Field-like interactions between motion-based reference frames

Mehmet N. Agaoglu^{1,2} · Michael H. Herzog³ · Haluk Öğmen^{1,2}

Published online: 17 April 2015
© The Psychonomic Society, Inc. 2015

Abstract A reference frame is required to specify how motion is perceived. For example, the motion of part of an object is usually perceived *relative* to the motion of the object itself. Johansson (*Psychological Research*, 38, 379–393, 1976) proposed that the perceptual system carries out a vector decomposition, which results in common and relative motion percepts. Because vector decomposition is an ill-posed problem, several studies have introduced constraints by means of which the number of solutions can be substantially reduced. Here, we have adopted an alternative approach and studied how, rather than why, a subset of solutions is selected by the visual system. We propose that each retinotopic motion vector creates a reference-frame field in the retinotopic space, and that the fields created by different motion vectors interact in order to determine a motion vector that will serve as the reference frame at a given point and time in space. To test this theory, we performed a set of psychophysical experiments. The field-like influence of motion-based reference frames was manifested by increased nonspatiotopic percepts of the backward motion of a target square with decreasing distance from a drifting grating. We then sought to determine whether these field-like effects of motion-based reference frames can also be extended to stationary landmarks. The results suggest that reference-field interactions occur only between motion-generated fields. Finally, we investigated whether and how different reference fields interact with each other, and found

that different reference-field interactions are nonlinear and depend on how the motion vectors are grouped. These findings are discussed from the perspective of the reference-frame metric field (RFMF) theory, according to which perceptual grouping operations play a central and essential role in determining the prevailing reference frames.

Keywords 2-D motion · Motion integration · Temporal processing

The relativity of perceived motion

By definition, *motion* is a change of position over time. Hence, the definition implies that to determine motion, one needs to have references (or coordinate systems) for position and time, and motion becomes relative to these coordinate systems. The choice of the coordinate systems and their scales depend on the phenomena of interest. For example, astronomical scales are used to characterize planetary motions, and the orbital speed of earth's motion according to a solar reference frame is about 30 km/s. However, in our ecological environment, for all practical purposes the earth appears stationary, and an earth-based (geocentric) reference frame prevails. As a result, explicitly or implicitly, geocentric (also called *spatiotopic*) motion is generally regarded as the “real,” “physical,” or “absolute” motion, whereas motion relative to other reference frames is considered “relative” or “illusory” motion (Swanston, Wade, & Day, 1987; Wade & Swanston, 1987). However, in analyzing its inputs, our visual system is faced by the complexity that arises at the individual stimuli level, and perceptual organizational principles, such as Gestalt grouping and figure-ground segregation are proposed to be fundamental requirements in processing visual inputs (Koffka, 1935; Wagemans, Elder, et al., 2012; Wagemans, Feldman, et al., 2012). The perception of a stimulus does not only depend on its own individual properties but also on the properties of

✉ Mehmet N. Agaoglu
mnagaoglu@gmail.com

¹ Department of Electrical and Computer Engineering, University of Houston, N308 Engineering Building 1, Houston, Texas 77204-4005, USA
² Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX, USA
³ Laboratory of Psychophysics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

other spatiotemporally neighboring stimuli (Koffka, 1935). It is not surprising, then, that relative motion was a central topic in Gestalt psychology (Duncker, 1929; Ellis, 1938). In one of his experiments, Karl Duncker used displays generated by point-lights attached to an otherwise invisible rotating and translating circular piece of cardboard (Duncker, 1929, p. 240). He found that when a single point-light was attached to the rim of the cardboard, observers reported seeing the point-light moving along a cycloidal trajectory. On the other hand, when he added a second point-light to the hub of the cardboard, observers often reported perceiving the point-light attached to the rim undergoing circular motion around the point-light on the hub, which itself was perceived to move horizontally. These results can be understood in terms of the reference frame against which the motion of the point-light is perceived. Cycloidal motion corresponds to a trajectory relative to a geocentric reference, whereas the rotation corresponds to a trajectory relative to a *moving* reference frame positioned on the point-light at the hub.

The theory of perceptual vector decomposition

In order to demonstrate the power of the visual system in determining behaviorally relevant reference frames, Johansson devised the “biological motion” paradigm, in which point-lights are attached to a few joints of humans undergoing complex motion (Johansson, 1973). In viewing these displays, observers readily perceived the underlying biological motion. In other words, observers were able to select the appropriate reference frame(s) that revealed the underlying biological motion. In order to explain the selection of the reference frame, Johansson proposed a theory of vector decomposition (Johansson, 1976). This theory is based on three principles: (i) elements under motion are always perceptually related, (ii) the simultaneous motion of elements forms rigid perceptual groups, and (iii) the decomposition of motion vectors into equal and simultaneous motion vectors leads to the perception of “common motion,” and the residual motion vectors will be perceived as “relative motion.” Figure 1 illustrates these concepts. Assume that the three vertically positioned dots shown in Fig. 1A have the corresponding motion vectors associated with them individually. The dots at the top and bottom move back and forth along the horizontal dimension, whereas the dot in the middle moves along an oblique trajectory. What observers mostly perceive is shown in Fig. 1B: all three dots moving in tandem to the right and to the left (blue arrows) as a group, and the one in the middle moving up and down (black arrows). According to the first principle, the individual trajectories of dots are not perceived in isolation, but in relation to each other. According to the second principle, the top and bottom dots form a perceptual group due to their simultaneous motions. Finally, according to the third and most important principle, the motion of the middle dot is

Fig. 1 (A) The three-element display used by Johansson (1976) to illustrate “perceptual vector decomposition.” Three dots with their corresponding motion vectors are shown. (B) Observers often report seeing all three dots moving in tandem horizontally (colored dashed ellipse and horizontal arrows). The middle dot is perceived to be moving up and down (black vertical arrows). (C) The vector decomposition corresponding to the percept in panel B: The simultaneous motion vectors of the dots at the top and bottom and the horizontal motion component of the dot in the middle together form a perceptual group. What is left is the middle dot’s relative motion, which results in an up-down motion percept (vertical dotted arrow). (D) An alternative way of decomposing the same motion vectors into common and relative motion components. (E) The physical motion trajectories in the two-dot display used by Johansson (1950). (F) Perfect vector decomposition predicts that the two dots will appear to move on a common diagonal axis toward each other, while the axis itself moves in the orthogonal direction. (G) Many observers perceive each dot moving along a different path, and these paths intersect with an angle around 30°–40° (Wallach, Becklen, & Nitzberg, 1985)

decomposed into a horizontal component, which is common to the other dots, and a vertical component, which makes it look like it is moving up and down. Figure 1C illustrates the common and relative motion components of each dot, resulting in the percept shown in Fig. 1B (for reviews, see Herzog & Ögmen, 2014; Shum & Wolford, 1983).

Vector decomposition: An ill-posed problem

Although the processes illustrated in Fig. 1 seem straightforward, in mathematical terms, vector decomposition is an ill-posed problem: Infinitely many pairs of common and relative motions can produce exactly the same absolute motion. Figure 1D shows an alternative set of common and relative motion components corresponding to the same physical motion as in Fig. 1A. Johansson recognized this ambiguity in some of his studies (Johansson, 1950, 1958; Johansson & Jansson, 1968). For example, in his Exp. 19 (Johansson, 1950, p. 89), in which he presented two-dot displays in which

one of the dots oscillated horizontally while the other oscillated vertically (see Fig. 1E), he reported that subjects did not always experience the same motion configuration. If they attended to one of the dots, they perceived the geocentric motion of that dot, while perceiving the other dot as moving along a slanted trajectory. There were even reports of 3-D rigid motion of a rotating rod. Similarly, with Duncker's (1929) wheel stimulus, some observers reported a rotating wheel, whereas others reported that the motion of two point-lights more resembled a tumbling stick (Cutting & Proffitt, 1982; Proffitt, Cutting, & Stier, 1979).

Since observers do not perceive all possible solutions, but instead a rather small subset (see, e.g., Johansson, 1950; Proffitt et al., 1979; Wallach, Becklen, & Nitzberg, 1985), the fundamental questions are to determine which subset of solutions is selected by the visual system, and why. Mathematically, the number of solutions of an ill-posed problem can be reduced by introducing additional information or constraints, an approach known as *regularization* (Marr & Ullman, 1981).

