PERGAMON

Vision Research 42 (2002) 2645–2650

Vision Research

www.elsevier.com/locate/visres

Rapid communication

Shifts in perceived position of flashed stimuli by illusory object motion

Katsumi Watanabe [a,*], Romi Nijhawan [a,b], Shinsuke Shimojo [a,c]

[a] Computation and Neural Systems, Division of Biology, California Institute of Technology, Mail Code 139-74, Pasadena, CA 91125, USA
[b] Department of Cognitive and Computing Sciences, University of Sussex, Falmer, Brighton, East Sussex, BN1 9QH, UK
[c] Human and Information Science Laboratory, NTT Communication Science Laboratories, Atsugi, Kanagawa 243-0198, Japan

Received 11 June 2002; received in revised form 23 July 2002

Abstract

Moving stimuli cause the position of flashed stimuli to appear shifted in the direction of motion (position capture). To examine whether position capture depends on low-level motion interactions or perception of integrated object motion, we employed a slit-view display. Two line-drawn diamonds translated horizontally in opposite directions, one above and one below the fixation cross, either behind an occluding surface with a narrow slit or without occluding surface. When the diamonds were in vertical alignment, two vertical bars were flashed, one in the center of each diamond. In the slit-view condition, the diamonds were visible through a 4-, 2-, or 1-pixel vertical slit; the width of the flashed bars always matched the width of the slit. Even though the horizontal component of physical motion was greatly reduced or absent in the slit-view conditions, observers perceived diamonds moving behind the occluding surface. Furthermore, the position of the flashed bar was captured by the moving diamonds such that each bar appeared shifted in the direction of perceived motion. We conclude that the position capture reported here has a component based on high-level motion processing that is responsible for dynamically integrating object motion and shape.

Keywords: Motion; Shape; Occlusion; Localization; Anorthoscopic perception

1. Introduction

One of the primary tasks of the visual system is to localize objects in visual space. Therefore, it is reasonable to assume that the visual system is equipped with highly efficient mechanisms of visual localization. Although visual localization can be done efficiently and usually without error, errors in visual localization do occur in various situations, particularly where object motion and eye movements are involved (see Schlag & Schlag-Rey, 2002 for review). This suggests that visual localization mechanisms are not impervious to influences from processes devoted to other perceptual and motor functions.

Previous studies have indicated that motion can significantly modulate visual localization. The relative position of a visual stimulus can be biased in the direction of motion signals contained within a stimulus (DeValois & DeValois, 1991; Ramachandran & Anstis, 1990). When a brief flash is to be localized, the position of a flashed target is 'captured' by motion signals that originate even in substantially distant regions of the visual field (position capture; Whitney & Cavanagh, 2000). Thus, transient positional signals seem to get integrated with motion signals. To what extent does position capture depend on low-level motion signals per se as opposed to perceptual (object) motion signals derived from 'higher'-level processes? [1] An answer to this question will be informative as to the level of visual processing at which motion and position signals are integrated. To address this issue, we utilized a slit-view motion display.

* Corresponding author. Present address: Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Building 49, Room 2A50, 49 Convent Drive, Bethesda, MD 20892-4435, USA. Tel.: +1-301-451-9903; fax: +1-301-402-0511.
E-mail address: kw@lsr.nei.nih.gov (K. Watanabe).

[1] In the present paper, we use 'low-level motion signal' to refer motion signal that can arise by simple pooling and interactions among outputs from nearby local motion detectors. 'High-level' motion requires additional constraints beyond local interactions (e.g., knowledge of the intrinsic characteristics of the world).

