Sensitivity of human visual cortical area V6 to stereoscopic depth gradients associated with self-motion

Velia Cardin and Andrew T. Smith
Department of Psychology, Royal Holloway University of London, Egham, United Kingdom

Submitted 27 December 2010; accepted in final form 3 June 2011

Cardin V, Smith AT. Sensitivity of human visual cortical area V6 to stereoscopic depth gradients associated with self-motion. J Neurophysiol 106: 1240–1249, 2011. First published June 8, 2011; doi:10.1152/jn.01120.2010.—The principal visual cue to self-motion (egomotion) is optic flow, which is specified in terms of local 2D velocities in the retinal image without reference to depth cues. However, in general, points near the center of expansion of natural flow fields are distant, whereas those in the periphery are closer, creating gradients of horizontal binocular disparity. To assess whether the brain combines disparity gradients with optic flow when encoding egomotion, stereoscopic gradients were applied to expanding dot patterns presented to observers during functional MRI scanning. The gradients were radially symmetrical, disparity changing as a function of eccentricity. The depth cues were either consistent with egomotion (peripheral dots perceived as near and central dots perceived as far) or inconsistent (the reverse gradient, central dots near, peripheral dots far). The BOLD activity generated by these stimuli was compared in a range of predefined visual regions in 13 participants with good stereovision. Visual area V6, in the parieto-occipital sulcus, showed a unique pattern of results, responding well to all optic flow patterns but much more strongly when they were paired with consistent rather than inconsistent or zero-disparity gradients. Of the other areas examined, a region of the precuneus and parietoinsular vestibular cortex also differentiate between consistent and inconsistent gradients, but with weak or suppressive responses. V3A, V7, MT, and ventral intraparietal area responded more strongly in the presence of a depth gradient but were indifferent to its depth-flow congruence. The results suggest that depth and flow cues are integrated in V6 to improve estimation of egomotion.

The neural correlates of stereopsis in humans have been assessed in a number of contexts such as the perception of 3D shape (e.g., Gilaie-Dotan et al. 2002; Tsao et al. 2003; Welchman et al. 2005). However, very little is known about the brain mechanisms responsible for encoding disparity gradients and integrating them with optic flow during the computation of self-motion. Rutschmann et al. (2000) used functional MRI (fMRI) to show that a region in BA19, which could correspond to V3B although identification of specific visual areas was not performed, revealed a preference for optic flow components over random noise, and this response was enhanced with dichoptic presentation of the stimuli and a disparity gradient. Recent studies have shown the involvement of several visual and multisensory areas in the representation of optic flow and self-motion (Peuskens et al. 2001; Cardin and Smith 2010; Wall and Smith 2008). In particular, the comparison of optic flow that is either compatible or incompatible with self-motion yields differential activation in area MST, V6, the cingulate sulcus visual area (CSv), the ventral intraparietal area (VIP), and multisensory/vestibular regions parietoinsular vestibular cortex (PIVC) and putative area 2v (p2v) (Cardin and Smith 2010; Wall and Smith 2008). There is also differential activation in a small part of the precuneus (PC; Cardin and Smith 2010) in the dorsal part of the ascending arm of the cingulate sulcus; here we refer to this as the precuneus motion area (PCm), to distinguish it from other parts of the precuneus. MST and VIP have been extensively characterized in macaque, and both contain many neurons that show preferences for optic flow components (Duffy and Wurtz 1991; Schaafmsa and Duysens 1996) and sensitivity to the direction of heading in the visual, somatosensory, and vestibular domains (Bremmer et al. 2002; Britten and van Wezel 1998; Duffy and Wurtz 1995; Gu et al. 2008; 2007; 2006; Page and Duffy 2003). In the human visual cortex, V6 responds more strongly to optic flow than incoherent motion (Pitzalis et al. 2010) and is the area that gives the strongest differential response to egomotion-compatible optic flow compared with egomotion-incompatible flow (Cardin and Smith 2010). Taking into account that V6 contains a representation of the contralateral visual field that extends up to 80 degrees, in which the center is not magnified relative to the periphery (Galletti et al. 1999; Pitzalis et al. 2006; Stenbacka and Vanni 2007), and that it is reciprocally connected to visual areas thought to be involved in egomotion perception, such as MST and VIP (Galletti et al. 2001; Shipp et al. 1998), V6 may therefore play a key role in the extraction of visual cues for self-motion.

Some of the cortical areas described above have been shown to be sensitive to stereoscopic cues or to the distance between object and observer (Roy and Wurtz 1990; Roy et al. 1992;
Quinlan and Culham 2007; Smith and Wall 2008), which suggests that they could be integrating motion and stereoscopic cues for self-motion. In this study we aimed to determine whether areas involved in the extraction of optic flow as a visual cue for self-motion also utilize disparity cues. For this purpose, we assessed the sensitivity of these areas to stereoscopic cues that are consistent with self-motion. By using the same field of expanding dots combined with either egomotion-consistent or egomotion-inconsistent disparity fields, we evaluated, not only the sensitivity of these areas to stereoscopic depth, but also the integration of disparity cues that are specifically relevant for the perception of self-motion.

