Auditory Spatial and Object Processing in the Human Planum Temporale: No Evidence for Selectivity

Kevin R. Smith, I-Hui Hsieh, Kourosh Saberi, and Gregory Hickok

Abstract

■ Although it is generally acknowledged that at least two processing streams exist in the primate cortical auditory system, the function of the posterior dorsal stream is a topic of much debate. Recent studies have reported selective activation to auditory spatial change in portions of the human planum temporale (PT) relative to nonspatial stimuli such as pitch changes or complex acoustic patterns. However, previous work has suggested that the PT may be sensitive to another kind of nonspatial variable, namely, the number of auditory objects simultaneously presented in the acoustic signal. The goal of the present fMRI experiment was to assess whether any portion of the PT showed spatial selectivity relative to manipulations of the number of auditory objects presented. Spatially sensitive regions in the PT were defined by comparing activity associated with listening to an auditory object (speech from a single talker) that changed location with one that remained stationary. Activity within these regions was then examined during a nonspatial manipulation: increasing the number of objects (talkers) from one to three. The nonspatial manipulation modulated activity within the "spatial" PT regions. No region within the PT was found to be selective for spatial or object processing. We suggest that previously documented spatial sensitivity in the PT reflects auditory source separation using spatial cues rather than spatial processing per se. ■

INTRODUCTION

Evidence for the existence of two broad processing streams has mounted steadily (Hickok & Poeppel, 2000, 2004, 2007; Alain, Arnott, Hevenor, Graham, & Grady, 2001; Maeder et al., 2001; Wise et al., 2001; Romanski et al., 1999), but the nature of the posterior/dorsal stream, in particular, has been the topic of discussion (Middlebrooks, 2002; Zatorre, Bouffard, Ahad, & Belin, 2002). In human research, this discussion has centered around cortical areas in the planum temporale (PT), which comprises auditory-related fields hypothesized to be a major correlate of the posterior/dorsal stream. Some authors have proposed that this region supports auditory–motor integration (Hickok & Poeppel, 2000, 2004, 2007; Warren, Wise, & Warren, 2005; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Wise et al., 2001), whereas others have questioned whether a pure spatial processing mechanism exists within posterior auditory cortex (Zatorre et al., 2002). One study parametrically varied the number of spatial locations from which a noise stimulus was presented (in sequence) and failed to find any area where there were corresponding increases in activity (Zatorre et al., 2002). An effect of spatial location was found only when location served as a cue for auditory source disambiguation.

Subsequent studies have explicitly examined the relation between auditory location and auditory object perception. Warren and Griffiths (2003) examined the response to tone sequences that either varied in pitch or location, and distinct regions were noted with pitch-related activations more lateral and anterior than the location-related activations, which were in the PT. In a similar experiment, Barrett and Hall (2006) found that largely distinct regions of human auditory cortex responded to changes of spatial location (nonprimary auditory cortex in the PT) compared with changes in pitch (Heschl's gyrus and more anterior regions). Altmann, Bledowski, Wibral, and Kaiser (2007) found a similar effect with natural sounds that either changed in content (sheep sound → dog sound) or in location. These studies report some degree of overlap between spatial and nonspatial activations, however, each identified regions within the PT that were selective for spatial stimuli. On the basis of these findings, it has been suggested that there are distinct auditory object and spatial processing pathways, and that a dedicated spatial processing system exists within the human PT. For example, Warren and Griffiths write, "The present study … suggests anatomically distinct spatial (posteromedial) and object (anterolateral) processing mechanisms within PT…" (p. 5803).

However, as these studies did not manipulate the number of auditory objects in the acoustic signal (different objects were presented in sequence), it remains a possibility that these spatially sensitive regions are not spatially selective. Such a possibility has important implications for understanding the function of the auditory dorsal stream.

University of California, Irvine

For example, Zatorre et al. (2002) noted that the interaction of spatial and object information is important for disambiguating overlapping auditory sources (i.e., auditory scene segregation). If putative spatial regions of the human PT are jointly sensitive to spatial and object information, this could indicate that these cortical regions support not spatial processing per se, but sound source segregation based on spatial cues.

With these questions in mind, the present study sought to assess the selectivity of the spatial responses within the human PT. Rather than manipulating pitch or pattern change as the nonspatial manipulation (Altmann et al., 2007; Barrett & Hall, 2006; Warren & Griffiths, 2003), we followed Zatorre et al. (2002) and varied the number of auditory objects present in the signal as our nonspatial manipulation. In addition, we also included an auditory motion condition to determine whether this kind of spatial signal may generate more selective responses (Krumbholz et al., 2005; Warren, Zielinski, Green, Rauschecker, & Griffiths, 2002; Lewis, Beauchamp, & DeYoe, 2000; Baumgart, Gaschler-Markefski, Woldorff, Heinze, & Scheich, 1999).

