

# Object-selective responses in the human motion area MT/MST

Zoe Kourtzi<sup>1,2</sup>, Heinrich H. Bühlhoff<sup>1</sup>, Michael Erb<sup>3</sup> and Wolfgang Grodd<sup>3</sup>

<sup>1</sup> Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tuebingen, Germany

<sup>2</sup> MIT, Department of Brain and Cognitive Science, NE20, 77 Massachusetts Ave., Cambridge, Massachusetts 02139, USA

<sup>3</sup> University Clinics, Hoppe-Seyler Str. 3, 72076 Tuebingen, Germany

Correspondence should be addressed to Z.K. ([zoe.kourtzi@tuebingen.mpg.de](mailto:zoe.kourtzi@tuebingen.mpg.de))

Published online: 10 December 2001, DOI: 10.1038/nn780

The perception of moving objects and our successful interaction with them entail that the visual system integrates shape and motion information about objects. However, neuroimaging studies have implicated different human brain regions in the analysis of visual motion<sup>1,2</sup> (medial temporal cortex; MT/MST) and shape<sup>3,4</sup> (lateral occipital complex; LOC), consistent with traditional approaches in visual processing that attribute shape and motion processing to anatomically and functionally separable neural mechanisms. Here we demonstrate object-selective fMRI responses (higher responses for intact than for scrambled images of objects) in MT/MST, and especially in a ventral sub-region of MT/MST, suggesting that human brain regions involved mainly in the processing of visual motion are also engaged in the analysis of object shape.

To test object-selective responses in MT/MST, we presented observers with intact and scrambled images of the same set of objects (Supplementary Fig. 1, available on the *Nature Neuroscience* web site) rendered in four different ways: static silhouettes of objects (two-dimensional (2D) objects), 2D translating objects (moving objects), static objects presented stereoscopically in front of their background (stereo objects), and static three-dimensional objects defined by shading (shaded objects).

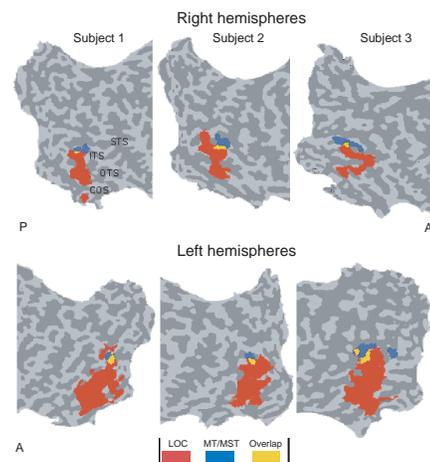
For each subject, we identified MT/MST, the LOC and the early visual areas as regions of interest (ROIs) (Fig. 1). Analysis of the responses in these ROIs for the different object types (Fig. 2, Supplementary Fig. 2) showed a significant three-way interaction ( $F_{3,69} = 2.28$ ,  $p < 0.05$ ) for ROI (MT/MST, LOC), image type

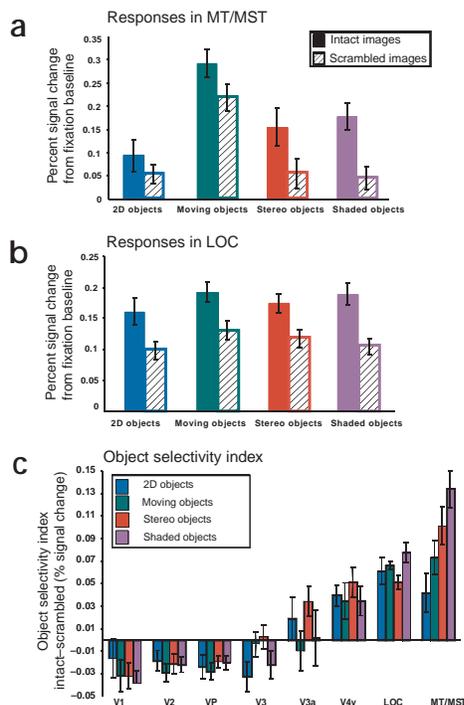
(intact, scrambled) and object condition (2D objects, moving objects, stereo objects, shaded objects) and two-way interactions for ROI and object condition ( $F_{3,69} = 30.34$ ,  $p < 0.001$ ), and image type and object condition ( $F_{3,69} = 9.55$ ,  $p < 0.001$ ).

