Apparent motion and reference frames
Michael H. Herzog¹ and Haluk Öğmen²

To appear in:
Oxford Handbook of Perceptual Organization
Oxford University Press
Edited by Johan Wagemans

¹Brain Mind Institute
École Polytechnique Fédérale de Lausanne
²Center for Neuro-Engineering and Cognitive Science
Department of Electrical & Computer Engineering
University of Houston

Abstract

This article presents a selective overview of motion perception starting with its early philosophical underpinnings. Its role in Gestalt psychology is highlighted including the discovery of the relativity of motion perception and form-motion interactions. The use of reference frames in the computation of motion is illustrated with examples, leading to its current implications on non-retinotopic processing.

Keywords: Apparent motion, motion perception, induced motion, relativity, reference frames, non-retinotopic processing.

1. History of apparent motion and its role in Gestalt psychology

1.1. Mathematical foundations of space and time, Zeno’s paradoxes and the implied psychological theory

By definition, motion is change of position over time. To understand motion from a psychological perspective, one needs to appeal to the concepts whereby space and time are defined from the perspective of physics (to express the stimulus) and from the perspective of psychology (to express the percept). Around 450 B.C., Zeno studied how motion can be expressed using the concepts of space and time available at this time (Kolers, 1972). Zeno’s analysis of physical motion led him to paradoxes that he could solve by suggesting that motion is a purely psychological construct. In one of these paradoxes, Achilles is trying to catch up Tortoise in a race where Tortoise starts with an initial advantage. Zeno argues that Achilles will never be able to catch up Tortoise because by the time Achilles reaches Tortoise’s starting point, Tortoise will have advanced to a new position; by the time Achilles reaches this new position, Tortoise will be yet at another position further down the road, and so on… Zeno thought that even if Achilles moves faster than Tortoise and reduces his distance at every iteration, he will still have
to do this infinitely many times. Lacking the concept of infinity and convergent series, he concluded that Achilles will never be able to catch Tortoise. A similar paradox arises if one wants to move from point A to point B. Zeno reasoned that infinitely many points need to be crossed and that one can never move between two points. When time is conceived as a continuous variable composed of infinitely short (i.e. duration-less) instants, one cannot be in motion because, by definition, the instant has no duration to allow change in position. If motion is not physically possible, what then explains our percepts of moving objects? Zeno thought that objects exist at different locations at different time instants. These percepts are stored in memory and compared over time. When a disparity in spatial position is detected, we create an illusion of motion to resolve this disparity. Progress in mathematics (the development of the concept of convergent series) removed the conceptual barriers in expressing motion as a physical stimulus. Armed with this new mathematics, naïve realistic approaches focused on how this real motion can be perceived as a veridical, as opposed to an illusory percept. Nevertheless, psychological implications of Zeno’s analysis have been enduring.

1.2. Exner’s and Wertheimer’s contributions, types of apparent motion, and Korte’s laws
About 2500 years later, an important advance occurred when Exner (1875) created a stimulus consisting of two brief flashes presented at two spatially neighboring locations. With proper selection of timing and separation parameters, this stimulus generated the perception of motion, the first flash appearing to move smoothly to the location of the second flash. Since there was no stimulation of the intermediate points between the two flashes, this was indeed an illusion created by the perceptual system. More generally, Exner found that when the inter-stimulus interval (ISI) between the flashes was 10 ms or shorter, the two flashes were perceived as simultaneous; subjects could not report reliably their temporal order. When ISI increased, the perception was that of a single object moving from one position to the other. At longer ISIs, the stimuli appeared as two temporally successive flashes without the perception of motion. The finding that the perception of motion occurred at ISIs at which the temporal order of stimuli cannot be resolved led Exner to reject Zeno’s memory explanation. Since the temporal order of the two stimuli cannot be determined, the contents of memory should appear simultaneous and no motion should be perceived. Hence, Exner defended the view that motion is not an indirect property inferred from the analyses of objects over time, but instead it is a basic dimension of perception.

