

Tactile–Auditory Shape Learning Engages the Lateral Occipital Complex

Jung-Kyong Kim and Robert J. Zatorre

Montreal Neurological Institute, McGill University, and Laboratory for Brain, Music, and Sound Research, Montréal, Québec H3A 2B4, Canada

Shape is an object property that inherently exists in vision and touch, and is processed in part by the lateral occipital complex (LOC). Recent studies have shown that shape can be artificially coded by sound using sensory substitution algorithms and learned with behavioral training. This finding offers a unique opportunity to test intermodal generalizability of the LOC beyond the sensory modalities in which shape is naturally perceived. Therefore, we investigated the role of the LOC in processing of shape by examining neural activity associated with learning tactile-shape-coded auditory information. Nine blindfolded sighted people learned the tactile–auditory relationship between raised abstract shapes and their corresponding shape-coded sounds over 5 d of training. Using functional magnetic resonance imaging, subjects were scanned before and after training during a task in which they first listened to a shape-coded sound transformation, then touched an embossed shape, and responded whether or not the tactile stimulus matched the auditory stimulus in terms of shape. We found that behavioral scores improved after training and that the LOC was commonly activated during the auditory and tactile conditions both before and after training. However, no significant training-related change was observed in magnitude or size of LOC activity; rather, the auditory cortex and LOC showed strengthened functional connectivity after training. These findings suggest that the LOC is available to different sensory systems for shape processing and that auditory–tactile sensory substitution training leads to neural changes allowing more direct or efficient access to this site by the auditory system.

Introduction

Our ability to perceive form is a crucial part of object recognition. The lateral occipital complex (LOC) is a higher-order region of the ventral visual pathway and is traditionally known as a shape processing center in human vision. The LOC is associated with responses to complex form regardless of visual properties such as size and position (Malach et al., 1995; Grill-Spector et al., 1999), luminance and motion (Grill-Spector et al., 1998), viewing condition (Kourtzi and Kanwisher, 2000), or contour and depth (Kourtzi and Kanwisher, 2001), therefore manifesting cue invariance.

Given the intramodal robustness of LOC, we may ask whether LOC activity also generalizes across nonvisual modalities. Studies examining tactile object recognition have shown that tasks related to shape discrimination by touch elicit activity in a subregion of LOC (Amedi et al., 2001, 2002; James et al., 2002; Pietrini et al., 2004; Prather et al., 2004; Peltier et al., 2007). Previous studies using sensory substitution algorithms have also demonstrated that shape information can be artificially coded by auditory input and extracted by blind (Amedi et al., 2007) and sighted

(Capelle et al., 1998; Arno et al., 1999; Amedi et al., 2007; Auvray et al., 2007; Kim and Zatorre, 2008) people. These studies showed that frequency- or frequency/time-based coding was effective in delivering form information and could be applied to convey new shapes (Capelle et al., 1998; Arno et al., 1999; Kim and Zatorre, 2008, 2010). Learning of tactile shape via auditory substitution also transfers to vision even when it is untrained (Kim and Zatorre, 2010), indicating shape representation at an abstract level. Is LOC then engaged in processing such auditory shape information? An answer to this question would address the extent of the intermodal generalizability of the LOC because shape is not inherent to hearing, and thus finding sound-induced LOC activity would demonstrate the highly robust multimodal nature of this region. Amedi et al. (2007) have demonstrated that, after extensive auditory-shape training based on the frequency–time conversion method (Meijer, 1992), sighted and blind individuals engaged the LOC in shape-based object recognition. However, since pretraining LOC activity was not examined, the neural mechanisms behind learning-induced cross-modal plasticity resulting in LOC activity are still unknown.

Here, we investigated the effect of auditory-shape learning on the recruitment of LOC by comparing the activity of this region before and after training. We predicted that LOC activity would emerge with sensory substitution training, reflecting shape processing within this area. We also addressed the issue of generalizability of LOC activity, which has not been done before, by examining the extent to which LOC could be accessed after auditory learning with new stimuli and by examining generalization to an untrained modality. An additional unknown is how the functional inputs to LOC may change with learning; we tested

Received June 29, 2010; revised Feb. 24, 2011; accepted March 8, 2011.

Author contributions: J.K.K. and R.Z. designed research; J.K.K. performed research; J.K.K. contributed unpublished reagents/analytic tools; J.K.K. and R.Z. analyzed data; J.K.K. and R.Z. wrote the paper.

This work was supported by operating funds from the Canadian Institutes of Health Research to R.Z. and by a graduate scholarship from the Natural Sciences and Engineering Research Council of Canada to J.K.K. We thank Dr. Peter Meijer for his helpful comments, and Patrick Bermudez, Nick Foster, Marc Bouffard, and the staff of the McConnell Brain Imaging Centre for their technical assistance.

Correspondence should be addressed to Jung-Kyong Kim, Montreal Neurological Institute, 3801 University, Room 276, Montréal, Québec H3A 2B4, Canada. E-mail: jkim@ego.psych.mcgill.ca.

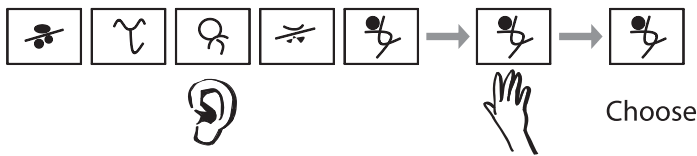
DOI:10.1523/JNEUROSCI.3399-10.2011

Copyright © 2011 the authors 0270-6474/11/317848-09\$15.00/0

A Examples of shapes



B Testing



C Training

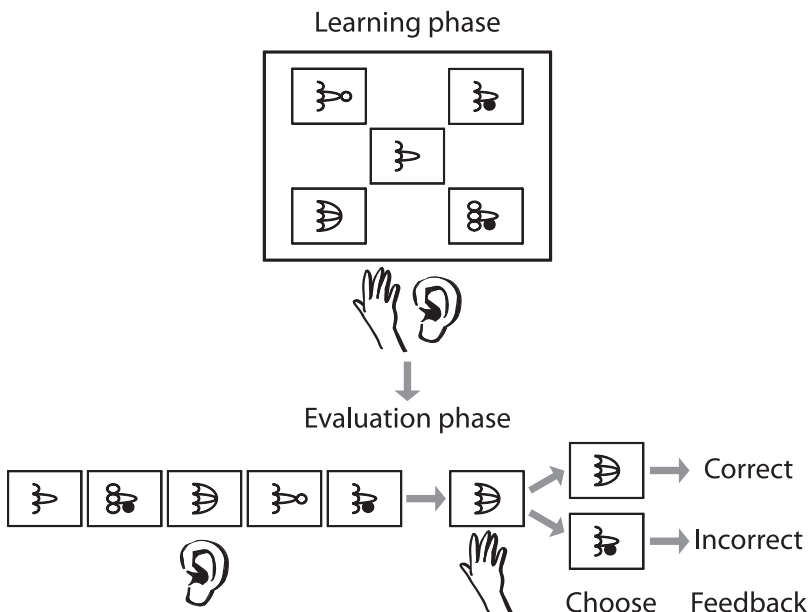


Figure 1. Behavioral paradigm. *A*, Examples of abstract shapes used in the study. *B*, Behavioral testing that took place in the laboratory was based on a forced-choice task where subjects heard five soundscapes and touched one shape, and chose the soundscape that matched the tactile shape. *C*, Training involved a learning phase where subjects explored five shapes and heard their corresponding soundscapes and an evaluation phase where they touched a shape and chose its corresponding soundscape from a set of five. Feedback was provided for each evaluation trial.

this question by examining functional connectivity of auditory and LOC regions, predicting that the pattern of covariation between these regions would change as a function of training, reflecting changes in neural information processing between these regions.

