

Human eye-head co-ordination in natural exploration

Wolfgang Einhäuser, Frank Schumann, Stanislavs Bardins, Klaus Bartl, Guido Böning, Erich Schneider & Peter König

To cite this article: Wolfgang Einhäuser, Frank Schumann, Stanislavs Bardins, Klaus Bartl, Guido Böning, Erich Schneider & Peter König (2007) Human eye-head co-ordination in natural exploration, *Network: Computation in Neural Systems*, 18:3, 267-297, DOI: [10.1080/09548980701671094](https://doi.org/10.1080/09548980701671094)

To link to this article: <https://doi.org/10.1080/09548980701671094>



Published online: 09 Jul 2009.



Submit your article to this journal [↗](#)



Article views: 395



View related articles [↗](#)



Citing articles: 15 View citing articles [↗](#)

Human eye-head co-ordination in natural exploration

WOLFGANG EINHÄUSER^{1,2}, FRANK SCHUMANN³,
STANISLAVS BARDINS⁴, KLAUS BARTL⁴, GUIDO BÖNING⁴,
ERICH SCHNEIDER⁴, & PETER KÖNIG³

¹*Institute of Computational Science, ETH Zürich, Switzerland,* ²*Division of Biology, California Institute of Technology, Pasadena, CA, USA,* ³*Institute of Cognitive Science, University of Osnabrück, Germany, and* ⁴*Neurology, Hospital of the Ludwig-Maximilians University Munich, Germany*

(Received 27 February 2007; accepted 9 September 2007)

Abstract

During natural behavior humans continuously adjust their gaze by moving head and eyes, yielding rich dynamics of the retinal input. Sensory coding models, however, typically assume visual input as smooth or a sequence of static images interleaved by volitional gaze shifts. Are these assumptions valid during free exploration behavior in natural environments? We used an innovative technique to simultaneously record gaze and head movements in humans, who freely explored various environments (forest, train station, apartment). Most movements occur along the cardinal axes, and the predominance of vertical or horizontal movements depends on the environment. Eye and head movements co-occur more frequently than their individual statistics predicts under an independence assumption. The majority of co-occurring movements point in opposite directions, consistent with a gaze-stabilizing role of eye movements. Nevertheless, a substantial fraction of eye movements point in the same direction as co-occurring head movements. Even under the very most conservative assumptions, saccadic eye movements alone cannot account for these synergistic movements. Hence nonsaccadic eye movements that interact synergistically with head movements to adjust gaze cannot be neglected in natural visual input. Natural retinal input is continuously dynamic, and cannot be faithfully modeled as a mere sequence of static frames with interleaved large saccades.

Keywords: *Natural scenes, visual system, attention, psychophysics*

Introduction

Understanding sensory coding under natural conditions requires knowledge of natural stimulus statistics. While the linkage of properties of the visual world to physiological and psychophysical phenomena of natural vision dates back at least to the early 20th century Gestalt psychologists (Wertheimer 1938), only recent advances in image processing technology and computational power made it feasible to investigate the statistics of large ensembles of natural scenes (Field 1987; Hancock et al. 1992; Ruderman and Bialek 1994; van der Schaaf and van Hateren 1996). Relating scene statistics to coding in the visual cortex has led to the fruitful concept of sparse coding (Olshausen and Field 1996) or to the characterization of V1 simple cells as independent components in natural scenes (Bell and Sejnowski 1997). In addition, various psychological and physiological effects have been explained by adaptation to scene statistics, such as the oblique effect (Boltz et al. 1979; Coppola et al. 1998), its inversion for broadband stimuli (Essock et al. 2003; Hansen et al. 2003), distance-size illusions (Howe and Purves 2002), and some aspects of color constancy (Golz and MacLeod 2002; Long and Purves 2003).

Despite their success, the aforementioned studies do not address an important aspect of natural scene statistics: Temporal dynamics. As a first approach, natural temporal dynamics is modeled by using videos of natural scenes, approximating passive viewing without eye- or head movements. The spatiotemporal power spectrum of such passively viewed natural scenes predicts human visual sensitivity (Dong and Atick 1995a; Dong 1997) and response properties of neurons in the lateral geniculate nucleus of the cat (Dong and Atick 1995b). Extending independent component analysis (ICA) from the purely spatial to the spatio-temporal domain on such data predicts the spatio-temporal properties of simple cells (van Hateren and Ruderman 1998). In analogy to sparse coding, different variants of the coding principle “temporal coherence”, all rooted in Földiák’s (1991) trace rule, have been successfully applied to natural stimuli. In such models, temporal coherence in natural scenes yields V1 complex cells (Einhäuser et al. 2002; Hashimoto 2003; Körding et al. 2004; Berkes and Wiskott 2005), intermediate-level cortical representations (Einhäuser et al. 2003), invariant object representations akin to those found in the inferotemporal cortex (Stringer and Rolls 2002; Einhäuser et al. 2005), as well as nonvisual sensory representations, such as somatosensory coding (Hipp et al. 2005) and place fields (Wyss et al. 2006).

These recent developments make high-quality dynamic stimuli from a truly natural setting necessary. Commercial movies are subject to filmmaker bias, and recordings from vehicles (van Hateren and Ruderman 1998) are restricted to a specific type of environment. To overcome these restrictions, Betsch et al. (2004) recorded movies using a camera mounted on the head of a freely behaving cat. Such head-centered recordings, however, neglect an important aspect of vision in many mammalian species: Eye movements. By simultaneously recording eye movements and movies from a head-centered camera, Moeller et al. (2003) showed that – even in the cat – eye movements have a profound impact on the dynamics of an animal’s retinal input. This surprising finding raises the question as to what extent eye movements influence the dynamics of visual input during free exploration in humans.

Some physiological studies have superimposed measured or simulated eye movements on static scenes to characterize sensory coding of natural stimuli (Vinje and Gallant 2000). However, only a few modeling approaches have explicitly integrated eye movements. Li and Clark (2004) demonstrated that explicit knowledge of saccadic eye movements improves the performance of a temporal coherence coding scheme by means of a nonuniform temporal sampling of the input. Rutishauser et al. (2004) showed that a model of eye movements facilitates subsequent object recognition. These studies restricted themselves to large volitional (saccadic) eye movements. During natural exploration, however, reflexive eye movements, such as optokinetic nystagmus (OKN) and vestibulo-ocular reflex (VOR), also have a nonnegligible impact on the direction of gaze relative to head and body (Carpenter, 1988). Reflexive movements are difficult to investigate under laboratory settings, since conventional eye-tracking techniques preferably use head-fixed setups or suffer from restricted mobility. Using sensor coils for both eye and head, Collewin et al. (1983) obtained extremely precise simultaneous measurements of eye and head movements. By a series of carefully designed experiments, these authors demonstrated that compensatory eye-movements are more accurate during active than during passive head movements and that the eye movement system can rapidly adjust its gain to optimally adapt for manipulations of the relation between visual input and head movements. A further development of this recording technique, the so-called Maryland revolving field monitor (MRFM, Collewin et al. 1992), allowed routine measurements of eye movements with unrestrained head in naturalistic tasks. Using this technique, Kowler et al. (1992) found an increased saccade frequency as compared to head-restricted settings, differences between natural and unnatural tasks, as well as evidence for “a single high-level motor program” controlling eye- and head rather than separate systems for either. These results stress the importance of naturalistic, minimally restrained, and simultaneous recordings of eye and head movements. In a pioneering series of studies, Land and colleagues went beyond the laboratory setting and characterized human eye movements during a variety of everyday activities, such as car driving (Land and Lee, 1994) making tea or preparing food (Land et al. 1999; Land and Hayhoe 2001). Similarly, by using a custom-developed portable eye-tracker, Pelz et al. (2000) characterized eye movements for specific tasks, such as map reading and hand washing. Nevertheless, it is difficult to generalize from these specific tasks to a general description of eye movements during natural behavior.

Despite the enduring interest in eye movements during natural behavior (Hayhoe and Ballard 2005 for review), to the best of our knowledge, human gaze during natural exploration within an unrestricted environment has not yet been continuously recorded. With the aforementioned exceptions, the majority of eye movements studies typically use head-restrained setups. Models of eye movements, such as saliency maps, are then validated – if at all – against such data. Here we used a novel, entirely mobile and portable setup, to simultaneously record gaze- and head-centered visual input for a prolonged period of time. We applied this unique setup to study the co-ordination of human eye and head movements during natural exploration. Irrespective of the classification of eye movements as volitional or reflexive, our aim was to determine to what extent co-occurring eye and head movements act synergistically to adjust gaze or

compensatorily to stabilize gaze. This knowledge is essential to estimate the effect of eye movements on natural stimulus statistics and consequently on sensory coding.

Methods

Setup

We developed a video camera setup (“EyeSeeCam”) for simultaneous recording of gaze- and head-centered videos, which is described in detail elsewhere (Schneider et al. 2005, 2006; Brandt et al. 2006). In brief, one head-mounted camera is continuously aligned with the observer’s orientation of gaze (“gaze camera”), while a second camera is fixed relative to the observer’s head (“head camera”, Figure 1). Swimming goggles together with a flexible, curved flat spring that fit the head in the mid-sagittal plane serve as a lightweight head mount. High-speed (100 Hz) video cameras that are laterally attached to the goggles track the eye movements (Figure 1b). These cameras digitally transfer video images to a wearable computer that detects pupil position in real-time. The eye tracker results drive two servo motors that rotate the pivotable gaze camera above the forehead. The camera is attached to the flat spring by means of a cardan joint and its top two corners are connected in a parallel kinematics setup by two lever arms to the servo motors. The distance of the camera to each eye was about 7 cm. For calibration purposes, a small laser diode is mounted on the gaze camera module. When switched on and projected onto an object like a wall, the laser dot is clearly visible for the observer and moves aligned with the gaze camera. During the calibration procedure, we ask the observer to fixate and follow the laser dot. The pupil position values, which are measured in image coordinates during these fixations, are linearly fitted by two third order 3D functions to the corresponding predefined motor command values. During normal operation these functions are used to map eye positions to motor commands. The head camera, which is of the same type as the pivotable one, is rigidly attached to the flat spring. The standard analog video signals of both cameras are first converted to digital video (DV) signals with DV-compression and then multiplexed in a single DV file. To avoid interlacing artifacts only odd lines are used for analysis and – to preserve the aspect ratio – only odd columns. This yields data at 288×360 pixel spatial and 25 Hz temporal resolution for each camera. For further analysis data are converted to grayscale using the default mapping of MatLab’s (Mathworks, Natick, MA) `rgb2gray` function. The raw video (DV) data are available from the authors on request.

