

# 1 Introduction

Accurate motion perception is essential for visually guided movement; complex behaviors such as chasing prey or catching a thrown ball require that an organism be able to rapidly determine the position and velocity of a moving object, and to anticipate its trajectory through space. Psychological research on motion perception has established that local and global motion are processed by separate mechanisms, an idea that dates back to Wertheimer's phenomenological distinction between fast "phi" and slower "beta" motion (Steinman et al., 2000). It is believed that local (or short range) mechanisms respond to the motion of visual features defined by luminance, or by motion energy in the Fourier domain (Adelson and Bergen, 1985), at short temporal and spatial scales. Higher-order motion does not necessarily require features to differ in mean luminance from the background, does not necessarily contain motion energy in the Fourier domain, and appears to operate over longer spatial and temporal scales than first-order motion. Strong global motion percepts can be elicited by stimuli that modulate in contrast or texture rather than in luminance. In general, higher-order motion seems to involve the change in position, over time, of some feature in the image (Lu and Sperling, 1995).

It is thought that the first-order motion mechanism originates with the direction selective response of cells in V1. Classic models of local motion include local filters that extract directional signals from space-time correlations in luminance contrast in a small region of space, in a way similar to what has been observed to drive the activity of cells in V1. Psychophysical data are consistent with the notion that the size of the local motion filters are similar in size of the classical receptive field of a V1 neuron. Long-range motion on the other hand seems to be capable of tracking a feature object over distances larger than what can be achieved through individual local filters. A strong sensation of motion can be created without any Fourier motion signal, so it is possible that local and global mechanisms have independent origins. The anatomical substrate of higher-order motion mechanism is unknown. While area MT and MST have been proposed as candidates, response properties of cells in those regions do not appear to explain our sensations of higher order motion (Ilg and Churan, 2004).

One thing that makes accurate estimation of motion a difficult problem is that the local motion signals are not always a reliable indication of the global motion. In a complicated visual world, motion can come from many sources, and accurate perception of the movement of objects requires disambiguating those motion signals attributable to the object from irrelevant motions in the background or of other objects. Consider the task of trying to track the movement of a zebra among a background of waving grass. While many of the local signals will correctly signal the motion of the zebra, the motion of the grass creates a subset of local motion signals that are substantially incorrect. Thus, a simple average of the motion over any receptive field large enough to cover more than a tiny portion of the scene will not accomplish the goal of tracking the motion of the zebra. On the other hand, an average of motion signals within

cites:(Reichardt;  
Adelson; Si-  
moncelli;  
Ullman \&  
Marr; others)  
cites

can provide  
more cites on  
this point.

very small regions will also be incorrect; small regions containing only one edge cannot uniquely determine the motion of an object, due to the aperture problem. To track the butterfly, then, requires integrating local motion signals over space and time in a way that is consistent with the global change in position of an object, while discarding local signals that are inconsistent with the object's trajectory.

cite re aperture?

In this paper we examine how first-order and higher order mechanisms interact in forming an overall perception of motion. We constructed stimuli that combined a first-order motion component with an independent higher-order (position-tracking) component.

Figure 1 demonstrates how first order and higher order motion can interact. The movie shows two wheels, each composed of five moving elements. The elements are Gabor-like stimuli that window a moving carrier inside of a spatial envelope. The envelope moves independently of the carrier, so that the carrier produces first order motion while the envelope produces higher order motion. On the right side of the display, the local and global motion components move in the same direction; on the left side the local and global components are in opposite directions. Full details of the construction of this display are given in section ??.

When the elements are viewed in isolation, as in Figure 1, the elements are seen to move consistent with the motion of the envelope. The carrier motion causes a change in the perception of "smoothness" but does not strongly affect the appearance of motion. However, when multiple elements are placed in proximity, but not overlapping, as in Figure 1, the appearance depends on whether the stimulus is viewed centrally or peripherally. When the five-element stimulus on the left is viewed centrally, the envelope motion is seen, but when it is viewed from the periphery, the carrier motion is seen. If the observer keeps attention on that stimulus while moving the eyes to place it in central and peripheral vision, its apparent motion appears to reverse.

From this demonstration it appears that spatial proximity of targets causes a change in how first-order and higher-order motion is processed. That the appearance changes with retinal eccentricity of the stimulus suggests that the range of spatial interaction scales with the retinal eccentricity. A plausible explanation could be that the presence of flanking objects limits the ability to see movement of the envelope, allowing the carrier motion to determine the percept; that is, crowding may be interfering with the processing of higher-order motion.

### 1.0.1 Crowding, a mini-review

Crowding is a phenomenon wherein identification or discrimination of an object presented in the visual periphery is impaired by the presence of nearby, but non-overlapping flanking objects. A finding characteristic of crowding is that critical spacing (usually a measure of the distance between target and flanker which achieves a particular elevation of threshold for recognition) scales linearly with retinal eccentricity (Bouma, 1970; Toet and Levi, 1992). Although most

