Optic flow is one of the most important sources of information for enabling human navigation through the world. A striking finding from single-cell studies in monkeys is the rapid saturation of response of MT/MST areas with the density of optic flow type motion information. These results are reflected psychophysically in human perception in the saturation of motion aftereffects. We began by comparing responses to natural optic flow scenes in human visual brain areas to responses to the same scenes with inverted contrast (photo negative). This changes scene familiarity while preserving local motion signals. This manipulation had no effect; however, the response was only correlated with the density of local motion (calculated by a motion correlation model) in V1, not in MT/MST. To further investigate this, we manipulated the visible proportion of natural dynamic scenes and found that areas MT and MST did not increase in response over a 16-fold increase in the amount of information presented, i.e., response had saturated. This makes sense in light of the sparseness of motion information in natural scenes, suggesting that the human brain is well adapted to exploit a small amount of dynamic signal and extract information important for survival.

Keywords: optic flow, fMRI, MT/MST, natural scenes, local motion


Introduction

The main brain areas specialized for visual motion processing, MT and MST (together also known as MT+ or V5), are well studied in monkey single-cell experiments (Duffy & Wurtz, 1991; Felleman & Kaas, 1984; Snowden, Treue, Erickson, & Andersen, 1991; Zeki, 1974) and human fMRI (Amano, Wandell, & Dumoulin, 2009; Bartels, Zeki, & Logothetis, 2008; Rees, Friston, & Koch, 2000; Smith, Wall, Williams, & Singh, 2006; Tootell et al., 1995; Zeki et al., 1991). However, the parallels across species as well as the connection between single-cell and population-level effects are not clear. Duffy and Wurtz (1991) investigated the response of neurons in monkey MST, an area selective for the visual motion patterns formed by the observer’s motion through an environment (optic flow). Using motion patterns formed by randomly placed dots, they found that increasing the dot density did not increase the response of individual neurons. A similar result exists for area MT (Snowden et al., 1991), and this saturation is also reflected in psychophysical experiments that have shown the magnitude of the motion aftereffect (Curran & Lynn, 2009) and heading estimation performance (Warren, Morris, & Kalish, 1988) saturates at low dot densities.

It is known that in area V1 BOLD response measured by fMRI increases with contrast (Tootell et al., 1995), in line with the selectivity of single neurons in this area for spatial changes in luminance. In the same study, it was found that area hMT+ was selective for motion (showing no response to stationary stimuli) but saturated rapidly with contrast level and did not increase its response over a wide range of the contrast of the grating stimulus used. A further fMRI study has shown that area hMT+ increases its response with the coherence (signal-to-noise ratio) of the moving stimulus (Rees et al., 2000), paralleling macaque single-cell data (Britten, Shadlen, Newsome, & Movshon, 1993). Less clear is how response scales with the amount of coherent motion per se, i.e., how much information is necessary to extract coherent motion in a more realistic setting, where the amount of coherent motion does not scale inversely with the amount of noise.
Stimuli such as dynamic natural scenes that cause large responses in the motion-sensitive areas of the human visual cortex provide the best opportunity to discern changes in activity. However, it can be difficult to quantify and control the amount of motion in these scenes. In a previous attempt, Bartels et al. (2008) used Hollywood movie clips with free viewing, using template matching to extract estimates of local and global motion to measure correlations with neural activity, finding that V5+ correlated better with local motion and areas further along the visual processing stream in the medial posterior parietal cortex correlated better with global motion.

We examined the dependence of BOLD response in visual areas on the amount of local motion in dynamic scenes, generated by moving forward through everyday environments. This is the most common type of motion experienced by humans and a motion pattern that MST appears to be specialized for (Duffy & Wurtz, 1991; Smith et al., 2006). In the first experiment, we used a simple, biologically motivated, computational model to quantify the amount of local motion across each movie clip — the motion energy content (as calculated by spatiotemporal correlation) — and examined how the magnitude of the cortical response depends on motion signal content. The 2DMD model (Zanker, Srinivasan, & Egalhaaf, 1999) has been successful in analyzing similar forward natural motion (Zanker & Zeil, 2005) and simulating psychophysical results (Zanker, 1997, 2004; Zanker & Braddick, 1999; Zanker, Hermens, & Walker, 2010). Motion correlation is formally equivalent to motion energy (Adelson & Bergen, 1985) under a range of conditions (Hildreth & Koch, 1987) and so we do not expect our results to be specific to the model used. We also tested the hypothesis that beyond the local motion output, familiarity may also exert a top-down influence; inverting the contrast polarity of the scenes reduces familiarity while preserving motion characteristics. Finally, we produced novel stimuli that used the same clips and manipulated how much visual motion information was presented. To this end, we manipulated the proportion of the visible area by masking these clips with an opaque gray layer (similar to Kane, Bex, & Dakin, 2011) and varying the number of hard-edged circular apertures through which they could be seen. This restricted all visual information, also resulting in a decrease in the amount of available motion signals, the attribute for which MT and MST are selective. We varied the number of apertures, the size of the apertures, the spatial arrangements of visible local motion, and the overall image contrast and also tested response to counter-phase flickering images. In this way, we ascertained how human visual response varies with the amount of dynamic information, i.e., coherent or incoherent motion, or flicker, in natural scenes.