METHODS

To address the questions outlined above, we conducted an fMRI study using speech stimuli that involved three spatial conditions (stationary sounds presented at one location, stationary sounds presented at three locations, and moving sounds) and two object conditions (one vs. three distinct individuals speaking). To approximate more naturalistic spatial percepts, individualized head-related transfer functions (HRTFs) were employed in stimulus development. Our primary analyses focused on the PT region.

Participants

Ten subjects (8 men) participated in this study. Subjects gave informed consent under a protocol approved by the Institutional Review Board at the University of California, Irvine.

Stimuli

Stimuli were sentences taken from the TIMIT sentence corpus (developed by Texas Instruments, Dallas, TX, and the Massachusetts Institute of Technology, Cambridge, MA). Sets of three to five sentences (depending upon length) were used to form 15-sec blocks. Each block of sentences was presented through loudspeakers and recorded inside each subject's ear canals using a pair of Etymotic ER-7C probe tube microphones (silicon tubes were placed 1–2 cm inside the ear canal; Middlebrooks, Makous, & Green, 1989; Wightman & Kistler, 1989). All sounds were recorded digitally at 44.1 kHz through 16 bit A-to-D converters using MATLAB software in a steel double-walled acoustically iso-lated chamber (Industrial Acoustics Company, New York, NY), the surfaces of which were covered with 10.2-cm acoustic foam wedges (Sonex, Seattle, Washington) to reduce reverberation during recordings. Stationary stimuli were played through loudspeakers positioned at 70 cm from the subject's head (either at −60°, 0, or 60°, with zero degrees defined as directly in front of the subject). Motion stimuli were played through a loudspeaker attached to a microprocessor controlled arc (Stepper motor system: Arrick Robotics, model MD-2, Tyler, TX), which rotated in a circular trajectory with a radius of 70 cm around the subject's head on the azimuthal plane. The motion stimuli covered 120° (from −60 to +60) in 15 sec, and thus, had a velocity of 8 deg/sec (this velocity has shown to produce low motion-detection thresholds in the free-field; Saberi & Perrott, 1990).

Once recorded, stimuli were additionally filtered off-line with the inverse transfer function of the insert headphones (Sensimetrics Corporation model V14 electrostatic transducers) used in the scanner to ensure that the HRTFs represent only the effects of the subject's own transfer functions and are not distorted by the headphones' transfer functions (separate inverse digital filters were used for the left and right channels). Each set of sentences was thus filtered only with a subject's individualized HRTFs to produce the percept of externalized sounds during playback in the MRI scanner.

Design

Two variables were manipulated in our study: one spatial variable and one auditory object variable. The object variable consisted of the number of talkers (1 vs. 3). The spatial variable had three levels: sound source at one location, three locations, or moving. This resulted in the crossed 3 × 2 design shown in Table 1. For the three-location conditions, when only one talker was presented, the location of that talker's voice randomly changed between locations throughout the block at a rate of one location/second. When three talkers were presented, the talkers' voices were presented simultaneously, each at a different location, which was maintained throughout the block (i.e., they did not change locations at all during the block). Total stimulus energy was equated within the one-talker conditions and within the three-talker conditions. We did not equate energy across the one- versus three-talker conditions because to do so would necessarily decrease the amplitude of each talker in the three-talker condition, which in turn would introduce a signal-to-noise ratio confound: A given speech stream (talker) in the three-talker conditions would have lower signal-to-noise than the speech stream of the one-talker conditions, and thus, reduce the ability to segregate one object (talker) from remaining objects (Kidd, Mason, & Gallun, 2005; Arbogast, Mason, & Kidd, 2002), particularly in the background scanner noise. We chose to equate signal-to-noise within each speech stream across conditions rather than overall acoustic energy. It is clear from our findings that

Table 1. Experimental Design Matrix

		Spatial Manipulation		
		1 Location	*3 Locations*	*Moving*
Number of talkers	1 talker	1 talker, 1 location	1 talker jumping between 3 locations	1 talker moving smoothly and continuously between −60° and +60°
	3 talkers	3 talkers simultaneously presented at 1 location	3 talkers, each at a different location	3 talkers moving smoothly and continuously between −60° and +60°

differences in the acoustic energy between the one- versus three-speaker conditions cannot explain our data (see below).

Stimuli were presented in a blocked design with 15-sec stimulation blocks alternating with 15-sec rest periods. A total of eight blocks per condition were presented across the entire session. The order of blocks was pseudo-randomized with the constraints that two blocks of the same condition could not be presented back to back, and each condition had to occur with equal frequency in each run. Block order within each run was fixed, and runs were counterbalanced across subjects.

