

## Multisensory perception of action in posterior temporal and parietal cortices

Thomas W. James<sup>a,b,c,\*</sup>, Ross M. VanDerKlok<sup>a</sup>, Ryan A. Stevenson<sup>a,b,d</sup>, Karin Harman James<sup>a,b,c</sup>

<sup>a</sup> Department of Psychological and Brain Sciences, Indiana University, United States

<sup>b</sup> Program in Neuroscience, Indiana University, United States

<sup>c</sup> Cognitive Science Program, Indiana University, United States

<sup>d</sup> Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, United States

### ARTICLE INFO

#### Article history:

Received 21 June 2010

Received in revised form

29 September 2010

Accepted 22 October 2010

Available online 29 October 2010

#### Keywords:

fMRI

Object recognition

Event perception

Auditory

Visual

### ABSTRACT

Environmental events produce many sensory cues for identifying the action that evoked the event, the agent that performed the action, and the object targeted by the action. The cues for identifying environmental events are usually distributed across multiple sensory systems. Thus, to understand how environmental events are recognized requires an understanding of the fundamental cognitive and neural processes involved in multisensory object and action recognition. Here, we investigated the neural substrates involved in auditory and visual recognition of object-directed actions. Consistent with previous work on visual recognition of isolated objects, visual recognition of actions, and recognition of environmental sounds, we found evidence for multisensory audiovisual event-selective activation bilaterally at the junction of the posterior middle temporal gyrus and the lateral occipital cortex, the left superior temporal sulcus, and bilaterally in the intraparietal sulcus. The results suggest that recognition of events through convergence of visual and auditory cues is accomplished through a network of brain regions that was previously implicated only in visual recognition of action.

© 2010 Elsevier Ltd. All rights reserved.

Recognizing events in a real environment is inherently multisensory (De Gelder & Bertelson, 2003; Gaver, 1993). Environmental events unfold over time and involve actions – either self-generated transitive movements of objects or object-generated movements (such as a human walking). In both cases, recognizing the object involved in the event is an important step toward understanding the event. Environmental events produce many sensory cues for identifying the objects and the actions involved in those events. The cues for recognizing objects and actions are usually distributed across multiple sensory systems. Thus, to understand how environmental events are recognized requires an understanding of the fundamental cognitive and neural processes involved in multisensory object and action recognition. Here, we investigated the neural substrates involved in audiovisual recognition of object-directed actions.

Although objects can be recognized without visual cues, a majority of work on the neural substrates of object recognition has been done using unisensory, visual presentation of familiar objects. A group of regions in the human brain that are selectively involved in object recognition are collectively known as the lateral occipital complex (LOC), which is a large area of cortex in the lateral and ventral occipito-temporal region (Grill-Spector, Kourtzi, &

Kanwisher, 2001; James, Culham, Humphrey, Milner, & Goodale, 2003; Malach et al., 1995). Activation in LOC is often defined as object-selective, that is, it is activated more with intact pictures of objects than with other classes of visual stimuli (Grill-Spector et al., 2001; Malach et al., 1995). Damage to the LOC causes impairments in object recognition, resulting in visual agnosia (James et al., 2003).

A typical fMRI study of visual object recognition uses static pictures of isolated objects as stimuli. This type of stimulus provides ample information for object recognition, but is impoverished with respect to the information needed for event recognition. With studies of visual action recognition, the tasks are focused on the event instead of the object; therefore, studies of action must use stimuli that unfold over time (dynamic stimuli). Studies of action recognition often use stimuli involving moving human bodies, hands, or faces and sometimes use stimuli involving human bodies or hands manipulating other objects. fMRI studies investigating the neural substrates of visual action recognition consistently find a network of brain regions that includes Broca's area (inferior frontal gyrus), several regions in the parietal lobe, including the intraparietal sulcus, the posterior middle temporal gyrus (pMTG), and superior temporal sulcus (pSTS, Caspers, Zilles, Laird, & Eickhoff, 2010). In the posterior temporal lobe, the pSTS is more selective for human actions (Beauchamp & Martin, 2007; Grossman & Blake, 2002; Puce & Perrett, 2003), whereas the pMTG is more selective for actions performed on other objects (Beauchamp & Martin, 2007; Valyear & Culham, 2010). The involvement of pMTG in action recognition, and especially recognition of actions involving non-human objects,

\* Corresponding author at: 1101 E Tenth St, Bloomington, IN 47405, United States.  
Tel.: +1 812 856 0841; fax: +1 812 855 4691.

E-mail address: [thwjames@indiana.edu](mailto:thwjames@indiana.edu) (T.W. James).

is of particular interest, because pMTG borders the LOC, which is involved in recognition of static isolated objects.