Materials and Methods

Subjects. Nine healthy, sighted individuals (five females; mean age, 24.11 years; SD, 3.55 years) gave consent to participate in this study, which was approved by the Montreal Neurological Institute Ethics Review Board. All participants were right-handed and had normal vision and hearing.

Study timeline. The study took place on 9 consecutive days. Days 1 and 2 involved pretraining testing in our laboratory and in the MRI machine, respectively. Days 3–7 consisted of five training sessions, each a separate day and lasting ~2 h. Days 8 and 9 involved posttraining testing in the laboratory and in the MRI machine, respectively.

Stimuli and apparatus. Tactile and auditory stimuli used for laboratory and fMRI testing and training were generated in the same way as in our previous study of auditory–tactile shape learning (Kim and Zatorre, 2010). The tactile stimuli consisted of abstract shapes that were embossed on paperboards (Fig. 1*A*). Each paperboard used for training consisted of

five tactile figures that were arranged in three rows, with two positioned in the top, one in the middle, and two in the bottom row (Fig. 1*C*), whereas the one used for testing consisted of one tactile figure in the center. Lines comprising the figures were 1.0 mm thick and filled-in elements were rendered as a raised surface with a coarse texture. Each figure was placed in the center of a raised rectangular frame consisting of lines 0.6 mm thick. Visual stimuli used for fMRI testing (in the posttraining session only) were the visual version of the type of abstract shapes used in the tactile condition. Auditory stimuli were generated from two-dimensional images of the shapes used in the study using Meijer's (1992) image-to-sound conversion algorithm. The vertical dimension of the image was coded into sound frequencies ranging from 500 to 5000 Hz, with higher spatial position corresponding to higher frequency, and the left-to-right dimension of the image was coded into 2 s left-to-right panning of sound (Kim and Zatorre, 2008, 2010). From here on, these shape-coded auditory transformations are referred to as "soundscapes."

Tactile, visual, and auditory stimuli used in the fMRI control conditions were created based on the stimuli used in the fMRI active conditions. The tactile and visual control stimuli were the tactile and visual versions, respectively, of 10×10 scrambled blocks of the shapes that were used in the active conditions. As with the normal shapes, these scrambled shapes were embedded in the same-dimensional rectangular frames. Auditory stimuli used for the fMRI auditory control condition were generated using 2-s-long white noise that was matched in terms of frequency and amplitude of the soundscapes. The noise was band-pass filtered so that only the frequencies in the range between the lowest and highest frequencies (500–5000 Hz) of the soundscapes were retained. The amplitude envelope of each soundscape was then applied to the filtered noise, creating a control stimulus corresponding to each soundscape.

The tactile stimuli during fMRI testing were delivered via a custom-made adjustable tray where a stack of shape-containing paperboards were placed. The auditory stimuli were delivered binaurally via MR-compatible earphones (Sensimetrics Model S14). The earphones were calibrated for loudness before starting using a standard tone to ensure that all subjects were tested at the same level.

Behavioral testing. Before the pretraining test, subjects underwent a 15 min familiarization session during which the rules of image-to-sound conversions were briefly explained and example soundscapes representing simple geometric shapes such as line, circle, square, and triangle were presented. The purpose of this session was to familiarize subjects with the procedure of the pretraining test that followed, but not with the types of sounds they were going to be presented with throughout the study. Thus, the soundscapes that were demonstrated during this session were much simpler, consisting of lines, circles, and geometric forms, and greatly different from those presented throughout later testing and training; it was thus unlikely to influence the subsequent pretraining performance.

Following the short familiarization session, subjects were evaluated on their ability to interpret shape-coded auditory information for complex shapes before and after training (on Days 1 and 8, respectively). Both the pretraining and posttraining evaluations consisted of 24 trials of a forced-choice task. Subjects were presented with a shape that they explored with

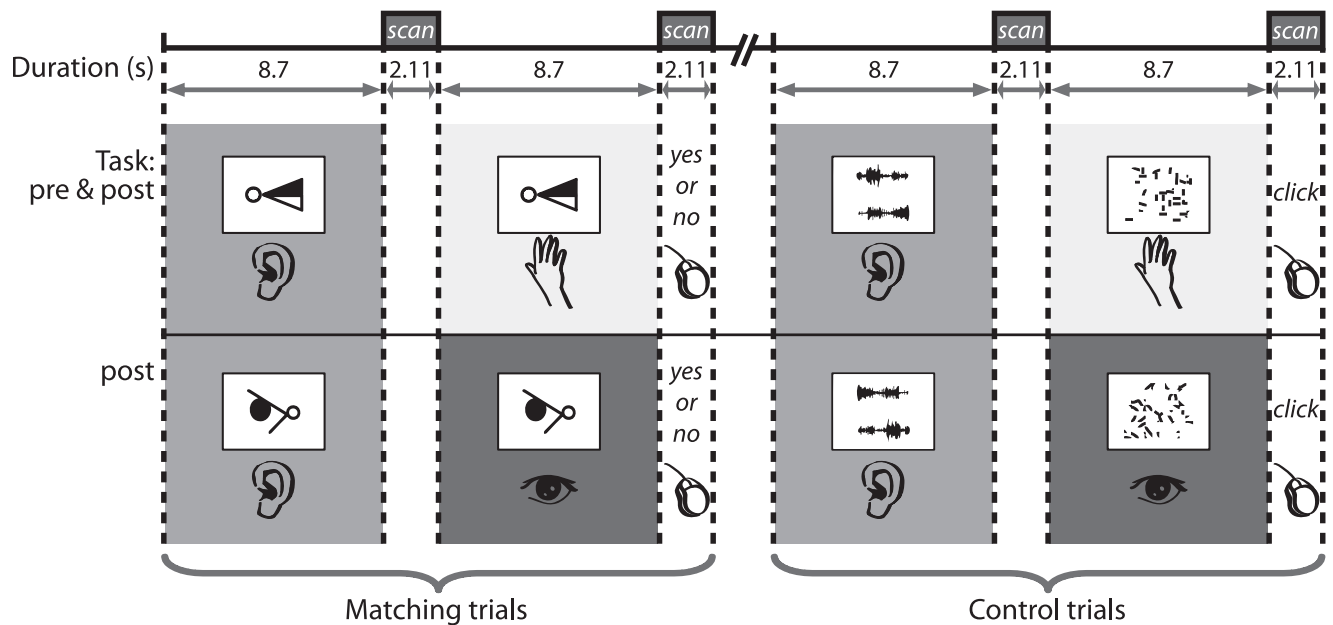


Figure 2. Pretraining and posttraining fMRI sessions involved matching and control tasks that were presented in paired trials. In the first part of each matching trial, blindfolded subjects listened to a soundscape repeated four times, and in the second part of the trial, touched a tactile shape that either matched or did not match the shape corresponding to the soundscape heard in the first trial. Subjects responded “yes” or “no” with a key press during the scan acquisition following the second part of the trial. A control task involved the same paired-trial procedure during which subjects first listened to an auditory stimulus that was matched for acoustic characteristics to the soundscapes, and then touched a scrambled tactile shape; they made a key press at the end but no yes/no response was required. The posttraining session also contained an additional condition where subjects saw a visual shape in the second half of the matching trial and responded whether or not the soundscape matched the visual shape. The corresponding control condition included a visual scrambled shape.

their right hand, and five soundscapes from which the sound corresponding to the shape touched was to be chosen (Fig. 1B). The tactile and auditory shapes that were tested posttraining did not overlap with the shapes used for the pretraining evaluation or training, and therefore were new to the subjects.

