The Integration of Colour and Motion by the Human Visual Brain

Objects in the visual scene are defined by different cues such as colour and motion. Through the integration of these cues the visual system is able to utilize different sources of information, thus enhancing its ability to discriminate objects from their backgrounds. In the following experiments, we investigate the neural mechanisms of cue integration in the human. We show, using functional magnetic resonance imaging (fMRI), that both colour and motion defined shapes activate the lateral occipital complex (LOC) and that shapes defined by both colour and motion simultaneously activate the anterior-ventral margins of this area more strongly than shapes defined by either cue alone. This suggests that colour and motion cues are integrated in the LOC and possibly a neighbouring, more anterior, region. We support this result using an fMR adaptation technique, demonstrating that a region of the LOC adapts on repeated presentations of a shape regardless of the cue that is used to define it and even if the cue is varied. This result raises the possibility that the LOC contains cue-invariant neurons that respond to shapes regardless of the cue that is used to define them. We propose that such neurons could integrate signals from different cues, making them more responsive to objects defined by more than one cue, thus increasing the ability of the observer to recognize them.

Keywords: adaptation, color and motion integration, fMRI, lateral occipital complex, object recognition

Introduction

The visual brain discriminates objects from their backgrounds by using many different signals, such as colour, motion, depth and luminance. But because many of these signals are similar for both the object and the background the visual system must rely on small differences in signal, or cues, for its processing. To allow rapid, and correct, discrimination of objects it must integrate different cues, utilizing every available source of information. Recent psychophysical studies (Rivet and Cavanagh, 1996; Rivet et al., 1997; Bach et al., 2000; Nothdurft, 2000; Kubovy and Cohen, 2001) have shown that the human visual system integrates information from different cues so that objects defined by several cues enjoy perceptual advantages over those defined by only one cue. In this study we explored the neural basis of this remarkable capacity, about which little is known.

In the following experiments we investigate the mechanisms of cue integration using the two cues that are known to be most separate in terms of their cortical representations, i.e. colour and motion. Early imaging studies supposed that shapes defined by different cues are processed by different visual areas (Gulyas et al., 1994). Similarly a region of occipital cortex, named ‘KO’, was proposed to respond specifically to kinetic contours over contours defined by other cues (Dupont et al., 1997; Van Oostende et al., 1997). However, more recent studies have found that human form areas are largely cue invariant, i.e. form-specific areas can respond to forms regardless of the cue that defines them. For example, the specialization of KO for kinetic contours has been questioned by Zeki et al. (2003), who found that there is no significant difference between KO’s response to motion shapes compared to its response to coloured shapes. Moreover, it is well documented that the lateral occipital complex, which is known to be specialized for processing the shape of objects (Malach et al., 1995; Kanwisher et al., 1996), responds well to shapes regardless of whether they are defined by motion or luminance (Grill-Spector et al., 1998) or depth or two-dimensional shading (Kourtzi et al., 2002).

What is not known is whether presenting shapes defined by two or more cues simultaneously produces any extra activity in cue-invariant areas. In other words, visual areas can respond to many different cues, but can they integrate them? If visual areas do integrate different cues, then it is of great interest to learn whether different types of information are brought together into individual, cue-invariant, neurons or whether they are processed by different subpopulations of neurons within the same visual area. We addressed these questions in three parts using functional magnetic resonance imaging (fMRI) for all.

Firstly, we localized regions of the brain that responded strongly to both colour and motion defined shapes. Secondly, in an event-related fMRI paradigm, we searched for areas where shapes defined by both colour and motion at the same time produced more activity than shapes defined by only one cue. We then used an fMR adaptation paradigm to determine whether these areas contain cue-invariant neurons or separate populations of neurons, each responding to shapes defined by one cue alone.

Materials and Methods

Subjects

Five subjects (three female, all right-handed, mean age 23.4 years) took part in experiments 1 and 2. Nine subjects took part in experiment 3 (all male, seven right-handed, mean age 23.8 years). All subjects gave informed consent in accordance with the Declaration of Helsinki, and the Ethics Committee of the National Hospital for Neurology and Neurosurgery, London, UK, granted ethical approval for the study.

Stimuli

In all the following experiments we used three types of simple shape stimuli composed of kinetic, coloured dots. We constructed these shapes by varying the amount of colour and motion coherence present in the shape (Fig. 1i-iii). Three different conditions were generated: shapes defined purely by colour coherence (C), those defined only by motion coherence (M), and those defined by both colour and motion coherence (CM). This results in two unimodal conditions in which only one source of information distinguishes the shape from its background (C, M) and one bimodal condition in which either of two sources of

Matthew W. Self and S. Zeki

Anatomy Department, Wellcome Department of Imaging Neuroscience, University College London, Gower Street, London WC1E6BT, UK
information, or both, could be used to identify the shape (CM). We also created control conditions that still contained high colour or motion coherence but no visible shape (Fig. 1i–iv). The three control conditions were: 100% coherent color with no shape present (CC), 100% coherent motion (MM), and 100% noise containing no form or property. For example if 50% of the dots inside the shape had certain colour and 33% of the dots outside the shape had the same colour, then the coherence of the shape was $100 \times (0.5 - 0.33) \times (1 - 0.33)^2 = 25\%$

Note that there was always motion and colour present in the stimuli regardless of the condition, the condition being varied only by varying the coherence of the two cues. In the bimodal condition, the same dots were chosen to have both coherent motion and colour. Dots were drawn with a limited lifetime; any one dot would be drawn for four frames then hidden for a further four frames.

