N94- 25510

Extracting Heading and Temporal Range from Optic Flow: Human Performance Issues

Mary K. Kaiser
John A. Perrone
Leland Stone
NASA Ames Research Center

Martin S. Banks
James A. Crowell
University of California, Berkeley

ABSTRACT

Pilots are able to extract information about their vehicle motion and environmental structure from dynamic transformations in the out-the-window scene. In this presentation, we focus on the information in the optic flow which specifies vehicle heading and distance to objects in the environment (scaled to a temporal metric). In particular, we are concerned with modeling how the human operators extract the necessary information, and what factors impact their ability to utilize the critical information. In general, the psychophysical data suggest that the human visual system is fairly robust to degradations in the visual display (e.g., reduced contrast and resolution, restricted field of view). However, extraneous motion flow (i.e., introduced by sensor rotation) greatly compromises human performance. The implications of these models and data for enhanced/synthetic vision systems are discussed.

INTRODUCTION

The out-the-cockpit scene provides a variety of visual cues to aid the pilot with vehicular control. As Walter Johnson discussed in his talk, some of these can be considered as static (e.g., horizon ratios), whereas others are dynamic or time-varying (e.g., change in the splay angle of the runway). Our research examines the control relevant information carried in the optic flow. Optic flow is the visual streaming of visible points, edges, and objects that results when one moves through a stationary, structured environment. During transport flight, relevant optic flow occurs primarily below the horizon line -- it is defined by textures and objects on the ground plane.

Optic flow is represented as a field of vectors, with the length of each vector representing the speed at which an element moves relative to the vantage point of the sensor (e.g., the human eye). For linear motion with a fixed-orientation sensor, the focus of expansion of the vector field defines the heading. If the sensor rotates as it translates (e.g., if it fixates on a point in the environment), this adds a common motion component to all the vectors which needs to be factored out before heading can be recovered. Once heading is extracted, the angle objects form relative to the heading (and the rate of change of this angle) define their temporal range. Thus, heading extraction is a critical component to range extraction as well. In this presentation, we describe a model of heading extraction by human observers which is both physiologically plausible and consistent with psychophysical data. We then discuss the psychophysical findings from our laboratories concerning what factors do and do not degrade heading and temporal range extraction.

HEADING EXTRACTION

Many algorithms have been proposed for solving the self-motion estimation problem (for reviews, Warren, Morris, & Kalish, 1988; Warren & Hannon, 1990). Some of these use the image motion from a small number of points to solve a set of nonlinear equations (e.g. Longuet-Higgins & Prazdny, 1980; Ballard & Kimball, 1983) . Such techniques tend to be sensitive to noise in the image motion measurements and must rely on iterative methods to arrive at a solution. Others make use of differential invariants of the flow field and are based on spatial derivatives (e.g. Koenderink & van Doorn, 1975). In addition to being sensitive to noise, these methods require locally continuous flow fields and a smoothness constraint for environmental surfaces. One of the more popular approaches to the self-motion problem makes use of the fact that image motion resulting from rotation is independent of the depth of points in the scene, while that resulting from translation is not (Longuet-Higgins & Pradzny, 1980). Therefore, the difference between flow-field vectors at adjacent points at different depths yields information related to the translation only. Rieger and Lawton (1985) developed a model which uses this principle, but which is able to use flow-field vectors from nearby points on the image plane rather than points that were exactly adjacent or overlapping. This "local differential motion model" is currently the most popular candidate for the algorithm underlying human self-motion perception (see Warren & Hannon, 1990; Hildreth, 1992). However, psychophysical studies at Ames Research Center by Perrone and Stone (Perrone & Stone, 1991; Stone & Perrone, 1991, 1993] have shown that heading can still be estimated correctly in situations that lack the local differential image motion necessary for the Reiger-Lawton model to work properly.

To explain their psychophysical findings, Perrone and Stone (Perrone, 1992; Perrone & Stone, 1992a, 1992b) have recently proposed an altogether different "physiologically-based" approach to solving the self-motion problem (Figure 1). The rationale for using a physiologically-based system is two-fold. First, it is more likely to allow extrapolation to a wider range of human performance and secondly, such "reverse engineering" will hopefully eventually lead to the design of artificial vision systems that are as robust and as fast as the human brain. One of the model's strengths is that it is based on known physiological properties of motion sensitive neurons in the Middle Temporal (MT) area of the primate visual cortex known to be involved in motion processing (Zeki, 1980; Maunsell & Van Essen, 1983; Albright, 1984; Newsome, Wurtz, Dursteler & Mikami, 1985; Newsome, Britten, & J. A. Movshon, 1989; Salzman, Britten, & Newsome, 1990) and proposes a theoretical framework for how neurons in the Medial Superior Temporal (MST) area might use the output from MT cells to extract heading. In the model, MT-like units carry out the local analysis of the 2-D image motion using direction and speed tuned "sensors" (Figure 2). The outputs from specific sets of MT-sensors are then summed to produce the output for a specialized MST-like "detector" which is "tuned" to a particular pattern of self-motion produced image motion and responds much like actual MST neurons (Saito, Yukie, Tanaka, Hikosaka, Fukada, & Iwai, 1986; Tanaka, Hikosaka, Saito, Yukie, Fukada, & Iwai, 1986; Duffy & Wurtz, 1991). These MST-like detectors sum MT-like sensor outputs over a large portion of the visual field and act as templates searching for specific patterns of global retinal image motion (Figure 3). The most active detector, within a map of possible combined translation-rotations, identifies what self-motion is most consistent with the image flow and, hence, solves the self-motion problem.

Comparison of human psychophysical data with simulations of the Perrone-Stone model (Figure 4) demonstrates that the model is consistent with known properties of visual heading perception and, in particular, that the model can provide a quantitative estimate of the break down of human performance at higher rotation rates seen by both Perrone and Stone (Perrone & Stone, 1991; Stone & Perrone, 1991) and Banks and colleagues (Royden et al., 1992). This approach is therefore very promising, although further psychophysical validation and refinement will be necessary before it can be used as an engineering design tool. In particular, the model does not attempt to include non-visual signals that are likely to contribute to human perception (Royden et al., 1992). However, the output-map structure of the Perrone-Stone model lends itself well to the incorporation of such additional non-visual information.