### General methods

#### Movie clips

Twelve AVI movie clips were recorded with a Panasonic 3CCD miniDV tape camera at a shutter speed of 1/50 s, with no compression or motion stabilization. The field of view was approximately 40° and the movie was viewed...
from the center of projection. The camera was attached to a trolley with wheels and sat 67 cm above the ground. The trolley was manually pushed along smooth wooden tracks with edges on to minimize side-to-side movement. The length of the track the trolley was pushed along was 200 cm, which took approximately 10 s, resulting in an average velocity of 0.2 m s\(^{-1}\). Movies were recorded at 25 frames s\(^{-1}\). The track could be moved to whichever location was needed. Movies were recorded indoors in office buildings and outdoors in both man-made (built) and natural (wooded) environments. The movies were then converted into grayscale images for input into the motion model and for presentation in the scanner (see Figure 1 for examples of single frames from these sequences and Movie 1 for a sample clip). Each frame was normalized to equal maximum contrast for each of the movies. There were 12 different grayscale movie clips of forward motion through natural scenes.

Experiment 1—Correlating motion energy content to cortical BOLD signal

We measured the magnitude of the BOLD signal change in human cortical areas V1, MT, and MST in response to each movie clip presentation and correlated this measurement with the overall local motion energy contained in these clips (averaged over each pixel and each frame), as measured with a two-dimensional motion detector (2DMD) model.

Methods

Motion analysis (2DMD simulations)

The 2DMD model is contrast dependent and can be varied in spatial (sampling distance) and temporal (time constant) scales to calculate the spatiotemporal correlation in a range of frequency channels. For this experiment, we used a fixed scale to consider relative magnitudes of local motion correlation compared between scenes. The following default 2DMD model settings were used (see Zanker et al., 1999 for a detailed description of these parameters) — input filter gain: 4, elementary motion detector (EMD) sampling base: 4 pixels, Difference of Gaussian (DoG) center width: 0.5 pixel, EMD time constant: 6.0 frames, DoG center gain: 1.0, EMD output gain: 2, EMD balance: 1. These parameters were chosen after some piloting to find the largest motion signal, i.e., choosing the spatial and temporal parameters that best match those of the clips, resulting in a relatively long time constant matched to the slow speed of movement forward. The model produced components of motion correlation magnitude (equivalent to motion energy) computed in the horizontal direction (hor) and in the vertical direction (ver) at each pixel for each frame, from which the direction and magnitude of the local motion signal can be derived. The first 6 frames of output are ignored in the analysis as there is some delay in the motion detector response. We calculated the magnitude of the motion signal as \(\sqrt{\text{hor}^2 + \text{ver}^2}\) at each pixel for each frame (within the boundary limits set by the size of the motion filters) and summed all these for each movie clip.

MRI acquisition and analysis

MRI BOLD images were obtained using a 3T Siemens Trio scanner and all data were pre-processed and analyzed using BrainVoyager QX (version 1.4; Brain Innovation, The Netherlands). Functional data were acquired using an 8-channel array head coil and a standard gradient-echo echoplanar (EPI) sequence (35 slices, 3-mm isotropic voxels, TR = 2.5 s, TE = 31 ms, flip angle = 80°, bandwidth = 1396 Hz/pixel). EPI data were corrected for head motion and slice timing and were filtered with a temporal high-pass filter of 0.014 Hz. No spatial smoothing was applied to avoid blurring the boundary between MT and MST. The functional data were also coregistered to a high-quality anatomical image to enable extraction of data from the V1–V4 and MT/MST ROIs (see details below). ROI-GLM analyses were performed on data from each participant and ROI (concatenated across all runs for each participant), using a standard hemodynamic response function. Participant head movement parameters derived from 3D motion correction were included as regressors in all statistical models. The amplitude of the fitted response was expressed as percent signal change from baseline response (as a more intuitive indicator of neural response) and was our dependent variable.