Environmental stimuli used in the investigation of auditory recognition always represent events. A majority of studies investigating the recognition of sounds use speech stimuli, however, there are several studies that investigate the recognition of environmental sounds more generally. When environmental sounds are contrasted with control stimuli such as white noise or scrambled nonsense sounds, neural activation is found in the middle and posterior aspects of the superior temporal gyrus (pSTG), the pSTS, and the pMTG (Amedi, Jacobson, Hendler, Malach, & Zohary, 2002; Beauchamp, Lee, Argall, & Martin, 2004; Doehrmann, Naumer, Volz, Kaiser, & Altmann, 2008; Lewis et al., 2004; Stevenson, Geoghegan, & James, 2007). However, when different categories of sounds are contrasted, some regions of the temporal lobe are found to be more selective for some sounds than others (Doehrmann et al., 2008; Lewis, Brefczynski, Phinney, Janik, & DeYoe, 2005). Sounds of animal vocalizations, including human speech, selectively activated the anterior and middle aspects of the STG and STS. Sounds made by machines and tools selectively activated the posterior aspect of the inferior temporal gyrus (pITG), pMTG, pSTS, and pSTG. Both studies showed a left-sided bias for tool-selective activation (Doehrmann et al., 2008; Lewis et al., 2005).

The study of the neural substrates of isolated object recognition has benefitted from testing for “object-selectivity” by contrasting intact images of objects with scrambled nonsense images (Grill-Spector et al., 2001; James et al., 2003; Malach et al., 1995). Scrambled images have many low-level visual properties in common with intact images, but are not recognizable. Investigations of recognition processes that are shared between the visual and somatosensory systems have used the “selectivity” method almost exclusively to define bi-modal visuo-haptic object-selective brain regions. Specifically, bi-modal regions are consistently found in the intraparietal sulcus and the LOC (Amedi, Malach, Hendler, Peled, & Zohary, 2001; James et al., 2002; Stilla & Sathian, 2008). The bi-modal intraparietal area is found on the anterior and middle aspects of the sulcus and the bi-modal lateral occipital area, called LOTv for tactile-visual, is found on the middle occipital gyrus (Amedi et al., 2002).

Unlike the study of tactile-visual or visuo-haptic convergence, the study of audiovisual convergence has largely eschewed the use of selectivity. Audiovisual convergence is usually assessed by measuring the enhancement of activation with a multisensory stimulus over and above that of unisensory stimuli from one or more sensory modalities (Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009). Both the selectivity and the enhancement methods of assessing multisensory convergence have their specific benefits and problems (Kim & James, 2010; Stevenson, Kim, & James, 2009). Thus, the current study sought to expand the use of the selectivity method with audiovisual stimuli. Because the focus of the research was on event perception, selective responses will be called “event-selective.”

The literature reviewed above suggests that the pMTG and pSTS are involved in recognition of environmental events through both visual and auditory sensory inputs. The pMTG is of particular interest with respect to investigations of object-directed actions, because it borders LOC, which is specifically involved in processing isolated objects. The findings reviewed above suggest that pMTG may be involved in visual and auditory processing of events generated by manual manipulation of tool-like objects in the environment. In other words, the pMTG/LOC junction may be a convergence zone for audiovisual recognition of object-directed actions.

To test this hypothesis, we presented subjects with video and audio of environmental events generated by manual manipulation of tool-like objects while they underwent functional MRI. To test for

event selectivity, intact and scrambled versions of video and audio sequences were contrasted. Bi-modal event selectivity was found bilaterally in the pMTG and in the left pSTS. Bi-modal selectivity was also found bilaterally in the posterior intraparietal sulcus and in the left insula. The results suggest that recognition of events through convergence of visual and auditory cues is accomplished through a network of brain regions that was previously implicated only in visual recognition of object-directed actions.

## 1. Methods and materials

### 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. Subjects were compensated for their time.

### 1.2. Stimuli

Experimental stimuli consisted of audio and video recordings of manual actions involving a moveable implement (e.g., hammer, paper cutter, paper towel dispenser, etc.). Hands were visible in the recordings. Recordings were made with a DCR-HC85 MiniDV Digital Handycam camcorder. Separate video and audio files were extracted from the raw recordings, such that they could be presented separately as visual and auditory stimuli. For audiovisual stimuli, the visual and auditory stimuli that were taken from the same raw recording were presented together. Video was acquired at the camera's native resolution of 1024 × 720. Audio was acquired with 16 bit at a sampling rate of 32 kHz with the camcorder's onboard microphone. Visual stimuli were cropped to square, down-sampled to a resolution of 200 × 200 pixels, and converted from color to greyscale. Audio was converted from stereo to mono. Pilot testing showed that these intact visual and auditory stimuli were very recognizable. Examples of two event stimuli are shown in Fig. 1A.