The Perrone-Stone model predicts, and psychophysical evidence demonstrates, that heading extraction is impaired when rotation (without non-visual information about rotation) is added to the visual display. Banks and his colleagues have also examined whether two aspects of display quality, resolution and contrast, affects people's ability to determine their heading from optic flow. Displays were presented both foveally and peripherally (40° nasal). Three levels of crab-angle (i.e., heading relative to the center of the display) were used: 0°, 20°, and 70°. In a reduced contrast study, Weber contrast was varied between 1 and 40 (0.85 is the contrast threshold for central vision, 3.10 is contrast threshold for 40° nasal). As shown in Figure 5, heading threshold varied as a function of crab angle; headings were harder to discriminate during higher crab angles. But heading extraction was fairly robust to contrast level, at least for supra-threshold contrast levels. For centrally viewed displays, performance did not improve with the Weber contrast levels increasing beyond five. In a visual acuity (resolution) study (Figure 6), there was a similar effect for crab angle, and some effect for resolution. Still, performance with the 0° crab angle, centrally viewed display was fairly accurate (threshold $< 2°$) even with 20/100 resolution.

TEMPORAL RANGE ESTIMATE

Given that people can extract heading from the optic flow, it is possible, in principle, to then determine the temporal range to any object in the environment (Kaiser & Mowafy, in press). For objects lying on the flight vector (Figure 7), the time to contact (TTC) is specified by the angular extent of the object, θ, divided by the rate of change of the angle, $\delta\theta/\delta t$. That is:

$$TTC = \theta \ / \ \delta\theta/\delta t \qquad (1)$$

For objects lying off the heading vector, an analogous derivation is possible, using the angle between the object and the tract vector, ϕ, and its rate of change, $\delta\phi/\delta t$. The ratio of these terms specifies time to passage (TTP), which is the time until the object intersects the eye-plane perpendicular to the heading vector (Figure 8):

$$TTP = \phi \ / \ \delta\phi/\delta t \qquad (2)$$

Most empirical work on people's sensitivity to this optical information has focused on the TTC situation, and the use of these cues for coordinating motor activity such as hitting and catching approaching objects (see Tresilian, 1991 for a review). However, the TTP case is more germane for most flight control regimes; the pilot needs to estimate the time to various way-points for navigation, control, and execution of maneuvers (e.g., flare). Kaiser and her colleagues (Kaiser & Mowafy, in press) have recently examined people's sensitivity to TTP information. In the experimental paradigm, observers viewed a translation through a volume of point lights, and either judged which of two targets would pass their eye plane first (relative judgment task) or indicated when a target which had left the field of view would pass their eye plane (absolute judgment task). In both relative and absolute judgment tasks, people were able to perform reliably. Judgments of relative TTP were precise to around 600 msec and were comparable for narrow (19°) and wide (46°) fields of view (Figure 9). Absolute TTP judgments were reliable even in the absence of feedback (Figure 10), indicating that people's temporal estimates are "pre-calibrated."

One manner in which pilots might use this TTP information for flight control is illustrated in Figure 11. For any assigned altitude, the distance along a particular gaze angle is constant in eye-heights (i.e., the ground plane along the 45° gaze angle is one eye-height distant, the ground plane along the 26.5° gaze angle is two eye-heights, etc.). Pilots may seek to maintain a constant temporal distance (i.e., lead time) to objects along a given gaze angle. This will result in appropriate flight control for some regimes (e.g., rotorcraft landing, where speed is reduced proportional to distance-to-go), but will cause an inappropriate bias when speed should be held constant during altitude change. Also, pilots may misjudge their taxi speeds if they perform ground operations in a variety of vehicles with very discrepant eye-heights (Figure 12).

IMPLICATIONS FOR ENHANCED/SYNTHETIC VISION SYSTEMS

Optic flow provides a critical source of visual information for vehicular control. If proposed sensor displays for enhanced/synthetic vision systems do not adequately preserve optic flow information, pilot performance may be impaired. Also, the noise from some sensor systems can mask or distort flow patterns. Empirical findings and performance models suggest that such extraneous pseudo-motion signals might seriously compromise human optical flow processing. In such cases where natural motion cues are degraded or distorted, pilots may require other visual cue augmentations (e.g., flare cues) to compensate.

REFERENCES

T. D. Albright, "Direction and orientation selectivity of neurons in visual area MT of the macaque," J. Neurophysiol. 52, 1106-1130 (1984).

D. H. Ballard and O. A. Kimball, "Rigid body motion from depth and optical flow," Comp. Vis. Graph. Image Process. 22, 95-115 (1983).

C. J. Duffy and R. H. Wurtz, "Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli," J. Neurophys. 65, 1329,1345 (1991).

E. Hildreth, "Recovering heading for visually-guided navigation," Vis Res. 32, 1177-1192 (1992).

M. K. Kaiser and L. Mowafy, "Optical specification of time-to-passage: Observers' sensitivity to global tau," J. Exp. Psychol. Hum. Percept. Perform. (in press).

J. J. Koenderink and A. J. van Doorn, "Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer," Optica Acta 22, 773-791 (1975).

H. C. Longuet-Higgins and K. Prazdny, "The interpretation of moving retinal images," Proc. R. Soc. London Ser B B208, 385-387 (1980).

J. H. R. Maunsell and D. C. Van Essen, "Functional properties of neurons in the middle temporal visual area of the Macaque monkey. I. Selectivity for stimulus direction, speed, orientation," J. Neurophysiol. 49, 1127-1147 (1983).

W. T. Newsome, K. H. Britten, & J. A. Movshon, "Neuronal correlates for a perceptual decision," Nature, 341: 52-54 (1989).

W. T. Newsome, R. H. Wurtz, M. R. Dursteler, & A. Mikami, "Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey," J. Neurosci. 5: 825-840 (1985).

J.A. Perrone, "Model for the computation of self-motion in biological systems," J. Opt. Soc. of Am. A, 9:177 (1992).

J. A. Perrone and L. S. Stone, "Using the properties of MT neurons for self-motion estimation in the presence of eye movements," Perception (Suppl) 21: 64-65 (1992a).

J. A. Perrone and L. S. Stone, "A possible role for the speed tuning properties of MT cells in the recovery of depth from motion," Invest. Opthalmol. & Vis. Sci. (Suppl) 33:1141 (1992b).

J. A. Perrone and L. S. Stone, " The perception of egomotion: Global versus local mechanisms," Invest. Opthal. & Vis. Sci. (Suppl) 32: 957 (1991).

J. H. Rieger and D. T. Lawton, " Processing differential image motion," J. Opt. Soc. Am. A 2, 354-360 (1985).

C. S. Royden, I. V. Laudeman, J. A. Crowell, & M. S. Banks, "The influence of eye movements on heading judgments," Invest. Opthal. & Vis. Sci. (Suppl) 33: 1051 (1992).

H. Saito, M. Yukie, K. Tanaka, K. Hikosaka, Y. Fukada, and E. Iwai, " Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey," J. Neurosci. 6, 145-157 (1986).

C. D. Salzman, K. H. Britten, & W. T. Newsome, "Cortical miscrostimulation influences perceptual judgments of motion direction," Nature 346, 174-177 (1990).

