Slant cue are combined early in visual processing: Evidence from visual search

Rita Sousa *, Eli Brenner, Jeroen B.J. Smeets

Faculty of Human Movement Sciences, Vrije Universiteit, Van Der Boechorstraat 9, 1081 BT Amsterdam, The Netherlands

ARTICLE INFO

ABSTRACT

When looking for a target with a different slant than all the other objects, the time needed is independent of the number of other objects. Surface slant can be inferred from the two-dimensional images on the retinas using various cues. The information from different cues is subsequently combined to get a single estimate of slant. Is information from the individual cues or from the combined percept responsible for us so easily finding the target? To find out we compared combinations of two slant cues. The cues that we chose are retinal shape and binocular disparity. We compared search times for conditions with the same differences between the target and the other objects in each individual cue, but for each object the two cues either indicated the same slant or opposite slants. Search times were independent of the number of other items if the target clearly differed in perceived slant from the other items. Subjects systematically found the target faster when the cues indicated the same slant. We conclude that slant cues are combined locally throughout the visual field before the search process begins.

1. Introduction

When people view a surface with a certain orientation in depth, they use several cues to infer the 3D slant from the 2D information on the retinas. Such cues include retinal shape and binocular disparity. The cues are then combined to get a single estimate of slant that is more accurate than the estimate based on each of the cues alone. Do the individual cues play any role in perception? It is generally believed that the slants indicated by the individual cues are not accessible (e.g. Hillis, Watt, Landy, & Banks, 2004). If the cues are in conflict, this conflict is visible as a change in some other attribute than slant (Muller, Brenner, & Smeets, 2007). Conflicts can arise naturally, through errors due to the limited resolution of each cue or when an assumption on which a cue is based is violated (e.g. Ames room).

We here use a visual search task to examine whether information from the individual cues is used in early visual processing. Visual Search is a task in which subjects have to find a target between other elements. They have to do so as quickly as possible. Obviously, if the target is salient (very different from the other elements) one will find it faster than if it is quite similar (see Rosenholtz, 1999). However, Treisman and Gelade's (1980) "feature integration theory of visual attention" claims that another more fundamental distinction can be made (see also Overvliet, Smeets, & Brenner, 2007). It proposes that there are simple features that are analyzed automatically early in visual processing, and that if the target differs from surrounding objects in such a simple feature, the number of other items is irrelevant (parallel search pattern). If the search requires combining features (conjunction search), or if one searches for the absence of a feature, it takes longer to find the target when there are more objects (serial search pattern).

Several studies revealed that there is more to visual search than this simple distinction. For instance, Nakayama and Silverman (1986) showed that conjunctions can sometimes be processed in parallel, whereas Wolfe (2001) discusses the many intricacies of asymmetries in visual search. Nevertheless there is a reasonable consensus about the fact that some distinctions are quite automatic and involve processing that takes place in parallel across the visual field, leading to a parallel search pattern.

Holliday and Braddick (1991) found that stereoscopically defined slants could be processed in parallel. They took various precautions to ensure that subjects had to rely on binocularly defined slant. Participants had to press one key if all the items had the same slant and another key if one item had a different slant from the others. The response time was almost identical when there were 3, 5 or 9 items in the display. Enns and Rensink (1991) found that the response times in a similar task with monocularly defined slants also hardly depended on the number of items. In their experiments there were 1, 6 or 12 items and subjects had to indicate whether or not a certain item was present. Their results suggest that the items' three-dimensional orientations were extracted from the two dimension drawings early in visual processing. However, Epstein and Babler (1990) found that even with both binocular and monocular cues

* Corresponding author.
E-mail address: ritass@gmail.com (R. Sousa).

present, search for a different slant was not always parallel. They suggested that the precise instructions mattered, implying that the type of processing was not only determined by the cues involved. In a different context, He and Nakayama (1992) showed that search efficiency depends on the 3D interpretation of the scene rather than the 2D image features (also see Aks & Enns, 1996). For slant such an interpretation involves combining the available slant cues, so performance when searching for a different slant may depend on the salience of the combined percept that arises from averaging the cues involved, and not on the salience of the individual cues.

Here, we use various combinations of two slant cues to examine whether the feature that allows one to directly find the target, irrespective of the number of other items, is one or both of the individual cues, or whether it is the slant that is determined from combining the cues. We compare search times for targets that differ from other items in the direction of slant. In a *consistent-cues* condition the two cues indicate the same slant value. The target is slanted with its top nearer than its bottom, whereas the bottom is nearer in the other items. In an *inconsistent-cues* condition the two cues always indicate opposite values for slant. The binocular cue has the same values for the target and the other items as in the *consistent-cue* condition, but the values of the monocular cue were switched between the target and the other items. In neither of these cases could the target be distinguished by its simulated shape.

