in (b). Obj indicates that the stimuli were familiar objects (haptic) or static pictures of familiar objects (visual), and Tex indicates that the stimuli were familiar textures (haptic) or static pictures of textures (visual). The protocols for runs testing audiovisual action selectivity are shown in (c). Act indicates that the stimuli were video or audio of object-directed actions, and Scr indicates that the stimuli were scrambled versions of the video or audio.

the rods and balls were dropped, they bounced and made multiple impacts with the surface they were dropped on. These impacts are seen in the sound waves and spectrograms (Fig. 1) as transients. The timing of the transients depended mostly on the shape of the object, rather than on its material. It seems likely that the important information in the sounds for identifying object shape was the pattern across time of the transients. The cues used to identify the material of the object are more ambiguous. The fundamental frequency of the wood balls was in a different range (800–900 Hz) than the other three stimulus types (1200–1300 Hz). Thus, fundamental frequency would help identify one of the four object types, but by itself would not help in the 2AFC shape or material tasks. Thus, it is likely that subjects used the timbre of the sounds to differentiate

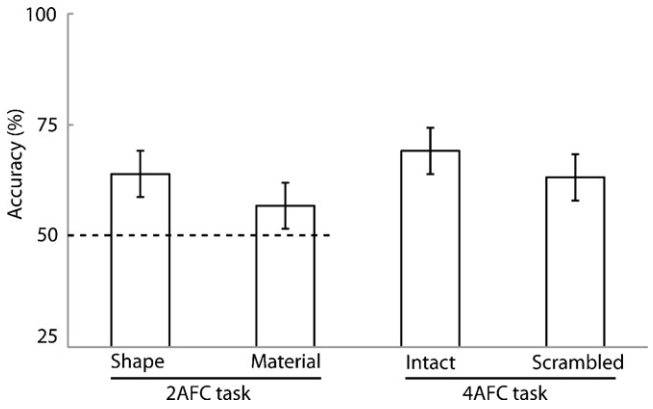

Fig. 4. Accuracy as a function of task for the impact sounds experiment. The dashed line through the 2AFC task represents chance performance (50%) for that task. Chance performance on the 4AFC task was 25%. Error bars are 95% confidence intervals.

the materials, but which aspect of the timbre was difficult for the subjects to articulate.

2.2. Imaging results

Fig. 5 shows the main results of the four contrasts of interest. As hypothesized, the activation in area LO/pMTG was greater when impact sounds were categorized based on shape compared to when they were categorized based on material (Fig. 5a). Specifically, activation was found at the junction between the posterior middle temporal gyrus and the anterior middle occipital gyrus in the right hemisphere. When subjects were allowed to categorize the sounds using all available information (i.e., 4AFC task), activation was found along the superior temporal sulcus (Fig. 5b), also in the right hemisphere. This cluster was clearly superior and anterior to the shape-selective area LO activation (Fig. 5e). More details of these and other clusters are shown in Table 1.

The difference between shape and material in area LO could have been due to the difference in behavioral performance across the two conditions. There is evidence that recognition accuracy can influence activation in area LO, with greater accuracy producing greater activation (James, Culham, Humphrey, Milner, & Goodale, 2003; James & Gauthier, 2006). The shape-matching task was performed more accurately than the material-matching task, which may explain the greater activation with shape matching. However, comparing the pattern of activation with the pattern of accuracy across the four impact sound conditions does not support this alternate hypothesis. Most strikingly, a contrast comparing the most accurate condition (4AFC intact) with the least accurate condition (2AFC material) produced no significant clusters, even at a very relaxed statistical threshold ($t = 2.0$, uncorrected).

Area LOtv was functionally localized using the established practice of a conjunction (logical AND) of two contrasts: visual objects minus textures AND haptic objects minus textures (Amedi et al., 2002; Amedi et al., 2001; Kim & James, 2010). This conjunction contrast produced significant activation in area LO (Fig. 5c), which overlapped with the shape-selective cluster in the right hemisphere (Fig. 5b and e). More details of these clusters are shown in Table 1.

Another conjunction contrast was performed for audiovisual action stimuli. The two contrasts were auditory actions minus scrambled and visual actions minus scrambled. This conjunction contrast also produced significant clusters in area LO in the left and right hemisphere (Fig. 5d). In the right hemisphere, the action-selective cluster in area LO overlapped with the shape-selective cluster in area LO and with area LOtv (Fig. 5e). More details of these clusters are shown in Table 1.