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. Scrambling the video prevented recognition of the objects and the actions performed with the objects. Using the intensity-matched exchange method preserved general changes in pixel intensities across frames, but rearranged the spatial locations of those changes. One result of this was a subjective perception of motion in the scrambled video. The motion percept in the scrambled videos, however, was not coherent like the motion percept in the intact videos. Although the strength of the motion percept was not measured in the scrambled videos, it was clear that a direction of motion was impossible to judge from the scrambled video. Examples frames of two scrambled videos are shown in Fig. 1B.

Audio sequences were also scrambled. 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 like noise. Using the amplitude-matched exchange method preserved the unfolding of general changes in amplitude across time. Examples of two scrambled waveforms are shown in Fig. 1B.

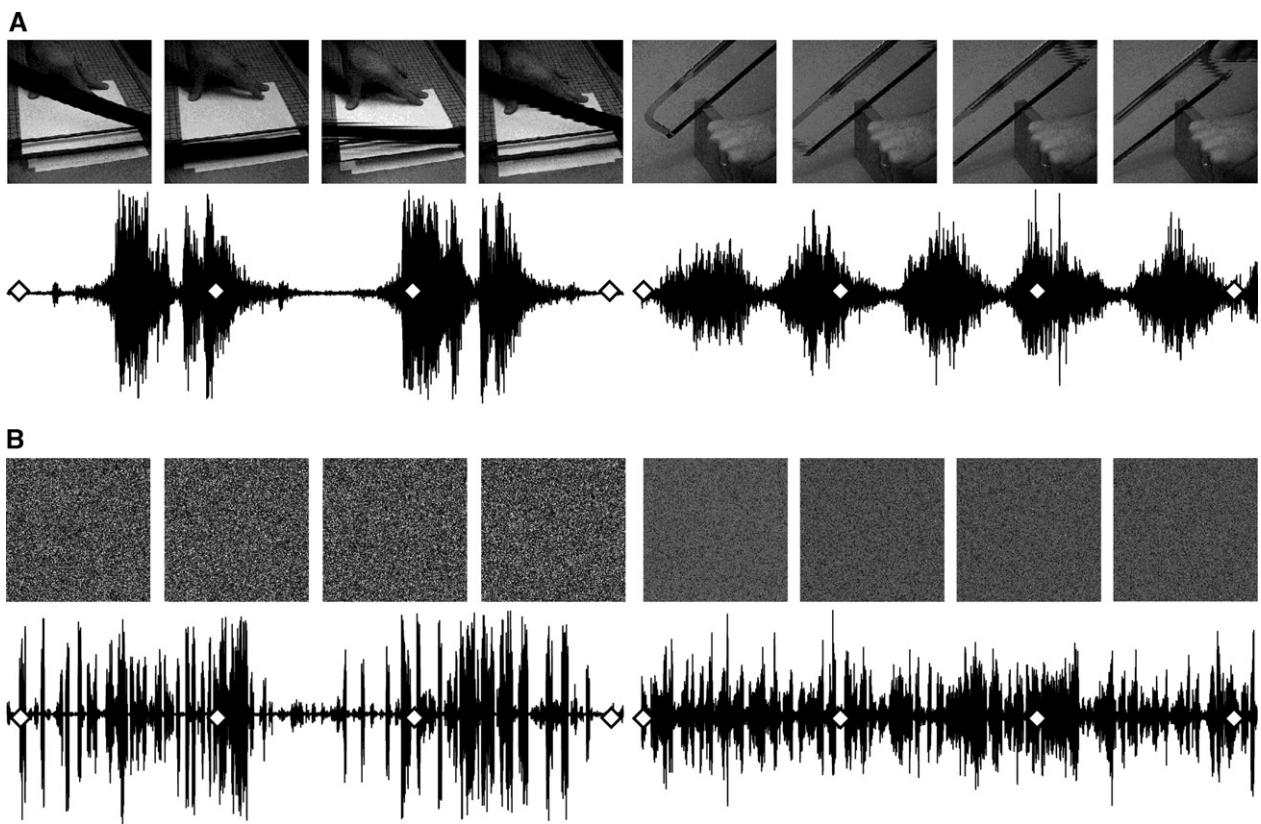
### 1.3. Procedures

Subjects lay supine in the bore of the MRI with their head in the head coil and a response pad placed on their right thigh. Intact and scrambled audio and video stimuli were presented 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 at 30 frames per second via 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 listened to through the pneumatic headphones. Foam was placed around the headphones inside the headcoil to reduce subject head movement.

BOLD fMRI measurements were collected in four runs, each 3 min long, two runs with visual stimuli and two with auditory stimuli. Stimuli were presented in a blocked design with 16-s stimulus blocks consisting of eight 2-s presentations of either intact or scrambled stimuli, interleaved by 12-s rest blocks during which the subjects fixated a central dot. Each run contained three intact and three scrambled blocks. Across the four runs, this resulted in six blocks of data for each of the four stimulus types.

During stimulus blocks, subjects performed a one-back perceptual matching task to maintain attention on the stimuli. Subjects responded with the right index for a duplicate stimulus and with their middle finger for a different stimulus.