MATERIALS AND METHODS

Participants. Fifteen healthy human volunteers took part in the study. They were screened for normal stereoeacuity before being invited to participate. Two were later excluded because of poor binocular fusion during scanning. Therefore, data from 13 participants (10 female; mean age 27 yr) were included in the analysis. All participants were scanned for the main experiment, 12/13 for retinotopic mapping and egomotion-compatible optic flow areas localizers, and 10/13 for MT/MST localizers (the missing localizer scans were due to unavailability of participants for rescanning). All participants were screened according to standard procedures, and written informed consent was obtained. The project was approved by the Royal Holloway ethics committee.

Data acquisition. MRI images were obtained with a 3-Tesla Siemens Magnetom Trio scanner (Siemens, Erlangen, Germany) and a custom-designed eight-channel posterior array (PA) head coil designed for imaging posterior regions of the brain (Stark Contrast, Erlangen, Germany). Scanning was conducted over at least two sessions on different days. In one session, the data for the main experiment were acquired; in the others, functional data for the localizers and a 3D, T1-weighted anatomical scan (modified driven equilibrium Fourier transform, MDEFT; 160 × 256 × 256 resolution; 1-mm isotropic voxels; Deichmann et al. 2004) were acquired. This anatomical image, which alone was acquired with a standard Siemens head array coil, was used as a reference to which all the functional images obtained with the PA coil were coregistered.

Functional data acquisition employed a standard gradient echoplanar sequence (main experiment and egomotion-sensitive area localizer: repetition time $T_R = 2.5$ s, 35 contiguous axial slices, interleaved acquisition order, 3-mm isotropic voxels, $T_E = 31$ ms; retinotopic mapping and MT/MST localizer: same parameters, except $T_E = 2$ s, 28 slices). Each scan run lasted 6 min 40 s for the main experiment, 8 min 56 s for retinotopic mapping, 4 min 47 s for egomotion-compatible optic flow areas, and 5 min 5 s for MT/MST localizers. For coregistration purposes (see Data analysis), at the beginning of each scanning session, we acquired two single-volume echo-planar imaging (EPI) sequences that had the same position parameters as the experimental runs: one (BC-EPI) using the scanner’s integral whole body coil (BC) to give uniform contrast, and another (PA-EPI), immediately after, acquired with the PA coil. Within each scanning session, a T1-weighted anatomical scan (MP-RAGE, Siemens; $1 \times 1 \times 1$ mm; field of view: $256 \times 256$) was also acquired to assist with coregistration of the functional data to the MDEFT scan.

Stimuli and procedure. The stimuli for the main experiment were viewed via binocular LCD goggles (NordicNeuroLab, Bergen, Norway) that were adjusted to match each participant’s interocular distance. The images subtended $30 \times 23$ degree visual angle and had a resolution of $800 \times 600$ pixels. The stimuli consisted of 600 moving white dots on a dark background. Each dot subtended 16-min arc diameter, and each eye’s display was updated at 30 Hz. The dots, which were initially randomly located, formed a circular patch of 23-degree diameter. Dot directions were arranged to produce radial motion away from the center of the patch (expansion, simulating forward motion of the observer), with dot speed increasing as a linear function of eccentricity ($\text{min} = 1.15$ degree/s, $\text{max} = 9.23$ degree/s). A central fixation cross was provided, and participants were instructed to maintain fixation at all times. The dots had a radial gradient of horizontal disparity that was centered at the fixation plane and spanned $\pm 52$-min arc disparity. There were three disparity-gradient conditions: 1) consistent disparity, 2) inconsistent disparity, and 3) zero disparity. In the consistent condition, horizontal disparity changed linearly with eccentricity in all directions from the center. Dots near the center of the display had the largest uncrossed disparity and were perceived as distant, whereas dots at the periphery had the largest crossed disparity and were perceived as closer to the observer, consistent with the forward self-motion indicated by the optic flow. The fixation cross appeared at a depth midway between the nearest and furthest dots. In the inconsistent condition, the disparity gradient was reversed, the perceptual result being that dots increased in distance from the observer as they moved away from the center. This is not consistent with natural self-motion. The zero disparity condition had no disparity gradient; all the dots were presented at the fixation plane. Thus the middle of the disparity range was zero for all three stimuli. Figure 1 shows sample stereopairs illustrating the consistent and inconsistent disparity gradients. Movies showing the complete, animated stimuli are provided as online supplemental material, available at the J Neurophysiology website.