Training procedure. Training involved learning the relationship between shape and sound by tactually exploring raised shapes and simultaneously listening to their soundscapes. In each training session, subjects were blindfolded and presented with eight sets of five tactile–auditory pairs of stimuli. Each set represented five shapes that were globally similar but locally different (Fig. 1C) (Kim and Zatorre, 2010). Subjects explored each shape with their right hand for 20 s while its corresponding soundscape repeated 10 times. They repeated this procedure four more times so that a total of 50 auditory repetitions for each shape were presented. Following this learning phase, subjects were given three forced-choice trials based on the stimuli they had just learned, where they touched one shape and chose the correct soundscape among five (Fig. 1C). Feedback was provided for each trial (for more training details, see Kim and Zatorre, 2010).

fMRI testing. Blindfolded subjects underwent scanning in a 3-T Siemens Trio with a 32-channel head coil. A $1 \times 1 \times 1 \text{ mm}^3$ high resolution T1-weighted anatomical scan was acquired before functional scanning. Pretraining functional scans consisted of three runs of T2*-weighted gradient echo-planar images; posttraining scans were identical but contained two additional runs. Sparse sampling was used where each volume of 40 slices was associated with a single trial lasting 8.7 s (repetition time, 10.81 s; echo time, 30 ms; voxel size, $3.4 \times 3.4 \times 3.4 \text{ mm}^3$; matrix size, 64×64); this procedure avoids contamination of the auditory cortex blood oxygenation level-dependent (BOLD) response with the scanner noise (Belin et al., 1999) and also affords time for the tactile scanning of the stimulus while avoiding any associated movement artifact.

Each pretraining functional scan consisted of four blocks of 14 trials alternating between shape and control conditions. In the shape condition, subjects were presented with paired trials. In the first part of each trial, they listened to a soundscape repeated four times (listen); in the second part, the subjects touched a shape with their right hand (touch). The task was to judge whether the sound heard in the first part of the trial

matched the shape touched in the second part. Subjects responded “yes” or “no” with their left hand on a two-button pad during the scan acquisition following the touch trial (Fig. 2). The control condition followed a similar paired-trial procedure. Subjects first listened passively to a frequency- and amplitude-matched noise burst that was repeated four times and then touched a scrambled shape with their right hand (Fig. 2). During the scan acquisition following the touch trial, they pressed a button with their index finger. Papers containing the tactile stimuli were stacked on a tray in the order of the trials. Each paper corresponded to one trial, and as each trial ended, it was removed by the experimenter during the following scan acquisition. During the listen trials, the paper was blank and subjects kept their hand still. A total of 14 silence trials were distributed in blocks throughout each run (for examples of sounds used in this study, visit <http://www.zlab.mcgill.ca/JNeurosci2011/>).

The posttraining scanning session was comprised of five functional runs. The first three runs followed exactly the same procedure as the pretraining runs. The last two runs involved removing the blindfold and presenting visual and auditory stimuli (Fig. 2). Each consisted of six blocks of trials alternating between shape and control conditions (with each condition presenting a total of 40 trials), and the procedure was same as that of the previous runs except that shapes (for the shape condition) and scrambled shapes (for the control condition) were presented visually through a mirror mounted on the head coil. This was the first time that subjects ever experienced visual shapes. During the listen trials of both shape and control conditions, subjects fixated their eyes on the cross located at the center of a blank screen. The shapes used for all posttraining testing were new, thus the posttraining shapes presented tactually and visually did not overlap with any previous stimuli or each other.

Note that subjects were not completely naive in both the pretraining laboratory and fMRI sessions, as the image-to-sound mapping rules were explained to them at the very beginning of the study. Our previous studies have shown that the mere explicit knowledge of the image–sound mapping relationship initially provides enough information to perform with some degree of accuracy in the shape–sound matching task (Kim and Zatorre, 2008, 2010). Thus, the performance based on this explicit knowledge was considered as the baseline performance and learning was

regarded as the improvement from this baseline after training was given. The term “pretraining” (for both the laboratory and fMRI sessions) refers to the experimental condition in which no training with feedback had been provided to allow the correct application of the image–sound mapping rules.

fMRI analyses. Images within each run were aligned to the fourth frame, motion corrected, and smoothed with an 8 mm full-width at half-maximum isotropic Gaussian kernel. Both functional and anatomical images were linearly registered to the symmetric ICBM 152 template with a 12-parameter linear transformation, and subsequently nonlinearly fit to the same template with an 8 mm node-spacing between vectors in the deformation grid (Collins et al., 1994). Statistical data analysis was based on a general linear model using fMRISTAT (Worsley et al., 2002) (www.math.mcgill.ca/keith/fmristat). The t threshold for search within the predefined LOC region was set at $t = 4.2$ (search volume = 34,576 mm³, number of voxels = 4322). Statistical significance for search within the rest of the brain was set at $t = 4.9$ ($p < 0.05$, corrected).

Two t statistic subtraction images were generated for each of the pretraining and posttraining sessions: shape–listen minus control–listen and shape–touch minus control–touch. Learning effect was examined by further contrasting the shape–listen minus control–listen subtraction of the pretraining and posttraining sessions with one another. Two additional contrasts were generated for the posttraining visual–auditory condition: shape–listen minus control–listen and shape–see minus control–see. The first contrast was denoted as shape–listen minus control–listen (V) to differentiate from the listening–associated contrast from the tactile–auditory condition.

The LOC region was determined using the contrast comparing the activity associated with seeing abstract versus scrambled shapes in the posttraining visual–auditory runs (from the last two functional runs). That portion of the lateral occipital cortex that met the corrected t threshold ($t = 4.9$) was defined as a search region for LOC activity associated with the shape task in the listen and touch trials. This LOC definition was also applied in the voxel selection for voxel-of-interest (VOI) analyses. Note that this LOC definition results in an independent identification of relevant voxels since the data from the visual runs were obtained in a separate run from the other data.

Conjunction analyses were performed to examine common regions of activity during the shape task in all three (auditory, tactile, visual) modalities. Using the contrasts comparing shape versus control for the three modalities, the minimum t value of 3.17 from each of the three contrasts was chosen for each voxel ($p < 0.001$, uncorrected) and those which survived this threshold were mapped out. This analysis was performed separately for the pretraining and posttraining sessions. Since there was no visual condition in the pretraining session, the posttraining contrast associated with the visual condition was used for the pretraining conjunction analysis.

