Calibration algorithm for eyetracking with unrestricted head movement

Jeffrey S. Johnson, Li Liu, Geb Thomas, and John P. Spencer
University of Iowa, Iowa City, Iowa

This article reports a calibration procedure that enables researchers to track movements of the eye while allowing relatively unrestricted head and/or body movement. The eye–head calibration algorithm calculates fixation point based on eye-position data acquired by a head-mounted eyetracker and corresponding head-position data acquired by a 3-D motion-tracking system. In a single experiment, we show that this procedure provides robust eye-position estimates while allowing free head movement. Although several companies offer ready-made systems for this purpose, there is no literature available that makes it possible for researchers to explore the details of the calibration procedures used by these systems. By making such details available, we hope to facilitate the development of cost-effective, nonproprietary eyetracking solutions.

Typical visual environments contain enormous amounts of information. Rather than attempting to process the full extent of this information in detail, the visual systems of most organisms have evolved to sequentially inspect only small portions of the visual world at a time (Treue, 2001). This is achieved primarily by movements of the eyes to regions of interest in the environment. Orienting the visual system in this way serves the purpose of centering these areas over the fovea, a region of the retina containing over 30,000 densely packed photoreceptors, which provides high-acuity information to the visual system.

The potential usefulness of recording and analyzing movements of the eyes while participants perform different tasks has been recognized for at least 100 years (see Richardson & Spivey, 2004). In the field of psychology, it has become clear that sequential fixations of the eyes are closely related to shifts of spatial attention (see Findlay & Gilchrist, 1998; Hoffman & Subramaniam, 1995), and their duration is correlated with cognitive processing in a number of different domains. For instance, the efficient recording and analysis of eye movements has provided insights into scene perception (Henderson, 2003), reading and language processing (Just & Carpenter, 1980; Rayner, 1998), memory and imagination (Spivey & Geng, 2001), and the use of memory in the performance of complex tasks (Ballard, Hayhoe, & Pelz, 1995).

Beginning in the early 1970s, a number of techniques were developed that allowed the efficient, noninvasive tracking of the eye and were tolerant to small movements of the head. However, the state of the art at the time, as reflected in systems such as the dual-Purkinje-image eyetracker (Cornsweet & Crane, 1973) and pupil/corneal-reflection eyetracking technologies, still necessitated that the head be immobilized by a chinrest or bite bar. This requirement limited the usefulness of these systems in

contexts in which natural head and/or body movements were desirable or unavoidable. It also limited the use of such technologies with children, who find it difficult to sit still for extended periods of time.

Recent technological advances in this area have effectively addressed this problem, and several ready-made systems are now available that provide high-precision point of regard (POR) data while allowing natural head and body movements (e.g., see systems by ISCAN, Inc., Applied Science Laboratories [ASL], Inc., and SensoMotoric Instruments). However, there is currently no literature available that allows researchers interested in eyetracking to explore the details of the calibration procedures used by these systems. Access to such details could be beneficial in at least three ways. First, access to nonproprietary eye- and head-tracking solutions would greatly facilitate the development and refinement of these procedures by allowing open access to the details of the calibration process. Second, a thorough understanding of how these procedures work could improve the quality of the research conducted using these methods, by suggesting how experiments might be modified to maximize the precision of existing systems. Finally, although many ready-made systems are relatively inexpensive (around $15,000–$30,000 for combined head- and eyetracking systems), such costs might be prohibitive for some would-be eye-movement researchers. One way to reduce these costs is to use existing technologies to solve the challenges of head-movement-tolerant eyetracking "by hand." At present, a head-mounted eyetracker (e.g., ISCAN Inc. Model RK-726 PCI) can be purchased for around $3,000–$4,000, and motion-tracking systems that would be suitable for the present purpose (e.g., Polhemus Inc. FASTRAK motion-tracking system) are available for under $10,000. In our case, we used an ISCAN head-mounted eyetracker

---

J. S. Johnson, jeffrey-johnson-2@uiowa.edu

and a high-end (around $50,000) motion-analysis system (OPTOTRAK), but we lacked an effective calibration algorithm that would make it possible to bring these technologies together to compute POR in a task in which free head movement was allowed.

The present article provides the details of the calibration algorithm we developed. Our hope is that the solution we developed will be useful for other researchers already equipped with some means of tracking head and/or eye movements. In these cases, the proposed algorithm might make head-movement-tolerant eyetracking more affordable, because one piece of the technological puzzle (e.g., a motion-analysis system) is already in place, and the missing piece (e.g., a head-mounted eyetracker) is relatively inexpensive. Moreover, we hope publication of this algorithm will help foster an open-access community in which to refine and extend such methods.

Overview of Lab Setup and Calibration Algorithm

The algorithm was developed in the context of experiments in our lab, and therefore, this section begins with a brief overview of the equipment and setup present in the lab, followed by an overview of the proposed calibration algorithm. The experimental context described in the following section can be seen in Figure 1.

Experiments in our lab require participants to view stimuli and point to target locations projected onto a large tabletop—a rear-projection surface that appears opaque when viewed from the participants' perspective, but that allows us to project images onto the surface from below. Pointing movements are tracked using an optical electronic motion-analysis system (OPTOTRAK, Northern Digital, Inc.) that tracks small infrared-emitting diodes (IREDs) that can be attached to the finger or other surfaces. The experimental table is quite large (0.921 m $\times$ 1.194 m), and the surface is homogeneous. The participants sit in a chair
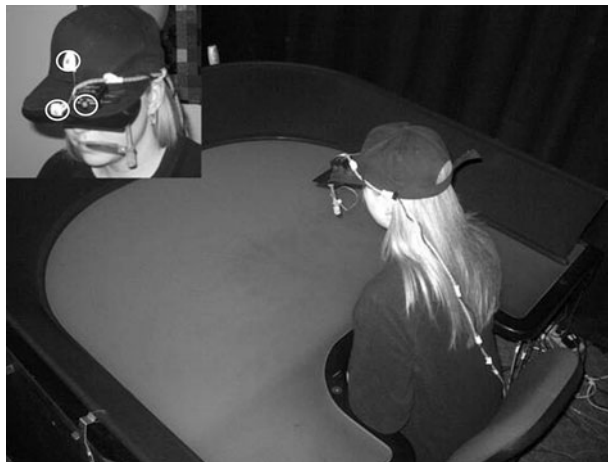

