Recent evidence has shown that early sensory cortex encodes not only low-level sensory properties of a stimulus, but also mid- to high-level perceptual information. For example, early visual cortex has been implicated in figure-ground discrimination (Heinen et al. 2005; Huang and Paradiso 2008; Hopé et al. 1998), shape and size perception (Fang et al. 2008; Murray et al. 2002, 2006), lightness constancy (Boyaci et al. 2007), attentional modulation (Datta and DeYoe 2009; Fischer and Whitney 2009; Ress et al. 2000), tracking stimulus reward history (Serences 2008), conscious perception (Hsieh et al. 2006; Hsieh and Tse 2009, 2010a, b), and even identification of a stimulus (Hsieh et al. 2010). Similarly, responses in early auditory cortex are modulated by attention (Alho et al. 1999; Grady et al. 1997; Hillyard et al. 1973; Jäncke et al. 1999; Lipschutz et al. 2002; Näätänen 1990; O’Leary et al. 1997; Tzourio et al. 1997; Woldorff et al. 1993; Woldorff and Hillyard 1991; Woodruff et al. 1996) and reflect cross-modal processing (Calvert et al. 1997; Calvert and Campbell 2003; Foxe and Schroeder 2005; Ghazanfar and Schroeder 2006; Meyer et al. 2010; Schroeder and Foxe 2005). However, this study was limited by its inability to determine whether the pattern information in auditory cortex actually reflects implied sound information or merely visual information. Here, we tested whether the representation of a stimulus in auditory cortex reflects contextual/implied auditory information or visual information per se.

Our experiment was conducted with functional MRI (fMRI) and included three conditions: silence, action control, and sound. In the silence condition, we presented different visual stimuli that implied similar knocking sounds. Our prediction was that the spatial pattern of blood oxygen level-dependent (BOLD) activation in auditory cortex would represent these visual stimuli differently. To distinguish further whether this pattern information reflects contextual/implied auditory information or visual information per se, we presented actions that do not imply sounds in the action control condition. If the pattern information observed in the silence condition reflects implied auditory information, we should expect the pattern information to be absent for the action control condition. However, if the pattern information observed in the silence condition reflects purely visual information, the pattern analysis should reveal information for the action control condition as well. Moreover, to determine whether this top-down signal is robust enough to persist with the addition of a salient bottom-up signal, we paired the visual stimuli from the silence condition with identical sounds in the sound condition.

METHODS

Participants. Ten volunteers between 18 and 30 yr old participated in the study. All of them were healthy and right-handed and had normal or corrected-to-normal visual and auditory acuity. All subjects gave informed written consent within a protocol passed by the Duke-NUS Graduate Medical School or the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects and were compensated with 60 dollars for their participation. Experimental procedures. Scanning was performed at the McGovern Institute at the Massachusetts Institute of Technology in
Cambridge, MA, with the Athinoula A. Martinos Imaging Center 3T Siemens Trio scanner. fMRI runs were acquired using a gradient-echo, echo-planar sequence (repetition time (TR) = 3 s, echo time (TE) = 30 ms, 2 × 2 × 2 mm + 20% spacing). Forty-six slices were collected with a 32-channel head coil. Slices were oriented roughly perpendicular to the calcarine sulcus and covered the whole brain.

Subjects viewed blocks of six 2-s color videos that corresponded to the three experimental conditions with two sequences of visual stimuli each (Fig. 1 and supplementary material available online at the Journal of Neurophysiology web site). The agent in the video, which was either a woodpecker or a human hand, distinguished the visual stimuli. Videos from the first condition (sound) visually depicted a sequence of generic knocking sounds in lieu of the original audio tracks, such that auditory stimulation was identical across visual stimuli. Videos from the second condition (“silence”) included identical visual stimuli without any auditory stimulation. Videos from the third condition (“action control”) depicted the same 2 agents in motion without performing any knocking actions and also without any auditory stimulation.

