Visual search for motion-in-depth: stereomotion does not 'pop out' from disparity noise

Julie M. Harris[1], Suzanne P. McKee[2] and Scott N. J. Watamaniuk[3]

[1] Department of Psychology, Ridley Building, Claremont Place, University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK

[2] The Smith-Kettlewell Eye Research Institute, 2232 Webster St., San Francisco, California 94115, USA

[3] Department of Psychology, Wright State University, Dayton, Ohio 45435, USA

Correspondence should be addressed to J.M.H. (Julie.Harris@ed.ac.uk)

In a visual search task, targets defined by motion or binocular disparity stand out effortlessly from stationary distractors ('pop-out'), suggesting that target and distractors are processed by different neural mechanisms. We used pop-out to explore whether motion directly toward or away from the observer (z-motion) is detected using binocular motion cues. A target moving laterally (x-motion) popped out amid stationary distractors with binocular disparity, but z-motion did not pop out. However, a small x-motion added to the target's z-motion caused it to pop out. We therefore suggest that the visual system may not be specifically sensitive to binocular motion differences.

The human visual system has evolved to enable us to perceive and act in a three-dimensional (3D) world. Both binocular disparity and motion are important sources of information for human visual perception, and specialized mechanisms are thought to be dedicated to their processing[1–5]. To perceive object position in a stationary scene, the stereoscopic system can use the small differences between the left and right eyes' views of an object (binocular disparity) to obtain information about depth[2,6]. However, the visual system is also particularly sensitive to image motion, which is unsurprising as the world outside the laboratory almost always contains moving observers watching moving objects. Here we test whether the visual system has specific sensitivity to the small binocular motion differences that occur when an object moves toward or away from an observer (z-motion). In this paper, we use the term stereomotion to refer specifically to stereoscopically defined motion-in-depth. Although there is abundant neurophysiological evidence that cortical motion neurons are tuned for both disparity and motion[7], there is scant evidence for neurons that respond specifically to stereomotion[8].

We used a visual search task in which an observer was asked to detect a moving target element among a group of stationary distractor elements. The visual search paradigm has been used to explore which stimulus dimensions (e.g. color, motion) are easily distinguished from others[9]. This technique can be thought of as probing whether target and distractor properties are processed by the same or different neural mechanisms[10]. For example, a target element defined only by motion is effortlessly detected (pop-out) amid identical but stationary elements[11–13], because moving and stationary targets are processed by different neural mechanisms. In this study, we manipulated the monocular motion components to generate the perception of either lateral (x) motion or stereo (z) motion. We report that whereas x-motion pops out among stationary distractors, z-motion does not.

Results

POP-OUT OF STEREOMOTION

Four experimental conditions were used (**Fig. 1**): a target dot moving laterally (x-motion) in a field of two-dimensional (2D) stationary distractors (zero disparity); a target dot undergoing x-motion in a field of 3D stationary distractors, each having a randomly chosen binocular disparity; a target dot undergoing 3D motion directly toward the observer (z-motion) in a field of 2D stationary distractors; and a target dot undergoing z-motion in a field of 3D stationary distractors. Observers were shown two stimuli in succession. One contained only distractors, the other distractors plus the target. Percent correct for detection of the interval containing the target was measured for a stimulus containing 100 dots.

By definition, we assumed that a specific motion-sensitive mechanism would not be affected by purely positional noise. Thus, pop-out should occur for any condition in which the target activates a motion-sensitive system that is distinct from the positional mechanism that registers the position of the distractors.

As expected, a target defined by x-motion was readily detected even in the presence of a large number of either 2D or 3D distractors (**Fig. 2**). Performance for a target undergoing z-motion in 2D noise was similar to that for x-motion (**Fig. 2**). However, in principle the target could have been distinguished from the distractors here either by its stereomotion or simply by its binocular disparity alone. When the target underwent z-motion, it started behind the fixation plane and moved to be in front of the fixation plane. Thus the target would have very different disparity from the distractors at the beginning or end of the trial and thus could be detected using disparity based on the beginning or end of the stimulus presentation.

Testing z-motion amid 3D distractors distinguished between the use of purely disparity-sensitive systems and systems that respond to stereomotion (which might rely on either registering changing disparity or a combination of monocular