Region-of-interest (ROI) definition was performed in a separate session from the experiment, wherein an anatomical (3D, T1-weighted) scan was acquired. A sequence (MDEFT; Deichmann, Schwarzbauer, & Turner, 2004) that gives exceptional contrast between white and gray matter was used to facilitate segmentation and flattening of the gray matter (176 axial slices, in-plane resolution = 256 × 256, 1-mm isotropic voxels, TR = 7.92 ms, TE = 2.45 ms, flip angle = 16°, bandwidth = 195 Hz/pixel). MT/MST ROIs were defined based on the presence of ipsilateral responses in MST when presented with random dot stimuli on one side of fixation, followed by the other side, as used previously by several authors (Dukelow et al., 2001; Huk, Dougherty, & Heeger, 2002; Smith & Wall, 2008; Smith et al., 2006; Wall & Smith, 2008). MST was defined as all contiguous voxels that were significantly active during ipsilateral motion stimulation. MT was defined as all contiguous voxels that were active during contralateral but not ipsilateral stimulation. In addition, since previous research (Huk et al., 2002; Smith et al.,...
has shown that the center of MST is located anteriorly with respect to the center of MT, any MT voxels situated further anterior than the median value of the MST ROI on the horizontal (axial) plane were removed from the MT ROI. Retinotopic data were analyzed by fitting an estimate of the BOLD response to the time course obtained with a rotating wedge stimulus. This estimate was a rectangular wave of appropriate duty cycle reflecting when the stimulus entered a particular portion of the visual field, convolved with a standard hemodynamic impulse response function. The phase of the fitted response was taken as an index of visual field location in terms of polar angle. Reversals of the direction of phase change across the cortical surface were taken as boundaries of visual areas (Sereno et al., 1995). ROIs (visual areas V1–V4) were segmented by eye based on these boundaries viewed on a flattened version of each hemisphere of each participant’s reference anatomy.

Stimulus presentation

In the experimental session, 7 functional scanning runs were acquired, each one lasting 5 min and 7 s. We used an event-related design in which 2-s clips were interspersed with a blank gray background (48 cd m \(^{-2}\)) of random duration between 5 and 16 s. Each of the 12 clips was shown once per run and once with an inverted contrast profile. The movies were projected on the back of a transparent screen, using a Sanyo projector at a refresh rate of 60 Hz, resulting in the clips being shown at around double speed as recorded, mimicking a faster forward speed. The size of each frame was 24.5° × 33.9° and the screen size was 31.6° × 45.2° (the movies appeared on the gray background). Participants viewed the stimuli through a mirror from 11.5-cm viewing distance. The minimum black contrast was 1.6 cd m \(^{-2}\) and the maximum white was 300 cd m \(^{-2}\). Clips were 2 s long (120 frames at 60 Hz). There was no other light source inside the scanning room. Mean luminance levels (approximately 50 cd m \(^{-2}\)) did not vary substantially throughout the clips (less than ±4% from overall mean of clip on any given frame) as there were no dramatic changes in lighting conditions and the scenes depicted remained very similar.

Procedure

Six volunteers participated (including the first author, all others naive to the purpose of the study, 5 females). Participants were required to fixate a central square and report how many times it turned blue over each run to ensure equal attention allocation and maintenance of fixation over the course of each scan (accuracy was recorded and used as an exclusion criterion). All of the experiments reported in this article were approved by the local ethics committee.

Results

For all participants, there was significant visual activation. Conventional local motion models such as the 2DMD do not predict a difference between the normal and inverted contrast conditions, and indeed, we found no difference in activation between these two conditions in V1 or MT/MST. Because of this, we used the averaged percent signal change in BOLD response across the two contrast conditions in further analyses. In Figure 2a, we plot the mean BOLD response for each participant over all clips, on separate graphs for V1, MT, and MST. There is some variation between participants, but all the above brain areas in all participants show a significant response to the movie clips. In Figure 2b, we remove the variation between participants as we did for our statistical analysis of the data and plot the BOLD response for each participant for each clip, normalized to the mean response of that participant over all the clips, as a function of the overall summed motion energy of each clip as estimated by the 2DMD model. Area V1 shows a significant correlation between the overall local motion magnitude estimate (summed over each spatial location and each frame) and the BOLD response, whereas areas MT and MST do not. Analysis of covariance was used to obtain correlation coefficients, while removing the variation due to subjects, as used by Bland and Altman (1995). This gave us a significant correlation in V1 (\(r = 0.64, p < 0.0005\)) but not in MT (\(r = 0.16, p = 0.19\)) or MST (\(r = 0.11, p = 0.37\)). Such a correlation is expected in V1, regardless of how well this area responds to motion, as the 2DMD model is contrast dependent. Although overall contrast was kept constant on a frame-by-frame basis, V1 will respond more to clips containing more local contrast, for instance, the more sparse indoor scenes have less local contrast. After correcting the results for individual variation, the correlation in V1 accounts for a large proportion of the variance, yet this is not evident in MT and MST, even though both areas respond robustly to these stimuli. This suggests that in this case hMT+ is not scaling with contrast or the amount of local motion.