Subjects/Environment

Four volunteers (three males, one female; ages 25–39) participated in the experiment. Two of the subjects (K, S) were among the authors. All participants were accustomed to wearing the recording equipment. Experiments were conducted in three different environments (Figure 2): A local forest, Munich’s central train station, and the apartment of one of the authors. Participants were instructed to behave naturally and freely explore their environment without paying specific

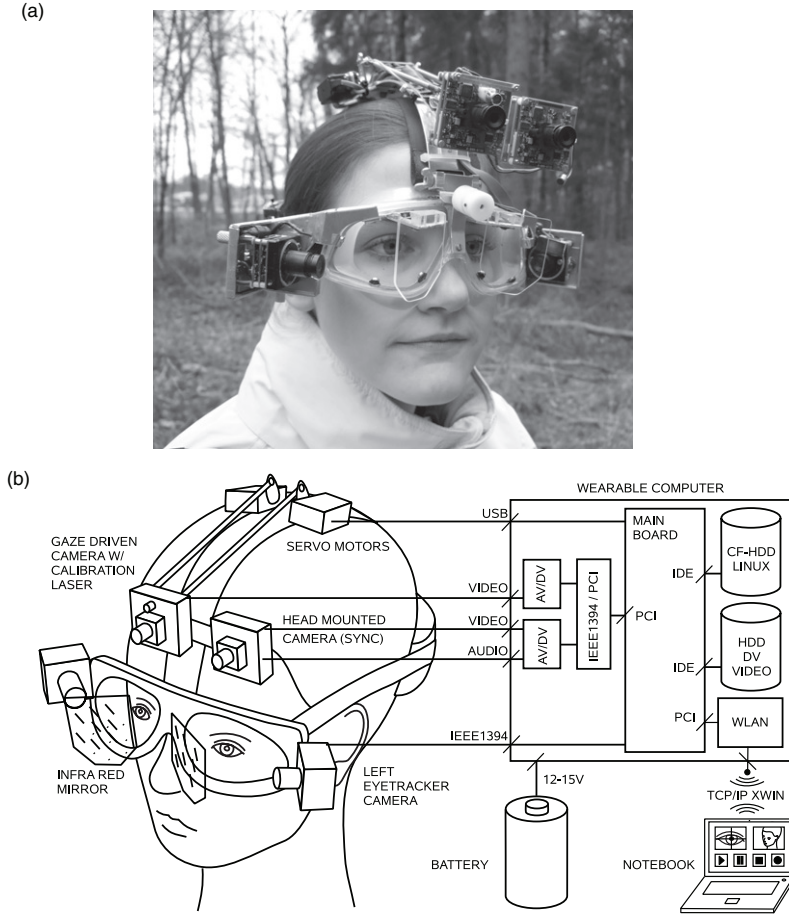


Figure 1. Video setup with gaze-driven and head-fixed scene cameras. (a) Setup during operation in natural environment. (b) Schematic view. Left: Swimming goggles and a lightweight mid-sagittal flat spring were used as a head mount for laterally attached eye tracker cameras and also for the gaze-fixed and head-fixed scene cameras that were attached above the volunteer's forehead. The gaze camera was rotated by servo motors. Right: A battery-driven wearable computer performed all required tasks: Eye tracking, servo motor control, and synchronized scene capturing of both gaze-fixed and head-fixed digitized video streams. The experimenter could follow the volunteer by carrying an additional notebook computer which enabled him to both operate the wearable computer and monitor the image quality of the eye tracking and scene cameras over a wireless remote control connection.

attention to the recording equipment. All procedures conformed with National and Institutional Guidelines for Experiments with human subjects and with the Declaration of Helsinki.

Velocity estimation

To estimate gaze (eye-in-world) velocity (\mathbf{v}_{gaze}) and head velocity (\mathbf{v}_{head}), the shift between subsequent frames of each camera was computed. The central

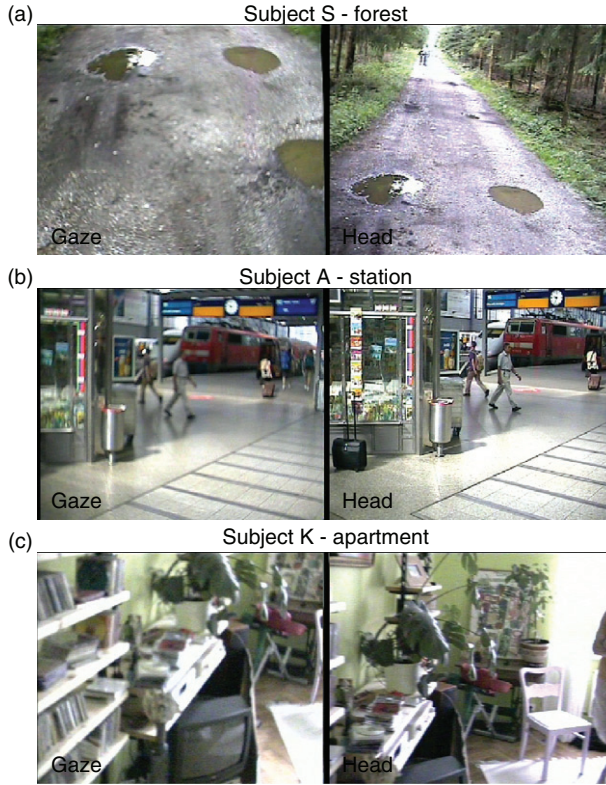


Figure 2. Still frames of different environments (a) forest (b) train station (c) apartment. Left: Head centered view, right: Gaze centered view.

128 × 128 pixels of each frame, which were down-sampled to 64 × 64 pixels for computational efficiency, were used. The two-dimensional cross-correlation between the current and the subsequent frame of the same camera was computed. The offset of the peak in the cross correlation from the center determines the shift between the two frames. The negative of this shift divided by the interval between the frames (40 ms) estimates the velocity of head ($\mathbf{v}_{\text{headcamera}}$) and gaze ($\mathbf{v}_{\text{gazecamera}}$), respectively.

By measuring the camera's aperture we computed the values in absolute coordinates of visual angle: 1 pixel/frame corresponds to 2.16°/s. As we compute flow only between subsequent frames this is the lower resolution limit on measured velocities. To obtain finer resolution, flow between images more than one sample apart need to be computed, which is not of interest to the current study. The upper limit on velocity is given by the size of the central patch. The maximum measurable velocity is thus given by 128 pixels/frame or 276°/s.

The true head velocity $\mathbf{v}_{\text{head}}(t)$ is well approximated by the measured $\mathbf{v}_{\text{headcamera}}(t)$. This velocity $\mathbf{v}_{\text{head}}(t)$ defines the complete movement of head in the world, irrespective of whether it results from a head-on-body or body-to-world movement. As the present study deals with the relative importance of eye movements to gaze direction, such a further distinction is not necessary.

In case of the gaze velocity the latency of the mechanical camera control has to be taken into account. In Appendix A we demonstrate using an artificial eye and two gyroscopes that this latency is 36 ms, independent of velocity. Hence $\mathbf{v}_{\text{gaze}}(t) = \mathbf{v}_{\text{gazecamera}}(t + 36 \text{ ms})$. Since 36 ms corresponds roughly to 1 frame (40 ms) and relevant accelerations on that time scale are small, we define the eye-velocity (eye-in-head) $\mathbf{v}_{\text{eye}}(t)$ as follows:

$$\begin{aligned}\mathbf{v}_{\text{eye}}(t) &= \mathbf{v}_{\text{gazecamera}}(t + 40 \text{ ms}) - \mathbf{v}_{\text{headcamera}}(t) \\ &= \mathbf{v}_{\text{gaze}}(t + 4 \text{ ms}) - \mathbf{v}_{\text{head}}(t) \approx \mathbf{v}_{\text{gaze}}(t) - \mathbf{v}_{\text{head}}(t)\end{aligned}$$

where the latter equality assumes that the gaze velocity changes little over the course of 4 ms, which we confirm by computing the time-constant of the velocities autocorrelation (see Results). In principle, the eye-velocity could also be estimated from the control signals for the motors steering the gaze camera. During calibration the (nonlinear) mapping between measured eye-position and servo-controls is estimated directly, without explicit representation of the eye-in-head in absolute coordinates. In Appendix B we demonstrate that reconstructing eye-in-head from the servo-controls yields results consistent with the flow-field estimates.

To further verify the method of eye, head and gaze velocity estimation, we instructed subjects to perform defined compensating movements under recording conditions identical to the main experiment. Appendix C demonstrates that the system reliably picks up compensatory movements, even in the presence of large head motion.

Results

Simultaneous head and gaze movements were recorded in four human observers during natural exploration of various environments. After the exclusion of periods of calibration, a total of 67 min of valid data were obtained. Forty-five minutes were recorded in a forest environment (Subject B: 9.3 min; K: 17.6 min; S: 18.5 min, Figure 2a), 13.3 min in a train station (Subject A, Figure 2b), and 8.6 min inside an apartment (Subject K, Figure 2c).

