

Perceptual interaction of local motion signals

Eyal I. Nitzany

Program in Computational Biology & Medicine,
Cornell University, Ithaca, NY, USA
Feil Family Brain and Mind Research Institute,
Weill Cornell Medical College, New York City, NY, USA
Department of Organismal Biology and Anatomy,
University of Chicago, Chicago, IL, USA



Maren E. Loe

Department of Organismal Biology and Anatomy,
University of Chicago, Chicago, IL, USA



Stephanie E. Palmer

Department of Organismal Biology and Anatomy and
Committee on Computational Neuroscience,
University of Chicago, Chicago, IL, USA



Jonathan D. Victor

Feil Family Brain and Mind Research Institute,
Weill Cornell Medical College, New York City, NY, USA



Motion signals are a rich source of information used in many everyday tasks, such as segregation of objects from background and navigation. Motion analysis by biological systems is generally considered to consist of two stages: extraction of local motion signals followed by spatial integration. Studies using synthetic stimuli show that there are many kinds and subtypes of local motion signals. When presented in isolation, these stimuli elicit behavioral and neurophysiological responses in a wide range of species, from insects to mammals. However, these mathematically-distinct varieties of local motion signals typically co-exist in natural scenes. This study focuses on interactions between two kinds of local motion signals: Fourier and glider. Fourier signals are typically associated with translation, while glider signals occur when an object approaches or recedes. Here, using a novel class of synthetic stimuli, we ask how distinct kinds of local motion signals interact and whether context influences sensitivity to Fourier motion. We report that local motion signals of different types interact at the perceptual level, and that this interaction can include subthreshold summation and, in some subjects, subtle context-dependent changes in sensitivity. We discuss the implications of these observations, and the factors that may underlie them.

Introduction

Motion is crucial for everyday tasks, such as navigation (Ullman, 1979) and figure/ground segregation (Grossberg, 1994). Motion analysis is generally considered to begin with the extraction of local motion signals, and subsequently, motion signals are integrated across space. Many kinds of local motion signals are recognized by human subjects, including Fourier (F), non-Fourier (NF), and glider (G), described in detail in Methods. These can be distinguished by the nature of the computations required to extract them. It is straightforward to construct stimuli that isolate each of these kinds of signals, enabling experimental analysis of visual responses to each. Yet natural scenes contain all of these signals (Nitzany & Victor, 2014), and they typically co-occur in the same location. Moreover, theoretical studies indicate that motion extraction can be made more efficient through the combined use of multiple local motion cues (Fitzgerald, Katsov, Clanin, & Schnitzer, 2011). In addition, since the relative contributions of these signals depend on the source of the motion (e.g., translation leads to primarily F motion, while looming of objects generates prominent G expansion signals and receding of objects generates prominent G contraction signals), biological motion processing may be context-dependent. Thus, it is of

Citation: Nitzany, E. I., Loe, M. E., Palmer, S. E., & Victor, J. D. (2016). Perceptual interaction of local motion signals. *Journal of Vision*, 16(14):22, 1–12, doi:10.1167/16.14.22.

doi: 10.1167/16.14.22

accepted September 24, 2016; published November 30, 2016

ISSN 1534-7362



interest to determine how local motion signals are jointly processed, and whether they are processed in a context-dependent fashion.

To address these questions, we asked subjects to report the perceived direction of synthetic motion signals that contained controlled mixtures of consistent F and G signals. Experiments were carried out in blocks—one consisting of mixtures of F with G contraction, one consisting of mixtures of F with G expansion. This enabled us to examine integration of motion signals that were simultaneously present, as well as context-dependence of processing of F motion.

To provide a conceptual framework for the different kinds of local motion signals used in these experiments, we first note that the archetypal local motion signal consists of spatiotemporal correlation between pairs of points (Reichardt, 1961); i.e., two points along a diagonal line in a space-time plot of the visual stimulus. This is often also called *Fourier motion*, as two-point correlations can be determined from the power spectrum. However, motion can also be signaled by multipoint correlations in any slanted region in space–time. These higher order motion signals include *glider motion* (Hu & Victor, 2010; Nitzany & Victor, 2014), which involves three points in a space–time triangle, and classic *non-Fourier motion*, which involves four points in a space–time parallelogram (Chubb & Sperling, 1988). The latter has been called *second-order motion*, a term that unfortunately obscures the fact that motion information is carried by four-point correlations. These different motion signals and stimuli are further discussed in Nitzany and Victor (2014; see Victor, Thengone, & Conte, 2013, for further discussion of the “order” terminology).

While several studies have examined the relationship between different kinds of local motion signals, they have focused on F and NF signals, and primarily on whether these two kinds of motion were processed by one or two systems (Lu & Sperling, 2001). Here, we are specifically interested in a different question: how distinct kinds of local motion signals (here, F and G) interact. Note that our focus is also distinct from questions about how local motion signals in different directions combine (i.e., plaids and one-dimensional vs. two-dimensional [2-D] motion; Movshon, Adelson, Gizzi, & Newsome, 1985; Wilson, Ferrera, & Yo, 1992); here we are concerned with how the visual system combines two kinds of signals in the same direction, rather than how it resolves a possible conflict of local motion cues in different directions.

A wide range of techniques are available to study processing of different kinds of cues, including neurophysiological experiments (Albright, 1992; Chaudhuri & Albright, 1997; O’Keefe & Movshon, 1998), studies of patients with neurologic disorders (Greenlee & Smith, 1997; Piponnier et al., 2015; Plant, Laxer,

Barbaro, Schiffman, & Nakayama, 1993; Plant & Nakayama, 1993; Vaina & Soloviev, 2004), imaging studies (Smith, Greenlee, Singh, Kraemer, & Hennig, 1998), and psychophysical experiments (Edwards & Badcock, 1995; Ledgeway, 1994; Piponnier et al., 2015; Victor & Conte, 1992; Yo & Wilson, 1992), with the latter approach either based on comparing the tuning of responses to different motion types (Piponnier et al., 2015; Yo & Wilson, 1992), or using adaptation paradigms (Edwards & Badcock, 1995; Ledgeway, 1994). In this study, we use a psychophysical strategy that explicitly combines F motion signals with several variants of G motion in a controlled fashion. This enables us to look directly at how F and G signals are integrated, and whether the responses to F signals depend on context.

We show that a combination of F and G signals can reach threshold even if both are below threshold, but this only occurs with G contraction. For F with G expansion, such integration was not found, and moreover, in some subjects, sensitivity to F motion signals was reduced in comparison to sensitivity to F motion signals in the context of G contraction.