Experiment 2—Manipulating the amount of visible information

One of the problems with natural images as experimental stimuli is that they are usually limited to a rather restricted range of stimulus parameters and are not well-controlled stimuli. Hence, there may be only a very limited range of overall motion magnitudes in the sample clips used so far. In Experiment 2, we systematically manipulated the amount of motion information shown from these
clips, creating large differences in the visible proportion of the scene, which also generated large differences in the overall combined amount of motion signals presented to participants. Additionally, we presented a rearranged version of scenes in which local motion was preserved, but global motion was eliminated, similar to the “scrambled” manipulation of artificial stimuli in Kane, Bex, and Dakin (2009).

Methods

MRI acquisition and analysis

fMRI data were collected as in Experiment 1 but using a custom-built posterior 8-channel array coil (23 slices, 3-mm isotropic voxels, TR = 2.5 s, TE = 31 ms). An MP-RAGE anatomical scan (Siemens, Germany) was collected at the beginning of each scanning session to enable coregistration into the MDEFT anatomical space (see above). Functional regions of interest (ROIs) were defined (as in Experiment 1, see Methods section) in a separate session for V1, V2, V3, V4, MT, and MST. ROI-GLM analyses were performed on data from each participant and ROI as in Experiment 1.

Stimulus presentation

Projector and luminance settings were the same as in Experiment 1. We presented 6 of the 12 grayscale movie clips from Experiment 1, in an event-related design. The 2-s clips were interspersed with a blank gray background (48 cd m$^{-2}$) of random duration between 5 s and 16 s. For each clip, we showed versions viewed through 10, 40, and 160 circular, non-overlapping hard-edged apertures (aperture diameter: 1.1), which were placed at random locations for each presentation. In a fourth condition, we “cut out” the motion visible through the 160 apertures and randomly rearranged the apertures, so that local motion was preserved, but no global motion associated with forward movement remained (“scrambled” condition). Each clip was shown once in each of the aperture conditions, making 24 clips per run, lasting 5 min and 8 s. There were 6 runs per participant, recorded in a single scanning session. Sample frames taken from the same clip under different conditions are shown in Figure 3 (Movies 1–4). For each clip on each trial, a different set of randomized locations was chosen for the 160 apertures. The 10 and 40 aperture conditions were a randomly chosen subset of these apertures, so that visual information is only removed and not added. As we compare the conditions over...
all clips, this makes $6 \times 6 = 36$ instances of each aperture condition, so we can assume that there will be negligible differences on average in the relative amount of peripheral versus central sampling between conditions. The average eccentricity of the apertures was 21° for each condition.

**Procedure**

Six volunteers participated (including the first author, all others naive to the purpose of the study, 4 females), and the same task was given and used as exclusion criterion as in Experiment 1.

**Results**

Mean percent signal change values across all participants for each region of interest (ROI) are shown in Figure 4. These values show the size of the variation of the BOLD response as a function of stimulus condition. For all participants, there was significant visual activation. It can be clearly seen that in areas V1, V2, V3, and V4, BOLD response increases as a function of the number of apertures (a repeated measures ANOVA shows a significant difference between the conditions in all four areas, $F_{2,10} > 22.31, p < 0.01$, with Greenhouse–Geisser correction where appropriate), each with significant differences between all the levels as tested by planned repeated contrasts ($F_{1,5} > 19.87, p < 0.01$ for all contrasts of 10 versus 40 and 40 versus 160 apertures). However, this is not the case in areas MT and MST, which show very similar results. In neither area MT nor MST is there a significant difference between conditions (MT: $F_{2,10} = 3.64, p = 0.065$; MST: $F_{2,10} = 1.15, p = 0.36$). This shows that while we were able to measure an increased response with the number of apertures in areas V1–V4, we could not...
find no evidence of such an increase in areas MT and MST, even though equally robust responses could be found in all areas. We found no significant activation difference in any of the participants between the two 160 aperture conditions (simple mask versus scrambled as tested by paired $t$-tests), apart from in V4 ($t_5 = 3.68$, $p < 0.05$), suggesting that the activation found in MT and MST is not in this case due specifically to the forward motion of the camera. Note also that the BOLD response in MT and MST is of a similar magnitude to Experiment 1. Although the two experiments involve different participants and protocols, it is interesting to note that BOLD response remains the same and is not any greater for the complete movie sequences.

### Experiment 3—Comparing responses to low and high contrasts

The third experiment aims to compare the pattern of results at high and low contrasts as it is known that responses in hMT+ saturate rapidly with contrast (Tootell et al., 1995). The results of Experiment 2 may therefore arise from ceiling effects, although many of the studies discussed above that have previously observed modulation of hMT+ activity were conducted using high-contrast stimuli. Additionally, we added a 2-aperture condition to try to detect the saturation point in hMT+ and showed counter-phase flickering stimuli to test for motion-specific effects.