An event-related design was used for the main experiment. Participants took part in six to nine experimental runs, each lasting 6 min 40 s. Within each run, there were 48 trials (16 per condition). During each trial, a stimulus was presented for 3 s, followed by an interstimulus interval (ISI) that varied between 2 s and 8 s in steps of 1 s, approximately following a Poisson probability distribution (median 4 s). During the ISI, the screen was a uniform gray apart from the central fixation cross, which was present continuously. A central fixation task was used to ensure attention to the depth structure of the motion stimuli and to monitor successful binocular fusion and stereopsis throughout the experiment. In this task, the color of the fixation cross was randomly reselected at a rate of 2 Hz from six easily distinguished hues, and participants were instructed to maintain fixation at all times and to monitor the color of the cross. Whenever a specific color (dark blue) appeared, the task was to indicate, by pressing one of four buttons, the nature of the displayed stimulus: “hollow” (consistent), “pointy” (inconsistent), “flat” (zero disparity), or “no stimulus” (ISI). A blue cross could occur at any time except 500 ms after the beginning and before the end of a stimulus or ISI. Only runs in which participants performed the task with at least 70% accuracy for each condition were included in the analysis. The task was possible only when fixation was close to the fixation cross. Occasional saccades resulting in small fixation errors may have occurred but are unlikely to affect the results because they do not alter the stereoscopic gradients.

On completion of the main scan, participants were asked whether they experienced vection (the illusory sensation that they were actually moving). All reported that they did not, at any time. This is expected when using brief presentations becausevection takes time to develop.

The strategy adopted was to quantify activity occurring in the main experiment within various regions of interest, some of which have previously been associated with optic flow or egomotion. These regions were identified with separate localizer scans that enabled us to demarcate their boundaries and also to avoid “double-dipping” bias when quantifying activity within them. Stimuli for the localizer scans were back projected onto a screen mounted in the rear of the scanner bore by a computer-controlled LCD projector. During the retinotopic mapping and MT/MST localizer scans, all stimuli were seen binocularly in free view via a mirror above the head coil. For retinotopic mapping, standard procedures were used (Sereno et al. 1995; rotating...
Fig. 1. A: diagram showing the depth structure of the 3 stimuli used in the main experiment. The arrows indicate local dot motion direction, arrow length indicates the presence of a dot speed gradient, and arrow color indicates the magnitude of horizontal binocular disparity (see key; units are min of arc). B–D: stereopairs illustrating the stereoscopic gradients that were applied to the stimuli. Each stereopair represents 1 frame of an animation in which dots moved outward from the center, simulating forward motion. When the right image is viewed by the right eye and the left by the left, the stereopair in B creates a zero gradient. That in C creates a consistent radial gradient, and D creates an inconsistent gradient. If free fused by converging the eyes, C and D will reverse.
24-degree flickering red and green “wedge”; 8 cycles; 64 s/cycle). MT and MST were mapped using a standard procedure (Huk et al. 2002) in which moving and static dots are contrasted in each hemifield. Areas V6, VIP, p2v, PIVC, and PcM were localized using egomotion-compatible optic flow as described in Cardin and Smith (2010). Stimuli for this localizer scan were viewed via a custom monocular compatible optic flow as described in Cardin and Smith (2010). Areas V6, p2v, and PIVC were localized using egomotion-in which moving and static dots are contrasted in each hemifield. The first three volumes of each run were discarded to allow for T1 gressor at every voxel. Six movement parameters were derived from a multiple-regression analysis to generate parameter estimates for each repetition of the canonical hemodynamic response function, and entered into a multiple-regressors representing the three stimulus categories and four button checked visually.

Reference MDEFT image, using parameters obtained from the coregistration of these images to the space of the intrasession inhomogeneity-corrected MP-RAGE and assumed no EPIs acquired with the PA head coil, coregistration of these images to the anatomy is poor. Therefore, we coregistered the BC-EPI to the intrasession homogeneity-corrected MP-RAGE and assumed no head movements between the acquisition of this BC-EPI and the PA-EPI. All images where then transformed to the space of the reference MDEFT image, using parameters obtained from the coregistration of the two anatomical scans. Coregistration accuracy was checked visually. Analysis was conducted by fitting a general linear model with regressors representing the three stimulus categories and four button presses. For every condition, each event was modeled as a boxcar with 3 s (stimuli) and 0.5 s (button presses) duration, convolved with a canonical hemodynamic response function, and entered into a multiple-regression analysis to generate parameter estimates for each repressor at every voxel. Six movement parameters were derived from the realignment of the images and were also included in the model. The first three volumes of each run were discarded to allow for T1 equilibration effects.

Effect sizes (β values) for the three experimental conditions were extracted for each independently defined ROI by averaging across all voxels in the ROI. The results were normalized, to account for overall amplitude differences across participants, by expressing each activation in a given ROI as a proportion of the largest activation. This was done independently for each ROI. Finally, the results were averaged across participants.

Data analysis: ROI definitions. Areas V6, VIP, p2v, PIVC, and PcM were defined as all contiguous voxels that were more active with a single pattern of expansion-contraction and rotation (egomotion compatible), than with nine simultaneous smaller patterns (egomotion incompatible) in the parieto-occipital sulcus (V6), the intraparietal sulcus (VIP), the junction of the intraparietal sulcus and the postcentral sulcus (p2v), the region of the precuneus dorsal to the ascending arm of the cingulated sulcus (PcM), and the posterior region of the insula, putatively the PIVC. Area V6 was successfully defined in 22/24 hemispheres, VIP in 14/24, p2v in 16/24, CSv in 20/24, PcM in 17/24, and PIVC in 19/24. The method is described in detail in our earlier work (Cardin and Smith 2010).