0042-6989/02/$ - see front matter
PII: S0042-6989(02)00296-1

Slit-view, or anorthoscopic perception, refers to perception of object motion when objects and their movements are visible only through a narrow slit (Parks, 1965; Rock, 1981). Hence, observers see only parts of a moving object at a given time. Many accounts of slit-view perception have been proposed (e.g., Morgan, Findlay, & Watt, 1982; Parks, 1965; Rock, Halper, Divita, & Wheeler, 1987; Shimojo & Richards, 1986). For the purpose of the present study, however, our interest was not to further the inquiry into understanding the mechanisms of slit-view perception, but to investigate the influence of 'higher'-level motion signals on localization of flashed objects. If an object viewed through a narrow slit moves perpendicularly to the slit (e.g., horizontal movement through a vertical slit), then physical (or low-level) motion signals that indicate the true direction of the object (i.e., horizontal component) are greatly reduced. Indeed, low-level motion signals orthogonal to true object motion (i.e., parallel to the slit) become dominant. Nonetheless, observers see a 'completed' object moving behind and laterally across the narrow slit.

The discrepancy between physical motion signals and object motion perception in slit-view displays provides an ideal tool to examine the contribution of low-level versus higher-level motion signals to the position capture effect and for analytical investigation of the integration process of transient and moving objects. If position capture depends exclusively on low-level motion signals, which are solely based on local motion interactions, the effect would be eliminated or reduced significantly under a slit-view condition. In contrast, if higher-level motion signals are successful in inducing position capture, then significant mislocalizations will be observed in the direction of perceived objects moving behind a slit.

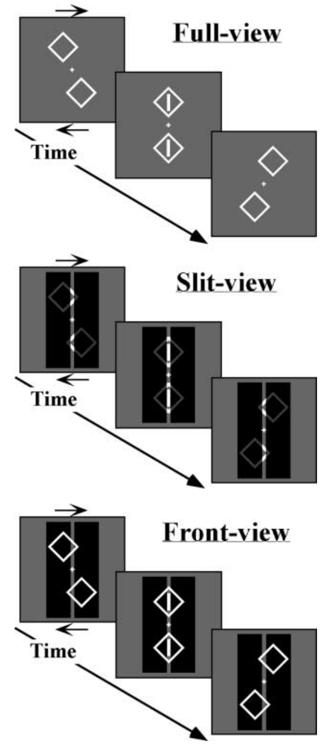

Fig. 1. Occlusion conditions: No occluding surface was presented in the full-view condition. In the slit-view condition, the occluding surface allowed the observer to see the diamonds motion only through the narrow slit. The diamonds are depicted in dim lines in the figure for viewing convenience but they actually were not visible in the experiment. The diamonds moved over the occluding surface in the front-view condition.

2. Methods

2.1. Observers

Six observers, including authors KW and SS, participated. Except for the authors, all other observers were naive. They all had normal or corrected-to-normal vision.

2.2. Stimuli

Visual stimuli were displayed on a color monitor (48° high, 60° wide, in visual angle) in a dimly lit room (about 0.5 lux). The frame rate of the monitor was 75 Hz. A white fixation cross (61.0 cd/m^2, 0.48°) was displayed against the gray (8.2 cd/m^2) background at the screen center throughout a session. Three conditions were used. (1) Full-view condition: at the beginning of a trial, two white outline diamonds (61.0 cd/m^2; 3° size, 0.17° line thickness) appeared 1.92° above and below and 4.8° to the left and right of the fixation cross (distance measure center-to-center). Immediately after their appearance, the diamonds started moving horizontally toward the opposite sides of the screen at 7.2°/s (Fig. 1). The diamonds traveled 9.6° and then disappeared. The motion direction was randomized within a session. When the diamonds were vertically aligned (on a straight line with the fixation cross), two white vertical bars (1° long) were flashed for one frame (~13.3 ms), one in the center of each diamond. [2] (2) Slit-view condition: the visual stimuli were similar to those in the full-view condition. However, two black opaque rectangles (0.02 cd/m^2, 8° × 6.5°) occluded the motion trajectory of the diamonds except for a narrow vertical slit at the screen's center. Consequently, the diamonds were visible only through the slit. (3) Front-view condition: the black rectangles were again presented, but they did not occlude the diamonds; the diamonds appeared to move in

[2] We presented the two rectangles moving in opposite directions in order to control the eye movement artifact. Also, the two bars were used to make the judgment of the perceptual (vernier) offset easier.

front of the rectangles. The thickness of the flashed bars was 4-pixels (5.76 min), 2-pixels (2.88 min), or 1-pixel (1.44 min), always matched the width of the slit. Thus, there were nine stimulus conditions; three occlusion conditions (full-view, slit-view, and front-view) $\times$ 3 slit/bar-width conditions.