**Pre-scanning Procedure and Difficulty Matching**

Naïve subjects were first trained on 100% coherent shapes from all four conditions until they felt confident about the task. The stimuli consisted of one of four shapes (Fig. 1i–iv). The subjects’ task was to indicate by pressing a button in a timely fashion whether the shape presented was vertically symmetrical or not (i.e. to discriminate between the square/cross and the T-shape/inverted T-shape). Subjects made the button press during the presentation of the shape and, if they failed to press the button during the presentation, the trial was discarded.

After training subjects were moved into a scanner (see scanner details below) where they viewed a screen onto which the stimuli were projected using an LCD projector directed onto an angled mirror. We first recalculated the equiluminant colours to be used during the experiment as described above. We then carried out a difficulty matching procedure to equate the two unimodal conditions in terms of performance. This was achieved by presenting shapes at various coherence levels between 0 and 100% in steps of 10%. Subjects viewed 220 trials covering each coherence value 10 times for C and M shapes.

We then constructed psychometric functions by averaging the number of correct shape discriminations at each coherence level. These curves were fitted with a Weibull function of the form: $y = a + b(1 - e^{-c})^e$, where $y$ is the probability of a correct response, $a$ is the probability of making a correct response by chance, $b = 1 - a$, $x$ is the percentage coherence of the shape, $d$ is the inverse of the central parameter, which corresponds to the coherence value at which the probability of making a correct response was halfway between 1 and chance, and $e$ is the steepness constant of the curve. The central parameter ($d^{-1}$) was taken as a measure of the centre of the function and was used to match the difficulty of the unimodal blocks during the scanning session in the following way. As the motion-defined stimuli were found to give higher $d^{-1}$ values than the colour-defined stimuli (i.e. they were more difficult to distinguish) the difference between the $d^{-1}$ parameters for the two conditions ($D$) was calculated as:

$$D = dm^{-1} - dc^{-1}$$

where $dm^{-1}$ is the central parameter of the motion-defined stimuli and $dc^{-1}$ is the central parameter of the colour-defined stimuli. The difficulty-matched coherence of the colour-defined stimuli ($C_{coh}$) was then calculated as follows:

$$C_{coh}(n) = 0.1 \times (n - 1) \times (100 - D)$$

where $n$ is the coherence level of the stimulus, which ranges from 1 to 11 in steps of 1. The coherence of the motion-defined condition ($C_{mot}$) was then simply calculated as:

$$C_{mot}(n) = C_{coh}(n) + D$$

The bimodal condition was created by combining two appropriate unimodal coherences, e.g. if $D$ was 10, a bimodal stimulus of coherence level 5 consisted of a colour shape of 36% coherence and a motion shape of 46% overlaid.
**Experiment 1 — Procedure**

The aim of experiment 1 was to localize the regions of the visual brain that are engaged when subjects view simple shapes and to determine which areas are activated by both colour- and motion-defined shapes. We used a block-based design consisting of five conditions (Figure 1i–vi): two shape conditions (C, M), two high-coherence conditions with no shape present (CC, MC) and one noise condition (N). The coherence levels used in the shape blocks were 100% for the M condition and the equivalent level of colour coherence needed to produce identical performance in the C condition as assessed by that subject’s psychometric function (see above). The same levels of coherence were used for the pure coherence blocks (CC and MC) as for the shape blocks, whereas 0% coherence was used in the noise blocks. The subjects’ task in the shape blocks was to fixate a central cross and perform the vertical-symmetry discrimination task, whereas in the non-shape blocks subjects fixated centrally and pressed a random button on each presentation. There were six blocks of each condition type, presented in a pseudo-randomized order. Each block lasted on average 15 s and contained eight presentations of shapes lasting 1.1 s each with a gap between presentations of, on average, 0.75 s. The gap was varied randomly between 0.5 and 1 s to produce some jitter between the onset of the shape and the slice acquired by the scanner.