Data Acquisition and Procedures

Thirty-two axial slices were collected using a 3-T Philips Intera Achieva 3T MR system (FOV = 256 mm, matrix = 64×64, size = 474 mm, TE = 40, slice thickness = 6 mm). For each subject, a high-resolution anatomical image was acquired with a magnetization-prepared rapid acquisition with gradient-echo pulse sequence.

Stimuli were presented in the scanner using Sensimetrics model V14 electrostatic headphones. Synchronization with the scanner was achieved manually. Subjects were instructed to fixate on a center cross, which was monitored using an eye-tracking system, and to attend to the stimuli, similar to previous studies, which have documented differences between spatial and object-based processing (Altmann et al., 2007; Warren & Griffiths, 2003). Subjects reported no apparent motion during the one-talker, three-location condition, and an externalized quality for all sound stimuli.

Analysis

To correct for subject motion artifacts, the image volumes of each subject were aligned to the sixth volume in the series using a 3-D rigid-body, six-parameter model in the AIR 3.0 program (Woods, Grafton, Holmes, Cherry, & Mazziotta, 1998). The volumes were then coregistered to the high-resolution anatomical image. After alignment, each volume was spatially smoothed (Gaussian spatial filter, 4 mm FWHM). For each subject, the data were normalized and corrected for variations in slice timing acquisition.

Due to high anatomical variability in the posterior Sylvian region (Steinmetz et al., 1990), we used an individual subject analysis. Regression analysis was performed separately on each subject using AFNI software. Predictor vectors, which represented the time course of stimulus presentations for each condition, were convolved with a standard hemodynamic response function and entered into the analysis to find the parameter estimates that best explain the variability in the data at each voxel.

ROIs were defined in each subject using two separate planned contrasts (*t* tests), each thresholded at $p < .0001$ (uncorrected). The first contrast aimed to identify subregions of the PT in each subject that were sensitive to spatial manipulations. These ROIs were defined by identifying contiguous voxels that were significantly ($p < .0001$) more active during the three-location/one-talker condition compared to the one-location/one-talker condition. The second contrast aimed to identify subregions of the PT in each subject that were sensitive to auditory object manipulations. These ROIs were defined by identifying contiguous voxels that were significantly more active during the three-talker/one-location condition compared to the one-talker/one-location condition. We probed activations in both the left and right PT of each subject, which we defined anatomically in reference to each subject's own structural MRI, as the cortical region on the supratemporal plane posterior to the Heschl's gyrus. Thus, all ROIs were confirmed to be within the PT in each subject. The time courses from the top five most highly activated voxels in each ROI were extracted and submitted to further group-level analyses to explore activation patterns across all conditions in both the spatially defined and object-defined ROIs. Specifically, time courses for each of the six conditions were extracted from each ROI in each subject, and activation amplitude estimates were computed for each condition by comparing the average signal amplitude during the middle 6 TRs (12 sec) of the activation blocks against the baseline activation (3 TRs prior to the activation blocks and 3 TRs following the activation blocks). This procedure yielded a single amplitude estimate for each condition in each subject in each ROI. These data were then submitted to ANOVAs to assess activation patterns within each ROI.

RESULTS

Overall, the two ROI-defining contrasts (3 minus 1 location; 3 minus 1 talker) yielded relatively consistent activation

across all subjects in the PT. For the 3-minus-1 location contrast, all 10 subjects showed significant activation in the left PT, and 9 out of 10 in the right PT. For the 3-minus-1 talker contrast, 8 out of 10 subjects showed significant activation in the left PT, but no right PT activation in any subject. These ROIs within the left hemisphere were not identical, but nonetheless, overlapped substantially (and, in fact, appear to be functionally indistinguishable; see below). For convenience, the normalized location of these activations across subjects can be visualized in Figure 1, which shows the activation peak in Talairach space for each subject on an average brain template. The fact that we were able to identify consistent activation in
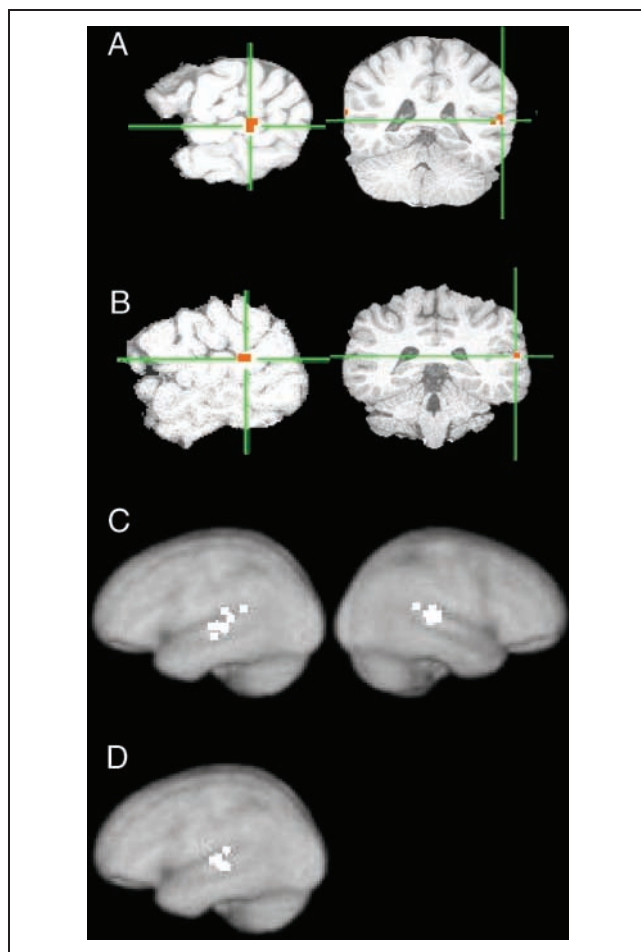