2.3. Procedure

Observers viewed the stimulus display binocularly from a distance of 57 cm while fixating at the white cross. After viewing the stimulus sequence, two continuously visible bars were presented at the same locations where the flashed bars had been presented. By using a computer-mouse, observers adjusted in a randomized order the horizontal location of each bar to indicate where the bars had been flashed relative to the fixation cross (localization task). After the localization task, the word "top" appeared on the screen center. Observers pressed the appropriate key to report the perceived direction (leftward or rightward) of the top diamond's motion (motion direction judgment). Following this the word "bottom" appeared for motion direction judgment of the bottom diamond. The adjusted (perceived) locations of the flashed bars were interpreted in terms of both the direction of the perceived motion of the diamond (based on the motion direction judgment) and the direction of the diamond's physical motion (based on the visual stimulus). [3] For example, if the diamond moved to the left but the observer perceived it moving to the right, and if the bar was perceived as shifted to the right by 9 min, the position capture effect on the basis of stimulus motion is −9 min and that on the basis of perceived motion is 9 min. Within a session, the width of the bar/slit was fixed and 40 trials were repeated randomly for each occlusion condition (resulting in 120 trials per session). A mean position-capture effect was calculated for each condition and for each observer.

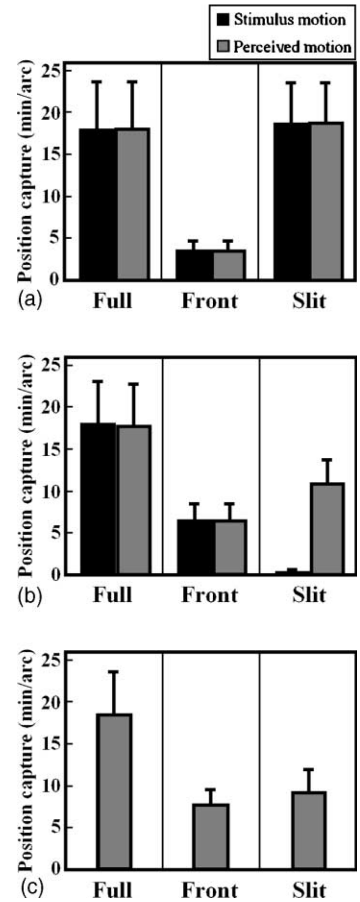

Fig. 2. (a) Position capture with 4-pixel (5.76 min) slit/bar. The position capture effects, averaged for observers, are plotted both in terms of the stimulus and perceived motion directions of the diamonds, with bars indicating 1 standard error. A larger positive value means a larger magnitude of position capture of the bar in the direction of stimulus/perceived motion of the diamond. (b) Position capture with 2-pixel (2.88 min) slit/bar. Note that the position capture effect in the slit-view condition is correlated with the perceived motion direction of the diamonds, not with the stimulus motion direction. (c) Position capture with 1-pixel (1.44 min) slit/bar. The effect is plotted only in terms of the perceived motion because there was no physical horizontal component in the slit-view condition. Yet, the sensation of the object motion of the diamonds remained, and the position capture effect significantly occurred in the same direction as the perceived direction of motion of the diamonds.

3. Results

In the full-view conditions, the observers were able to report the direction of the diamond's motion almost perfectly irrespective of the bar width. The motion direction judgment was on average correct on 98.3%, 98.8% and 99.2%, of the trials in the 4-, 2-, and 1-pixel bar conditions, respectively. Fig. 2 shows the mean position capture effect in terms of stimulus (not applicable in the 1-pixel condition) and perceived motion directions. For all the bar width conditions, the position capture was significant in the full-view condition ($t(5) > 3.12$, $p < 0.05$). The position capture was virtually identical whether it was interpreted with respect to the stimulus or perceived motion because the observers correctly perceived the stimulus motion direction. An ANOVA indicated the position capture was affected by neither the bar width nor the way position capture was calculated (stimulus or perceived motion) ($F(4, 16) = 0.97$, $p = 0.98$).