L. S. Stone and J. A. Perrone, "Human heading perception during combined translational and rotational self-motion," Soc. for Neursoci. Abstr. 17: 847 (1991).

L. S. Stone and J. A. Perrone, "Human heading perception cannot be explained using a local differential algorithm," Invest. Opthalmol. & Vis. Sci. (Suppl) 34: 1229 (1993).

K. Tanaka, K. Hikosaka, H. Saito, M. Yukie, Y. Fukada, and E. Iwai, " Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey," J. Neurosci 6, 134-144 (1986).

J. R. Tresilian, "Empirical and theoretical issues in the perception of time to contact," J. Exp. Psychol. Hum. Percept. Perform. 17, 865-876 (1991).

W. H. Warren and D. J. Hannon, " Eye movements and optical flow," J. Opt. Soc. Am. A 7, 160-168 (1990).

W. H. Warren, M. W. Morris, and M. Kalish, " Perception of translational heading from optical flow," J. Exp. Psychol. Hum. Percept. Perform. 14, 646-660 (1988).

S. M. Zeki, " The response properties of cells in the middle temporal area (area MT) of owl monkey visual cortex," Proc. R. Soc. London Ser B B 207, 239-248 (1980).

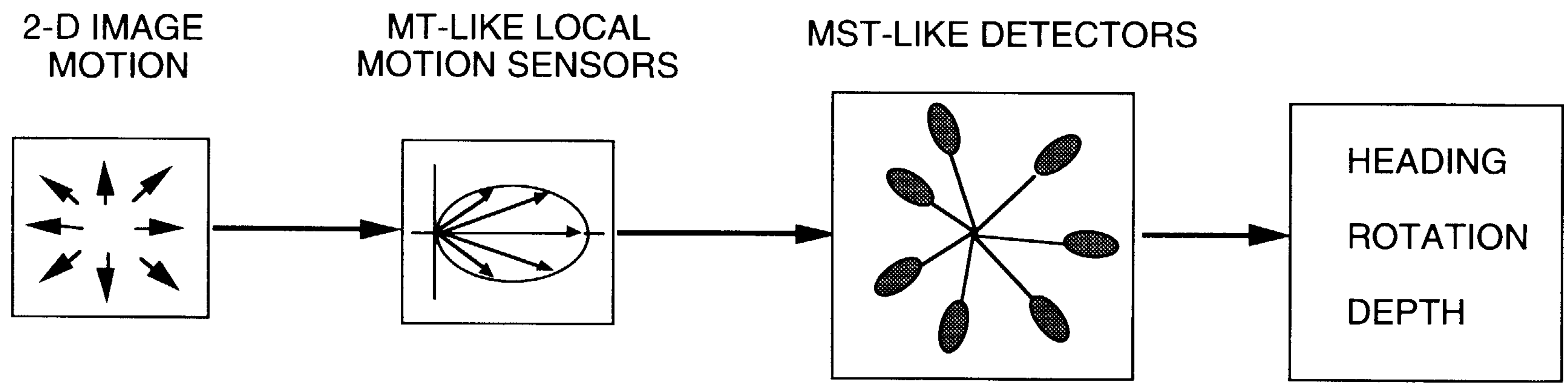

Figure 1. Overall structure of template model.

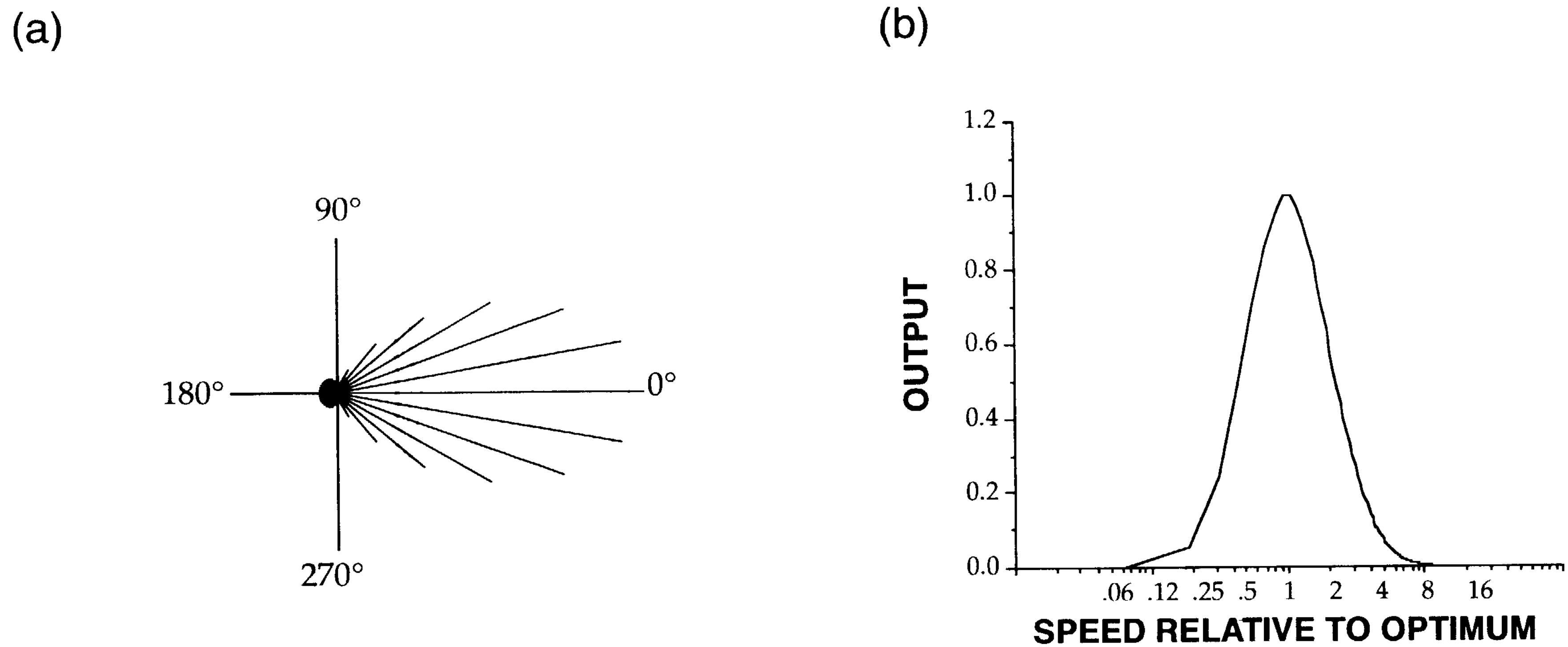

Figure 2. Idealized MT neuron responses. **a)** Direction tuning curve in polar plot form. **b)** Speed tuning curve.