The experimental technique developed by Exner was essential in Max Wertheimer’s influential study that led to the development of Gestalt psychology (Wertheimer, 1912; for a review of the development of Gestalt psychology, see Wagemans, this volume). Using a borrowed tachistoscope and with Wolfgang Köhler and Kurt Koffka as his subjects, Wertheimer extended Exner’s study by creating a richer and more nuanced phenomenology. Exner’s three stages (simultaneity, motion, succession) were refined further by describing different types of perceived motion: one type of perceived motion was smooth movement of the object as described by Exner. This was called beta motion. A second type is partial movement, i.e., the object appears to move up to a certain point along the trajectory between the flashes, disappears, and reappears in movement again at a further point along the trajectory. Finally, a third type of movement, called the phi motion, corresponded to the percept of movement without any specific form, i.e., “figureless movement”. Wertheimer used phi motion to argue
that the perception of motion does not emerge from the comparison of objects in memory but it is a fundamental dimension of perception in its own, separate from form perception.

In terms of terminology, the perception of motion generated by two flashes is called apparent motion. Phi and beta motions are sub-types of apparent motion. They are distinguished from real motion, which refers to the perception of motion generated by a smoothly moving object. Following Wertheimer's study, Gestalt psychologists Korte and Neuhaus explored further the effect of various stimulus parameters leading to the so-called “Korte’s laws” (Korte, 1915; Neuhaus, 1930). These “laws” can rather be viewed as rules of thumb since the relationship of the percept with the parameters is rather complex (e.g., Kolers, 1972; Gepshtein & Kubovy 2007). In short, Korte’s laws state that to obtain the percept of apparent motion between flashes: 1) larger separations require higher intensities, 2) slower presentation rates require higher intensities, and 3) larger separations require slower presentation rates.

Since this early work, there have been a large number of studies investigating systematically the dependence of motion perception on a broader range of stimulus parameters. Around 1980s, research focus shifted from explaining the complex phenomenology of motion to the more basic question of how we detect motion. Several computational models have been proposed and eventually united under a broad umbrella. In the next section, we briefly review these models after which we will return to the main theme of our chapter, viz., phenomenal and organizational aspects of motion.

2. Computational basis of motion detection

2.1. Motion detection as orientation detection in space-time
As shown in Fig. 1A, the real (continuous) motion of an object with a constant speed can be described by an oriented line in a space-time diagram. An apparent motion stimulus is a sampled version of this stimulus consisting of two (or more) discrete points on the pathway (Fig. 1B). Motion detection mechanisms have been described as filters tuned to orientation in space-time. Among the earliest models, the Barlow-Levick model (Barlow & Levick, 1965) takes its input from one point in space, delays it, and compares it (with Boolean “AND” operation) to the input from another point in space. The Hassenstein-Reichardt correlation model (Hassenstein & Reichardt, 1956) works on a similar principle but the comparison is carried out by the correlation integral (Fig. 1C). Since these models sample space at two discrete spatial and temporal positions, they respond to apparent and real motion in the same way. The more elaborate

---

1 Note that the terms apparent/real motion may refer to the stimulus or to the percept generated by the stimulus depending on context. Stroboscopic motion and sampled motion are synonymous terms for apparent motion; the former derived from the equipment used to generate it (stroboscope), while the latter to highlight its relation to real motion (see Section 2.1).

2 Several demos can be found in the movies page. We have also included the source Powerpoint files. We encourage the reader to change various parameters (spatial separation between stimuli, shapes, geometrical organization) and experiment. As mentioned in the text, time parameters are important in determining the percept. An easy way to modify time parameters is to go to “Transition” tab of Powerpoint and change the time parameter under “Advance Slide”. The modified file can be saved as a movie and played as a movie file.
versions of these models include denser sampling to build a space-time receptive field as shown in Fig. 1D. These spatio-temporal models have been further extended by introducing nonlinearities at early stages so that they can respond to second-order stimuli (i.e., defined by stimulus dimensions other than luminance, such as texture). Finally, a third order motion system has been proposed that requires attention (for review, see Lu & Sperling, 2001). Salient features are detected and tracked over time. One implication of spatio-temporally localized receptive fields is that each motion detecting neuron “views” a small part of the space via its receptive field which acts as an “aperture”. When a uniform surface or edge moves across the viewing aperture, only the motion component perpendicular to the edge can be measured by a local motion detector, a problem known as the aperture problem (for a review, see Bruno & Bertamini, this volume). The solution of the aperture problem requires integration of motion signals across space. The motion integration problem will be discussed in the following sections within a broader context, viz., even when each local measurement is accurate.