Methods

Experiments were organized into trials in which one or more cues (see below) indicated motion to the left or the right, with equal probability. When two motion cues were present, they always carried motion in the same direction.

Subjects performed a two-alternative forced-choice task, indicating their judgment of the direction of motion of a 1-s movie segment via a button-press (they were not asked to indicate or attend to the kind of motion signal). Movie segments consisted of 10 frames, 100 ms each, each composed of a 16×16 array of black and white checks, with a fixation point superimposed on the center of the display prior to the onset of the movie. Experiments took place in two laboratories, with different stimulus parameters in each location. Overall stimulus sizes ranged from $7.5^\circ \times 7.5^\circ$ (28-min checks) to $14.5^\circ \times 14.5^\circ$ (54-min checks) and viewing distances ranged from 30 to 85 cm (see “Subjects and display” section below). Trials were self-paced, and no feedback was given. Data were collected after a small number of practice trials to ensure that the subject understood the task.

Visual stimuli

Movies were categorized by the strength of the F motion signal (C_F) and the strength of the G signal

Kind	Subtype	Template	Subtype	# black	# white
2-point (Fourier)			Standard Reverse phi	0 or 2 1	0 or 2 1
	3-point (Glider)	Contraction		White exp. Black exp.	0 or 2 1 or 3
Expansion			White cont. Black cont.	0 or 2 1 or 3	1 or 3 0 or 2

Table 1. Examined motion types. *Notes:* Fourier (F) motion is characterized by two-point correlations along a spatiotemporal diagonal, and glider (G) motion is characterized by three-point motion in a spatiotemporal triangle. G motion has contraction and expansion subtypes, depending on the orientation of the triangle in space–time. Motion signals are further subdivided by the parity of the number of checks of each color.

(C_G). Either signal could range from 0 (absent) to 1 (maximal). As noted above, when both signals were present, the motions they defined were in the same direction (randomly left or right).

As mentioned above, both kinds of local motion are defined by correlations in slanted spatiotemporal regions, which we designate as their “templates” (Nitzany & Victor, 2014). For F motion, the template consists of two checks in a space–time diagonal; for G motion, the template consists of three checks in a space–time triangle (see Table 1). Since the templates contain either black or white checks, their correlations are defined by parity rules—whether the number of black (or white) checks is an even or an odd number. Thus, for F motion, full strength ($C_F = 1$) corresponds to all templates containing matching checks (either two white checks or two black checks; see Table 1 and Figure 1A); at zero strength ($C_F = 0$), half of the templates contain matching checks and half contain one check of each color. For G motion, full strength ($C_G = 1$) stimuli can be of two polarities—black triangles, in which all triangles have either three black checks or one black check, and white triangles, in which all triangles have either three white checks or one white check (see Table 1 and Figure 1A). Additionally, the G template can have two different orientations in space–time, expanding or contracting. We emphasize that these terms are simply descriptors of 2-D movies, not three-dimensional (3-D) shape changes. (For further information about G motion, we refer the interested reader to Hu & Victor, 2010, and Nitzany & Victor, 2014).

To make a stimulus that combines an F motion signal of strength C_F with a G motion signal of strength C_G , we adapted the maximum-entropy texture generation algorithms of Victor and Conte (2012) to spatiotemporal stimuli (see below). The maximum-entropy property means that the movies are as random as possible, given the two specified component motion signal strengths. That is, they contain no other cues to

motion other than what is implied by the component motion signals.

In detail, stimulus generation is as follows. We began with the algorithms of Victor and Conte (2012), which create 2-D visual textures. We then used these 2-D spatial (XY) arrays as XT slices of a movie (where X is the horizontal axis). With this reassignment, the spatial second-order image statistics β_l and β_r of the Victor and Conte (2012) algorithm capture the pairwise correlations along diagonals in space–time. Thus, to generate an F motion signal to the left or right, we set one of these β s equal to C_F . Similarly, the spatial third-order image statistics θ_l , θ_r , θ_{\downarrow} , and θ_{\uparrow} of Victor and Conte (2012) capture three-point correlations within a triangular G template. Thus, to generate a G motion signal, we set one of the four θ s equal to $\pm C_G$. $\theta = -C_G$ generates a preponderance of templates with an odd number of black checks (black Gs), while $\theta = +C_G$ generates a preponderance of templates with an odd number of white checks (white Gs). The selection among the four θ s is determined by whether the G subtype is expansion or contraction, and whether the motion signal is to the left or the right. Specifically, θ_l and θ_r correspond to G contraction, while θ_{\downarrow} and θ_{\uparrow} correspond to G expansion; θ_l and θ_{\downarrow} are leftward motions, while θ_r and θ_{\uparrow} are rightward motions.

Note that because the θ textures have no second-order correlations, the G motion stimuli derived from them are “microbalanced” in the sense of Chubb and Sperling (1988). They have no pairwise correlations, either in space or in time. The appearance or disappearance of an element on one frame is uncorrelated with any of the single-pixel values on subsequent frames.

The algorithms of Victor and Conte (2012) include procedures for making maximum-entropy textures for any pairwise combination of values of the β s and θ s, provided that the total strength does not exceed 1. These algorithms translate directly into making XT slices of maximum-entropy movies containing mixtures of F and G motion signals. We used this approach to generate 16 independent XT slices for each movie, and then stacked them along the vertical (Y) dimension to make a single XYT movie.

In this manner, F and G motion signal strengths could be varied independently, and thus constituted coordinates in a planar combination space of motion signals. We carried out experiments in two such planes: F combined with G contraction, and F combined with G expansion. Since pure F stimuli were common to both experiments, this design enabled us to study two kinds of interactions: direct interactions between F and G motion signals that were simultaneously presented, and modulatory interactions (i.e., changes in sensitivity to F motion signals that depended on whether they were in the context of G contraction vs. G expansion).

