OXFORD

Cerebral Cortex, 2023, **33**, 9280–9290
https://doi.org/10.1093/cercor/bhad200
Advance access publication date 6 June 2023
Original Article

Simple shape feature computation across modalities: convergence and divergence between the ventral and dorsal visual streams

Shuang Tian[1,2], Yuankun Chen[1,2], Ze Fu[1,2], Xiaoying Wang[1,2,*], Yanchao Bi[1,2,3,*]

[1]State Key Laboratory of Cognitive Neuroscience and Learning & IDG, McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China,
[2]Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China,
[3]Chinese Institute for Brain Research, Beijing 102206, China

*Corresponding authors: Xiaoying Wang, State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China. Email: wangxiaoying@bnu.edu.cn and Yanchao Bi, State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal Universtiy, Beijing 100875, China. Email: ybi@bnu.edu.cn

Shape processing, whether by seeing or touching, is pivotal to object recognition and manipulation. Although the low-level signals are initially processed by different modality-specific neural circuits, multimodal responses to object shapes have been reported along both ventral and dorsal visual pathways. To understand this transitional process, we conducted visual and haptic shape perception fMRI experiments to test basic shape features (i.e. curvature and rectilinear) across the visual pathways. Using a combination of region-of-interest-based support vector machine decoding analysis and voxel selection method, we found that the top visual-discriminative voxels in the left occipital cortex (OC) could also classify haptic shape features, and the top haptic-discriminative voxels in the left posterior parietal cortex (PPC) could also classify visual shape features. Furthermore, these voxels could decode shape features in a cross-modal manner, suggesting shared neural computation across visual and haptic modalities. In the univariate analysis, the top haptic-discriminative voxels in the left PPC showed haptic rectilinear feature preference, whereas the top visual-discriminative voxels in the left OC showed no significant shape feature preference in either of the two modalities. Together, these results suggest that mid-level shape features are represented in a modality-independent manner in both the ventral and dorsal streams.

Key words: fMRI; cross-modal; multimodal; haptic; visual cortex; shape.

Introduction

How the brain computes different types of signals (e.g. optic vs. haptic), initially processed by distinct modality-specific neural systems, into more abstract, modality-independent information contents, is a fundamental question in cognitive neuroscience. One such example is shape computation. Although shape computation in human is commonly considered in the visual processing context for object recognition (e.g. Kourtzi and Kanwisher 2000; Kourtzi et al. 2003; Haushofer et al. 2008), it can also be achieved through the haptic system. Multimodal shape representation has been extensively studied at the object level (Amedi et al. 2001; Amedi et al. 2002; Amedi et al. 2007; Stilla and Sathian 2008; Lacey et al. 2009b; Amedi et al. 2010; Lacey and Sathian 2014; Lee Masson et al. 2016a). Whether the neural computation for earlier-level shape features independent of object context could occur in a multimodal fashion remains uninvestigated and is the question of the current study.

The neural representation of object shape accessed from visual and haptic inputs has been extensively investigated using univariate and multivariate analyses, revealing distributed shape representations in the brain (e.g. Stilla and Sathian 2008; see Lacey and Sathian 2014 for a review; Lee Masson et al. 2016b; Erdogan et al. 2016). For visual shape processing, the involvement of the ventral visual pathway is well established, with extensive evidence from functional neuroimaging, neuropsychological, and neurophysiological studies (Haushofer et al. 2008; Peelen and Caramazza 2012; Bracci and Op de Beeck 2016; see reviews in Grill-Spector et al. 2001; Grill-Spector and Weiner 2014). Within the ventral visual pathway, the lateral occipital cortex (LOC) also showed shape sensitivity when sighted people haptically explored object shapes (Amedi et al. 2001, 2002) and even when early/congenitally blind individuals, who primarily construct object shape knowledge through haptic experiences, process objects through touch, verbal inputs or sound-substitute devices (Amedi et al. 2007, 2010; Striem-Amit et al. 2012; Peelen et al. 2014; Striem-Amit and Amedi 2014; Xu et al. 2022; see reviews in Lacey et al. 2009b; Bi et al. 2016). Although spatially overlapping activations may house functionally independent neural populations (e.g. Wurm and Caramazza 2019), one study further showed cross-modal neural similarity between visual and haptic object representation in the LOC, providing evidence for shared neural representation of objects in this territory (Erdogan et al. 2016).

The dorsal visual pathway along the superior occipital and parietal cortex, whose functionality has been classically assumed to be the processing of spatial- and action-related visual information (Ungerleider and Mishkin 1982; Goodale and Milner 1992), also showed sensitivity to visual object shapes (see Freud et al. 2016 for a review). Neurons selective to simple visual 2D geometric shapes (e.g. triangle, square, circle) were found in the macaque lateral intraparietal area (LIP) (Sereno and Maunsell 1998); the human superior intraparietal sulcus (IPS) represents

Received: March 1, 2023. **Revised:** May 16, 2023

various object shape features (e.g. absence or presence of a hole, different shape outlines, different combination of simple objects) in a visual short-term memory task (Xu and Chun 2006); human IPS1/IPS2 exhibited adaptation to visual 2D and 3D objects independently of image transformations, mirroring the pattern of LOC (Konen and Kastner 2008; see Xu 2018 for a review). Haptic inputs also elicit shape-related responses in the parietal cortex, with activity strength in the bilateral posterior central sulci and superior parietal lobule (SPL) being modulated by haptic shape complexity (number of curves in a haptic curve counting task, Yang et al. 2021). The neural representational similarity of haptic objects in the SPL and anterior IPS correlated with the object shape similarity (Fabbri et al. 2016; Lee Masson et al. 2016).