Figure 1. A photograph of a research participant wearing the head-mounted eyetracking unit and seated at a display table used in the experiments in our lab. Inset: Frontal view of the eyetracker with the attached infrared-emitting diodes (IREDs, circled in white) used to track movements of the head.

positioned within an arc cut out from one edge of the table, and view stimuli presented on the surface of the table directly in front of them. For studies in which eye movements are monitored, the participants wear a baseball cap with a small infrared camera mounted to the brim that records the position of the eye and transmits this information to an eyetracking system (ISCAN, Inc. Model RK-726 PCI) installed on a local computer. Additionally, the OPTOTRAK motion-analysis system is used to track movements of the head, which is made possible by affixing three IREDs to the brim of the baseball cap (see Figure 1 inset). Under these experimental conditions, it is reasonable to assume that the participant's POR always lies in the plane defined by the surface of the table. In other words, the fixation point is the intersection of the participant's instantaneous line of sight and the horizontal visual display surface.

The eye–head calibration procedure reported here estimates the necessary parameters for the calculation of fixation point based on 12 samples of visual angles acquired from the head-mounted eyetracking system and their corresponding head positions acquired by the 3-D motion-tracking system. In particular, the calibration algorithm uses estimates of visual angle in azimuth and elevation and the global coordinates of the three IREDs on the visor that determine the head's position and orientation with respect to the visual display (in our case, the horizontal table). This information alone is not enough, however, to determine the 3-D ray representing the line of sight. The other required parameters are the spatial relationship between the visor reference frame and the eye reference frame. This relationship can be expressed as a linear transformation represented by a rotation matrix and a translation matrix. These parameters can be estimated by analyzing a series of visor positions and visual eye angles obtained when the participant fixates targets at known positions in the global reference frame. After this estimation, a nonlinear least square optimization method is used to calculate the parameters. The line of sight and the ultimate fixation point are then calculated by making use of the acquired parameters.

Plan of the Present Article

The primary goal of the present article is to report and test an algorithm developed to accurately determine participants' PORs without requiring head-movement restraint. The remainder of the article is organized as follows. First, we describe the proposed eye–head calibration algorithm and nonlinear least squares optimization procedure. Second, we test the effectiveness of the proposed method in an experiment. The results demonstrate that the eye–head calibration algorithm reported here serves as an effective procedure for calculating POR without requiring head-movement restraint. Finally, we discuss limitations of the proposed solution and future approaches that might improve the algorithm, as well as the generalizability of the algorithm to experimental contexts other than our own.

Eye–Head Calibration Algorithm

The eye frame calibration algorithm described in this section requires accurate data for movements of the eye and the head, as well as the ability to transform target locations

measured in a global reference frame into head- and eye-centered coordinates. Effective use of this algorithm, therefore, begins with the calibration of the eyetracking system. The details of this procedure will differ, depending on the specific model of eyetracker being used (calibration of the ISCAN RK-726 PCI is described in a later section). After the eyetracking system has been calibrated, it can be used to acquire samples of visual angles while participants view a series of targets placed at known locations in the global reference frame (e.g., on the surface of a display table). The eye-position data and corresponding head positions while participants view each target can then be used to calculate the line of sight and the ultimate fixation point using the nonlinear least squares optimization procedure. This section begins with a description of the three relevant frames of reference and the translations among them, followed by an account of the nonlinear least squares procedure.

Three Coordinate Systems: Global Frame, Head Frame, and Eye Frame

Three different reference frames are involved in the present problem: a fixed, global reference frame, a head-centered reference frame, and an eye-centered reference frame. The *global reference frame* is the coordinate system defined with respect to fixed objects in the room—in particular, the horizontal surface of the display table and the position of the 3-D motion-tracking camera mounted to the laboratory ceiling. The calibrated fixation points and the three IREDs attached to the eyetracker's visor are all measured in the global frame. Note that other types of head-tracking systems also provide information about the position of markers on the head (i.e., the head reference frame) relative to a fixed global frame of reference. Moreover, the locations of targets in this global frame can easily be determined by placing sensors (e.g., IREDs) at the target locations ahead of time, to identify their spatial coordinates within the global frame.

The *head-centered reference frame* is defined by the positions of the three IREDs (P1, P2, P3) affixed to the eyetracker's visor (see Figure 2). The position of each IRED is given in Cartesian coordinates in the global reference frame (e.g., P1 = $[x_{P1}, y_{P1}, z_{P1}]^{T}$). The origin of the head-centered reference frame, defined by three Cartesian axes (x_h, y_h, z_h), is the geometric center of these three points. The x_h-axis extends from the origin to P1. The y_h-axis is perpendicular to x_h and normal to the plane containing P1, P2, and P3. y_h is defined as the normalized cross-product of (P1 − P3) × (P2 − P3). The z_h-axis is defined by the cross-product of x_h and y_h.

The *eye-centered reference frame*, defined by three Cartesian axes (x_e, y_e, z_e), is located at the center of the eye and has a fixed translation and rotation with respect to the head-centered reference frame. This translation and rotation are unchanged even when the eye or head is moving. The z_e-axis lies along the optical axis when the participant is looking directly forward. The x_e-axis points leftward along the line defined by the center of both eyes. The y_e-axis points vertically upward, perpendicular to the other two axes. Figure 3 illustrates a POR to a target (t_{ex}, t_{ey}, t_{ez}) in the eye frame and the corresponding azimuth,
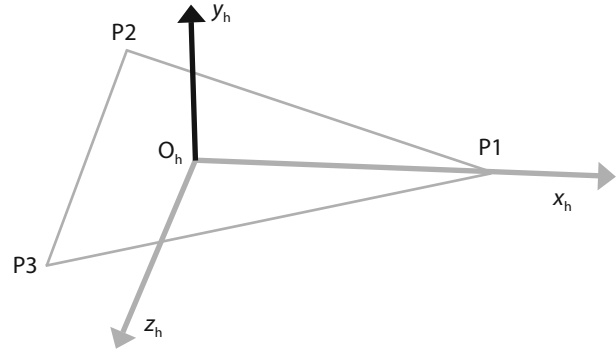
Figure 2. The head-centered coordinate system formed by the three IREDs (points P1, P2, and P3) attached to the brim of the eyetracking visor. The position of each IRED is given in Cartesian coordinates in the global reference frame (e.g., P1 = $[x_{P1}, y_{P1}, z_{P1}]^{T}$). The origin of the head-centered reference frame defined by three Cartesian axes (x_h, y_h, z_h) is the geometric center of these three points. The x_h-axis extends from the origin to P1. The y_h-axis is perpendicular to x_h and normal to the plane containing P1, P2, and P3. The z_h-axis is defined by the cross-product of x_h and y_h.