article

Fig. 1. Experimental design. **(a)** Observers viewed an array of 3D distractor dots, randomly arranged in a notional end-on cylinder. The fixation point was identical to the distractor dots and positioned at the center of the cylinder (shown here as a gray dot for illustrative purposes). **(b)** Schematic of the four stimulus condition used: target undergoing x-motion amid 2D distractors; target undergoing x-motion amid 3D distractors; target undergoing z-motion amid 2D distractors; and target undergoing z-motion amid 3D distractors.
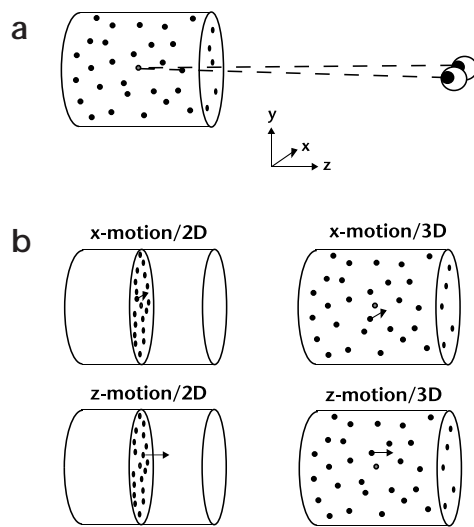
motion signals[14,15]) in detecting the target. Because the distractors had randomly chosen disparities, the target could not be detected on the basis of its disparity alone in any one frame of the motion sequence. The only feature distinguishing the target from the distractors was its z-motion. For this condition, the target was very much harder to detect, suggesting that there is no specific z-motion sensitivity (**Fig. 2**).

If target and distractors are detected by a common neural mechanism, then detection of the target will become more difficult as the number of distractors increases. Therefore, we repeated the experiment for x-motion with 2D distractors and z-motion with 3D distractors, but measured percent correct for detection of the target dot as a function of the number of distractors (**Fig. 3**). Note that with only one distractor present (the fixation point), all observers had over 90% correct performance, which was similar for the x-motion and z-motion conditions (although there are inter-observer differences: JMH found x-motion slightly easier, SPM found z-motion easier, and SNW's performance was the same for both conditions). However, z-motion does not pop out for a large range of different conditions; it is much harder to detect than x-motion, even for a display containing as few as eight distractors.

Given that z -motion in the world corresponds to approximately equal and opposite x-motion in the left and right eye, it is startling that the z-motion is almost undetectable, whereas the monocular components are highly visible. If the observer were to view only the signal presented to one eye, the x-motion of the target would make it instantly pop out. In fact, the x-motion stimulus with 2D distractors consists of one stereo half-image from the z-motion stimulus with 3D distractors, but viewed by both eyes. (Note that very similar results were found when this stimulus was shown to one eye only.) Thus, simply changing the direction of the signal in one eye from rightwards to leftwards (and hence from x-motion to z-motion) greatly degrades performance. This is even more surprising when one considers that there are at least two potentially independent sources of information that the observ-

er could use to detect the motion in this display[8]. First, it is possible to measure the disparity of the target through the motion sequence and use the fact that disparity has changed to detect the target. Second, each eye received a motion signal that would have been easily detectable if presented alone. Either eye's signal, or a comparison of the two, could have been used to detect the motion. Therefore, although each eye in the z-motion condition received a target signal that would pop out when viewed monocularly, showing the views together to produce a fused 3D percept prevented those monocular signals from being available for target detection.

Our data raise two questions. First, why does performance for z-motion fall as the number of 3D distractors is increased? We interpret this as meaning that the target and distractors are processed by the same visual mechanism. The distractors in our display consist of stationary 3D noise, which implies that detection of z-motion uses 3D position-based mechanisms, suggesting that there is no special sensitivity to stereomotion.

Second, why does the visual system not have access to the monocular motion signals that impinge on each eye? One possibility is that incoming signals from the right and left eyes are fed into a relatively simple mechanism that cannot use the subtle interocular motion differences, but simply averages them. Our experiment used z-motion, which has the unique property of giving rise to almost equal and opposite motion signals in the two eyes. If averaging occurred, it would result in zero motion for such a stimulus. An averaging mechanism for binocular motion would stand in contrast to the more sophisticated position system, which uses the difference between left and right eye images (binocular disparity) as a cue for relative depth. Our second experiment addressed whether the theoretically useful difference information was being discarded in this way.

POP-OUT WITH COMBINED X-MOTION AND Z-MOTION
A small amount of x-motion was added to the z-motion of the target, resulting in a slightly different 3D trajectory (**Fig. 4**). The motion signals were still of opposite sign in each eye, but now of slightly different magnitude; signal averaging would result in a detectable motion signal, so this hypothesis predicts that the target should pop out.