MST was defined as all contiguous voxels within MT+ that were significantly active during ipsilateral motion stimulation. MT was defined as all contiguous active voxels that were active during contralateral but not ipsilateral stimulation, with the proviso that 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. MT and MST were successfully defined in all the scanned hemispheres (n = 20).

Retinotopic data were analyzed conventionally. ROIs (visual areas V1-V7) were drawn by eye on the basis of boundaries formed by phase reversals, viewed on a flattened version of each participant’s reference anatomy. V3B was defined according to the original criteria of Smith et al. (1998) and corresponds to LO1 of Larsson and Heeger (2006). Retinotopic areas were successfully defined in all the scanned hemispheres (n = 24).

RESULTS

The aim of this study was to determine sensitivity to stereoscopic depth gradients in areas of the brain putatively involved in the processing of egomotion. We localized egomotion-sensitive cortical areas V6, PIVC, VIP, p2V, PcM, and CSv using data from independent scans in which activations elicited by egomotion-consistent and egomotion-inconsistent optic flow were contrasted (see MATERIALS AND METHODS). Figure 2 shows the locations of these regions in an inflated representation of one hemisphere, and their locations are described in detail in our earlier work (Cardin and Smith 2010). There is some uncertainty concerning the identity and labeling of VIP and p2V. The average coordinates of our VIP are rather more posterior than those reported by Bremmer et al. (2001), Sereno and Huang (2006), and Huang and Sereno (2007), which are closer to our p2V. The coordinates of our VIP are somewhere in between topographic maps IPS3 and IPS4 indented by Swisher et al. (2007) and Hagler et al. (2007); IPS3 seems also to correspond to the area described as the homologue of macaque’s lateral intraparietal area (Sereno et al. 2001; Hagler et al. 2007). For clarity and for consistency with our previous work, we have simply adopted here the same conventions and labels that we have used previously.

We also localized MT and MST and retinotopic areas V1-V7 with standard procedures. We then determined the sensitivity to stereoscopic egomotion cues in each of these areas by measuring the response to a constant optic flow stimulus with zero disparity and comparing it to the responses with added disparity gradients that were either consistent or inconsistent with egomotion.

Table 1. Mean Talairach coordinates of ROIs identified with egomotion-compatible optic flow

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Right</th>
<th>Number of voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
</tr>
<tr>
<td>V6</td>
<td>−17</td>
<td>−78</td>
<td>26</td>
</tr>
<tr>
<td>PcM</td>
<td>−12</td>
<td>−49</td>
<td>46</td>
</tr>
<tr>
<td>VIP</td>
<td>−27</td>
<td>−53</td>
<td>51</td>
</tr>
<tr>
<td>2v</td>
<td>−30</td>
<td>−46</td>
<td>47</td>
</tr>
<tr>
<td>CSv</td>
<td>−11</td>
<td>−19</td>
<td>40</td>
</tr>
<tr>
<td>PIVC</td>
<td>−40</td>
<td>−33</td>
<td>19</td>
</tr>
</tbody>
</table>

The numbers of voxels are mean numbers (± SE) of functional voxels included in each region of interest (ROI). PcM, precuneus motion area; VIP, ventral intraparietal area; CSv, cingulate sulcus visual area; PIVC, parietoinsular vestibular cortex.
between two conditions (1 patch and 9 patches, see MATERIALS). Should also be remembered that PIVC is defined by a contrast response apparently drops to near zero (Fig. 3). Second, it Smith 2008 for VIP and CSv), and in the case of PIVC the main experiment, gives weaker responses in many areas (see Cardin and Smith 2010). Pure expansion, as used in the case, involves time-varying flow that cycles through spiral space, but two points should be born in mind. First, the localizer zero-disparity stimulus is similar to that used for the localizer, zero (Fig. 3) - entropy gradients than enhancement by consistent gradi-ents. With no disparity gradient, the average response is near consistent but not inconsistent disparity gradients. Figure 3A shows that V6 activity is similar for the zero-disparity and inconsistent conditions, but in the consistent condition it is much higher. This difference was significant both in relation to the zero-disparity condition $[t(21) = 4.85, P < 0.0001]$, and in relation to the inconsistent condition $[t(21) = 3.78, P < 0.002]$. When activity in the zero-disparity condition is subtracted off (Fig. 3B), a large residual activation remains, but only for consistent depth. One interpretation is that V6 encodes optic flow for the purpose of determining egomotion trajectory (Cardin and Smith 2010) and that it uses stereoscopic depth gradients, where available, to assist.