ϕ_t, and elevation, θ_t, angles measured by the eyetracking device. Note that these angles are different than the Euler angles we describe below, which determine the transformation between the head- and eye-coordinate systems.

Transformations Among the Three Coordinate Systems

An isometric transformation between two arbitrary coordinate systems can be defined by a 3×3 rotation matrix, **R**, and a 3×1 translation vector, **T**. For example, a point in the coordinate system A ($\mathbf{P}_A = [x_A, y_A, z_A]^{T}$) can be transformed to coordinate system B ($\mathbf{P}_B = [x_B, y_B, z_B]^{T}$) by the following equation.

$$\mathbf{P}_B = \mathbf{R} \times \mathbf{P}_A + \mathbf{T}$$

$$\mathbf{P}_B = \begin{bmatrix} x_B \\ y_B \\ z_B \end{bmatrix} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix} \times \begin{bmatrix} x_A \\ y_A \\ z_A \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

By introducing a dummy number "1" into the coordinate vector, the above transformation is equivalent to the following equation:

$$\mathbf{P}_B = \begin{bmatrix} x_B \\ y_B \\ z_B \\ 1 \end{bmatrix} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_A \\ y_A \\ z_A \\ 1 \end{bmatrix}$$

The latter expression is convenient because it combines the rotation matrix and a translation vector into a single 4×4 transformation matrix.

A series of transformations among the three coordinate frames described above—global frame, head frame, and eye frame—can determine the coordinates of the POR in the global frame based on the real-time visual angles provided by an eyetracker and real-time information about
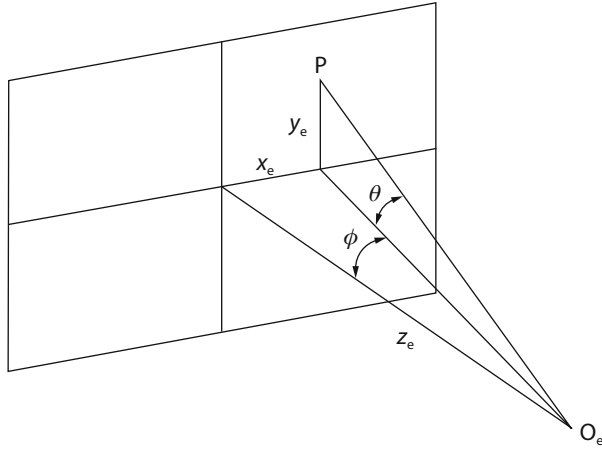
Figure 3. Illustration of a fixation point, P, with coordinates x_e, y_e, and z_e in the eye-centered reference frame with corresponding azimuth, ϕ_t, and elevation, θ_t, angles.

head location, assuming that all of the PORs lie on a known surface. Each transformation is defined by a 4×4 matrix, $_A\mathbf{T}_B$ (see Hapgood, 1992, for a comprehensive description of such transformations). A transformation matrix, $_A\mathbf{T}_B$, can transform any point or vector, $\mathbf{P}_B$, defined with respect to coordinate frame B, to $\mathbf{P}_A$, the same position or vector represented in coordinate frame A. Let $\mathbf{P}_B$ and $\mathbf{P}_A$ be 4×1 column vectors $[\mathbf{P}_x, \mathbf{P}_y, \mathbf{P}_z, 1]^\mathbf{T}$. Then:

$$\mathbf{P}_A = {_A\mathbf{T}_B}*\mathbf{P}_B.$$

The head-tracking system provides an estimate of the head-centered reference frame's position in the global coordinates, defined by $_g\mathbf{T}_h$, the transformation matrix from the head coordinate frame to the global coordinate frame. Head movements change the components in $_g\mathbf{T}_h$. In the system described here, the components of $_g\mathbf{T}_h$ may be measured directly using P1, P2, and P3 (the positions of the sensors on the visor).

Assuming that neither the participant nor the researcher shifts the position of the visor on the participant's head during the experimental session, the sensors on the visor will reflect the head position, and the transformation, $_h\mathbf{T}_e$, between the head-centered reference frame and the eye-centered reference frame should remain constant. The components of this transformation may not be directly measured, but must be derived indirectly from a calibration process.

Nonlinear Least Squares Method

The eye frame calibration procedure uses a nonlinear least squares method to determine the transformation matrix, $_h\mathbf{T}_e$, from the eye-centered reference frame to the head-centered reference frame. All 16 elements of this matrix can be determined from six unknowns: the position of the center of the eye in the head coordinate frame (x_e, y_e, z_e) and three rotation angles (θ, ϕ, and ψ). The three angles are the Euler angles: yaw, θ; pitch, ϕ; and tilt, ψ, representing the rotation of the eye-centered reference frame around the y-, x-, and z-axes of the head-centered reference frame, respectively. These parameters are de-

termined by iteratively improving their estimation while minimizing the error between the predicted and observed gaze angle as the observer fixates a set of known target positions. The optimization is implemented using MATLAB software (version 6.5—source code can be obtained from the Psychonomic Society Archive of Norms, Stimuli, and Data at www.psychonomic.org/archive). In addition to the manufacturer-recommended eyetracker calibration, the procedure has seven steps.

(1) Collect data. A consistent set of data must be collected that includes the head position and visual angles as the participant fixates a series of known target positions. Theoretically, three fixation points are enough to calculate the six unknown parameters, since each fixation point has 2 *df* and gives two equations to calculate the unknowns. In practice, however, it is advisable to use more target positions to cancel the effects of noise, because each point has measurement errors associated with the systems used to measure the position of the head and visual angles. To randomize the noise in the data, both the target positions and the head positions should be varied across the range that will be used during the experiment. In the experiment described below, data were collected for 12 gaze positions. At each fixation position, we recorded the target's global coordinate (t_{gx}, t_{gy}, and t_{gz}), the global coordinate of each of the three visor IREDs (P1, P2, and P3), and the visual angles (ϕ_t and θ_t) from the eyetracking system.