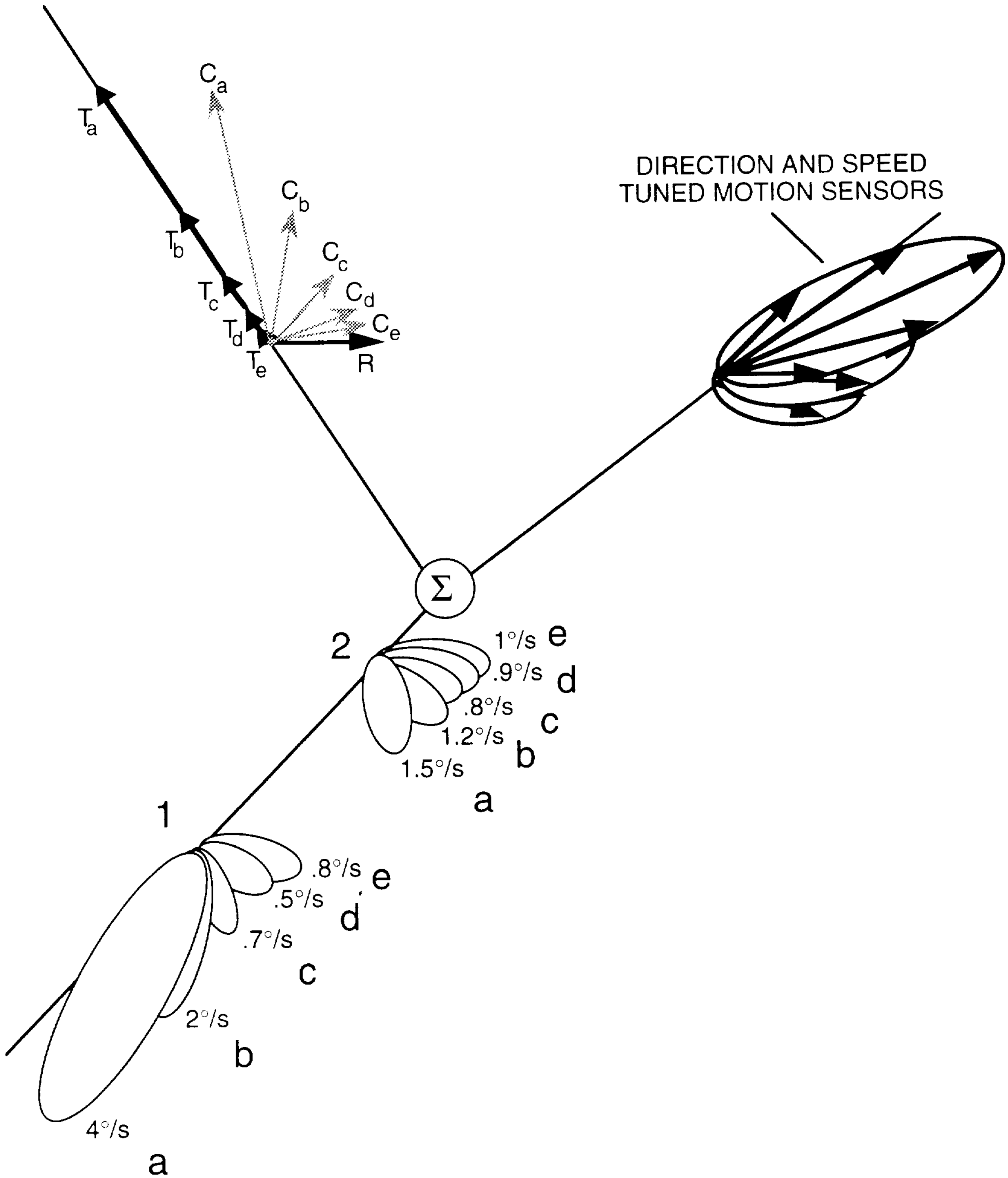

Figure 3. MST-like detector which acts as a template for a specific heading-rotation combination. The activity of groups of MT-like sensors at various locations in the visual field is summed, with the speed and direction-tuning of each sensor set to respond to the image motion, C = T (translation) + R (rotation), associated with a specific depth plane (a through e).

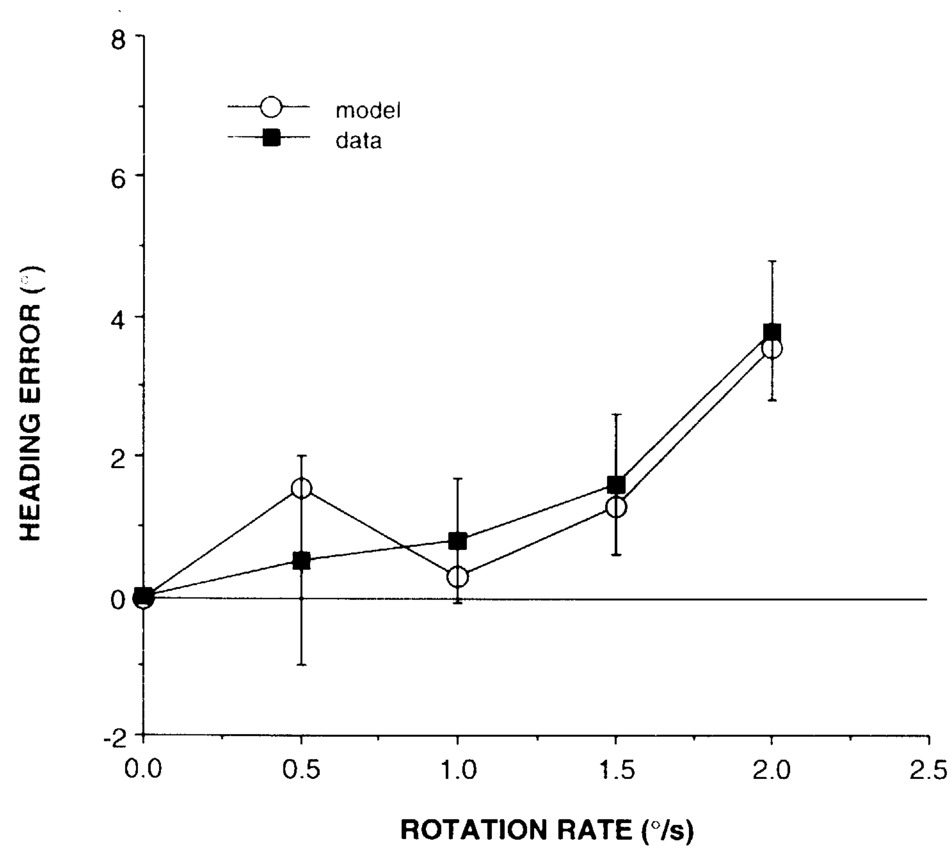

Figure 4. Comparison of heading error vs rotation rate for human observers and for the Perrone-Stone model.