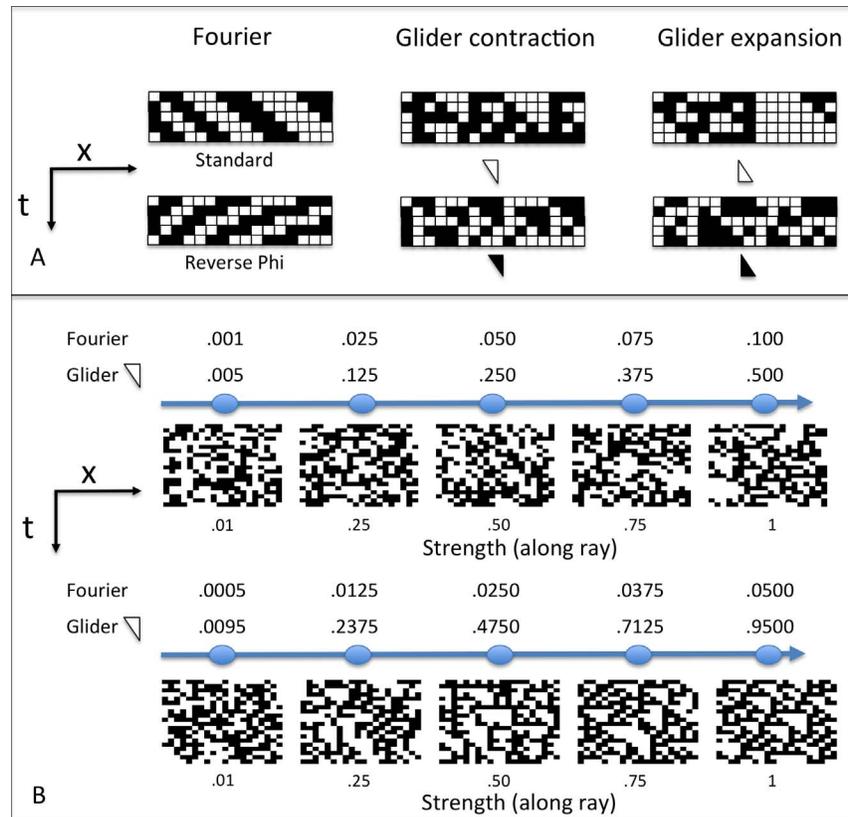


Figure 1. Stimulus examples. Panel A: Space–time slices of stimuli containing F (left column), G contraction (middle column), and G expansion (right column) signals at maximum correlation strength. Top row: Even parity, corresponding to standard F motion, white G contraction, and white G expansion (see Table 1). Bottom row: Odd parity, corresponding to reverse-phi F motion (not used here), black G contraction, and black G expansion. Panel B: Space–time slices of stimuli containing mixtures of F and white G contraction, in the proportions used in these experiments. Top row: Five example stimuli along a ray ending with $C_F = 0.1$ and $C_G = 0.5$. Bottom row: Five example stimuli along a ray ending with $C_F = 0.05$ and $C_G = 0.95$.

The first two experiments were organized in these two combination spaces (F combined with G contraction, F combined with G expansion), as follows. In each block, test points were located along seven rays emanating from the origin of the combination space. Along each ray, five points were examined. Specifically, each ray was defined by a maximum motion strength of each signal type ($C_{F,\max}$, $C_{G,\max}$), and the stimuli along each ray were defined by $(C_F, C_G) = R_i \times (C_{F,\max}, C_{G,\max})$, where $R_i = 0.01, 0.25, 0.5, 0.75$, and 1. Trials with these 35 different combinations of cues and cue strengths were randomly interleaved in each block. Figure 1B shows examples of combined stimuli along two rays.

A third experiment was designed to assay sensitivity to F motion signals across three contexts: contexts dominated by G contraction, by G expansion, and by random movies. Here, each block consisted of 400 trials, 80 of which (20%) contained F motion, with $C_F = 0.1$. The remaining 320 trials (80%) consisted either of G contraction ($C_G = 0.95$), G expansion ($C_G = 0.95$), or random movies (no motion signal), depending on the block. When G contraction or

expansion was used for context, randomly half of the trials consisted of white versus black Gs. These three block types were presented with 5–10-min breaks within a session, and sessions containing each of the six possible permutations of block orders were run on the four subjects who participated in the first two experiments. Subjects were blinded to the identity of each block.

The above stimuli were generated via software written in Matlab (Natick, MA; version 2010a), which also recorded the subjects' responses.

The supplementary information contains examples of movie clips. Other examples may be found in the supplements of Hu and Victor (2010) and Nitzany and Victor (2014).

Data analysis

The two kinds of motion signal combinations studied (F combined with G contraction, and F combined with G expansion) were analyzed separately

based on data in the first two experiments. In each combination plane, the first step was to fit the measured fraction correct with a Weibull function along each ray, as in Victor, Chubb, and Conte (2005), using the maximum likelihood approach:

$$FC(x) = \frac{1}{2} + \frac{1}{2} \left(1 - 2^{-\left(\frac{x}{a_r}\right)^{b_r}} \right) \quad (1)$$

In this equation, x is the distance along the ray from the origin, given by $x = \sqrt{C_F^2 + C_G^2}$, where C_F and C_G are the individual motion strengths; b_r is the Weibull shape parameter; and a_r is the motion strength at which the fraction correct is 0.75. The value of the Weibull shape parameter (b_r) was found to be in the range of 0.85 to 2.96, across all rays, interaction planes, and subjects. We then fit each dataset (all the rays within a single interaction plane for a single subject) with a uniform shape parameter value b , allowing the threshold parameter a_r to vary across rays. This yielded consensus values of b in the range 1.5 to 2.3 (across subjects and G contraction vs. expansion). The fitted value of a_r for each ray was then taken to be the threshold along that ray. Ninety-five percent confidence intervals (CIs) were determined from the empiric distribution of 1,000 bootstrapped samples.

In the third experiment, data analysis consisted of tallying the fraction correct for the Fourier motion trials and each of the context trials (G contraction, G expansion, and random), followed by standard statistics.

Subjects and display

Subjects (four male, three female) had either normal or corrected-to-normal vision, and ranged in age from 20 to 38. Subjects EIN, ML, and SEP were authors; subject PSS was not an author but was aware of the purpose of the experiments. For the first two experiments (combinations of F and G motion signals), displays were as follows: EIN, AB, and TS: 17-in. Retina display on a Macbook Pro, 60-Hz frame rate and 2880×1800 pixel resolution, mean luminance 33 cd/m^2 , viewing distance 57 cm, stimulus size $8.15^\circ \times 8.15^\circ$; PSS: 15-in. Retina display on a Macbook Pro, 60-Hz frame rate and 2880×1800 pixel resolution, mean luminance 33 cd/m^2 , viewing distance 57 cm, stimulus size $7.5^\circ \times 7.5^\circ$; ML, SEP, and S4: 22-in. CRT monitor, 100-Hz frame rate and 1024×768 pixel resolution, viewing distance 85 cm, stimulus size $14.5^\circ \times 14.5^\circ$, mean luminance 75 cd/m^2 . For the third experiment (F motion in three contexts), all subjects viewed the latter 22-in. monitor.