In what processing stage(s) does the transition from separate primary sensory cortices to modality-independent neural convergence happen? Recent visual research has focused on mid-level visual shape features (e.g. curvature), which were visual features of intermediate complexity, considered as conjunctions of low-level visual features (e.g. luminance, contrast, orientation, spatial frequency) and building blocks that underlie higher-level visual cortex domain organizations (Nasr et al. 2014; Peirce 2015; Long et al. 2018; Tang et al. 2018). The object domain distribution has been shown to be (at least partly) multimodal (e.g. Wolbers et al. 2011; Wang et al. 2015b; van den Hurk et al. 2017). Is it possible that the shared neural structures across visual and haptic modalities are already present at the mid-level shape feature processing stage before the recognition of a holistic object? To test this possibility, we chose curvature/rectilinear lines and sphere/cube plastic models as visual and haptic stimuli, respectively, in consideration that they correspond to more natural basic components for object recognition within each sensory modality respectively. We conducted fMRI experiments in which subjects visually perceived and haptically explored these stimuli and performed analyses to examine the following questions: (i) Are there brain regions showing shape sensitivity (i.e. successful within-modality decoding of curvature vs. rectilinear) for both visual and haptic inputs? (ii) Do these regions represent visual shape and haptic shape with common neural structures (i.e. successful cross-modal decoding for curvature vs. rectilinear)? (iii) Do the multimodal/cross-modal regions have specific shape preferences in a multimodal fashion (i.e. univariate response magnitude differences)?

Materials and methods

Participants

Twenty-one healthy subjects (8 females, age: mean ± SD. 23 ± 2 years, range 19–26 years) were recruited among the students at Peking University and Beijing Normal University. Out of the 21 subjects, 14 participated in both the haptic and visual versions of the shape feature experiments, and all analyses were performed on their data. The other seven subjects only took part in the haptic experiment for their unwillingness to participate in the other one. All subjects were right-handed, had normal or corrected-to-normal vision and hearing, and had no history of psychiatric or neurological disease. All of them provided informed consent and received monetary compensation for their participation. All experimental protocols were approved by the Human Subject Review Committee at Peking University.

Stimuli

The visual stimuli (see Fig. 1B) were made following a previous study looking at visual mid-level features (Nasr et al. 2014). There were three conditions: rectilinear, curvature, and straight-lines (the straight-line condition was not included in the analysis because it did not correspond to either of the haptic shape features and was designed for another study). Each condition included four patterns of arrays. Each pattern of the array contained 40 nonoverlapping shapes distributed randomly across the display. The orientations of arrays for each pattern varied in 22.5° steps, resulting in 16 pictures for each pattern and 64 pictures in total for each condition. The thickness and total length of the lines were equal across conditions.

For the haptic stimuli, basic 3D components with rectilinear and curvature features were chosen—cube and sphere. The cube and sphere models (see Fig. 1A) were made by a 3D printer (JGMAKER, Z-603S, industrial-grade precision). The bases of the models were fixed on an elongated cardboard that was used for passing the stimuli to the subjects during fMRI scanning. To control the range of hand motion, the great circle of the sphere and each surface of the cube were made with the same perimeter of 18 cm, resulting in relatively comparable total touchable surface areas (cube: 101.25 cm^2, sphere: 102.73 cm^2) and volumes (cube: 91.13 cm^3; sphere: 97.94 cm^3) between the two models.

Procedures

The haptic experiment (Fig. 1A) consisted of two functional runs, each lasting for 270 s. In each run, after a 12-s-long fixation, the experimenter presented each stimulus to the subjects' right hand in a pseudorandom order. The subjects were instructed to explore each shape when they heard "Start" and stop exploring when they heard "Stop." There were three conditions: the haptic exploration of a cube, a sphere, and the motor control. In the motor control condition, the subjects were asked to pantomime the haptic exploration of the previous stimulus with no stimulus presented to them. Each exploration or pantomiming lasted 6 s and was followed by an 8-s interstimulus interval. Each condition was presented six times per run. During the whole haptic experiment, subjects were blindfolded and wore rubber gloves to prevent the undesired effect of visual input or tactile sensation of surface texture. The haptic experiment was always conducted first to prevent the subjects from recalling and imagining the specific visual stimuli when haptically exploring.

The visual fMRI experiment (Fig. 1B) consisted of four runs. Each run lasted for 300 s and included 12 blocks (three conditions, each presented four times). Each block lasted 16 s, and the inter-block interval was 8 s. The order of the blocks was counterbalanced across conditions within each run, and the order of the four runs was counterbalanced across subjects. Within each block, there were 16 trials, consisting of an 800-ms presentation of a picture followed by a 200-ms interval. The subjects were asked to fixate on a cross in the center of the screen for 10 s at the beginning of each run and then perform a one-back task that required them to press a button with their right index finger when the current stimulus was the same as the one right before it. For each condition, there were six blocks where the one-back event appeared once, three blocks where the one-back event appeared twice, and no one-back events in the remaining three blocks. The total number of one-back events was matched across conditions.

Functional imaging

Images were acquired using a Siemens Prisma 3-T scanner with a 20-channel phase-array head coil at the Imaging Center for MRI Research, Peking University. The participants lay supine with their heads fixed with foam pads to minimize head movement. Functional imaging data were acquired with a simultaneous