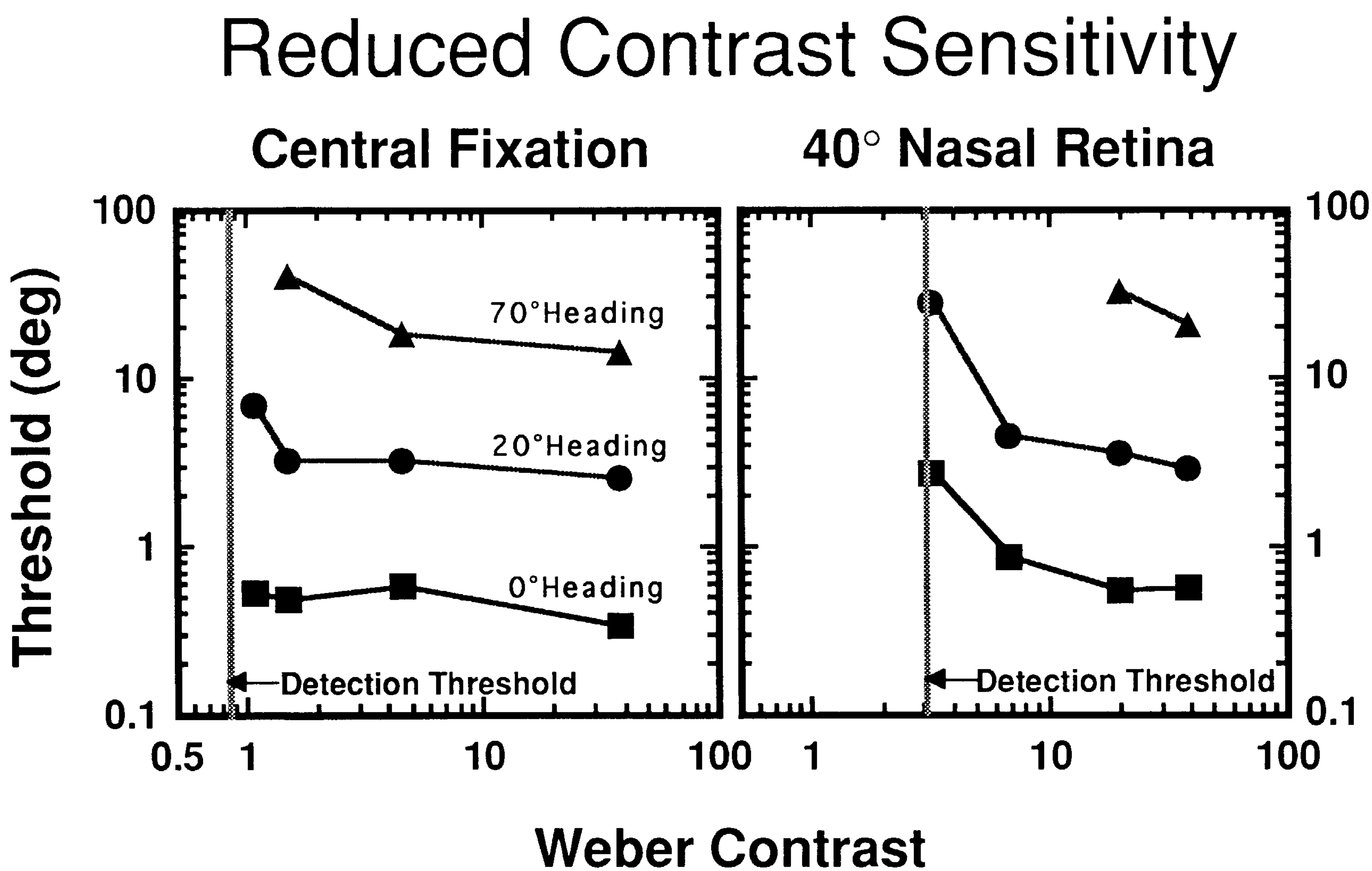

Figure 5. Heading threshold as a function of Weber contrast, eccentricity, and crab angle.