Which subset and why: Regularization approach in vector decomposition

A variety of constraints have been proposed to explain how the visual system regularizes vector decomposition. Hochberg and McAlister (1953) argued that the perceptual system chooses the simplest solution in terms of the information required to define the pattern when it encounters an ambiguous stimulus with multiple potential interpretations. In other words, the alternative in Fig. 1D will be rejected because it necessitates additional motion vectors rather than just one, as in the case of Fig. 1C. Börjesson and von Hofsten (1972) proposed as a constraint that residual motion vectors should sum to zero. Gogel proposed the “adjacency principle,” according to which the relative motion determination is restricted only to nearby objects (Gogel, 1974; Gogel & Koslow, 1972). Proffitt and colleagues proposed that the common motion is determined by the motion of the center of gravity of the dots (Proffitt et al., 1979). Restle (1979) proposed “information load” as the constraint to be minimized in determining the prevailing solution. A hybrid system that minimizes either the common or relative component, depending on which process (assuming that the common and relative motion calculations are done via independent processes) is completed first, has been shown to account for some of the classical findings in dot motion experiments (Cutting & Proffitt, 1982). More recently, building upon Johansson's (1950) original study of vector analysis, a Bayesian framework with a set of probabilistic constraints has also been introduced (Gershman, Jäkel, & Tenenbaum, 2013). In sum, the constraints introduced in regularization approaches to vector decomposition provide heuristics to explain, at least partially, *why* the human visual

system selects a particular vector decomposition in motion perception.

To put this approach in perspective, consider its use in physics. In physics, the principle of minimum total potential energy (which states that particles move so as to minimize the total potential energy) is formulated to explain *why* things move the way they do. Hence, on the basis of this constraint, a global energy function can be minimized to determine the motion of particles in a medium. An alternative perspective is to express *how* this particular solution emerges in real time through interactions between particles. In this case, one uses forces and fields applied to particles. In mathematical terms, these two approaches are related, and can be expressed as the energy (or Ljapunov) function of a system and the differential equations governing the system.

Receptive-field-based approaches

In the 1980s and 1990s, significant advances occurred in characterizing the neurophysiological and computational properties of neurons that show selectivity to motion (Albright & Stoner, 1995; Lu & Sperling, 2001). These neurons can be described as retinotopically anchored spatio-temporal filters with receptive fields that are oriented in space–time and that contain center–surround antagonism (Adelson & Bergen, 1985; Watson & Ahumada, 1985). Hence, various approaches have been proposed to explain relative-motion perception using these concepts. For example, the mutual influences between two moving stimuli can be explained by center–surround antagonism in the receptive fields of the motion mechanisms (e.g., Kim & Wilson, 1997; Murakami & Shimojo, 1996; Nawrot & Sekuler, 1990; Tadin, Lappin, Gilroy, & Blake, 2003). Although these approaches can explain how we perceive the motion of relatively simple stimuli, such as the motion of a central patch in the presence of a surrounding stimulus, and have the advantage of relating perception to the single-neuron level, they cannot explain complex motion patterns such as the biological motion in Johansson's studies. For this, one needs to consider the activities of networks of neurons, to determine the reciprocal influences between multiple stimuli and the emergent motion percepts for each individual stimulus. In building large-scale network models, a common theoretical view is to adopt the atomic hierarchical approach, in which a first, bottom-up stage computes basic features (properties), such as motion, orientation, and so forth, through receptive fields acting as selective filters, and a second stage combines the outputs of these filters to form more-complex percepts (Riesenhuber & Poggio, 1999). However, recent findings suggest that the computation of basic features cannot be abstracted from Gestalt organizational processes such as perceptual grouping (Herzog, Hermens, & Ögmen, 2014; Thunell, Herzog, & Ögmen, 2015). For example, vernier acuity, crowding, attention, motion, and form perception cannot

be explained by simple receptive-field filtering approaches. Consider, for example, the stimulus shown in Fig. 2 (Boi, Ögmen, Krummenacher, Otto, & Herzog, 2009). The three black disks form a Ternus–Pikler display (Pikler, 1917; Ternus, 1926). Depending on the value of the interstimulus interval (ISI), the disks are perceived to be in “group motion” or “element motion” (Figs. 2a and b; see the demos in Boi et al., 2009). The data in Fig. 2d indicate that observers can

judge the direction of rotation of the dot in the central disk only when the disks are perceived to be in group motion. Since the perception of rotation requires the use of a reference frame that moves along with the perceived motion of Ternus–Pikler elements, this finding indicates a *perceptual-grouping-based moving reference frame* and illustrates the importance of perceptual grouping in computing basic motion. Receptive-field-based models of motion cannot explain these perceptual-

Fig. 2 Nonretinotopic motion perception in the Ternus–Pikler paradigm. (a) When the interstimulus interval (ISI) is long (210 ms), the three disks appear to move in tandem as a group (“group percept”). (b) When a short ISI is used (e.g., 0 ms), the percept becomes “element motion,” in which the leftmost disk in the first frame appears to jump to the rightmost disk, and vice versa, whereas the other two disks are perceived to be stationary. (c) Removal of the flanking disks leads to the perception of two stationary disks. (d) Performance is judging the direction of rotation for the dot in the central disk. (e) Average horizontal eye movement pattern during stimulus presentation for one observer under group-motion conditions. The light

gray rectangles represent the space–time diagram of the horizontal position of the central disk, whereas the white ones represent the positions of the lateral disks. The dot position within the disk is shown by the small dark gray rectangles; 95 % confidence intervals are shown by the light gray lines flanking the dark gray lines. From “A (Fascinating) Litmus Test for Human Retino- vs. Non-Retinotopic Processing,” by M. Boi, H. Ögmen, J. Krummenacher, T. U. Otto, and M. H. Herzog, 2009, *Journal of Vision*, 9(13), Article 5, page 4. Copyright 2009 by the Association for Research in Vision and Ophthalmology (ARVO ©). Reprinted with permission. Source: http://www.journalofvision.org/site/misc/terms_conditions.xhtml

grouping-dependent motion percepts (Clarke, Repnow, Ögmen, & Herzog, 2012). Hence, an alternative approach to reference-frame synthesis is needed to take into account perceptual-grouping operations.

Reference-frame metric field (RFMF) theory

The early visual system is retinotopically organized. Furthermore, a briefly presented stimulus remains visible for about 100 ms after its offset, a phenomenon known as *visible persistence* (Coltheart, 1980; Haber & Standing, 1970). On the basis of retinotopic representations and visible persistence, one would expect moving objects to appear extensively blurred, without any form information (Ögmen, 2007; Ögmen & Herzog, 2010). However, under normal viewing conditions, the moving objects appear relatively sharp and clear (Hammett, 1997; Ramachandran, Rao, & Vidyasagar, 1974). To explain how dynamic form is computed while avoiding these problems, we proposed the RFMF theory (Ögmen, 2007; Ögmen & Herzog, 2010; Ögmen, Herzog, & Noory, 2013). The main idea is to compute dynamic form not according to a retinotopic reference frame, but instead according to a reference frame that moves according to the motion of the stimuli. Figure 3 provides a schematic description. The plane at the bottom depicts retinotopic representations. Two groups of dots, shown in different colors, move in different directions in retinotopic coordinates. The first step in RFMF is to determine perceptual grouping of the motion vectors in the retinotopic space. Accordingly, the different sets of dots are distinguished into two different groups, and a common motion is determined for each group. The requirement for perceptual grouping at this stage comes from the fact that, in a scene,

multiple objects can move in different directions, and hence multiple reference frames need to be established. Grouping motion vectors first allows the determination of a common reference frame for each group. This common motion is used to map the retinotopic representations into distinct nonretinotopic representations, depicted as spheres at the top of the figure. According to RFMF theory, each retinotopic motion vector creates a field in the retinotopic space (like an electromagnetic field). The fields created by different motion vectors interact in order to determine a motion vector that will serve as the reference frame at a given point and time in space. Previously, we have demonstrated the field-like nature—that is, the distance-dependent influence—of motion-based reference frames in the perception of motion direction (Agaoglu, Herzog & Ögmen, 2015; Noory, Herzog, & Ögmen, 2013, 2015; Ögmen et al., 2013). This concept of distance-dependent reference-frame determination is also consistent with earlier observations of relative motion. For example, Mori (1979) found a linear decrease in the perceived horizontal component of the middle dot in Fig. 1A with increased separation between the dots. However, an important difference between RFMF and other approaches is that perceptual grouping plays a critical and essential role during the process of reference-frame extraction. A second important difference is that RFMF is *not* designed to explain only the perception of relative motion, but to explain the perception of *all* attributes of the stimuli. It can also explain how and why form is perceived when the retinotopic layout of the stimulus is reduced to a very narrow slit (anorthoscopic perception; Agaoglu, Herzog, & Ögmen, 2012; Aydın, Herzog, & Ögmen, 2008, 2009).

