Frame and metrics for the reference signal

Victor I. Belopolsky

vbelop@ipras.msk.su; Institute of Psychology, Russian Academy of Science, Moscow 129366, Russia

In the attempt to reinforce the behavioral validity of the traditional inferential theory of perception, Wertheim has reconsidered the nature of the extraretinal signal. The reference signal, instead of the extraretinal one, is assumed to encode information about eye movement relative to external space together with visual spatiotemporal and vestibular motion information. Although this novel position appears reasonable and is capable of explaining some contradictory experimental findings, it faces serious problems, both old and new.

The basic assumption of the traditional inferential theory is the egocentric, mostly oculocentric, representation of the external world; that is, to be compatible, the metrics of both the retinal and the hypothesized extraretinal signals must be defined in visual angles. Wertheim accepts that "we see a stable world during eye movements because retinal and extraretinal signals are equal: the velocity of the image of the world across the retinae equals the velocity of the eyes" (sect. 1, para. 1). However, even under conditions in which the eye rotates around a fixed center, the two signals do not operate in the same metric.

The above claim is derived from the fact that the eye's rotation center does not coincide with the eye's nodal point. So when the eye rotates a certain angular distance, the corresponding shift in the retinal image of a stationary object will actually be a fraction of that angle. In terms of visual directions, the ratio could be up to 1:2. The same ratio will hold for foveal velocity during smooth pursuit of a moving target and for the target's retinal velocity relative to the stationary eye. Jung (1972), who first advanced this line of argument, used it to explain the Aubert-Flieschl phenomenon. This can also be considered the source of the Filehne illusion. As a consequence, we must assume that recalibration of the extraretinal signal is a precondition for visual stability during eye movements.

The discrepancy between the center of rotation and the nodal point of the eye high-lights the role of another factor that is usually underestimated by the inferential theory, the target's distance. Two targets at different distances which were previously on one visual axis lose this identity of their visual directions after an eye movement (Howard 1982, p. 278, Fig. 7.1). This means that retinal velocity during eye movement is a function of the target's distance. This factor could play an important role in contradictory demonstrations of the Filehne illusion: it is much easier to experience the illusion when the moving target and the background are located in different planes than when they are in the same plane.

The situation becomes more complicated when someone tries to define the metrics of eye movement relative to external space. Wertheim introduces the vector signal of head movement in coordinates of 3-D "Newtonian" space (sect. 5.3, para. 1). It is not difficult to show that the metrics of head rotation differ from the oculomotor and retinal vector metrics even when a subject is sitting upright and turning his head and eyes around a vertical axis. If we also take into account translational head movements (because the head has 6 degrees of freedom), this raises the question of the metrics of the reference signal. We must agree that Wertheim's Equation 9 is accurate only if the dimensionalities of its terms are the same. However, $V_{est.s}$ and V_{ref} are encoded in the metrics of the exocentric coordinate system whereas V_{ret} is encoded in oculocentric terms. Moreover, in general, there is no universal transfer rule for transforming one coordinate system into another: the rule depends on the relative positions and movements of the observer and the external objects. This point was in fact crucial for Gibson's rejection of geometrical optics in favor of his ecological position (Gibson 1979).

The acceptance of a visual component of the reference signal reflects the proved significance of visual feedforward in visual stability processing in addition to visual feedback (e.g., Belopolsky 1978; MacKay 1973). Unfortunately, this claim cannot be formalized in terms of vector algebra (see sect. 6.5, para. 1) and leads to the redundant duplication (or triplication?) of visual pathways in the proposed model in the target article.

According to the proposed model, the reference signal has no sensory correlate; it is used only to cancel, completely or partially, the retinal signal. As a result, its role is purely visual rearrangement. The sensations of self-movement or self-stability come through a parallel branch of the information processing system, although the reference signal itself contains all necessary data. The model does not provide any special mechanism for the coordination of visual and ego movement in space. For example, the model is rather efficient in explaining the time course of background motion perception (assuming variability in V_{ref} gain) but fails to explain movement of eye/head/body egocenters during the circular vection illusion. It is worth noting that direct perception theory manages this problem by considering the observer's body parts as the context of the optical array (Gibson 1979).

The most challenging problem for both theories remains the nature of eye or, more appropriately, gaze positional sense, by which I mean the human's ability to hold, reorient, and locate the position of attentional focus in space. Direct perception theory does not indicate the body's landmark connected with the gaze direction (this cannot be derived simply as the center of the optical array). On the other hand, this sense cannot be identified with an extraretinal or a proposed reference signal. Two examples will illustrate this idea. First, when the stabilized retinal image subtends a large visual angle (> 40 deg of arc), eye movements do not change its apparent spatial location, as occurs with a smaller image. Especially interesting is the fact that, in both cases, subjects experience their gaze as moving in space (Belopolsky 1985; Zinchenko & Vergiles 1972). Second, when voluntary eye movements are made as the subject examines meaningless texture patterns through an artificially reduced (up to 3–5 deg of arc) central visual field, the visual world is perceived as movable relative to the stationary gaze (Belopolsky 1978).

In summary, Wertheim's target article provides a subtle analysis of the direct versus inferential perception controversy. However, the attempt to resolve the controversy on the basis of inferential theory has made it too complicated and flexible to be an effective tool for predicting perceptual experience in certain conditions.