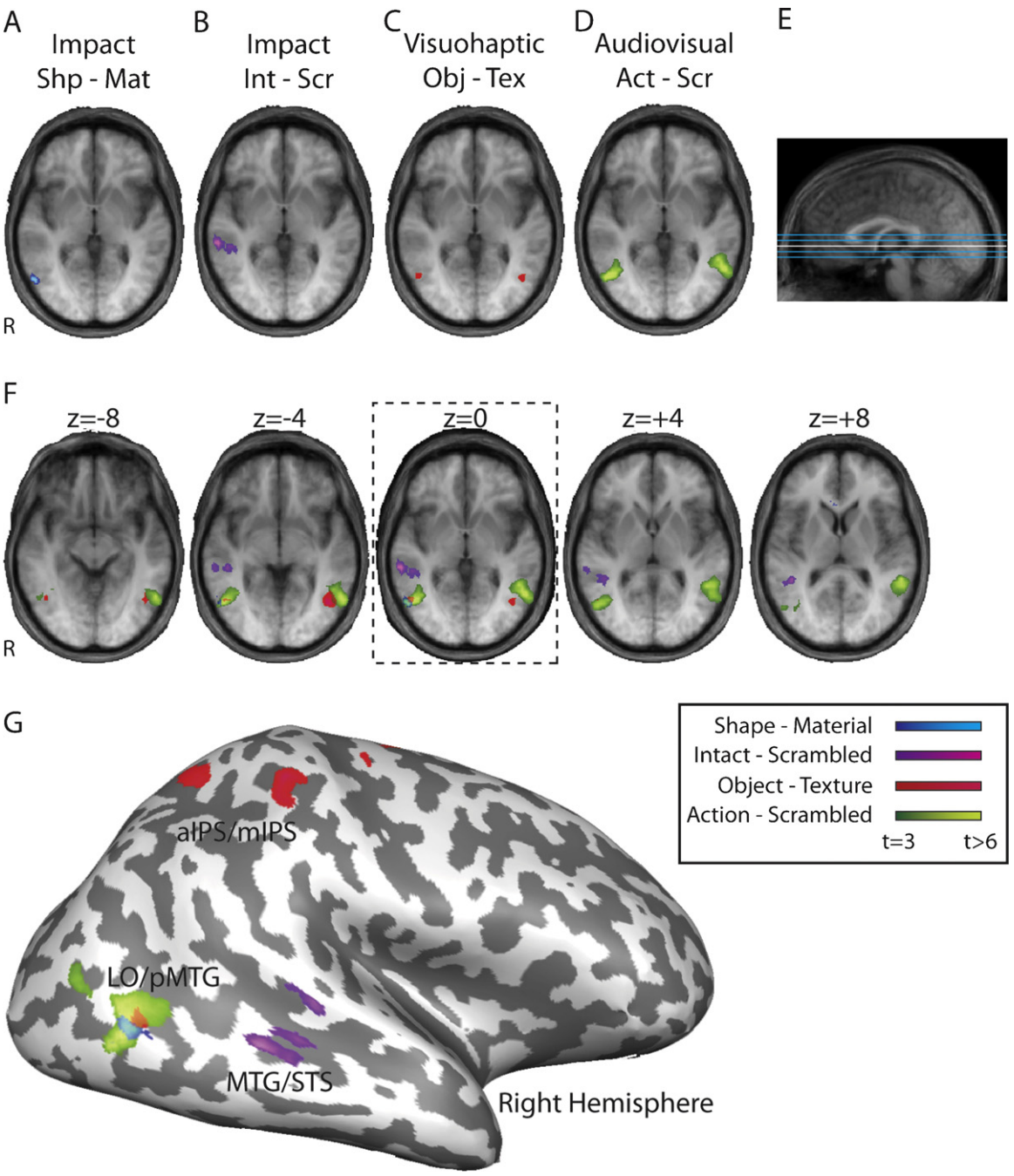

Fig. 5. Clusters from whole-brain contrasts. The heights of the axial slices are shown on a mid-sagittal image (e). The white line indicates the coordinate $z = 0$, which is the height of the four images in panels (a–d) and the image in panel (f) enclosed in the box. The other four images in panel (f) are shown at 4 mm intervals above and below the $z = 0$ slice. Each of the four image in (a–d) depicts a different contrast of interest, which is described in the label above each image and by the color look-up-table in the legend. The five images in (f) show all four contrasts of interest superimposed to assess their overlap. The five images represent five slice heights, which are indicated by the *z*-coordinates above each image. The image in (g) is a 3-D rendering of the inflated cortical surface of the right hemisphere of a representative subject. It shows the four contrasts of interest superimposed with the same four look-up-tables shown in the legend. aIPS/mIPS = anterior/middle intraparietal sulcus; LO/pMTG = lateral occipital cortex/posterior middle temporal gyrus; MTG/STS = middle temporal gyrus/superior temporal sulcus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Discussion