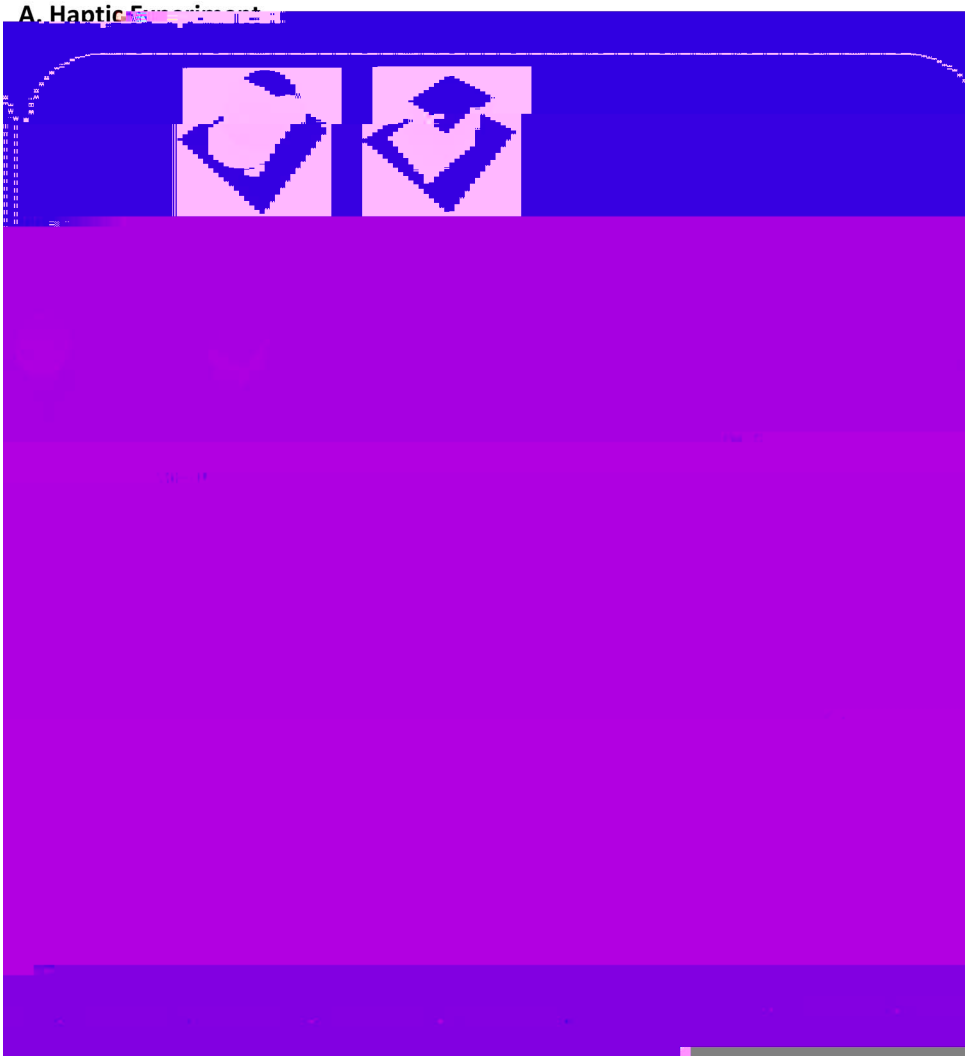

Fig. 1. Procedure of the haptic experiment (A) and the visual experiment (B). In the haptic experiment, subjects were blindfolded and cued by auditory signals to start and stop exploring the stimulus with their right hand. A baseline motor control condition was included, where the subjects pantomimed the haptic exploration of the former stimulus without any actual stimulus input. In the visual experiment, subjects were asked to perform a one-back task that required them to press a button with their right index finger when the current stimulus was the same as the one immediately before it.

multi-slice (SMS) echo-planar sequence: 62 axial slices, 2.0 mm thickness; 0.3 mm gap; multi-band factor = 2; TR = 2,000 ms; TE = 30 ms; FA = 90°; matrix size = 112 × 112; FoV = 224 × 224; voxel size = 2 × 2 × 2 mm^3.

T1-weighted anatomical images were acquired using a 3D MPRAGE sequence: 192 sagittal slices; 1 mm thickness; TR = 2,530 ms; TE = 2.98 ms; inversion time = 1,100 ms; FA = 7°; FoV = 224 × 256 mm^2

were convolved with a canonical hemodynamic response function (HRF). The beta images for multivariate analysis underwent a moderate level of smoothing with a 2 mm FWHM Gaussian kernel (Op de Beeck 2010; Gardumi et al. 2016; Hendriks et al. 2017).

. We constrained analyses within a whole-brain gray matter mask. The mask was defined as the voxels whose probability was higher than 1/3 in the SPM12 gray matter template and within the cerebral regions (1#–90#) in the Automated Anatomical Labeling (AAL) template, resulting in 122,694 voxels (981,552 mm^3). We contrasted haptic exploration with motor control to identify the regions related to haptic perception, and contrasted visual feature viewing with fixation to identify the regions related to visual perception. The group-level statistical parametric maps were thresholded at < 0.05 (voxel-level FDR-corrected, cluster size > 100 voxels). The regions that survived the statistical tests (see Results) were defined as the regions of interest (ROIs) and further underwent multivariate and univariate analyses.

. The estimated beta values of the conditions (i.e. curvature and rectilinear) in each perception ROI were extracted, standardized across voxels within each subject and fed into the support vector machine (SVM) classifier implemented in the LIBSVM library (Chang and Lin 2011) in MATLAB. In the within-modality decoding analysis, the classifier was trained and tested to discriminate between curvature and rectilinear features using the leave-one-run-out cross-validation method (see Fig. 2B–C for the schematics), and the accuracy of each ROI were tested against chance-level (i.e. 50%) using a one-sample -test (one-tailed) with R (R Core Team 2019). In this step, all ROIs underwent both within-haptic and within-visual decoding analyses. Any ROIs whose decoding accuracies were significantly above chance within either of the modalities were considered to be processing the shape features of that modality. As no ROIs showed successful decoding within both haptic and visual modalities (see Results), we conducted a more sensitive top-discriminative-voxel-based decoding analysis.