Specifically, in MT/MST, a repeated ANOVA showed main effects of image type ( $F_{1,69} = 46.89$ ,  $p < 0.001$ ) and object condition ( $F_{3,69} = 58.44$ ,  $p < 0.001$ ) and an interaction between these two variables ( $F_{3,69} = 5.97$ ,  $p = 0.001$ ). A similar statistical analysis in the LOC showed main effects of image type ( $F_{1,69} = 169.74$ ,  $p < 0.001$ ) and object condition ( $F_{3,69} = 9.47$ ,  $p < 0.001$ ) but no significant interaction between these two variables ( $F_{3,69} = 1.98$ ,  $p = 0.15$ ). In particular, object-selective responses (stronger responses to intact than to scrambled images of objects) were observed in both MT/MST and the LOC. Moreover, the overall responses to moving objects were higher than the responses to stereo, shaded and static 2D objects in both ROIs. These differences across object types may have been due to the fact that motion and depth cues (stereo, shading) provide more information about the shape of objects than the 2D silhouettes. Furthermore, contrast analysis in MT/MST showed significantly stronger responses to intact than to scrambled images for moving ( $F_{1,69} = 20.54$ ,  $p = 0.001$ ), stereo ( $F_{1,69} = 56.29$ ,  $p = 0.001$ ) and shaded objects ( $F_{1,69} = 86.52$ ,  $p = 0.001$ ), but not for 2D objects ( $F_{1,69} = 2.1$ ,  $p = 0.15$ ). That is, object-selective responses were observed in MT/MST for moving objects as well as for static objects defined by binocular or monocular depth cues, but not for static 2D objects. In contrast, the lack of interaction between image type and object condition in the LOC suggests that object-selective responses in this region were similar across object types. These results indicate that shape information about objects defined by different visual cues may be processed differentially in MT/MST and the LOC (Supplementary Fig. 4). Finally, responses in the early retinotopic regions were not object-selective (Fig. 2c).

Interestingly, we observed an overlap on the activation maps between MT/MST and the LOC (Fig. 1). Analysis of the responses (Supplementary Fig. 3) within this overlap region in the ventral part of MT/MST and a non-overlap region (voxels in MT/MST that did not overlap with voxels in the LOC) showed a significant three-way interaction ( $F_{3,69} = 2.2$ ,  $p = 0.05$ ) for ROI (overlap, non-overlap region), image type and object condition, and two-way interactions for ROI and object condition ( $F_{3,69} = 14.98$ ,  $p < 0.001$ ), and image type and object condition ( $F_{3,69} = 18.38$ ,  $p < 0.001$ ). No significant interaction was observed for ROI and image type ( $F_{1,69} < 1$ ,  $p = 0.75$ ), suggesting that both subregions in MT/MST show shape-selective responses. However,

**Fig. 1.** Functional activation maps. Functional activation maps for 3 subjects showing MT/MST (Talairach coordinates, mean  $\pm$  s.d. for right hemispheres,  $42.9 \pm 5.5$ ,  $-72.5 \pm 4.7$ ,  $4.4 \pm 4.3$ , and left hemispheres  $-45.8 \pm 6.3$ ,  $-67.4 \pm 7.1$ ,  $0.2 \pm 6.9$ ) and the LOC (Talairach coordinates, right hemispheres,  $41.4 \pm 3.8$ ,  $-62 \pm 3.8$ ,  $-11.9 \pm 3.1$ , and left hemispheres,  $-37.9 \pm 3.6$ ,  $-66.1 \pm 4.1$ ,  $-7.2 \pm 5.1$ ). The functional activations are superimposed on flattened cortical surfaces of the right and left hemispheres. The sulci are coded in darker gray than the gyri and the anterior–posterior orientation is noted by A and P. STS, superior temporal sulcus; ITS, inferior temporal sulcus; OTS, occipitotemporal sulcus; COS, collateral sulcus. MT/MST was defined as the set of all contiguous voxels (indicated in blue) in the lateral occipital region activated more strongly ( $p < 10^{-4}$ ) by moving than by stationary rings (Supplementary Methods). The LOC was defined as the set of all contiguous voxels (indicated in red) in the ventral occipitotemporal cortex that were activated more strongly ( $p < 10^{-4}$ ) by intact than by scrambled images of objects (Supplementary Methods). Overlapping voxels between the LOC and MT/MST are shown in yellow.