Figure 1. Spatial- and object-based activations. (A) Sample activation focus in coronal and sagittal views from one participant for the spatially defined ROI. (B) Sample activation focus in coronal and sagittal views from one participant for the object-defined ROI. (C and D) Normalized activation foci from each participant displayed on a normalized MNI template brain. (C) Activation foci for the spatially defined ROIs for each subject in the left (mean: −52, −26, 8) and right (mean: 55, −27, 10) hemispheres. (D) Activation foci for the object-defined ROIs for each subject in the left hemisphere (mean: −55, 23, 7; object defined ROIs were not consistently found in the right hemisphere). Note: Normalized activations are approximate as the normalization process introduces localization error. For analysis purposes, we defined ROIs relative to each subjects own unnormalized structural MRIs.
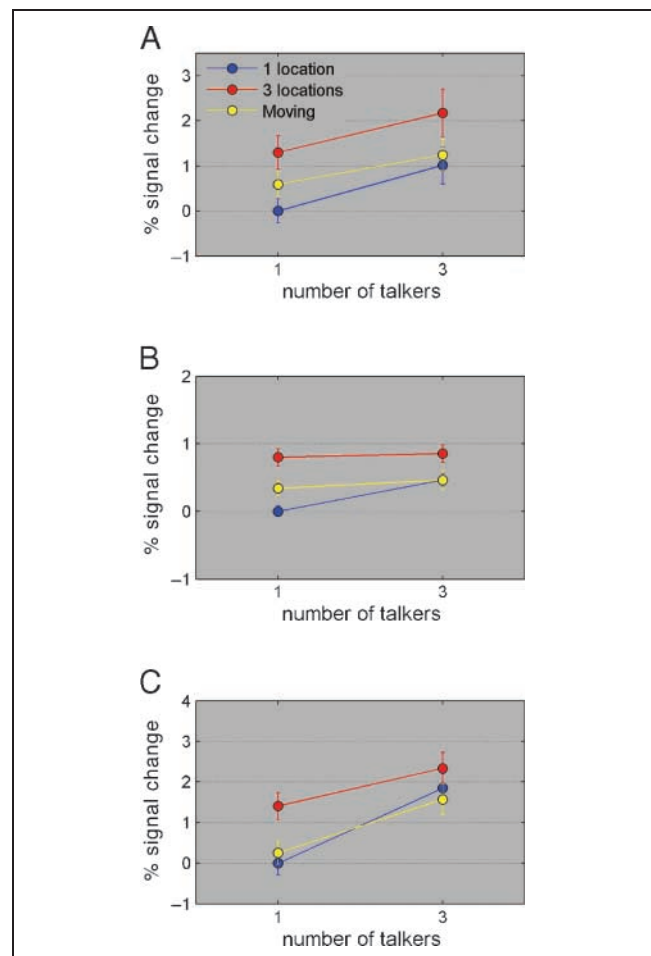

Figure 2. Mean signal amplitudes and standard error bars for voxels in the left spatially defined ROI (A), in the right spatially defined ROI (B), and in the left object-defined ROI (C).
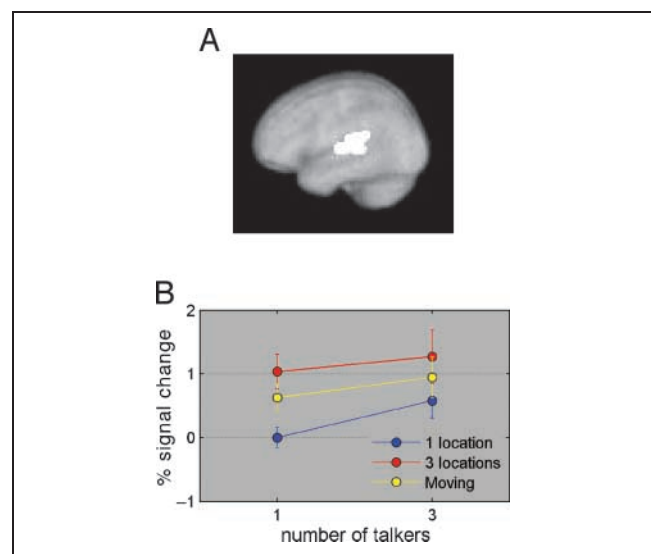