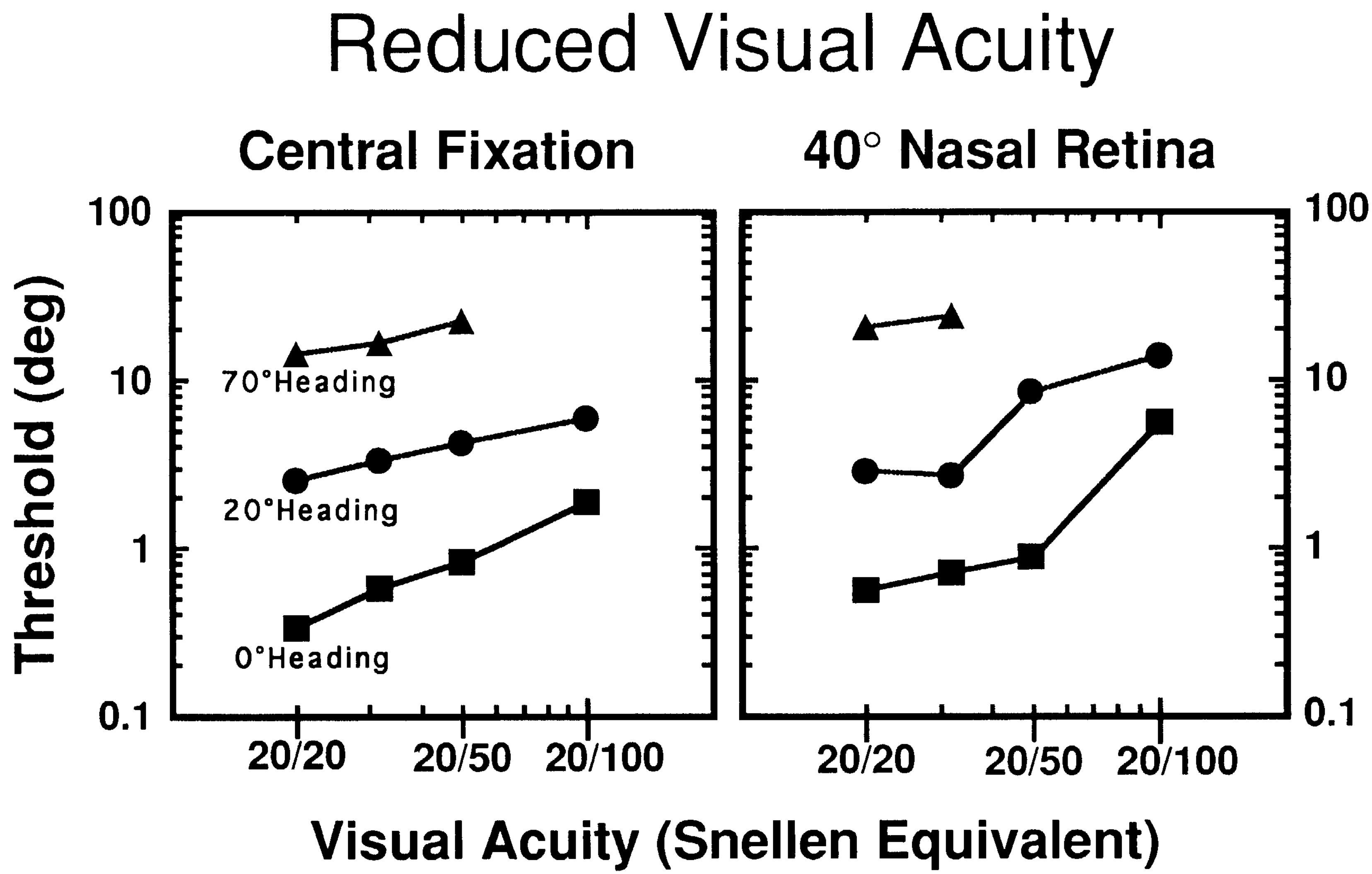

Figure 6. Heading threshold as a function of visual acuity, eccentricity, and crab angle.

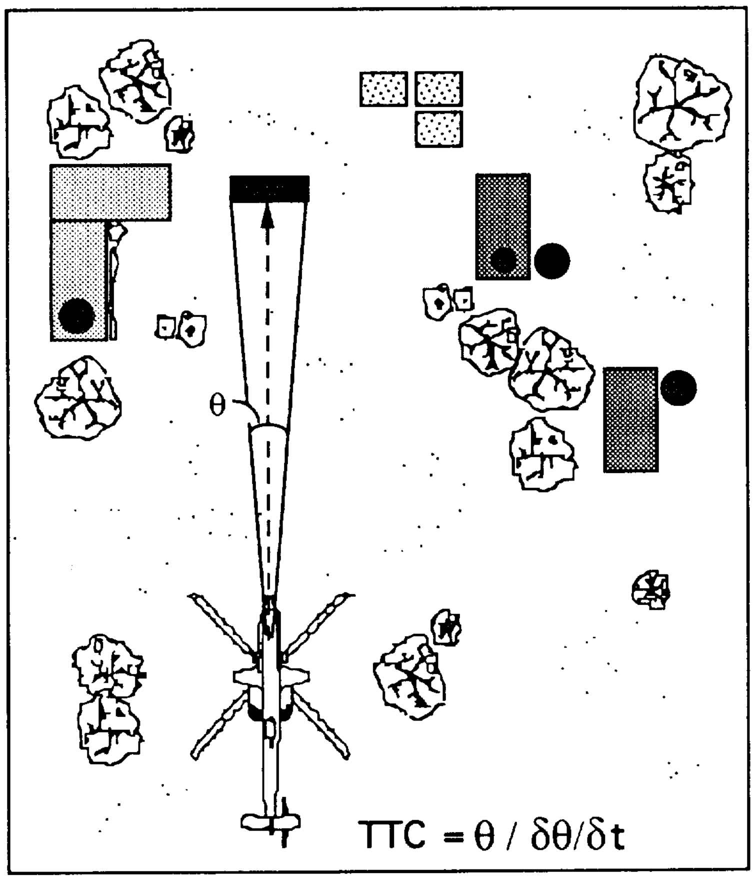

Figure 7. Geometry of the Time-to-Contact (TTC) situation. θ is the visual angle an object subtends.

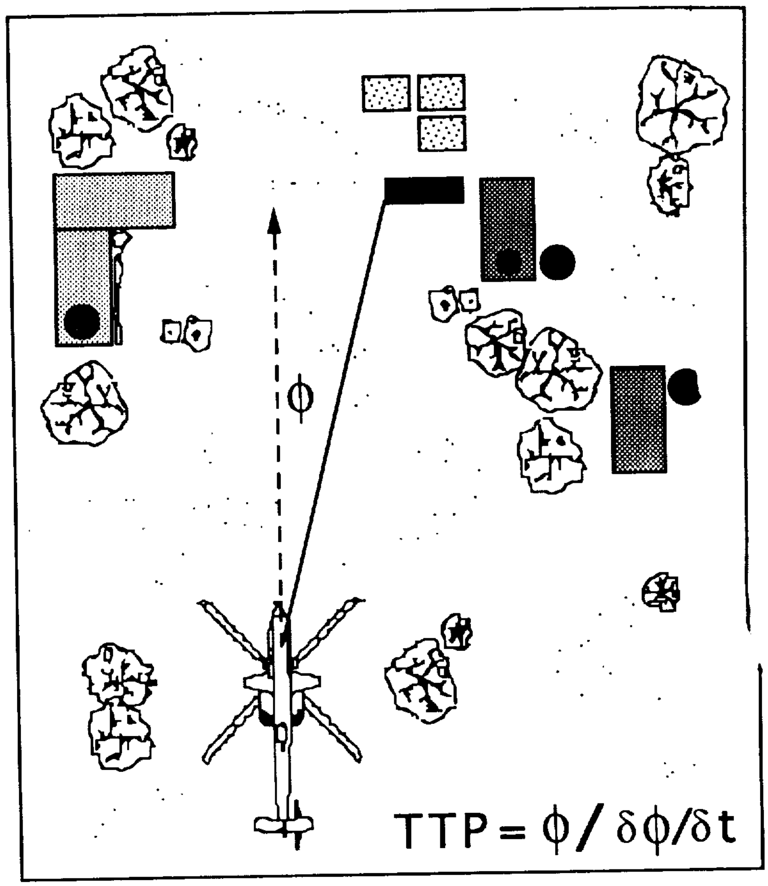

Figure 8. Geometry of the Time-to-Passage (TTP) situation. ϕ is the visual angle between an object and the heading vector.