The goal of the present study was to investigate whether and how different reference-frame fields interact. We sought to determine (i) whether reference-frame fields are actually generated by local motion vectors (i.e., to replicate previous findings with a variant of the stimuli previously used), (ii) whether stationary landmarks generate reference fields as well, (iii) whether the fields created by different motion-based reference frames actually interact, and if so, (iv) how they interact. To this end, we probed the perceived direction of a moving dot with and without drifting gratings (which produce motion-based reference frames) at various distances from the dot, and we quantified the field effect and relativity of the perceived motion.

Fig. 3 Schematic illustration of the reference-frame metric field theory. The motions of several differently colored dots are shown in the retinotopic space (the plane). Dots are grouped into two groups based on their motion vectors. A reference motion vector is extracted and serves as the reference by which dots are mapped into nonretinotopic representations (spheres at the top). The reference motion vectors' effects spread over space and time much like an electromagnetic field (dotted ellipses, whose thicknesses symbolize field strength). The interactions between different reference fields, if there are many, determine the resultant nonretinotopic representations. For more details, see Ögmen (2007) and Ögmen and Herzog (2010)

General method

Participants

Three naive observers and one of the authors (M.N.A.) participated in the study. The age of the participants ranged from 26 to 29 years, and all participants had normal or corrected-to-

normal vision. All experiments followed a protocol approved by the University of Houston Committee for the Protection of Human Subjects, and each observer gave written informed consent before the experiments.

Apparatus

All visual stimuli were created via the VSG2/5 visual stimulus generator card (Cambridge Research Systems) and displayed at a resolution of 800×600 pixels, with a refresh rate of 100 Hz. Gaze position monitoring for both eyes was performed by means of an EyeLink II eyetracker at a 250-Hz sampling rate. The distance between observers' eyes and the display was 1 m, and the dimensions of the display at this distance were 22.7×17.0 deg. A head/chin rest was used to help stabilize fixation, and observers reported their responses via a joystick. All experiments were done in a normally illuminated room.

Experiment 1: The reference-field effect

The purpose of the experiment was to demonstrate the distance-dependent effects of a motion-based reference frame—that is, the reference-field effect. In Experiment 1a, we established a baseline for the judgments of motion direction in the absence of any dynamic reference frames. In Experiment 1b, we added a motion-based reference frame to the display and varied the distance of the target from this reference.

Method

Stimuli and procedures Previously, we have used a variant of the induced motion paradigm to see how various reference frames influence the perception of motion (Agaoglu et al., 2015). We used a small white square (56 cd/m^2), similar to the stimuli used in that study, as the target on a black background ($<0.5 \text{ cd/m}^2$; see Fig. 4A). Each trial started off with presentation of a fixation cross at the center of the display for a randomly chosen duration (400–700 ms). As soon as the target appeared, the fixation cross was turned off, to avoid additional reference frames that might be used in motion judgments. Observers were required to keep fixation at the remembered location of the fixation cross during the trial. Trials during which the gaze positions of observers deviated more than 2 deg from the fixation cross were discarded and repeated immediately. The target square moved horizontally at a fixed vertical eccentricity in a given trial. Five possible vertical eccentricities for the target square were used in all conditions: 0, ± 2.75 , and ± 5.5 deg (positive values represent the upper visual field), and the direction of the target motion was randomized across trials. An example velocity profile of the target, in

which it moved from left to right, is given in Fig. 4F (thin curves). The velocity of the target was constant during the first and last 630 ms. From 630 to 1,890 ms, the velocity of the target square was modulated by a sine wave. As long as the amplitude of the sine wave was smaller than the average speed of the target ($9^\circ/\text{s}$), the target might either decelerate or accelerate, but it would never move backward in spatiotopic (e.g., screen-based) coordinates (light thin curves in Fig. 4F). However, if the amplitude of the sine wave exceeded the average speed, the target stopped and reversed its direction for a short amount of time (dark thin curves). Figure 4G shows the space–time diagram of the target's motion for various amplitudes of the sine modulation. Note that the x -axis represents roughly the second quarter of the target motion on the display. When the amplitude of the sine wave is equal to the average speed of the target—that is, $9^\circ/\text{s}$ —the target slows down and stops completely, and starts accelerating until it reaches $9^\circ/\text{s}$ speed again. When the amplitude of the sine modulation exceeds $9^\circ/\text{s}$, dips in the space–time curves, indicating *spatiotopic* backward motion, become apparent (modulation amplitudes 12 and 15 in Fig. 4G).

In Experiment 1a (Fig. 4A), the target square was presented alone at various vertical eccentricities across blocks (0, ± 2.75 , and ± 5.5 deg; positive and negative values indicate the upper and lower visual fields, respectively). In Experiment 1b, a square wave grating (dimensions: 23×1 deg, spatial frequency: 0.25 cpd, duty cycle: 50 %, Michelson contrast: 0.98) was also presented, always at 6.5-deg vertical eccentricity (in either the upper or the lower part of the display). The distances between the target square and the grating were, therefore, 1, 3.75, 6.5, 9.25, and 12 deg. The vertical eccentricities of the target (i.e., target–grating distances) were blocked, and the order of blocks was randomized. The drift speed of the grating was equal to the average speed of the target square ($9^\circ/\text{s}$), and the drift was always in the same direction as the target square's motion (see the thick arrow in Fig. 4B and the thick top line in Fig. 4F).

Observers were asked to spread their attention to the entire display and, as soon as the target (and the grating) completed their motion and disappeared, to report via a joystick whether the target had ever moved backward during the trial (yes/no). The amplitude of the sine modulation in the target's velocity profile was varied across trials by an adaptive staircase algorithm (see the various thin curves showing different modulation amplitudes in Fig. 4F). For each reversal in observers' responses, the step size in the staircase was halved. Four independent staircases with randomly chosen initial amplitudes (within the range 0° – $19^\circ/\text{s}$) were interleaved in a block of trials. A single staircase was completed in 15–25 trials. A staircase was considered “converged” when it had encountered ten reversals in an observer's responses and the last eight reversals were used to calculate the threshold for perceiving backward motion. The minimum velocity of the target

Fig. 4 Spatial and temporal characteristics of the stimuli. In all experiments, a target dot moved horizontally. The velocity of the dot was modulated by a sine wave during its motion. (A) In Experiment 1a, the target dot was presented alone. (B) In Experiment 1b, in addition, a drifting grating was presented at various vertical eccentricities. The drift velocity of the grating matched the average velocity of the target dot. (C) In Experiment 2, a static grating at the same vertical eccentricity as the

dynamic one was added, but in the opposite half of the screen. In Experiment 3a (D), both gratings drifted in the same direction as the target dot, whereas in Experiment 3b (E), one of them drifted in the opposite direction. (F) Velocity profiles of the target dot and drifting gratings. (G) Close-up view of the space–time graphs for the horizontal position of the target dot for various magnitudes of velocity modulation. Position is given in terms of distance from the edge of the display

corresponding to this threshold amplitude was taken as the *reference-field effect*. For instance, if the staircase converged to 9°/s amplitude for sine modulation, it corresponded to the minimum target velocity of $(9 - 9) = 0^\circ/\text{s}$ (reference-field effect = $0^\circ/\text{s}$). This would mean that backward motion is *perceived* only when the target velocity goes below $0^\circ/\text{s}$ (veridical percept). On the other hand, if, for instance, the staircase converges to 6°/s, corresponding to the minimum target velocity of $3^\circ/\text{s}$, it would mean that as soon as the target velocity fell below $3^\circ/\text{s}$ (reference-field effect = $3^\circ/\text{s}$), backward motion would be *perceived* (illusory percept), although it might never spatiotopically move backward. For each vertical position of the target, each observer ran one block of trials (four staircases).