On a separate day, all subjects underwent a short imaging session to collect data to functionally localize brain regions involved in visuo-haptic convergence. An



**Fig. 1.** Examples of visual and auditory stimuli. Each panel shows two examples of environmental events used as stimuli. The top row in each panel shows four frames taken from the video stream. The bottom row in each panel shows the auditory waveform. White diamond symbols superimposed on the waveform show the time at which the four video frames were taken. Examples of intact stimuli are shown in (A). Examples of scrambled stimuli are shown in (B).

established localizer task was used that has been described elsewhere (Amedi et al., 2001; Kim & James, 2010). Briefly, subjects viewed and felt objects and textures. Textures were used as control stimuli for assessing bi-modal visuo-haptic object-selectivity.

#### 1.4. Imaging parameters and analysis

Imaging was carried out using a Siemens Magnetom TRIO 3-Tesla 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, which 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 3D 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 Brain Voyager™ 3D analysis tools, using linear trend removal, 3D spatial Gaussian filtering (FWHM 6 mm), slice scan-time correction, and 3D motion correction. Anatomical volumes were transformed into the common stereotactic space of Talairach and Tournoux using an 8-parameter affine transformation. Functional volumes were then coregistered to the anatomical volume, and transformed into Talairach space. Data were analyzed using a general linear model with predictors generated based on the timing of the blocked design protocol for placement of canonical hemodynamic response functions (Glover, 1999).