### Methods

#### MRI acquisition and analysis

This was the same as for Experiment 2, except that due to more conditions, the gaps between the stimuli were reduced; they were chosen randomly and ranged between 2 s and 13 s.

#### Stimulus presentation

The same clips were used as in Experiment 2. We measured response with 2, 10, and 160 apertures (using the same aperture positions as before, the average eccentricity of the 2 apertures over all trials was 19°). A new Sanyo projector had been installed since the first experiment and was linearized. In the high 100% contrast condition, maximum luminance was 162 cd m$^{-2}$, minimum luminance was 0 cd m$^{-2}$, and the movies were shown on a gray 58 cd m$^{-2}$ background. The low-contrast condition was 10% contrast (keeping mean luminance the same). Additionally, we showed the first frame from the high-contrast movies counter-phase flickering at 1 Hz with 10 and 160 apertures. This was chosen as a temporal frequency known to activate MT and MST (Singh, Smith, & Greenlee, 2000) and roughly matching the main temporal frequency found in the movies, estimated by manually tracking contours over the frames.

#### Procedure

The task remained the same as before. Each run lasted approximately 8 min and there were 6 runs as before. Seven participants took part, including two of the authors. After Experiment 4, it became clear that one of the authors was a clear outlier in terms of the percent activity above baseline in Experiments 3 and 4. This could be due to attentional effects as this author wrote the stimulus presentation program. It was deemed safer to not use these data in these experiments, leaving $N = 6$ (4 females).

### Results

The averages over all six participants for the motion conditions are shown in Figure 5. By considering the
pattern over the averages, it seems that apart from a reduction in the response in areas V1–V4 and MT, there is no clear difference in the pattern over apertures between high- and low-contrast responses. For each ROI, we carried out a 2 (levels of contrast) × 3 (number of apertures) repeated measures ANOVA. In all areas V1–V4, there is a significant effect of aperture number (V1: $F_{2,10} = 53.83, p < 0.0005$; V2: $F_{2,10} = 63.09, p < 0.0005$; V3: $F_{2,10} = 60.01, p < 0.0005$; V4: $F_{2,10} = 23.96, p < 0.0005$) and all have a significant main effect of contrast apart from V2 (V1: $F_{1,5} = 7.74, p < 0.05$; V2: $F_{1,5} = 5.51, p = 0.066$; V3: $F_{1,5} = 9.58, p < 0.05$; V4: $F_{1,5} = 10.33, p < 0.05$). In area MT, there is again a significant main effect of contrast apart from V2 (V1: $F_{1,5} = 11.52, p < 0.05$; V2: $F_{1,5} = 5.51, p = 0.066$; V3: $F_{1,5} = 9.58, p < 0.05$; V4: $F_{1,5} = 10.33, p < 0.05$). In area MST, there is no effect of contrast or aperture size ($F_{1,5} = 2.44, p = 0.18$; $F_{1,5} = 0.289, p = 0.76$). Moreover, it is clear that the lack of difference in response between the different numbers of apertures is not caused by a low response in these areas as, if anything, the response in MT and MST is greater to two apertures than in the other areas. In addition, all MT and MST responses are significantly above baseline (1-tailed one-sample test) apart from the 10-aperture low-contrast condition in MT.

Figure 6 illustrates the results for high-contrast motion versus flicker. We can see that there is no difference in motion versus flicker response in any of the ROIs. We confirmed this with 2 (number of apertures) × 2 (motion versus flicker) ANOVAs for all the ROIs and found a significant main effect of the number of apertures in V1–V4 (V1: $F_{1,5} = 83.40, p < 0.0005$; V2: $F_{1,5} = 41.59, p < 0.005$; V3: $F_{1,5} = 37.37, p < 0.005$; V4: $F_{1,5} = 20.62, p < 0.01$), but no other significant effects. This implies that the flattening out of responses across apertures that we observe in MT and MST is not specific to motion but rather to dynamic images.

**Experiment 4—Comparing different types of motion**

In this experiment, we wished to further investigate whether MT/MST response could be modulated by some properties of motion using our manipulated dynamic
stimuli. Previous reports suggest that area MT+ increases in response with motion coherence (Rees et al., 2000), which we did not find in Experiment 2, and Experiment 3 also found no difference in response between flickering and moving stimuli. Experiment 4 involved more comparisons between different types of motion using stimuli more similar to those used by Rees et al. (2000) but still using natural scenes. The stimuli used by Rees et al. were small (2° diameter) random dots, and they compared different proportions of uniform planar translating motion, mixed in with dots randomly redrawn on each frame. Accordingly, we used smaller stimuli and smaller apertures and also used uniform planar translation.