Figure 1. A. Trajectory of a stimulus moving with a constant speed can be described as an oriented line in a space-time diagram. B. Apparent motion stimulus is a sampled version of continuous motion. C. A motion detector samples space-time Delay Compare space time A B C D
the input at two spatial locations and carries out a delay-and-compare operation. D. The denser sampling in space-time yields an oriented receptive field for the motion detector. This detector will become maximally active when the space-time orientation of the motion stimulus matches the orientation of its receptive field.

2.2. Is motion an independent perceptual dimension?

Given this background, we can now return to one of the original questions about motion perception: is it derived from object comparisons over time through memory or is it a fundamental dimension of perception? At a first glance, all models discussed above involve memory (e.g., delay or temporal filtering operations) and carry out comparisons (e.g., AND gate or correlation). However, first-order and second-order models compare relatively raw inputs without prior computation of form. As such, they constitute models that represent motion as an independent dimension. The third order motion system, however, identifies and tracks features; this system is, at least partially, built on form analyzers.

From the neurophysiological perspective, motion sensitive neurons have been found in many cortical areas. In particular, visual areas MT and MST are highly specialized in motion processing (for review, see Albright & Stoner, 1995). These areas are located in the dorsal stream as opposed to form related areas located in the ventral stream. In sum, there is a broad range of evidence for different systems dedicated to the processing of motion and form and that motion constitutes an independent perceptual dimension. However, there is also evidence that these systems are not strictly independent, but rather interact.

3. The problem of phenomenal identity and the correspondence problem

After Wertheimer’s pioneering work on apparent motion, while the major focus of Gestalt psychology shifted to static images, there was still a strong emphasis on motion. In his 1925 dissertation, with Wertheimer as his second reader, Joseph Ternus took the task of studying how grouping principles can be applied to stimuli in motion. The fundamental question he posed was what he termed the problem of phenomenal identity: “Experience consists far less in haphazard multiplicity than in the temporal sequence of self-identical objects. We see a moving object, and we say that “this object moves” even though our retinal images are changing at each instant of time and for each place it occupies in space. Phenomenally the object retains its identity” (Ternus 1926). He adopted a stimulus previously used by Pikler (1917), shown in Fig. 2A.
Figure 2. A. A simple Ternus-Pikler display. B. An apparent motion stimulus with two different shapes. C. The influence of shape is strong in correspondence matching when there is overlap between stimuli (left) and becomes weaker as the overlap is eliminated (right). D. A stimulus configuration used by Ternus to investigate the relationship between local motion matches and global shape configurations.

The first frame contains three identical elements. In the second frame, these elements are displaced so that some of them overlap spatially with the elements in the previous frame. In the example of Fig. 2A, the three disks are shifted by one inter-disk distance so that two of the disks overlap across the two frames. Given all identical elements in the two frames, one can then ask how will the elements be grouped across the two frames? This question has been later termed the “motion correspondence” problem. If we consider the central disk in Frame 2 (Fig. 2A), will this disk be grouped with the rightmost disk of the first frame based on their common absolute spatial location, i.e, the same retinal position, or will it be grouped with the central disk of the first frame based on their relative position as the central elements of spatial groups of three elements? The answer to this question turned out to be quite complex with several variables influencing the outcome. For example, when the ISI between the two frames is short, the leftmost element in the first frame appears to move to the rightmost element in the second frame while the spatially overlapping elements in the center appear stationary (i.e., they are grouped together). For longer ISIs, a completely different organization emerges: the three elements appear to move in tandem as a group, i.e., their relative spatial organization prevails in the spatiotemporal organization. These two distinct percepts are called element and group motion, respectively. Many other variables, such as inter-element separation, element size, spatial frequency, contrast, inter-stimulus interval (ISI), luminance, frame duration, eccentricity, and attention influence which specific organization emerges as the prevailing percept (e.g., Alais & Lorenz, 2002; Aydin et al., 2011; Breitmeyer & Ritter, 1986a, 1986b; Casco & Spinelli,