Percent correct for detection of the target was measured for the x-only, z-only and x- plus z-motion conditions (**Fig. 4**).
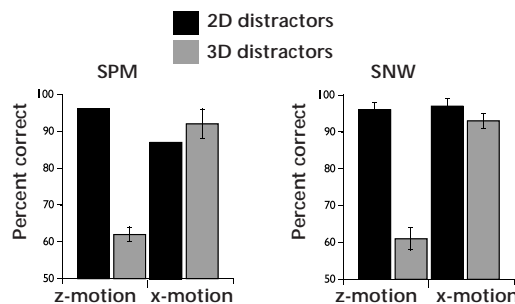
Fig. 2. Percent correct for detecting the interval containing the moving target dot for two observers, SPM and SNW. For each condition, retinal motion was at 4.13 min/s (corresponding to 0.04 m/s in depth for the z-motion condition). Error bars show standard errors.
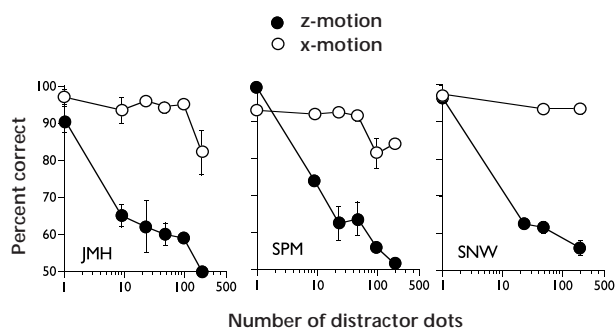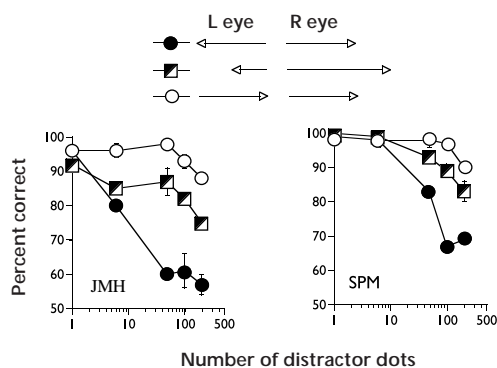
Fig. 3. Percent correct as a function of the number of distractors for x-motion amid 2D distractors (ᴏ) and z-motion amid 3D distractors (●). Each graph is for a different observer. Performance falls dramatically for the z-motion condition as the number of distractors is increased. Error bars show standard errors.

As predicted, adding even a tiny x-motion to the z-motion greatly improved performance. However, note that x-motion alone remained easier to detect than the combined signal. This is consistent with the averaging hypothesis because the average motion for the x- plus z-motion condition had a magnitude that was half that of the x-motion alone. In current experiments, we are using a wide range of x- and z-motion combinations to explore this issue further.

Discussion

To summarize, we suggest that z-motion may be detected by 3D position-based mechanisms (stationary binocular disparity), rather than by distinct mechanisms that respond to changing 3D position. Recent work has debated whether a ratio of left and right eye signals, or rate-of-change of disparity is the binocular motion cue used for 3D motion perception[14,15]. Our

Fig. 4. Percent correct as a function of the number of distractors in three experimental conditions for two observers. For JMH, a motion-in-depth of 0.04 m/s was used; for SPM, motion-in-depth of 0.06 m/s was used. Conditions were z-motion only amid 3D distractors (●), z-motion plus x-motion amid 3D distractors (◨), and x-motion amid 3D distractors (ᴏ). The combined x- and z-motion signal was constructed so that one eye's motion signal had half the magnitude it had in the x-motion condition, and the other had 1.5 times that magnitude. Note that this results in an average retinal motion of half that for the x-motion condition. The two graphs show that adding a small amount of x-motion to the z-motion condition greatly improved performance. Error bars show standard errors.

results suggest that although both sources of information are available in our stimulus, neither is used by the visual system as an explicit source of 3D motion information for this task. Our explanation could account for Tyler's classic finding[16] that the absolute threshold for z-motion is higher than for its monocular x-motion components. Westheimer suggested that the monocular motion signals were canceled in z-motion[17]. We also suggest that the local motion system averages the left- and right-eye velocity signals, thereby discarding binocular differential information about motion.

Many studies in the current neurophysiological literature suggest that motion-sensitive cells are also tuned for binocular disparity[7, 18]. However, this does not necessarily mean that these neurons are specifically responsive to stereomotion. Rather, they may be involved in identifying the position in depth of a moving object. For example, at any particular time, the interocular position difference of an object (whether moving or not) specifies its position in depth. A neuron that is responsive to this position difference may also be responsive to motion. However, our data suggest that if such neurons are involved in the perception of both motion and disparity, then any differences in monocular speeds or directions are not used to detect stereomotion. A system based on such neurons would allow a target moving laterally in front of another moving surface to be highly visible, whereas a small z-motion of the target would hardly be noticed.

Methods

STIMULI. Stimuli were generated and displayed using an Amiga 3000 computer. Each stereo half-image was presented on a x-y CRT screen with P4 phosphor. Images were superimposed for stereo viewing using a beamsplitter and viewed at 1 m. Polarizing filters in front of eyes and screens guaranteed that only one screen was visible to each eye.