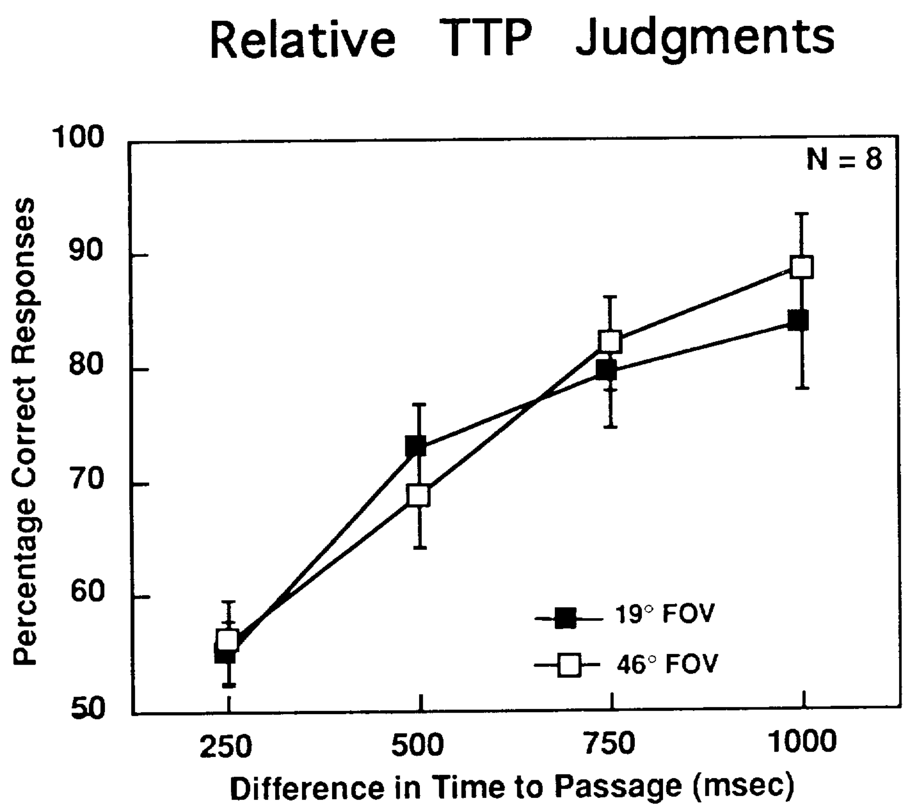

Figure 9. Relative Time-to-Passage (TTP) judgments for narrow (19°) and wide (46°) fields of view (FOV).

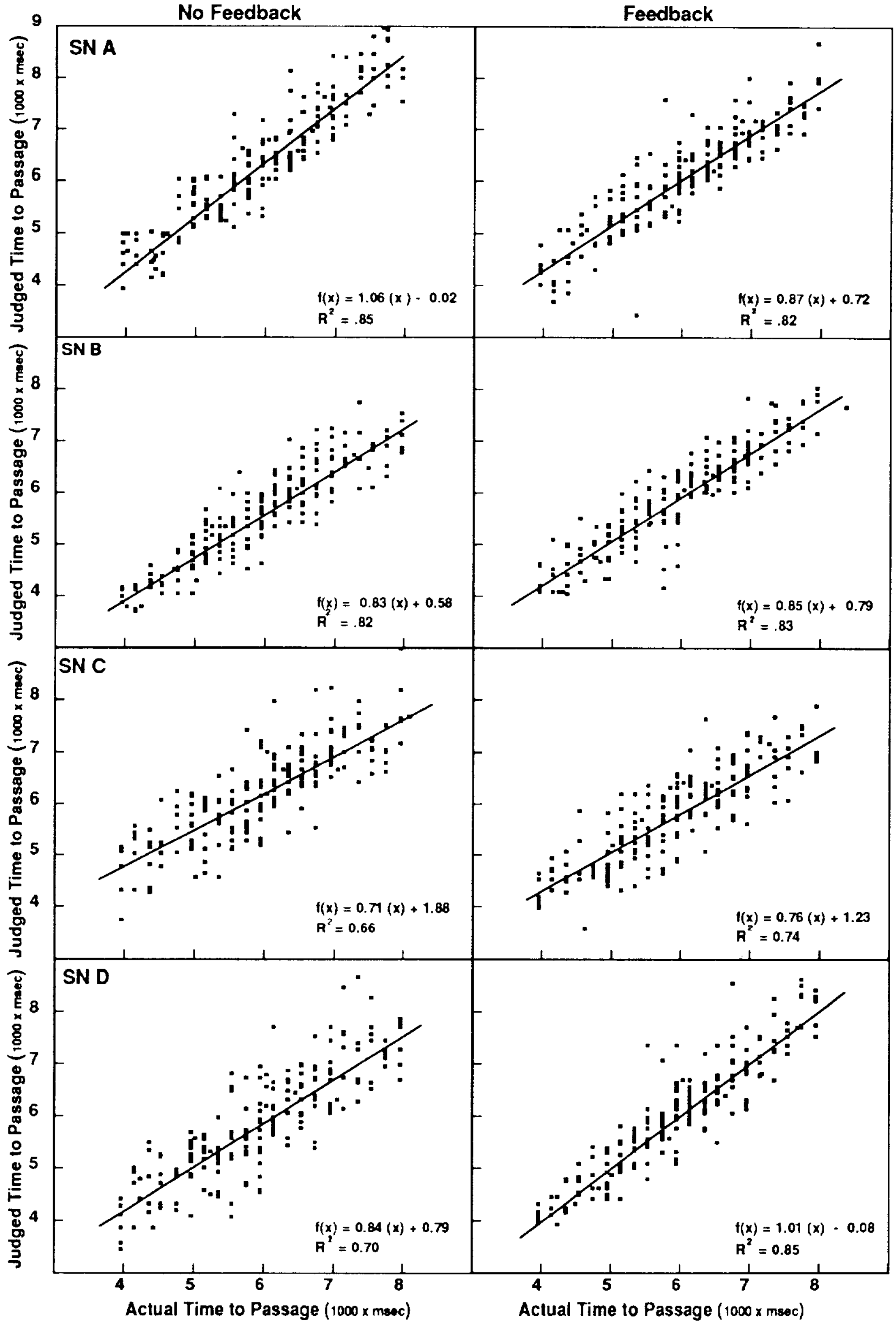

Figure 10. Relative Time-to-Passage (TTP) judgments in the presence and absence of feedback.