It is well established that BOLD activation in area LO is shape selective with visual and haptic sensory inputs (James et al., 2003; James et al., 2002; James et al., 2007; Stilla & Sathian, 2008; Tal & Amedi, 2009). Area LO/pMTG is also object selective with auditory inputs (Amedi et al., 2007; Beauchamp et al., 2004; James et al., 2011; Lewis et al., 2004). Although this suggests that area LO may be the site of a multisensory shape operator, auditory shape selectivity had not been explicitly tested in area LO until now. Here, we showed that area LO was more strongly recruited when subjects discriminated impact sounds based on the shape of the object that made them, compared to when subjects discriminated those same sounds based on their material. Thus, the previous findings combined with the current findings suggest that activation in area LO is shape selective across the three sensory input modalities that carry useable shape information about objects. The results are consistent with an emerging theory of perceptual functional specialization of

Table 1
Stereotactic coordinates for regions of interest.

Contrast	Brain region label	Coordinates	BA
Impact sounds			
Shape – material			
	Middle temporal gyrus	51, −62, 0	19
	Anterior cingulate	1, 31, 13	24
Intact – scrambled			
	Culmen (cerebellum)	28, −27, −27	
	Superior temporal sulcus	48, −36, 1	21
	Precentral gyrus	−39, −2, 29	6
	Precuneus	−14, −64, 30	31
	Posterior cingulate	−27, −43, 30	31
	Angular gyrus	−32, −52, 35	39
	Precuneus	−14, −62, 40	7
	Medial frontal gyrus[a]	−6, 47, 25	9
Audiovisual			
Action – scrambled			
	Middle temporal gyrus	−50, −52, 1	21
	Middle temporal gyrus	−42, −61, 3	37
	Precuneus	22, −70, 29	31
	Inferior parietal lobule	−53, −34, 30	40
	Precuneus	−8, −65, 36	7
	Precuneus	−13, −67, 37	7
Visuohaptic			
Object – texture			
	Middle occipital gyrus	45, −59, −3	19
	Middle occipital gyrus	−42, −59, −4	19
	Postcentral gyrus	−42, −29, 44	40
	Paracentral lobule	−2, −10, 47	31
	Inferior parietal lobule	33, −32, 48	40
	Precuneus	22, −49, 50	7

BA: Brodmann area.
[a] Activation in the opposite direction (negative) of the specified contrast.

the brain that is task-based rather than sensory modality-based (James et al., 2011; Lacey et al., 2009; Pascual-Leone & Hamilton, 2001).

Sounds made by environmental events, such as dropping an object from a height, can provide a wealth of information about the source of the sound (in this case the object), including its shape, size, length, and material (Gaver, 1993a). This ability of sounds to provide such information is evidenced by the accuracy shown by subjects on the shape and material tasks, despite the fact that they were listening to the sounds in a noisy environment. Previously, we argued that processing in area LO may be driven by coherent perception of environmental events (James et al., 2011). The current findings suggest that the role of area LO may be more specialized than event perception. A more specific hypothesis is that area LO is recruited for event perception when understanding the event relies on shape information about the objects in the event. Other regions may be recruited for processing the other multisensory characteristics of objects that are also important for understanding environmental events. For instance, in addition to a multisensory shape operator, there may also be a multisensory texture or roughness operator. Because shape information plays such a large role in visual object understanding, it is logical that the convergence zone for shape lies in what has traditionally been considered visual cortex. Further research is needed to discover the other nodes in the multisensory neural network responsible for event perception.