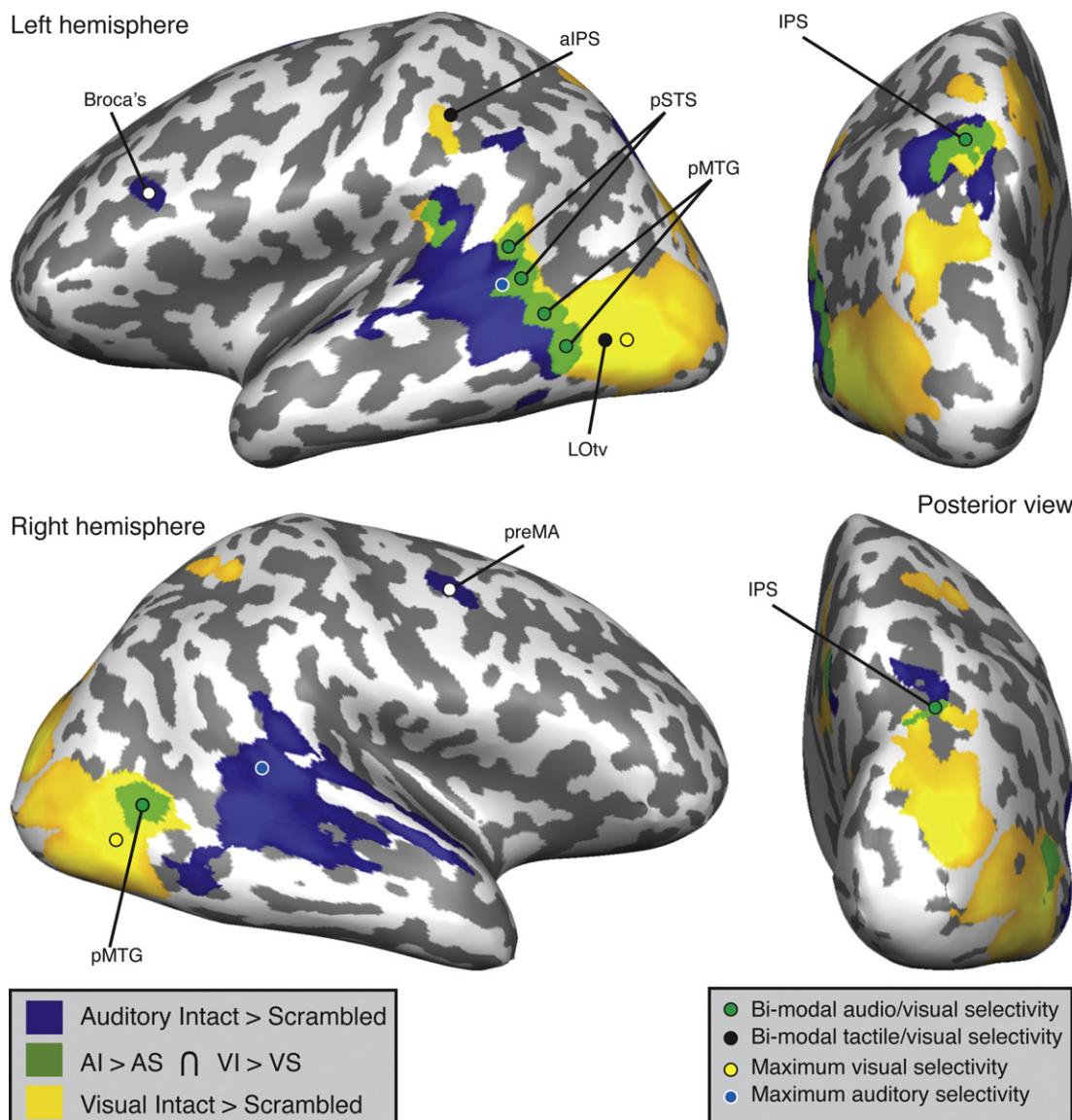
## 2. Results

To assess auditory and visual event selectivity, a whole-brain group-average analysis was performed using a random-effects general linear model with predictors representing audio and visual, intact and scrambled stimuli. These predictors were combined to perform two specific contrasts. The first contrast, audio intact > audio scrambled, identified auditory event-selective brain

regions, while the second contrast, visual intact > visual scrambled, identified visual event selective brain regions (Fig. 2). Correction for multiple tests was done for both contrasts using a False Discovery Rate (FDR) of  $q = 0.05$  combined with a cluster threshold of 15 voxels. The cluster-threshold technique controls false positives, with a relative sparing of statistical power (Forman et al., 1995; Thirion et al., 2007). Thus, the combination of FDR and cluster threshold produced a more conservative threshold than FDR alone.

It was expected that the unisensory contrasts would activate extensive areas of cortex involved in recognition of object, actions, and sounds. It was also expected that not all of these areas would be involved strictly in event perception. Consistent with previous work on recognition of environmental tool sounds (Doehrmann et al., 2008; Lewis et al., 2005), auditory event selectivity (blue) was found in the middle and posterior aspects of the STG, STS, and MTG. Auditory selectivity was also found in Broca's area and the pre-motor area. The locations of these areas are indicated with white dots in Fig. 2. The location of the maximum statistical value for the auditory contrast is indicated by the blue dot in Fig. 2.

Consistent with previous work on visual recognition of action (Beauchamp & Martin, 2007; Caspers et al., 2010; Grossman & Blake, 2002), visual event selectivity (yellow) was found in the pMTG and pSTS. Similar to previous work on visual recognition of isolated objects (Grill-Spector et al., 2001; James et al., 2003; Malach et al., 1995), visual event selectivity (yellow) was also found in areas of the middle and inferior occipital lobe, in the known location of the LOC. Visual selectivity was also found in the motion-selective region known as the human MT complex (hMT+). This region was likely recruited because the motion signals in the intact videos were more coherent than the motion signals



**Fig. 2.** Whole-brain map of audiovisual event selectivity. Group-average maps are shown on an inflated cortical representation of a single subject shown from the left, right, and posterior views. Maps represent contrasts of intact and scrambled conditions for visual (yellow) auditory (blue) stimuli. Green areas indicate the overlap (intersection) of the two maps. Abbreviations: posterior superior temporal sulcus (pSTS), posterior middle temporal gyrus (pMTG), intraparietal sulcus (IPS), anterior IPS (aIPS), lateral occipital tactile-visual area (LOTv), pre-motor area (preMA). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

in the scrambled videos. Finally, visual selectivity was found in the body part-selective region known as the extrastriate body-part area (EBA). This region was likely recruited because the hands performing the actions were recognizable in the intact videos, but not in the scrambled videos. The location of the maximum statistical value for the visual contrast is indicated by the yellow dot in Fig. 2.

Clusters of voxels that showed statistical significance with both contrasts (i.e., audio intact > audio scrambled  $\cap$  visual intact > visual scrambled), however, were labeled bi-modal event-selective brain regions. The locations of these regions are indicated by the green dots in Fig. 2. These overlapping regions were considered to be specifically involved in event perception. Because the bi-modal contrast uses a logical AND operation, voxels shown in the bi-modal map actually have a more conservative threshold than voxels shown in the two unisensory maps from which it was generated.

Multisensory audiovisual event selectivity (green) was found in regions along the occipito-temporal junction. In the left hemisphere, clusters corresponded specifically to the pSTS and pMTG.

In the right hemisphere, only pMTG was found. The lack of overlap in the right hemisphere was due to hemispheric differences in the activation pattern produced by the auditory stimuli. Auditory selectivity was not found on the posterior aspect of the right STS/STG, but was found on the posterior aspect of the right MTG. These differences across hemispheres in auditory activation in pSTS and pMTG with tool stimuli match well with previously reported patterns (Doehrmann et al., 2008; Lewis et al., 2005). Coordinates and Brodmann areas for all unimodal and bimodal regions of interest are shown in Table 1.

The visuo-haptic functional localizer data were analyzed using a random-effects general linear model and a similar bi-modal contrast as the audiovisual data (i.e., tactile object > tactile texture  $\cap$  visual object > visual texture). With an FDR threshold ( $q < .05$ ), two regions were found in the left hemisphere, the anterior aspect of the intraparietal sulcus (aIPS) and the “tactile-visual” part of the lateral occipital area (LOTv). The locations of these two regions are indicated by the black dots in Fig. 2.