4. Form motion interactions

4.1. How local form information influences motion perception

The apparent motion stimulus lends itself nicely to the study of form motion interactions (for other examples of form motion interactions, see Blair et al., this volume). Remember that Zeno claimed that motion is an illusion created by the observer in order to reconcile the existence of an object at two different spatial locations at two different time instants. The observer would compare the two stimuli from memory and if a suitable match is found a phenomenal identity will be attributed to these two stimuli as two instances of the same object. Perceived motion from one object to the other would signal the conclusion that these two objects are one and the same. Thus, according to this view, form analysis is a precursor of motion perception and the match of the form of the two objects is a prerequisite for motion perception. This can be tested directly by creating an apparent motion stimulus where the shapes presented in the two frames are different (Fig. 3; see also the demo “AM – different shapes”). Many such experiments have been carried out showing that form has little effect on the perception of apparent motion, i.e., motion percepts between the two stimuli are strong (Kolers, 1972). In the example of Fig. 3, one perceives the square morphing into a circle along the path of apparent motion. That the shape of an object in apparent motion should remain constant can, in general, be expected to hold only for small displacements. This is because, the proximal stimulus is a two-dimensional projection of a three-dimensional object and during motion, one experiences perspective changes resulting in different views of the object. It is this very fact that Ternus used in defining the problem of phenomenal identity.

In the case of the example shown in Fig. 2B, there is no motion ambiguity and the interpretation of an object whose form changes (presumably due to perspective change) appears to be a natural solution. What happens, however, if the correspondences in the display are more complex and represent ambiguities such as the ones shown in Fig. 2C? Results indicate that form information (or in general feature information such as color, texture) can be used in resolving ambiguities in the case where there is physical overlap between elements of the two frames (Ternus-Pikler displays; see for example the demo “TP – feature bias”) but this influence becomes weaker when the overlap is reduced and the distance between the elements is increased (Hein & Cavanagh, 2012). Taken together, all these results indicate that motion and form are separate but interacting systems.

4.2. How local motion information influences form perception

Having answered the question of how local form information can influence motion perception, one can ask the converse question, viz., how local motion information can influence form perception. Figure 2D shows one of Ternus’ displays where in each static display dots group into global shapes. One can see a vertical line and a diamond shape which are moved left to right
and vice-versa respectively. However, the strength of static groups cannot predict the perceived forms in motion; i.e. the percept in Fig. 2D does not correspond to a line moving right and a diamond moving left. Instead, at short ISIs, the three horizontally-aligned central dots appear stationary while the outer dots appear to move rightwards. For longer ISIs, the percept appears to be that of a single object rotating 180 degrees in 3D (Ternus, 1926). Note that in these complex displays, multiple possible motion correspondences exist (e.g., Dawson & Wright, 1994; Otto, Ogmen, & Herzog, 2008) and the percept may vary from subject to subject, or even from trial to trial for the same subject. The reader can experiment with the demo “TP complex configuration”.

Having established that form and motion information interacts, the next question is to understand how. Combining signals from form and motion systems require a common basis upon which they can be expressed. In other words, what is the reference frame that allows interactions between these two systems? We will proceed first by discussing reference frames within the motion system and then by extending these reference frames to form computations.