2D-Velocity distributions of eye, head and gaze

The velocities of gaze (i.e., eye-in-world) were measured by computing optic flow for the gaze centered camera. The technical analysis with an artificial eye (Appendix A) ensured that the true gaze direction was reflected faithfully in these camera movements up to eye-velocities of 200°/s, while maintaining a constant lag of 36 ms. It is important to note that the absolute pupil position was not represented explicitly in the nonlinear mapping from eye measurement to camera control (cf. Appendix B). Instead, this nonlinear mapping was adjusted for each individual at the start of each session. Hence the gaze camera provides the best estimate of eye-in-world velocities.

For all individuals and environments, most gaze shifts were oriented along one of the cardinal axes (Figure 3a). There was a pronounced and remarkable anisotropy, which depended on the type of environment. In the indoor environment

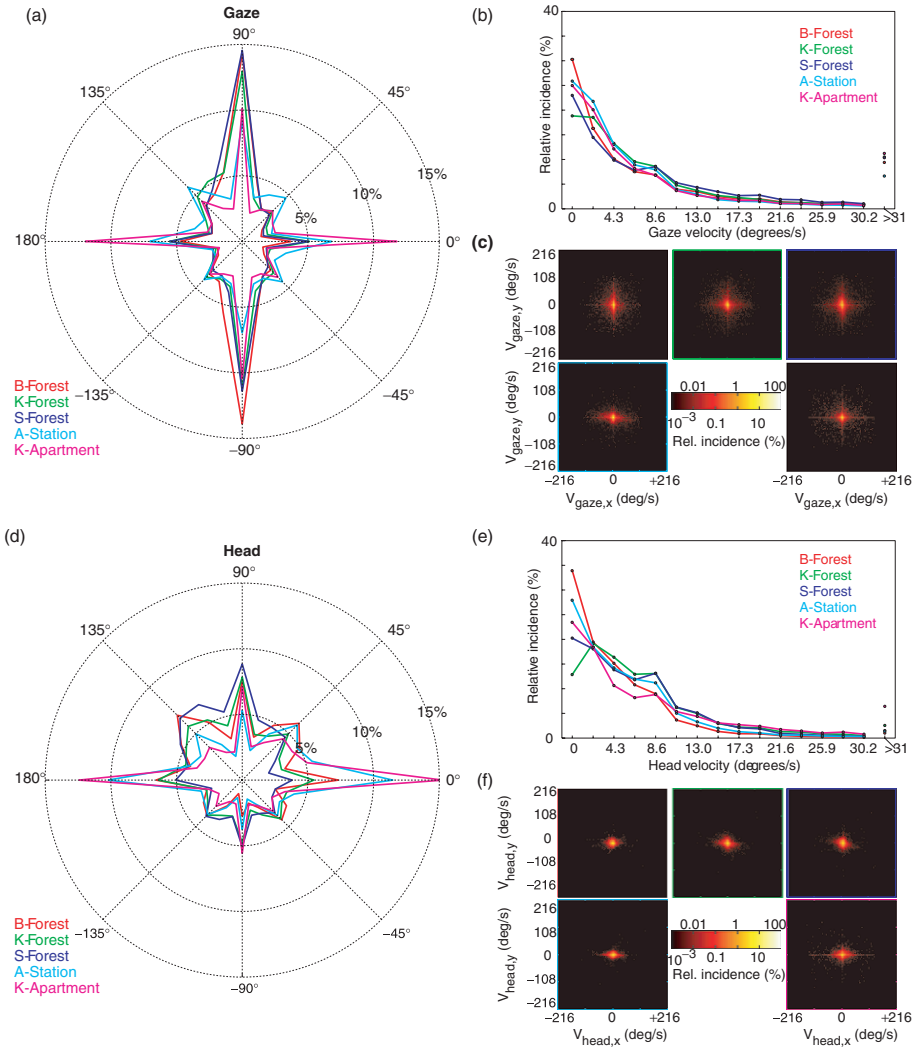


Figure 3. Gaze and Head velocity distributions: (a) Histograms of gaze velocity angles for all nonzero speeds; normalized to unit integral for comparison, bin width 15° ($[-7.5^\circ, +7.5^\circ)$, $[+7.5^\circ, +22.5^\circ)$, ..., $[172.5^\circ, \dots -172.5^\circ)$, ..., $[-22.5^\circ, -7.5^\circ)$). Colors denote subjects and environments. Directions in the panel (left-right, up-down) correspond to direction of gaze movement (b) Histograms of gaze velocity absolute values, bin-width according to pixel resolution and sampling rate 1 pixel/frame = $2.2^\circ/\text{s}$. (c) 2D-histograms of gaze velocity, frame color denotes subject and environment according to colors in panels a and b (top row: Forest B, K, S; bottom row A-station; K apartment). Note the logarithmic scale of the histogram color-code. (d) Histogram of head velocity angles analogous to panel A. (e) Histogram of head speeds, analogous to panel b. (f) 2D-histograms of head velocity, analogous to panel f.

(apartment) horizontal movements were more frequent, while in the forest vertical movements dominated. The station environment fell in between those two extremes and of all environments had the largest contribution from oblique orientations. It is likely that the vertical component is attributable to movement planning and gait

control during walking in the forest (cf. Figure 2b), while in the apartment exploration has to take place in a vertically restricted field of view (ceiling, floor).

Absolute gaze velocities were similar in all environments tested (Figure 3b) and there was a substantial fraction ($9.6\% \pm 1.8\%$) of very rapid movements beyond $31^\circ/\text{s}$. In turn, times of nearly no measurable movement (smaller than $2.2^\circ/\text{s}$) accounted for $24.6\% \pm 4.2\%$ of the time. The two-dimensional analysis of gaze shifts (Figure 3c) showed an angular distribution similar to the overall angular distribution of each individual, i.e., there was no noticeable preference for faster or slower movements along one of the axes.

By analyzing head movements, which integrate movements of the body in the world and of the head relative to the body, we judged the role of eye movements in the anisotropic distribution of gaze shifts. Although the pronounced dominance for horizontal movements in indoor environments was still present for the head alone, the vertical dominance was substantially reduced (Figure 3d). Furthermore, upward head movements were more frequent than downward movements. This implies that the vertical components of gaze shifts during walking in outdoor environments were predominantly carried by eye movements, while head movements contributed substantially to horizontal shifts. There was no systematic difference between different environments with respect to head-movement speeds (Figure 3e). In all cases, head movements showed less high velocities ($2.6\% \pm 2.2\%$ beyond $31^\circ/\text{s}$) than gaze movements, which implies that fast eye movements are the main contributor to rapid gaze-shifts. The two-dimensional analysis of head movements (Figure 3f) again was dominated by the marginal distributions in angle and value, such that there was no noticeable dependence of the two. In summary, there was a difference between vertical and horizontal gaze-shifts in relative contributions of head and eye to gaze allocation. This is the rationale to perform separate analyses of these components below.

Difference of head- and gaze shifts: Eye movements

To assess how eye- and head movements interact for gaze adjustment, first the relative angle between simultaneous gaze and head movements was measured. The enclosed angle peaked at multiples of 45° (Figure 4a). This relation was expected in the light of dominant cardinal orientations for the individual angles. While head and gaze movements most often pointed in the same direction ($14.7\% \pm 2.7\%$ have angles smaller than 15° , Figure 4a), there was a substantial fraction of time, when gaze and head moved in opposite directions, i.e. eye movements overcompensate for head movements ($9.8\% \pm 1.8\%$ for angles larger than 165°). Hence an analysis of relative movements of eye and head was necessary to understand their relative contributions to gaze.

We defined the contribution of eye movements to gaze shifts as the difference between gaze and head velocity (see Appendix B for validation). While the bulk of time there were no or comparably slow eye movements ($52.0\% \pm 5.6\%$ below $10^\circ/\text{s}$), $14.2\% \pm 3.8\%$ were faster than $30^\circ/\text{s}$ and $4.4\% \pm 1.8\%$ were even faster than $60^\circ/\text{s}$ (Figure 4b). Hence eye movements were the main contributor to rapid gaze shifts (Figure 3b).

The angular distribution of eye movements (Figure 4c) exhibited a strong bias towards the cardinal axes, also a pronounced difference between indoor and

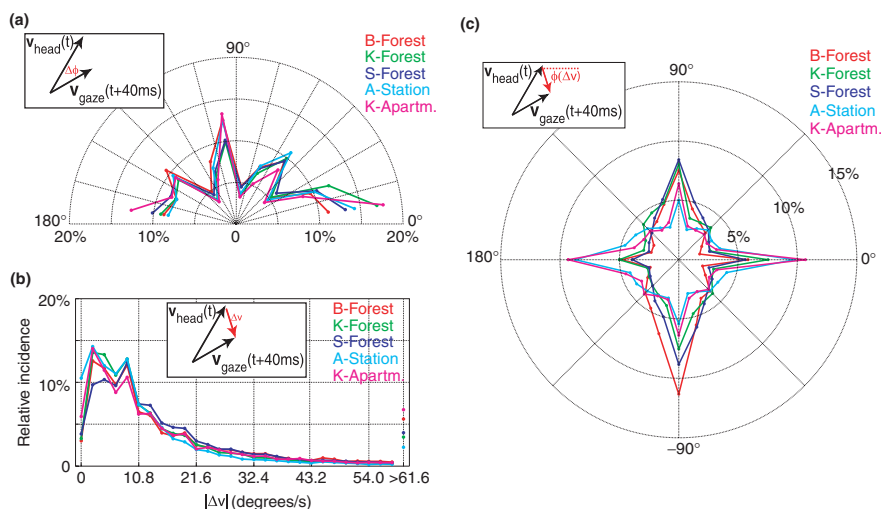


Figure 4. Relative movement of head and gaze; eye movement (a) Histogram of angles between gaze velocity and head velocity. Note that gaze velocity is taken a frame later to account for the system latency of about 40 ms (Appendix A). Color codes as in Figure 3. Binning in 15° steps, [0°, 15°), ..., [90°, 105°), ..., [150°, 165°), [165°, 180°], i.e., last bin includes 180°. (b) Eye velocity as difference between gaze and head velocity; histogram over speeds. (c) Histogram of eye-movement directions. Binning and colors as in Figures 3a and d.

outdoor environment. During walking in the forest vertical movements were dominant, whereas horizontal movements dominated for apartment and station. Two hypothesis on the origin of vertical eye movements during walking are possible, and not mutually exclusive: On the one hand, walking in the open induces a strong vertical optic flow that eye movements may partly compensate for to stabilize gaze; on the other hand, movement planning and gait control on uneven terrain is likely to require alternate looking on the ground and ahead, i.e. eye movements support head and body movements for gaze adjustment. Below we thus estimate the relative contributions of the former, *compensating* movements, to the latter *synergistic* movements for all environments. The observed difference between the horizontal and the vertical requires separate analysis along the cardinal axes in addition to analysis along the full angular range.