**Experiment 2 — Procedure**

Experiment 2 was an event-related study of cue integration. It was designed to localize regions that are more active when subjects view bimodal shapes than unimodal shapes. This experiment was carried out directly after experiment 1 using the same subjects. It consisted of three conditions: the two unimodal conditions (C, M) and the bimodal condition (CM). The coherence of each shape presented was varied between 0 and 100% as in the difficulty matching procedure. The coherence values were chosen in a pseudo-random order so that 10 presentations of each coherence value were shown for each condition. The stimuli were grouped into blocks consisting of five shapes, each block containing stimuli of only one type (C, M or CM). Each shape presentation lasted 1.6 s and was followed by a period displaying just the central fixation cross for 0.9 s. The average block length was 12.5 s and the order of the blocks was again pseudo-randomized. In total the experiment consisted of 60 blocks, 20 blocks of each of the three conditions. The subjects performed the same shape discrimination task as before and again were required to press a button indicating their judgement within the 1.6 s shape presentation.

**Experiment 3 — Procedure**

Experiment 3 was an fMRI adaptation study of cue-invariance in the human visual system. It was designed to test whether similar levels of adaptation could be observed in the lateral occipital cortex (LOC) regarding of the cue that was used to define the shape or even if the cue that defined the shape was alternated. We hoped that this experiment might provide evidence for integration of different visual cues in a single population of neurons in the LOC. All stimuli were made using the same procedure as in experiment 1. There were six condition types, summarized in Figure 1viii–xiii. Briefly, 12 shapes were presented in each block, each presentation lasting 0.87 s and being followed by a period of 0.13 s containing just the fixation cross. Blocks were either adapting, when the same shape was repeatedly presented, or non-adapting, when the shape was varied between 10 alternatives. Each of the main block types lasted 12.5 s and was repeated four times; a shorter block of noise, lasting 3.5 s, separated the blocks (to allow for recovery from adaptation between blocks). The order of the blocks was pseudo-randomized (so that each block was displayed before a block-type was repeated). The subjects’ task was to fixate the central cross and press a button if they detected a small change in thickness in the fixation cross. These changes occurred on one sixth of the shape presentations.

**Scanning Details**

In experiments 1 and 2 subjects were scanned in a 2 T Magnetom Vision fMRI scanner with a head-volume coil (Siemens, Erlangen, Germany). A gradient echo-planar imaging (EPI) sequence was used to maximize blood oxygen level dependent (BOLD) contrast ($T_E = 40$ ms, repeat time 2.89 s). Each brain image was acquired in a descending sequence comprising 38 axial slices, each 1.8 mm thick with 1.2 mm gaps in-between, and consisting of 64 x 64 voxels. In total, 250 volumes were acquired in experiment 1 and 325 volumes in experiment 2; in both cases the first four were discarded to allow for $T_1$ equilibration effects. Each subject was also scanned after the main scanning session with a $T_1$-weighted structural sequence to obtain a high-resolution structural image.

In experiment 3 subjects were scanned in a 1.5 T Sonata scanner (Siemens) with a head volume coil. A gradient EPI sequence was used ($T_E = 50$ ms, repeat time 3.42 s). Each brain image was acquired in a descending sequence comprising 38 slices, each 2 mm thick with 1 mm gaps in-between, and consisting of 64 x 64 voxels. A total of 190 acquisitions was made and the entire scan lasted 12 min, with each subject taking part in two sessions. Each subject was again scanned after the main scanning session with a $T_1$-weighted structural sequence.

**Data Analysis**

The data were analysed using the SPM2 software (Wellcome Department of Imaging Neuroscience, London, UK). The echo planar images were spatially realigned to the first image and then resliced to produce a final voxel size of 2 x 2 x 2 mm. In experiments 1 and 2 these images were realigned in time using sinc interpolation as if every slice was acquired at the same time as the middle (19th) one. All images were then spatially normalized to the Montreal Neurological Institute template provided in SPM2, which approximates to the atlas of Talairach and Tournoux (1988). The images were then spatially smoothed with a Gaussian kernel of 10 mm full width at half maximum. The resulting images were temporally filtered using a high-pass filter with a low-frequency cut off of 1/400Hz to remove slowly varying scanner noise. Serial autocorrelations were modelled using an AR(1) method.

The shape-localizer study (experiment 1) was a simple block-based design, modelled using box-car regressors. Each box-car was convolved with SPM2’s canonical haemodynamic response function (HRF) and then entered into a multiple linear regression. The head movement parameters obtained whilst realigning each image, and the subjects’ button presses, were included in the analysis as events of no interest.

Experiment 2 was an event-related study in which we presented subjects with shapes of one of three conditions at 11 separate coherence levels. We modelled this experiment by grouping shapes of one condition and coherence value together to produce 11 regressors per condition, and 35 regressors in total for each subject. We then assessed the effects of cue integration by summing across coherence levels for each condition. To study the BOLD responses produced by shapes at different coherence levels (Fig. 4ii) we also grouped shapes into 4 coherence groups to increase the number of shape presentations in each regressor. The first group contained the 0% coherence stimuli across all three conditions, and the remaining groups each contained stimuli of two coherence values (the composition of these groups is shown in the key of Fig. 4ii). In all analyses, each shape presentation was modelled as a stick-function (essentially a boxcar of duration 1/16th of a TR) then convolved with the HRF and entered into a multiple regression analysis as above. The subjects’ reaction times, which were also measured, showed no significant differences between conditions. Head movement parameters were included in the analysis as events of no interest.