Two other areas also show sensitivity to the polarity of the disparity gradient (Fig. 3). The first is area PIVC. In this case, however, the results are more suggestive of suppression by inconsistent gradients than enhancement by consistent gradients. With no disparity gradient, the average response is near zero (Fig. 3A). This may seem surprising, given that the zero-disparity stimulus is similar to that used for the localizer, but two points should be born in mind. First, the localizer involves time-varying flow that cycles through spiral space, and this reliably gives positive, if weak, activation in PIVC (see Cardin and Smith 2010). Pure expansion, as used in the main experiment, gives weaker responses in many areas (see Fig. 4 of Smith et al. 2006 for MT and MST, Fig. 2 of Wall and Smith 2008 for VIP and CSv), and in the case of PIVC the response apparently drops to near zero (Fig. 3). Second, it should also be remembered that PIVC is defined by a contrast between two conditions (1 patch and 9 patches, see MATERIALS AND METHODS), and a significant contrast does not necessarily imply positive responses; it could be that one is positive and one negative, or even that both are negative but one more so than the other. In the consistent condition, the response in PIVC is also near zero, but, in the inconsistent condition, it is negative, suggesting suppression. The difference between consist-ent and inconsistent conditions is significant $[t(18) = 2.80, P < 0.02]$. The difference is also clear when the results are considered in the format of Fig. 3B. This result is consistent with the idea that PIVC, which receives excitatory vestibular input, is suppressed by visual motion information that suggests a motion source other than egomotion and uses stereoscopic depth gradients in the computation.

Another variation on sensitivity to the gradient polarity is apparent in area PcM. Here the activity with zero disparity is again near zero, as for PIVC (but not V6). However, in this case, activity seems to increase to positive levels with a consistent gradient and can be suppressed below baseline by an inconsistent gradient. Caution is required because the effects are small, and there is always uncertainty about the location of the baseline in fMRI studies. Moreover, neither the consistent nor the inconsistent condition is significantly different from zero disparity. However, inconsistent and consistent conditions are significantly different from each other $[t(16) = 3.11, P < 0.01]$, suggesting that there is sensitivity to the polarity of the disparity gradient. Thus three areas (V6, PIVC, PcM) show this characteristic, but it is manifest in different ways. Any or all of these three areas may be involved in integrating optic flow and stereoscopic signals to estimate egomotion.

Visual areas V3A, V7, MT, and VIP responded well to optic flow in all three conditions, responded significantly more strongly when a stereoscopic gradient was added than when flow was presented in the fixation plane, but were indifferent to the polarity of that gradient. Figure 4A shows the results from these areas in the same format as Fig. 3A. In each case, the response is reduced in the zero-disparity condition compared with the consistent and inconsistent conditions, which were similar to each other. This pattern of results indicates the presence of sensitivity to stereoscopic depth but not necessarily sensitivity to global depth structure; it is possible that the same result would be obtained if the depth gradient were spatially scrambled to create a depth “cloud”. Figure 4B shows, for each area, the mean additional activity obtained in the consistent and inconsistent conditions relative to the zero-disparity condition,

---

**Fig. 2.** An inflated representation of the right hemisphere of a representative participant (medial and lateral views) marked with the locations of the regions of interest (ROIs) that were studied. CSv, cingulate sulcus visual area; PcM, precuneus; VIP, ventral intraparietal area; p2v, putative area 2v; PIVC, parietoinsular vestibular cortex.

Uniquely among all the regions examined, area V6 showed a clear response to optic flow with zero disparity that was substantially enhanced (more than doubled) by the addition of consistent but not inconsistent disparity gradients. Figure 3A shows that V6 activity is similar for the zero-disparity and inconsistent conditions, but in the consistent condition it is much higher. This difference was significant both in relation to the zero-disparity condition $[t(21) = 4.85, P < 0.0001]$, and in relation to the inconsistent condition $[t(21) = 3.78, P < 0.002]$. When activity in the zero-disparity condition is subtracted off (Fig. 3B), a large residual activation remains, but only for consistent depth. One interpretation is that V6 encodes optic flow for the purpose of determining egomotion trajectory (Cardin and Smith 2010) and that it uses stereoscopic depth gradients, where available, to assist.

Two other areas also show sensitivity to the polarity of the disparity gradient (Fig. 3). The first is area PIVC. In this case, however, the results are more suggestive of suppression by inconsistent gradients than enhancement by consistent gradients. With no disparity gradient, the average response is near zero (Fig. 3A). This may seem surprising, given that the zero-disparity stimulus is similar to that used for the localizer, but two points should be born in mind. First, the localizer involves time-varying flow that cycles through spiral space, and this reliably gives positive, if weak, activation in PIVC (see Cardin and Smith 2010). Pure expansion, as used in the main experiment, gives weaker responses in many areas (see Fig. 4 of Smith et al. 2006 for MT and MST, Fig. 2 of Wall and Smith 2008 for VIP and CSv), and in the case of PIVC the response apparently drops to near zero (Fig. 3). Second, it should also be remembered that PIVC is defined by a contrast between two conditions (1 patch and 9 patches, see MATERIALS AND METHODS), and a significant contrast does not necessarily imply positive responses; it could be that one is positive and one negative, or even that both are negative but one more so than the other. In the consistent condition, the response in PIVC is also near zero, but, in the inconsistent condition, it is negative, suggesting suppression. The difference between consist-ent and inconsistent conditions is significant $[t(18) = 2.80, P < 0.02]$. The difference is also clear when the results are considered in the format of Fig. 3B. This result is consistent with the idea that PIVC, which receives excitatory vestibular input, is suppressed by visual motion information that suggests a motion source other than egomotion and uses stereoscopic depth gradients in the computation.