Figure 3. (A) Activation maps depicting the center coordinate of activation for each subject on a template brain for the motion defined ROI. (B) Mean signal amplitudes with standard error bars for voxels in the motion-defined ROI (left hemisphere).

all participants in the PT for the spatial manipulation (with object information held constant) and for the object manipulation (with location information held constant) suggests that the PT generally is sensitive to both of these sources of information, consistent with previous studies. However, previous studies found that subparts of the PT were selective for spatial manipulations, and this question is the focus of subsequent analyses.

Spatially Defined ROI (3 Minus 1 Location)

Data were examined from all conditions within these location-defined PT ROIs, one in each hemisphere. Figure 2A and B shows the mean signal amplitudes for each condition in each hemisphere in these location-defined ROIs. A 2 (1 vs. 3 locations) × 2 (1 vs. 3 talkers) repeated measures ANOVA was carried out with signal amplitude entered as the dependent variable. Analyses were carried out separately in each hemisphere. Motion activations were not included in the analysis but are plotted in Figure 2 for comparison. For both hemispheres, there was a significant main effect for the spatial manipulation—not surprisingly because the ROIs were defined by this contrast—with greater activation in the three-location than one-location condition [left: $F(1, 9) = 18.721, p < .005$; right: $F(1, 8) = 21.902, p < .005$]. However, we also found a significant main effect of the object manipulation with greater activation for the three-talker versus the one-talker condition [left: $F(1, 9) = 5.492, p < .05$; right: $F(1, 8) = 7.649, p < .03$]. These factors did not interact in the left hemisphere [$F(1, 9) = .04, ns$], but we did find a significant interaction in the right hemisphere [$F(1, 8) = 6.509, p < .04$]: The effect of the spatial manipulation was reduced (but still significant in a paired t test, $p < .04$) in the three-talker condition (see Figure 2B). Overall, this analysis showed that even in the spatially defined ROIs, auditory object manipulations had an effect on activation levels.

Auditory Object-defined ROI[1]

Data were examined from all conditions within this auditory object-defined ROI within the left PT. Figure 2C shows the mean signal amplitudes for each condition in each hemisphere in these location-defined ROIs. A 2 (1 vs. 3 locations) × 2 (1 vs. 3 talkers) repeated measures ANOVA was carried out with signal amplitude entered as the dependent variable. Motion activations were not included in the analysis but are plotted in Figure 2 for comparison. As in the previous contrast, significant main effects were found not only for the object manipulation [$F(1, 7) = 17.168, p < .005$] with three talkers producing greater activation than one talker (which defined the ROI), but also for the spatial manipulation [$F(1, 7) = 16.789, p < .005$], with three locations yielding greater activation than one location. A significant interaction was also observed [$F(1, 7) = 22.408, p < .005$]: The spatial manipulation only significantly modulated signal amplitude in the one-talker

condition. Overall, this analysis showed that even in the auditory object-defined ROI, spatial manipulations had an effect on activation levels, although only in the one-talker condition.

Motion-defined ROI

As is clear from the graphs in Figure 2, neither the spatial-defined nor object-defined PT ROIs are motion-selective. In fact, nonmoving, but spatially varying, signals produce more activation in these regions than do the moving stimuli. To assess the possibility that a motion-selective region might exist in some other region of the PT, we used the one-talker motion condition contrasted with the one-talker/one-location condition (i.e., stationary) to define PT ROIs. A motion-defined ROI based on this contrast was identified in 9 out of 10 subjects in the left PT, largely overlapping the other ROI locations (Figure 3). The right PT was not consistently activated across subjects in this contrast. As with other ROIs, data were examined from all conditions within this motion-defined ROI. However, because we were specifically interested in the relation between the motion conditions and the non-motion conditions, we conducted a 3 × 2 ANOVA that contained a three-level spatial variable (1 location vs. 3 locations vs. motion) and a two-level object variable (1 vs. 3 talkers). This analysis revealed a main effect of the spatial variable [$F(2, 8) = 7.8, p = .004$]. Post hoc analyses showed that this main effect was attributable to the motion and three-location conditions yielding significantly more activation than the one-location condition [motion > 1-location: $t(8) = 2.86, p = .02$, two-tailed; 3-location > 1-location: $t(8) = 3.15, p = .01$, two-tailed]. There was a trend for *greater* activation in the three-talker condition than in the motion condition ($p = .1$, two-tailed). There was no main effect of the object manipulation ($p = .21$) and no interaction ($p = .14$) (see Figure 3B). It is clear from this analysis that even within a motion-defined PT ROI, responses are not selective for motion. The nonmoving, three-location conditions produced an equal amount of (and perhaps even more) activation than the motion conditions. Also of interest is the observation that activation associated with three talkers presented at one location yielded the same amount of activation as the one-talker moving stimulus. Thus, spatial variation is not even necessary to activate this motion-defined ROI to levels equal to a moving stimulus.