Appearing at the beginning of a 3-s TR, the 2-s videos had a resolution of 960 × 540 pixels and subtended 12.25° × 6.75° of visual angle against a black background. A white fixation cross was always present in the center of the display and subtended 0.75° × 0.75°. Seven repetitions of a given condition-stimulus pair occurred within each of the 12 21-s blocks that were interleaved among 15-s fixation periods within each run (Fig. 2). Six seconds of fixation at the beginning and end of a run were added to yield a total duration of 429 s for each run. For each video presentation, the overall luminance of the video was randomly selected to be 25% greater or lesser than its original value. Subjects were required to press 1 of 2 buttons on a response box (2-alternative forced choice) to indicate whether each individual 2-s video appeared atypically bright or dim, and every subject’s performance was near ceiling. The presentation order of the 6 condition-stimulus pairs was randomized within each run. While being scanned, all subjects completed between 9 and 11 runs.

**ROI identification.** Functional localization of the region of interest (ROI) was based on three runs of a separate auditory localizer. It consisted of 30-s blocks in which repeated 100-ms pure tones were presented with an interstimulus interval of 400 ms while subjects fixated. Conditions were defined by the three frequency ranges of the blocks, namely low (340–870 Hz), middle (880–2,170 Hz), and high (2,370–5,900 Hz). The frequencies of tones within each block increased linearly with modulo 5.2 s, corresponding to slopes of 102, 248, and 679 Hz/s, respectively. For example, the first tone in a low-frequency block was always 340 Hz, and the frequency increased by 51 Hz for the next tone 500 ms later. If, at the end of a 5.2-s loop, the frequency exceeded 870 Hz by \( x \), it would be reset to 340 + \( x \) Hz and looped again throughout a 30-s block. Since we did not find a significant difference between the BOLD responses to any of the three conditions, auditory cortex was defined bilaterally as the contiguous regions in the vicinity of the superior and transverse (Heschl’s) temporal gyri that responded significantly more robustly to auditory stimulation, combined across the three conditions, than to background noise alone (e.g., from the scanner; \( P < 10^{-12} \)). ROI sizes ranged from 177 to 432 voxels with a mean of 331.4 voxels and a standard deviation of 70.7.

**Data analysis.** Data analysis was conducted using the fMRI software package FreeSurfer (http://surfer.nmr.mgh.harvard.edu) and MATLAB (The MathWorks). The preprocessing steps for both the localizer and experimental runs included motion correction and intensity normalization. Preprocessing for the localizer runs also included spatial smoothing using a Gaussian kernel with a full-width at half-maximum of 6 mm. A gamma function with \( \delta = 2.25 \), \( \tau = 1.25 \), and \( \alpha = 2 \) was used to estimate the hemodynamic response for each condition in both the experiment and the localizer.
In addition to a standard univariate analysis of the mean BOLD response to each condition, a correlational analysis was performed on the $\beta$-weights between visual stimuli for each condition in each voxel with a standard multivariate pattern analysis method (Haxby et al. 2001). Data were split into odd and even runs, and spatial patterns of response were extracted from each subset of data for the six condition-stimulus pairs. The patterns were first normalized, such that the mean response in each voxel across the visual stimuli to be compared was subtracted from the response to each stimulus for each half of the data before the correlation values were calculated. Within each ROI, we then computed the split-half correlations as Pearson correlation coefficients between the normalized activity patterns in response to the two sequences of visual stimuli for each of the three experimental conditions, that is, sound, silence, and action control. These correlations were computed for each subject and then averaged across subjects by condition.

RESULTS

The results of the univariate analysis (Fig. 3) revealed increased mean BOLD activation in auditory cortex for the sound condition relative to baseline ($P < 10^{-4}$) and to the silence and action control conditions ($P < 10^{-3}$). However, no differential mean activation was observed between the silence and action control conditions ($P > 0.05$).