(2) Calculate $_g\mathbf{T}_h$, the transformation from the head-centered reference frame to the global frame. Although different head-tracking systems may provide this information in different ways, here we describe the calculation necessary to derive the transformation using the three points on the visor.

$$_g\mathbf{T}_h = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \sum_{i=1}^{3} Pi/3, \begin{bmatrix} r_{1,1} \\ r_{2,1} \\ r_{3,1} \end{bmatrix} = \frac{P1 - \sum_{i=1}^{3}(Pi/3)}{\text{norm}\left[P1 - \sum_{i=1}^{3}(Pi/3) \right]},$$

$$\begin{bmatrix} r_{1,2} \\ r_{2,2} \\ r_{3,2} \end{bmatrix} = \frac{(P1 - P3) \times (P2 - P3)}{\text{norm}\left[(P1 - P3) \times (P2 - P3) \right]},$$

$$\begin{bmatrix} r_{1,3} \\ r_{2,3} \\ r_{3,3} \end{bmatrix} = \begin{bmatrix} r_{1,1} \\ r_{2,1} \\ r_{3,1} \end{bmatrix} \times \begin{bmatrix} r_{1,2} \\ r_{2,2} \\ r_{3,2} \end{bmatrix}$$

(3) Provide initial estimates of the six transformation parameters from the head-centered reference frame to the eye-centered reference frame:

$$\left(x'_e, y'_e, z'_e, \theta'_e, \phi'_e, \psi'_e \right).$$

$$
{}_h\mathbf{T}_e = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

$$
\begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix} = \begin{bmatrix} \cos\psi'_e\cos\theta'_e & \sin\psi'_e\cos\theta'_e & -\sin\theta'_e \\ -\sin\psi'_e\cos\phi'_e + \cos\psi'_e\sin\theta'_e\sin\phi'_e & \cos\psi'_e\cos\phi'_e + \sin\psi'_e\sin\theta'_e\sin\phi'_e & \cos\theta'_e\sin\phi'_e \\ \sin\psi'_e\sin\phi'_e + \cos\psi'_e\sin\theta'_e\cos\phi'_e & -\cos\psi'_e\sin\phi'_e + \sin\psi'_e\sin\theta'_e\cos\phi'_e & \cos\theta'_e\cos\phi'_e \end{bmatrix}
$$

$$
\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} x'_e \\ y'_e \\ z'_e \end{bmatrix}
$$

(4) Determine $_h\mathbf{T}_e$ from the estimates in Step 3 (see the expression at the top of the page).

(5) Calculate target locations in the eye frame. Given a target position $t_g = [t_{gx}, t_{gy}, t_{gz}, 1]$ in the global frame, $_g\mathbf{T}_h$, and $_h\mathbf{T}_e$, a target point in the eye frame, $t_e = [t_{ex}, t_{ey}, t_{ez}, 1]$, is given by:

$$
t_e = {}_h\mathbf{T}_e^{-1}\,{}_g\mathbf{T}_h^{-1}\,t_g,
$$

where the negative exponent indicates the inverse matrix operation.

(6) Calculate the visual angle for each fixation point (ϕ'_t and θ'_t) based on the coordinates of these points in the eye frame $[t_{ex}, t_{ey}, t_{ez}]$.

$$
\phi'_t = \tan^{-1}\left(-\frac{t'_{ex}}{t'_{ez}}\right)
$$

$$
\theta'_t = \sin^{-1}\left(\frac{t'_{ey}}{\sqrt{t'^2_{ex} + t'^2_{ey} + t'^2_{ez}}}\right)
$$

(7) Iteratively refine the estimates of (x'_e, y'_e, z'_e, θ'_e, ϕ'_e, ψ'_e) by minimizing the sum of the squares of the predicted and measured visual angles:

$$
\min\left[\sum_{\text{all targets}}(\theta_t - \theta'_t)^2 + (\phi_t - \phi'_t)^2\right].
$$

This optimization can be accomplished using the nonlinear estimation function provided in the MATLAB Optimization Toolbox, which can be purchased as an add-on to MATLAB from The MathWorks (www.mathworks.com) for around $200.

Determining the POR in the Global Coordinate Frame

This calibration determines $_e\mathbf{T}_h$, the transformation matrix from the head-centered reference frame to the eye-centered reference frame, which is not expected to change during the experiment. During the experiment, the eyetracker provides (ϕ_e, θ_e), the line of sight in the eye frame. The line of sight in the global frame may be determined by selecting two arbitrary points on the line of sight in the eye frame, translating the two points into the head frame, and then translating them into the global frame. This results in a line of sight expressed in the global frame. Because the z-coordinate in the global frame is zero for every fixation point (i.e., all fixation points lie within the plane of the display table), the fixation point in the global frame can be calculated using the line of sight. The following steps specify this sequence of calculations:

1. Select two arbitrary points A and B along the line of sight in the eye frame, where A is 1 unit from the origin and B is 10 units from the origin. Calculate their coordinates in the eye frame

$$
A_e = \begin{bmatrix} A_{ex} \\ A_{ey} \\ A_{ez} \\ 1 \end{bmatrix} = \begin{bmatrix} -\cos\theta_e\sin\phi_e \\ \sin\theta_e \\ \cos\theta_e\cos\phi_e \\ 1 \end{bmatrix},\ B_e = 1 * \begin{bmatrix} 10 * A_{ex} \\ 10 * A_{ey} \\ 10 * A_{ez} \\ 1 \end{bmatrix}.
$$

2. Translate the two points to the global frame

$$
A_g = {}_g\mathbf{T}_h\,{}_h\mathbf{T}_e\,A_e
$$
$$
B_g = {}_g\mathbf{T}_h\,{}_h\mathbf{T}_e\,B_e.
$$

3. Calculate the line of sight in the global frame

$$
\frac{x - B_{gx}}{A_{gx} - B_{gx}} = \frac{y - B_{gy}}{A_{gy} - B_{gy}} = \frac{z - B_{gz}}{A_{gz} - B_{gz}}.
$$

4. In the global frame, the fixation point $\mathbf{T}$ is along the line of sight and in the plane of the display table with $z = 0$.

$$
\mathbf{T}_g = \begin{bmatrix} \mathbf{T}_{gx} \\ \mathbf{T}_{gy} \\ \mathbf{T}_{gz} \\ 1 \end{bmatrix} = \begin{bmatrix} B_{gx} - \dfrac{B_{gz}}{A_{gz} - B_{gz}}\left(A_{gx} - B_{gx}\right) \\ B_{gy} - \dfrac{B_{gz}}{A_{gz} - B_{gz}}\left(A_{gy} - B_{gy}\right) \\ 0 \\ 1 \end{bmatrix}
$$

Similar functions can be calculated for other geometric conditions by finding the intersection of the line of sight with the specific display surface being used. For instance,

if the display surface was vertical (as is the case in many laboratories), the above equation could be modified, setting $y = 0$ instead of $z = 0$.