5. Reference frames

5.1. Relativity of motion and reference frames

The Gestalt psychologist Karl Duncker’s work was instrumental in highlighting the importance of reference frames in perception (Duncker, 1929; for review, see Wallach, 1959; Mack, 1986). In one of his experiments, he presented a small stimulus embedded in a larger one (Fig. 3A, left panel). He moved the large surrounding stimulus while keeping the smaller one stationary. Observers perceived the surrounding stimulus as stationary and the smaller stimulus as moving in the direction opposite to the physical motion of the surrounding stimulus (for a recent paper with demos, see Anstis & Casco, 2006). To account for this illusory induced motion, he proposed that the larger surrounding stimulus served as the reference frame against which the position of the embedded stimulus is computed. The right panel of Fig. 3A shows another configuration studied by Duncker, the “rolling wheel”. If a light dot stimulus is placed on the rim of a wheel rolling in dark, the perceived trajectory of this dot is cycloidal. If a second dot at the center of the wheel is added to the display, one perceives the central dot to move in a linear trajectory and the dot on the rim is perceived to rotate around the central dot. In other words, the central dot serves as a reference against which the motion of the second dot is computed (for demos on the relativity of motion using the Ternus-Pikler paradigm, the reader is referred to demos in Boi et al., 2009).
To explain these effects, Johansson (1973) proposed a theory of vector analysis based on three principles. The first principle states that elements in motion are always perceptually related to each other. According to his second principle, simultaneous motions in a series of proximal elements perceptually connect these elements into rigid perceptual units. Finally, when the motion vectors of proximal elements can be decomposed to produce equal and simultaneous motion components, per the second principle, these components will be perceptually united into the percept of common motion. Fig. 3B illustrates these concepts. Fig. 3B-a shows the stimulus. By the first principle, the motion of these dots are not perceived in isolation but are related to each other. By the second principle, the top and bottom dots are connected together as a single rigid unit moving together horizontally. By the third principle, a horizontal component equal and simultaneous with the horizontal motion of the top and bottom dots is extracted from the motion of the central dot (Fig. 3B-b). The resulting percept is the horizontal movement of three dots during which the central dot moves up and down between the two flanking dots (Fig. 3B-c) (Johansson, 1973).

In a more natural setting, the distal stimulus generates a complex optic flow pattern on the retina. For example, while watching a street scene, one perceives the background (shops, houses, etc.) as stationary, cars and pedestrians moving with respect to this stationary background, the legs and arms of pedestrians undergoing periodic motion with respect to their body, their hands moving with respect to the moving arms, etc. Thus, the stimulus can be analyzed as a hierarchical series of moving reference frames and motions are perceived with

Figure 3. A. Two stimulus configurations studied by Duncker. The top panels represent the stimuli and the bottom panels depict the corresponding percepts. **Left panels:** Induced motion, **Right panels:** Rolling wheel illusion. B. An example illustrating Johansson’s vector decomposition principles. a. The stimulus. b. The decomposition of the motion of the central dot so as to identify common vector components for all three dots. c. The resulting percept.
respect to the appropriate reference frame in the hierarchy (e.g., hand with respect to the arm, the arm with respect to the body). While powerful and intuitively appealing, the basic principles of this theory are not sufficient to specify unambiguously how vectors will be decomposed in complex naturalistic stimuli. In fact, a vector can be expressed as the sum of infinitely many pairs of vectors, and it is not clear a priori how to predict which combination will prevail for complex stimuli. The difficulty faced here is similar to the one when we attempt to apply the Gestalt “laws” derived from simple stimuli to complex stimuli. To address this issue, Gestaltists put forth the “law of Prägnanz” (or the law of good Gestalt) which states that among the different possible organizations, the one that is the “simplest” is the one that will prevail (Koffka, 1935; Cutting & Proffitt, 1982; for a review, see van der Helm, this volume). However, the criterion for “simplest” remains arbitrary and elusive. The same concept has been adopted by other researchers who tried to quantify the simplicity of organizations. For example, Restle (1979) adopted the coding theory where different solutions are expressed as quantifiable “codes”. A stimulus undergoing circular motion can be described by three parameters: amplitude, phase, and wavelength. Restle used the number of parameters describing a configuration as the “information load” and predicted that the configuration with lowest information load would be the preferred (i.e., perceived) configuration. Dawson (1991) used a neural network to combine three heuristics in solving the correspondence problem. However, these approaches all suffer from the same general problems: As acknowledged by Restle, the method does not have an automatic way for generating all possible interpretations. Moreover, the choice of parametrization and its generality, the heuristics, their benefit and costs as well as the optimization criteria remain arbitrary.