Perhaps contrary to the metamodal brain hypothesis, the results in Fig. 5 show no evidence of a "material" operator. However, when the map in Fig. 5a was reproduced with a more liberal threshold ($p < .05$, uncorrected), distinct clusters appeared in the right lingual gyrus (+4, −70, 0) and bilateral anterior insula/claustrum (±32, 15, 12). The anterior insula has been implicated in a variety of perceptual tasks, and may be recruited when a task is especially effortful (Ho, Brown, & Serences, 2009). Material judgments were more difficult than shape judgments, which may explain the insula activation. The lingual gyrus cluster, on the other hand, is close to regions reported in previous studies of auditory, tactile, and visual texture perception (Cant & Goodale, 2007; Stilla & Sathian, 2008; Tal & Amedi, 2009). It is not clear from this combination of studies, however, whether or not these brain regions are merely close to each other or overlapping. If they are overlapping, then the ventromedial occipitotemporal cortex may be a candidate as the site of a multisensory texture or material operator. More studies that consistently vary the texture or material information of objects (in addition to other types of information) and that test those manipulations across multiple sensory systems are needed if we are to further explore the utility of the metamodal brain hypothesis as a framework for understanding cortical specialization.

Finding that area LO was recruited for visual, haptic, and auditory shape processing is consistent with a "metamodal" view of cortical organization (Pascual-Leone & Hamilton, 2001). The metamodal view is an alternative to the long-standing view that the cortex is organized as multiple parallel sensory systems that eventually converge onto multisensory cortical areas. There are two main tenets to the metamodal brain hypothesis. First, the metamodal view suggests that multisensory processing is not restricted to special multisensory regions of cortex. Instead, much of the cortex, including putative primary sensory areas, is multisensory and is organized based on "operators". Operators are specialized for performing specific calculations or cognitive operations, rather than for processing specific sensory inputs. The fact that much of the cortex originally appeared to be unisensory can be explained if it is assumed that most operators have a preferred modality of sensory input. In the case of area LO, it is activated more strongly with visual input than with haptic, and activated more strongly with haptic input than with auditory. This led researchers in the earliest reports to consider area LO a visual region (Malach et al., 1995), and in later reports to consider it a bi-modal visuohaptic region (Amedi et al., 2002). We suggest that area LO is the site of a multisensory operator that processes shape information regardless of sensory input modality (Amedi et al., 2007). The second tenet of the metamodal view is that even operators that do not appear multisensory have the latent capacity for multisensory processing. This aspect of the metamodal view was not tested in this experiment, but could form the impetus for future studies on the functional organization of the brain through early and late development.

As the site of a multisensory operator for shape, area LO would represent a highly specialized perceptual processing unit that would require very specific inputs to successfully complete its operations. Based on the current findings, it is likely that area LO receives inputs from at least three different sensory modalities. It is unlikely that these inputs come directly from the primary sensory cortices. If the calculations or operations that area LO performs are being performed similarly across sensory modalities, then the input from those separate modalities must undergo considerable sensory input-specific transformation before reaching the shape operator. Some of the intermediate stages of processing between primary sensory representations and shape representations have been described for the visual system (for example, see Wilkinson et al., 2000), but they are much less understood for the haptic and auditory systems. For haptic inputs, it is possible that the secondary somatosensory cortex in the posterior insula/parietal operculum may be involved at an intermediate stage of processing (Stilla & Sathian, 2008). For auditory inputs, it is possible that a specific sub-region of the posterior superior temporal sulcus plays an intermediate role (Beauchamp et al., 2004; Doehrmann, Naumer, Volz, Kaiser, & Altmann, 2008; James et al., 2011; Lewis et al., 2005; Lewis et al., 2004; Stevenson & James, 2009). Another aspect of the highly specialized role of the shape operator is that it would adapt to the distribution of inputs that it receives. If shape processing is required more frequently with visual inputs than haptic, then the operator would develop a greater representation for vision than haptics.

Likewise, if shape processing is required more frequently for combinations of visual and haptic inputs than for combinations of visual and auditory inputs, then the operator may develop a greater capacity to integrate visual and haptic signals than visual and auditory signals.

Previous work has reported a dissociation between the neural substrates that are recruited for recognition of vocalizations as compared to tool sounds (Doehrmann et al., 2008; Lewis et al., 2005). These studies found that tool sounds activated area pMTG more than vocalizations, whereas vocalizations activated the middle to anterior superior temporal gyrus and sulcus more than tools sounds. The location of the tool-selective activation in these studies is overlapping with the action-selective activation in area LO shown in the current study, which was also assessed using sounds made by manual tools. The action/tool-selective area LO/pMTG activations from the previous and current studies overlapped with the shape-selective activation shown in the current study with impact sounds. The overlap between auditory action/tool-selectivity and auditory shape-selectivity suggests that auditory action/tool-selectivity may be a byproduct of shape selectivity. More specifically, the dissociation between the neural substrates for tool sounds and vocalizations may be based on the processing of acoustic shape information. Although there is evidence that vocal sound characteristics are influenced by the shape and size of the vocal apparatus (von Kriegstein, Smith, Patterson, Ives, & Griffiths, 2007), tool sounds may contain more cues to shape than vocalizations. Also, subjects may need to rely more on acoustic shape information when recognizing tools from sound than when recognizing vocalizations. One or both of these factors may lead to the dissociation in the neural substrates underlying auditory recognition of tools and vocalizations.