These analyses were carried out both at an individual subject level and the group fixed effects level whereby the reliability of the measurements was assessed in relation to the within subject variance (Friston et al., 1999).

Data from experiment 3 were analysed in a similar manner to those in experiment 1. Each block was modelled by a simple boxcar design to allow us to test the overall amplitude of the response in different conditions. We also used a more complex model in which the response during a block was modelled by three cosine functions rather than a single boxcar so that fewer a priori assumptions had to be made about the shape of the response from a particular voxel under adapting conditions. The cosine functions allowed a number of possible response shapes to be modelled so that the modelled response more closely resembled the actual response. The responses from different voxels were then assessed using an F-test that used contributions from any of the three cosine components. Results using this technique were
compared to the simpler boxcar design and were found to be qualitatively similar. We therefore report results from the boxcar design alone in this paper. Subjects’ button presses and head movements were modelled out of the data. The actual BOLD response of interesting voxels revealed by the above approach was assessed using a block triggered averaging approach. The raw signal was pre-processed in an identical fashion to the data used in the SPM analysis and was mean-corrected. The BOLD responses for each epoch type were then aligned to the start of each block and binned into 2.05 s bins. These values were then used to extract the average BOLD response across subjects from the beginning of the block until 26.7 s after block-onset.

As we were uncertain whether adaptation effects might be reliably found in individual subjects, we changed our approach from a fixed-effects analysis to a random effects approach (RFX) where activations were assessed in relation to their variability across subjects (Friston et al., 1999).

Results

Experiment 1: Shape Localising Block-based fMRI Study

We were first interested in localizing regions of the cortex that respond strongly to the simple-shape stimuli used in our study. We adopted a block-based localizer approach similar to the object-based localizer scans used in other studies (Malach et al., 1995; Grill-Spector et al., 1999; Avidan et al., 2002; Kourtzi et al., 2002, 2003b; Lerner et al., 2002; Murray et al., 2002). In these studies, areas of the brain responding strongly to objects were revealed by comparing activity when subjects viewed blocks of greyscale photographs (or line drawings) of objects to blocks of scrambled objects. We revealed shape-specific activity by comparing blocks of 100% coherent shapes defined by either colour or motion with blocks of just 100% colour or motion coherence with no visible shape present (to control for regions that might respond to high motion or colour coherence). The results are shown in Figure 2a–iii. In all subjects, we observed a region of lateral occipital cortex that responded significantly more strongly to shapes than high-coherence controls ($P < 0.05$, corrected for multiple comparisons). The Talairach coordinates of the centre of this region were (46, -72, -10) in the right hemisphere and (-44, -72, -10) in the left hemisphere. We compared the location of this region to sulcal landmarks in individual subjects and determined that it always fell around the occipito-temporal sulcus (OTS), largely on the lateral surface where the OTS meets the inferior temporal sulcus (ITS). The major focus of shape activity was ventral to V5, although some activity was found more posterior to it, extending back to the lateral occipital sulcus. The location of this region was consistent with the location of the LOC, known to give strong responses to objects compared to scrambled controls (Malach et al., 1995, Grill-Spector et al., 1998, 1999; Kourtzi and Kanwisher, 2001; Avidan et al., 2002; Lerner et al., 2002; Kourtzi et al., 2003b). We did not observe much shape-specific activity extending anteriorly into the fusiform gyrus.

The responses of the LOC in our study were equal to both colour- and motion-defined shapes, as has been reported elsewhere for other cues (Grill-Spector et al., 1998). We also observed reasonable shape responses in V5, as has been reported previously (Kourtzi et al., 2002), although these responses were significantly stronger to motion-defined shapes than to those defined by colour ($P < 0.05$, corrected). Further regions giving strong shape responses included the superior parietal lobe and the superior occipital lobe, possibly at the location of V3A (all $P < 0.05$, corrected). This area has also been previously reported to respond to objects (Grill-Spector et al., 1999; Kourtzi and Kanwisher, 2001; Avidan et al., 2002; Lerner et al., 2002; Murray et al., 2002).
Apart from V5, the responses to colour or motion-defined shapes were equal throughout this network of areas.

Experiment 2: Cue Integration Event-related fMRI Study

In this experiment we used fMRI to localize regions that are activated more strongly by shapes defined by both colour and motion than by shapes defined by either cue in isolation, i.e. areas that may bring together form information from different submodalities. We presented subjects with shapes defined by one cue (C or M shapes) or shapes defined by two cues (CM shapes). We presented shapes at every coherence level between 1 and 11. The subjects’ task was to determine if the shape was vertically symmetrical, that is, to discriminate between the square/cross and the T-shape/inverted T-shape (Fig. 1i–iii).