Testing the Eye–Head Calibration Procedure

We tested the effectiveness of the calibration procedure in three steps. First, the eyetracker was calibrated by recording eye-position data while each participant viewed a series of targets at known locations with his or her head fixed in a chinrest. Second, we completed the eye–head calibration procedure described above. Third, we tested the robustness of the calibration by having participants fixate a series of targets presented on the surface of the display table with unrestricted head movement. We used the difference between the actual locations of the targets and the estimated POR provided by the eye–head calibration as the primary index of the effectiveness of the procedure.

METHOD

Participants

The participants were 11 undergraduate and graduate students, 8 of which were working in our lab, and 3 of which were recruited from other labs within the University of Iowa Department of Psychology. The participants' ages ranged from 18 to 35 years, with a mean age of 28 years. All reported having normal or corrected-to-normal visual acuity. All of the participants gave informed consent before participating in this study.

Apparatus

Eyetracking. The participants' left eyes were tracked with a head-mounted, ISCAN, Inc. Model RK-726 PCI dark-pupil-to-corneal-reflection eyetracking system (accuracy = 0.5º, sample rate = 60 Hz) interfaced with a PC driven by a Pentium IV processor. The RK-726 PCI was configured to accept an NTSC signal from the eye camera mounted on the cap's visor, and the eye image was displayed on a 15-in. Sony Trinitron monitor. Visual angles from the eyetracker were sent in analog form to a data acquisition unit (ODAU, Northern Digital, Inc). This unit sampled the eyetracking signal at 150 Hz in sync with data from the OPTOTRAK system (see below).

Head tracking. Movements of the participants' heads were recorded with an optical electronic motion-analysis system (OPTO-TRAK, Northern Digital, Inc.). OPTOTRAK tracks small (radius = 3.5 mm), individually pulsed IREDs within a specified 3-D volume with better than 1-mm precision. Three of these IREDs were attached to the brim of the head-mounted eyetracker (see Figure 1 inset). Head-position data were collected at 150 Hz in a predefined coordinate system (the global reference frame). The origin of this coordinate system was centered at an arbitrary start location on the surface of the display table used in experiments in our lab. The x-axis goes from left to right. The y-axis goes from the front to the back of the table. The z-axis is perpendicular to the plane of the table and goes upward.

Display table and stimuli. Throughout each of the following procedures, the participants were seated on a chair placed within an arc cut out from the side of a large (0.921 m × 1.194 m) table (see Figure 1). The surface of the table was homogeneous, and three edges of the table were covered with a curved border to occlude the corners from the participants' view. Sessions were conducted in a dimly lit room with black curtains covering the walls and ceiling. Stimuli were projected onto the surface of the table from below using a Barco 708 data projector. The screen resolution was 1,024 × 768 pixels.

Stimuli used in the first calibration phase consisted of five small black dots printed on a white background and mounted to an appara-

tus that was attached to the chinrest. The five-dot array consisted of a single central dot surrounded by four additional dots; one directly above, one directly below, one directly to the left, and one directly to the right of the central dot. Each of these dots was positioned 7º of visual angle away from the central dot at a viewing distance of 30 cm.

Stimuli used in the eye–head calibration phase consisted of small white target dots (luminance = 0.1462 cd/m²; color index on the CIE 1932 system was $x = 0.266$, $y = 0.156$). The participants viewed three different arrays consisting of four dots arrayed in a cross, for a total of 12 targets. The first array was presented in the upper left corner of the display area (x-, y- target coordinates in pixels = [384, 320], [512, 312], [448, 256], [448, 384], where [0, 0] = the top left corner of the display table), the second was presented in the center of the display area (target coordinates = [448, 384], [576, 384], [512, 320], [512, 448]), and the third array was presented in the lower right of the display area (target coordinates = [512, 448], [640, 448], [576, 384], [573, 512]). We varied the spatial position of these arrays to ensure good coverage of the likely display area used in experiments in our lab, and to sample across a wide range of head positions.

Stimuli used in the calibration test phase were the same size, luminance, and color as those used for the eye–head calibration. For the test phase, the participants viewed a total of 16 dots positioned in a 4 × 4 array on the surface of the table. Targets 1–4 occupied the top row of the array (x-, y-coordinates in pixels = [205, 154], [410, 154], [615, 154], [820, 154]), targets 5–7 were in the second row ([205, 308], [410, 308], [615, 308], and [820, 308]), targets 8–11 were in the third row ([205, 462], [410, 462], [615, 462], and [820, 462]), and targets 12–16 appeared in the bottom row ([205, 616], [410, 616], [615, 616], and [820, 616]).

Procedure

Eyetracker calibration. Each experimental session began by calibrating the RK-726 PCI eyetracker using the manufacturer-recommended procedure. The participants were seated at the display table with the eyetracker positioned to track the left eye. Next, each participant placed his or her head in the chinrest and sequentially fixated each of five targets while we recorded their eye positions using the eyetracking system. The participants were instructed to keep their heads as still as possible and to carefully fixate each target without blinking until prompted to move on to the next target location. The experimenter viewed each eye fixation on the display monitor. For each target, the experimenter waited for the eye to stabilize at the fixated location before pushing a button on the keyboard, thereby capturing the eye-position data. The participants were then prompted to move their eyes to the next target location, and the procedure was repeated.