**Fig. 2.** Object-selective responses in MT/MST, LOC and retinotopic regions. Average percent signal increases (from the fixation baseline trials) for images of 2D, moving, stereo and shaded objects in MT/MST (a) and the LOC (b). Shown is the average signal from time points 3–5 s that were selected as the peak of the event-related responses (Supplementary Methods). Error bars, standard errors on the percent signal change averaged across scans and subjects. (c) An object-selectivity index (percent signal change for intact images – percent signal change for scrambled images) is plotted for each object type in each ROI. The error bars indicate standard errors on the percent signal change averaged across scans and subjects. This analysis illustrates that object selectivity was observed in the LOC and in MT/MST. An interaction between ROI and object condition ( $F_{3,69} = 2.3$ ,  $p < 0.05$ ), and a main effect of object condition ( $F_{3,69} = 9.55$ ,  $p < 0.001$ ) with the highest selectivity for shaded objects ( $F_{1,69} = 15.35$ ,  $p < 0.001$ ) and the lowest for 2D objects ( $F_{1,69} = 22.31$ ,  $p < 0.001$ ) were observed. Finally, responses in the early retinotopic regions were not object-selective. That is, retinotopic regions responded to both intact and scrambled images of objects but did not show stronger responses to intact than scrambled images. Specifically, we observed stronger activations for scrambled than for intact images of objects (V1,  $F_{1,69} = 14.17$ ,  $p < 0.001$ ; V2,  $F_{1,69} = 24.50$ ,  $p < 0.001$ ; VP,  $F_{1,69} = 22.64$ ,  $p < 0.001$ ), or no significant differences between intact and scrambled images (V3,  $F_{1,69} = 1$ ,  $p = 0.588$ ; V3a,  $F_{1,69} < 1$ ,  $p = 0.453$ ). Interestingly, stronger responses to intact than to scrambled images of objects were observed in area V4v ( $F_{3,69} = 8.27$ ,  $p < 0.01$ ) for the type of objects presented in this experiment.

er, a contrast analysis showed that the responses to moving objects in the non-overlap region were not object-selective. That is, the non-overlap region responded strongly to both intact and scrambled images of moving objects ( $F_{1,69} < 1$ ,  $p = 0.49$ ). This analysis suggests two functionally separable subregions in MT/MST: a region that responds to motion independent of shape properties (MT/MST proper), and a region that responds to both the shape and the motion of objects (overlap region between the LOC and MT/MST). The human MT/MST is a complex of regions that have not yet been conclusively characterized as separable areas with different anatomical loci and functional properties. It is possible that the overlap region we observed in our experiments (Talairach coordinates, mean  $\pm$  s.d. for right hemispheres,  $43.3 \pm 7.7$ ,  $-71.2 \pm 6.7$ ,  $1.6 \pm 5.7$ , and left hemispheres,  $-41.3 \pm 5.6$ ,  $-74.4 \pm 6.1$ ,  $4.7 \pm 5.2$ ) is anatomically the same as a region adjacent to MT/MST that responds to structure-from-motion stimuli<sup>5,6</sup> or kinetic motion boundaries<sup>7,8</sup> or a ventral subregion within MT/MST that responds to optic flow stimuli that provide 3D information about objects<sup>9</sup>. Further studies of higher spatial resolution are needed to test more precisely the anatomical and functional localization of these regions.