Another variation on sensitivity to the gradient polarity is apparent in area PcM. Here the activity with zero disparity is again near zero, as for PIVC (but not V6). However, in this case, activity seems to increase to positive levels with a consistent gradient and can be suppressed below baseline by an inconsistent gradient. Caution is required because the effects are small, and there is always uncertainty about the location of the baseline in fMRI studies. Moreover, neither the consistent nor the inconsistent condition is significantly different from zero disparity. However, inconsistent and consistent conditions are significantly different from each other $[t(16) = 3.11, P < 0.01]$, suggesting that there is sensitivity to the polarity of the disparity gradient. Thus three areas (V6, PIVC, PcM) show this characteristic, but it is manifest in different ways. Any or all of these three areas may be involved in integrating optic flow and stereoscopic signals to estimate egomotion.

Visual areas V3A, V7, MT, and VIP responded well to optic flow in all three conditions, responded significantly more strongly when a stereoscopic gradient was added than when flow was presented in the fixation plane, but were indifferent to the polarity of that gradient. Figure 4A shows the results from these areas in the same format as Fig. 3A. In each case, the response is reduced in the zero-disparity condition compared with the consistent and inconsistent conditions, which were similar to each other. This pattern of results indicates the presence of sensitivity to stereoscopic depth but not necessarily sensitivity to global depth structure; it is possible that the same result would be obtained if the depth gradient were spatially scrambled to create a depth “cloud”. Figure 4B shows, for each area, the mean additional activity obtained in the consistent and inconsistent conditions relative to the zero-disparity condition,
obtained by subtracting off the latter to isolate activity attributable to the presence of the stereoscopic depth. The plot suggests that, of the four areas, disparity sensitivity is greatest in VIP and least in MT. Pairwise $t$-tests showed that, in each of these four areas, there was a significant difference between the zero-disparity condition and the consistent condition [VIP: $t(13) = 5.88, P < 0.0001$; V7: $t(23) = 5.68, P < 0.0001$; V3A: $t(23) = 3.04, P < 0.01$; MT: $t(19) = 2.07, P < 0.05$] and also between the zero-disparity condition and the inconsistent condition [VIP: $t(13) = 4.85, P < 0.0001$; V7: $t(23) = 4.38, P < 0.001$; V3A: $t(23) = 3.50, P < 0.002$; MT: $t(19) = 2.55, P < 0.05$]. However, in no case was there a significant difference between consistent and inconsistent conditions.

Also shown in Fig. 4 are results for vestibular area 2v. This resembles the other areas in the figure when viewed in terms of activity relative to zero disparity (Fig. 4B), suggesting that, like those areas, it has a nonspecific preference for disparity over no disparity. However, when the three conditions are compared directly (Fig. 4A), it can be seen that the preference for the presence of disparity arises from a negative (suppressive) response in the absence of a disparity gradient. Consistent [$t(15) = 2.70, P < 0.02$] and inconsistent [$t(15) = 2.75, P < 0.02$] gradients are both significantly more positive than zero disparity, but, as with PIVC, the results may reflect suppressive as well as facilitatory influences with reliable interpretation difficult.

Many of the regions studied were indifferent to the presence of stereoscopic depth gradients. Figure 5 summarizes the results for these areas. They include all the retinotopically defined regions except V3A and V7, together with CSv and, perhaps surprisingly, MST. For each area, the figure shows the
normalized response to each of the three conditions, averaged across participants. There are no significant differences between consistent, inconsistent, and zero-disparity conditions in any of these brain regions. The optic flow generates large activations in most cases, but these are not influenced by the stereoscopic gradient. This does not mean that there is no sensitivity to local stereoscopic depth in neurons in these areas; indeed there is good reason to believe that there is (see DISCUSSION). It means only that there is no sensitivity to global disparity structure, at least when considering the two structures employed here. It can be seen that the response in CSv is relatively small. This is not inconsistent with our earlier work (Wall and Smith 2008), in which we claim not that CSv responds more strongly than MST but instead that the observed response is more selective for egomotion compatibility than in MST.