**Methods**

**fMRI acquisition and analysis**

This was the same as for Experiment 3.

**Procedure**

The task remained the same as before. Each run lasted approximately 8 min and there were 6 runs as before. Seven participants took part, including two of the authors. After Experiment 4, it became clear that one of the authors was an extreme outlier in terms of the percent activity above baseline in both Experiments 3 and 4. This could be due to attentional effects as this author wrote the stimulus presentation program. It was deemed safer to not use these data, leaving \( N = 6 \).

**Stimulus presentation**

The same clips were used as in Experiments 2 and 3. Maximum luminance was 162 cd m\(^{-2}\), minimum luminance was 0 cd m\(^{-2}\), and the movies were shown on a gray 58 cd m\(^{-2}\) background.

The eight conditions consisted of:

- **Three motion types**—scrambled (explained below), uniform (explained below), expanding (as before);
- **Two aperture sizes**—large (1.1° diameter) and small (0.25° diameter);
- **Two image sizes**—large (as before, small apertures cover the same area as large, making for 3004 small apertures) and small (not used for expanding motion, and only containing small apertures, explained further below).

The uniform motion is produced by translating the first frame of each image to the left or right (randomly chosen on each trial) at 1 pixel/frame resulting in a speed of 2.5° s\(^{-1}\).

The scrambled motion was created in the same way as in Experiment 2, except that each image sample was randomly rotated to ensure that there would be no consistent side-to-side motion in this condition.

The small images were a (4.1° × 5.7°) rectangular subarea of the movie clip. They contained 160 small apertures, so in effect the small mask is a version of the full large aperture mask scaled down to around a sixth size. They were randomly located in a position that lay within the large image. For the small image sizes, the scrambled motion is rearranged within the subarea visible through the mask (i.e., the random motion has the same features/contrast as the uniform and expanding motion for the small image size).

### Results

For each region of interest, we considered whether the type of motion had an effect, considering first only the full (large) image conditions. We tested this for both of the aperture sizes. See Figure 7 for results.

Figure 7a suggests that in areas V1–V4 the variation in response was caused by the increase in response for the uniform motion condition with small apertures. This was confirmed when we found no effect of motion for large apertures (V1: \( F_{2,5} = 0.022, p = 0.978 \); V2: \( F_{2,5} = 0.27, p = 0.628 \); V3: \( F_{2,5} = 0.30, p = 0.751 \); V4: \( F_{2,5} = 1.325, p = 0.30 \). In area V1, there was also no significant effect of type of motion for small apertures (V1: \( F_{2,5} = 3.86, p = 0.057 \)) either as we might expect as local image features remain very similar across all conditions. For V2, V3, and V4, however, an effect of motion type with small apertures was found (V2: \( F_{2,5} = 4.26 p < 0.05 \); V3: \( F_{2,5} = 5.574 p < 0.05 \); V4: \( F_{2,5} = 5.94 p < 0.05 \)), accompanied by a significant contrast in V2 between uniform and expanding motion (\( f_{1,5} = 25.62, p < 0.005 \)) and in V3 and V4 by significant contrasts between uniform and both other types of motion (V3: \( F_{1,5} = 8.30, p < 0.05 \); \( F_{1,5} = 14.61, p < 0.05 \); V4: \( F_{1,5} = 8.39, p < 0.05 \); \( F_{1,5} = 6.89, p < 0.05 \)) with a one-way repeated measures ANOVA.
Conversely, MT and MST appear to show a somewhat opposite pattern of response from the other areas. They show no difference when viewing through small apertures and seem to prefer uniform motion less than the other two types when viewing through large apertures. Testing the pattern of response over motion conditions reveals no significant effects for either area with small apertures (MT: $F_{2,5} = 0.78, p = 0.48$; MST: $F_{2,5} = 1.13, p = 0.36$). With large apertures, MT displays an effect of motion type ($F_{2,5} = 6.16, p < 0.05$) with a significant contrast between uniform motion and the other two types ($F_{1,5} = 11.22, p < 0.05$; $F_{1,5} = 8.35, p < 0.05$), but MST shows no significant effects ($F_{1,5} = 5.39, p = 0.068$).

In this experiment, we were primarily interested in seeing if our stimulus could be used to demonstrate an effect of different types of motion in MT or MST. We do observe a significant effect in MT, although it is an unexpected one. This shows that although we fail to replicate Rees et al.’s (2000) findings with our stimulus (in fact, we find the opposite effect in the large aperture condition), it is not due to the fact that there is an insufficient response in MT to produce significant effects. The patterns observed across areas are rather complex and more experiments would be needed to untangle them further, which is beyond the scope of this article. What remains clear, however, is our main finding that areas V1–V4 scale with the amount of visual information presented, whereas MT and MST saturate very rapidly. The rest of the conditions illustrated in Figure 8 demonstrate this clearly again, now with a smaller aperture size.