Experiment 1a was designed to determine quantitatively for each observer the ability to detect a reversal in the direction of motion of a simple stimulus as a function of eccentricity. Ideally, the thresholds for detecting backward motion should be zero or close to zero. However, since we used a yes/no task, bias might occur. In Experiment 1b, the reference field generated around the drifting grating should cause illusory percepts. To be more specific, the RFMF theory predicts that backward motion should be perceived even at positive minimum target velocities (field effects $> 0^\circ/\text{s}$). In order to show that percepts were not simply due to subtraction of a common motion component and that a motion-based reference frame is effective

only within a limited spatial region, we analyzed the field effects as a function of the grating-to-target distance.

Results and discussion

Figure 5A shows the baseline thresholds as a function of vertical eccentricity in Experiment 1a. There is no indication of perceptual biases at a particular eccentricity. A one-way repeated measures analysis of variance (ANOVA) confirmed that the vertical eccentricity had no significant effect [$F(2, 6) = 0.338, p = .726, \eta_p^2 = .101$] on the perceived direction of the target's motion. However, we found a small but significant bias toward negative reports—that is, to say “No, it did not move back” [mean across observers and vertical eccentricities = $-0.620^\circ/\text{s}$, one-sample t test: $t(11) = -4.018, p = .002$].

The presence of the drifting grating significantly changed the pattern of results. Figure 5B shows the reference-field effects (defined as the minimum speed at which the target was perceived to be moving backward; see the Stimuli and Procedures section above) as a function of the target–grating distance. Although we found a small but significant perceptual bias in Experiment 1a in the absence of the drifting grating, we did not take this bias into account for the results of Experiment 1b, since the bias was not eccentricity dependent. Factoring out these biases thus would only cause an overall upward shift, not affect the statistical analysis. The drop in the effect

Fig. 5 (A) In Experiment 1a, the minimum velocity of the target dot at which it appeared to move backward is taken as the motion reversal threshold and plotted against vertical eccentricity. (B) Motion reversal thresholds (referred to as the *reference-field effects*) in the presence of a

drifting grating are plotted against the target dot's distance from the grating. Each gray line represents a different observer, and black lines show the averages across observers. The dotted horizontal line in panel B represents the prediction of perfect vector decomposition

size with increasing target–grating distance, shown in Fig. 5B, is significant [$F(4, 12) = 13.550$, $p < .001$, $\eta_p^2 = .819$]. The results replicate the distance-dependent influence of motion-based reference frames on the perceived motion of nearby objects that has already been shown by several studies (DiVita & Rock, 1997; Hochberg & Fallon, 1976; Mori, 1979; Shum & Wolford, 1983). All of these studies lend support to the claim that common motion might serve as a reference frame, but its effectiveness is limited to a spatial region.

Experiment 2: No motion, no interaction

The RFMF theory predicts that reference-field interactions will occur only when there is motion. In this experiment, we tested this prediction by presenting a static grating in addition to the dynamic grating. The static grating was in the other half of the visual field than the dynamic grating. It has been shown that the presence of a stationary landmark (such as a fixation point, surrounding frame, etc.) substantially influences the perceived motion (e.g., Wallach et al., 1985). The static grating used in Experiment 2 provided an additional reference frame (along with the display borders) for motion computations; however, whether it could generate interactions with a reference field needed to be investigated. If the presence of the static grating modulated the strength of a motion-based reference field, we should see distance-dependent drops in the reference-field effect, as compared to the case in which only the dynamic grating was presented.

Method

Stimuli and procedures The stimuli and procedures were identical to those used in Experiment 1b, with the following

exceptions. In addition to the dynamic grating, which always moved in the same direction as the target, a second grating was presented at the same vertical eccentricity as the dynamic one, but in the opposite half of the screen (Fig. 4C). The second grating was stationary (drift velocity = 0°/s). Which one of the gratings was presented in the upper visual field was randomized across trials.

Results and discussion

Figure 6A shows the reference-field effects measured in the presence (Exp. 2) and the absence (replotted from Exp. 1b) of the static grating. Similar to that from Experiment 1b, a one-way repeated measures ANOVA on the reference-field effects in Experiment 2 revealed a significant effect of the distance between the target and the dynamic grating [$F(4, 12) = 18.355$, $p < .001$, $\eta_p^2 = .860$]. A two-way ANOVA with Target-to-Dynamic-Grating Distance and Experiment as the main factors showed, once again, a significant effect of distance [$F(4, 12) = 18.792$, $p < .001$, $\eta_p^2 = .834$]. However, the main effect of experiment did not reach significance [$F(1, 3) = 6.346$, $p = .086$, $\eta_p^2 = .679$]. More importantly, the interaction of the main factors was not significant, either [$F(4, 12) = 0.824$, $p = .535$, $\eta_p^2 = .215$]. These results suggest that static reference frames do not generate field-like spatial zones within which their influence is modulated by distance. The failure to obtain a significant main effect of experiment (i.e., the addition of a static grating) might have been due to a floor effect: The experiments were done in a normally illuminated room where screen edges and other potential spatial landmarks were visible. Adding another spatial landmark might not have been very effective.

One might also argue that the failure to find a significant interaction between distance and experiment here might have

Fig. 6 (A) Reference-field effects averaged across observers ($n = 4$) in Experiment 1b and Experiment 2 are plotted against the target dot's distance from the drifting grating. Square markers represent reference-field effects in Experiment 1b, whereas circular markers represent the results of Experiment 2. The horizontal line represents the predictions

been due to the floor effect. At around 9–12 deg of distance from the dynamic grating, the effect size is already very close to zero, and any potential drop in the effect size due to the presence of the static grating might be clouded. In order to address this issue, we took the three closest distances to the dynamic grating and repeated the statistical analysis. The main effect of distance was again significant [$F(2, 6) = 10.240, p = .012, \eta_p^2 = .773$], whereas the main effect of experiment was not [$F(1, 3) = 4.775, p = .117, \eta_p^2 = .614$]. The interaction of the main factors was, once again, not significant [$F(2, 6) = 0.358, p = .713, \eta_p^2 = .106$]. Moreover, within-observer differences between the two experiments did not show any trends whatsoever (see Fig. 6B). If a field was indeed associated with static references, we would see positive slopes—that is, increased effect size difference with distance.

Experiment 3: Interacting dynamic reference fields

In the first experiment, we replicated the basic finding that the perceived motion of a stimulus can be influenced by the motions of nearby objects, and that this effect spreads over space in a field-like manner. In the second experiment, we showed that the reference fields are generated only when motion is present. In ecological viewing conditions, a multitude of objects might move in various directions. According to the RFMF theory, in order to perceive sharp and clear forms of these objects, each object needs to be processed in a proper reference frame (determined by the local motion vectors). Furthermore, the selection of a certain reference frame is not done in an all-or-none manner. Instead, each reference frame has a reference field associated with it and exerts its effect within this spatiotemporally limited field. A question, then,

of perfect vector decomposition. Error bars represent $\pm SEMs$. (B) Change in effect size with the addition of a static grating. Each symbol represents a different observer. There is no discernable pattern in the changes of the effect sizes

arises: What if several reference fields come into close proximity with each other? The RFMF theory suggests that the reference fields would interact to reach an equilibrium in the retinotopic space. In this experiment, we tested this prediction.

Method

Stimuli and procedures The stimuli and procedures were identical to those used in Experiment 2, with the following exceptions. The static grating was replaced by a drifting grating having the same spatial characteristics as the dynamic grating in Experiments 1b and 2. In Experiment 3a, the drift velocities of the two gratings were identical and equal to the average velocity of the target (Fig. 4D). Since these gratings were identical in all respects, the target–grating distances of 1 and 12 deg and 3.75 and 9.25 deg were essentially the same. Therefore, we effectively had only three target–grating distances in Experiment 3a (1, 3.75, and 6.5 deg). In Experiment 3b, one of the gratings (primary) always drifted in the target's motion direction, while the other (secondary) drifted in the opposite direction, but with the same speed (Fig. 4E). As in Experiment 2, the target was presented at five different vertical eccentricities (the corresponding target primary distances were 1, 3.75, 6.5, 9.25, and 12 deg). Which one of the gratings was presented in the upper visual field was randomized across trials.