5.2. Object file theory
Kahneman and colleagues addressed the problem of phenomenal identity by adapting two concepts from computer science, viz., addresses and files (Kahneman et al., 1992). The fundamental building blocks of their theory are “object files”, each containing information about a given object. These files establish and maintain the identities of objects. According to their theory, an object file is addressed, not by its contents, but by the location of the object at a given time. This location-based index is a type of reference frame discussed in the previous section. However, by restricting the file addressing mechanism to a spatial location, this theory faces many shortcomings. In the object file theory, features are available on an instant-by-instant basis and get inserted into appropriate files. On the other hand, feature processing takes time. Without specifying the dynamics of feature processing, the theory ends up in a bootstrapping vicious circle. When and how is the opening of an object file triggered? Since an object is defined by features, an initial evidence for opening a file for an object necessitates that at least some of the relevant features of the object are already processed; however, the processing of features for a specific object requires that a file for that object is already opened. Typical experiments used within the context of the object file theory include static preview conditions whose “main end product (...) is a set of object files” (Kahneman et al., 1992).

---

3 A similar concept was also proposed by Pylyshyn in his FINST theory (Pylyshyn, 1989). Several extensions and variants of the object file theory have been proposed, including the detailed analysis of object updating (Moore & Enns, 2004; Moore et al., 2007) and hierarchies in object structures (Lin & He, 2012).
However, under normal viewing conditions objects often appear from our peripheral field or behind occlusions necessitating mechanisms that can operate in the absence of static preview conditions. Another problem with the object file theory is that while vision has geometry, “files” do not specify a geometric structure. Objects have spatial extent and thus the location of an object cannot be abstracted from its features. Assume that the centroid of an object is used as its location index. To put features in the file indexed by this location, one needs to know not just one location index but the retinotopic extent of the object, which in turn necessitates surface and boundary features. Moreover, as we will discuss below (feature attribution and occlusion problems), objects may occlude each other. The insertion of correct features to correct object files cannot be accomplished by location indices alone, spatial extent and occlusion information needs to be represented as well.

In sum, while all this work highlights the importance of motion grouping and motion based reference frames, a deeper understanding of why the visual system needs reference frames may provide the constraints necessary to determine how and why reference frames are established.

6. The need for reference frames

6.1. The problems of motion blur and moving ghosts

In order to appreciate why reference frames are needed, consider first the fact that humans are mobile explorers and interact constantly with other moving objects. The input to our visual system is conveyed following the optics of the eye. The geometry of image formation can be described by projective geometry. Neighboring points in the environment are imaged on neighboring photoreceptors in the retina. The projections from retina to early visual cortical areas preserve these neighborhood relationships creating a retinotopic representation of the environment. To analyze the impact of motion on these representations, we need to consider the dynamical properties of the visual system.

A fundamental dynamical property of vision is visible persistence: Under normal viewing conditions, a briefly presented stationary stimulus remains visible approximately 120 ms after its physical offset (e.g., Haber & Standing, 1970; Coltheart, 1980). Based on this duration of visible persistence, we would expect moving objects to appear highly blurred. For example, a target moving at 10 deg/s would generate a trailing smear of 1.2 deg. The situation is similar to taking pictures of moving objects with a camera at an exposure duration that mimics visible persistence. Not only do the moving objects exhibit extensive motion smear, they also have a ghost-like appearance without any significant form information. This is because static objects remain long enough on a fixed region of the film to expose sufficiently the chemicals while moving objects expose each part of the film only briefly thus failing to provide sufficient exposure to any specific part of the film. Similarly, in retinotopic representations, a moving object will stimulate each retinotopically localized receptive-field briefly and an incompletely processed form information would spread across the retinotopic space just like the ghost-like appearances in pictures (Ogmen, 2007). Unlike photographic images, however, in human vision objects in motion typically appear relatively sharp and clear (Bex, et al., 1995; Burr, 1980; Burr &
Morgan, 1997; Burr et. al, 1986; Hammett, 1997; Ramachandran et al., 1974; Westerink & Teunissen, 1995).