. To reduce the potential noise introduced by uninformative voxels, feature selection (Norman et al. 2006) was performed before testing multimodal decoding for each ROI. We began by running a searchlight-based within-modality decoding across the perception regions. A search sphere was centered on each voxel, with a radius of 10 mm, resulting in 515 voxels. The estimated beta weights within the sphere were extracted and standardized across voxels, and underwent the within-modality decoding analysis using the same method as in the ROI-based decoding. The predicted accuracy was assigned to the centered voxel. Following the searchlight analysis, the voxels with the top % highest accuracies in each ROI were selected as the most discriminative voxels for individual subjects. Specifically, to determine the number of , we plotted the within-modal decoding accuracy averaged across subjects against the percentage of top accuracy voxels which varied from 5 to 100% with a step of 5% (Supplementary Fig. 3A–B). The percentage (%) at which the mean accuracy reached its maximum was selected for subsequent analyses. It should be noticed that selecting the top discriminative voxels to decode within the same modality could have the problem of circular reasoning (Kriegeskorte et al. 2009). Thus, we selected voxels within one modality and tested them in the other modality. Note that here we only focused on regions which already showed successful within-modality decoding in the above ROI-based analysis. Once the decoding accuracy of the selected voxels in an ROI was significantly above chance, the ROI was considered to contain multimodal neural populations, and cross-modal decoding was then tested with these voxels. Cross-validation for cross-modal decoding was performed using the data of all runs from one modality as the training set and all runs of the other modality as the testing set, and vice versa (see Fig. 2D for the schematic). The mean accuracy across all subjects was tested against chancel-level (i.e. 50%) using a one-sample -test (one-tailed) with R.

. To test whether the multimodal/cross-modal ROIs showed any shape feature preference and whether the preference is multimodal, we also performed univariate contrast analysis for the curvature and rectilinear conditions in both the visual and haptic modalities. The whole-brain preference effects were analyzed using the group-level analysis module in SPM12. The estimated beta values for curvature and rectilinear conditions of the top discriminative voxels identified as multimodal/cross-modal were extracted and averaged across voxels for each ROI and each subject, then were tested using a 2×2 (modality $\times$ feature) repeated-measures ANOVA.

Results

To test whether and how the ventral and dorsal visual pathways support the visual and haptic processes of the curvature/rectilinear features, the following analyses were performed: (i) multimodal decoding analyses, which tested whether regions showing shape decoding in one modality also show shape decoding in the other modality; (ii) cross-modal decoding analyses, which tested whether regions identified as multimodal also have common neural shape representations across visual and haptic inputs; and (iii) univariate shape feature preference analyses, which explored whether there is any region showing specific shape preference in both modalities.

Multimodal shape feature decoding results

We first defined haptic and visual perceptual ROIs using univariate contrasts. We identified the bilateral postcentral gyri (postCG, extending to the precentral gyrus on the left side), right precentral gyrus (preCG), bilateral parietal opercula (PO), and the posterior parts of the bilateral supplementary motor areas (pos-SMA) by contrasting haptic exploration with motor control in the haptic experiment. By contrasting shape feature viewing with fixation in the visual experiment, we identified the bilateral occipital cortices (OCs) in the ventral visual pathway, the bilateral posterior parietal cortices (PPCs) within the dorsal visual pathway, as well as bilateral middle frontal gyri (MFG), the anterior parts of the bilateral supplementary motor areas (ant-SMA), right inferior frontal gyrus (IFG), and right insula (see Fig. 2A). We could only observe small overlapping clusters between visual and haptic modalities in the left OC, bilateral middle OCs, and bilateral postCG at a rather lenient threshold (< 0.05, uncorrected, Supplementary Fig. 1A). No overlapping region between modalities was observed after multiple-comparison correction. We also contrasted haptic exploration with fixation to define the haptic regions. The result is highly similar with those defined by contrasting haptic exploration with motor control after multiple-comparison correction (Supplementary Fig. 1B).

We performed shape feature decoding (i.e. curvature vs. rectilinear) within each modality in the above-identified ROIs. As shown in Fig. 2(E) and Table 1, the bilateral OC could decode visual shape features (left: mean accuracy $\pm$ SD $= 0.69 \pm 0.15$, $t_{(13)} = 4.58$, $= 0.003$; right: mean accuracy $\pm$ SD $= 0.64 \pm 0.16$, $t_{(13)} = 3.31$, $= 0.01$, FDR-corrected), whereas the bilateral PPC

Fig. 2. Shape feature decoding in perceptual ROIs. (A) the regions of interest (ROIs) defined by univariate activity contrasts at the threshold of < 0.05, voxel-wise FDR-corrected, cluster size > 100 voxels; (B–D) the analysis schemes of the within-modality and cross-modality SVM decoding. The classifier was trained and tested using a leave-one-run-out cross-validation method for the within-visual (B) and within-haptic (C) decoding. For cross-modal decoding (D), the classifier was trained with the data of all runs from one experiment and was tested with the data from the other, and vice versa; (E) results of ROI-based shape feature decoding of the perceptual ROIs. For each region, the accuracies were averaged across subjects and then tested against the chance-level accuracy (50%) using one-tailed one-sample -tests. * < 0.05, ** < 0.01, *** < 0.001, FDR-corrected.

(left: mean accuracy ± SD = 0.65 ± 0.16, $t_{(13)} = 3.73$, $= 0.01$; right: mean accuracy ± SD = 0.64 ± 0.10, $t_{(13)} = 5.44$, $= 0.0008$, FDR-corrected), bilateral MFG (left: mean accuracy ± SD = 0.59 ± 0.10, $t_{(13)} = 3.51$, $= 0.01$; right: mean accuracy ± SD = 0.59 ± 0.12, $t_{(13)} = 2.84$, $= 0.03$, FDR-corrected), ant-SMA (mean accuracy ± SD = 0.57 ± 0.09, $t_{(13)} = 2.66$, $= 0.03$, FDR-corrected), and bilateral postCG (left: mean accuracy ± SD = 0.63 ± 0.08, $t_{(13)} = 6.27$, $= 0.0004$; right: mean accuracy ± SD = 0.62 ± 0.13, $t_{(13)} = 3.33$, $= 0.01$, FDR-corrected) could decode shape features significantly above chance within the haptic modality. None of the ROIs showed significant shape feature decoding in both modalities at the entire ROI level.