Results and discussion

The reference-field effects averaged across observers in Experiments 3a and 3b are plotted in Fig. 7. The primary x-axis represents the distance to the primary grating (which moved in the same direction as the target's average velocity), and the

Fig. 7 Results from Experiments 3a and 3b. Circles represent the results from Experiment 3a, whereas square symbols represent the data from Experiment 3b. The data at distances of 1 and 3.75 deg in Experiment 3a are replotted with open symbols at distances of 12 and 9.25 deg, respectively, to facilitate comparisons. The average baseline threshold obtained in Experiment 1a is also shown by a dotted horizontal line. The solid horizontal line represents the predicted effect when perfect vector decomposition occurs. Error bars represent $\pm$ SEMs ($n = 4$)

secondary x -axis represents the distance to the secondary grating. In the case of Experiment 3a, the distinction between the two gratings was void, since they both drifted in the same direction, and in order to facilitate visual comparison of the results of Experiments 3a and 3b, the data at distances of 1 and 3.75 deg are replotted at distances of 12 and 9.25 deg, respectively (Fig. 7, dashed line and open symbols). However, all of the following statistical analyses were carried on the reference-field effects at the three closest distances (i.e., 1, 3.75, and 6.5 deg). A one-way repeated measures ANOVA showed that the effect of distance is not significant [$F(2, 6) = 4.774, p = .057, \eta_p^2 = .614$]. When the second grating drifted in the opposite direction, the effect of distance became highly significant [$F(4, 12) = 27.265, p < .001, \eta_p^2 = .901$]. Moreover, a two-way repeated measures ANOVA with Distance to the Primary Grating and Drift Direction of the Secondary Grating as the main factors yielded significant main effects of both factors [distance, $F(2, 6) = 37.715, p < .001, \eta_p^2 = .921$; drift direction, $F(1, 3) = 24.517, p = .016, \eta_p^2 = .891$]. The interaction between the factors was also significant [$F(2, 6) = 7.450, p = .024, \eta_p^2 = .713$]. The significant main effect of drift direction of the second grating suggests that the reference fields generated by two gratings interact, and the nature of this interaction (constructive or destructive) depends on the direction of the motion within the second grating. The significant interaction of the main factors, as manifested by distance-dependent differences between the reference-field effects in Experiments 3a and 3b, also lends support to the reference-field hypothesis.

Experiments 3a and 3b allow us to draw conclusions about the interactions taking place. Figure 8A illustrates the reference fields generated by the drifting gratings. Assuming that the different fields add up linearly, when both gratings drift in the same direction, their respective reference fields should lead to accumulation, as is illustrated in Fig. 8B. However, when one of the gratings drifts in the opposite direction, the respective fields should not add up linearly. The “sign” of the reference field of the second grating is negative, and the corresponding fields are illustrated in Fig. 8C. If two fields were to simply add up, the illusory perception of backward motion should have decreased with distance, and perceptual responses to say “No, the target did not move backward” (although it physically did) should have taken over and become stronger as the target became closer to the second grating [Fig. 8C(i)]. The resultant reference-field effects, hence, should start at a positive value (indicating illusory percepts) and go down to negative values (indicating perceptual biases to say “no”), roughly linearly. However, our results did not confirm these predictions: A faster drop in the field effect was followed by veridical percepts (i.e., a zero effect size) as the distance between the target and the second grating became smaller. This effect is illustrated in Fig. 8C(ii). Therefore, the interactions of different reference fields cannot be explained by simple linear summation of the individual fields, as one would expect from receptive-field-based center-surround models. Rather, the asymmetry underlying the nonlinear effect needs to be understood in terms of the coherence between the motion of the target and those of the reference frames. As is illustrated in Fig. 8, the average motion of the target is in the same direction as one of the two gratings, but in the opposite direction from the other grating. As is depicted in Fig. 3, according to RFMF theory, motion-based grouping takes place first. Hence, the target is grouped with the grating drifting in the same direction, but not with the grating drifting in the opposite direction. Hence, the results can be incorporated into the theory as follows: A motion signal that is not grouped with the target can decrease the strength of the field, but it cannot revert it so as to cause a sign change in the effect. In other words, it can nullify the effect of a reference frame, but it cannot become the reference frame itself, since it is not grouped with the target. This can be experimentally tested by manipulations that will allow the target to be grouped with different stimuli. Whether a stimulus can revert a field should be directly dependent on whether or not observers report that stimulus to be grouped with the target.

General discussion

When we compute the motion of an object in everyday life, we generally use the static environment as a reference frame. However, the perceived motion of an object corresponds to its

Fig. 8 Illustration of the reference fields and their interactions. (A) Motion within each grating generates a reference field, which spreads over space and time and is stronger at the spatiotemporal position of the motion. Here only the fields generated by the motion vectors of the central part of each grating are shown, for clarity. (B) When both gratings drift in the same direction, their strengths add up. The results from Experiment 3a are consistent with this prediction. (C) When the gratings move in opposite directions, the signs of the fields generated by them will be different. If interactions of these two fields occur as the mathematical summation of their strengths, we should get a linear

decrease in the effect sizes in Experiment 3b (i). However, what we found more resembles graph (ii): The field generated by the grating that drifts in the same direction as the target gets truncated by the field of the grating drifting in the opposite direction. Simple summation of field strengths does not explain this interaction between reference fields. (D) The results of Experiments 1b, 2, 3a, and 3b are plotted on the same axis for comparison. The effect of adding another field with opposing forces can also be seen by comparing Experiments 1b and 3b. Error bars are removed to avoid clutter

motion with respect to a static reference frame only in special, simple cases. When we see a friend waving his or her hand on a moving bicycle, the hand undergoes a complex motion trajectory with respect to the static background. But, in fact, we perceive the hand moving vertically up and down, discounting the motion of the bicycle. Likewise, we see the wheels of the bicycle rotating around their axles, with the horizontal motion of the axles discounted. Hence, the circular motion of the wheel is perceived with respect to the translatory motion of the bicycle. The inadequacy of using the static environment as the single reference frame and the roles of moving reference frames have been systematically investigated by Johansson (1950, 1958, 1973, 1976, 1986) and many others (Duncker, 1929; Gogel, 1974; Hochberg & Fallon, 1976; Mori, 1979; Wallach et al., 1985; Wallach, O’Leary, & McMahon, 1982).

Johansson claimed that the rotary motion of any point on a wheel can be deduced by perceptual subtraction of the translatory motion vector, common to both the hub and the wheel, from its “real” cycloidal motion (Johansson, 1950, 1976). The

theory of perceptual vector analysis can explain rapid perception of highly complex motion patterns, such as biological motion displays, by a hierarchy of moving reference frames, thus simplifying the motions of the knees and feet as the simple harmonic motion of a pendulum (Johansson, 1973, 1976). The gist of Johansson’s theory lies in the extraction of common and relative motion components. However, many studies have demonstrated that the extraction of common motion is not always perfect (DiVita & Rock, 1997; Gogel, 1974; Hochberg & McAlister, 1953; Johansson, 1974; Mori, 1979, 1984; Wallach, 1959). More importantly, in mathematical terms, vector decomposition is an ill-posed problem and needs additional information to be solved. Several constraints to limit the number of solutions have been introduced (Börjesson & von Hofsten, 1972; Cutting & Proffitt, 1982; Gershman et al., 2013; Gogel, 1974; Gogel & Koslow, 1972; Hochberg & McAlister, 1953; Proffitt et al., 1979; Restle, 1979). In short, these constraints provide heuristics to explain *why* the human visual system selects a particular solution. In the

present study, we have taken the alternative perspective, and looked at *how* a particular solution emerges through interactions between motion vectors.

Previously, we have shown that the perceived motion of a target stimulus can be influenced by nearby motion of another object, and that this object need not be surrounding the target stimulus, as in the induced-motion paradigm (Agaoglu et al., 2015; Noory et al., 2015). In this study, we started off by replicating our previous findings that each local motion vector has a reference field associated with it, and this is manifested by increased illusory percepts of backward motion with decreasing distance to this moving reference frame. We then sought to determine whether these field-like effects of motion-based reference frames can also be extended to stationary landmarks. We presented a highly salient stationary grating along with a drifting one to examine whether the effect of the latter would be in any way modulated by the presence of the former. Although there was a consistent trend of reduced effect sizes in the presence of a stationary grating, this reduction did not reach significance. More importantly, we did not find any significant interaction between distance and the presence/absence of the stationary grating, suggesting that reference fields interact only when there is motion.