Using the auditory trials involving the shape–listening task, functional connectivity analyses were performed to examine cortical regions whose BOLD activities covaried with that of a prechosen seed over time. We applied the model $Y_{ij} = X_i\beta_{1j} + R_i\beta_{2j} + \epsilon_{ij}$, where, for a given i th frame for a j th voxel, Y is the BOLD signal value, X represents the dummy variable coding the experimental condition, β the parameter estimate, R the BOLD signal value from the seed voxel, and ϵ the error term. We mapped out the connectivity of the seed region with the rest of the brain by testing the significance of the parameter estimate of R on a voxel-by-voxel basis. In the same way that simple subtraction analyses were performed, this analysis was first performed at the single subject level, and the results then were combined across subjects by pooling the effect of R at the multisubject level. The final result was a statistical map showing the t value of the group effect of R at each voxel. The seed region was selected to be the voxel yielding the highest activity in the auditory cortex based on the shape–listen minus control–listen contrast in the entire group. As the seed voxel was chosen for each hemisphere, two analyses were performed for each subject for each of the pretraining and posttraining sessions.

To test directly training-induced changes in the relationship between the auditory cortex and LOC, BOLD signal time series were extracted at the auditory cortex and LOC from the auditory trials involving the

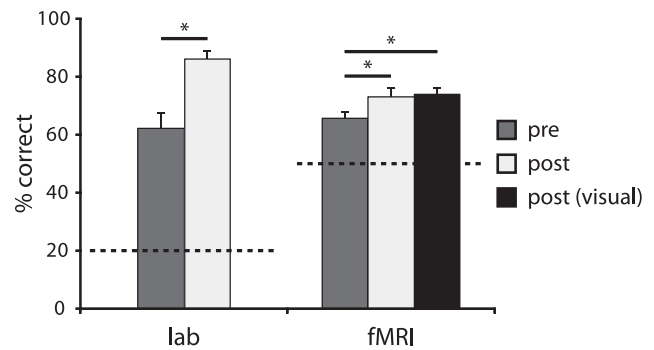


Figure 3. Behavioral results from the auditory–tactile shape-matching task (percentage correct shown). Performance in both the laboratory and the scanner was above chance before training (dark gray) and improved significantly after training (light gray). Performance on the auditory–visual task (black) was equivalent to that on the auditory–tactile task at posttraining. Dotted lines indicate chance level of performance. * $p < 0.05$; significance based on paired-samples t tests.

shape–listening task. The voxels of interest were selected using those that yielded the highest auditory and LOC peaks from the shape–listen minus control–listen contrast of each session. The time series were then used for VOI analyses, which examined change in the BOLD signal correlation between the auditory region and LOC as a function of training. Analysis of covariance (ANCOVA) was performed using the BOLD signal value from the chosen LOC voxel as the dependent variable, session (fixed; pretraining and posttraining) and subject (random; nine subjects) as categorical variables, and BOLD signal value from the auditory seed region as a continuous variable. We used the ANCOVA model that tests heterogeneous (or nonparallel) slopes across different levels of each of the categorical variables (Johnson and Neyman, 1936; Kowalski et al., 1994a,b). We tested the significance of three-way interaction (auditory BOLD \times subject \times session) to examine the change of correlation in BOLD activity between the auditory and LOC regions as a function of subject and session. A significant three-way interaction represents a change in the interaction between two variables across different levels of the third variable, and therefore would suggest that individually varying correlations between the auditory and LOC regions (i.e., two-way interactions between auditory BOLD and subjects) change across sessions.

Results

Behavioral performance

Performance on all the tasks was measured in terms of the percentage of correct responses. Scores obtained from the tactile–auditory matching task indicated that subjects performed above chance pretraining (compared with the 20% chance level, $t = 7.43$, $p < 0.001$), and showed significant improvement after training ($t = 6.02$, $p < 0.001$). Scores obtained during the fMRI scans showed the same behavioral pattern where subjects performed above chance in the pretraining session (compared with the 50% chance level; $t = 7.18$, $p < 0.001$) and showed significant improvement ($t = 2.32$, $p < 0.05$) after training. The posttraining visual–auditory test showed that subjects also performed better than on the pretraining tactile–auditory task ($t = 2.98$, $p < 0.05$) and equally well on this as on the posttraining tactile–auditory task, as no score difference was found between the two conditions (Fig. 3).

fMRI results

Regions of activity during auditory–shape processing

Our main hypothesis of interest was that the LOC would be associated with shape processing of shape-coded auditory input. The shape–listen minus control–listen contrast based on the auditory–tactile task showed activity in this region both before (x , y , $z = -50$, -56 , 4 ; $t = 4.45$) and after (x , y , $z = -48$, -54 , 2 and 58 ,

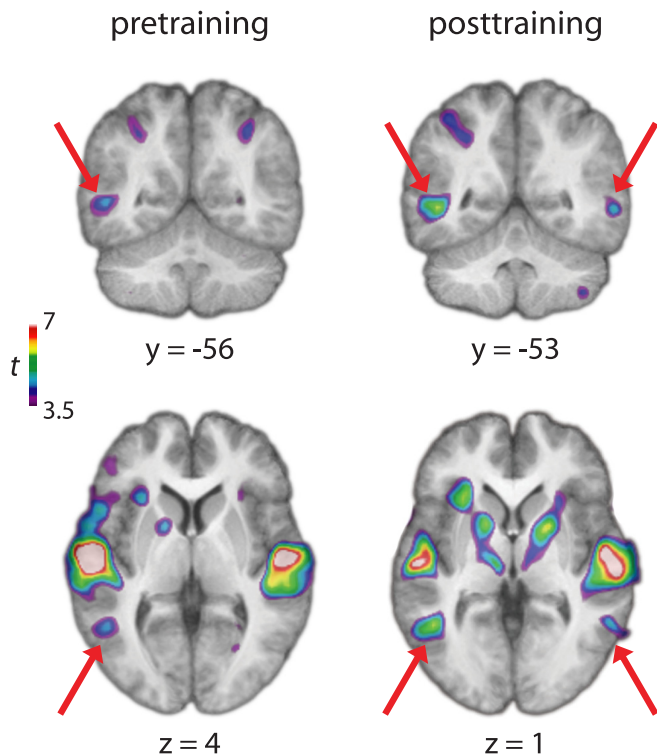


Figure 4. The shape-listen minus control-listen contrast yielded activity in the LOC both before (left) and after (right) training in addition to expected activity in auditory cortical areas. Red arrows indicate LOC activity on the coronal and transverse slices.

$-52, 0; t = 5.88, 4.78$, respectively) training (Fig. 4). When the pretraining contrast was compared with the posttraining contrast, there was a weak trend for greater activity posttraining compared with pretraining in the LOC, but no statistical difference was found between the sessions in this region ($x, y, z = -46, -58, 0$ and $62, -60, 2; t = 2.78, 2.27$).

To test the possibility that this lack of session difference in LOC activity resulted from the use of a conservative localizer (i.e., the entire area of the LOC associated with visual shape recognition), we performed an additional analysis by further limiting the region of interest (ROI) to the subregion of LOC that has been found to be associated with visual and tactile object recognition (Amedi et al., 2001). For each fMRI session, the lateral occipital tactile–visual (LOtv) region was defined as the common area of activity based on the shape minus control contrast for the visual and tactile conditions by choosing the voxels in the LOC exceeding the t value of 3.17 (resulting in 555 and 431 pretraining voxels and 618 and 548 posttraining voxels in the left and right hemispheres, respectively). The average percentage signal change in this ROI was calculated using the shape-listen minus control-listen contrast in each hemisphere for each subject. The average values were compared between the sessions for each hemisphere using the paired-samples t test. Results showed no significant difference between the two sessions for either hemisphere, although the left hemisphere showed a slight trend for a greater response for the posttraining than pretraining sessions ($t = 2.20, p = 0.06$).

