NeuroImage 122 (2015) 298–305

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Visual search in depth: The neural correlates of segmenting a display into relevant and irrelevant three-dimensional regions

Katherine L. Roberts [a,*], Harriet A. Allen [b], Kevin Dent [c], Glyn W. Humphreys [d]

[a] Department of Psychology, University of Warwick, Coventry, UK
[b] School of Psychology, University of Nottingham, Nottingham, UK
[c] Department of Psychology, University of Essex, Colchester, UK
[d] Department of Experimental Psychology, University of Oxford, Oxford, UK

ARTICLE INFO

Article history:
Received 1 July 2015
Accepted 18 July 2015
Available online 26 July 2015

Keywords:
Visual search
Segmentation
Depth
fMRI

ABSTRACT

Visual perception is facilitated by the ability to selectively attend to relevant parts of the world and to ignore irrelevant regions or features. In visual search tasks, viewers are able to segment displays into relevant and irrelevant items based on a number of factors including the colour, motion, and temporal onset of the target and distractors. Understanding the process by which viewers prioritise relevant parts of a display can provide insights into the effect of top-down control on visual perception. Here, we investigate the behavioural and neural correlates of segmenting a display according to the expected three-dimensional (3D) location of a target. We ask whether this segmentation is based on low-level visual features (e.g. common depth or common surface) or on higher-order representations of 3D regions. Similar response-time benefits and neural activity were obtained when items fell on common surfaces or within depth-defined volumes, and when displays were vertical (such that items shared a common depth/disparity) or were tilted in depth. These similarities indicate that segmenting items according to their 3D location is based on attending to a 3D region, rather than a specific depth or surface. Segmenting the items in depth was mainly associated with increased activation in depth-sensitive parietal regions rather than in depth-sensitive visual regions. We conclude that segmenting items in depth is primarily achieved via higher-order, cue invariant representations rather than through filtering in lower-level perceptual regions.

Introduction

Being able to selectively attend to relevant aspects of the world is critical for efficient information processing (Broadbent, 1958; Tsotsos, 1990). Prioritisation of items of interest can be based on low-level visual features such as colour (Wolfe et al., 1989) or motion (McLeod et al., 1988), or on more complex features such as common temporal onset (Watson and Humphreys, 1997). Understanding the process by which prioritisation is achieved can provide insights into the mechanisms by which cognitive control influences perceptual representations. Neuroimaging has revealed that a region of the precuneus is involved in segmenting a scene into relevant and irrelevant items based on different features (motion, temporal onset; Dent et al., 2011). Activation is also found in the relevant feature-specific regions, such as those representing motion (Dent et al., 2011). Here, we extend this work to investigate the mechanisms involved in selectively attending to items in a relevant 3D region of space. Segmenting a scene into relevant and irrelevant 3D regions can help distinguish steps, kerbs and other hazards, or help find objects in a crowded shop display. We ask whether the same precuneus region involved in segmenting items by motion and time is also involved in segmenting items in depth. We also ask if segmenting items in depth is associated with activation in visual areas tuned to disparity or surfaces, or parietal regions containing higher-order 3D representations.

Visual search tasks have proved to be a valuable tool for evaluating the ability to segment a visual scene into relevant and irrelevant regions. In visual search tasks, participants search for a target item while ignoring irrelevant (distractor) items. When the target is defined by a single feature (e.g. colour), search is highly efficient and not dependent on the number of distractors in the display ('pop-out' search). When the target is defined by a conjunction of features (e.g. colour and form), search time increases with increasing numbers of non-target distractors (Treisman and Gelade, 1980). These data indicate that search is facilitated if participants can segment the scene into relevant and irrelevant items, and can direct their attention to only the relevant subset of items (see Wolfe and Horowitz, 2004, for a review). When participants are able to segment the search display in this way, search is more efficient, with search time reflecting the number of distractors in the attended subset rather than the number of distractors in the entire display. This has been demonstrated with temporal segmentation ('preview search', in which a subset of distractors are presented in advance; Theeuwes,

* Corresponding author at: Department of Psychology, University of Warwick, Coventry CV4 7AL, UK. Fax: +44 024 765 24225.
E-mail address: k.roberts@warwick.ac.uk (K.L. Roberts).

http://dx.doi.org/10.1016/j.neuroimage.2015.07.052

Kramer, & Atchley, 1998; Watson and Humphreys, 1997), colour segmentation (Wolfe et al., 1989), and motion segmentation (in which a subset of items are moving, Dent et al., 2011; McLeod et al., 1988; von Muhlenen & Muller, 2000). There is also evidence that depth cues can be used to segment items in a display (Finlayson et al., 2013; Nakayama and Silverman, 1986), with participants able to perform an efficient 'pop-out' search within an attended depth plane.

Numerous studies have demonstrated that attention can be directed to a specific location in 3D space (e.g. Anderson and Kramer, 1993; Nakayama and Silverman, 1986). However, there is debate over whether attention can be directed to a specific depth (disparity), or whether attention is in fact allocated to surfaces within 3D space (He and Nakayama, 1995). It also seems that there must be considerable separation between the depth planes in order for them to be separately attended (more than 6 min of arc), even though perceptual stereo thresholds are considerably smaller (on the order of seconds rather than minutes; de la Rosa et al., 2008). He and Nakayama (1995) found that participants were unable to attend to items that shared a common disparity if the individual items were tilted forwards or backwards, preventing them from appearing to fall on a common surface. In contrast, the participants were able to attend to items on a plane that was tilted in depth, so that the items formed a surface but were at different disparities. It may be that separate mechanisms are engaged when attention is directed to a specific depth or to a surface in depth. He and Nakayama (1995) found that increasing the separation in depth (disparity) between target and distractor items impaired selective attention when the items were on different planes, but had no effect when those same items appeared to be on a surface that was tilted in depth.