In the front-view conditions, again the observers reported the direction of the diamond's motion correctly with all the bar/slit widths tested (99.2%, 100%, and 99.6% for the 4-, 2-, 1-pixel bar/slit conditions, respectively). The position capture effect was significant in all

[3] Note the distinction between physical motion and perceived motion is not valid for the 1-pixel slit-view condition.

bar/slit conditions ($t(5) > 2.79$, $p < 0.05$). An ANOVA showed significant effect of bar/slit width ($F(4, 16) = 6.26$, $p < 0.05$); the position capture effect became slightly larger as the bar/slit width decreased. Notably, with all bar/slit widths, the position capture effects in the front-view conditions were significantly smaller than those in the full-view conditions ($t(5) > 2.75$, $p < 0.05$).

In the 4-pixel slit-view condition, the position capture effect was significant ($t(5) > 3.8$, $p < 0.05$), and the pattern of the position capture effect was almost the same as that in the full-view condition. Likewise, the motion direction judgment was near perfect (99.2%).

However, when the bar/slit width was reduced to 2-pixel, the perception of the moving diamonds became ambiguous. Although in after-session interviews the observers reported vivid horizontal motion of the diamonds, their motion direction judgments in the slit-view condition were correct only 51.7% of the trials (not different from chance; $t(5) = 0.76$, $p = 0.49$). This was expected because the horizontal motion component was less available as the slit became narrower. Moreover, the bar position was perceived as shifted in the direction of *perceived* motion of the diamond ($t(5) = 3.63$, $p < 0.01$) rather than stimulus motion ($t(5) = 0.53$, $p = 0.3$). It is worth mentioning that, in the 2-pixel slit-view condition, the observers saw the two diamonds moving in the same direction on a majority of the trials (91.7%). Yet, they did perceive the diamonds moving in opposite directions on the remaining trials (8.3%) and, on these trials, they saw the two bars shifted in opposite directions. The perceived shift of the flashed bar on those trials was not correlated with the direction of stimulus (physical) motion ($R = 0.06$, $p = 0.82$) but significantly correlated with the direction of perceived, namely opposing, motion ($R = 0.71$, $p < 0.01$). Thus, the position capture occurred in the direction of the perceived direction of the diamond's motion, regardless of whether that perceived direction was consistent or inconsistent with the stimulus motion.

In the 1-pixel slit-view condition, there was no horizontal component in stimulus motion. Nevertheless, in after-session interviews, most of the observers confirmed that they did perceive horizontal object motion, though it was much weaker. The position capture effect in the 1-pixel slit/bar condition was still significant and in the same direction as the perceived motion of the diamonds ($t(5) = 4.03$, $p < 0.01$). In 98.2% of the trials, on average, the direction of the perceived motion of the diamond and the displacement of the flashed bar were in the same direction. As with the 2-pixel slit, the observers tended to report that the two diamonds appeared to move in the same direction on a majority of the trials (93.3%), but did report the opposing object motion on the remaining trials (6.7%). On these trials, the position capture was again correlated with the perceived direction of motion of the diamonds ($R = 0.64$, $p < 0.01$).