For Experiments 1 and 2, subjects EIN, TS, ML, SEP, and S4 participated in four 1-hr sessions, and PSS

and AB participated in two 1-hr sessions; sessions contained approximately 3,500 trials. For the third experiment, subjects ML, SEP, and S4 participated in six sessions and EIN participated in 12 such sessions; sessions contained 1,200 trials. The datasets EIN1 and EIN2 are simply repeats of the entire protocol for subject EIN for each experiment; they were paired in Experiments 1 and 2, but there is no correspondence between these pairs of datasets and those of Experiment 3.

Human subject procedures were approved by the Institutional Review Committee of Weill Cornell Medical College and by the Institutional Review Committee of University of Chicago.

Results

Previous theoretical work (Fitzgerald et al., 2011) points out advantages in integrating several local motion signals, and in particular emphasizes the role of F and G motion signals in integrated motion processing. To determine how F and G motion cues are combined perceptually, we measured the ability of a subject to determine the direction of motion in a stimulus containing both kinds of motion cues in a range of proportions. Figure 2 shows how performance depended on F and G strength for G contraction (first and third rows), and G expansion (second and fourth rows). The most obvious feature is that when motion cues were presented in isolation, there was a much greater sensitivity to F cues than to G cues: When only the F cue was present (along the abscissa), a strength $C_F = 0.1$ was typically sufficient to produce a fraction of correct trials of 0.75, but when only the G cue was present (along the ordinate), a strength of $C_G = 0.5$ or greater was required to achieve the same level of performance. As shown in Hu and Victor (2010), three-point motion cues in other template configurations are also much weaker than F cues.

For combinations of F and G contraction (first and third rows), the consistently curved contour lines showed that F and G contraction signals are integrated—a given level of performance can be achieved by a combination of signals, even if neither signal by itself would have supported that performance level. If, alternatively, there were no integration of the motion signals, the contours would have been rectilinear because a criterion performance level would only be reached in the combination stimulus when one of its components meets the given performance threshold. The elliptical shape of these contours suggests that the signals are combined in an approximately quadratic fashion, with F signals given a stronger weight.

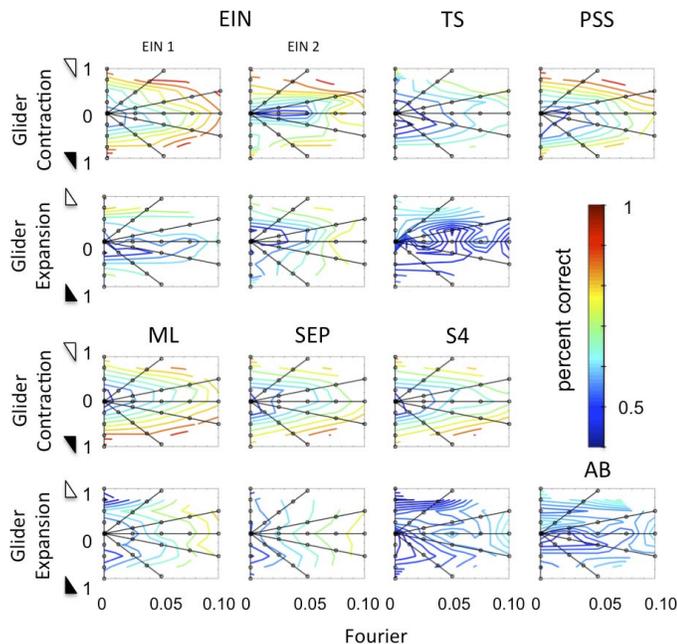


Figure 2. Psychophysical performance for direction judgments for stimuli containing mixtures of F and G signals. Contour maps show fraction correct as a function of their signal strengths, C_F (abscissa) and C_G (ordinate). Upper quadrant shows responses for stimuli containing white Gs; lower quadrant shows responses for stimuli containing black Gs. The abscissa corresponds to pure F stimuli. First and third rows: Mixtures of F and G contraction. Second and fourth rows: Mixtures of F and G expansion. The lines indicate the rays that were studied, and the points on the rays the specific signal combinations. When two motion cues were present, they were always in consistent directions.

For combinations of F signals with G expansion, the pattern of interaction is different. For subjects ML, TS, SEP, and S4—who had elliptical contours for G contraction—contour lines are approximately horizontal or vertical—that is, showing that performance depended primarily on one component of motion (G expansion or F). For subject EIN (both experiments), the corners of the contours were somewhat rounded but not nearly as much as for G contraction.

For subjects TS and AB, low levels of black G expansion led to performance that was below chance (i.e., they perceived the stimulus as moving in the direction consistently opposite to its true motion—a finding also reported in Hu & Victor, 2010, for a pure G expansion stimulus).

It is not the case that G expansion is simply a weak motion signal that is antagonized by a stronger F motion signal. If subjects were responding to the G motion direction correctly and this signal is antagonized by adding F motion, then moving parallel to the abscissa in Figure 2 would result in a decrease in fraction correct, and this is not what the data show:

There are increases in four of the datasets (EIN1, EIN2, ML, and SEP), and in the other three, performance is generally close to chance (TS, S4, and AB). Alternatively, if G expansion were seen as motion in the “wrong” direction and it is overcome by an F signal in the “right” direction, performance would be less than chance along the entire ordinate, and then increase above chance with movement parallel to the abscissa—but none of the subjects show this behavior either.

Finally, these data suggest that the responses to the pure F stimuli (plotted on the abscissa) depended on context (i.e., whether it was presented in the G contraction experiment vs. the G expansion experiment). In the G contraction experiment, they elicited strong motion percepts with low detection thresholds. However, in the context of stimuli containing mixtures of F and G expansion, sensitivity to F appeared to be reduced. We pursue this possibility further below.

Across subjects, F and G motion signals (and their combinations) were perceived more accurately in the G contraction experiments compared to the expansion experiments (paired t test for each subject on all data points; $p < 0.001$ for all subjects). The observed differences might be due to either: (a) differences in threshold detection levels of the relevant signals (G contraction and G expansion, which have different detection thresholds when presented alone) or (b) a result of the context in which they were presented, or a combination of (a) and (b).

To quantify these observations, we fit the psychophysical data along each ray to a Weibull function (see Methods). For G contraction, this provided a reasonable fit to the data, as shown in Figure 3A. Thus, psychophysical performance for G contraction in the entire C_F , C_G domain can be summarized by the Weibull threshold parameter (a_r) along each ray, with a consensus value of the Weibull shape parameter b (here, 1.65) used for all rays.