In order to investigate whether and how different reference fields interact with each other, we presented two drifting gratings at various distances from the target square. In different experiments, we manipulated the drift direction of the secondary grating while the primary grating always drifted in the same direction as the average target velocity. We found that when both gratings drifted in the same direction, their effects combined and strengthened the illusory backward motion percepts. When the secondary grating drifted in a direction opposite to the direction of both the target and the primary grating, we found a significant drop in the illusory percept—that is, in the reference-field effect. These drops, however, cannot be explained by linear summation of the reference fields. Taken together, these findings suggest that reference fields do interact, and the way that their effects combine is nonlinear and depends on how the motion vectors are grouped.

These results clarify the details of the interactions posited in the RFMF theory. The RFMF theory was developed to explain how the visual system computes the attributes of stimuli under ecological conditions—that is, when the observer and the objects in the environment are in motion. Due to visible persistence, moving objects should appear extensively smeared, but under normal viewing conditions they do not (Hammett, 1997; Ramachandran et al., 1974). In addition, a moving object activates a retinotopically anchored receptive field only briefly, which does not allow sufficient time for computation of the stimulus attributes. As a result, one would expect moving objects to have a featureless “ghost-like” appearance (Ögmen, 2007; Ögmen & Herzog, 2010); however, under normal viewing conditions, moving objects appear

sharp and clear. RFMF suggests that the visual system avoids these problems by computing the attributes of moving objects, not based on retinotopic coordinates, but instead according to a reference frame that moves with the object. To achieve this, as is depicted in Fig. 3, a first stage of processing groups motion information and extracts reference frames that are used to compute the attributes of the moving objects. The use of nonretinotopic, motion-based reference frames has been supported by several studies (Agaoglu et al., 2012; Boi et al., 2009; Hisakata, Terao, & Murakami, 2013; Kawabe, 2008; Nishida, Watanabe, Kuriki, & Tokimoto, 2007; Ögmen, Otto, & Herzog, 2006; Yamada & Kawabe, 2013). On the basis of our results, we can summarize the reference-frame rules as follows: First, individual motion vectors are grouped according to their similarities (law of common fate). Each individual motion vector creates a field whose effect decays with distance. The fields of vectors that are grouped together reinforce each other. The field of a vector can weaken the effects of the fields of vectors that form a different group; however, without being grouped with these motion vectors, it cannot revert their effects.

Tynan and Sekuler (1975) argued that theories of motion perception based on reference frames have two limitations: (i) They do not state unambiguously which objects will form the reference frame in a complex scene, and (ii) they do not account for the dependence of suppressive interactions on stimulus velocities. Regarding the first point, we have recently implemented RFMF theory as a computational model that specifies mechanistically how reference frames are determined (Clarke, Ögmen, Herzog, 2015). Regarding the second point, additional experiments will be needed to determine whether the dependence of suppression on stimulus velocities arises from the center-surround organization of basic motion detectors, motion grouping, or other factors. RFMF theory takes into account basic motion detectors' receptive-field properties at the retinotopic level, and the grouping operations at the level of retinotopic representations are mapped into nonretinotopic representations (Fig. 3).

Other studies have also focused more on how, rather than why, a particular solution emerges as a result of vector decomposition. Wallach and colleagues (1985) rejected the idea of perceptual vector analysis and interpreted the percepts that supposedly result from imperfect vector decomposition as a consequence of “process combination” (see also Johansson, 1985, and Wallach & Becklen, 1985). They claimed that there is no need for the extraction of common motion, and that the perceived motion patterns are nothing but incidental results of the sensory apparatus. In other words, component motions activate different kinds of motion processes, and sometimes these processes can combine, resulting in motion percepts that deviate from what vector decomposition theory would predict. Take, for example, the two-dot display shown in Fig. 1E. According to process combination theory, the individual motions

of the dots, the displacements of the group as a whole, and motion within the group activate different motion sensors. When a stationary landmark (a fixation point or a rectangular frame) is presented along with the two-dot display shown in Fig. 1E, observers mostly perceive the absolute motion paths—that is, one dot oscillating vertically and the other horizontally—because motions with respect to the stationary reference are enhanced and grouping of the dots is weakened (Wallach et al., 1985), which is also in line with RFMF theory.

Although it is indirect, some partial evidence supports this position. The existence of motion sensors that are tuned to various types of moving patterns has been shown: The dorsal portion of the medial superior temporal area is known to be sensitive to global motion patterns, whereas the ventrolateral portion is more sensitive to within-configuration or object motions in the scene (Duffy & Wurtz, 1995; Eifuku & Wurtz, 1998), and some medial temporal (area V5) neurons are responsive to the global motion of a plaid, whereas others respond to the motion of its individual sinusoidal components (Rust, Mante, Simoncelli, & Movshon, 2006). If a given type of motion sensor is activated more than others, perceived motion can be mostly determined by the outcome of this process as a result of process combination. For instance, during steady fixation, relative motion determines the percepts at slow speeds, whereas absolute motion takes over at high speeds (Baker & Braddick, 1982; Snowden, 1992); this might be due to different levels of activation of the corresponding motion sensors at different speeds. However, in contrast to this perspective built on hardwired motion mechanisms, we suggest that the formation of reference frames is a dynamic process that is *adaptable* in real time. The rationale for this is that under ecological viewing conditions, trajectories can be arbitrarily complex, and it is not possible for the visual system to build hardwired motion sensors for all possible trajectories. Hence, real-time grouping operations and field interactions between the activities generated by a small set of canonical motion mechanisms can represent a *neural-network state* solution that prevails under the given specific stimulus conditions.

A constraint that can play an important role in disambiguating vector decomposition is prior knowledge among the observers. For example, observers can readily recognize biological motion when the stimulus is presented in its correct orientation, but they fail to do so when it is inverted (Pavlova & Sokolov, 2000). This suggests that templates from memory can also help resolve ambiguity and prime the grouping process. Part of the reason why stimuli like the simple dots shown in Fig. 1 generate multiple percepts may be due to the fact that they are not rich enough to engage specific memory patterns, and thus form unambiguous groups. Hence, in general, learned figural configurations need also to be taken into account. Recently, Grossberg, Léveillé, and Versace (2011) proposed a neural-network model to explain how

vector decomposition might occur by taking into account figural factors. According to their model, figure-ground separation and inhibition between neural populations, which represent motion at different depths, play the critical role; near-to-far inhibition and the resultant peak-shift in the population activity leads to vector decomposition. Consistent with this model, it has been shown that surface decomposition leads to velocity decomposition (Watanabe, 1997). An important, but rather implicit, assumption of the model, which has not been tested formally, is that common (or coherent) motion is perceived to be at a different depth or scale than relative (or incoherent) motion: The former resides at a nearer depth or a larger scale. Moreover, as was acknowledged by the authors, their model in its current form cannot account for induced motion, in which a stationary object is seen to be moving in the opposite direction from a surrounding (or neighboring) moving object. This is partly due to a claim that vector decomposition and induced motion arise from different neural mechanisms (DiVita & Rock, 1997). For instance, induced motion is not perceived after the motion threshold (up to 3°/s) for the frame is reached. Also in induced motion, the inducer is perceived to be either moving at a lower speed or not moving at all, whereas in vector decomposition stimuli, the common and relative parts are perceived simultaneously. However, we obtained strong illusory perception of backward motion in the presence of a grating with a drift velocity of 9°/s. Therefore, one can argue that the vector analysis effect and induced motion stem from the same neural mechanism, and that the reference-field effect reported here constitutes a special case of induced motion. Last but not the least, figure-ground segregation via perceptual grouping operations requires at least some form of computation. Grossberg and colleagues' model extracts the form of the group of two dots via illusory contours; only then can common motion be calculated. It would be interesting to see how their model would respond to a “formless” motion stimulus. The RFMF theory predicts that retinotopic motion (without any form) is sufficient to generate a reference field.