DISCUSSION

Use of stereoscopic cues for egomotion. Our key finding concerns human visual area V6, or hV6, which was identified only quite recently (Pitzalis et al. 2006) and has been associated with encoding coherent global motion (Cardin and Smith 2010; Pitzalis et al. 2010; von Prössl et al. 2009). We have hypothesized (Cardin and Smith 2010) that it is one of several areas that encode optic flow in connection with monitoring and controlling egomotion. We show here (Fig. 3) that, not only is V6 responsive to conventional optic flow patterns presented in 2D, but also that, when optic flow is presented as a 3D array of moving dots, it is sensitive to the depth structure of the array. When a radial stereoscopic depth gradient (Fig. 1) is added to the optic flow, responses in V6 are enhanced if the structure is consistent with the interpretation that the moving dots reflect egomotion relative to a static environment (consistent condition) and are not enhanced if the dots are more likely to reflect moving objects (inconsistent condition).

Previous studies have shown that stereoscopic cues contribute to the judgment of heading direction (Macuga et al. 2006; van den Berg and Brenner 1994a; 1994b) and enhance the sensation of vection (Palmisano 2002; 1996). It should be noted that a consistent depth gradient by itself does not indicate egomotion. It simply suggests that the observer is, or could be, in a typical environment with near and far objects. The gradient does not change as one moves through an environment with regular features, so its presence is equally consistent with forward egomotion, backward egomotion, and no egomotion. Only when combined with optic flow is it informative about egomotion. The overall depth structure is informative in determining whether the objects (in this case dots) forming the flow field are at depth positions consistent with the interpretation that the objects are static and the observer is moving, or whether a more probable interpretation is that dot motion arises from the motion of objects relative to a static observer. Consistent with this idea, van den Berg and Brenner (1994a) concluded that binocular disparities improved judgments of heading by imposing a depth order on the elements of the scene and not by providing information on the objects’ motion in depth.

If a given brain region responds well to optic flow and the response is systematically enhanced when depth cues are consistent with observer motion, it is likely that the region is concerned in some way with encoding egomotion. We found this behavior only in V6. This, of course, does not preclude the processing of optic flow in other areas (e.g., MST), but it suggests that these areas do not make great use of stereoscopic depth cues in disambiguating the two possible causes of retinal motion.

Areas PIVC and PcM also differentiate between consistent and inconsistent gradients but on the basis of weak or suppressive responses that are harder to interpret. It seems that PIVC, which is associated primarily with vestibular activity, is actively suppressed in the presence of incompatible visual cues but unresponsive in the presence of compatible stereoscopic cues or in the absence of stereoscopic cues. Macaque PIVC does not respond to optic flow (Chen et al. 2010), and several previous studies have reported suppression of PIVC by visual motion (Brandt et al. 1998; Dieterich and Brandt 2001; Kikuchi et al. 2009). Area PcM does appear to respond positively, though weakly, to optic flow, at least when stereoscopic cues are consistent with egomotion. Previous studies have shown that a region in the precuneus, the anatomical location of which is similar to our PcM, is involved in spatial representation of the world, showing differential activations in the retrieval of an event or person in a spatial context (Burgess et al. 2001), when subjects have to update the spatial location of objects, taking into account a self-displacement (Wolbers et al. 2008), or when future path information allows for greater anticipation of a heading response (Billington et al. 2011). Furthermore, preliminary data from our laboratory show that this area is also activated by galvanic vestibular stimulation, which results in the perceptual experience of self-motion (roll or yaw).

Stereoscopic gradients are useful mainly for objects or textures that are located within about 10 m of the observer. This is because, for distant objects, the magnitude of the change in disparity with distance becomes vanishingly small. Because many objects are distant during locomotion, stereoscopic cues may not always add much information about egomotion. They will be useful only to the extent that the scene contains near objects. When walking across a flat surface with no obstacles, the peripheral lower visual field contains useful stereoscopic cues, but other parts of the image do not. As the environment becomes more cluttered with objects, such as trees or foliage, they become much more useful. It is unclear why V6 shows strong sensitivity to depth gradients while other areas associated with optic flow processing do not, but one possibility could be that V6 is specialized for navigating in dense or cluttered environments. Given its proximity to parietal areas involved in motor planning and control, it is possible that V6 is not involved in the perception of egomotion in general but instead is concerned specifically with egomotion relative to objects and obstacles that are amenable to motor interventions, such as pushing aside branches, during locomotion. In support of this, Quinlan and Culham (2007) showed that a region in the human parieto-occipital sulcus (POS), the location of which matches that of V6, shows stronger activation during vergence in near than far space. VIP has also been attributed with a specialization for encoding near space (Colby et al. 1993).

In macaques, area V6 is adjacent to area V6A, which is located more dorsally but still within the POS. Area V6A plays a key role in reaching and grasping, has cells that have larger receptive fields and are less visually responsive than cells in V6, and does not have a clear retinotopic map (Galletti et al. 2010).
In a previous study in which we show that a differential activation for egomotion-compatible optic flow overlaps with retinotopically mapped area V6 in the human brain (Cardin and Smith 2010), we mention that it is difficult to rule out the possibility that the region we define as V6 includes at least a portion of human V6A, given the variability and sometimes ill-defined nature of retinotopic maps. This is also the case in this study, in which we use a functional localizer scan to identify V6, and, even though we restricted the inclusion criteria to highly significant voxels for our contrast, it is possible that activations spread to adjacent areas including V6A. This is a particularly important possibility to consider in the light of recent results showing that V6A cells have a preference for targets fixated in reachable space, especially at the beginning of a fixation (Hadjidimitrakis et al. 2010). However, it is difficult to be definitive about the role of this area until a reliable way of defining V6A is found for the human brain.