Figure 4 shows this summary of thresholds for all four subjects for the combination of F and G contraction signals, and confirms the observations made from Figure 2. There is integration between the two types of motion signals, as manifest by the curved trajectory of the performance threshold.

For the experiment examining combinations of F with G expansion, Weibull fits (e.g., see Figure 3B) could not capture qualitative aspects of the performance, since (as mentioned above) two subjects perceived motion opposite to the true direction for low levels of black G expansion. However, the Weibull fits along the F ray allowed for a direct comparison to the parallel condition in the G contraction experiment (Figure 5). In three datasets (EIN 1, TS, and S4) sensitivity to F in the context of the G expansion experiment was significantly decreased compared to

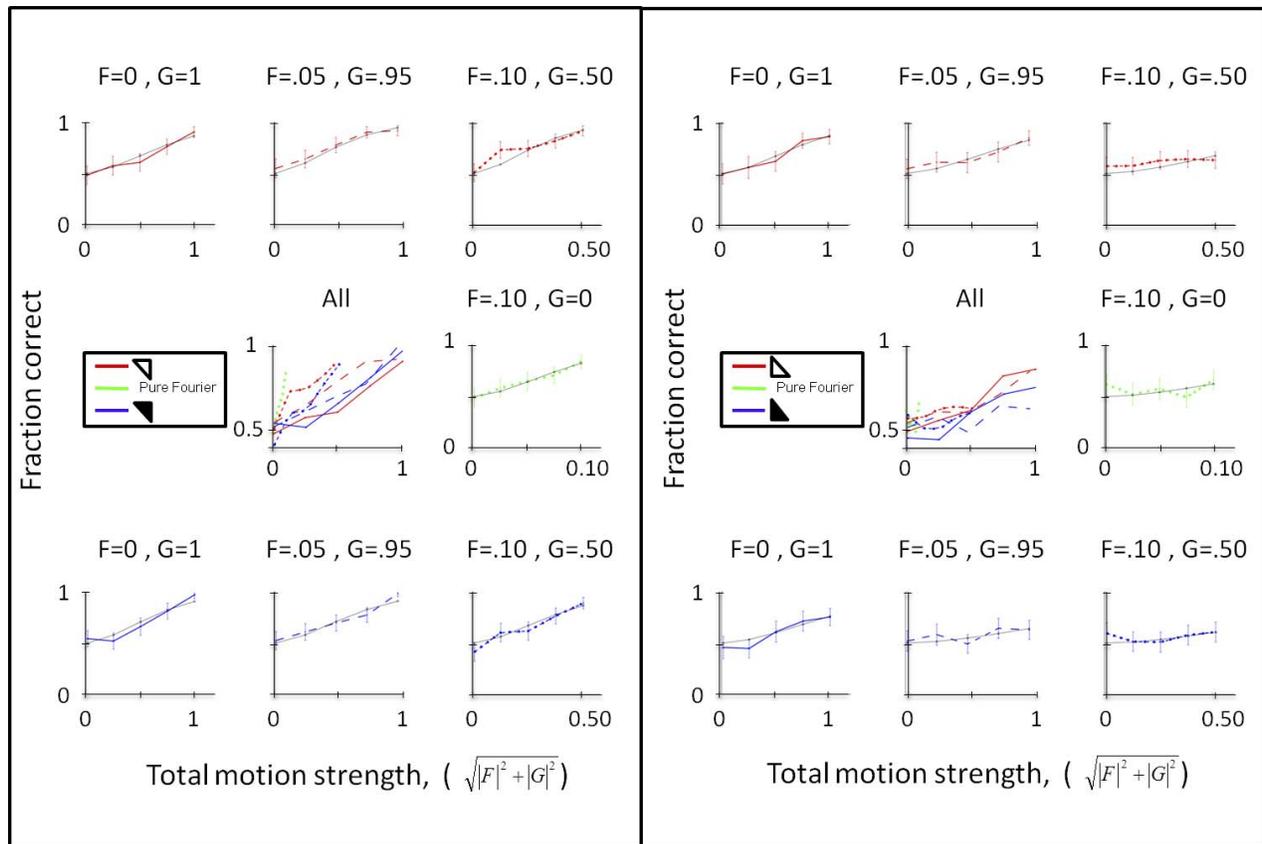


Figure 3. Fits of Weibull functions to performance along each ray. Panel A: F and G contraction; Panel B: F and G expansion. In each panel, the seven individual plots show the measured fraction correct along a single ray and the fitted Weibull function. The abscissa extends to the maximal total motion strength on that ray $\sqrt{|C_F|^2 + |C_G|^2}$; error bars indicate 95% CIs determined by bootstrap. Data are shown with a line color indicating glider parity (red: white G; blue: black G; green: pure F [no G]) and line style indicating the maximal motion strength at the end of the ray (solid: $C_F = 0, C_G = 1$; dashed: $C_F = 0.05, C_G = 0.95$; dot-dash: $C_F = 0.1, C_G = 0.5$; dotted: $C_F = 0.1, C_G = 0$). Black solid lines show the Weibull fit. The central plot shows all of the fits superimposed. In each panel, Weibull functions had the same shape parameter: $b = 1.65$ in Panel A, $b = 1.70$ in Panel B. Subject EIN, dataset 1.

sensitivity to F in the G contraction experiment. Results from three other datasets (EIN 2, ML, and SEP) are consistent with this notion: Performance in the expansion context never exceeds performance in the contraction context (within error bars). Two other subjects (PSS and AB) were only tested under one condition, so no within-subject comparison could be made. We note that the dataset of EIN 2, which did not show a significant difference between contexts, included fewer repeats.

In these experiments, the finding that the response to F motion is weaker in the context of G expansion than in the context of G contraction could also be explained by a nonspecific effect, i.e., a change in policy or attention driven by the greater difficulty of the blocks with G expansion (the overall fraction correct is lower). To examine this possibility, we carried out a further experiment in which we embedded weak pure F motion stimuli ($C_F = 0.1$) in three kinds of blocks, with a more profound difference in task difficulty. In each block,

only 20% of the blocks contained F motion; the other 80% of the trials were “context” trials that also controlled the overall difficulty of the block, containing just G contraction ($C_G = 0.95$), just G expansion ($C_G = 0.95$), or dynamic random checkerboards ($C_F = C_G = 0$).