A true integration area should bring together information from both cues, as well as responding to each cue in isolation. It should therefore be significantly more active in the bimodal condition than either the C or M unimodal conditions (averaged across all coherence values). To reveal such areas we used a conjunction analysis (Nichols et al., 2004) to search for voxels that show significant effects in both the [CM-C] and the [CM-M] comparisons. This type of analysis can be thought of as a logical AND operator, since if a voxel is significant at a \( P < 0.001 \) uncorrected level, then both parts of the conjunction would be deemed to have reached this level of significance, i.e. the probability of observing this result by chance was \( 0.001^2 = 10^{-6} \). This is therefore a very conservative technique.

Our a priori hypothesis was that the regions revealed to be shape-selective in experiment 1 were most likely to be involved in the integration of colour and motion shape information. We therefore constructed a region of interest (ROI) consisting of the regions that were significantly activated by shapes compared to coherence at the \( P < 0.05 \) level (corrected) in experiment 1 using MARSBAR (Rorden and Brett, 2000) and restricted our search to these regions. This ROI contained bilateral LOC, V3A, V5 and the superior parietal lobe. The conjunction analysis revealed that out of all these regions only one distinct locus, centered on coordinates (44, -66, -18) in Talairach space (Fig. 2iv), was significantly more active in the bimodal condition than in either of the two unimodal conditions (\( P < 0.05 \), corrected). The coordinates of this region fell within the right LOC. A previous study (Grill-Spector et al., 1999) divided the LOC into two subdivisions, a dorsal-posterior region called LO and a ventral-anterior region entitled the LOa/pFs. Based on its coordinates, our activation was found to be overlapping but slightly ventral to the region defined as LO (Malach et al., 1995; Grill-Spector et al., 1999).

We compared the location of our activation with the shape-selective regions revealed in the shape localizer study in two representative subjects (Fig. 2v). In both, the integration-related activity fell partially within the shape-responsive area, although it extended into the anterior-ventral margins of this region. The pattern of activation was quite similar in both subjects, falling on the anterior bank of the OTS. In subjects CB and LD we can therefore demonstrate a region of the LOC that responds strongly to shapes, gives equal responses to colour- and motion-defined shapes and also shows significant increases in activity in the bimodal condition. This region lies partially in the shape-responsive LOC, but extends in a more anterior-ventral direction onto the anterior bank of the OTS.

To ensure we did not exclude other interesting activations outside of shape-selective regions we widened our search to include the whole brain but no further regions were revealed, even at uncorrected significance levels (\( P < 0.001 \)). The increased activation of the LOC under bimodal conditions was quite reliable across subjects (Fig. 3). We looked at the responses of the most significantly activated voxel in the LOC and found that in four out of five subjects it was more active in the bimodal condition than in either the C or the M conditions. One subject showed equal activations across all three conditions.

We obtained psychophysical data from each subject during their scans (Fig. 4i). Subjects correctly identified more shapes in the bimodal condition than in the unimodal conditions (380 shapes for the CM condition versus 323 for M and 314 for C), reflecting the perceptual advantage enjoyed by bimodal over unimodal shapes. We studied the BOLD responses from the voxel showing the most significant integration-related activity (Fig. 4ii) and found that they were qualitatively similar to the psychophysical performance of the subjects at different coherence levels, as has been reported previously (Grill-Spector et al., 2000). Because subjects identified more bimodal shapes, the extra activity we observed in the LOC under this condition could possibly have been due to improved image segmentation and/or increased attentional modulation in the bimodal condition. To ensure that this effect was not contributing to our results, we analysed each condition separately according to the response of the subject. We created two regressors for each condition, one containing the correctly identified shapes and one containing the shapes that were not identified by the subjects. We then restricted our analysis to shapes that were correctly identified, thus ensuring both that image-segmentation processes must have been successfully completed and that levels of attentional modulation were balanced across conditions. This analysis did generally reduce the increased activity in the bimodal condition in the LOC, but it was still significantly more active in the CM condition than both the C and M conditions, as assessed by a conjunction analysis at the same statistical threshold as used above. This indicates that the extra

![Figure 3. BOLD responses from the LOC. Parameter estimates from the most significantly activated voxel in the LOC at (44, -66, -18). These values are the amount by which each individual regressor was scaled to best fit the data. In this sense they represent an estimate of the mean BOLD signal change produced by an event. Here we show the difference in parameter estimates between the bimodal condition and the unimodal condition (black) and the unimodal coherence condition (grey). These values are shown here to indicate the reliability of the increased activity in the bimodal condition across subjects. Responses from five individual subjects are shown; four out of five had greater levels of activity under the CM condition than either the C or M conditions.](image-url)
activity seen in the LOC under the CM condition was not due to improved image-segmentation or attentional amplification of the signal. Further support for this view comes from the fact that, whilst shape responses are seen bilaterally in the LOC, the integration-related activity was strongly lateralized to the right hemisphere. If our result had been due to attentional amplification of the signal, we should have observed bilateral increases.