If a single cue allows subjects to quickly find the target (this would be the cue that is processed fastest), there should be no difference in the time taken to find the target between these two conditions, because the difference between the conditions is not in the single cues but in the combination. Conversely, if there is a difference in performance between these conditions, the subjects are probably relying on the combined percept. If so, we may expect the condition with the cues in conflict to give slower parallel search, because there will still be a difference in perceived slant but it will be smaller. The above reasoning is valid as long as the combined slant estimate is not dominated by a single cue, because if it is, then relying on the combined percept also predicts that there will be no difference between the conditions. We therefore included two more conditions in which only one of the cues was varied, which we will refer to as *binocular* and *monocular* conditions to indicate the cue that specifies the target.

2. Methods

2.1. Subjects

Ten subjects participated in this study. Two of them were authors and the other eight were experienced with psychophysical tasks but naive as to the purpose of the experiment. All the subjects had normal binocular vision. Four other subjects and one author also took part in a control experiment.

2.2. Experimental setup

The setup consisted of an Apple G5 computer that generated the images and processed the responses, a 57 cm (diagonal) Sony Trinitron monitor (resolution 1096 × 686 pixels), and Crystal Eyes stereo shutter spectacles. Images were generated at a refresh rate of 160 Hz (80 Hz per eye) and alternate images were presented to the two eyes with the help of the shutter spectacles. The spectacles isolate the images for the two eyes best for red images so we used red items (except for the 'cursor' that was always at the same position on the screen for both eyes). Observers sat 70 cm from the screen.

2.3. Stimuli

The items were based on simulated red squares with sides of 4 cm on a black background (Fig. 1). All the items were on an invisible 10 cm radius circle around a 4-mm diameter green dot at the center of the screen. They were spaced regularly along this circle, but not at fixed positions. Each square was rotated by a random amount between −5 and 5 degrees within its own plane so that the edges of adjacent squares were not parallel to each other. Each square's slant was defined by two cues: binocular disparity and retinal shape. The retinal shape (a monocular cue) was manipulated independently of the binocular disparity by considering the viewing geometry from a position midway between the subject's eyes. The binocular disparities were in accordance with the actual viewing geometry. Each cue always indicated a slant of either 70 or −70 degree, where 0 degrees is the frontoparallel plane and positive angles indicate that the target is slanted so that its top is closer to the observer (forwards) and negative ones that the target is slanted in the opposite direction (backwards).

Besides items in which the cues indicated the same slant (i.e. simulations of a slanted square) we also used items in which the cues indicated different slants. Although each cue had equivalent values to those used to simulate the squares, the simulations correspond with slanted non-squares; and indeed looked like slanted non-squares to our subjects. Different subjects probably judged these shapes and slants slightly differently, because the weights given to these cues differ between people, but all items always clearly looked slanted. No subjects ever reported difficulties in judging the surfaces' slants or seeing the slant change during a trial.

Fig. 2 summarizes the four conditions of the main experiment. In the *consistent* condition, the target was slanted forward according to both cues (70°/70°) and all the other items were slanted backwards according to both cues (−70°/−70°). In the *inconsistent* condition, the target was slanted backwards according to the monocular cue and slanted forwards according to the binocular cue (−70°/70°). The other items had the opposite slant for each cue: forwards for the monocular cue and backwards for the binocular cue (70°/−70°). For both these conditions the task was to find the surface with a different slant.

In the two other conditions only one of the cues differed between the items. Since the difference in shape was more conspicuous than the difference in slant in these conditions, we instructed the subjects to look for a particular shape, specifically to find a square among non-squares. The non-squares were identical to the non-targets in the inconsistent condition. The square target was one of the two kinds of items in the consistent condition. In the *binocular* condition, the target only differs from the other items in the binocular cue. In the *monocular* condition the target only differs from the other items in the monocular cue.