The stimuli consisted of a variable number of stationary noise dots in a 3D cylinder of diameter 5.66 cm (subtending 3.24 degrees at the 1 m viewing distance) and simulated length in depth of 5.66 cm (giving 11.64 arcmin binocular disparity). The stationary distractor dots and a single target dot each subtended approximately 2 minutes and had a luminance of 6 cd/m^2 (for details of luminance measurement see ref. 19). Observers fixated a point in the center of the 3D cylinder.

TASK. In a two-alternative, forced-choice procedure, the observer was shown two intervals, one containing only the stationary noise dots, the other containing the moving target dot plus the stationary distractors, and was asked to judge which interval contained the target dot. Each interval was presented for 500 ms (at a frame rate of 30 Hz).

EXPERIMENTAL CONDITIONS. In the first motion frame, the target appeared behind the fixation point. Its position was arranged such that halfway through the motion sequence the target would pass through the fixation plane. We used relatively slow target speeds (typically 0.04 m/s for z-motion) to ensure that the stimuli were binocularly fused. At faster speeds, a target undergoing z-motion was easier to detect. However, there could be at least two reasons for this. First, the fusion range for small foveal targets can be 10 arcmin or less[20]. Any failure of fusion could introduce a small amount of x-axis motion, which would be readily detected among 3D distractors. Second, we were interested in comparing the sensitivities of putative z-motion and x-motion mechanisms. The large change in disparity associated with faster speeds might have been detected by a mechanism responsive to disparity or position. Target motion was well above motion detection threshold, as is demonstrated by almost perfect detection for the conditions with no stationary noise. In the first experiment, the z-motion speed was 0.04 m/s, corresponding to 4.13 minutes/sec on each retina. The target dot thus covered 2 cm in depth and 2.07 minutes on each retina during the duration of the stimulus presentation.

article

Acknowledgments

1. Sekuler, R. & Ganz, L. Aftereffect of seen motion with a stabilized retinal image. *Science* **139**, 419–420 (1963).
2. Wheatstone, C. Contributions to the physiology of vision. Part the first, on some remarkable and hitherto unobserved phenomena of binocular vision. *Phil. Trans. R. Soc.* **148**, 371–394 (1838).
3. Julesz, B. *Foundations of Cyclopean Perception* (University of Chicago Press, Chicago, 1971).
4. Nakayama, K. & Tyler, W. Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Res.* **21**, 427–433 (1981).
5. Westheimer, G. & McKee, S.P. What prior uni-ocular processing is necessary for stereopsis? *Invest. Ophthalmol. Vis. Sci.* **18**, 614–621 (1979).
6. Berry, R.N. Quantitative relations among vernier, real depth and stereoscopic depth acuities. *J. Exp. Psychol.* **38**, 708–721 (1948).
7. Maunsell, J.H.R. & Van Essen, D.C. Functional properties of neurons in middle temporal visual area of the macque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.* **49**, 1148–1166 (1983).
8. Cumming, B. in *Visual Detection of Motion* (eds. Smith, A.T. & Snowden, R.J.) 333–366 (1994).
9. Treisman, A.M. & Gelade, G. A feature integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980).
10. Verghese, P. & Nakayama, K. Stimulus discriminability in visual-search. *Vision Res.* **34**, 2453–2467 (1994).
11. Nakayama, K. & Silverman, G.H. Serial and parallel processing of visual feature conjunctions. *Nature* **320**, 264–265 (1986).
12. McLeod, P., Driver, J. & Crisp, J. Visual-search for a conjunction of movement and form is parallel. *Nature* **332**, 154–155 (1988).
13. Verghese, P. & Pelli, D.G. The information capacity of visual attention. *Vision Res.* **32**, 983–995 (1992).
14. Cumming, B.G. & Parker, A.J. Binocular mechanisms for detecting motion-in-depth. *Vision Res.* **34**, 483–496 (1994).
15. Regan, D. Binocular correlates of the direction of motion in depth. *Vision Res.* **33**, 2359–2360 (1993).
16. Tyler, C.W. Stereoscopic depth movement: Two eyes less sensitive than one. *Science* **174**, 958–961 (1971).
17. Westheimer, G. Detection of disparity motion by the human observer. *Optom. Vis. Sci.* **67**, 627–630 (1990).
18. Roy, J.-P. and Wurtz, R.H. The role of disparity-sensitive cortical neurons in signaling the direction of self-motion. *Nature* **348**, 160–162 (1990).
19. Harris, J.M. & Watamaniuk, S.N.J. Speed discrimination of motion-in-depth using binocular cues. *Vision Res.* **35**, 885–896 (1995).
20. Tyler, C.W. in *Binocular Vision: Vision and Visual Dysfunction* (ed. Regan, D.) 19–37 (CRC Press, Boca Raton, Florida, 1992).