**Table 1**  
Regions of interest.

Brain region	Coordinates	BA
Bi-modal event-selective activation		
Left		
pMTG	-59, -61, 2	37/21
pSTS	-64, -44, 8	22
Insula	-47, -34, 20	13
IPS	-21, -69, 40	7
Right		
pMTG	61, -59, 1	37/21
IPS	21, -62, 32	7
Unisensory activation		
Left		
LOtv/MOG	-55, -64, -4	19
STS/STG	-52, -40, 6	22
Broca	-44, 18, 12	45
Right		
LOtv/MOG	56, -64, -4	19
STS/STG	68, -40, 8	22
preMA	43, -4, 43	6

Notes: Coordinates are in Talairach space in the order X, Y, Z. BA = Brodmann's area  
pMTG = posterior middle temporal gyrus; pSTS = posterior superior temporal sulcus;  
IPS = intraparietal sulcus; LOtv = lateral occipital tactile-visual; MOG = middle occipital gyrus; STG = superior temporal gyrus; Broca = Broca's area; preMA = pre-motor area.

The unisensory selectivity maps in Fig. 2 were produced from a group-average analysis. The assumption is that overlap of these maps reflects consistent bi-modal overlap in every individual subject. An alternative, though, is that overlap in the group-average map simply reflects blurring of unisensory selective regions that vary in location across individuals. To ensure that the overlap shown in Fig. 2 was not spurious, an analysis of individual subjects was conducted. Whole-brain fixed-effects general linear models were fit to individuals' data instead of to the entire group. Individuals' data were transformed into the same standard space as the group analysis, such that coordinates could be compared across individuals and with the group analysis. The same contrasts performed on the group were performed on the individuals. The contrasts were thresholded using FDR ( $q < .05$ ) and a 15-voxel cluster threshold.

Fig. 3 shows the same axial slice from each individual with a map created from the bi-modal contrast (i.e., audio intact > audio scrambled  $\cap$  visual intact > visual scrambled). The slice coordinate (height on z-axis) was set to the center of the pMTG clusters found in the group analysis. White dots on the images show the in-plane coordinates of the center of the pMTG clusters found in the group analysis. In the left hemisphere, 10 of 12 subjects showed overlap between their own bi-modal cluster and the group cluster ( $p = .003$ ). The two subjects without overlap each showed a bi-modal region anterior to the group cluster. In the right hemisphere, 9 of 12 subjects showed overlap between their own bi-modal cluster and the group cluster ( $p = .02$ ). One of the three without overlap showed a bi-modal region medial to the group cluster. The other two without overlap did not have bi-modal clusters anywhere along the MTG/MOG.

### 3. Discussion

To our knowledge, this is the first fMRI study of auditory and visual recognition of events produced by object-directed actions. Consistent with our hypothesis, multisensory audiovisual event-selective activation was found in pMTG, at the junction of the MTG and LOC. Previous work has shown that the LOC is involved in recognition of isolated objects (Grill-Spector et al., 2001; James et al., 2003; Malach et al., 1995), that the pMTG is involved with recognition of visual actions (Beauchamp & Martin, 2007; Caspers et al., 2010; Grossman & Blake, 2002; Valyear & Culham, 2010), and that