The shape-listen minus control-listen contrast also yielded activity in other regions that were likely to be related to the task both before and after training. The areas included the auditory cortex ($x_{pre}, y_{pre}, z_{pre} = -62, -14, 4$ and $54, -10, 0; t = 9.93$ and 9.46 , respectively; $x_{post}, y_{post}, z_{post} = -58, -16, 6$ and $56, -8, 0$;

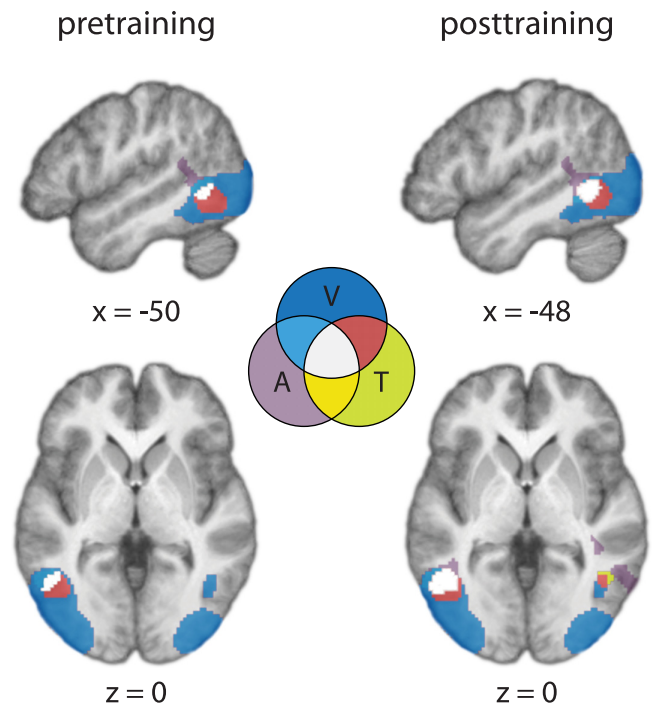


Figure 5. The shape minus control contrast involving each of the auditory and tactile modalities commonly yielded a subregion of the LOC as an area of activity both before and after training. Top, Left hemisphere sagittal sections; bottom, horizontal sections. The LOC was defined as the occipitotemporal area of visual activity resulting from the seeing-shape minus scrambled-shape (data acquired only posttraining). Regions of activity involving one, two, or all three modalities were coded differentially by different colors as indicated in a Venn diagram. A, Auditory; T, tactile; V, visual.

$t = 9.20$ and 7.63 , respectively), intraparietal sulcus (IPS; $x_{pre}, y_{pre}, z_{pre} = -38 -46 42$ and $40 -44 48, t = 5.15, 4.89$; $x_{post}, y_{post}, z_{post} = 42, -42, 46$ and $-38, -46, 48; t = 5.14$ and 4.89 , respectively), and prefrontal cortex (PFC; $x_{pre}, y_{pre}, z_{pre} = -46, -2, 44$ and $42, -2, 36; t = 7.63$ and 6.50 , respectively; $x_{post}, y_{post}, z_{post} = 44, 8, 30$ and $-52, 0, 44; t = 7.16$ and 6.92 , respectively). No session difference was found for activity in any of these regions in the direct contrast (for the whole-brain presentation of these contrast results, see supplemental Fig. 1, available at www.jneurosci.org as supplemental material).

As expected, the control-listen minus silence contrast yielded the highest peaks in the auditory cortex both before and after training ($x_{pre}, y_{pre}, z_{pre} = 44, -26, 8$ and $-40, -30, 10; t = 10.02$ and 8.53 , respectively; $x_{post}, y_{post}, z_{post} = 50, -22, 8$ and $-48, -24, 8; t = 9.04$ and 9.00 , respectively). However, no significant peaks were found in the LOC. There was also no difference in this contrast when pretraining was compared with posttraining datasets.

Conjunction analyses show LOC as a common area of activation during auditory, tactile, and visual shape tasks

The shape minus control contrast involving each of the auditory, tactile, and visual modalities was used in a conjunction analysis to examine activity in all three modalities. Since there was no visual testing in the pretraining session, the posttraining (shape-see minus control-see) contrast was used for the pretraining analysis. The analysis showed the LOC as a common region of activity during the shape task involving all the modalities both before ($x, y, z = -50, -56, 0; t = 4.08$, just below the a priori t threshold) and after ($x, y, z = -48, -54, 0; t = 5.70$) training (Fig. 5; for the whole-brain presentation, see supplemental Fig. 2, available

at www.jneurosci.org as supplemental material). To examine whether the trimodal activity was due to spatial smoothing and/or group averaging, we performed the conjunctive analysis on unsmoothed data in each individual and searched for voxels in the LOtv region that responded to all three modalities using the t threshold of 3.17. Results showed that seven of nine subjects before training and eight of nine subjects after training showed voxels that were overlapped by all three modalities (supplemental Fig. 3, available at www.jneurosci.org as supplemental material). Furthermore, the voxels in the few subjects who did not show the trimodal activity responded to at least two modalities (mostly visual–tactile). Thus, these results support the conclusion that the overlap in activity by the three modalities was not an artifact of averaging or spatial smoothing.

We also examined the possibility that the learning effect might have manifested itself as a change in the size of LOC activity (as opposed to magnitude) by quantifying the trimodal overlap in each hemisphere and examining the change in this variable. Using the same LOtv search regions as before, we calculated the percentage of trimodal voxels in each subject for each of the two sessions. Percentage (i.e., proportion of trimodal voxels in the search volume) instead of absolute number of trimodal voxels was used to measure the size of activation to take into account the change in the volume of the LOtv after training (from 555 and 431 pretraining voxels to 618 and 548 posttraining voxels in the left and right hemispheres, respectively). Paired-samples t tests comparing the pretraining and posttraining sessions showed no significant difference for either hemisphere in percentage of voxels active to all three modalities, although the comparison on the left hemisphere yielded a trend for a greater percentage of LOtv voxels after training ($t = 2.23$, $p = 0.06$).

Training is associated with strengthened functional connectivity between auditory cortex and LOC during auditory-shape processing

Although no significant difference was observed in the overall LOC activity before and after training, it is possible that change is shown in the relationship between the LOC and the original source of the shape input (i.e., auditory cortex). Therefore, seed regions were chosen for functional connectivity analysis based on the site of maximum auditory activity from the shape-listen minus control-listen contrast ($x_{\text{pre}}, y_{\text{pre}}, z_{\text{pre}} = -62, -14, 4$ and $54, -10, 0$; $x_{\text{post}}, y_{\text{post}}, z_{\text{post}} = -58, -16, 6$ and $56, -8, 0$). The posttraining connectivity analysis showed that the BOLD signal of the left auditory seed was positively correlated with that of the left LOC during the shape-listening trials ($x, y, z = -54, -56, 0$; $t = 5.15$) (Fig. 6A). Pearson correlation of the BOLD signal between these two voxels for individual subjects ranged from slightly to highly positive ($r = 0.26, 0.06, 0.24, 0.41, 0.23, 0.06, 0.20, 0.17$, and 0.22). Positive connectivity was also shown between the auditory seed and the right LOC at subthreshold level ($x, y, z = 58, -52, 0$; $t = 3.97$) (Fig. 6A). However, no such correlation was present pretraining, even at subthreshold levels. A similar but weaker pattern was found for the right auditory seed, which was positively correlated with the right LOC, albeit at subthreshold level, posttraining ($x, y, z = 60, -46, -6$; $t = 3.18$) (for whole-brain presentation of these results, see supplemental Fig. 4, available at www.jneurosci.org as supplemental material).