We can see that there is no difference between the response to the small subsection and the whole image in MT and MST, and there is clearly a much bigger response here to the small subimages than in V1–V4, so the lack of difference is not caused by less response in MT and MST. All MT and MST responses are significantly above baseline (1-tailed one-sample test) apart from the small scrambled apertures in the large image condition in MST. Statistical tests confirm a significant effect of image size in areas V1–V4 (V1: $F_{1,5} = 89.77, p < 0.0005$; V2: $F_{1,5} = 53.12, p < 0.005$; V3: $F_{1,5} = 33.46, p < 0.005$; V4: $F_{1,5} = 20.66, p < 0.01$) and no effects in MT and MST (MT: $F_{1,5} = 0.31, p = 0.34$; MST: $F_{1,5} = 4.12, p = 0.55$).

**Discussion**

In the first part of this work, we found that BOLD response in area V1 increases with the amount of local motion detected by the 2DMD model in natural dynamic self-motion scenes, but this was not true in areas MT and MST, and response in these areas was also unaffected by familiarity. The results for V1 are not surprising given that BOLD response in this area is known to be dependent on contrast (Tootell et al., 1995), as is the 2DMD output; however, it seems that despite the known selectivity for motion of areas MT and MST their response did not
increase with the amount of local motion signal. When we introduced apertures to manipulate the amount of visual information available from these scenes, we found that as aperture density increases, the increase in the amount of contrast (the sum of local contrast over space), number of contours, shape information, and other visual attributes causes V1–V4 response to increase. However, even though human areas MT and MST are mostly involved in the processing of motion (Smith et al., 2006; Tootell et al., 1995; Zeki et al., 1991), again, as the amount of visual information including motion signal density increases in the scene, we find that the response in these areas remains approximately constant. The surprising lack of decrease in response in MT and MST even when only two apertures were introduced in Experiment 3 confirms that we are witnessing very rapid saturation with the amount of dynamic information. Importantly, the same pattern occurs at low-contrast levels in MT and MST as at high contrast. This means that the saturation observed at low contrast is not due to the fact that hMT+ has reached maximum possible response but only maximum response to all stimuli at that level of contrast, regardless of how much more dynamic information is added.

Both MT and MST receptive fields are relatively large in non-human primates (around 5° in MT; Felleman & Kaas, 1984) and possibly even larger in humans (Amano et al., 2009). In MST, receptive fields extend into the ipsilateral visual field in monkeys (Duffy & Wurtz, 1991) and humans (Smith et al., 2006). This large size suggests that the flat response may be due to the same group of neurons being recruited across conditions and each neuron exhibiting a similar response magnitude for each condition. This is reinforced by the fact that in Experiments 2 and 3 we see a flatter response in MST, where receptive fields are larger (covering over half the visual field according to the way we have defined this area) and so more likely to be recruited even with only two apertures. However, as MST also does not differentiate between the two different levels of contrast in Experiment 3 it is harder to rule out an effect of response saturation due to contrast here; at 10% contrast, this seems unlikely. It appears that even a small amount of the appropriate type of dynamic pattern within the receptive field of these neurons is enough to almost fully activate them. This would explain why in our first experiment response remained constant on average in MT/MST, as it would seem that these areas reached response saturation to all the stimuli. In a brain area specialized for global motion, one might not expect a change in response in Experiment 1 as this should be similar across conditions; however, it is interesting that the restricted 2- and 10-aperture conditions show no significant difference from 160 apertures in Experiments 2 and 3.

It would appear that although response in these areas has been found to increase the more uniform motion becomes (going from random to coherent; Rees et al., 2000), it does not increase with the amount of coherent motion per se. Certainly, if BOLD response is simply a scaling up of MT or MST single-cell response, one would expect to see an increase of response with coherence (Britten et al., 1993). Interestingly, in contrast to the findings with planar random dot motion, we do not always find a difference here between the incoherent and coherent motion conditions. In Experiment 4, we created similar conditions to those used in Rees et al. (2000) and found the opposite effect, which has also been documented by Kayser et al. (2010). This result could be due to several potential factors resulting from the use of different types of motion and stimuli. For instance, our scrambled motion condition retains more coherent signals than the random motion dot displays, which refresh with random positions on every single frame, providing very different kinds of signal than the coherent motion. Furthermore, Rees et al. manipulated attention, asking participants to detect the direction of motion on a given side. Recent work shows that the amount of attention paid to a motion coherence type stimulus can have large effects on the pattern of activity as a function of coherence (Kayser et al., 2010). We attempted to remove effects of attention as much as possible in order to be able to make parallels with the monkey physiology. Although our test of scrambled stimuli was carried out at high contrast, so were those of Rees et al. The robust activation with only two apertures suggests that our pattern of response is a result of the recruitment of several (in fact, we suggest the majority of) MT receptive fields with only a few patches of visual information. It remains to be seen how motion density and motion coherence interact in terms of BOLD response amplitude when varied together in random dot stimuli of different sizes. A recent computational analysis of the differences between dynamic continuous contours viewed through apertures and scrambled versions of the images reveals differences in direction bandwidth (Kane et al., 2009) that the overall average BOLD response in MT/MST may not be sensitive to but may explain observed single-cell differences (Britten et al., 1993).