Table 1. ROI-based decoding results.

	Within visual modality			Within haptic modality		
ROI	Mean Accuracy ± SD	(13)	-value (FDR-corrected)	Mean accuracy ± SD	(13)	-value (FDR-corrected)
Visually defined ROIs						
Left OC	0.69 ± 0.15	4.58**	0.003	0.53 ± 0.07	1.46	0.24
Right OC	0.64 ± 0.16	3.31*	0.01	0.51 ± 0.08	0.66	0.43
Left PPC	0.57 ± 0.19	1.42	0.24	0.65 ± 0.16	3.73**	0.01
Right PPC	0.51 ± 0.17	0.19	0.57	0.64 ± 0.10	5.44***	0.0008
Left MFG	0.54 ± 0.19	0.72	0.43	0.59 ± 0.10	3.51*	0.01
Right MFG	0.48 ± 0.15	−0.46	0.81	0.59 ± 0.12	2.84*	0.03
Right IFG	0.50 ± 0.10	0	0.63	0.51 ± 0.14	0.24	0.57
Right insula	0.46 ± 0.14	−0.94	0.91	0.48 ± 0.10	−0.79	0.90
SMA-ant	0.51 ± 0.22	0.15	0.57	0.57 ± 0.09	2.66*	0.03
Haptically defined ROIs						
Left postCG	0.54 ± 0.19	0.72	0.43	0.63 ± 0.08	6.27***	0.0004
Right postCG	0.54 ± 0.17	0.77	0.43	0.62 ± 0.13	3.33*	0.01
Left PO	0.41 ± 0.13	−2.5	0.99	0.51 ± 0.10	0.45	0.50
Right PO	0.46 ± 0.14	−1.16	0.93	0.53 ± 0.11	0.99	0.36
Right preCG	0.56 ± 0.18	1.34	0.24	0.46 ± 0.09	−1.5	0.95
SMA-pos	0.53 ± 0.16	0.61	0.44	0.53 ± 0.08	1.35	0.24

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$

Considering that uninformative voxels can induce noise and reduce classifier performance (Norman et al. 2006) and thus lead to false negatives, we carried out a further analysis to address this issue. Specifically, to increase the potential signals, we selected the top % discriminative voxels within each ROI in its dominant modality (i.e. the modality in which ROI-based decoding was significantly above chance at the entire ROI level) and used these voxels to decode the shape features from the other modality for individual subjects (see Materials and Methods). We plotted the decoding accuracies across the full range of top voxel percentages in Supplementary Fig. 3(A–B), and identified the top % voxels corresponding to the optimal decoding accuracy for the visual/haptic dominant ROIs (see Table 2 for the specific top % value for each ROI), and then performed the analyses in the other modality on the identified voxels (see Fig. 3A). For the visual-dominant bilateral OC ROIs, only the activity pattern of the visual-discriminative voxels (top 20%) in the left OC could decode haptic shape features significantly above chance (mean accuracy ± SD = 0.57 ± 0.11, $t_{(13)} = 2.43$, = 0.03, FDR-corrected; see also Table 2). Among the haptic shape feature-sensitive ROIs, the haptic-discriminative voxels in the left PPC (top 30%) could decode visual shape features (mean accuracy ± SD = 0.65 ± 0.18, $t_{(13)} = 3.08$, = 0.03, FDR-corrected; see also Table 2). Thus, the left OC and PPC top informative voxels could discriminate shape features in both visual and haptic modalities. We also presented the results of the full range of the discriminative voxels (see Fig. 3B). This full-range picture again showed that only the left OC and the left PPC exhibited successful shape feature decoding in both modalities across a relatively large top-voxel percentage range, which indicated that the findings were not specific to any particularly highlighted percentage.

Cross-modal shape feature decoding results

As the top discriminative voxels in the left OC and left PPC showed multimodal effects, we further tested whether these voxels exhibited shared neural population coding of shape features across visual and haptic modalities, or they coded shape features in an independent way within each modality (Wurm and Caramazza 2019). We used the above-identified top discriminative voxels to perform the cross-modal decoding analysis. As shown in Fig. 3(A) and Table 2, the top visual-discriminative voxels in left OC and top haptic-discriminative voxels in left PPC could also decode shape features cross-modally (OC: mean accuracy ± SD = 0.53 ± 0.06, $_{(13)} = 2.25$, = 0.05, FDR-corrected; PPC: mean accuracy ± SD = 0.52 ± 0.04, $_{(13)} = 2.13$, = 0.04, FDR-corrected). By separating the two ways of cross-modal decoding (i.e. training with the haptic data and testing with the visual data, and vice versa), we found that the successful cross-modal decoding was mainly driven by the visual-to-haptic direction (Fig. 3A, bottom panel; also see Supplementary Fig. 3C for confusion matrices).

We also performed whole-brain searchlight analyses for within-modal and cross-modal decoding (see Supplementary Material for the method). Significant visual shape feature decoding was observed in the bilateral LOC, and haptic shape feature decoding in the bilateral PPC (voxel-level < 0.001, cluster-level FWE corrected < 0.05; Supplementary Fig. 4A). However, no overlapping clusters between modalities were observed under any thresholds adjusted for multiple-comparison. No clusters survived multiple-comparison correction for cross-modal shape feature decoding, either. Under a lenient threshold (uncorrected < 0.05), we found overlapping areas between the two modalities along both ventral and dorsal pathways, including the left OC and left PPC observed as multimodal and cross-modal in the ROI analyses, as well as in the left postCG, right preCG, and bilateral MFG (Supplementary Fig. 4B).