**Experiment 3: fMR Adaptation Study of Cue Invariance**

In experiment 1 we observed that the LOC (among other areas) gave equal responses to both colour- and motion-defined shapes. Similar responses in LOC have also been observed for objects defined by other cues (Grill-Spector et al., 1998). In experiment 2 we observed that a sub-region of the LOC gave increased responses to bimodal over unimodal shapes. These findings point towards a sub-region of the LOC as being particularly involved in the integration of colour and motion signals to define shapes and also raises the question of whether this integration occurs at the level of individual neurons in the LOC (which would make them cue-invariant cells), or whether the LOC contains a mixture of two separate neuronal populations, each responding to shapes defined by specific cues (we call such an area cue-variant). In the latter case a bimodal shape would activate both populations of neurons. Given the poor spatial resolution of the BOLD signal, a cue-variant area could give exactly the same response as an area containing cue-invariant neurons. This would make it difficult to discriminate between a single population of neurons responding more strongly and a mixed population of neurons responding in greater numbers.

To determine whether the increased response in the LOC in experiment 2 was due to integration of colour and motion information in a single neural population we used an fMR adaptation paradigm (Grill-Spector et al., 1999). We presented subjects with blocks of the same stimuli used in experiment 1 but this time divided into two main types. In adapting blocks the shape presented was always the same for the entire duration of the block, whereas in non-adapting blocks the shape was varied with every presentation. In both the adapting and the non-adapting blocks we varied the direction of motion or the colour of the dots inside the shape to prevent adaptation in the areas specialized for the detection of the cues themselves (see Fig. 1vii,viii for details of these conditions). If an area contains neurons that respond specifically to a shape regardless of the cue from which it is derived (i.e. cue-invariant neurons), then such neurons should adapt if the same shape is repeatedly presented, even when the cue that defines it is varied. We tested this view by using blocks containing only colour-defined stimuli (AC, adapting colour), only motion-defined (AM, adapting motion) and a condition in which the cue that defined the shape alternated between colour and motion (AA, adapting alternating). Each of these blocks had an accompanying non-adapting control condition in which the shape was varied (NC, NM and NA respectively).

A truly cue-invariant area should show similar levels of adaptation for repeated presentations of colour, motion, and alternating colour and motion shapes since the underlying neurons should be blind to variations in the cue. We searched for such areas using a conjunction analysis, looking for areas that showed reduced responses in all the adapting conditions compared to their relevant non-adapting controls. As we had a specific *a priori* hypothesis that the ventral part of the LOC would contain cue-invariant neurons, we used a small search volume correction of a 10 mm sphere centred on the coordinates of the LOC activations obtained in experiment 2. In this experiment we assessed the reliability of our findings by comparing their magnitude to their variability across subjects in an RFX analysis. We found that a region of the LOC, with a peak at (52, −68, −14), showed a significant conjunction (*P* < 0.05, small volume corrected, RFX). The BOLD responses from this voxel, averaged over subjects, show an obvious adaptation effect in the AA block but not in the NA block, suggesting that this area does contain shape-selective, cue-invariant neurons (Fig. 5ii). The location of this region in three individual subjects
Results from the fMR adaptation study. (i) Parameter estimates from voxels from two different areas. The LOC data comes from the voxel with the most significant decreases in the AA condition compared to its non-adapting control, NA. The ventral LOC data (LOCv) comes from the voxel with the maximum difference between the AA condition and both the AM and AC conditions as assessed by a conjunction analysis. The three colours represent the three main condition types (blue = alternating cue, red = motion defined, green = colour defined). Darker colours represent the non-adapting conditions, whereas lighter colours represent the adapting conditions. The error bars are the mean standard errors within subjects. (ii) Plots showing the average BOLD response across all subjects aligned to the start of each block (see Materials and Methods) from the same voxels as shown above. The error-bars show the random effects standard errors across subjects. These plots are shown to demonstrate the variability of the cue-variant adaptation across all subjects and were not significantly different from one another (Fig. 5i). This is inconsistent with the attentional modulation view and suggests that this region may contain a mixture of form-selective neurons responding to specific cues. We checked the location of this region in single subjects (Fig. 5ii) and found that, in some, it partially overlapped with the region showing cue-invariance (BR), whereas in others there was no overlap between the cue-variant and cue-invariant regions (PL) or simply no evidence for any cue-variant region (KG). The variability of the cue-variant region means we cannot be entirely confident of its exact location. However, we can state that the region with the strongest cue-variant adaptation across all subjects was more ventral to the site showing the strongest cue-invariance.