Other Analyses

Although our focus is on activation patterns in the PT, we did explore our dataset for consistent activations in other brain regions. For the 3-minus-1 location contrast, 9 out of 10 subjects had significant activation in the left posterior superior temporal sulcus ($x = -52, y = -30, z = 1$; no consistent patterns were found in the right hemisphere). At lower thresholds, this superior temporal

sulcus activation merged with the PT activation noted above. For the 3-minus-1 talker contrast, no other brain area in either hemisphere, including Heschl's gyrus, showed a consistent pattern across subjects.

DISCUSSION

Our primary goal in this study was to assess sensitivity in the PT to auditory spatial and auditory object manipulations. Previous studies had questioned the existence of a purely spatial processing mechanism (Zatorre et al., 2002), arguing instead that spatial effects may be attributed to the use of spatial information as a cue to object identification. Although a number of studies, including the present experiment, have shown that a pure spatial manipulation can activate the PT (Altmann et al., 2007; Smith, Saberi, & Hickok, 2007; Barrett & Hall, 2006; Smith, Okada, Saberi, & Hickok, 2004; Warren & Griffiths, 2003), in contrast to the findings of Zatorre et al. (2002), none of these studies assessed the selectivity of the spatial response by manipulating the number of simultaneously presented auditory objects in the signal. The present study aimed to address this issue.

Our results showed that spatially sensitive regions of the human PT, namely, those regions that responded more for a spatially varying stimulus than a spatially nonvarying stimulus, were also modulated by varying the number of auditory objects simultaneously present in the signal. The same held true if regions sensitive to auditory motion were examined. That is, in contrast to previous findings (Altmann et al., 2007; Barrett & Hall, 2006; Warren & Griffiths, 2003), we did not find any region of the PT that responded selectively to spatial (including motion) over object manipulations. We attribute this discrepancy to the object manipulation used in this study (number of object present in the signal) compared to others (pitch or pattern change). Likewise, auditory object-sensitive regions of the PT, namely, those regions that responded more for three objects presented simultaneously at one location than one object presented at one location, were also modulated by spatial variation. Thus, there was no region within the PT that responded selectively to auditory object information.

When activation patterns across all conditions were examined within both the spatially defined and object-defined PT ROIs, we found evidence confirming sensitivity to both classes of information in the PT in each ROI. For example, in both the left- and right-hemisphere spatially defined ROIs, the object manipulation had significant effects on activation level (3 talkers > 1 talker), and this was true even in the one-location condition. Likewise, within the object-defined ROI, we found significant effects of the spatial manipulation (3 locations > 1 location, in the 1-talker condition). More generally, the overall pattern across the ROIs (although less robust in the object-defined ROI) was that the lowest activation levels were found for the one-talker/one-location condition, and the highest for the three-talker/three-location condition, with the other two conditions (1-talker/3-locations and 3-talkers/1-location) falling in between. A particularly interesting aspect of this pattern, especially in relation to the spatially defined ROIs, concerns the fact that the greatest activation was for the three-talker/three-location condition. What is interesting is that this condition did not involve any spatial variation. We had previously suggested (Smith et al., 2004, 2007) that auditory spatial activations in the PT—whether produced via moving or nonmoving but spatially varying stimuli—resulted from processes involved in detecting spatial change. However, the present study found that even regions that respond more to a spatially varying sound source compared to nonspatially varying sound source respond even more vigorously to multiple auditory objects at distinct but nonvarying spatial locations. This suggests that it is not the computation of spatial location alone that is driving activation in these regions.

We interpret these data as support for the view that putative spatial regions of the PT are more concerned with using spatial information to inform processes underlying auditory stream segmentation (Zatorre et al., 2002). This result is not inconsistent with previous demonstrations in showing distinctions between object- and location-based responses in auditory cortex (Lomber & Malhotra, 2008; Altmann et al., 2007; Alain et al., 2001; Maeder et al., 2001; Rauschecker, 1998). Indeed, like previous studies, we find evidence of spatially sensitive cortical areas in the PT, even when object-based information is held constant. Rather, our findings go beyond previous results by showing that if one introduces object manipulations to these spatial hearing experiments, joint sensitivity to these classes of information is clearly evident. Of course, it remains a possibility that within these ROIs the two types of information rely on different populations of intermixed cell ensembles. This possibility is an important topic for future studies.