Cardinal components of eye-, head, and gaze velocities

Since the cardinal movement directions were dominant for gaze, eye and head shifts, and the first analysis implied difference between horizontal and vertical movements, their respective distributions were analyzed separately. Stable gaze does not necessarily imply the absence of head and eye movements; instead, head movements and eye movements may compensate each other to stabilize gaze. To quantify this relation, the fraction of time, in which no measurable movement occurred in a given cardinal direction (horizontal or vertical), was assessed (throughout the paper we define “no (measurable) movement” along a cardinal

axis as velocities of 1 pixel/frame, i.e. $2.2^\circ/\text{s}$, or less for the respective axis). In all subjects and environments, the horizontal component of gaze was stable more often than eye and head alone (Figure 5a). This indicates that co-ordination between eye and head movements serves in part to stabilize gaze beyond the stability of either movement alone. Vertical eye movements were less frequent than vertical head or gaze movements, as was the case for horizontal movements. In contrast to the horizontal case, however, vertical gaze shifts were more frequent than vertical head movements (Figure 5b). This difference between the two cardinal directions might result from vertical eye and head movements being generally more synergistic (less compensatory) than horizontal movements. Alternatively, compensatory mechanisms could persist longer in the horizontal than in the vertical direction, yielding more aggregate horizontal than vertical compensatory movements. Below these two possibilities were dissociated by measuring the time-course of compensatory and synergistic eye movements, respectively. The individual analysis of gaze, head, and eye movements, respectively, showed some evidence for gaze stabilization through co-ordinated eye and head movements. However, periods of stable gaze exceeded periods of stable head by only a small fraction in the horizontal and not at all in the vertical direction. This leaves room for eye movements that are not primarily compensatory, but act synergistically with head movements in adjusting gaze. Consequently, the co-occurrence of eye and head movements was quantified, and their relative direction investigated. As only opposing eye and head movements *can* contribute to gaze stabilization, the fraction of such (potentially) compensatory eye movements provides an upper bound for the time in which the retinal image is stabilized.

Dependence of head and eye movements

In order to quantify the mutual dependence of different eye and head movements, the fraction of isolated and combined eye and head movements was computed and compared to a prediction of those values if independent. As above, an eye or head movement was defined as any period of time, in which the respective velocity exceeded $2.2^\circ/\text{s}$. For example, Subject S in the forest environment, the eyes moved horizontally 61.0% of the time, and the head 49.5% of the time. An eye-movement was termed “isolated”, if there was no simultaneous head movement and vice versa. For horizontal movements in S, isolated eye movements accounted for 19.3% of the measured time, isolated head movements for 7.8% of the measured time. For 31.2% of the time no measurable movement occurred; for 41.7% of the time eye – and head movements co-occurred (Figure 5c). If eye and head movements were independent, such combined movements would happen only 30.2% of the time $[61.0\% \times 49.5\%]$, fixations 19.7% $[(100\% - 61.0\%)(100\% - 49.5\%)]$, isolated eye movements 30.8% $[(100\% - 49.5\%) \times 61.0\%]$, and isolated head movements 19.3% $[(100\% - 61.0\%) \times 49.5\%]$ (Figure 5d). This shows that eye and head movements occur in isolation about half as often (27.1%) than a null hypothesis of independence would predict (50.1%). Despite some variation in the overall fraction of eye and head movements, this dependence of eye and head movements held for all subjects and environments (Figure 5e). The same also held for vertical eye and head movements: They were not independent for any subject or

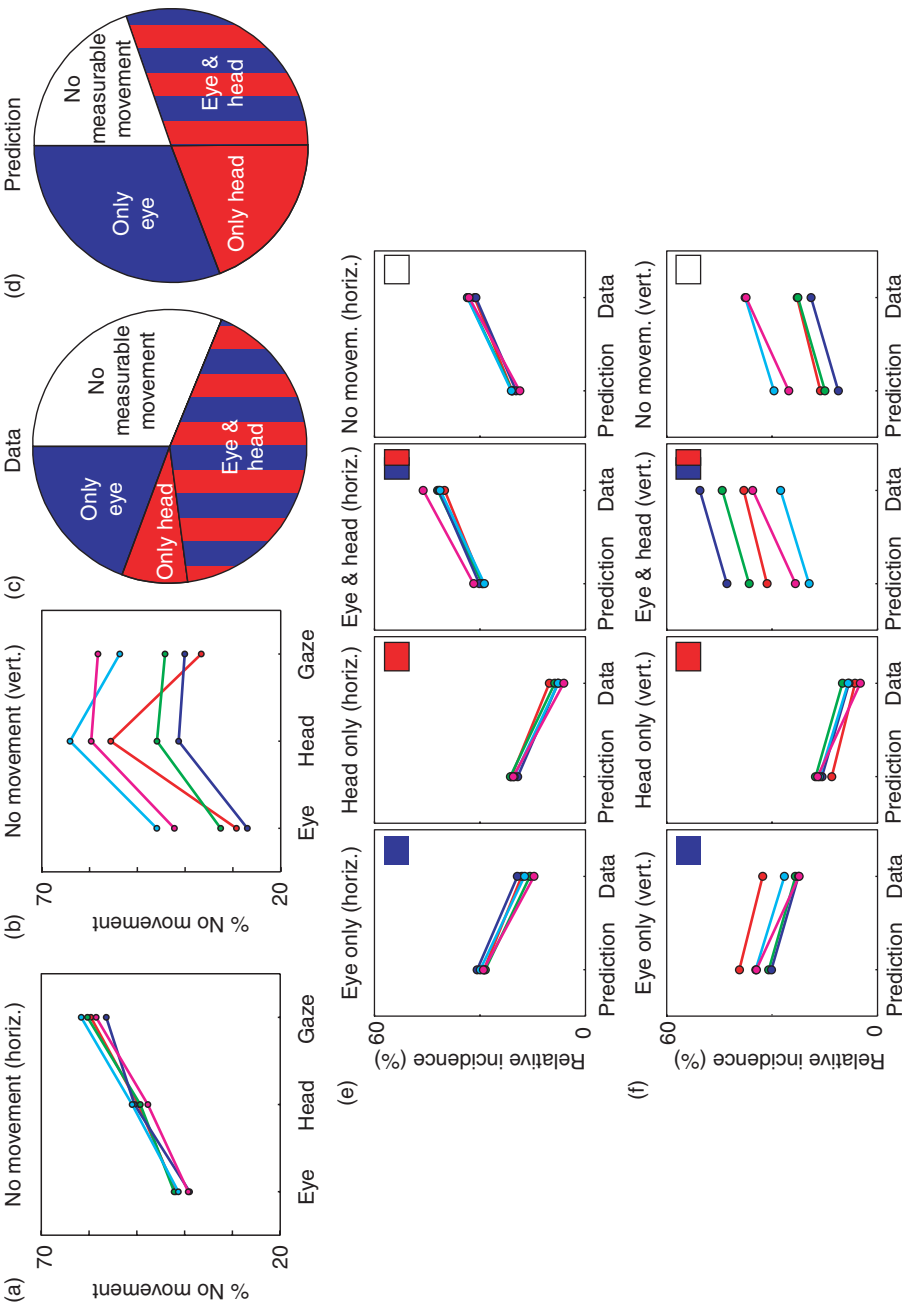


Figure 5. Co-ordination of eye and head movements. (a) Percentage of time, in which no horizontal movement occurred for eye, head, and gaze, respectively. Different colors denote subjects and environments. (b) Percentage of time in which no vertical movement occurred for eye, head, and gaze, respectively. (c) Fraction of isolated horizontal eye movements (blue), isolated horizontal head movements (red), combined horizontal eye-head movements (blue/red) and no horizontal movements (“fixations”, white) for example subject (S) in the forest environment. (d) Prediction if horizontal eye and head movements were independent. Note that isolated movements would be much more frequent. Colors as in panel c. (e) Data of panels c and d for all subjects. Color codes as in panel a. In all subjects there were less isolated movements (left panels) and more combined movements (right panels) than independence would predict. (f) Analogue data to panel e for vertical eye movements.

condition (Figure 5f). These results demonstrate that eye and head movements are highly co-ordinated and provide a first quantification of this dependence.

Relative direction of eye and head movements

As stated above, frequent co-occurrence of eye and head movements can have two reasons, which are not mutually exclusive. First, eyes and head may move synergistically; second, eye movements may counteract head movements to stabilize gaze. If gaze stabilization were the predominant function of nonvolitional eye movements under natural viewing conditions, modeling retinal input as a sequence of static images interleaved by saccades would receive some support. If, however, there were a substantial fraction of eye movements that were not gaze stabilizing, models would have to take gaze dynamics into account. Consequently, we analyzed the relative *direction* of eye and head movements. Counting all co-occurring eye and head movements that oppose each other as compensatory irrespective of speed, we overestimated the fraction of stabilizing movements. If this conservative measure still yielded a substantial fraction of noncompensatory (synergistic) movements, we can thus safely conclude that gaze stabilization is *not* the only dominant function of eye-head co-ordination, and gaze dynamics needs to be explicitly modeled.