Univariate results: shape preference across modalities

The previous SVM analyses examined whether neural activity patterns could discriminate between different shape features, without distinguishing the potential functional preference for each shape feature (i.e. curvature vs. rectilinear). Here, we tested whether the shape feature preference was similar or different across visual and haptic modalities in the top discriminative voxels in the left OC and left PPC (Fig. 4A). As shown in Fig. 4(B), in the left OC, there was a significant main effect of modality ($_{(1, 13)} = 147.3$; $= 1.82 \times 10^{-8}$; $\eta^2 = 0.92$; stronger responses in visual modality), but neither a significant main effect of

Table 2. Top-discriminative-voxel-based decoding results.

ROI	Decoding modality	Top % (number) of voxels	Mean accuracy ± SD	$_{(13)}$	-value (FDR-corrected)
Top visual-discriminative voxels					
Left OC	Haptic	20% (475)	0.57 ± 0.11	2.43*	0.03
Right OC	Haptic	20% (622)	0.51 ± 0.07	0.3	0.39
Top haptic-discriminative voxels					
ant-SMA	Visual	55% (136)	0.51 ± 0.21	0.16	0.58
Left PPC	Visual	30% (154)	0.65 ± 0.18	3.08*	0.03
Left postCG	Visual	30% (603)	0.56 ± 0.16	1.45	0.30
Left MFG	Visual	60% (184)	0.50 ± 0.16	0	0.58
Right PPC	Visual	60% (277)	0.50 ± 0.10	0	0.58
Right MFG	Visual	65% (137)	0.53 ± 0.15	−1.47	0.92
Right postCG	Visual	35% (215)	0.46 ± 0.09	0.67	0.58
Multimodal voxels					
Left	Cross-modal	20% (475)	0.53 ± 0.06	2.25*	0.05
OC	Visual-to-haptic		0.55 ± 0.09	2.01*	0.05
	Haptic-to-visual		0.52 ± 0.05	1.47	0.08
Left PPC	Cross-modal	30% (154)	0.52 ± 0.04	2.13*	0.04
	Visual-to-haptic		0.54 ± 0.07	2.31*	0.04
	Haptic-to-visual		0.50 ± 0.07	0	0.50

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$

feature ($_{(1, 13)} = 0.03$; $= 0.88$; $\eta^2 = 1.88 \times e^{-3}$) nor an interaction ($_{(1, 13)} = 0.09$; $= 0.78$; $\eta^2 = 6.54 \times e^{-3}$) was observed. For the left PPC, the main effect of feature ($_{(1, 13)} = 23.77$; $= 3.03 \times 10^{-4}$; $\eta^2 = 0.64$; stronger responses to rectilinear features) and modality ($_{(1, 13)} = 52.53$; $= 6.48 \times 10^{-6}$; $\eta^2 = 0.79$; preferring visual modality) as well as the interaction between feature and modality ($_{(1, 13)} = 16.41$; $= 1.37 \times 10^{-3}$; $\eta^2 = 0.60$) were all significant. The interaction was in the direction of stronger rectilinear preference in the haptic than in the visual experiment. Indeed, post-hoc tests showed that the rectilinear-preference effect was significant in the haptic modality ($_{(13)} = 8.29$, $= 3.02 \times 10^{-6}$, FDR-corrected), but not in the visual modality ($_{(13)} = 1.73$, $= 0.11$, FDR-corrected).

We also ran a voxel-wise contrast analysis across whole-brain for each experiment. No voxels survived multiple-comparison correction in the visual experiment. In the haptic experiment, bilateral SPL and postCG showed significant preference to rectilinear feature (FDR-corrected; see Supplementary Fig. 5 for the visualization of the statistical maps), converging with the ROI-wise result.

Discussion

To investigate where and at what processing stage neural shape computations become multimodal and cross-modal, we conducted visual and haptic fMRI experiments using contrasting shape features (rectilinear vs. curvature). Our results indicate that voxels in the ventral visual region left OC and the dorsal visual region left PPC exhibit multimodal shape representation properties, as they were able to discriminate rectilinear vs. curvature shape features from both visual and haptic inputs, based on at least partly common neural representations as shown by cross-modal decoding success. Additionally, our specific shape feature preference test revealed different relationships between modality and shape pattern across regions, with the PPC preferring the haptic rectilinear shape, whereas OC showing no significant shape preference in either of the two modalities.

The findings that OC and PPC showed shape feature representation in both visual and haptic experiments are generally consistent with previous studies examining object shape processing in these territories across multiple modalities (see Introduction). Importantly, they indicate that sensitivity to shape across modalities is already present at the level of simple shape features without real object contexts, and the positive cross-modal decoding results demonstrate shared neural shape representation across modalities. It is worth noting that the visual and haptic stimuli in our experiments were not identical, given the intrinsic difference of the two sensory systems. However, both types of stimuli included curved vs. rectilinear geometry elements, albeit in an abstract manner. The positive cross-modal decoding effects—classifiers trained on one modality (i.e. visual rectilinear vs. curvature) successfully decoded corresponding features in the other modality (i.e. haptic cube vs. sphere)—in OC and PPC were thus particularly informative, indicating that the representations shared between modalities were rather abstract. The imbalance of cross-modal decoding (i.e. strong visual-to-haptic decoding but insignificant haptic-to-visual decoding) is also informative, indicating that the haptic perception may involve multiple processing components not shared by the visual modality (discussed further below).