Consistent with our two previous studies (Smith et al., 2004, 2007), we again found no evidence for a motion-selective region within the PT (or anywhere in the cortex). Within our (nonmoving) spatially defined ROIs, moving stimuli produced significantly more activation than one-location/one-talker stimuli. However, nonmoving stimuli of a variety of sorts yielded activation equal to or greater than the moving stimuli (Figures 2 and 3). Further, even within a motion-defined PT ROI (moving vs. nonmoving talker), we found equal or greater activation for a variety of nonmoving conditions, including a condition that involved three talkers presented simultaneously at one location. Thus, it is now quite clear that the putative human auditory motion area (Krumbholz et al., 2005; Pavani, Macaluso, Warren, Driver, & Griffiths, 2002; Warren et al., 2002; Lewis et al., 2000; Baumgart et al., 1999)—defined by contrasting moving with nonmoving stimuli—is in fact, not motion-selective at all, as we have previously argued (Smith et al., 2004, 2007). Furthermore, and in contrast to our previous suggestion, this region is not even

selective for auditory spatial information: The present study found equivalent levels of activation during the perception of a moving voice as was found during the perception of three voices presented simultaneously at a single location.

The ROIs identified by the various contrasts (3 vs. 1 location, 3 vs. 1 talker, and motion vs. stationary) were largely overlapping but not identical. It is unclear whether the differences reflect true functional variation in the underlying neural systems, or just random variation. The qualitatively similar response pattern found across all ROIs, and the extensive overlap of the ROIs suggests some uniformity of function. One clear difference, however, was the lack of consistent right-hemisphere activation across our subjects in the object- and motion-defined ROIs, in contrast to the (nonmotion) spatially defined ROI which was identified in 90% of subjects (at the threshold we used). This hints at some functional distinction within the PT, although we note that even within the right-hemisphere spatially defined ROI, activation patterns across the other conditions was similar to the left-hemisphere ROIs. Additional studies are needed to sort out the finer-grained functional organization of this region.

The present failure to find spatially selective responses in the human PT raises questions about the functional role of the auditory dorsal stream. If we are correct about this region playing a primary role in auditory source separation, such a function would seem to align more closely with a "what" stream than a "where" stream. An alternative proposal for the function of the auditory dorsal stream is that it supports sensory–motor integration (Hickok & Poeppel, 2000, 2004, 2007; Warren et al., 2005) analogous to sensory–motor integration regions in posterior parietal cortex (Andersen, 1997). Several recent experiments have shown that a region at the posterior extent of the PT region (area Spt) exhibits response properties similar to those found in posterior parietal regions including both sensory and motor responsivity (Buchsbaum, Olsen, Koch, & Berman, 2005; Hickok et al., 2003; Buchsbaum, Hickok, & Humphries, 2001), functional connectivity with motor areas (Buchsbaum et al., 2001), motor-effector selectivity (Pa & Hickok, 2008), and multisensory responsivity (Okada & Hickok, 2009). Based on these findings, we have argued that the auditory dorsal stream is not so much an "auditory" pathway as it is a sensory–motor integration network for the vocal tract effectors which happens to receive a great deal of input from the auditory system because auditory information is most critical to vocal tract behaviors such as speech and vocal music production (Pa & Hickok, 2008; Hickok & Poeppel, 2007). This system appears to be quite distinct anatomically from regions of the PT that are spatially responsive based on within-subject fMRI studies of spatial and sensory–motor functions, the latter being more posterior and perhaps extending into the parietal operculum (Okada & Hickok, unpublished data). On this view, the anterior PT comprises auditory

fields, some of which support auditory source separation, whereas the posterior PT is not part of auditory cortex and supports sensory–motor integration for the vocal tract effectors. This division of the PT into auditory and nonauditory sectors fits nicely with available cytoarchitectonic data (Hackett, De La Mothe, et al., 2007; Hackett, Smiley, et al., 2007; Galaburda & Sanides, 1980).

In summary, we have found that although spatial manipulations of an acoustic signal (speech in this case) can modulate activity in the human PT, this activity is not selective for spatial information as these regions are also sensitive to the number of auditory objects present in the signal. We suggest that this region of the cortex is not functioning as a dedicated auditory spatial processing system, but instead is integrating spatial location information with auditory object information as part of a sound source separation process (Zatorre et al., 2002). It will be instructive in future work to determine whether such a mechanism is informed preferentially by spatial cues, or whether other cues to sound source segregation might also be processed in this region.