Results are shown in Figure 6. As expected, the fraction correct for the random context trials was approximately 0.5, much lower than for either G context, as no motion signal was present (abscissa). If the reduction in the sensitivity to F in the context of G expansion seen above were due to a change in policy or attention because of the general lack of motion signals in these blocks, then F sensitivity in the context of random blocks should be lower than in either kind of G block (ordinate). This was seen in subject SEP, but not in the other four experiments: In those cases, including the two subjects (EIN and S4) who showed a context effect in Figure 5, fraction correct for the F trials in the random blocks was either the same as (EIN2 and S4) or

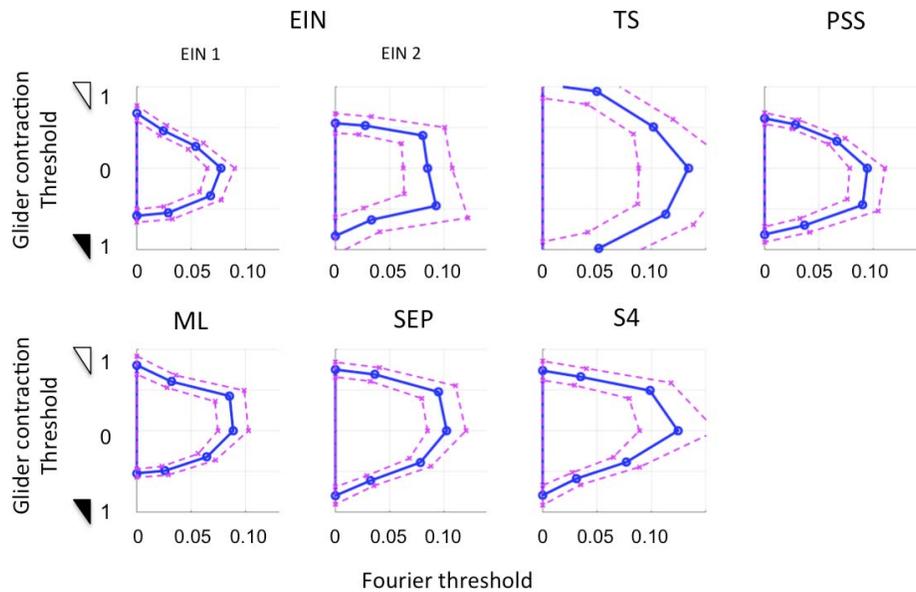


Figure 4. Isodiscrimination curves for combinations of F and G contraction. Blue curves connect the distances a_r along each ray at which a fraction correct of 0.75 is reached (see Methods). Purple dashed lines: 95% confidence limits. Ordinate shows G contraction strength, upper quadrant for white Gs and lower quadrant for black Gs.

slightly higher (EIN1) than the fraction correct for F trials in the G expansion blocks, even though the random-motion blocks were substantially harder (overall fraction correct was lower, and motion signals were more infrequent). In sum, while the data from SEP suggests a contribution of nonspecific effects to the sensitivity changes observed above, the data from the other subjects indicate that this is unlikely to be the sole explanation.

Discussion

Local motion signals can be distinguished by mathematical characterization of the underlying spatiotemporal correlations (Chubb & Sperling, 1988; Lu & Sperling, 2001; Reichardt, 1961). This approach identifies several kinds of motion elements (F, NF, and G) that are mathematically independent, and are also independent in an operational sense: Each kind of motion signal can be isolated experimentally, by creating artificial stimuli that drive it and none of the others. These different kinds of motion signals have been shown to elicit behavioral responses in humans (Hu & Victor, 2010; Lu & Sperling, 1995) and several other species (*Drosophila*: Clark et al., 2014; zebrafish: Orger, Smear, Anstis, & Baier, 2000; dragonfly: Nitzany et al., 2014; macaque: Nitzany et al., 2014; O’Keefe & Movshon, 1998).

However, this subdivision of motion signal types is arguably unnatural, in that individual motion types rarely occur in isolation in natural sensory inputs. Rather, as we recently showed (Nitzany & Victor, 2014), several kinds of local motion signals occur together in naturalistic inputs, and typically in overlapping locations. While different naturalistic inputs have similar mixtures of motion signal types, and the co-occurrence of these signal types is correlated on a scene-by-scene basis, they are not redundant: The presence of one motion signal only predicts the strength of another within about a factor of two.

In addition, although recent theoretical work (Fitzgerald et al., 2011) indicates that motion analysis can

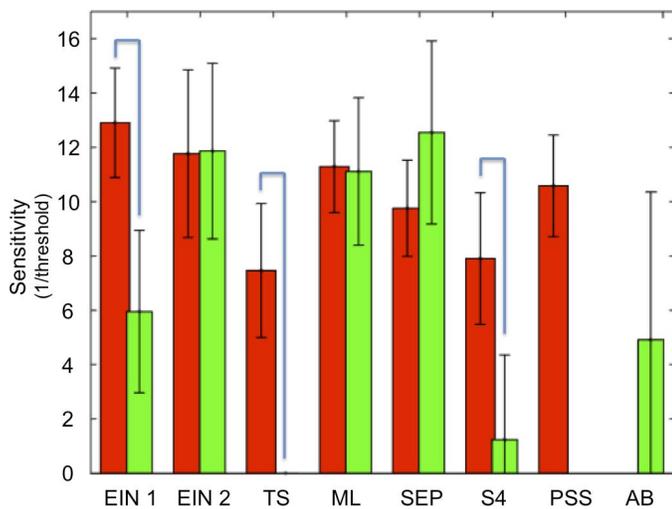


Figure 5. Comparison of sensitivity ($1/\text{threshold}$; i.e., $1/a_r$ in Equation 1) to F motion in the context of G contraction (red) and expansion (green). Error bars indicate 95% CIs. Braces connect data from subjects run in both kinds of experiments. Note that for subject TS, there was no measurable sensitivity to F motion in the context of G expansion.