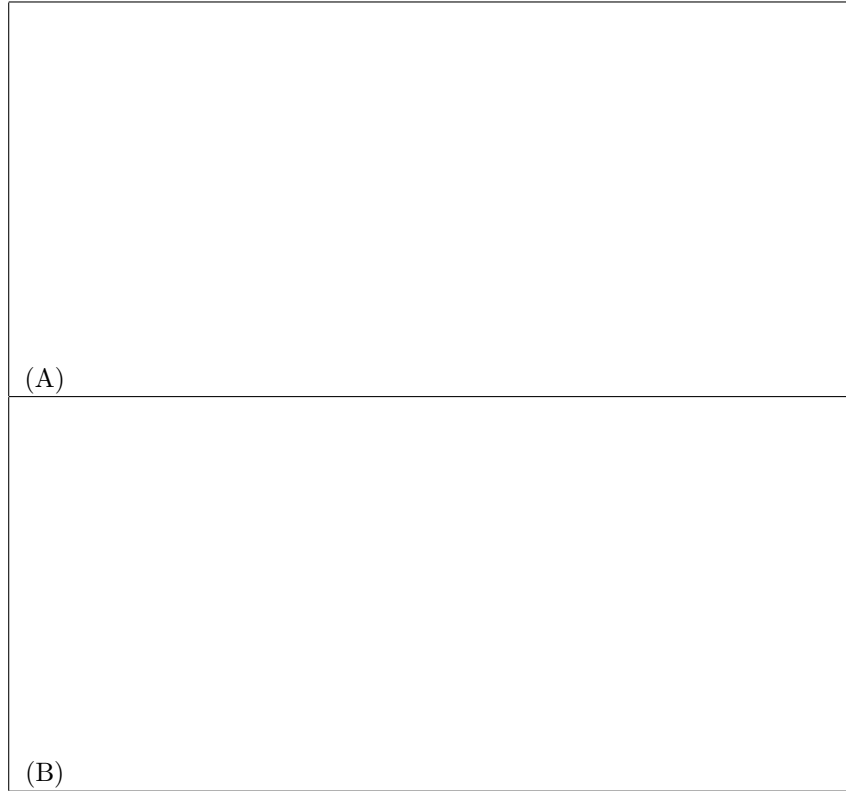


Figure 1: (A) Single motion elements, moving alone. At left, the carrier motion is opposite the envelope motion. At right, the carrier motion is in the same direction as the envelope motion. (B) Five elements, each identical to the single element in A, distributed around each fixation point. When elements are placed in close proximity, there is an eccentricity-dependent change in appearance. The appearance of the wheel on the left (with opposing carrier and envelope motions) will change depending on where the eyes fixate.

studies of crowding focus on its effect of impairing the recognition of shapes (e.g. letters) in parafoveal vision, it has become apparent that crowding is a more general phenomenon, extending to many different types of visual features (e.g. (van den Berg et al., 2007; for review, see Levi, 2008.)) It is thought that crowding is characteristic of some cortical mechanism that integrates signals from low-level feature detectors, a so-called “integration field” (Pelli et al., 2004). Because the scaling of critical distance with spacing mirrors the variation of cortical magnification with eccentricity, the integration field is thought to be a process that subsumes a constant distance on the cortical surface (Pelli, 2008).

The next couple of paragraphs refer to a construction of my experiment I’m not doing any more...

(Pelli et al., 2004) proposed that the crucial diagnostic test for crowding as opposed to masking or other forms of spatial interference is that the critical spacing scales with eccentricity and is relatively unaffected by signal size. Accordingly, we set out to determine which target spacing and motion parameters are necessary to drive the reversal of apparent motion as various eccentricities. In \autoref{sec:constant} below, we determine the relationship between critical spacing and target spacing, which satisfies Bouma’s law. In \autoref{sec:occlusion} we show that the critical spacing is unaffected by the presence of an occluder which covers 2/3 of the visible circle, meaning that it is the spacing which is relevant and not the number of visible targets. In \autoref{sec:grid} we show that the critical distance and scaling property is robust to the size of the stimuli. We also test its robustness to variations in temporal frequency, step size, and step interval.

While most studies of crowding involve stationary stimuli, motion stimuli add a temporal component. In Experiments 1 through 3 we consistently find that for stimuli near the crowding distance, the trials for which the subject took longer in responding were more likely to correctly reflect the global direction of motion. In Experiment 4 we use an auditory cue to vary the subjects’ response time to investigate this effect in more detail.

Our results reinforce the idea that global motion processing is the result of an integration of the output of low-level feature detectors, and that in fact the process subserving detection of global motion might be identical to the processes underlying object recognition and target selection. We discuss the implications for possible mechanisms of higher order motion perception and speculate on their possible physiological implementations.

## References

- Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A*, 2(2):284–299.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(5241):177–8.

- Ilg, U. J. and Churan, J. (2004). Motion perception without explicit activity in areas mt and mst. *Journal of Neurophysiology*, 92(3):1512–1523.
- Levi, D. M. (2008). Crowding—an essential bottleneck for object recognition: a mini-review. *Vision Res*, 48(5):635–654.
- Lu, Z. L. and Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Res*, 35(19):2697–2722.
- Pelli, D. G. (2008). Crowding: a cortical constraint on object recognition. *Curr Opin Neurobiol*, 18(4):445–451.
- Pelli, D. G., Palomares, M., and Majaaj, N. J. (2004). Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J Vis*, 4(12):1136–69.
- Steinman, R. M., Pizlo, Z., and Pizlo, F. J. (2000). Phi is not beta, and why Wertheimer’s discovery launched the gestalt revolution. *Vision Research*, 40(17):2257–2264.
- Toet, A. and Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Res*, 32(7):1349–1357.
- van den Berg, R., Roerdink, J. B. T. M., and Cornelissen, F. W. (2007). On the generality of crowding: visual crowding in size, saturation, and hue compared to orientation. *J Vis*, 7(2):14.1–11.