The analysis so far treated each measured time-point as an individual measurement. In fact, however, both eye- and head movements do not start or end instantaneously on the sampled timescale, but may have characteristic time-courses over several data points, which are e.g., limited by the inertia of the individual systems. While we did not aim at segmenting the signal in individual motion segments, we nevertheless analyzed the time-course of relative movements. As first quantification of the time-course of compensatory and synergistic eye and head co-ordination, we compared the relative movement direction of eye and head at different temporal lags. First the time-course of the relative direction of head and eye movements was analyzed pooled over all movement directions. For this purpose we defined a movement as *compensatory* if the relative angle between eye and head fell above 165° and as *synergistic* if the angle was below 15° . At 0 lag¹ $35.8\% \pm 6.4\%$ of the eye movements were compensatory, whereas $4.6\% \pm 0.7\%$ synergistic (Figure 6a). At a lag of 1s, these values were $9.1\% \pm 0.6\%$ and $10.7\% \pm 0.5\%$, respectively. As individual gazeshifts may safely be regarded as independent over this 1s period (see also discussion of autocorrelations below), this provided a baseline for the number of compensatory and synergistic movements expected by chance. Significant deviations from this baseline (at $p = 0.05$, t -test) were found between -240 ms and 80 ms lag for compensatory movements, and between -240 ms and 0 ms lag for synergistic movements. The asymmetry suggested preparation of a head movement by eye movements ($t < 0$) to start about half a second before a head movement, but to end less than a quarter second afterwards. One hypothesis for the source of this asymmetry is the difference of head and eye inertia, which forces velocities to be correlated over extended periods of time. A first hint in this direction is provided by separating the analysis for different head speeds (Figure 6b). The faster the head-speed, the stronger and earlier the onset of compensatory eye movements occurred. Below the contributions of head and eye inertia are quantified by the respective

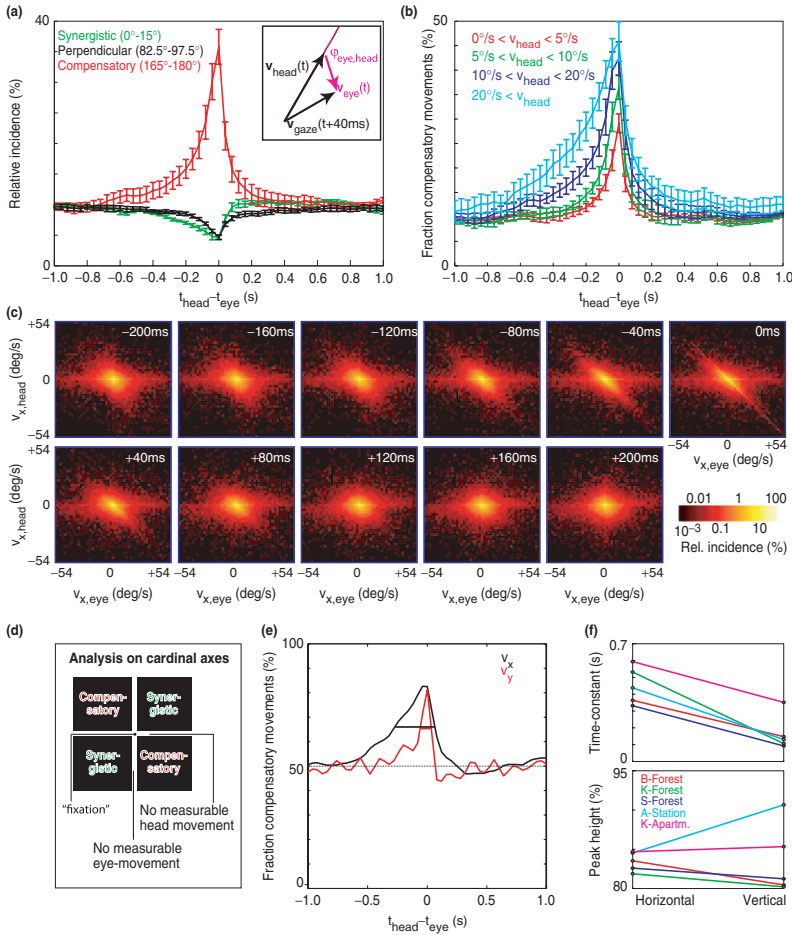


Figure 6. Time-course of eye-head co-ordination. (a, b) Two-dimensional analysis of "synergistic" and "compensatory" movements, irrespective of movement direction. Angle between eye and head movement defines movement type. (a) Relative number of compensatory (red), perpendicular (black) and synergistic (green) movements as fraction of total time with nonzero movement for different lags between head and eye movement. (b) Relative number of compensatory movements (165 – 180°) for different head speeds as fraction of total time in this velocity range. (c–f). Analysis of "synergistic" and "compensatory" along cardinal axis (horizontal or vertical). (c) 2D histograms of horizontal eye and head velocity for different lags between eye and head movement. All axes denote velocity in degrees/second. Figure shows data for Subject S in environment forest. Note the logarithmic scale. (d) Illustration of diagrams of panel c: The inner bins denote velocities below the $2.2^\circ/\text{s}$ threshold, which are counted as no eye or head movement, respectively; their union (neither eye nor head movement) is denoted as "fixation". Eye and head movements in the same direction are defined as "synergistic" in the opposite direction as "compensatory". (e) Fraction of compensatory eye movements in horizontal (black) and vertical (red) directions for different lags between eye and head movements, i.e., the incidences in the lower right and upper left quadrant divided by the incidences in the four quadrants. Dashed lines denote full width at half maximum (FWHM relative to 50% level), which is used for computing the time constants. (f) Time constants (FWHM) for all subjects in vertical and horizontal direction. Colors denote subjects and environments as in Figure 3c. (g) Peak height (maximum fraction of compensatory movements) for horizontal and vertical movements, respectively.

velocity autocorrelations. For the interpretation of the present data at non-zero lags, it is important to note, that a single movement segment contributed to data across several lags, as the autocorrelation of all velocities was larger than the sampling rate.

Next synergistic and compensatory movements along each cardinal axis were separately analyzed. For this 1D analysis, we defined *synergistic* movements along an axis as movements of the opposite direction (sign) and *compensatory* movements as movements of the same sign, again irrespective of absolute speed. We computed 2D histograms for horizontal eye and head velocities. At 0 lag there was not only a clear peak at the central bins of no measurable movement, but also a preference for the secondary diagonal, i.e., relative eye and head movements that were compensatory in direction and size. This dominance also prevailed for nonzero lags about 200 ms around the 0 lag (Figure 6c). When counting the fraction of all fully or partly compensatory movements, i.e. velocities falling in the upper left or lower right quadrant of the 2D-histogram (Figure 6d), relative to all movements, the horizontal movements of Subject S exhibited a peak of compensatory relative direction at 0 lag (Figure 6d). At this time point, 82.6% of the co-occurring eye and head movements pointed in opposing directions. This is, maximally 82.6% of eye movements can act to horizontally stabilize gaze. Importantly, the true number of fully stabilizing movements, which also require the same absolute velocity in addition to opposing direction, is considerably lower, on the order of maximally 55% (see analysis below). Even with the most conservative estimate, however, at least about 20% of eye movements cannot be explained by gaze stabilization alone.

Whereas eye movements that oppose head movements might contribute to gaze stabilization, it was unlikely that all of them perfectly stabilize gaze. In particular, in the histograms of Figure 6c many entries were counted as “compensatory”, although they were far off the second diagonal, where – under ideal conditions – all stabilizing movements would fall. When restricting the definition of “compensatory” to co-occurring movements that were within 1 pixel/frame ($2.2^\circ/\text{s}$) of this diagonal, i.e., opposed each other and had about the same absolute value, $54.8\% \pm 3.6\%$ for horizontal and $41.3\% \pm 8.4\%$ for vertical compensatory movements (mean \pm SD across subjects for peak values, which all occurred within one frame of 0 lag) was obtained. If exactly the same absolute value was required, i.e., only occurrences on the second diagonal were counted, the figures dropped to $33.8\% \pm 3.1\%$ and $25.3\% \pm 6.2\%$, respectively. Hence the figure of about 80% was only an upper limit for the fraction of stabilizing movements, with the true value of perfectly stabilizing movements being considerably smaller. However, even if all compensatory eye movements were perfectly stabilizing, still on the order of 20% of eye movements would not even be potentially compensating, but would act in a synergistic fashion with head movements to direct gaze. In Appendix C instructed compensatory behavior is used to show that a worst-case estimate of maximally 6% of compensating movements could remain undetected or misclassified, with an average estimate of less than 4%. This number gets even smaller, if it is assumed that even under instruction compensation is not perfect. Consequently, at least 20% of eye movements that co-occur with head movements can safely be assumed to not compensate but rather synergistically support the head movement in directing gaze.

Time constants of relative eye head movements

The fraction of compensatory movements peaked at lag 0 but extended for a finite amount of time. We quantified this temporal extent by the full-width at half maximum (FWHM) relative to the chance level of 50% compensating movements. In the example subject, half maximum for horizontal movements was thus 66.3% (horizontal line in Figure 6e), and the FWHM time constant was 331 ms (linear interpolation between adjacent data points achieves sub-frame resolution). The time constant for vertical eye movements was substantially lower (92.4 ms, Figure 6e red). Note that both time constants were substantially longer than the sampling rate, which ensured that compensatory eye movements were not underestimated due to technical limitations.