In the control experiment there were two conditions. The first condition, *control same*, was similar to the monocular condition described above, except that the angles were reduced from ±70° to ±30° from frontal, and the target was the non-square and the other items were squares (rather than the other way around). As before the non-target items all had the same orientation. The second
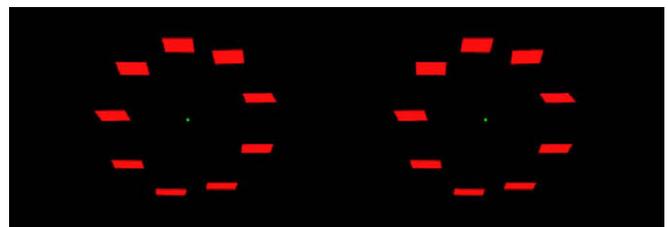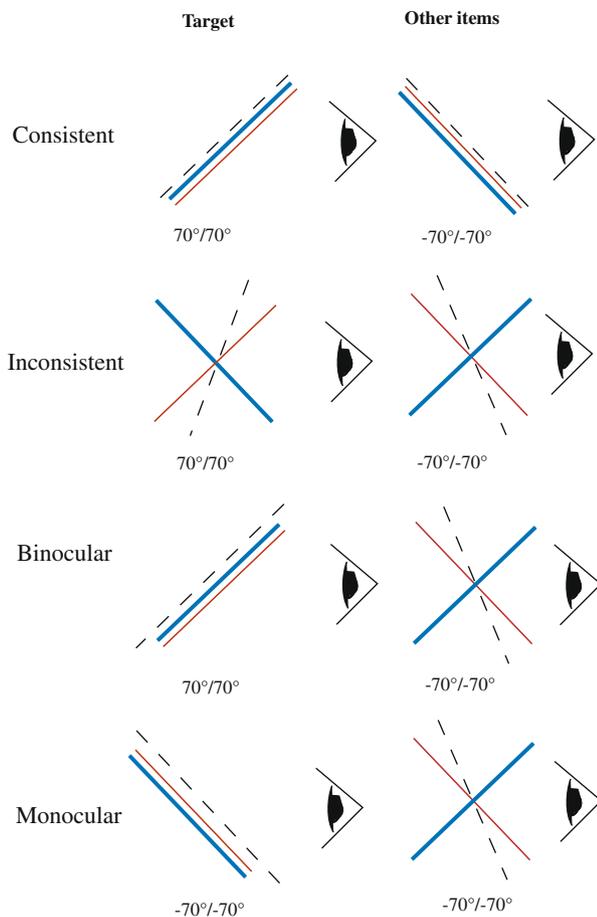

Fig. 1. A stereogram illustrating the display.

Fig. 2. Schematic representation the four simulated viewing conditions of the main experiment. The thick blue lines represents the monocular cue and the thin red ones the binocular cue. The dashed lines represent the perceived slant. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

condition, *control different,* was the same except that the (non-target) squares could have either orientation.

2.4. Procedure

The experiment was a visual search task. Subjects had to find the target among the other elements. The response was made by moving the green dot (the cursor) with the mouse. When the green dot was 7.5 cm from the center (which was its initial position on each trial), the nearest item was considered to have been chosen. The dot then jumped back to the center and only this dot was visible for 1 second before the next trial began.

Two variables were manipulated: the number of non-targets (2, 4, 6 and 8) and the cues that indicate slant (the four conditions). Since we had two different tasks in the main experiment, to find the differently slanted surface or to find the square, half the participants started with the consistent and inconsistent conditions (and the slant task) and the other half with the monocular and the binocular conditions (and the shape task). Each condition within these pairs of conditions had a block of 40 trials for each number of items. These eight blocks were presented in a random order, and subjects did three complete series for each condition before switching to the other task.

2.5. Predictions

Independent of whether search takes place after or before the cues are combined, we expect the consistent and inconsistent con-

ditions to both give a parallel search pattern. If we do not find this, we will be unable to interpret our results in terms of automatic simultaneous processing throughout the visual field.

If the difference in the value of an individual cue allows people to directly find the target, the time taken to find the target will be the same for the consistent and inconsistent condition, because for each cue individually the two conditions are equivalent. If the binocular cue is processed much faster than the monocular cue, the consistent, inconsistent and binocular conditions should take about the same amount of time. If the monocular cue is processed much faster, the consistent, inconsistent and monocular conditions should have about the same timing. If both cues are processed about equally fast, responses to the combination may be faster than for each individually because when there are two cues the target will be found when a difference in either is detected.