More fine-grained location comparison also suggests potentially interesting differences. By plotting the distribution of the top-discriminative voxels in the left OC across subjects (Fig. 4A), we observed that the highly overlapping voxels located at the ventral occipital area, which is consistent with the peak cross-modal cluster in the left OC in the whole-brain searchlight results (peak coordinates: −37, −77, −11; see Supplementary Fig. 4B). This area is more ventral and posterior to the commonly proposed multimodal (object) shape area (LOtv), which has been previously shown to be activated by both visual and haptic perception of objects in sighted people and by haptic object perception in people without visual experience (Amedi et al. 2001, 2002, 2010; Snow et al. 2014; Erdogan et al. 2016; Lee Masson et al. 2016). For the parietal lobe, although the shape representations probed using haptic objects were observed in SPL (Erdogan et al. 2016; Lee Masson et al. 2016; Yang et al. 2021) and IPS (Amedi et al. 2002; Stilla and Sathian 2008; Lacey et al. 2009a; Snow et al. 2014; Fabbri et al. 2016), those probed using visual objects were mainly distributed in IPS1/IPS2/superior IPS (Freud et al. 2017; Konen and Kastner 2008; Xu and Chun 2006; see Xu 2018 for a review). The PPC cluster we observed as multimodal and cross-modal located mainly in the SPL (also see the only parietal cluster showing

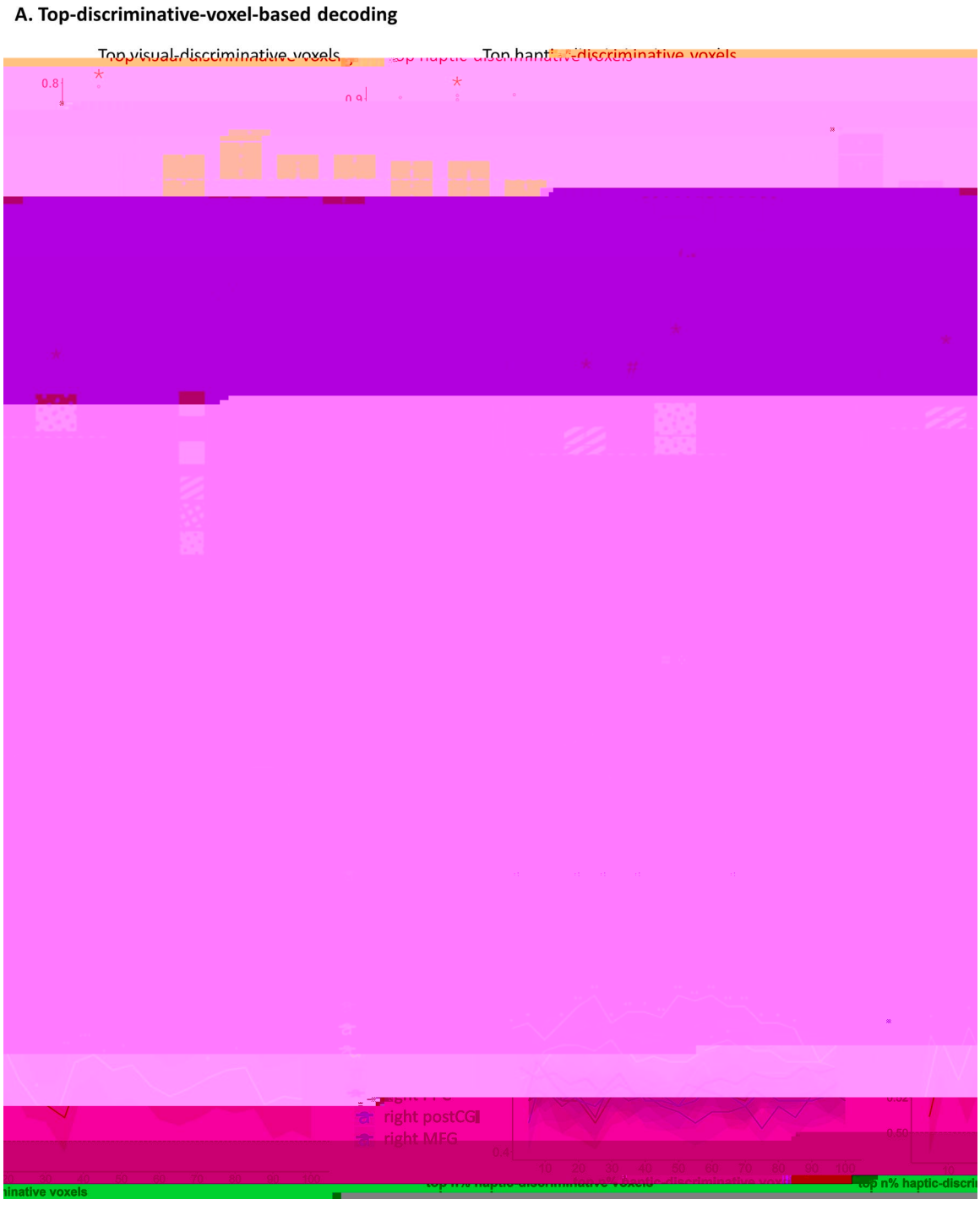

Fig. 3. Results of top-discriminative-voxel-based multimodal and cross-modal decoding. (A) Accuracies of decoding with the optimal number of voxels in each shape feature-sensitive ROI. The top discriminative voxels were defined in individual subjects (see Materials and Methods), and the obtained accuracies were tested against the chance-level accuracy (50%) using one-tailed one-sample -tests. The cyan boxplots indicate within-haptic decoding with the top visual-discriminative voxels; the orange indicatenge w 1of dnge 390.963.d tt.7(t)-.6(o4(T)-322.7(sha)1ha).6(visual-discriminati)6.7(v)16(nge)-39