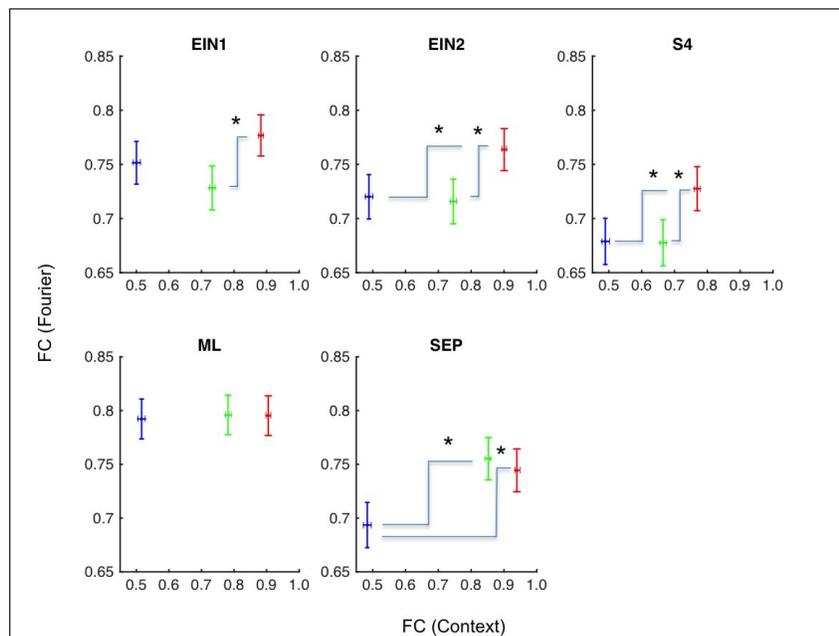


Figure 6. Comparison of fraction correct (FC) for F motion, with strength $C_F = 0.1$, in three contexts: G contraction (red), G expansion (green), and random movies (blue). Error bars indicate 1 SEM. Brackets indicate significant ($p < 0.05$) differences in FC for F (ordinate) motion via a one-tailed (G contraction > G expansion > random) t test, paired across blocks. FC for the context trials (abscissa) is significantly different at $p < 0.05$ for every within-subject comparison.

benefit from integrating different types of motion signals, it is unclear whether our visual systems take advantage of this strategy. Furthermore, because different motions are generally associated with different natural phenomena (e.g., translation leads to primarily F motion, while looming objects generate prominent G expansion signals and receding objects generate prominent G contraction signals), biological motion processing may be context-dependent. Here, using the simple psychophysical task, we studied the integration of these two motion signal types, and tested for context-dependence of the processing of F motion. We found that integration of subthreshold signals occurred when F and G signals were simultaneously present, as manifest in elliptical isodiscrimination contours (Figure 3). This is the behavior expected of computational models that make optimal use of two kinds of signals that are nonredundant (Fitzgerald et al., 2011). But surprisingly, this integration only occurred for G contraction. For G expansion, we observed a context effect in some subjects. In blocks in which G expansion was present, sensitivity to F signals was slightly reduced.

What factors may affect our ability to perceive different kinds of local motion signals (F, G contraction, and G expansion signals)? Two nonexclusive options are immediately apparent: First, each kind of signal may have different threshold detection levels, as they are different stimuli, which are also associated with different natural phenomena (translation, receding

from and looming towards objects, respectively—though we emphasize that the stimuli used here are planar, and do not elicit a 3-D percept). Second, the context in which each kind of signal appears may affect these thresholds. Natural stimuli tend to include a mixture of local motion signals (Nitzany & Victor, 2014). Most likely, both factors play a role. On one hand, threshold levels for the different motion signal kinds (i.e., F and G) are different, as are sensitivities for G contraction and expansion (Figure 6, abscissa, and Hu & Victor, 2010). On the other hand, some subjects showed a context effect on F sensitivity—reduced in the context of G expansion compared to G contraction (Figure 6, ordinate)—and this context effect may also have influenced the sensitivities to the G stimuli as well (see below).

The greater sensitivity to G contraction than G expansion is puzzling, and raises questions that we cannot at present resolve. Since G expansion is associated with looming and G contraction is associated with receding, it is surprising that observers are less sensitive to G expansion than to G contraction. More work needs to be done on this issue. We may speculate that the reason is that objects undergoing 3-D motion typically generate a strong F motion cue along with the G motion cue; when only one cue is present, the resulting stimulus is not as likely to be interpreted as true motion. But this is at best a partial answer, since one would then expect a cooperative interaction

between G expansion and F motion, rather than a weak suppression.

Our findings are in line with those of Hu and Victor (2010), but the correspondence is not complete, and we suspect that this difference reflects aspects of contextual modulation. As in Hu and Victor, 2010, we found that sensitivity to G motion was substantially less than sensitivity to F motion, and that responses to G contraction were veridical (i.e., that inversion of the sign of contrast [black vs. white Gs] did not invert the perceived direction). For G expansion, however, the findings differed somewhat. Hu and Victor (2010) found that inversion of contrast led to a reversal of the perceived motion direction: White (“even” parity) G expansion elicited perceived motion in the veridical direction, while black G expansion elicited perceived motion opposite to the veridical direction. Here, we only found this inversion at low levels of G strength, and only in the two naive subjects (TS and AB, Figure 2). We hypothesize that this difference may result from a difference in the experimental paradigm. Specifically, Hu and Victor, 2010, presented stimuli with many kinds of Gs at full correlation strength, but randomly interleaved on a trial-by-trial basis, so there was no establishment of a G expansion or contraction context. Here, only G contraction (or only G expansion) was presented, for blocks of up to 3,500 trials, across sittings with a total duration of several hours, enabling a context to develop.

The lack of inversion between black and white G motion, along with differences in sensitivity to black versus white G motion seen in some subjects, are both forms of black-white asymmetry: An opponent model with multipoint nonlinear interactions would predict inversion of perceived motion direction when polarity inverts, and no change in sensitivity (see table 1 of Hu & Victor, 2010). This asymmetry also deserves further exploration. We speculate that it may have a functional role related to black-white asymmetries in the natural visual environment (Fitzgerald et al., 2011), and related black-white asymmetries have been found in studies of G motion in *Drosophila* (Clark et al., 2014; Fitzgerald & Clark, 2015).

Summary

Using novel synthetic stimuli that contain controlled amounts of two kinds of local motion signals, we find two kinds of interactions at the perceptual level. For combinations of F and G contraction signals, we find subthreshold integration. For combinations of F and G expansion signals, subthreshold integration is much less prominent. In some subjects, sensitivity to F signals is mildly reduced in comparison to sensitivity to these signals in the context of G expansion.

Keywords: local motion, motion perception, non-Fourier motion

Acknowledgments

We are very grateful to Leslie Osborne for running some of the experiments in her lab space. We are also grateful to two anonymous reviewers whose insightful comments helped us improve this manuscript, and to James Fitzgerald for some very helpful discussion. This work was supported in part by National Institute of Health Grant No. EY7977 to JDV. SEP was supported by Alfred P. Sloan Foundation Research Fellowship and Big Ideas Generator Vision grant.

Commercial relationships: none.