Eye–head calibration. After successful calibration of the eyetracker, we removed the chinrest (thereby allowing unrestricted head movement) and completed the eye–head calibration procedure. In this phase, the participants fixated a series of dots on the display table. As before, the experimenter waited for the eye to become stable before pressing a button on the computer keyboard, capturing the eye-position data for that target. For each target, eye- and head-position data were collected across 29 successive frames within a 200-msec time window centered on the buttonpress. We used these data to compute mean (and SD) eye position (in degrees of visual angle) and mean (and SD) x-, y-, and z-coordinates (in mm) for the three IREDs attached to the visor. Note that we eliminated blinks from the eye-position data by excluding any visual angles outside of the range ±40º. These data were output to a text file along with the global x-, y-, and z-coordinates of each of the 12 fixated targets (in mm). This file was then read by the eye–head calibration algorithm running in MATLAB. The MATLAB code generated an error term reflecting the difference between the calculated visual angle and the visual angle recorded by the eyetracking system. The calibration phase was considered successful when the root mean square error (RMSE) across targets was < 0.30º of visual angle. Nine par-

ticipants were successfully calibrated on their first time through the eye–head calibration phase, whereas the 2 remaining subjects required one repetition of this phase before reaching criterion.

Eye–head calibration test. Following successful completion of the eye–head calibration procedure, we tested the robustness of the computed solution by having the participants sequentially fixate each of 16 different target dots (see above) presented in a random order on the surface of the display table. Eye- and head-position data were captured in the same way as described for the eye–head calibration phase.

RESULTS

We analyzed the test data with four primary goals in mind. First, we wanted to assess the overall effectiveness of the calibration procedure within the experimental context described above. Second, given the large size of our task space, we wanted to determine the range of spatial locations across which the calibration procedure provided accurate POR estimates. Third, we were interested in how robust the procedure was across different experimental participants. Finally, we were interested in the range of head positions sampled in the present experiment to assess the tolerance of the calibration procedure to head movements. To examine these issues, we looked at both eye-position and head-position data while participants fixated each of the 16 targets used in the calibration test phase.

POR analyses. We calculated POR while each participant fixated the 16 calibration test targets using a modified version of the eye–head calibration algorithm running in MATLAB (code is available upon request). This algorithm required the same inputs described previously: mean eye- and head-position values during each fixation, along with the global x-, y-, and z-coordinates (in mm) of each test target. For each fixation, the MATLAB code generated an error term reflecting the difference (in degrees of visual angle) between the calculated POR and the actual location of the target. We used these error values to calculate the RMSE for each target across all participants and for each participant across all targets. In addition, we calculated an overall RMSE for both azimuth and elevation.

The full set of POR estimation errors is presented in Figure 4, panels A and B. As can be seen in Figure 4A, the algorithm's performance in azimuth was excellent overall, with RMSE = 0.93º (σ = 0.18º) across targets. Azimuth errors were generally consistent across targets with low variability, with the highest error produced for Target 1 (topmost left corner of the display table). Elevation errors were also reasonably accurate overall, producing an RMSE of 1.65º (σ = 0.34º). As with azimuth errors, elevation errors were generally commensurate across targets, with the notable exception of Targets 1 and 9, which were both at the far left edge of the task space. Note, however, that elevation errors were nearly twice the magnitude of those observed for azimuth, with twice the variability. This was the case across all targets, with the exception of Targets 2, 12, and 16, where azimuth and elevation errors were comparable. Although substantially higher than the azimuth errors, the obtained RMSE for elevation was relatively good, given the horizontal orientation of the display table. At this orientation, the participants' eyes were gen-

erally rotated downward. As this rotation angle increased, estimates of eye elevation from the eyetracker we used became less reliable. Despite this, the calibration algorithm was still able to estimate the target location in the task space with a resolution near 1.5º of visual angle.

Figure 4B shows the overall azimuth and elevation errors across the 11 participants. Once again, azimuth errors were lower and less variable than elevation errors. The azimuth error values were consistently good across all participants, with accuracy generally near 1.0º (overall RMSE = 0.95º and σ = 0.34º across participants). The largest errors occurred for Participant 5 (RMSE = 1.56º). In contrast to the azimuth data, elevation errors were more variable (RMSE = 1.6º, σ = 0.86º), with Participants 2 and 8 producing elevation errors nearly three times as large as the average error produced by the remaining 9 participants (2.83º and 3.26º, vs. 1.28º, respectively). Thus, elevation errors were particularly poor for 2 participants, but were nearly as impressive as azimuth errors for the remaining 9 participants.

Head-position data. To analyze the head-position data, we calculated the mean and range of head positions sampled while the participants fixated each of the calibration test targets. Recall that movement of the head within the global reference frame was tracked by an infrared camera that detected the outputs of three small IREDs attached to the brim of the eyetracker. We averaged the x, y, and z spatial position recorded from each IRED to obtain a single set of x, y, and z values that reflected the mean position of the head in relation to the display table while the participants fixated each of the targets. These data are shown in Figure 4, panels C and D. To demonstrate the robustness of our POR estimates across variations in head position, error bars reflect the range (max and min) of head positions sampled during the calibration test phase.

Average head position above the table (z-axis), as well as the participants' distance from the table (y-axis), remained relatively consistent across targets, whereas the left/right translation of the head (x-axis) varied substantially, depending on which target the participant was fixating. The ranges were highest along the y-dimension (toward/away from the table), where head positions ranged from −56.88 cm to −15.58 cm (average range = 38.39 cm, σ = 1.61). In contrast, ranges were lower across participants in the x- and z-dimensions, with mean ranges of 11.02 cm (σ = 3.25 cm), and 9.67 cm (σ = 0.71 cm), respectively. The substantially higher range for the y-dimension likely reflects the large difference between the average position of Participants 6 and 7, with the former sitting very close to the table, and the latter sitting relatively far from the table throughout the testing session (see Figure 4D).