To test whether there was a statistically significant change in the functional connectivity between the auditory and LOC regions when comparing pretraining with posttraining conditions, we extracted, for each subject, the BOLD signal time series at the left pretraining and posttraining auditory seeds ($x_{\text{pre}}, y_{\text{pre}}, z_{\text{pre}} =$

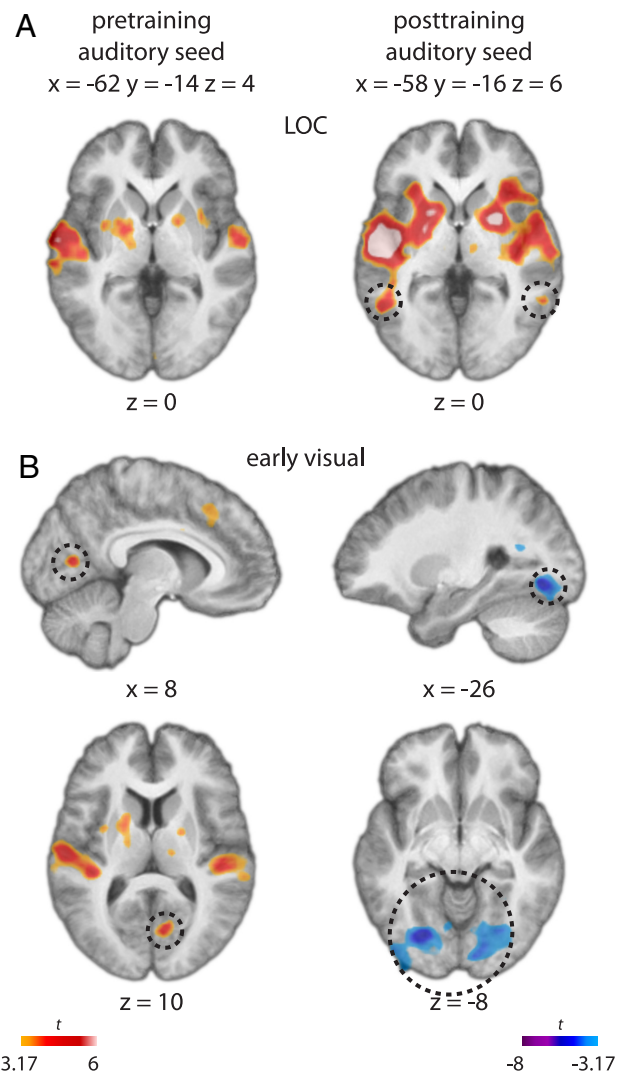


Figure 6. *A*, Results of the functional connectivity analysis shown in a t map. The analysis showed that BOLD activity at an auditory cortex seed region, selected as the peak from the shape-listen minus control-listen contrast ($x_{\text{pre}}, y_{\text{pre}}, z_{\text{pre}} = -62, -14, 4$; $x_{\text{post}}, y_{\text{post}}, z_{\text{post}} = -58, -16, 6$), was correlated with that of the LOC (dotted circles) after but not before training (activity shown at $z = 0$). *B*, The same analysis also yielded positive correlation with the early visual cortex before training (dotted circles at $x = 8, z = 10$) but negative correlation in more ventral portions of early visual cortex after training (dotted circles at $x = -26, z = -8$).

$-62, -14, 4$; $x_{\text{post}}, y_{\text{post}}, z_{\text{post}} = -58, -16, 6$) and at the left LOC with the highest peaks from the pretraining and posttraining conjunctive analyses ($x_{\text{pre}}, y_{\text{pre}}, z_{\text{pre}} = -50, -56, 0$; $x_{\text{post}}, y_{\text{post}}, z_{\text{post}} = -48, -54, 0$). The LOC BOLD values were then assigned as the dependent variable and the auditory BOLD signal, subject, and session as predictor variables in an ANCOVA. The three-way interaction (auditory BOLD \times subject \times session) was found to be significant ($F_{(18,720)} = 1.77$, $p = 0.02$), indicating a change in the linear relationship between the auditory and LOC BOLD signal as a function of subject and session. Based on the functional connectivity results, this change in the slope indicates a positively strengthened correlation between the two regions across subjects after training.

Training is also associated with changes in auditory connectivity with early visual regions

The functional connectivity analyses based on the same auditory seed regions also showed a change in the BOLD correlation with

other visual regions. The left auditory and early visual regions were shown to be positively related before training ($x_{\text{visual}}, y_{\text{visual}}, z_{\text{visual}} = 8, -72, 10; t = 4.51$) but negatively correlated after training ($x_{\text{visual}}, y_{\text{visual}}, z_{\text{visual}} = -26, -76, -8; t = -5.54$) (Fig. 6B). According to cytoarchitectonic probability maps (Amunts et al., 2000), the visual peak of connectivity with auditory cortex shown pretraining was in or close to V1 and posttraining in early extrastriate cortex (approximately in the region of V2). The right auditory and early visual regions showed a weaker but similar pattern; the two regions were not related before training but were negatively correlated after training ($x_{\text{visual}}, y_{\text{visual}}, z_{\text{visual}} = 22, -82, 4; t = -3.40$). Compared with the strengthened functional connectivity between the auditory and LOC regions after training, these results show the opposite pattern where the early visual areas show decreasing activity with increasing auditory activity.

Patterns of activity in the auditory-visual task: further evidence for LOC as independent shape region

The pattern of activity during the auditory–visual task, performed only at posttraining, resembled the pattern found during the auditory–tactile task. The shape-listen minus control-listen (V) contrast showed bilateral activity in the LOC at subthreshold level ($x, y, z = -54, -58, 0$ and $62, -56, -4; t = 4.02, 3.67$, respectively) and other regions such as auditory ($x, y, z = 54, -12, 2$ and $-56, -16, 6; t = 9.65$ and 8.46), IPS ($x, y, z = -40, -38, 42$ and $38, -42, 40; t = 5.16$ and 4.85), and PFC ($x, y, z = -22, -4, 56; t = 4.99$). The only difference was negative activity in other areas of the visual ventral stream ($x, y, z = -36, -88, 0$ and $38, -70, -4; t = -6.08$ and -5.51). Functional connectivity analysis was performed based on the auditory regions of the maximum activity from this contrast ($x, y, z = 54, -12, 2$ and $-56, -16, 6$). We found that the BOLD signal of the left auditory seed was positively correlated with that of both the left and right LOC at subthreshold level ($x, y, z = -56, -60, 2; t = 2.74; x, y, z = 58, -54, 0; t = 3.63$). No such correlation was present for the right auditory seed.

Discussion

Behavioral performance

The pretraining performance was well above chance level, indicating subjects were able to extract shape information to some extent just by explicitly knowing the image-to-sound conversion rules. Our previous studies also have consistently found better-than-chance performance before training, whether subjects are tested in auditory–visual (Kim and Zatorre, 2008) or auditory–tactile modalities (Kim and Zatorre, 2010). Also consistent with our previous findings, the ability to match auditory input with tactile shape improved significantly after training, showing that subjects could learn to use the auditory input to interpret shapes with better accuracy after training. Importantly, the posttraining testing involved new auditory and tactile stimuli; consequently, the improved performance indicates generalized learning of image-to-sound mapping rules, which subjects were able to apply to new shapes. Furthermore, the score from the auditory–visual transfer task in the posttraining session was equivalent to that from the auditory–tactile task, indicating that the rule-learning was not only generalized to new stimuli, but also to the untrained visual modality. These behavioral findings confirm that auditory input can indeed be used to convey shape information and, furthermore, the complete transfer to the untrained modality suggests that shape can be represented at an abstract level.