In summary, in contrast to traditional theories of visual processing, our findings suggest functional interactions between the neural mechanisms involved in the processing of shape and motion information about objects. Consistent with this hypothesis, previous studies have shown responses in motion areas for structure-from-motion displays<sup>5,6</sup> and surfaces in different depth planes defined by motion<sup>10–12</sup>. Our findings provide evidence that MT/MST, and especially a ventral subregion of MT/MST, is involved in the processing of shape properties of both moving and static 3D objects defined by binocular or monocular (shading) depth cues. These findings are consistent with neurophysiological studies showing that two-thirds of the neurons in MT are tuned for binocular disparity<sup>13</sup> and may mediate depth perception for both moving and static stimuli<sup>14,15</sup>. The object-selective fMRI responses that we observed for shaded objects

suggest that these binocular disparity-tuned neural populations in MT/MST are not only involved in the analysis of local disparity signals but may also be engaged in the processing of the perceived 3D shape of objects independent of the cues (binocular or monocular) defining the object structure. Consistent with this interpretation, the strongest object-selective responses in MT/MST were observed for objects defined by shading cues that specify 3D structure than for objects defined by motion or stereo that simply facilitate segmentation of objects from their background (Fig. 2c). Future studies will be required to fully explore the role of MT/MST in mediating the perception of objects and our interaction with them.

*Note: Supplementary Figs. 1–4 and Methods are available on the Nature Neuroscience web site ([http://neuroscience.nature.com/web\\_specials](http://neuroscience.nature.com/web_specials)).*

#### Acknowledgements

We would like to thank A. Dale, B. Fischl, D. Greve, A. van der Kouwe and T. Kammer for their help with imaging. A. Höpfer for technical help with the data collection and N. Aguilar and M. Thangarajh for their help with data analysis. We would also like to thank the following people for comments and suggestions: D. Cunningham, N. Kanwisher, N. Logothetis, M. Sereno, N. Sigala, S. Smirnakis, and A. Tolias. This work was supported by the Max Planck Society and a Mc Donnell-Pew grant # 3944900 to Z.K.

RECEIVED 13 AUGUST; ACCEPTED 5 NOVEMBER 2001

- Zeki, S. *et al.* *J. Neurosci.* **11**, 641–649 (1991).
- Tootell, R. B. H. *et al.* *J. Neurosci.* **15**, 3215–3230 (1995).
- Kanwisher, N., Chun, M. M., McDermott, J. & Ledden, P. J. *Brain Res. Cogn. Brain Res.* **5**, 55–67 (1996).
- Malach, R. *et al.* *Proc. Natl. Acad. Sci. USA* **92**, 8135–8138 (1995).
- Orban, G. A., Sinaert, S., Todd, J. T., Van Hecke, P. & Marchal, G. *Neuron* **24**, 929–940 (1999).
- Paradis, A. L. *et al.* *Cereb. Cortex* **10**, 772–783 (2000).
- Orban, G. A. *et al.* *Proc. Natl. Acad. Sci. USA* **92**, 993–997 (1995).
- Tootell, R. B. H. & Hadjikhani, N. *Cereb. Cortex* **11**, 298–311 (2001).
- Morrone, M. C. *et al.* *Nat. Neurosci.* **3**, 1322–1328 (2000).
- Qian, N. & Andersen, R. A. *J. Neurosci.* **14**, 7367–7380 (1994).
- Bradley, D. C., Qian, N. & Andersen, R. A. *Nature* **373**, 609–611 (1995).
- Xiao, D. K., Marcar, V. L., Raiguel, S. E. & Orban, G. A. *Eur. J. Neurosci.* **9**, 956–964 (1997).
- Maunsell, J. H. & Van Essen, D. C. *J. Neurophysiol.* **49**, 1148–1167 (1983).
- DeAngelis, G. C., Cumming, B. G. & Newsome, W. T. *Nature* **394**, 677–680 (1998).
- DeAngelis, G. C. & Newsome, W. T. *J. Neurosci.* **19**, 1398–1415 (1999).