Based on previous studies of visuohaptic shape processing that found bilateral activation in the LOtv (Amedi et al., 2002; Amedi et al., 2001; Sathian & Zangaladze, 2002; Stilla & Sathian, 2008), it was expected that if auditory shape selectivity was found, it would be found bilaterally. However, auditory shape-selective activation with impact sounds was found only in the right hemisphere. Even at much more liberal statistical thresholds, no differences were found between shape and material judgments in left area LO—the lack of an effect in the left hemisphere was not imposed by overly conservative statistical thresholds. The result raises the possibility that auditory shape processing is lateralized to the right hemisphere. However, a second alternative possibility is that the pattern of individual differences in the location of activation was more diffuse in the left hemisphere than the right hemisphere. An example of this was described in two previous reports examining activation in STS with either speech sounds or other environmental sounds (Stevenson, Altieri, Kim, Pisoni, & James, 2010; Stevenson & James, 2009). The authors of those reports hypothesized that the variable location of the clusters in the left hemisphere led to less overlap across individuals, which led to a lack of an effect in the group-average contrast. A similar effect may have occurred in the present study, producing right-hemisphere activation with no corresponding left-hemisphere activation in the group-average analysis. Although the design of the impact sounds experiment did not allow for reliable single-subject analysis, we nevertheless perform an examination of the individuals using relatively liberal statistical criteria. Of the subjects that showed shape-selective activation in area LO, half showed bilateral activation, while the other half showed right-hemisphere activation only. This suggests that lateralization of the shape selective cluster was not just a statistical artifact, however, it also shows that lateralization is not consistent across subjects.

One consideration that must be addressed whenever activation is found in putative visual areas with non-visual stimuli is whether or not the activation is due to visual mental imagery. The results of previous reports of visuohaptic processing in LO converge to rule out the possibility that activation in area LO with haptic stimuli is due *only* to visual imagery (James, James, Humphrey, & Goodale, 2005; Lacey et al., 2009; Stilla & Sathian, 2008). In other words, it is not possible to explain all of the previous results by suggesting that visual imagery is the only mechanism by which area LO is activated with haptic stimulation. The results of those previous studies, however, do not rule out the possibility that visual imagery is *involved* in the activation of area LO. In fact, one theory of multisensory activation in area LO suggests that it receives both bottom-up (sensory) and top-down (imagery) inputs and that the weighting of these inputs changes depending on the task (Lacey et al., 2009). This view is consistent with the metamodal brain hypothesis – the functional organization of the brain is based on cognitive operations, not on sensory modalities. Operators receive bottom-up inputs from multiple sensory modalities and also receive top-down inputs. Whether or not those top-down inputs include imagery signals and whether or not those imagery signals are unisensory, multisensory, or amodal is a question for future research. Regardless, the distinguishing feature of an operator is that if the input signals contain the appropriate information (e.g., shape), the operator will process it, regardless the sensory modality or even whether they are bottom-up or top-down.

In conclusion, the current results show evidence of auditory shape-selectivity in area LO, suggesting that area LO is recruited for shape processing regardless of the modality of sensory input. The results suggest that previous reports of auditory object-selective activation in posterior aspect of area MTG and the anterior aspect of area LO may constitute the same underlying shape-selective process. The results converge with previous views (Amedi et al., 2007; James et al., 2011; Lacey et al., 2009; Pascual-Leone & Hamilton, 2001) suggesting that LO (and specifically LOtv) is the site of a metamodal shape operator. This operator may be one of several in a multisensory network involved in the coherent perception of environmental events.

Acknowledgments

This research was supported by NIH grant DC00012, the IUB Faculty Research Support Program, and by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc. We thank Thea Atwood and Becky Ward for their assistance with data collection and Christine White for her assistance with data analysis.

References

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex*, *12*, 1202–1212.

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, *4*(3), 324–330.

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, *10*(6), 687–689.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809–823.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Cant, J. S., & Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex*, *17*(3), 713–731.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: Functional anatomy by positron emission tomography. *The Journal of Neuroscience*, *11*(8), 2383–2402.