4. Discussion

In order to examine the contribution of low-level versus higher-level motion signals to the position capture effect we employed a slit-view display in which physical motion signals are decoupled from perceived motion (Parks, 1965; Rock, 1981). Our results demonstrate that position capture does occur in the slit-view conditions on the basis of perceived motion. Intriguingly, even when the slit width was only 1-pixel, and hence physically no horizontal motion component existed, our observers perceived the object motion of the diamonds. [4] This observation suggests that the perception of 'completed' object is more than the recovery of true two-dimensional velocity, and presumably requires an interaction with the shape processing. Despite the fact that this object motion perception was totally 'illusory', the flashed bars were nonetheless perceived as shifted in the direction of 'illusory' motion. This one pixel case can hardly be explained by any account based on recovery of lateral motion behind the slit, unless one appeals to some top-down (i.e., knowledge-based) mechanism. This result suggests that the position capture reported here has a component attributable to high-level motion processing responsible for dynamically integrating object motion and shape. However, it is still contentious if various sorts of motion-induced position capture share a common mechanism (DeValois & DeValois, 1991; Nishida & Johnston, 1999; Ramachandran & Anstis, 1990; Snowden, 1998; Whitney & Cavanagh, 2000; Watanabe, Sato, & Shimojo, 2002).

Our previous study indicated that visual localization involves high-level visual processes that cannot be reduced to low-level signal interactions (Watanabe, 2002a,b; Watanabe, Nijhawan, Khurana, & Shimojo, 2001). Additionally, recent studies have shown that position capture can be caused by motion perception due to motion aftereffect, where no physical motion signal exists when a stationary stimulus appears (Nishida & Johnston, 1999; Snowden, 1998). The position capture with illusory object motion reported here is consistent with these studies in that physical motion signals in the direction that the position of flashed stimuli is shifted are not necessary for the position capture effect. Although the present study does not exclude the contribution of low-level motion signals to the position capture, high-level motion signals as revealed by the observers' perceptions appear sufficient to produce the position capture effect.

Compared with the full-view condition, the front-view condition consistently led to a reduced position-capture effect. This may be because the occluding

[4] Nakayama and Shimojo had observed a similar effect some years ago (unpublished observation).

surface added a reference frame for visual localization, similar to the reduction of saccadic mislocalization with a structured background (e.g., Honda, 1999). However, it should be noted that the occluding surface did not diminish the position capture effect when the visual stimulus was interpreted as diamonds moving behind the surface (i.e., in the slit-view conditions) particularly when the slit was wide (i.e., with the 4-pixel slit in the present study). This observation suggests that the horizontal stimulus motion signals, which are strong in the front-view condition, are by themselves not sufficient to fully account the position capture effect, and further supports the involvement of high-level motion processes. If position capture depended *exclusively* on low-level motion signals, it should have been larger in the front-view condition than the slit-view condition.

The illusory mislocalization phenomenon that we report here suggests that visual localization, a task that appears relatively simple at the outset, requires elaborate computations beyond positional information on the retina (Schlag & Schlag-Rey, 2002). Mislocalization effects may reflect the underlying neural processes that typically yield an accurate perception of position (Gregory, 1997; Nijhawan, 1994). In the slit-view condition, the visual stimuli permit two interpretations: (1) a diamond with a vertical bar at the center moves behind the occluding surface, (2) a diamond moves behind the surface and a vertical bar happens to flash at the slit independently. The latter is more accidental because it requires the flashed bar to exactly fit inside the slit despite the fact that there are many other possible locations where the flash could occur. Furthermore, it requires the flashed bar to be presented at the moment the center of the diamond arrives at the slit despite the fact that there are many other possible times at which the flash could occur. The first alternative is much simpler: the bar and the diamond are part of the same stimulus, and the bar appears through the slit when the center of the diamond is coincident with the slit. On this view, although the slit in the slit-view condition does provide a visual reference frame, it also allows an interpretation that the moving diamond and the bar form a single unit. In contrast, in the front-view condition, the visual reference frame does add further information regarding the position of the flash relative to the slit (i.e., it is on the slit). However, if the vertical bar is a part of the diamond and flashes while the diamond in motion, the probability that the flash location coincides with the slit location is very small. Thus, the visual system would take a more generic interpretation; the flash is an independent object that occurs at the slit.

We suggest that the position of the flashed bar is captured by the moving diamond when the flashed bar is integrated into a unitary percept with the moving object. Therefore, the positional shifts of the flashed bar observed here may be related not only to other position capture effects, but may also shed light on motion-shape integration mechanisms that commonly underlie anorthoscopic perception. In particular, our present findings, together with previous ones (Watanabe, Nijhawan, et al., 2001; Watanabe et al., 2001), indicate that position perception of a flashed object may be determined at or later than the level of perceptual grouping.