**Discussion**

**Cue Invariance in the Primate Visual System**

Electrophysiological studies in monkeys have found cue invariance at the very earliest stages of cortical visual processing. Studies have shown that V1 and V2 neurons are capable of responding to an oriented border regardless of what cue is used to define it (Leventhal et al., 1995, 1998; Zipser et al., 1996) and that neurons in the inferior temporal cortex are largely cue invariant (Sary et al., 1993, 1995). Similarly, studies of V3/V3A suggest that neurons can respond in the same specific way to multiple cues but sometimes prefer one cue over the others (Zeki et al., 2003). In contrast, imaging studies in humans have found limited cue invariance in early visual areas. Grill-Spector et al. (1998) found that, in addition to the LOC, V3A and possibly ‘KO’ showed identical responses to objects defined by motion, luminance and texture. Similarly Zeki et al. (2003) found that KO gave identical responses to shapes defined by either colour or motion. In our simple-shape localizer study (experiment 1) we found a network of areas including the LOC, V3A and parietal areas that gave equal responses to shapes defined by either colour or motion. All of these imaging experiments, including our experiment 1, suffer from the problem that they compare activity during the presentation of shapes to some non-shape-containing control. Therefore regions activated by attention-to-shapes will also be revealed in this study. This may have given a false impression of the degree of cue-invariance in the human visual system. Our study is the first to use fMR adaptation to investigate cue invariance. This

is shown in Figure 5iii and was consistently found on the lateral surface in the OTS extending onto the posterior and anterior banks of this sulcus.

We also searched for areas that might contain a mixture of neurons responding to shapes defined by only one cue. We call such areas cue-variant. We would expect such areas to adapt only if presented with shapes defined by the same cue. If the cue is alternated, adaptation should not occur, as the same neurons are no longer being driven by every stimulus. To search for these regions we used a conjunction analysis to reveal areas that were more active in the AA condition than both the AM condition and the AC condition (at $P < 0.001$, uncorrected, RFX). This revealed a bilateral, more ventral portion of the LOC with maxima at (−42, −66, −24) and (40−70 −24), just posterior to the region termed LOa/pFs by Grill-Spector et al. (1999). The responses from this region (Fig. 5i) suggest that it contains a mixture of cue-variant neurons, but this response could also be explained by attentional modulation. If attention increases the responses of the region only in the alternating condition then the contrasts [AA− AM] and [AA− AC] would become significant. If this is the case, then the NA condition should also produce more activity than either the NM or the NC conditions. We examined the parameter estimates from this region and found that the responses in the non-adapting control conditions were not significantly different from one another (Fig. 5i). This is inconsistent with the attentional modulation view and suggests that this region may contain a mixture of form-selective neurons responding to specific cues. We checked the location of this region in single subjects (Fig. 5ii) and found that, in some, it partially overlapped with the region showing cue-invariance (BR), whereas in others there was no overlap between the cue-variant and cue-invariant regions (PL) or simply no evidence for any cue-variant region (KG). The variability of the cue-variant region means we cannot be entirely confident of its exact location. However, we can state that the region with the strongest cue-variant adaptation across all subjects was more ventral to the site showing the strongest cue-invariance.
The relatively new technique of fMR adaptation allows the poor spatial resolution of fMRI to be side-stepped by targeting specific groups of neurons within a particular area (Grill-Spector et al., 1999; Grill-Spector and Malach, 2001). This has commonly been achieved by repeatedly showing the same stimulus to subjects and measuring the amount of adaptation (i.e. the reduction in response) compared to a non-adapting control, in which a particular attribute of the stimulus is varied (Grill-Spector et al., 1999; Kourtzi and Kanwisher, 2001; Avidan et al., 2002; Kourtzi et al., 2003a,b). We found that a sub-region of the LOC, just ventral to LO, adapted if the cue that defined the shape alternated between colour and motion. The level of adaptation produced in this region by shapes defined by alternating cues was the same as that produced when the cue remained constant. This raises the possibility that this region does contain cue-invariant populations of neurons and provides evidence in favour of the supposition that the extra activity seen in the same region in the bimodal condition in experiment 2 was due to cue integration rather than the summed response of separate populations of cue-variant neurons. Though likely in view of other evidence reported here, the above interpretation is not conclusive. One could, for example, envisage an area containing a mixed population of neurons that still adapt to their preferred stimulus even if the intervening stimulus is non-preferred. Such a population could produce the responses we observed in our study. Even so, one would have to account for the difference in response to adapting stimuli between this region and the neighbouring region lying just posterior to the LOa/pFs. The responses from this latter region are consistent with a mixed population of colour and motion-form selective neurons. Here adaptation is evident only if the form is defined by one cue alone, and absent if defined by alternating cues. So it seems in this region that adaptation cannot be maintained across a stimulus. This strengthens the possibility that the adaptation we observe ventral to LO is due to adaptation in a cue-invariant population and not to maintained adaptation over an intervening stimulus in a mixed population. Overall, although we cannot conclusively state that the LOC contains cue-invariant populations of neurons, we believe the balance of evidence (increased responses to bimodal stimuli, identical levels of adaptation regardless of the defining cue and differences in responses to adapting stimuli between two contiguous subdivisions of the LOC) points in its favour.