Figure 4D shows the mean and range of head positions for each participant across all 16 calibration targets. Means along the z-axis primarily reflect differences in the height of participants, with the range for a given participant reflecting up/down movements of the head. Means along the y-axis reflect how close/far different participants were sitting from the display table, and the range reflects movements toward or away from the table. Not surprisingly,

POR Estimation Errors

Sampled Head Positions
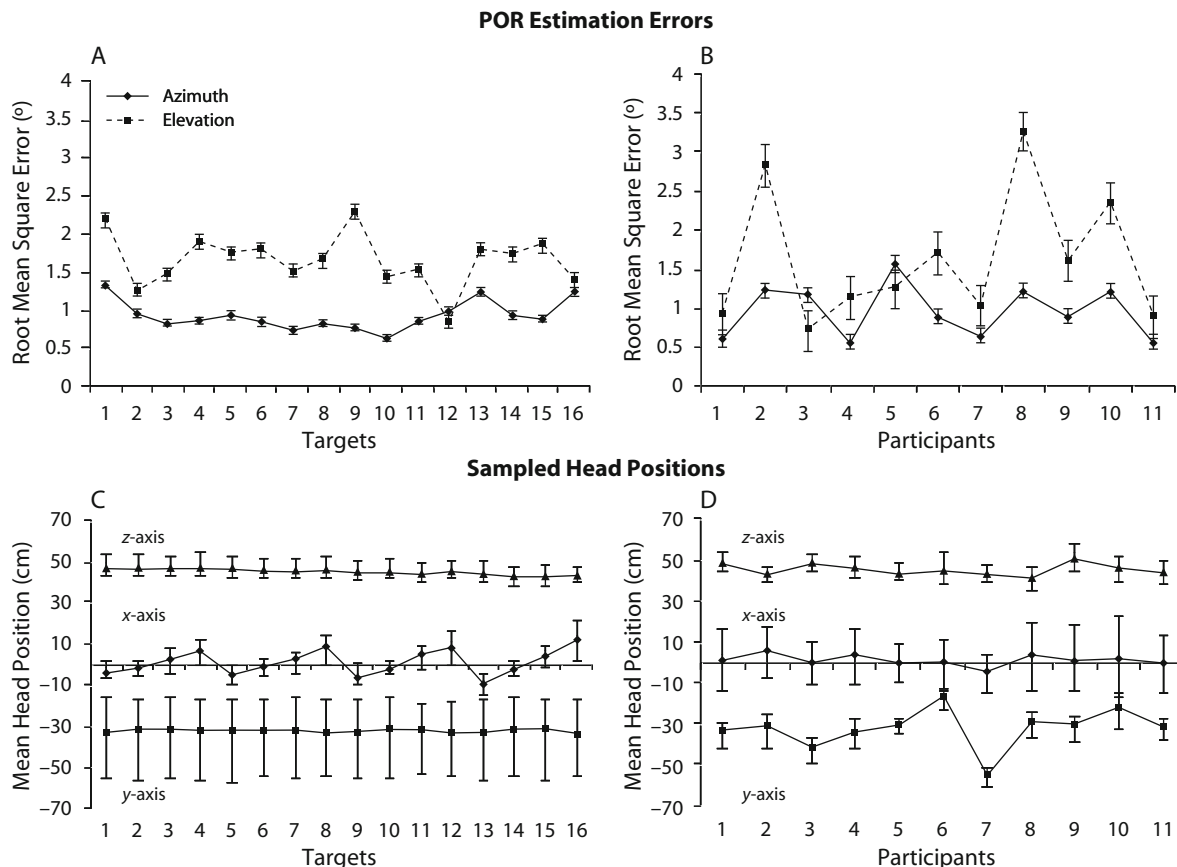
Figure 4. POR estimate error data and head-position data recorded at each target and for each participant in the calibration test phase. Panel A shows the root mean square error (RMSE) across participants for each of the 16 fixated targets. Panel B shows the RMSE across targets for each of the 11 participants. Error bars for these panels reflect the standard error of the mean. Panels C and D contain the associated head-position data across participants and targets, as above. Because the primary motivation for analyzing the head-position data was to demonstrate that the calibration procedure remains robust across a wide sampling of head positions, error bars for these data show the range. See the text for discussion.

head positions were sampled from a relatively narrow range along the y- and z-dimensions, with mean ranges across participants of 12.45 cm ($\sigma = 3.14$) and 10.82 cm ($\sigma = 2.53$ cm), respectively. Head position was sampled most widely along the x-dimension, because participants moved their heads from left to right to facilitate fixations of the calibration targets. Head positions along this dimension ranged from −17.36 cm (for the leftmost targets) to 22.81 cm (for the rightmost targets), with an average range of 27.11 cm—a relatively large range, given that participants were sitting at a fixed location viewing targets on a tabletop.

To explore the relationship between POR estimation errors and variation in head position, we conducted two multiple regression analyses predicting POR error estimates (for both azimuth and elevation), using the range of head movements in the x-, y-, and z-dimensions as predictors. The regressions accounted for very little variance in azimuth and elevation estimation errors ($R^2_{adj} = -392$ and .096, respectively), with nonsignificant overall R^2 values in both cases [$F(3,10) = 0.061$, n.s., and $F(3,10) = 1.36$, n.s., respectively]. Moreover, none of the individual predictors were

significantly related to POR error estimates in either analysis. In short, we found no systematic relationship between variation in head position and POR estimation errors.

DISCUSSION

The calibration procedure reported here provides a robust and accurate estimate of participants' PORs within the experimental context used in our laboratory. This estimate remained reliable over a substantial viewing area on the surface of our large display table, despite the fact that unrestricted movements of the head were allowed throughout the calibration test phase. The algorithm performed remarkably well in estimating azimuthal PORs, with accuracy better than 1° of visual angle across all participants and targets. POR estimates for elevation were also relatively accurate, particularly given the orientation of our display surface. How does the accuracy of this eye- and head-tracking system compare with other ready-made systems? A number of commercially available systems (i.e., ISCAN, Inc., ASL, Inc.) report system accuracy of around 0.5° in both dimensions, while allowing complete,

unrestricted movements of the head (specifications for a wide range of eye- and head-tracking solutions offered by ASL, Inc. can be found at www.a-s-l.com). Thus, the system reported here gets within the ballpark of these state-of-the-art systems but is less accurate overall, particularly with regard to the estimation of elevation POR.