Finally, it is interesting to note that the bar was perceived as shifted by about 9–17 min on average in the slit-view conditions while the slit width was at most 5.76 min (= 4-pixels). This implies that the observers might perceive the bar as *on*, or *in front of*, the occluding surface. However, when interviewed after the experiment, our observers reported that they *perceived the bars as being flashed behind the surface*. This is analogous to the phenomenology of the integrated shape perception itself (an object such as the diamond appears *behind* the surface), again consistent with our interpretation that the flashed bar is interpreted as a part of the diamond. The nature of this 'anomalous' perception (i.e., perceiving a flash occurring behind an occluding surface) requires further investigation.

Acknowledgements

This work was partly supported by JSPS and by NIH. The part of the present study has been reported in abstract form (Watanabe, Nijhawan & Shimojo, 2001).

References

DeValois, R. L., & DeValois, K. K. (1991). Vernier acuity with stationary moving gabors. *Vision Research, 31*, 1619–1626.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 352*, 1121–1127.

Honda, H. (1999). Modification of saccade-contingent visual mislocalization by the presence of a visual frame of reference. *Vision Research, 39*(1), 51–57.

Morgan, M. J., Findlay, J. M., & Watt, R. J. (1982). Aperture viewing: a review and a synthesis. *Quarterly Journal of Experimental Psychology Section A Human Experimental Psychology, 34*, 211–233.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature, 370*, 256–257.

Nishida, S., & Johnston, A. (1999). Influence of motion signals on the perceived position of spatial pattern. *Nature, 397*, 610–612.

Parks, T. E. (1965). Post-retinal visual storage. *American Journal of Psychology, 78*, 145–147.

Ramachandran, V. S., & Anstis, S. M. (1990). Illusory displacement of equiluminous kinetic edges. *Perception, 19*, 611–616.

Rock, I. (1981). Anorthoscopic perception. *Scientific American, 244*(3), 145–153.

Rock, I., Halper, F., Divita, J., & Wheeler, D. (1987). Eye-movement as a cue to figure motion in anorthoscopic perception. *Journal of Experimental Psychology: Human, Perception and Performance, 13*(3), 344–352.

Schlag, J., & Schlag-Rey, M. (2002). Through the eye, slowly: delays and localization errors in the visual system. *Nature Reviews Neuroscience, 3*, 191–200.

Shimojo, S., & Richards, W. (1986). Seeing shapes that are almost totally occluded: a new look at parks camel. *Perception and Psychophysics, 39*(6), 418–426.

Snowden, R. J. (1998). Shifts in perceived position following adaptation to visual motion. *Current Biology, 8*, 1343–1345.

Watanabe, K. (2002a). Time course of perceptual grouping revealed by the asymmetric mislocalization effect. Paper presented at Visual Localization in Space–Time, Sussex, England, August 2002.

Watanabe, K. (2002b). Perceptual grouping by motion precedes relative localization of visual stimuli. *Perception*, *31* (Suppl.), 109 (Abstract).

Watanabe, K., Nijhawan, R., Khurana, B., & Shimojo, S. (2001). Perceptual organization of moving stimuli modulates the relative position of a visual flash. *Journal of Experimental Psychology: Human Perception and Performance, 27*(4), 879–894.

Watanabe, K., Nijhawan, R., & Shimojo, S. (2001). Position capture by object motion through a slit. *Journal of Vision, 1*(3) (15a (abstract)).

Watanabe, K., Sato, T., & Shimojo, S. (2002). Perceived shifts of flashed stimuli by visible and invisible object motion. Paper presented at the 2nd Asian Conference on Vision, Gyeongju, South Korea, July 2002.

Whitney, D., & Cavanagh, P. (2000). Motion distorts visual space: shifting the perceived positions of remote stationary objects. *Nature Neuroscience, 3*, 954–959.