A comparison of the time constants of horizontal and vertical movements showed that compensatory vertical movements peaked slightly lower than horizontal ones: At 81.2%. The pattern that compensating for horizontal eye movements was lasting longer than for vertical movements, was consistent across all subjects (Figure 6f). Despite substantial inter-observer variability, the ($452 \text{ ms} \pm 111 \text{ ms}$) and the mean vertical ($166 \text{ ms} \pm 106 \text{ ms}$) time constant was significant ($p = 0.003$, t -test). Consequently, horizontal compensatory movements tended to persist longer than vertical ones.

This difference between vertical and horizontal time constants raises two questions: First, could the difference be a consequence of different peak fractions of compensatory movements for the two directions. Such a difference was, however, not observed: neither was the direction (vertical larger than horizontal or vice versa) consistent across observers (Figure 6f), nor did the means for the two directions differ significantly ($p = 0.95$, t -test). For all subjects and both directions the maximum fell between 80.2% and 90.7%, with an average of $83.4\% \pm 1.2\%$ (horizontal) and $83.6\% \pm 4.5\%$ (vertical). Hence – as one would expect – potentially compensatory movements accounted for the majority of eye movements in natural exploration. Second, can the fraction of synergistic eye movements be accounted for solely by saccadic eye movements? This seems unlikely for several reasons. First, the synergistic eye movements did not exhibit particularly high velocities, especially compared to eye movements occurring in absence of head movements (Figure 6c). Second, even if the system were assumed to fail to detect too fast saccades and the subject made five saccadic eye movements per second, 20% ($5^\circ/\text{s}$ by 25 Hz sampling rate) of all recorded eye movements (including those in absence of head movement) would be saccades. Under the assumption that all synergistic relative eye and head movements are explained by saccades, this would leave no saccades possible in the absence of head movements. Not only is this unlikely, but the velocity distribution in the absence of head movements does not support this conclusion. Thus it can safely be concluded that a substantial fraction of the synergistic eye relative to head movements, which account for at least 20% of the total data according to a most conservative estimate, is not explained by saccadic eye movements. Instead, these movements are evidence of a synergistic interaction between head movements and nonsaccadic eye movements for gaze allocation.

Individual time constants of gaze-, eye and head movements

The observed time constants for compensatory relative movements could in principle be the result of the compensatory mechanism itself or of a slow individual mechanism. Hence we quantified the inertia of all eye- head and gaze movements by measuring their autocorrelation. When the time constants of eye, head, and gaze were analyzed individually, comparably small time constants were found for gaze (vertical: $70 \text{ ms} \pm 11 \text{ ms}$; horizontal: $110 \text{ ms} \pm 29 \text{ ms}$ Figures 7a and b) and eye (vertical: $72 \text{ ms} \pm 15 \text{ ms}$; horizontal: $115 \text{ ms} \pm 19 \text{ ms}$; Figures 7e and f) movements. In contrast, head movements exhibited time constants on the order of the compensatory movements (vertical: $317 \text{ ms} \pm 185 \text{ ms}$; horizontal: $611 \text{ ms} \pm 130 \text{ ms}$; Figures 7c and d). For all movements, however, mean time constants were significantly smaller for vertical than for horizontal movements (gaze: $p=0.02$; head: $p=0.02$; eye: $p=0.004$). This suggests that the difference between vertical and horizontal time constants in compensation may be caused by different time constants of the individual movements. It seems likely that the comparably smaller vertical oculomotor range does not allow (or require) the movement to begin earlier and stop later to cause a prolonged shift of the eye, the head, or the gaze.

Time constants for compensatory and synergistic movements

After having analyzed the time course of individual movements and the time course of the relative direction of movements, next the question as to whether compensatory and synergistic eye movements themselves have different time constants was addressed. The conditional probability was measured to find exactly k synergistic movements in the n frames following a synergistic movement as well as the conditional probability to find k compensatory movements following a compensatory movement. Figure 8a depicts this analysis for horizontal movements of the example subject S in the forest. The probability of finding a second compensatory movement after a first compensatory movement is 62.1% (Figure 8a, left), while the same probability is less than 50% for synergistic movements (29.6%; Figure 8a, right). In the 14 frames (600 ms) following a synergistic movement it is the most likely event to find no additional synergistic movement, while for compensatory movements finding about $n/2$ further compensating events is most likely. To quantify this difference further, the special case $n=k$ was analyzed, i.e., the probability that all n frames following a compensatory frame were again compensatory, and analogously for synergistic movements. This conditional probability decays much slower with increasing n for compensatory than for synergistic movements (Figure 8b). This difference was more pronounced for horizontal than for vertical movements, which decay faster. If the conditional probability of an $(n+1)$ th compensating frame to follow an n th compensating frame is assumed to be independent of n , an exponential can be fitted to the decay over n . This fit and the fact that one frame corresponds to 40 ms, allowed us to define the time constant of synergistic and compensatory movements as the life-time of these exponential functions (i.e., the time they take to decay by a factor of e). As we fitted to 10 datapoints, the estimates are robust even for life-times on the order of the temporal resolution, as long as the model assumption of a singly exponential

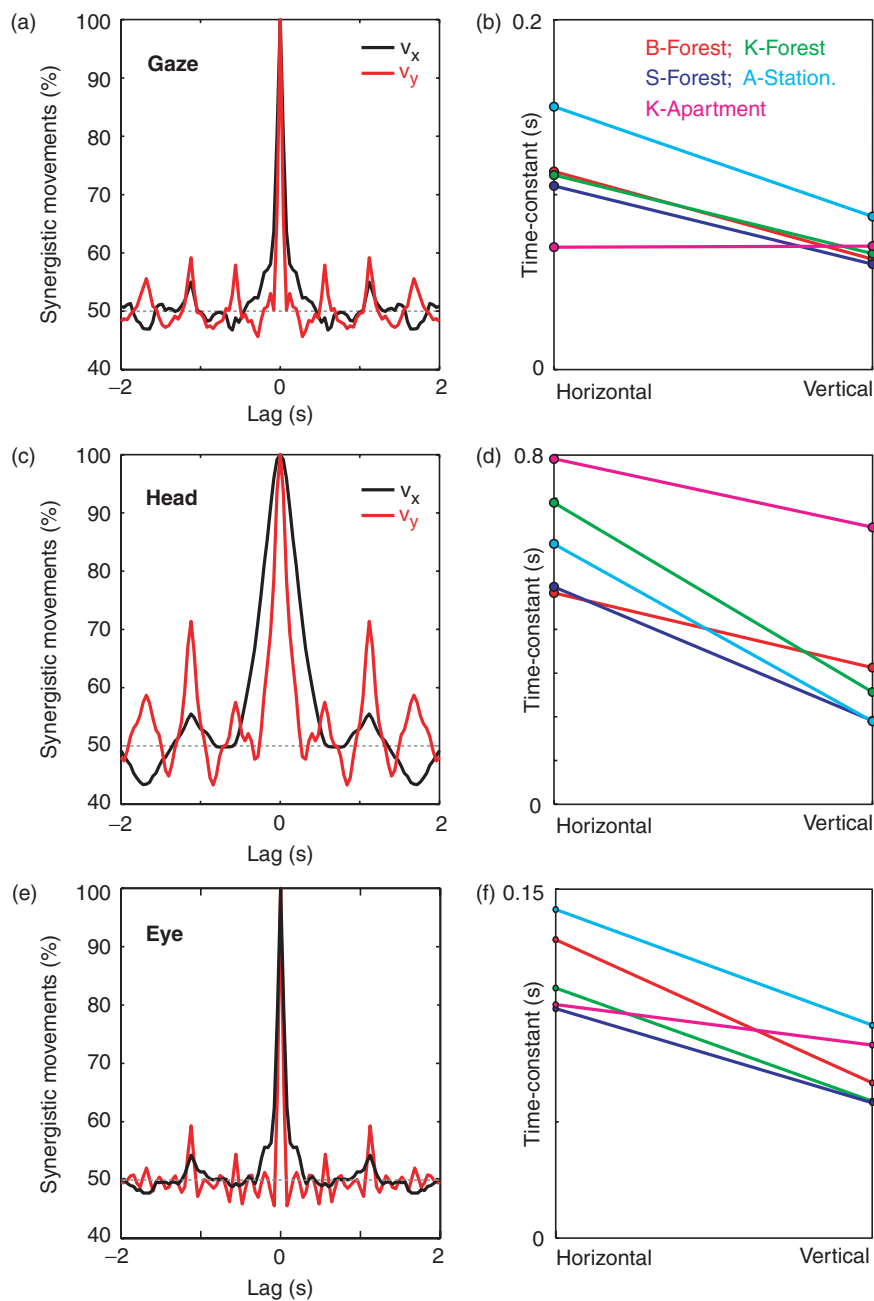


Figure 7. Time course of eye, head and gaze movements. In analogy to Figure 5, the fraction of synergistic movements (in the same direction) for different temporal lags. Panels on the left (a, c, e) depict example Subject S in the forest environment; panels to the right (b, d, f) the time constants (FWHM) for all subjects. Colors as in Figure 3c. (a–b) gaze, (c–d) head, (e–f) eye movements.