Note

1. Although this contrast confounds low-level acoustic factors with number of auditory objects (more talkers = more spectro-temporal variation, higher acoustic energy), our observed effect in the PT cannot be explained by such acoustic differences because (i) if low-level acoustic differences were driving changes in brain activity, such an effect should be apparent throughout many auditory regions such as primary auditory cortex, which was not the case; (ii) such an effect should be equally robust in both hemispheres, which was not the case; and (iii) as can be seen clearly in Figure 2A, conditions with less acoustic energy (1 talker, 3 locations) can produce as much or more activation than conditions with more acoustic energy (3 talkers, 1 location).

REFERENCES

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 12301–12306.

Altmann, C. F., Bledowski, C., Wibral, M., & Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *Neuroimage, 35,* 1192–1200.

Andersen, R. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 352,* 1421–1428.

Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic

masking of speech. *Journal of the Acoustical Society of America, 112,* 2086–2098.

Barrett, D. J., & Hall, D. A. (2006). Response preferences for "what" and "where" in human non-primary auditory cortex. *Neuroimage, 32,* 968–977.

Baumgart, F., Gaschler-Markefski, B., Woldorff, M. G., Heinze, H. J., & Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature, 400,* 724–726.

Buchsbaum, B., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science, 25,* 663–678.

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron, 48,* 687–697.

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology, 190,* 597–610.

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex: II. Thalamocortical connections of the caudal superior temporal plane. *Journal of Comparative Neurology, 502,* 924–952.

Hackett, T. A., Smiley, J. F., Ulbert, I., Karmos, G., Lakatos, P., de la Mothe, L. A., et al. (2007). Sources of somatosensory input to the caudal belt areas of auditory cortex. *Perception, 36,* 1419–1430.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience, 15,* 673–682.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4,* 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92,* 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8,* 393–402.

Kidd, G., Mason, C. R., & Gallun, F. J. (2005). Combining energetic and informational masking for speech identification. *Journal of the Acoustical Society of America, 118,* 982–992.

Krumbholz, K., Schonwiesner, M., Rubsamen, R., Zilles, K., Fink, G. R., & von Cramon, D. Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *European Journal of Neuroscience, 21,* 230–238.

Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex, 10,* 873–888.

Lomber, S. G., & Malhotra, S. (2008). Double dissociation of "what" and "where" processing in auditory cortex. *Nature Neuroscience, 11,* 609–616.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: A human fMRI study. *Neuroimage, 14,* 802–816.

Middlebrooks, J. C. (2002). Auditory space processing: Here, there or everywhere? *Nature Neuroscience, 5,* 824–826.

Middlebrooks, J. C., Makous, J. C., & Green, D. M. (1989). Directional sensitivity of sound–pressure levels in the human ear canal. *Journal of the Acoustical Society of America, 86,* 89–108.

Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters, 452,* 219–223.

Pa, J., & Hickok, G. (2008). A parietal–temporal sensory–motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. *Neuropsychologia, 46,* 362–368.

Pavani, F., Macaluso, E., Warren, J. D., Driver, J., & Griffiths, T. D. (2002). A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Current Biology, 12,* 1584–1590.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology, 8,* 516–521.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience, 2,* 1131–1136.

Saberi, K., & Perrott, D. R. (1990). Minimum audible movement angles as a function of sound source trajectory. *Journal of the Acoustical Society of America, 88,* 2639–2644.

Smith, K. R., Okada, K., Saberi, K., & Hickok, G. (2004). Human cortical motion areas are not motion selective. *NeuroReport, 9,* 1523–1526.

Smith, K. R., Saberi, K., & Hickok, G. (2007). An event-related fMRI study of auditory motion perception: No evidence for a specialized cortical system. *Brain Research, 1150,* 94–99.

Steinmetz, H., Rademacher, J., Jancke, L., Huang, Y., Thron, A., & Zilles, K. (1990). Total surface of temporoparietal intrasylvian cortex: Diverging left–right asymmetries. *Brain and Language, 39,* 357–372.

Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *Journal of Neuroscience, 23,* 5799–5804.

Warren, J. D., Zielinski, B. A., Green, G. G., Rauschecker, J. P., & Griffiths, T. D. (2002). Perception of sound-source motion by the human brain. *Neuron, 34,* 139–148.

Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: Auditory–motor transformations and the posterior temporal plane. *Trends in Neurosciences, 28,* 636–643.

Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening: II: Psychophysical validation. *Journal of the Acoustical Society of America, 85,* 868–878.

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural sub-systems within "Wernicke's area". *Brain, 124,* 83–95.

Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., & Mazziotta, J. C. (1998). Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography, 22,* 141–154.

Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is "where" in the human auditory cortex? *Nature Neuroscience, 5,* 905–909.