The results of the multivariate pattern analysis (Fig. 4) indicated that the spatial patterns of BOLD activation in auditory cortex contained information that distinguished between the visual stimuli (hand and woodpecker) for the silence condition ($P = 0.015$). However, this was not the case for the sound ($P > 0.05$) and action control conditions ($P > 0.05$). Moreover, decoding accuracy for which visual stimulus was presented in the silence condition was significantly greater than decoding for the sound condition ($P = 0.005$) and the action control condition ($P = 0.039$). These findings demonstrated that 1) top-down information may dominate processing in auditory cortex only when bottom-up auditory input is absent, and that 2) there may be no such top-down information in auditory cortex when the visual stimuli do not imply sound. As such, these findings suggest that spatial patterns of BOLD activation in auditory cortex do not reflect purely visual information per se, but rather visually induced top-down information.

DISCUSSION

Our results show that the spatial patterns of BOLD activation in auditory cortex reflected a cross-modal influence of visual information when bottom-up auditory input was absent. In contrast, no top-down information was found in auditory cortex when the visual stimuli did not imply sound (i.e., in the action control condition). These findings suggest that 1) the activity being captured in the pattern of responses across auditory cortex reflects high-level, top-down, and cross-modal information, and that 2) stimuli from one modality (e.g., vision) that do not provide a context relevant to another modality (e.g., audition) will not exert a cross-modal top-down influence. Note, however, that we cannot completely rule out the possibility of such top-down effects in the action control condition insofar as top-down activity might simply be weaker, less spatially consistent (i.e., requiring greater spatial resolution), or otherwise structured in a way that cannot be detected with fMRI and multivariate pattern analysis.

Our data did not reveal any visually induced pattern information for the sound condition when the bottom-up auditory signal, which was identical across visual stimuli, was present. One possible explanation is that modulatory top-down information is overshadowed by the more robust bottom-up auditory information. An alternative account is that top-down modulatory signals might be assigned less weight or even disappear altogether when bottom-up signals are strong and unambiguous. Increased mean BOLD activation was observed for the sound condition, however, indicating that the information captured with multivariate analysis, which is encoded in the spatial pattern of activation, is distinct from that captured with univariate analysis, which is encoded in the overall activation across voxels.

The difference in spatial patterns observed in auditory cortex between visual stimuli in the silence condition is likely due to some combination of contextual information and implicit auditory imagery. For example, previous findings have shown that auditory imagery activates secondary auditory cortex (Bunzeck et al. 2005; Halpern et al. 2004; Yoo et al. 2001; Zatorre and Halpern 2005) and that sound-implying visual
stimuli can be decoded on the basis of activity in auditory cortex (Meyer et al. 2010). However, it is worth noting that our ROIs were mainly within Heschl’s gyrus and primary auditory cortex. This possible discrepancy with work identifying the neural correlates of auditory imagery in secondary auditory cortex is likely due to some combination of the following. First, primary auditory cortex may actually be involved in imagery, but the information contained within this region could only be detectable with pattern analysis. Second, the patterns of responses to the silence condition may not be driven by imagery at all, but rather by visually induced (audition-relevant) contextual information. Little activity in secondary auditory cortex is to be expected in either case, so determining the validity of these possibilities and dissociating their contributions to our main results will require further research.

To conclude, our data show that 1) auditory cortex contained information about visual stimuli when bottom-up auditory input was absent, and that 2) no contextual top-down information was observed in auditory cortex when the visual stimuli did not provide a context relevant to audition. Our findings attest to the capacity of early auditory cortex to be affected by high-level, top-down, and cross-modal information and indicate that the spatial patterns of activity in auditory cortex reflect contextual/implied auditory information but not visual information per se.

ACKNOWLEDGMENTS

We thank Mark Tomkins for providing the woodpecker footage.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

P.-J.H., J.T.C., and N.K. conception and design of research; P.-J.H. and J.T.C. performed experiments; P.-J.H., J.T.C., and N.K. interpreted results of experiments; P.-J.H. and J.T.C. prepared figures; P.-J.H., J.T.C., and N.K. drafted manuscript; P.-J.H., J.T.C., and N.K. reviewed and revised manuscript; P.-J.H., J.T.C., and N.K. approved final version of manuscript; J.T.C. analyzed data.

REFERENCES