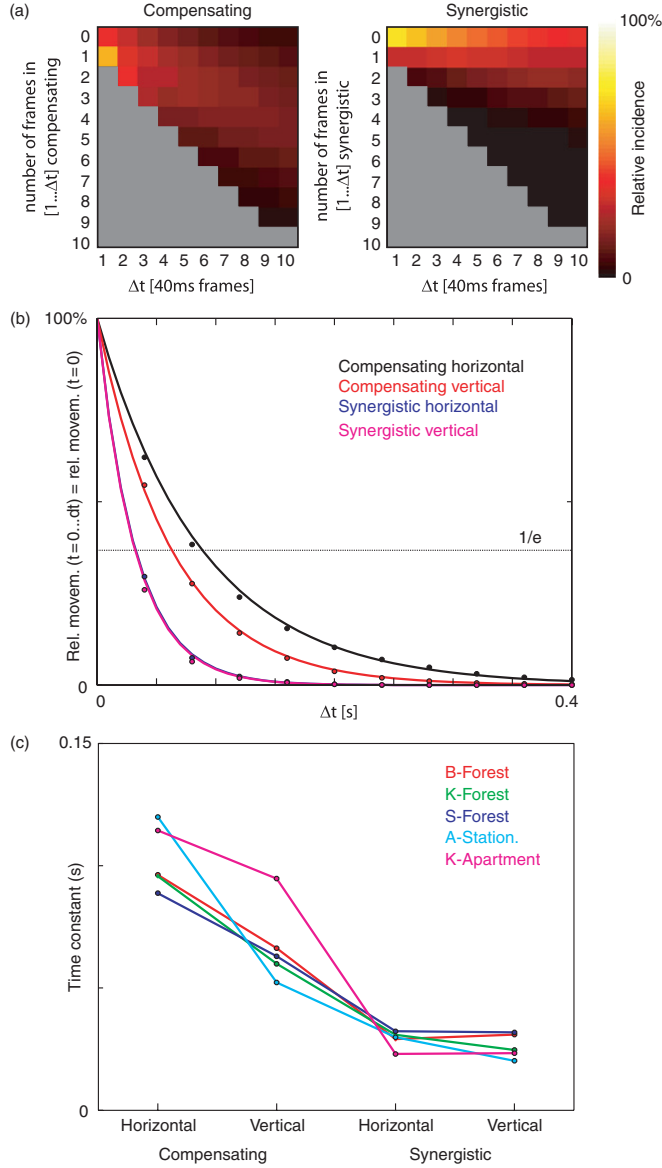


Figure 8. Time constants of compensatory and synergistic movements. (a) Left: Given a compensatory movement at time $t=t_0$, the n th column depicts the histogram of compensatory movements between $t=t_1$ and $t=t_n$, i.e., each row corresponds to the number of compensatory frames k between t_1 and t_n , histogrammed over the whole experiment for Subject S in the forest environment. Right: Same for synergistic movements. (b) Black: Given that the horizontal movement at $t=0$ was compensatory, the probability that all frames between $t=0$ and $t=\Delta t$ contained a compensatory horizontal movement; corresponds to $n=k$ diagonal of panel a. Red: Vertical compensatory movements. Blue: Synergistic horizontal movements (corresponds to diagonal of panel a, right). Magenta: Synergistic vertical movement. Dots: Data, solid lines: Exponential fit to data points shown. Time constants reported are time constants for the exponential fit, i.e., time at which the e^{-1} line is crossed. (c) Time constants (of exponential fit) for all subjects. Colors as in Figure 3c.

decay holds. For both, synergistic and compensatory movements, the vertical time constants were consistently smaller than the horizontal ones (Figure 8c), for compensating movements horizontal time constants averaged $103 \text{ ms} \pm 13 \text{ ms}$ and vertical time constants $67 \text{ ms} \pm 16 \text{ ms}$ ($p=0.005$, t -test); for synergistic movements these values reached $29 \text{ ms} \pm 3.6 \text{ ms}$ and $26 \text{ ms} \pm 0.5 \text{ ms}$, respectively ($p=0.33$, t -test). More importantly, there was a large difference between compensating and synergistic movements, which was highly significant for horizontal ($p=2 \times 10^{-6}$) and vertical ($p=6 \times 10^{-4}$) movements (Figure 8c). Thus, compensating movements had a circa 3-times longer time constant than synergistic movements. While synergistic movements were rapid, compensating movements tended to be prolonged. A single compensating epoch therefore yielded a larger fraction of time than a single synergistic epoch. This may have further overrated the role of compensating movements when only the aggregate durations of compensating and synergistic movements were measured.

Discussion

We simultaneously recorded gaze and head movements during free exploration of various natural environments. We showed that gaze, eye and head preferentially move along the cardinal axes, with the relative frequency of horizontal and vertical movements depending on the environment. When analyzing both directions separately, we found gaze to be slightly more stable than the eyes alone; for horizontal movements gaze was also more stable than the head by itself. Eye and head movements were not independent but co-occurred frequently. While about 80% of co-occurring eye and head movements opposed each other, a substantial fraction of nonsaccadic eye movements acted synergistically with head movements to adjust gaze. Such synergistic epochs were of shorter duration than compensatory movements, which may partly account for their rarer observation. Thus, nonsaccadic eye movements have a substantial synergistic component during natural exploration behavior. Consequently, periods in which eye movements perfectly compensate head and body movements are rare and natural visual input cannot be modeled as a sequence of static images interleaved by saccades. Instead the complex dynamics of gaze allocation through combined eye and head movements has to be taken into account when modeling the temporal statistics of natural stimuli.

The co-ordination of eye and head movements has been well studied in a variety of mammalian species. Collewyn (1977) did not observe any eye movements in the absence of head movements in freely behaving rabbits, suggesting that eye movements solely act to compensate or facilitate head movement-induced gaze shifts. Similarly, Guitton et al. (1984) showed that cats shift their gaze by an integrated movement of eye and head, but rarely exhibited isolated eye movements. While their results emphasize, as our present findings, the co-ordination of head and eye movements to adjust gaze, studies in rabbits and cats are obviously not directly transferable to humans. Rabbits do not have front-facing eyes and therefore only little binocular overlap. Cats do not have a pronounced area of high acuity like the primate fovea, but a wide “area centralis” of relatively high receptor density (Rapaport and Stone 1984). In nonhuman primates, eye-head co-ordination has

been well studied under laboratory conditions. Bizzi et al. (1971) demonstrated that monkeys who shift their gaze towards a target, first perform a voluntary eye-movement, which is followed by a head movement (although the movement signal arrives at the neck muscles earlier than at the eye) and then compensate for the moving head by reflexive eye movements. If the target location is predictable, however, the head movement (together with the compensatory eye-movement) precedes the saccadic eye-movement. Hence, task and target predictability play a decisive role in eye-head-co-ordination, which highlights the need for natural settings.

The choice of whether to orient the eye or the head towards a target depends in unrestrained monkeys on the amplitude of the gaze shift (Freedman and Sparks 1997). For small shifts ($<20^\circ$) eye movements are used exclusively. For larger shifts, the contribution of head movements linearly increases with the shift size. Humans generally use a combination of eye and head movements to rapidly change the direction of the line of sight (Barnes 1979; Guitton and Volle 1987). During large gaze shifts, the eye movement starts together with the head movement (but see Pelz et al. 2001). The eye, however, reaches the target faster, and is subsequently stabilized by the VOR, while the head continues to move. Under laboratory conditions, human eye-head co-ordination is characterized by an eye-only range for small gaze shifts on the order of 35° (Stahl 1999). This number is associated with a high inter-subject variability of 32° (SD), which – together with a high intra-subject consistency – suggests a division of humans into head movers and nonmovers (Fuller 1992).

The necessity for unrestrained recordings of eye movements to understand natural behavior has been noted before. Regarding recording stability under relatively free conditions, plenty of custom-developed and commercial eye-trackers available today have come a long way since the room-filling devices Buswell (1935) and his contemporaries needed to use. The developments of Collewijn, Steinman and colleagues took head-free recordings from rabbits (Collewijn 1977) and nonhuman primates to humans (Collewijn et al. 1992). With the advent of portable computers and digital cameras, several wearable systems have been developed (e.g., Pelz and Canosa 2001; Patla and Vickers 2003). Of the few laboratories that have taken eye-tracking technology into the field to study natural behavior, Land and colleagues pioneered eye-tracking in a variety of real-world situations, such as driving (Land and Lee 1994), tea making (Land et al. 1999), cricket playing (Land and McLeod 2000), and race car driving (Land and Tatler 2001). Similarly, Pelz and colleagues have investigated tasks like copying Lego structures (Pelz et al. 2001), putting in golf (Fairchild et al. 2001), and taking photographs (Babcock et al. 2002). The task constitutes a major factor in guiding eye movements (Buswell 1935; Yarbush 1967), and all the aforementioned studies ingeniously investigated the task's role in specialized situations, and often in highly trained individuals and in some cases compared to nonexperts: T. Scheckter driving his race car, batsmen hitting a ball, the British making tea, and so forth. Here we instead focused on free exploratory behavior in a variety of natural environments. By controlling a gaze-contingent camera online and in real time, we assessed gaze and head velocities with the same technique, which facilitated their direct comparison.

A robust system for prolonged viewing that minimally affects natural behaviour by size or weight and is developed at reasonable costs has to naturally accept some compromises with respect to spatial and temporal precision. Therefore, we did not aim at segregating saccadic eye movements from reflexive mechanisms at the present stage. Such a classification would require a higher sampling rate for the cameras to be reliable, which is currently prohibited by the cost of lightweight fast cameras but not incompatible with the setup in principle. Hence, we did not base our investigation on a classification of eye movements into saccades, reflexive movements, and so forth, but computed the relative movement irrespective of such categories. While this may be disadvantageous for a precise dissociation of different eye-movement systems, it presents a clear advantage for the question of interest: The contribution of (any type of) eye movements to the dynamics of sensory stimuli under natural conditions. Obviously, we do not argue against the role of large volitional eye movements in exploratory behavior. Our data show, however, that under natural conditions isolated, voluntary movements only account for a small part of the rich human gaze allocation behavior. Similarly, the spatial resolution of our robust outdoor eye-tracking device is below that of high precision measurements under very controlled laboratory conditions. Recently it has been demonstrated that small fixational eye movements play an important role in the processing of fine spatial detail in stimuli with natural scene like power spectra (Rucci et al. 2007). An interesting issue for further research is thus to clarify how different eye movement systems precisely interact during natural exploration.