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*(11), 2776–2786.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance Medicine*, *33*(5), 636–647.

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *NeuroReport*, *16*(5), 419–423.

Frackowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R., Price, C. J., Zeki, S., et al. (2004). *Human brain function* (Second ed.). Amsterdam: Elsevier Academic Press.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *Journal of the Acoustical Society of America*, *87*(1), 311–322.

Gaver, W. W. (1993a). How do we hear in the world?: Explorations in ecological acoustics. *Ecological Psychology*, *5*(4), 285–313.

Gaver, W. W. (1993b). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, *5*(1), 1–29.

Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, *27*(5), 392–401.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception and Psychophysics*, *67*(2), 274–284.

Ho, T. C., Brown, S., & Serences, J. T. (2009). Domain general mechanisms of perceptual decision making in human cortex. *Journal of Neuroscience*, *29*(27), 8675–8687.

James, T. W., Culham, J. C., Humphrey, G. K., Milner, A. D., & Goodale, M. A. (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: An fMRI study. *Brain*, *126*, 2463–2475.

James, T. W., & Gauthier, I. (2006). Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Human Brain Mapping*, *27*, 37–46.

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., & Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, *40*, 1706–1714.

James, T. W., James, K. H., Humphrey, G. K., & Goodale, M. A. (2005). Do visual and tactile object representations share the same neural substrate? In M. A. Heller, & S. Ballesteros (Eds.), *Touch and blindness: Psychology and neuroscience*. Mahwah, NJ: Lawrence Erlbaum.

James, T. W., Kim, S., & Fisher, J. S. (2007). The neural basis of haptic object processing. *Canadian Journal of Experimental Psychology*, *61*(3), 219–229.

James, T. W., VanDerKlok, R. M., Stevenson, R. A., & James, K. H. (2011). Multisensory perception of action in posterior temporal and parietal cortices. *Neuropsychologia*, *49*(1), 108–114.

Kim, S., & James, T. W. (2010). Enhanced effectiveness in visuo-haptic object-selective brain regions with increasing stimulus salience. *Human Brain Mapping*, *31*, 678–693.

Kolb, B., & Whishaw, I. Q. (2003). *Fundamentals of human neuropsychology* (5th ed.). New York: W.H. Freeman and Company.

Lacey, S., Tal, N., Amedi, A., & Sathian, K. (2009). A putative model of multisensory object representation. *Brain Topography*, *21*(3–4), 269–274.

Lazar, N. A. (2010). *The statistical analysis of functional MRI data*. New York: Springer.

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*(21), 5148–5158.

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*(9), 1008–1021.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(18), 8135–8139.

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research*, *134*, 427–445.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Sathian, K., & Zangaladze, A. (2002). Feeling with the mind's eye: Contribution of visual cortex to tactile perception. *Behavioural Brain Research*, *135*(1–2), 127–132.

Stevenson, R. A., Altieri, N. A., Kim, S., Pisoni, D. B., & James, T. W. (2010). Neural processing of asynchronous audiovisual speech perception. *NeuroImage*, *49*, 3308.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–1223.

Stilla, R., & Sathian, K. (2008). Selective visuo-haptic processing of shape and texture. *Human Brain Mapping*, *29*(10), 1123–1138.

Tal, N., & Amedi, A. (2009). Multisensory visual-tactile object related network in humans: Insights gained using a novel crossmodal adaptation approach. *Experimental Brain Research*, *198*(2–3), 165–182.

Thirion, B., Pinel, P., MÈriaux, S., Roche, A., Dehaene, S., & Poline, J.-B. (2007). Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *NeuroImage*, *35*(1), 105–120.

Van Essen, D. C., Casagrande, V. A., Guillery, R. W., & Sherman, S. M. (2005). Corticocortical and thalamocortical information flow in the primate visual system. *Progress in Brain Research*, (149), 173–185 [Elsevier]

von Kriegstein, K., Smith, D. R., Patterson, R. D., Ives, D. T., & Griffiths, T. D. (2007). Neural representation of auditory size in the human voice and in sounds from other resonant sources. *Current Biology*, *17*(13), 1123–1128.

Warren, W. H., Jr., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 704–712.

Wilkinson, F., James, T. W., Wilson, H. R., Gati, J. S., Menon, R. S., & Goodale, M. A. (2000). An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Current Biology*, *10*(22), 1455–1458.