Perhaps even more surprising is that Experiment 3 reveals no differences between counter-phase flickering and moving stimuli at low or high contrast in any of the visual areas, even though Experiment 4 did show some modulation of response with motion properties. However, similar results have been found using fMRI, where only modest differences have been found (Heeger, Boynton, Demb, Seidemann, & Newsome, 1999; Singh et al., 2000). Even Tootell et al.’s (1995) fMRI study shows that MT responds almost equally as vigorously to 2–3 Hz flickering stimuli as moving stimuli. Again, this does not appear to mirror what is known from single-cell physiology, where MT and MST cells are found to be highly tuned for certain directions of motion (Albright, 1984; Felleman & Kaas, 1984; Maunsell & Van Essen, 1983). Although as Singh et al. (2000) point out, this is usually found by comparing response between a cell’s preferred direction...
and anti-preferred direction, which does not entirely predict how a population response may look when contrasting natural optic flow versus counter-phase flicker. The fact that we cannot even detect a difference in response to coherent motion from counter-phase flicker seems to indicate that the fMRI subtraction method used here (which is aimed at detecting mean differences at the population level) may not be the most suitable for detecting differences in selectivity in these areas. We merely report here the rapid saturation of these areas with dynamic stimuli and use different examples to examine the specificity of the observed response pattern. The fact that we do not find differences does not imply that single cells do not exhibit this kind of selectivity. From our point of view, the lack of difference found for the scrambled and flickering conditions suggests that the rapid saturation we see with the number of apertures both at low and high contrasts in areas MT and MST is not specific to motion but rather applies to all dynamic stimuli.

Cinematic movie clips were also used by Bartels et al. (2008); however, it must be considered that the free-viewing of these stimuli adds unpredictable motion components to the movement contained in the film sequences, making it difficult to determine the amount of motion in the information picked up by the retina. They found a strong response in area MT+ corresponding to local motion in these clips as analyzed by a local motion pattern matching model. This correlation over time could simply reflect the selectivity for motion of these areas and would yield similar results with rapid saturation or linear scaling. They did not separate out MT and MST but found that the amount of global motion did not correlate with the BOLD response over the combined MT+ area. This corresponds with the lack of difference we found between the coherent global motion and scrambled motion in the two 160 large aperture conditions, suggesting possibly that the role of these areas in optic flow processing may not be as clear as suggested by past studies in which optic flow motion versus random dot motion caused increased activity in areas MT and MST (Morrone et al., 2000; Smith et al., 2006). Although single cells do display motion coherence preference, it is possible that conflicting results concerning difference at a population level suggests that these areas represent a step along the way from local motion to self-motion detection, integrating signals over larger areas and combining motion cues with disparity cues (Smith & Wall, 2008) to which these areas are also responsive and which is crucial for deducing self-motion (Calow & Lappe, 2008; Roy & Wurtz, 1990).

We initially chose natural motion, as it was assumed to be an optimal stimulus for driving activity in motion-sensitive areas, in order to be able to discern any small changes in activity and to provide a more realistic signal-to-noise ratio in the input. This efficiency that causes a maximal response to relatively small amounts of motion may be necessary in light of the sparse motion content found in natural dynamic scenes (Zanker & Zeil, 2005).

Our present study estimates the amount of local motion signals in a range of natural moving scenes and also directly manipulates the amount of information visible in such stimuli and, for both conditions, finds no increase in MT or MST as a function of the amount of dynamic information provided in the stimulus. This suggests that human motion processing areas are very efficient; in other words, the full processing capacity of these areas is dedicated to extracting dynamic information, no matter how small an amount is present.

Acknowledgments

During some of this work, Szonya Durant was supported by Leverhulme Trust Early Career Fellowship ECF/2007/0326.

Many thanks to Velia Cardin, Jac Billington, and Michele Furlan for help with data analysis, Andy Smith for advice on the experimental design, Philip Roberts for help with recording the movies, and Ari Lingeswaran for help with running the experiments.

Commercial relationships: none.
Corresponding author: Szonya Durant.
Email: Szonya.durant@rhul.ac.uk.
Address: Department of Psychology, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK.

References