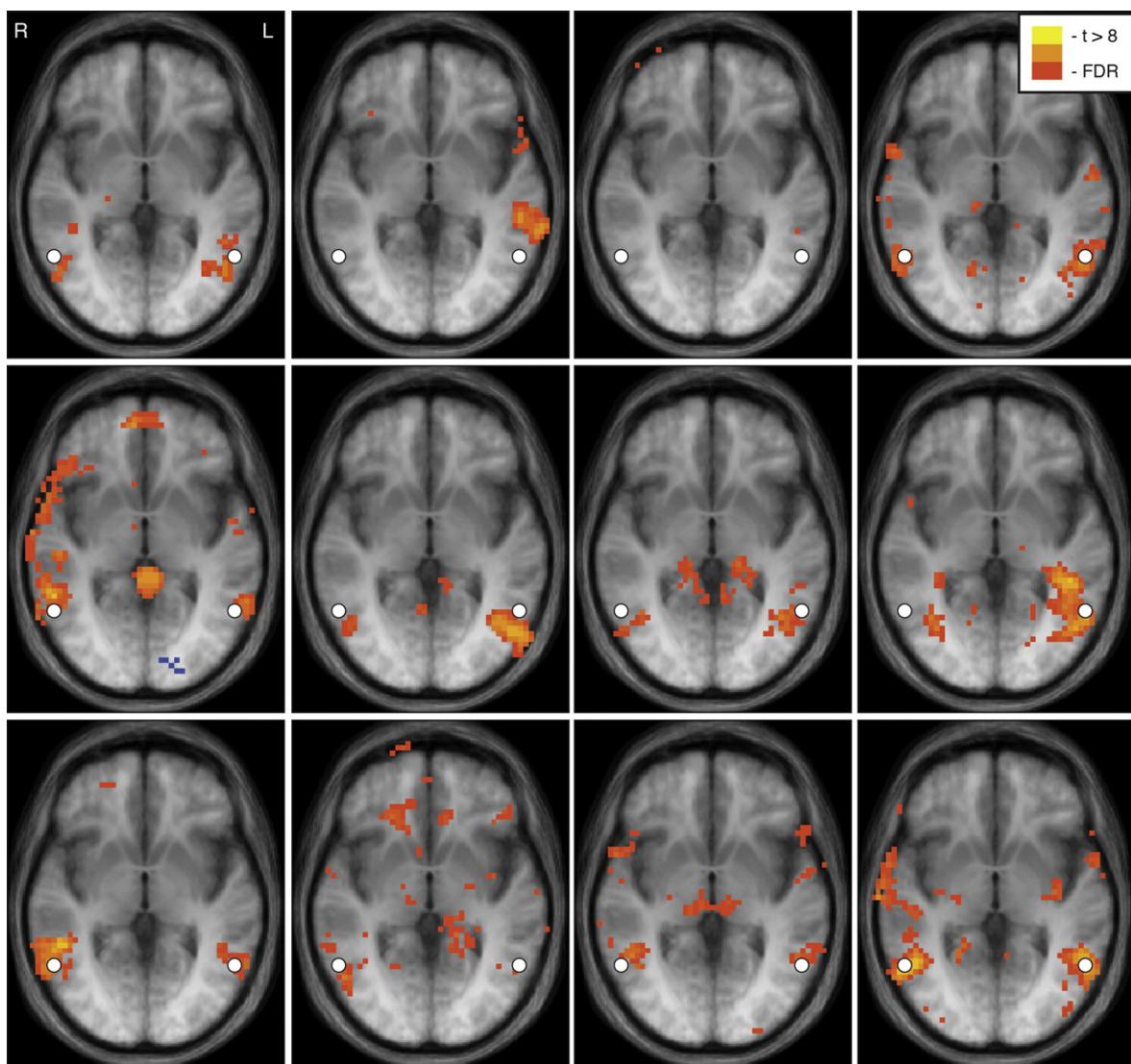
pMTG is also involved in recognition of environmental tool sounds (Doehrmann et al., 2008; Lewis et al., 2005). The current findings bring together these divergent fields of inquiry and suggest that the junction of pMTG represents a convergence zone for auditory and visual information, the purpose of which is to identify or categorize environmental events.

Bi-modal event-selectivity was also found in the pSTS and the posterior insula in the left hemisphere, and in the intraparietal sulcus, bilaterally. The pSTS has been implicated in both visual action perception (Beauchamp & Martin, 2007; Grossman & Blake, 2002; Puce & Perrett, 2003), and audiovisual integration (Beauchamp et al., 2004; Calvert, Campbell, & Brammer, 2000; Stevenson & James, 2009). Thus, it is not surprising that pSTS contributes to event recognition. It is worthwhile noting that pSTS was activated only on the left, whereas pMTG was activated bilaterally. This may suggest that pSTS represents a more specialized form of processing or a higher level of hierarchical processing than pMTG.

The left posterior insula activation was found on the lateral bank of the parietal operculum. This is near to, or possibly overlapping with, the secondary somatosensory cortex (SII). The role of SII is not clear, but it has been implicated in multisensory processing, at least for haptic and visual sensory modalities (Binkofski et al., 1999; Still & Sathian, 2008). Based on the current data and previous data, it is possible that SII is a site of tri-modal processing, but that determination will require a more systematic study of its response with stimuli from the three sensory systems.

There is good evidence that the anterior aspect of the intraparietal sulcus (IPS) integrates visual and haptic signals (Binkofski, Kunesch, Classen, Seitz, & Freund, 2001; Bodegard, Geyer, Grefkes, Zilles, & Roland, 2001; Bohlhalter, Fretz, & Weder, 2002; Culham & Valyear, 2006). Fig. 2 shows visually selective activation in the anterior IPS, but not audiovisual event-selective activation. The audiovisual site was much more posterior along the IPS, and even more posterior in the right hemisphere than in the left. Many areas of the posterior IPS have been identified that are involved in different aspects of visuomotor control (for review, see Culham & Valyear, 2006). It seems likely that at least one of these sites overlaps with the audiovisual site identified in Fig. 2. The tasks in the current experiment, however, did not involve visuomotor control; therefore, it is an open question why this area of the IPS is recruited for bi-modal perception of object-directed action. One hypothesis is that the areas of the IPS that are involved in visuomotor control and planning are also involved in the recognition of those actions (Culham & Valyear, 2006; Valyear & Culham, 2010). Motor actions are not only controlled by visual signals, but also by haptic and auditory signals. Synchronization of movements with sounds – such as with playing music, but also with simpler tasks such as finger tapping – demonstrates the use of auditory signals to control the timing of movements (for review, see Reppe, 2005). The influence of auditory signals on motor movements could be partially mediated by processes dedicated to integrating auditory and visual signals. The type of information that can be integrated with vision and touch is different from the type of information that can be integrated with vision and audition. Anterior IPS may be specialized for integrating visual and haptic information, whereas more posterior areas of IPS may be specialized for integrating visual and auditory information. These integration sites may play a role in motor control, and may also play a role in the recognition of actions.