As we have mentioned before, theories of motion perception based on center-surround antagonisms within receptive fields have proposed inhibitory interactions between motions across space (e.g., Kim & Wilson, 1997; Murakami & Shimojo, 1996; Nawrot & Sekuler, 1990; Tadin et al., 2003; Tynan & Sekuler, 1975). Field-like effects of motion-based reference frames and inhibitory interactions can be predicted from these theories to some extent. The RFMF theory differs from these approaches by putting perceptual-grouping operations as an essential component of reference-frame extraction; in this context, the nonlinearity in the antagonistic interactions in Experiment 3b can be explained as an effect of perceptual grouping. Moreover, most theories of vision, including the aforementioned approaches to motion perception, are based on retinotopic representations, although mounting evidence

is showing that perception is highly nonretinotopic (Agaoglu et al., 2012; Boi et al., 2009; Hisakata et al., 2013; Kawabe, 2008; Nishida et al., 2007; Ögmen et al., 2006; Shimozaki, Eckstein, & Thomas, 1999; Yamada & Kawabe, 2013). Hence, a fundamental gap exists between retinotopic theories and the nonretinotopic percepts that these theories attempt to explain. The RFMF theory offers a unified solution by constructing nonretinotopic representations according to motion-based reference frames. For instance, anorthoscopic perception (i.e., perceiving an object as a whole when it moves behind a narrow slit) had no viable explanation based on retinotopic theories, whereas it can readily be explained by the nonretinotopic representations of RFMF (see Ögmen, 2007). We recently tested this theory and showed how a motion-based reference frame can construct space and enable form perception (Agaoglu et al., 2012; Aydın et al., 2008, 2009; see also Nishida, 2004). Therefore, all of the theories employing center-surround antagonism that are listed above are tailored specifically for motion, whereas the RFMF theory accounts for all attributes of the stimuli. Hence, the selection and construction of motion-based reference frames are needed not only to perceive motion, but also to perceive the complete stimulus. Therefore, RFMF theory can make predictions on perceived form. For instance, in slit viewing, objects appear to be compressed along the axis of motion. The RFMF theory predicts that this apparent compression results from perceived speed differences between the different parts of the moving object. These predictions have also been tested formally, and the results lend further support to the RFMF theory (Aydın et al., 2008). Moreover, why and how attention is allocated to moving stimuli (Boi et al., 2009; Boi, Vergeer, Ögmen, & Herzog, 2011) and why masking is retinotopic but form perception can escape masking when a motion is predictable are other predictions of the RFMF theory that we have recently tested experimentally (Noory, Herzog, & Ögmen, *in press*), and we have demonstrated that RFMF can be implemented computationally to explain data at the quantitative level (Clarke et al., 2015).

Author note M.H.H. is supported by the Swiss National Science Foundation (SNF) project “Basics of Visual Processing: From Retinotopic Encoding to Non-Retinotopic Representations.” We thank the reviewers for helpful comments and suggestions.

References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2, 284–299.
- Agaoglu, M. N., Herzog, M. H., & Ögmen, H. (2012). Non-retinotopic feature processing in the absence of retinotopic spatial layout and the construction of perceptual space from motion. *Vision Research*, 71, 10–17. doi:10.1016/j.visres.2012.08.009
- Agaoglu, M. N., Herzog, M., & Ögmen, H. (2015). The effective reference frame in perceptual judgments of motion direction. *Vision Research*, 107, 101–112. doi:10.1016/j.visres.2014.12.009
- Albright, T. D., & Stoner, G. R. (1995). Visual motion perception. *Proceedings of the National Academy of Sciences*, 92, 2433–2440.
- Aydın, M., Herzog, M. H., & Ögmen, H. (2008). Perceived speed differences explain apparent compression in slit viewing. *Vision Research*, 48, 1603–1612. doi:10.1016/j.visres.2008.04.020
- Aydın, M., Herzog, M. H., & Ögmen, H. (2009). Shape distortions and Gestalt grouping in anorthoscopic perception. *Journal of Vision*, 9(3), 8:1–8. doi:10.1167/9.3.8
- Baker, C. J., & Braddick, O. (1982). Does segregation of differently moving areas depend on relative or absolute displacement? *Vision Research*, 22, 851–856.
- Boi, M., Ögmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision*, 9(13), 5:1–11. doi:10.1167/9.13.5
- Boi, M., Vergeer, M., Ögmen, H., & Herzog, M. H. (2011). Nonretinotopic exogenous attention. *Current Biology*, 21, 1732–1737. doi:10.1016/j.cub.2011.08.059
- Börjesson, E., & von Hofsten, C. (1972). Spatial determinants of depth perception in two-dot motion patterns. *Perception & Psychophysics*, 11, 263–268.
- Clarke, A. M., Ögmen, H., & Herzog, M. H. (2015). *A computational model for reference-frame synthesis with application to motion perception*. Manuscript submitted for publication.
- Clarke, A. M., Repnow, M., Ögmen, H., & Herzog, M. H. (2012). Does spatio-temporal filtering account for nonretinotopic motion perception? Comment on Pooresmaeili, Cicchini, Morrone, and Burr (2012). *Journal of Vision*, 13(10), 19. doi:10.1167/13.10.19
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, 27, 183–228. doi:10.3758/BF03204258
- Cutting, J. E., & Proffitt, D. R. (1982). The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology*, 14, 211–246.
- DiVita, J. C., & Rock, I. (1997). A belongingness principle of motion perception. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1343–1352. doi:10.1037/0096-1523.23.5.1343
- Duffy, C. J., & Wurtz, R. (1995). Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *Journal of Neuroscience*, 15, 5192–5208.
- Duncker, K. (1929). Über induzierte Bewegung. *Psychologische Forschung*, 12, 180–259. doi:10.1007/BF02409210
- Eifuku, S., & Wurtz, R. (1998). Response to motion in extrastriate area MSTl: Center-surround interactions. *Journal of Neurophysiology*, 282–296.
- Ellis, W. D. (1938). *A source book of Gestalt psychology*. London: Paul, Trench, & Trübner.
- Gershman, S., Jäkel, F., & Tenenbaum, J. (2013). Bayesian vector analysis and the perception of hierarchical motion. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Cooperative minds: Social interaction and group dynamics. Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 489–494). Austin, TX: Cognitive Science Society.
- Gogel, W. C. (1974). Relative motion and the adjacency principle. *Quarterly Journal of Experimental Psychology*, 26, 425–437. doi:10.1080/14640747408400432
- Gogel, W. C., & Koslow, M. (1972). The adjacency principle and induced movement. *Perception & Psychophysics*, 11, 309–314. doi:10.3758/BF03210385
- Grossberg, S., Léveillé, J., & Versace, M. (2011). How do object reference frames and motion vector decomposition emerge in laminar cortical circuits? *Attention, Perception, & Psychophysics*, 73, 1147–1170. doi:10.3758/s13414-011-0095-9