LOC is involved in shape processing via auditory input

Our main question was whether the LOC is engaged in processing shape information that has entered via the auditory pathway. Our results showed that for both pretraining and posttraining sessions, part of the LOC was involved in processing of sound-coded shapes. No learning effect was found in the magnitude or size of LOC activity, as no significant session difference was found on these measures. However, there was a slight but nonsignificant posttraining increase in the magnitude and size of LOC activity in the left hemisphere, suggesting that this trend could lead to a significant difference if more extensive training were given. Further testing is necessary to address this possibility.

It was somewhat surprising to find LOC activity even before training, given that subjects were only given verbal explanations about the image-to-sound conversion rules and a few example soundscapes representing simple shapes. Nonetheless, the pretraining performance was above chance and the LOC activity is probably related to this level of performance. The fact that the LOC can be accessed even in the absence of feedback-based training suggests that activity in this region may not solely be a result of functional plasticity during sensory substitution learning (Amedi et al., 2007); rather, the present evidence indicates that the LOC can be readily accessed by auditory input.

Furthermore, our conjunction results (Fig. 5) show that part of the LOC is commonly active in response to shape presented in all three (visual, tactile, and auditory) modalities. These results held true even in the individual subject analysis with unsmoothed data (supplemental Fig. 3, available at www.jneurosci.org as supplemental material). This finding of conjunction is consistent with the literature showing that the LOC is responsive to shape in vision (Malach et al., 1995; Grill-Spector et al., 2001; Vinberg and Grill-Spector, 2008) and touch (Amedi et al., 2001, 2002; Stilla and Sathian, 2008) and our findings extend this concept to all three modalities. In addition to the robust activity to various visual cues (Malach et al., 1995; Grill-Spector et al., 1998, 1999; Kourtzi and Kanwisher, 2000), our findings indicate that LOC activity can be generalized across different sensory modalities, supporting the notion that the LOC is a region in the ventral stream of object recognition that is responsive specifically to shape information at an abstract, amodal level of representation. The present findings thus support the theory of metamodal organization of the brain (Pascual-Leone and Hamilton, 2001), positing that modularity is not determined by the sensory modality per se, but rather by optimal matching between the demand of the input task and the capability of a specific brain region to process the type of afferent information. Our results can be explained by this model in that the required nonvisual shape task involved fine spatial analysis, which the visual cortex is best capable of, and consequently, the auditory input requiring such processing gained access to this region.

Training associated with modulation of connectivity between auditory, LOC, and early visual areas

We reasoned that if the behavioral change associated with learning was not reflected in the average magnitude of LOC activity, changes perhaps occurred in the way that auditory cortex interacts with the LOC. The functional connectivity analysis indeed showed that posttraining activity in the auditory region was positively correlated with that of the LOC, contrasting with the absence of such a pattern before training and indicating strengthened connectivity between the two regions as a result of cross-modal training. This learning-related change in functional connectivity (also shown by the modulated correlation between

the two regions by training as demonstrated by the covariance analyses) is consistent with findings of studies in which learning effects were best characterized by the changes in functional connections among the brain regions involved in the given tasks (McIntosh et al., 1999, 2003; DeGutis and D'Esposito, 2009). Similarly, the fact that the LOC was engaged during the auditory–shape task both before and after training, but connectivity with the auditory region was shown only after training, suggests that the two regions operate relatively independently from each other before training but become part of the same network and work together after training. Thus, with training, the auditory–shape information becomes more directly communicated between the two regions. The behavioral improvement is likely a reflection of this direct and therefore more efficient communication between the auditory and LOC regions. The lack of pretraining connectivity between the two regions despite the pretraining LOC activity indicates that these areas are not temporally coupled, or perhaps that indirect routes are involved in how the auditory–shape information reaches and is processed at the LOC site.

The connectivity analysis also showed that the same auditory region was positively correlated with early visual areas before training, but the correlation became negative after training. The connectivity results indicate a training-related change in the relationship between the auditory and visual regions, as increasing auditory activity was accompanied by increasing activity in the early visual cortex before training but by increasing activity in the LOC and decreasing activity in the early visual cortex after training. Reduced activity in early visual areas (most strongly in V1), accompanied by increased activity in LOC, has been observed during tasks where subjects viewed coherent shape versus randomly arranged local features of the shapes (Murray et al., 2002; Fang et al., 2008). These studies also recorded subjects' ratings of their percept of shape on the basis of whether the stimulus looked like a coherent structure, and showed increasing LOC and decreasing V1 activity with higher rating for perceived coherence. It has been argued that LOC is associated with perception of whole and sends feedback signal to lower visual areas that are associated with perception of parts and not in demand when the task is to focus on the global shape of the stimulus, therefore resulting in the net negative V1 activity (Murray et al., 2004). We believe that this visual mechanism of perceptual grouping might be relevant to our finding of training-induced change in connectivity between the auditory and early visual cortex. We speculate that the change in connectivity between auditory and early visual regions from positive to negative might reflect improved perceptual grouping of the local features of the auditory shape into a coherent structure as a result of training. When subjects perform our experimental task without much experience, the interpretation of global shape is not an immediate process because they have to keep track of the local features of the soundscape played over the 2 s time frame and assemble them to form a global shape. Training presumably makes the analysis of global shape easier and more efficient, as indicated by the improved performance, and might reflect the reduced connectivity between early visual cortex and auditory cortex.

Transfer of learning to vision provides further support for LOC as a robust site for multimodal shape processing

The behavioral data showed that learning acquired in the auditory–tactile modalities could be transferred completely to vision, in agreement with our previous study (Kim and Zatorre, 2010). Consistent with the results from the auditory–tactile task, listening to auditory-encoded shapes compared with listening to con-

trol sounds during the auditory–visual task was associated with LOC activity. Furthermore, there was some evidence that the auditory cortex was functionally connected with the LOC during this transfer task, as it was in the auditory–tactile task. These findings suggest that the transfer to the untrained modality was based on the recruitment of the auditory–LOC network that had already been established after auditory–tactile training. The involvement of the same functional network in the transfer task as in the auditory–tactile task suggests that the auditory shape information could be extracted and applied in vision via the shape representation that is necessarily abstract to be accessed by different modalities. Therefore, this finding further supports the role of the LOC in abstract representation of shape.

Visual imagery?

One issue that is often raised in the vision substitution literature concerns visual imagery as a possible means for the substitution performance (for discussion, see Poirier et al., 2007). This interpretation would suggest that the soundscapes in the auditory shape task of our study were visualized and the sound-evoked LOC recruitment resulted from activity associated with visual imagery. Although we cannot exclude this possibility, we propose several points to argue against it. First, the visual activity observed in our study was limited mainly to the LOC and not other visual regions. This pattern is different from that normally observed during visual imagery that engages earlier visual regions (Kosslyn et al., 1995, 1999). Indeed, in the present study, early visual regions were relatively suppressed in terms of connectivity after training, arguing against their playing an important role. Second, if visual imagery is the dominant mechanism for the substitution performance, imagining shapes via auditory or tactile means should consistently lead to visual recruitment. However, not all vision substitution studies have shown nonvisually evoked activity in the visual cortex of sighted individuals (Arno et al., 2001; Ptito et al., 2005; Collignon et al., 2007; Kupers et al., 2010). Third, it has been demonstrated that blind people using vision substitution systems consistently show occipital recruitment (Arno et al., 2001; Ptito et al., 2005; Amedi et al., 2007; Collignon et al., 2007; Kupers et al., 2010), which could not have been due to visual imagery. These points suggest that visual imagery cannot be the only explanation for vision substitution, as visual cortical recruitment is not a consistently paired phenomenon to substitution performance.