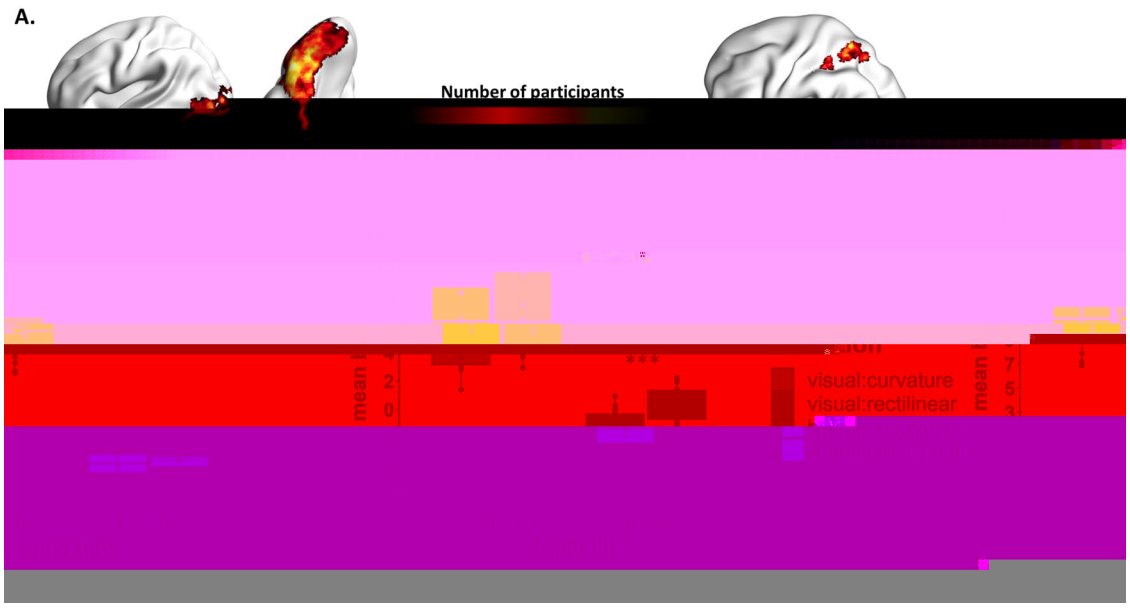

Fig. 4. Shape feature preference in the left OC and left PPC. (A) The visualization of the distributions of the top 20% visual-discriminative voxels in the left OC and the top 30% haptic-discriminative voxels in the left PPC, which exhibited both multimodal and cross-modal multivariate representation of shape

patterns. This is in line with the classical role of the ventral visual pathway supporting object shape identification (Amedi et al. 2001, 2002; Kourtzi and Kanwisher 2001; Kourtzi et al. 2003; Haushofer et al. 2008), and offers further evidence for the representation of highly abstract shape feature knowledge that is extracted from different sensory modalities and very different geometric configurations. Note that although no significant shape feature preference was observed for the left OC multimodal voxels, we observed a trend of shape feature preference pattern under a lenient threshold (uncorrected < 0.05) within the visual cortex, with bilateral lateral fusiform gyri showing a trend of stronger responses to curvature, and the very posterior parts of the bilateral LOC preferring rectilinear. These findings were largely in line with previous studies (Nasr et al. 2014; Yue et al. 2014, 2020; Fan et al. 2021). In contrast, although the dorsal pathway also exhibits evidence for (at least partly) shared neural representation of cross-modal shape features, there is an additional functional preference difference. PPC multimodal voxels show an intriguing interaction between modality and shape, with a stronger preference for the rectilinear feature when touching. The stronger preference for rectilinear stimuli in haptic processes here is potentially aligned with one of the primary functionalities of the parietal lobe as processing sensory information for action (reaching and grasping; Binkofski and Buxbaum 2013; Culham et al. 2003; Fabbri et al. 2016; Goodale and Milner 1992; Rizzolatti and Matelli 2003), as objects that humans tend to manipulate (i.e. haptically interact with) associate with more rectilinear features (Fan et al. 2021).

In conclusion, we observed both convergent and different profiles in the ventral and dorsal visual pathways for mid-level geometric shape feature processing. Both streams contained voxels whose neural patterns could distinguish between geometric shape features within both visual and haptic modalities, with sufficient shared neural representation to enable successful cross-modal decoding. Furthermore, the discriminative voxels in the dorsal stream had a preference for rectilinear shape features in the haptic modality, whereas those in the ventral stream had no specific shape feature preference in either modality. Together, these findings suggest that both ventral and dorsal visual streams process mid-level shape features invariant to input modality, albeit potentially for different computational goals (i.e. recognition vs. action).

Acknowledgments

We thank Dr. Shahin Nasr for kindly sharing the visual stimuli from their study. We thank Chunfang Yan and Dr. Weiwei Men for the help with the haptic fMRI experiment settings.

Author contributions

Shuang Tian: Data curation, Formal Analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing—original draft, Writing—review and editing. Yuankun Chen: Investigation, Methodology, Resources. Ze Fu: Methodology, Resources. Xiaoying Wang: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing—original draft, Writing—review and editing. Yanchao Bi: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing—original draft, Writing—review and editing.

Supplementary material

Supplementary material is available at

Gardumi A, Ivanov D, Hausfeld L, Valente G, Formisano E, Uludağ K. The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. 2016:132:32–42.

Goodale MA, Milner AD. Separate visual pathways for perception and action. 1992:15(1):20–25.

Grill-Spector K, Weiner KS. The functional architecture of the ventral temporal cortex and its role in categorization. 2014:15:536–548.

Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object recognition. 2001:41(10–11):1409–1422.

Haushofer J, Livingstone MS, Kanwisher N. Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. 2008:6(7):1459–1467.

Hendriks MHA, Daniels N, Pegado F, de Beeck HPO. The effect of spatial smoothing on representational similarity in a simple motor paradigm. 2017:8:1–11.

van den Hurk J, Van Baelen M, Op de Beeck HP. Development of visual category selectivity in ventral visual cortex does not require visual experience. 2017:114(22):E4501–E4510.

James TW, Humphrey GK, Gati JS, Servos P, Menon RS, Goodale MA. Haptic study of three-dimensional objects activates extrastriate visual areas. 2002:40(10):1706–1714.

Konen CS, Kastner S. Two hierarchically organized neural systems for object information in human visual cortex. 2008:11(2):224–231.