In normal viewing, we tend to track moving stimuli with pursuit eye movements and thereby stabilize them on the retina. While pursuit eye movements can help reduce the perceived blur of a moving object (Bedell & Lott, 1996), the problem of motion blur remains for other objects present in the scene, since we can pursue only one object at a time. Eye movements also cause a retinotopic movement for the stationary background creating the blur problem for the background. Furthermore, the initiation of an eye movement can take about 150-200 ms during which a moving object can generate considerable blur. How does the visual system solve the problems of motion blur and moving ghosts? A potential solution to the motion blur problem is the use of mechanisms that inhibit motion smear in retinotopic representations (Chen et al., 1995; Ogmen, 1993, 2007; Purushothaman et al., 1998). A potential solution to the moving ghosts problem is the use of reference frames that move along with moving objects rather than being anchored in retinotopic coordinates (Ogmen, 2007).

6.2. The problems of dynamic occlusions and feature attribution
When an object moves, a variety of dynamic occlusions happens. The object occludes different parts of the background and, depending on depth relations, either occludes or gets occluded by other objects in the scene. Moreover, as its perspective view changes with respect to the observer, its visible features also change due to self-occlusion. All these dynamic considerations lead to two inter-related questions: First, as highlighted by Ternus, how does the object maintain its identity despite the changes in its features? Second, due to these occlusions, features of different objects become dynamically entangled. How does the visual system attribute features to the various objects in a consistent manner? As discussed in the previous sections, a possible solution to maintain object identities is to establish motion correspondences and to arrange the resulting motion vectors as a hierarchical set of reference frames. These exo-centered reference frames establish and maintain the identity of objects in space and time. As we discuss below, these reference-frames can also provide the basis for feature attribution.

7. Non-retinotopic feature attribution

7.1. Sequential metacontrast and non-retinotopic feature attribution
Earliest studies of motion blur and deblurring can be traced back to McDougall (1904) and Piéron (1935). Figure 9 depicts the stimulus arrangements used by these researchers. As mentioned in Section 6.1, the motion blur generated by a moving stimulus can be “deblurred” by inhibitory mechanisms in retinotopic representations. In fact, McDougall reported that the blur generated by the leading stimulus “a” in Fig. 4A could be curtailed by adding a second stimulus, labeled “b” in Fig. 4A in spatiotemporal proximity. The specific type of masking where

---

4 Reference frames can be broadly classified into two types: Ego-centered reference frames are those centered on the observer (e.g., eye-centered, head-centered, limb-centered). Exo-centered reference frames are those centered outside the observer (e.g., centered on an object in a scene).
the visibility of a target stimulus is suppressed by a spatially non-overlapping and temporally lagging stimulus is called metacontrast (Bachmann, 1994; Breitmeyer & Ogmen, 2006).

Piéron (1935) modified McDougall’s stimulus to devise a “sequential” version as shown in Figure 4B. This sequential stimulus provides a temporally extended apparent motion and metacontrast stimulus that can be used to illustrate the motion deburring phenomenon. It can also be used to study the feature attribution problem. Fig. 4C shows a version of sequential metacontrast where the central line contains a form feature: A small Vernier offset is introduced by shifting the upper segment of the line horizontally with respect to the lower segment (Otto et al., 2006).