- Haber, R., & Standing, L. (1970). Direct estimates of the apparent duration of a flash. *Canadian Journal of Psychology*, 24, 216–229.
- Hammett, S. T. (1997). Motion blur and motion sharpening in the human visual system. *Vision Research*, 37, 2505–2510. doi:10.1016/S0042-6989(97)00059-X
- Herzog, M. H., Hermens, F., & Öğmen, H. (2014). Invisibility and interpretation. *Frontiers in Psychology*, 5, 975. doi:10.3389/fpsyg.2014.00975
- Herzog, M. H., & Öğmen, H. (2014). Apparent motion and reference frames. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization*. Oxford, UK: Oxford University Press.
- Hisakata, R., Terao, M., & Murakami, I. (2013). Illusory position shift induced by motion within a moving envelope during smooth-pursuit eye movements. *Journal of Vision*, 13(12), 21:1–12. doi:10.1167/13.12.21
- Hochberg, J., & Fallon, P. (1976). Perceptual analysis of moving patterns. *Science*, 194, 1081–1083. doi:10.1126/science.982065
- Hochberg, J., & McAlister, E. (1953). A quantitative approach, to figural “goodness”. *Journal of Experimental Psychology*, 46, 361–364. doi:10.1037/h0055809
- Johansson, G. (1950). *Configurations in event perception: An experimental study*. Stockholm, Sweden: Almqvist & Wiksell.
- Johansson, G. (1958). Rigidity, stability, and motion in perceptual space. *Acta Psychologica*, 14, 359–370. doi:10.1016/0001-6918(58)90028-3
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14, 201–211. doi:10.3758/BF03212378
- Johansson, G. (1974). Vector analysis in visual perception of rolling motion. *Psychologische Forschung*, 36, 311–319. doi:10.1007/BF00424568
- Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research*, 38, 379–393. doi:10.1007/BF00309043
- Johansson, G. (1985). Vector analysis and process combinations in motion perception: A reply to Wallach, Becklen, and Nitzberg (1985). *Journal of Experimental Psychology: Human Perception and Performance*, 11, 367–371. doi:10.1037/0096-1523.11.3.367
- Johansson, G. (1986). Relational invariance and visual space perception: On perceptual vector analysis of the optic flow. *Acta Psychologica*, 63, 89–101. doi:10.1016/0001-6918(86)90057-0
- Johansson, G., & Jansson, G. (1968). Perceived rotary motion from changes in a straight line. *Perception & Psychophysics*, 4, 165–170. doi:10.3758/BF03210461
- Kawabe, T. (2008). Spatiotemporal feature attribution for the perception of visual size. *Journal of Vision*, 8(8), 7:1–9. doi:10.1167/8.8.7
- Kim, J., & Wilson, H. R. (1997). Motion integration over space: Interaction of the center and surround motion. *Vision Research*, 37, 991–1005. doi:10.1016/S0042-6989(96)00254-4
- Koffka, K. (1935). *Principle of Gestalt psychology*. New York, NY: Harcourt.
- 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, 2331–2370.
- Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society B*, 211, 151–180. doi:10.1098/rspb.1981.0001
- Mori, T. (1979). Relative locations among moving spots and visual vector analysis. *Perceptual and Motor Skills*, 48, 587–592.
- Mori, T. (1984). Change of a frame of reference with velocity in visual motion perception. *Perception & Psychophysics*, 35, 515–518. doi:10.3758/BF03205947
- Murakami, I., & Shimojo, S. (1996). Assimilation-type and contrast-type bias of motion induced by the surround in a random-dot display: Evidence for center-surround antagonism. *Vision Research*, 36, 3629–3639.
- Nawrot, M., & Sekuler, R. (1990). Assimilation and contrast in motion perception: Explorations in cooperativity. *Vision Research*, 30, 1439–1451. doi:10.1016/0042-6989(90)90025-G
- Nishida, S. (2004). Motion-based analysis of spatial patterns by the human visual system. *Current Biology*, 14, 830–839.
- Nishida, S., Watanabe, J., Kuriki, I., & Tokimoto, T. (2007). Human visual system integrates color signals along a motion trajectory. *Current Biology*, 17, 366–372. doi:10.1016/j.cub.2006.12.041
- Noory, B., Herzog, M. H., & Öğmen, H. (in press). Retinotopy of visual masking and non-retinotopic perception during masking. *Attention, Perception, & Psychophysics*. doi:10.3758/s13414-015-0844-2
- Noory, B., Herzog, M. H., & Öğmen, H. (2013). Temporal characteristics of non-retinotopic reference frames in human vision. *Journal of Vision*, 13(15), 13:1–13. doi:10.1167/13.15.13
- Noory, B., Herzog, M. H., & Öğmen, H. (2015). *The non-retinotopic reference-frame field effect in human vision*. Manuscript submitted for publication.
- Öğmen, H. (2007). A theory of moving form perception: Synergy between masking, perceptual grouping, and motion computation in retinotopic and non-retinotopic representations. *Advances in Cognitive Psychology*, 3, 67–84. doi:10.2478/v10053-008-0015-2
- Öğmen, H., & Herzog, M. H. (2010). The geometry of visual perception: Retinotopic and nonretinotopic representations in the human visual system. *Proceedings of the IEEE*, 98, 479–492. doi:10.1109/JPROC.2009.2039028
- Öğmen, H., Herzog, M., & Noory, B. (2013). Reference-Frame Metric Field (RFMF) theory of non-retinotopic visual perception. *Perception*, 42(ECVP Abstract Suppl.), 160. Retrieved from www.perceptionweb.com/abstract.cgi?id=v130065
- Öğmen, H., Otto, T. U., & Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research*, 46, 3234–3242. doi:10.1016/j.visres.2006.04.007
- Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics*, 62, 889–899. doi:10.3758/BF03212075
- Pikler, J. (1917). *Sinnesphysiologische Untersuchungen*. Leipzig, Germany: Barth.
- Proffitt, D. R., Cutting, J. E., & Stier, D. M. (1979). Perception of wheel-generated motions. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 289–302. doi:10.1037/0096-1523.5.2.289
- Ramachandran, V. S., Rao, V. M., & Vidyasagar, T. R. (1974). Sharpness constancy during movement perception (short note). *Perception*, 3, 97–98.
- Restle, F. (1979). Coding theory of the perception of motion configurations. *Psychological Review*, 86, 1–24. doi:10.1037/0033-295X.86.1.1
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025. doi:10.1038/14819
- Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, 9, 1421–1431. doi:10.1038/nn1786
- Shimozaki, S. S., Eckstein, M., & Thomas, J. P. (1999). The maintenance of apparent luminance of an object. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1433–1453. doi:10.1037/0096-1523.25.5.1433
- Shum, K. H., & Wolford, G. L. (1983). A quantitative study of perceptual vector analysis. *Perception & Psychophysics*, 34, 17–24. doi:10.3758/BF03205891
- Snowden, R. J. (1992). Sensitivity to relative and absolute motion. *Perception*, 21, 563–568.
- Swanston, M., Wade, N., & Day, R. (1987). The representation of uni-form motion in vision. *Perception*, 16, 143–159.

- Tadin, D., Lappin, J. S., Gilroy, L. A., & Blake, R. (2003). Perceptual consequences of centre-surround antagonism in visual motion processing. *Nature*, 424, 312–315. doi:[10.1038/nature01800](https://doi.org/10.1038/nature01800)
- Ternus, J. (1926). Experimentelle Untersuchungen über Phänomenale Identität. *Psychologische Forschung*, 7, 81–136.
- Thunell, E., Herzog, M. H., & Ögmen, H. (2015). *Putting low-level vision into global context: Can we reduce vision to basic circuits?* Manuscript submitted for publication.
- Tynan, P., & Sekuler, R. (1975). Simultaneous motion contrast: Velocity, sensitivity and depth response. *Vision Research*, 15, 1231–1238. doi:[10.1016/0042-6989\(75\)90167-4](https://doi.org/10.1016/0042-6989(75)90167-4)
- Wade, N., & Swanson, M. (1987). The representation of nonuniform motion in vision. *Perception*, 16, 555–571.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012a). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138, 1172–1217. doi:[10.1037/a0029333](https://doi.org/10.1037/a0029333)
- Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J., van der Helm, P., & van Leeuwen, C. (2012b). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, 138, 1218–1252.
- Wallach, H. (1959). The perception of motion. *Scientific American*, 201, 56–60.
- Wallach, H., & Becklen, R. (1985). Response to Gunnar Johansson's critical commentary. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 372–373. doi:[10.1037/0096-1523.11.3.372](https://doi.org/10.1037/0096-1523.11.3.372)
- Wallach, H., Becklen, R., & Nitzberg, D. (1985). Vector analysis and process combination in motion perception. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 93–102. doi:[10.1037/0096-1523.11.1.93](https://doi.org/10.1037/0096-1523.11.1.93)
- Wallach, H., O'Leary, A., & McMahon, M. L. (1982). Three stimuli for visual motion perception compared. *Perception & Psychophysics*, 32, 1–6. doi:[10.3758/BF03204861](https://doi.org/10.3758/BF03204861)
- Watanabe, T. (1997). Velocity decomposition and surface decomposition—Reciprocal interactions between motion and form processing. *Vision Research*, 37, 2879–2889. doi:[10.1016/S0042-6989\(97\)00101-6](https://doi.org/10.1016/S0042-6989(97)00101-6)
- Watson, A. B., & Ahumada, A. J., Jr. (1985). Model of human visual sensing. *Journal of the Optical Society of America A*, 2, 322–341.
- Yamada, Y., & Kawabe, T. (2013). Localizing non-retinotopically moving objects. *PLoS ONE*, 8, e53815. doi:[10.1371/journal.pone.0053815](https://doi.org/10.1371/journal.pone.0053815)