Several previous studies have investigated the neural substrates of action recognition in the framework a 'mirror' system (for review, see Fabbri-Destro & Rizzolatti, 2008; Iacoboni, 2009). The mirror system is seen as a mechanism for the imitation of other's action through observation. The observation of action has been consistently shown to activate a network of brain regions including Broca's area, several regions of the parietal cortex, and the



**Fig. 3.** Individual subject analysis of audiovisual event selectivity in the pMTG. The 12 images represent one axial slice from each individual subject. The white dots indicate the coordinates of the bilateral pMTG clusters from the group analysis. The maps show the bi-modal contrast only (i.e., audio intact > audio scrambled  $\cap$  visual intact > visual scrambled).

pMTG/pSTS (Caspers et al., 2010). Our findings are consistent with the previous work on observation of action. A network of regions was found that included Broca's area, the pre-motor area, areas of the intraparietal sulcus, the pMTG, and the pSTS. Several areas of the network showed bi-modal event-selective activation. Our findings do not speak to the contribution of those areas to the imitation of actions. What they do indicate is the existence of processes that are involved in more than analyzing isolated sensory channels. These processes combine information across sensory channels, with the ultimate goal of understanding events in the environment.

#### Acknowledgments

This research was supported by NIH grant T32DC00012; the Faculty Research Support Program, administered through the IUB Office of the Vice President of Research; and by the Indiana META-Cyt 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.

#### 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.
- 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.
- Beauchamp, M. S., & Martin, A. (2007). Grounding object concepts in perception and action: Evidence from fMRI studies of tools. *Cortex*, 43(3), 461–468.
- Binkofski, F., Buccino, G., Posse, S., Seitz, R. J., Rizzolatti, G., & Freund, H. J. (1999). A fronto-parietal circuit for object manipulation in man: Evidence from an fMRI-study. *European Journal of Neuroscience*, 11, 3276–3286.
- Binkofski, F., Kunesch, E., Classen, J., Seitz, R. J., & Freund, H. J. (2001). Tactile apraxia: Unimodal apractic disorder of tactile object exploration associated with parietal lobe lesions. *Brain*, 124(1), 132–144.
- Bodegard, A., Geyer, S., Grefkes, C., Zilles, K., & Roland, P. E. (2001). Hierarchical processing of tactile shape in the human brain. *Neuron*, 31, 317–328.
- Bohilhalter, S., Fretz, C., & Weder, B. (2002). Hierarchical versus parallel processing in tactile object recognition: A behavioural-neuroanatomical study of aperceptive tactile agnosia. *Brain*, 125, 2537–2548.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- 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(11), 649–657.

- Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage*, 50(3), 1148–1167.
- Culham, J. C., & Valyear, K. F. (2006). Human parietal cortex in action. *Current Opinion in Neurobiology*, 16(2), 205–212.
- De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Science*, 7(10), 460–467.
- 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.
- Fabbri-Destro, M., & Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology (Bethesda)*, 23, 171–179.
- 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.
- Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1), 1–29.
- Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, 9(4), 416–429.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10–11), 1409–1422.
- Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, 35(6), 1167–1175.
- Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60, 653–670.
- 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., 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.
- 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.
- 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.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of Royal Society of London B: Biological Science*, 358(1431), 435–445.
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6), 969–992.
- Stein, B., Stanford, T., Ramachandran, R., Perrault, T., & Rowland, B. (2009). Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Experimental Brain Research*, 198(2), 113–126.
- Stevenson, R. A., Geoghegan, M. L., & James, T. W. (2007). Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Experimental Brain Research*, 179(1), 85–95.
- 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.
- Stevenson, R. A., Kim, S., & James, T. W. (2009). An additive-factors design to disambiguate neuronal and areal convergence: Measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Experimental Brain Research*, 198(2–3), 183–194.
- Still, R., & Sathian, K. (2008). Selective visuo-haptic processing of shape and texture. *Human Brain Mapping*, 29(10), 1123–1138.
- Thirion, B., Pinel, P., Mérioux, 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.
- Valyear, K. F., & Culham, J. C. (2010). Observing learned object-specific functional grasps preferentially activates the ventral stream. *Journal of Cognitive Neuroscience*, 22(5), 970–984.