One final point of interest is whether the region showing cue-invariant type responses in this experiment overlaps with the region showing integration-related activity in experiment 2. From the coordinates given here it appears that the peak of the integration-related activity is located between the peaks of the cue-invariant and cue-variant regions. It is therefore possible that the integration-related activity revealed in experiment 2 contains contributions from both the cue-invariant region and the cue-variant region. We can, however, draw much stronger conclusions about the contribution of the cue-invariant region, as it was so much more reliably located across subjects than the cue-variant region. Further work would be required to be certain of the exact location of the cue-variant region in individual subjects.

**Cue Integration in the Primate Visual System**

There have been very few previous studies of the integration of shape information from different sub-modalities. One study has shown that cue-invariant cells in V2 show stronger responses to oriented bars defined by more than one cue (Leventhal et al., 1998) suggesting that cue integration may begin as early as area
V2 in the monkey. Other than this the mechanisms of cue integration in the primate visual brain are unknown. Our study shows that a sub-region of the LOC gives a greater response if a shape is defined by more than one cue at the same time. We propose that cue-invariant neurons in this region are the likeliest locus of integration of colour and motion form cues in our study. This result is consistent with previous views of the function of the LOC, which have suggested that it integrates local border cues together to form a representation of the global shape (Lerner et al., 2002), although two recent studies have suggested that early visual areas may also play a role in this process (Altmann et al., 2003; Kourtzi et al., 2003).

It is interesting to note that we find no integration-related activity in V2, unlike the studies of Leventhal et al. (1998) in monkeys and cats. There are several differences between these two studies, not least the difference in species and experimental technique that might account for the different result. Importantly, the large shape stimuli used in our experiment are quite poor stimuli for activating early visual areas (Murray et al., 2002). It may be that only the LOC contains receptive fields of a suitable size to process such large shapes. This may also explain why shape-related activity was observed in a large region of cortex extending posterior to V5, whereas integration-related activity was only observed at the very anterior margins of this region. Psychophysical results from our subjects indicate that cue integration improves performance mainly at mid-range coherence values. At these coherence levels the amount of form information available is limited and therefore needs to be summed from a larger region of space to allow correct discrimination of a border and consequently a shape. Therefore integration-related activity should be found in regions where receptive fields are larger than those absolutely necessary to detect high-coherence shapes. As receptive fields tend to become larger as one progresses from posterior to anterior visual areas (Smith et al., 2001) this may be the reason why integration-related activity was not found in posterior parts of the LOC.

It is interesting to speculate whether the focus of the integration-related activity would shift to earlier visual areas or whether it would remain in the LOC if the stimuli were changed to a simpler (or higher spatial frequency) form. This raises the question of whether the LOC acts as a general site for cue integration for all shapes, with more anterior regions receiving this cue-integrated shape information as the basis for their own processing, or whether the integration site for different shapes and objects will be found in the area that contains cells with receptive fields of suitable complexity for processing that object (e.g. the fusiform face area will be the site of integration for cues which define faces). We believe the latter is more likely as it would be consistent with the general rule that visual areas can draw on any source of information they need to perform their task within anatomical constraints (Zeki and Shipp, 1988). We therefore propose that the brain uses a simple solution to the problem of integration of colour and motion form cues in which the area that performs the integration is the same as the one that is specialized for processing the current visual object. Using this logic we predict that the integration of form cues other than colour and motion (depth, luminance, etc.) should also lead to increased responses in the area specialized for processing the current visual object. Equally, if colour and motion are used as the sources for constructing other visual stimuli, e.g. faces, then the integration related activity would shift to the cortical area(s) specialized for the processing of faces.

Conclusion

The human visual system is able to integrate the different visual cues that define a shape. For shapes defined by colour and motion we suggest that this integration process is undertaken by the LOC and this allows the visual system to use multiple sources of information available to it about a shape in the visual scene. We propose that cue-invariant neurons in the LOC are able to integrate across different cues and are therefore activated to a higher degree by shapes defined by more than one cue. Given the relationship between the strength of activity in the LOC and the psychophysical performance of the subject we propose that it is this increased activity of neurons in this region that results in the perceptual advantage enjoyed by bimodal shapes over unimodal ones. This integration process therefore plays an important role in the remarkable ability of our visual system to rapidly discriminate an object from its background and must also contribute to the uniform, seamless nature of our perception of objects in the visual world.

Notes

We would like to thank John Romaya for his technical assistance and Drs. Andreas Bartels and Stewart Shipp for their helpful comments on this manuscript. This work was supported by a grant from the Wellcome Trust, London.

Address correspondence to Matthew W. Self, Anatomy Department, Wellcome Department of Imaging Neuroscience, University College London, Gower Street, London WC1E6BT, UK. Email: m.self@ucl.ac.uk.

References