If search takes place after the cues are combined, it may take longer to find the target in the inconsistent condition because the conflict reduces the apparent difference in slant. The weights given to the cues probably also differ slightly at the different positions in the display, because of the different viewing geometry, so there may be some variability between the combined slant estimates for the different non-target items (as well as the clear differences between the perceived shapes) in this condition, which may also decrease the detectability of the target (Duncan & Humphreys, 1989). In the monocular and binocular conditions we may find a serial search pattern, because one is then looking for a slanted square among various other four-sided slanted objects, which is likely to be less efficient (although we could not be sure of this). In these two conditions subjects' performance should be identical because the simulated shapes are the same (assuming that subjects only rely on the shape and not on the different slant). We will return to the issues of slant magnitude, the heterogeneity of the non-targets, and the independence of judgments of shape and slant, when discussing the results of the control experiment.

2.6. Analysis

The first of the three series for each task was considered to be a practice series and was not analyzed. Trials in which the subjects selected the wrong item were also excluded. We determined the median search time for each participant, condition and number of items on the remaining trials of the second and third series and averaged these two median values. A repeated measures ANOVA was performed to evaluate the consistency across subjects of the effect of the condition and number of items on these average values. Slopes of search time versus number of items (linear fit) were also determined per subject and condition, and differences in slope between the conditions were evaluated with second repeated measures ANOVA.

3. Results

The total percentage of errors was 5.6% (3.7% in the monocular condition, 2.6% in the binocular condition, 6.3% in the consistent condition and 9.8% in the inconsistent condition) independent of the number of items in the display. Fig. 3 shows the average of the ten subjects' median search times in the four conditions. It is evident from the figure that the search pattern is parallel for the consistent (slope ± SE: −27 ± 11 ms/item) and inconsistent conditions (−27 ± 7 ms/item) while it is serial for the binocular (73 ± 18 ms/item) and monocular conditions (164 ± 22 ms/item). It also takes less time to find the target in the consistent condition than in the inconsistent condition. All these findings are consistent with search taking place after the cues are combined. The ANOVA on the data of the main experiment confirmed that search times
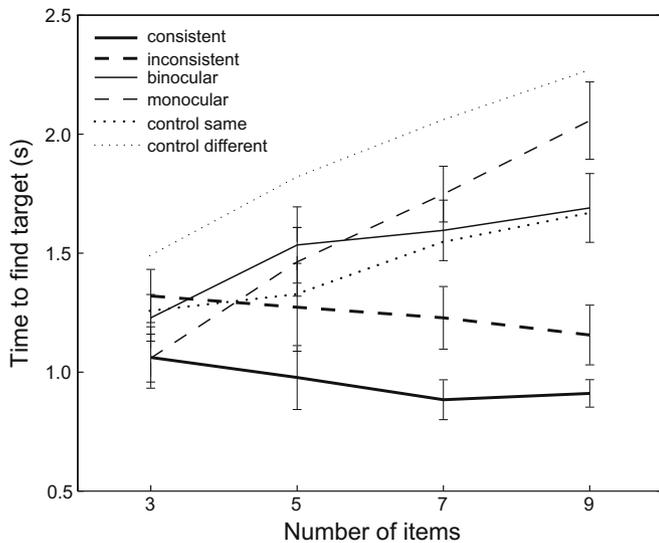
Fig. 3. Average time taken to find the target in the four main conditions (with standard errors) and in the two control conditions.

differed between the conditions, depended on the number of items in the display and that there was an interaction between these two factors (all $p < .0001$). Post hoc tests (Fisher PLSD) showed that all the averages for all the conditions differed from each other except for the binocular and monocular conditions. There were also significant differences between the slopes (ANOVA; $p < .0001$), whereby only those for the consistent and inconsistent conditions were not significantly different from each other (Fisher PLSD). The results of the control (dotted lines in Fig. 3) show that even when searching for a non-square among objects that are all squares, the slant of the squares does matter (difference between the two conditions). The search times for the *control same* condition (all squares slanted the same way) were very similar to those for the *monocular* condition of the main experiment, so neither the smaller difference in slant nor the fact that the target was not always the same shape seems to matter very much.

4. Discussion

The subjects took more time to find the target in the inconsistent condition than in the consistent one. They had a parallel search pattern for both these conditions while they had a serial search pattern when looking for the square in the binocular and monocular conditions. Thus the search time is clearly not only determined by the values of the individual cues. When considered in terms of perceived slant after cue combination, the difference in slant in the consistent condition is larger than that in the inconsistent condition, so our findings are in agreement with the general principle that the bigger the difference between target and distracters, the less time it takes to find the target (Duncan & Humphreys, 1989). Moreover the perceived slants may also be more variable when they are combinations of inconsistent cues, because the weights given to the cues may not be the same throughout the stimulus. However, the relevant difference may not be constrained to the key dimension for the task, because the control experiment showed that adding variability in slant increased the search times for a non-square shape. Thus the longer search times for the slant in the inconsistent condition may also be caused by the larger variability in perceived shape. Note that all these explanations rely on the perceived slant or shape, not on the values of the individual cues, so they are consistent with

the notion that the search is based on the combined percept and not on the individual cues.