In this stimulus, the central line containing the Vernier offset is invisible to the observer because it is masked by the two flanking lines. One perceives two streams of motion, one to the left and one to the right. The question of feature attribution is the following: What happens to the feature presented in the central invisible element of the display? Will it also be invisible, or will it be attributed to motion streams? The results of experiments using various versions of this sequential metacontrast stimuli show that features of the invisible stimuli are attributed to motion streams and integrated with other features presented within each individual motion stream. In other words, features are processed according to reference frames that move according to the motion vector of each stream (Otto et al., 2006, 2008, 2009, 2010a,b).
7.2. **Ternus-Pikler display and non-retinotopic feature attribution in the presence of retinotopic conflict**

Ternus-Pikler displays are designed to pit directly retinotopic relations against non-retinotopic grouping relations. This property offers the advantage of directly assessing whether features are processed according to retinotopic or grouping relations (Ogmen et al., 2006). Fig. 5 shows an example of how the Ternus-Pikler display is used for studying feature attribution. As a feature, a Vernier offset, called the “probe Vernier” is inserted to the central element of the first frame (Fig. 5). Observers were asked to report the perceived offset direction for elements in the second frame, numbered 1, 2, and 3 in Fig. 5D-left. None of these elements contained a Vernier offset and naïve observers did not know where the probe Vernier was located. Consider first the control condition in Fig. 5E, obtained by removing the flanking elements from the two frames. In this case no motion is perceived. Based on retinotopic relations, the probe-Vernier should be integrated with element 1 in the second frame and the agreement of observers’ responses with the direction of probe-Vernier offset should be high for element 1 and low for element 2. If processing of Vernier were to take according to retinotopic relations, one would predict the same outcome for the Ternus-Pikler display regardless of whether element or group motion is perceived. On the other hand, if feature processing and integration take place according to motion grouping relations (Fig. 5B and 5C), instead of retinotopic relations, one would expect the probe Vernier to integrate with element 1 in the case of element motion (Fig. 5B) and with element 2 in the case of group motion (Fig. 5C). The results of this experiment along with those conducted with more complex combination of features show that form features are computed according to motion grouping relations, in other words, according to a reference frame that moves according to prevailing motion groupings in the display (Ogmen et al., 2006).
Figure 5. The Ternus-Pikler Display (A) and the associated percepts of “Element Motion” (B) and “Group Motion” (C). Dashed arrows in panels B and C depict the perceived motion correspondences between the elements in the two frames. Experimental results for Ternus-Pikler stimulus (D) and the control stimulus (E). From (Ogmen et al., 2006).

In follow-up studies, this paradigm has been applied to other visual computations and it has been shown that form, motion, visual search, attention, binocular rivalry all have non-retinotopic bases (Boi et al., 2009; 2011b). Non-retinotopic computation of various stimulus features has also been supported by other paradigms using motion stimuli (Shimozaki et al., 1999; Nishida, 2004; Nishida et al., 2007; Kawabe, 2008) or attentional tracking (Cavanagh et al., 2008). On the other hand, not all processes are non-retinotopic; motion and tilt adaptation have
been found to be retinotopic (Boi et al., 2011a; Knapen et al., 2009; Wenderoth & Wiese, 2008) indicating that they are by-products of computations occurring prior to the transfer of information from retinotopic to non-retinotopic representations.

8. Concluding remarks

Motion is ubiquitous in the ecological environment and most biological systems devote extensive neural processing to its analysis. This importance has been recognized by philosophers and scientists who carried out extensive studies on how motion is processed and perceived. While there has been convergence in the types of computational models that can detect motion, the broader issue of how motion is organized as a spatiotemporal Gestalt remains a challenging question. The discovery of the relativity of motion led to the introduction of hierarchical reference frames according to which part-whole relations can be constructed. This chapter provided a review of why reference frames are needed from ecological and neurophysiological (retinotopic organization) perspectives. These analyses show that reference frames are needed not just for motion computation but for all stimulus attributes. We expect future research to develop more in depth the properties of these reference frames which will provide a common geometry wherein all stimulus attributes can be processed jointly.
9. References