Conclusions

Our results show that the LOC was engaged in processing shape-driven auditory input. The LOC activity itself was not necessarily learning-dependent, as it was present both before and after training. What changed was the relationship between the auditory and LOC regions, which showed strengthened functional connectivity after training. These findings suggest that the LOC is a highly robust region that is available to different sensory systems for shape processing, and that auditory sensory substitution training leads to neural changes, allowing more direct or efficient access to this site by the auditory system.

References

- Amedi A, Malach R, Hendler T, Peled S, Zohary E (2001) Visuo-haptic object-related activation in the ventral visual pathway. *Nat Neurosci* 4:324–330.
- Amedi A, Jacobson G, Hendler T, Malach R, Zohary E (2002) Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cereb Cortex* 12:1202–1212.
- Amedi A, Stern WM, Camprodon JA, Bermpohl F, Merabet L, Rotman S,

- Hemond C, Meijer P, Pascual-Leone A (2007) Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat Neurosci* 10:687–689.
- Amunts K, Malikovic A, Mohlberg H, Schormann T, Zilles K (2000) Brodmann's areas 17 and 18 brought into stereotaxic space: where and how variable? *Neuroimage* 11:66–84.
- Arno P, Capelle C, Wanet-Defalque MC, Catalan-Ahumada M, Veraart C (1999) Auditory coding of visual patterns for the blind. *Perception* 28:1013–1029.
- Arno P, De Volder AG, Vanlierde A, Wanet-Defalque MC, Strel E, Robert A, Sanabria-Bohórquez S, Veraart C (2001) Occipital activation by pattern recognition in the early blind using auditory substitution for vision. *Neuroimage* 13:632–645.
- Auvray M, Hanneton S, O'Regan JK (2007) Learning to perceive with a visuo-auditory substitution system: localisation and object recognition with 'the vOICe'. *Perception* 36:416–430.
- Belin P, Zatorre RJ, Hoge R, Evans AC, Pike B (1999) Event-related fMRI of the auditory cortex. *Neuroimage* 10:417–429.
- Capelle C, Trullemans C, Arno P, Veraart C (1998) A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Trans Biomed Eng* 45:1279–1293.
- Collignon O, Lassonde M, Lepore F, Bastien D, Veraart C (2007) Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cereb Cortex* 17:457–465.
- Collins DL, Neelin P, Peters TM, Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* 18:192–205.
- DeGutis J, D'Esposito M (2009) Network changes in the transition from initial learning to well-practiced visual categorization. *Front Hum Neurosci* 3:44.
- Fang F, Kersten D, Murray SO (2008) Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *J Vis* 8:2.1–2.9.
- Grill-Spector K, Kushnir T, Edelman S, Itzhak Y, Malach R (1998) Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* 21:191–202.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203.
- Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. *Vision Res* 41:1409–1422.
- James TW, Humphrey GK, Gati JS, Servos P, Menon RS, Goodale MA (2002) Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* 40:1706–1714.
- Johnson PO, Neyman J (1936) Tests of certain linear hypotheses and their applications to some educational problems. *Stat Res Memoirs* 1:57–93.
- Kim JK, Zatorre RJ (2008) Generalized learning of visual-to-auditory substitution in sighted individuals. *Brain Res* 1242:263–275.
- Kim JK, Zatorre RJ (2010) Can you hear shapes you touch? *Exp Brain Res* 202:747–754.
- Kosslyn SM, Thompson WL, Kim IJ, Alpert NM (1995) Topographical representations of mental images in primary visual cortex. *Nature* 378:496–498.
- Kosslyn SM, Pascual-Leone A, Felician O, Camposano S, Keenan JP, Thompson WL, Ganis G, Sukel KE, Alpert NM (1999) The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science* 284:167–170.
- Kourtzi Z, Kanwisher N (2000) Cortical regions involved in perceiving object shape. *J Neurosci* 20:3310–3318.
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506–1509.
- Kowalski CJ, Schneiderman ED, Willis SM (1994a) Assessing the effect of a treatment when subjects are growing at different rates. *Int J Biomed Comput* 37:151–159.
- Kowalski CJ, Schneiderman ED, Willis SM (1994b) ANCOVA for nonparallel slopes: the Johnson-Neyman technique. *Int J Biomed Comput* 37:273–286.
- Kupers R, Chebat DR, Madsen KH, Paulson OB, Ptito M (2010) Neural correlates of virtual route recognition in congenital blindness. *Proc Natl Acad Sci U S A* 107:12716–12721.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135–8139.
- McIntosh AR, Rajah MN, Lobaugh NJ (1999) Interactions of prefrontal cortex in relation to awareness in sensory learning. *Science* 284:1531–1533.
- McIntosh AR, Rajah MN, Lobaugh NJ (2003) Functional connectivity of the medial temporal lobe relates to learning and awareness. *J Neurosci* 23:6520–6528.
- Meijer PB (1992) An experimental system for auditory image representations. *IEEE Trans Biomed Eng* 39:112–121.
- Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002) Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci U S A* 99:15164–15169.
- Murray SO, Schrater P, Kersten D (2004) Perceptual grouping and the interactions between visual cortical areas. *Neural Netw* 17:695–705.
- Pascual-Leone A, Hamilton R (2001) The metamodal organization of the brain. *Prog Brain Res* 134:427–445.
- Peltier S, Stilla R, Mariola E, LaConte S, Hu X, Sathian K (2007) Activity and effective connectivity of parietal and occipital cortical regions during haptic shape perception. *Neuropsychologia* 45:476–483.
- Pietrini P, Furey ML, Ricciardi E, Gobbi MI, Wu WH, Cohen L, Guazzelli M, Haxby JV (2004) Beyond sensory images: object-based representation in the human ventral pathway. *Proc Natl Acad Sci U S A* 101:5658–5663.
- Poirier C, De Volder AG, Scheiber C (2007) What neuroimaging tells us about sensory substitution. *Neurosci Biobehav Rev* 31:1064–1070.
- Prather SC, Votaw JR, Sathian K (2004) Task-specific recruitment of dorsal and ventral visual areas during tactile perception. *Neuropsychologia* 42:1079–1087.
- Ptito M, Moesgaard SM, Gjedde A, Kupers R (2005) Cross-modal plasticity revealed by electro-tactile stimulation of the tongue in the congenitally blind. *Brain* 128:606–614.
- Stilla R, Sathian K (2008) Selective visuo-haptic processing of shape and texture. *Hum Brain Mapp* 29:1123–1138.
- Vinberg J, Grill-Spector K (2008) Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. *J Neurophysiol* 99:1380–1393.
- Worsley KJ, Liao CH, Aston J, Petre V, Duncan GH, Morales F, Evans AC (2002) A general statistical analysis for fMRI data. *Neuroimage* 15:1–15.