We mentioned finding a parallel search pattern in the consistent and inconsistent conditions, but it would appear from Fig. 3 that the slopes are negative, indicating that there is an improvement in search as the number of other items increases. This pattern has previously been reported in studies in which subjects were looking for the odd one out, rather than for a particular shape or slant (Song & Nakayama, 2006; Song, Takahashi, & McPeek, 2008). Perhaps the fastest responses in our study were guided by a certain object not appearing to fit within the pattern, rather than it specifically being recognized as having a certain shape or orientation. This would explain why differences in slant influenced search based on shape (control experiment), and possibly vice versa (difference between consistent and inconsistent conditions), and could even be responsible for the fact that searching for a different shape (monocular and binocular conditions) led to a serial search pattern, because the cue conflict stimuli differed considerably in shape. One way of seeing this is that even if subjects are trying to find a square, they may find it on the basis of its slant if the square always has a very different slant than the other items. Not that subjects neglect the instructions, but they may be guided to the square by its different slant, and then check whether it is indeed a square. Indeed, some subjects in our study noticed (as revealed by later questioning) that the square that they were looking for also had a different orientation. We find a steeper slope for the monocular than for the binocular condition, which is consistent with this proposal considering the subjective impression that the perceived slant was strongly biased towards the binocular cue in our experiment. The above reasoning could also account for Epstein and Babler's (1990) findings, because although in their experiments the target had a different orientation from the other items, all items in the display had different shapes. The different shapes probably made it much harder to detect the target so that the subjects had to check each item to determine whether it was the target (serial search) until they were trained to focus on the slant.

The difference between the search patterns for slant and shape cannot simply be explained as slant being easier to judge than shape, because with two distracters the shape and slant judgments take about the same amount of time. Thus shape does not take more time to process than slant. The difference between the search patterns is presumably the result of all other items than the target having the same slant, whereas each had a different simulated shape (Rosenholtz, 1999), so we cannot consider this as evidence for a fundamental distinction between judgments of slant and shape. As argued above, the distinction probably depends on both slant and shape judgments for both tasks. Importantly, it is these judgments that determine the search pattern, and not the values of the cues before they are combined.

Our research confirms that the surface representation and not image features determine the search pattern in a visual search task (Aks & Enns, 1996; He & Nakayama, 1992). Our study is similar to that of He and Nakayama (1992) in that they too varied the surface representation across conditions, in their case through perceptual completion, without changing the basic features. In their experiment perceptual completion was critical for finding a parallel search pattern, and any information that interfered with it influenced this search pattern. In our study a clear difference in perceived slant was critical, and the image features were the two slant cues. Our study obviously does not prove that we have lost access to these individual cues, but it supports the notion that the cues are combined at a very early stage of visual processing. In particular, that these slant cues are combined before (parallel) search, and therefore simultaneously throughout the visual field.

References

Aks, D. J., & Enns, J. T. (1996). Visual search for size is influenced by a background texture gradient. *Journal of Experimental Psychology: Human Perception and Performance, 22*(6), 1467–1481.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433–458.

Enns, J. T., & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review, 98*(3), 335–351.

Epstein, W., & Babler, T. (1990). In search of depth. *Perception & Psychophysics, 48*(1), 68–76.

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 359*, 231–233.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*, 967–992.

Holliday, I. E., & Braddick, O. J. (1991). Pre-attentive detection of a target defined by stereoscopic slant. *Perception, 20*, 355–362.

Muller, C. M. P., Brenner, E., & Smeets, J. B. J. (2007). Living up to optimal expectations. *Journal of Vision, 7*(3), 1–10. 2.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320*, 264–265.

Overvliet, K. E., Smeets, J. B. J., & Brenner, E. (2007). Parallel and Serial Search in Haptics. *Perception & Psychophysics, 69*(7), 1059–1069.

Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research, 39*, 3157–3163.

Song, J. H., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision, 6*, 982–995.

Song, J. H., Takahashi, N., & McPeek, R. M. (2008). Target selection for visually guided reaching in macaque. *Journal of Neurophysiology, 99*, 14–24.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics, 63*(3), 381–389.