Corresponding author: Eyal I. Nitzany.

Email: eyalni@gmail.com.

Address: Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL, USA.

References

- Albright, T. D. (1992). Form-cue invariant motion processing in primate visual cortex. *Science*, 255(5048), 1141–1143, doi:10.1126/science.1546317.
- Chaudhuri, A., & Albright, T. D. (1997). Neuronal responses to edges defined by luminance vs. temporal texture in macaque area v1. *Visual Neuroscience*, 14(5), 949–962, doi:10.1017/S0952523800011664.
- Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli- a general basis for studying non-fourier motion perception. *Journal of the Optical Society of America, A: Optics and Image Science*, 5(11), 1986–2007, doi:10.1364/JOSAA.5.001986.
- Clark, D. A., Fitzgerald, J. E., Ales, J. M., Gohl, D. M., Silies, M. A., Norcia, A. M., & Clandinin, T. R. (2014). Flies and humans share a motion estimation strategy that exploits natural scene statistics. *Nature Neuroscience*, 17(2), 296–303, doi:10.1038/nn.3600.
- Edwards, M., & Badcock, D. R. (1995). Global motion perception: no interaction between the first- and second-order motion pathways. *Vision Research*, 35(18), 2589–2602, doi:10.1016/0042-6989(95)00304-5.
- Fitzgerald, J. E., & Clark, D. A. (2015). Nonlinear circuits for naturalistic visual motion estimation. *eLife*, 4, e09123, doi:10.7554/eLife.09123.

- Fitzgerald, J. E., Katsov, A. Y., Clandinin, T. R., & Schnitzer, M. J. (2011). Symmetries in stimulus statistics shape the form of visual motion estimators. *Proceedings of the National Academy of Sciences*, *108*(31), 12909–12914, doi:10.1073/pnas.1015680108.
- Greenlee, M. W., & Smith, A. T. (1997). Detection and discrimination of first- and second-order motion in patients with unilateral brain damage. *The Journal of Neuroscience*, *17*(2), 804–818.
- Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, *55*(1), 48–120, doi:10.3758/BF03206880
- Hu, Q., & Victor, J. D. (2010). A set of high-order spatiotemporal stimuli that elicit motion and reverse-phi percepts. *Journal of Vision*, *10*(3):9, 1–16, doi:10.1167/10.3.9. [PubMed] [Article]
- Ledgeway, T. (1994). Adaptation to second-order motion results in a motion aftereffect for directionally-ambiguous test stimuli. *Vision Research*, *34*(21), 2879–2889, doi:10.1016/0042-6989(94)90056-6.
- Lu, Z. L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research*, *35*(19), 2697–2722, doi:10.1016/0042-6989(95)00025-U.
- Lu, Z. L., & Sperling, G. (2001). Three-systems theory of human visual motion perception: Review and update. *Journal of the Optical Society of America, A*, *18*(9), 2331–2370, doi:10.1364/JOSAA.18.002331.
- Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. *Pattern Recognition Mechanisms*, *54*, 117–151, doi:10.1007/978-3-662-09224-8_7.
- Nitzany, E. I., Menda, G., Shamble, P. S., Golden, J. R., Hoy, R. R., & Victor, J. D. (2014). Evolutionary convergence in computation of local motion signals in monkey and dragonfly. Presented at CoSyNe, Salt Lake City, Utah.
- Nitzany, E. I., & Victor, J. D. (2014). The statistics of local motion signals in naturalistic movies. *Journal of Vision*, *14*(4):10, 1–15, doi:10.1167/14.4.10. [PubMed] [Article]
- O’Keefe, L. P., & Movshon, J. A. (1998). Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey. *Visual Neuroscience*, *15*(2), 305–317, doi:10.1017/S0952523898152094.
- Orger, M. B., Smear, M. C., Anstis, S. M., & Baier, H. (2000). Perception of Fourier and non-Fourier motion by larval zebrafish. *Nature Neuroscience*, *3*(11), 1128–1133, doi:10.1038/80649.
- Piponnier, J.-C., Forget, R., Gagnon, I., McKerral, M., Giguère, J.-F., & Faubert, J. (2015). First- and second-order stimuli reaction time measures are highly sensitive to mild traumatic brain injuries. *Journal of Neurotrauma*, *33*, 242–253, doi:10.1089/neu.2014.3832.
- Plant, G. T., Laxer, K. D., Barbaro, N. M., Schiffman, J. S., & Nakayama, K. (1993). Impaired visual motion perception in the contralateral hemifield following unilateral posterior cerebral lesions in humans. *Brain*, *116*(6), 1303–1335, doi:10.1093/brain/116.6.1303.
- Plant, G. T., & Nakayama, K. (1993). The characteristics of residual motion perception in the hemifield contralateral to lateral occipital lesions in humans. *Brain*, *116*(6), 1337–1353, doi:10.1093/brain/116.6.1337.
- Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. *Sensory Communication*, 303–317.
- Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M., & Hennig, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). *The Journal of Neuroscience*, *18*(10), 3816–3830.
- Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London B: Biological Sciences*, *203*(1153), 405–426, doi:10.1098/rspb.1979.0006.
- Vaina, L. M., & Soloviev, S. (2004). First-order and second-order motion: neurological evidence for neuroanatomically distinct systems. *Progress in Brain Research*, *144*, 197–212, doi:10.1016/S0079-6123(03)14414-7.
- Victor, J. D., Chubb, C., & Conte, M. M. (2005). Interaction of luminance and higher-order statistics in texture discrimination. *Vision Research*, *45*(3), 311–328, doi:10.1016/j.visres.2004.08.013.
- Victor, J. D., & Conte, M. M. (1992). Evoked potential and psychophysical analysis of Fourier and non-Fourier motion mechanisms. *Visual Neuroscience*, *9*(2), 105–123, doi:10.1017/S0952523800009573.
- Victor, J. D., & Conte, M. M. (2012). Local image statistics: Maximum-entropy constructions and perceptual salience. *Journal of the Optical Society of America, A*, *29*(7), 1313–1345, doi:10.1364/JOSAA.29.001313.
- Victor, J. D., Thengone, D. J., & Conte, M. M. (2013). Perception of second- and third-order orientation signals and their interactions. *Journal of Vision*, *13*(4):21, 1–21, doi:10.1167/13.4.21 [PubMed] [Article]

Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, 9(1), 79–97, doi:10.1017/S0952523800006386.

Yo, C., & Wilson, H. R. (1992). Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research*, 32(1), 135–147, doi:10.1016/0042-6989(92)90121-X.