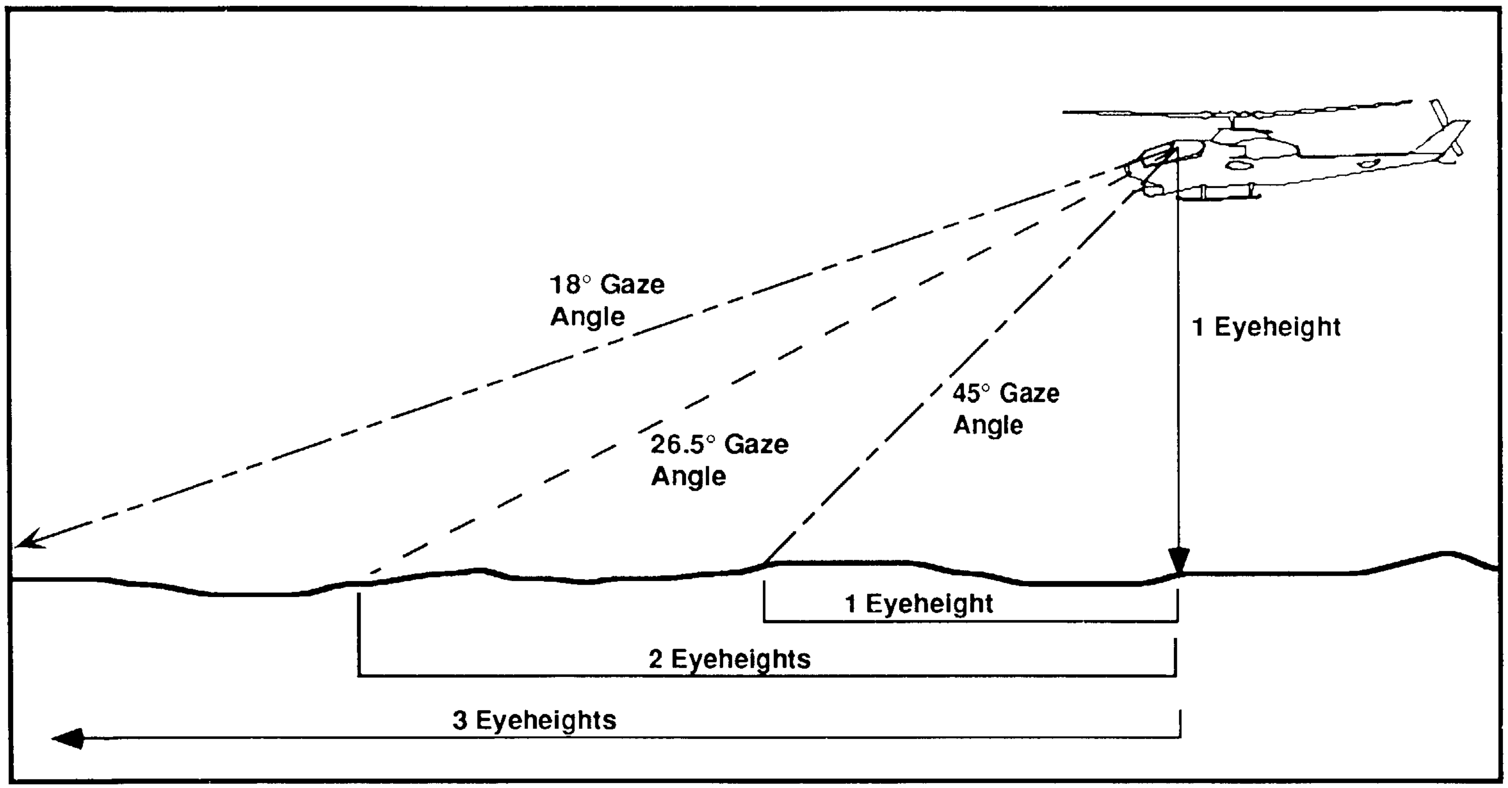

Figure 11. Eyeheight geometry. Distance along a given gaze angle is constant in eyeheight units.

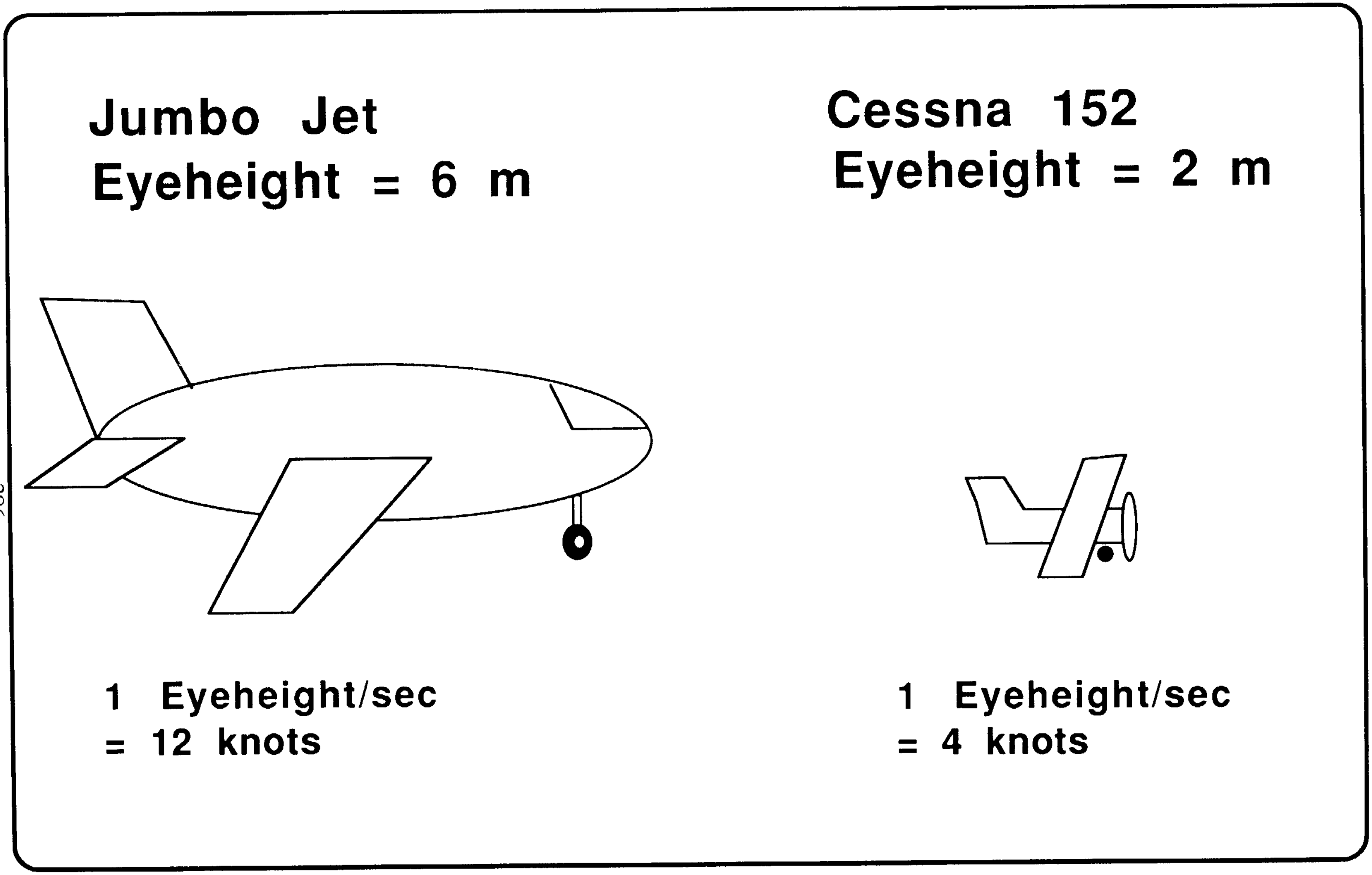

Figure 12. Speed corresponding to 1 eyeheight/second for two sample eyeheights.