Wheatley et al. (2004) suggested that different surfaces are preattentively segregated. Participants were asked to detect the number of targets that differed in depth from background items. Search was efficient when the target items fell on the same surface, even when it was tilted in depth, but inefficient when items appeared on different depth planes. Preattentive segregation of depth planes has also been demonstrated in multiple-object tracking tasks (Haladjian et al., 2008). Viswanathan and Mingolla (2002), for example, found that performance on a tracking task was improved when targets and distractors were presented in two depth planes rather than one, and when items appeared on tilted surfaces. Interestingly, unlike in the visual search studies described above, a benefit was also found when items appeared within depth-defined volumes. This finding is consistent with results on flanker interference (Anderson and Kramer, 1993), where there is an attentional gradient in depth, with flanker interference decreasing as the separation in depth of targets and flankers increases. These studies indicate that it may be possible to selectively attend to items within a depth-defined region of space, even when those items do not form a common surface. This is in keeping with real-world examples of segmentation search, such as searching for a friend arriving at a train station where we may exclude from search (a) people who have been on the station for some time (segmentation by time/preview search); (b) people who are stationary (segmentation by motion); and (c) people who are nearer or further down the platform, who are unlikely to be coplanar (segmentation in depth).

In the present study, we examined the neural basis of segmenting items in depth, using fMRI while participants performed a difficult search task, in which they did or did not know the likely 3D location of the target. Participants searched for a target among distractors, with items appearing in front and behind the fixation plane. Displays were identical in the two top-down segmentation conditions, the only difference being that when the 3D location of the target was known, participants could use this information to segment the scene into relevant and irrelevant items, searching only the relevant items (Factor 1. Target depth known versus unknown; Fig. 1A). Two further factors were included to separate effects of attending to depths and surfaces. Displays were either vertical (fronto-parallel) or tilted backwards 45° (Factor 2. Display type: vertical or tilted; Fig. 1B). Within the display condition, letters in front and behind fixation were either presented with a common disparity (so that letters formed planes at different depths) or within depth-defined volumes (so that letters were jittered in depth and did not form planes) (Factor 3. Letter placement: planes or jittered; Fig. 1C). Depth regions were therefore defined by either common disparity (vertical displays, planes), common surfaces (vertical and tilted displays, planes) or a depth-defined region of space only (vertical and tilted displays, jittered). Comparing activation when target depth was known versus unknown, for the different display types and letter placement conditions, allows us to isolate activation associated with selectively attending to a specific region of 3D space, whether defined by a common disparity, common surface, or depth-defined regions.

Previous work has indicated that segmentation by time and motion activates a common region of the precuneus, as well as task-specific regions (Dent et al., 2011). The precuneus is likely to be involved in maintaining a spatial representation of distractor locations, and activity in this region is correlated with the amount of benefit obtained from segmenting the display (Dent et al., 2011). During segmentation by motion, activation was also found in motion-processing areas (Dent et al., 2011). In this case, segmentation may be at least partially based on a motion filter in the feature-specific region MT/MT+, which is used to guide attention to moving items and to filter out stationary items (Ellison et al., 2007; McLeod et al., 1988). We hypothesise that segmenting items according to their 3D location will recruit the same region of the precuneus as that identified by the previous segmentation tasks, demonstrating the supramodal nature of visual segmentation in this brain region (Dent et al., 2011). We are also interested in whether segmenting items according to their 3D location leads to increased activation in visual regions sensitive to depth perception (kinetic occipital area (KO), motion area MT/MT+, and lateral occipital cortex (LO)), or in higher-order depth-sensitive regions along the intraparietal sulcus (IPS) (Preston et al., 2008). If segmenting items in depth is achieved through filtering in depth-sensitive visual regions, we might expect increased activation in the target-known condition, as is the case in MT/MT+ when participants attend to motion (e.g. Dent et al., 2011; Saenz et al., 2003). An alternative possibility is that visual regions may show reduced activation due to attention being focused on only part of the display.

We are also interested in whether activity in the visual and parietal depth-sensitive regions is cue invariant, or if it depends on the cues available to target depth (i.e., common disparity, common surface or depth-defined regions). Note that we use the term 'segmentation' to refer to dividing a search display into relevant and irrelevant items according to a specified feature; in this case, their 3D location. This is distinct from the perceptual process of segmenting a visual scene into surfaces and objects. Similarly, 'cues to target depth', refers to the cues available to the participant to aid them in dividing the scene into relevant and irrelevant items (e.g. the possible range of target disparities). This is not the same as 'cues' available for depth perception, such as occlusion and motion parallax.

Materials and methods

Participants

Seventeen participants took part. Data from one participant had to be excluded due to excessive movement during the imaging session. The analyses are based on data from the remaining 16 participants (5 male, mean age 23 years (19 to 33 years), all reported being right-handed). All participants gave written informed consent and received £20 compensation.

Stimuli and design

The task was to search for a target letter (Z or N) among distractor letters (H, I, V, X), and indicate with a button-press response whether