Stereopsis in other areas. Other areas traditionally (MST, VIP) and more recently (CSv) associated with visual cues to egomotion do not show the same characteristics as V6. MST appears to be invariant, in terms of total population response, to our stereoscopic manipulations (Fig. 5). It is known that most neurons are sensitive to disparity in macaque MSTd (Roy et al. 1992). There is even some evidence for sensitivity to specific 3D depth planes (Sugihara et al. 2002) although these were not defined stereoscopically. A recent fMRI adaptation study shows that human MST is sensitive to the stereoscopic plane in which rotational flow is presented (Smith and Wall 2008). Thus it is highly probable that our depth manipulations do alter the pattern of activity within MST. The fact that they do not alter the total population response indicates that all disparities are equally represented and that there is probably no specialization for egomotion-consistent global depth configurations. There is no evidence for sensitivity to radial (as opposed to planar) depth gradients in macaque MSTd and no reason to suppose that stereoscopic depth is used in calculating heading in MSTd.

Similar arguments apply to area CSv. Whether this region has any sensitivity to stereoscopic depth is unknown, but certainly it appears that there is no sensitivity to the polarity of a radial stereoscopic gradient, or at least no preference for egomotion-compatible gradients. This may seem surprising in light of our previous claims that CSv may be specialized for encoding self-motion, but the result is clear, allowing us to conclude that CSv is concerned with self-motion cues derived from 2D optic flow but not with those derived from stereoscopic depth. This is supported by work from Fischer et al. (2010), showing that the response profile of CSv is consistent with a role in integration of visual cues for self-motion, because of its enhanced response to coherent, full-field and real (as opposed to retinal) motion signals.

Area VIP (ventral intraparietal) is one of several areas that respond more strongly to flow in the presence of depth gradients than in their absence but with no sensitivity to gradient polarity (Fig. 4). Indeed, the enhancement is particularly strong in VIP. Other areas in this category are V7, V3A, 2v, and (weakly) MT. The difference between MT and MST is subtle; both show the same trends, but only MT reaches statistical significance. The reality could be that they show very similar behavior, but they are assigned here to different groups on the basis of statistics. Conversely, it is also possible that they are genuinely different but that the difference is eroded by imperfect separation of the two regions during localization. Only those voxels responding to ipsilateral stimulation are included in the MST ROI. Therefore, if the receptive fields of some MST neurons include mainly contralateral regions, the voxels containing them may be misclassified as MT. There is also a possibility of MT receptive fields extending over the midline. However, this is less likely to be a confound because we did not stimulate the foveal region in our localizer. In VIP, V7, and V3A, the effect is much more clear cut than in MT. However, despite a strong preference for the presence of stereoscopic gradients, the lack of sensitivity to gradient polarity again indicates that there is probably no specialism for egomotion-consistent depth structure in these areas. The general sensitivity to stereoscopic depth evident in VIP, V7, and V3A is in line with previous human fMRI studies. Backus et al. (2001) demonstrated that the response to two transparent fronto-parallel surfaces formed by dynamic random-dot stereograms is larger than that to a single plane with the same total number of dots. This difference was seen particularly strongly in V3A compared with V1–V3 (V7 and VIP were not examined). Other fMRI studies (Merboldt et al. 2002; Neri et al. 2004; Nishida et al. 2001; Tsao et al. 2003) all found sensitivity in many areas, but collectively they highlight V3A, V7, and the intraparietal sulcus as possible foci of stereoscopic depth processing. Neurophysiological studies also identify V3A as strongly sensitive to stereoscopic depth (e.g., Poggio et al. 1988; Tsao et al. 2003), whereas Shikata et al. (1996) found sensitivity to stereoscopically defined depth surfaces in neurons in a caudal region of the macaque intraparietal sulcus, adjacent to V3A. Another possible interpretation of the results in V3A, V7, and VIP is that the greater response in the presence of depth compared with a flat plane reflects processing of the 3D shape (convex or concave cone) created by the radial disparity gradient. A partially overlapping set of cortical areas has been implicated in this process (Georgieva et al. 2009). However, it should be noted that the cone in our experiment subtended 23-degree diameter, requiring rather large receptive fields to detect it.

In summary, we find sensitivity to stereoscopic depth in various cortical regions, but one region, V6, stands out as a candidate for processing depth in the context of egomotion. We show a clear functional difference between V6 and neighboring areas V3A and V7, supporting the view that V6 has a specialized role in the extraction of visual cues for self-motion.

GRANTS
This work was supported by a research grant to ATS from the Wellcome Trust.

DISCLOSURES
No conflicts of interest, financial or otherwise, are declared by the authors.

REFERENCES