A critical question is whether the errors reported here were due to errors in our calibration algorithm, as opposed to errors stemming from the OPTOTRAK and/or eyetracking systems. One method of estimating the error introduced by the OPTOTRAK system is to contrast the actual distances among the three IREDs fixed to the eyetracker's visor with the computed IRED distances derived from the real-time motion data. The actual distances between the three IREDs were 91 mm, 88 mm, and 43 mm. These values varied by 0.6 mm, 0.4 mm, and 0.2 mm, respectively, during the calibration test phase. Based on these errors, we estimate that OPTOTRAK contributed around 0.5º of visual angle error. As mentioned previously, the eyetracker reportedly produces an overall error of approximately 0.5º. Thus, together, the eyetracker and the OPTOTRAK produce a total estimation error of approximately 1º of visual angle. This value accounts for the entirety of the estimation error for azimuth reported here, but leaves approximately 0.65º of error in elevation estimates unaccounted for. Which of these two systems could be contributing to the additional estimation error?

There are several ways in which the eyetracking and OPTOTRAK systems could have contributed to larger errors in estimating elevation. First, as suggested previously, the eyetracker could become less reliable at larger eye rotations. Given the horizontal orientation of the display table, the eyes would tend to be rotated downward during the testing session, which might lead to larger overall elevation errors. Second, errors could arise if there were a shift in the relative position of the IREDs mounted on the eyetracker's visor and the participant's head. For instance, a fit loose enough to allow the visor to shift by only 1º during the testing session would produce systematic errors in estimating POR that would account for the 0.65º of error in elevation estimates unaccounted for. The likelihood that such visor shifts could have a differential impact on elevation error estimates seems high, given the uneven distribution of weight on the front of the visor, where the infrared camera and the eyepiece are mounted, and the presence of wires hanging from the back of the visor. Although small visor shifts cannot be avoided completely, especially during lengthy experimental sessions, their occurrence can be minimized by ensuring that the visor is securely fastened to the participant's head at the beginning of the session, and by using hairpins or some other means to limit slippage. Additionally, Hornof and Halverson (2002) have developed a method that reduces systematic error in eye-movement data by having participants fixate a number of required fixation locations during the course of the experiment and using these data to correct eye-position data offline. Use of this error-correction method together with the eye-calibration procedure developed here could largely eliminate systematic POR estimation errors caused by movement of the eyetracking visor.

Finally, the arrangement of the IREDs on the eyetracker's visor, relative to the motion-tracking cameras, may have limited the OPTOTRAK system's ability to accurately determine head elevation. This might occur because the relative change in the vertical spacing of the IREDs from the perspective of the camera is much smaller for vertical head motions than for horizontal head motions. Given that the cameras have a fixed angular resolution, a small relative motion would lead to less precision in the elevation angle. On the basis of this analysis, it is likely that elevation errors could be reduced by either repositioning the IREDs on the eyetracker's visor to achieve greater vertical spacing, or by placing the motion-tracking cameras above, rather than in front of, the participant. Additionally, given these considerations, we would expect that overall estimation errors, and elevation errors, in particular, would likely be reduced if one were to use a standard, vertically oriented CRT or LCD monitor to present stimuli.

We are still in the process of determining which of these possibilities contributed to the relatively high elevation errors in the calibration test phase. Nevertheless, the detailed discussion of these alternatives illustrates one of the central points of this article: Having access to the details of the calibration algorithm underlying eyetracking can suggest a number of possibilities for optimizing eyetracking accuracy within the constraints of a particular experimental setup. Such refinements are not nearly as obvious when the details of the calibration procedure remain a mystery, as they do with most ready-made eye- and head-tracking solutions.

Finally, we should mention the difficulties that can sometimes arise in getting the optimization algorithm to converge on a particular solution, and how these difficulties can be minimized. The convergence on a particular solution is related to the initial estimate of the six unknown parameters and the upper and lower bounds defining the range of acceptable values, and the number of sampled fixation points. Because there may be more than one solution to the nonlinear least square, it is possible for the algorithm to converge on a solution that is not reasonable. The likelihood of convergence on a reasonable value can be increased by limiting the allowable search region of the algorithm to a practically reasonable range. Although determination of what constitutes a reasonable range can be a somewhat subjective decision, we did not find this aspect of the calibration procedure to be problematic. Additionally, the calibration algorithm is very sensitive to the spatial distribution of the calibration targets, and must sample a wide range of head positions to allow convergence. Thus, it is very important that an evenly distributed set of target locations is chosen so that the participant is required to look up and down, left and right at various points during the calibration sequence. It should be noted that although a wide range of spatial positions were sampled during the calibration phase of the present study, the particular spatial distribution may not have been optimal, and limitations on the required up and down movement of the head may have been an additional contributor to the magnitude of the elevation errors reported. This possibility should be carefully considered when determining the

spatial distribution of targets used in the eye–head calibration phase.

Because of its affordability and transparency, the eye–head calibration procedure reported here holds great promise for use in a wide variety of experimental contexts other than our own. For example, we would expect this procedure to work very well in more standard setups, in which participants view stimuli presented on vertically oriented displays. Additionally, having the details of the calibration procedure freely available can be particularly helpful in adapting this approach to other nonstandard experimental setups, such as our own. It is our hope that by making these procedures available, we will encourage the development of an open-access community of researchers interested in eyetracking, who can build on and refine the present approach.

AUTHOR NOTE

REFERENCES

BALLARD, D. H., HAYHOE, M. M., & PELZ, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, **7**, 66-80.

CORNSWEET, T. N., & CRANE, H. D. (1973). Accurate two-dimensional eye tracker using first and fourth Purkinje images. *Journal of the Optical Society of America*, **63**, 921-928.

FINDLAY, J. M., & GILCHRIST, I. D. (1998). Eye guidance and visual search. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 295-312). Oxford: Elsevier.

HAPGOOD, M. A. (1992). Space physics coordinate transformations: A user guide. *Planetary Space Science*, **40**, 711-717.

HENDERSON, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences*, **7**, 498-504.

HOFFMAN, J. E., & SUBRAMANIAM, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, **57**, 787-795.

HORNOF, A. J., & HALVERSON, T. (2002). Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers*, **34**, 592-604.

JUST, M. A., & CARPENTER, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, **87**, 329-354.

RAYNER, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**, 372-422.

RICHARDSON, D. C., & SPIVEY, M. J. (2004). Eye tracking: Characteristics and methods. In G. E. Wnek & G. L. Bowlin (Eds.), *Encyclopedia of biomaterials and biomedical engineering* (Vol. 1, pp. 568-572). New York: Dekker.

SPIVEY, M. J., & GENG, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, **65**, 235-241.

TREUE, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*, **24**, 295-300.