In the present study we only addressed the relative contribution of eye movements to gaze, and did thus not separate different sources of head-in-world movements. In fact, the differences between angular distributions of head velocities for in- and outdoor environments point towards distinctive roles of body-in-world and head-on-body contributions. Marigold and Patla (2007) demonstrated that subjects who walk on uneven terrain, frequently fixate areas of the path they subsequently step on, try to maximize information to optimize safe foot placement, and use eye movements to continuously update ground information in a task-modulated fashion. By using additional body-fixed cameras, the present setup can be readily extended to distinguish different contributions to head movements. Although this is beyond the scope of the present study, this will be a fruitful step towards a complete understanding of naturalistic eye movements.

Quantitative models of human eye movements under natural conditions are usually based on models of attention, such as Koch and Ullman's (1985) saliency map. Hence they typically focus on predicting large voluntary eye movements from stimulus properties. Consequently, approaches that combine eye movements with models of early sensory coding (Li and Clark, 2004) or object recognition (Dickinson et al. 1997; Rutishauser et al. 2004) are also restricted to such large attention-driven gaze shifts. Our demonstration of the importance of synergistic eye and head movements in adjusting gaze and thus for the temporal dynamics of natural stimuli requires integrated models of gaze allocation and sensory coding. Since gaze allocation is a specific – and low-dimensional – motor task, this presents a special case of a more general challenge to close the loop between sensory coding and motor output. Future approaches thus need to combine coding principles like temporal coherence or spatio-temporal variants of sparse coding and ICA, with action control. Recent studies that use temporal coherence (Wyss et al. 2006) or

sparseness to train a robot's sensory system in a complex environment offer a promising option. To operate such systems as realistic models of human sensory coding, it is imperative to understand the temporal statistics of the input signal arriving at the human retina.

Acknowledgements

This work was financially supported by the Bavarian-Californian Technology Center (BaCaTeC), the Swiss National Science Foundation (grant no. PA00A-111447 to WE), Bayerische Forschungsförderung (FORBIAS, Az 567/03), and in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*. We would like to thank Thomas Dera for his work on the eye-tracker algorithms and David Wittmann for assembling the wearable computer. We are grateful to Judy Benson for critically reading the manuscript.

Note

- [1] “0 lag” refers to the true eye, head and gaze movements, not the camera signal. This means, the 36 ms lag have been taken into account by shifting the gaze video by one frame (40 ms).

Appendix

A. Validation of the gaze-driven camera

Several tests were performed in order to validate the gaze-driven camera system (Wagner et al. 2006 for details). In a first test the total latency was measured. Latency was defined as the time difference between the movement of the camera and of an artificial eye that was rotated by an additional actuator. This actuator was controlled independently by a profile generator. We used sinusoidal profiles at five frequencies ranging from 1 to 10 Hz. The eye tracker detected the artificial eye's pupil as it would with a real eye and moved the head-mounted camera accordingly. Two gyroscopes (ADXRS300, Analog Devices, Norwood, MA) were attached to the artificial eye and to the head-mounted camera in order to synchronously measure their angular velocities (Figure 9a). The phase shift of both angular velocities was calculated by fitting sine curves to the measured (noisy) signals by means of regression calculation (Figure 9b). Using the known frequency, the latency was computed from the phase shift. The results at each tested frequency are contained in Table I. Overall, a mean latency of 36 ms was obtained, which is constant within 2% for the frequency range tested (Table I).

The precision and resolution of the camera positioning was determined from the distance (in degrees) between fixation dots on a wall and the corresponding camera positioning targets. These targets were indicated by the luminous dot of the camera-attached calibration laser. After the calibration procedure, subjects were positioned 3 m in front of a poster and were asked to sequentially make saccades to 25 equidistant (8.5° distance) fixation dots arranged in a regular 5×5 matrix

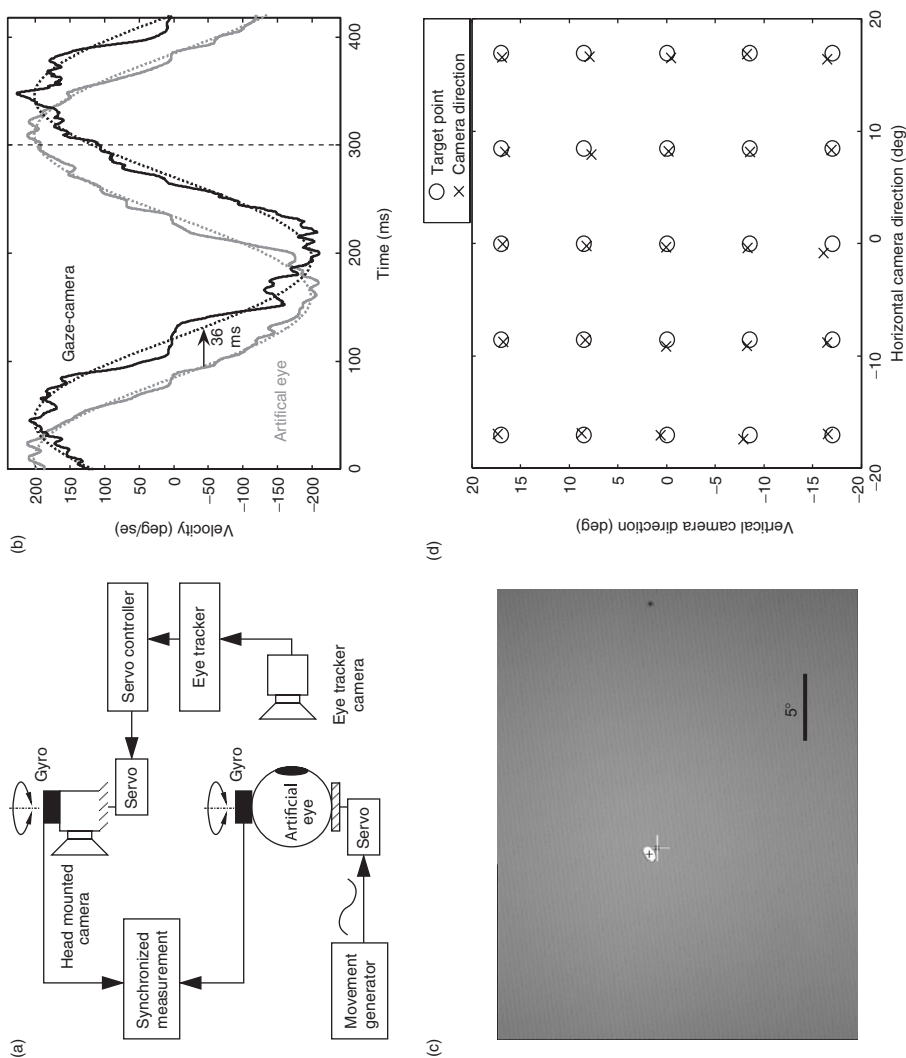


Figure 9. Accuracy and resolution of gaze camera positioning (adapted from Wagner et al. 2006). (a) A servo motor (right) rotates an artificial eye whose pupil is detected by an eye tracker system. Control commands for the head mounted camera are calculated from the pupil movements. Gyroscopes measuring the angular velocities are attached on both the artificial eye and the head mounted camera. (b) Solid black: Measured (raw) velocity of the gaze-camera; dashed gray: Corresponding sine fit. (c) Gaze camera view of the luminous calibration laser dot 2 out of 25 dark fixation dots (right) can be seen. Marker crosses show the results of automated target detection. The distance between the laser and the nearest fixation dot was used to estimate accuracy. (d) 25 fixations of one subject during calibration. Circles indicate the target points, which have to be fixated and crosses show the actual gaze camera alignment. The mean distance between targets and corresponding camera positions is a measure of the spatial accuracy of the complete system.

Table I. Latency of gaze centered camera in dependence of oscillation frequency.

Frequency [Hz]	Latency [ms]
1.0	35.6
2.0	35.7
3.3	36.3
5.0	36.1
10.0	36.1

(Figure 9c). The calibration laser was switched on during this task, and the video signal of the head-mounted camera was recorded to a video file. This file was later analyzed with an interactive video processing software to measure the distance (in degrees) between the projected laser dot and the black fixation dots (Figure 9d). This distance was used as the measure of accuracy. This measurement yielded a mean precision of 0.5° and a resolution (SD) of 0.25° . The resultant accuracy is on the order of the accuracy that can be expected from video eye tracking technology (van der Geest and Frens 2002).

B. Pupil position from camera-control signal

Whereas eye-position was not recorded explicitly in the nonlinear mapping from eye measurement to camera control, a posthoc remapping could be achieved from the recorded servo controls of the gaze camera. Figure 10 shows such a remapped trace for a validation sequence (Subject A, environment station). In these validation sequences, which we recorded at the beginning, in the middle and the end of each recording session, the observer was asked to successively fixate a grid depicted on a cardboard in approximately 2m distance. The measurement of the vertical component (Figure 10a, top) followed the eye, while the subject was subsequently fixating the equally spaced marks on the fixation grid (Figure 10b). The velocity estimated by the derivative of this signal (Figure 10a, middle) shows the same time course as the eye velocity estimated from the flow-field (Figure 10a, bottom), which in addition provides an absolute scale. This verifies the flow field estimation method used throughout. By manually inspecting all validation periods we verified that no shifts of the calibrated eye measurement to gaze camera occurred during a recording session. The calibration of the recording system remained stable during the prolonged periods of recording.

C. Validation of flow-field estimates under realistic conditions

In order to validate our velocity measurements and to obtain baseline results for optimally compensatory movements, we performed the following control measurements for two of the subjects (A and K) in the forest environment. The subjects fixated a point in the environment (a mark on a tree), while moving their head up and down (vertical compensation) or left-right (horizontal compensation) with increasing velocity. For the thus obtained videos, we performed exactly the same analysis as in the main experiment. Figure 11a depicts

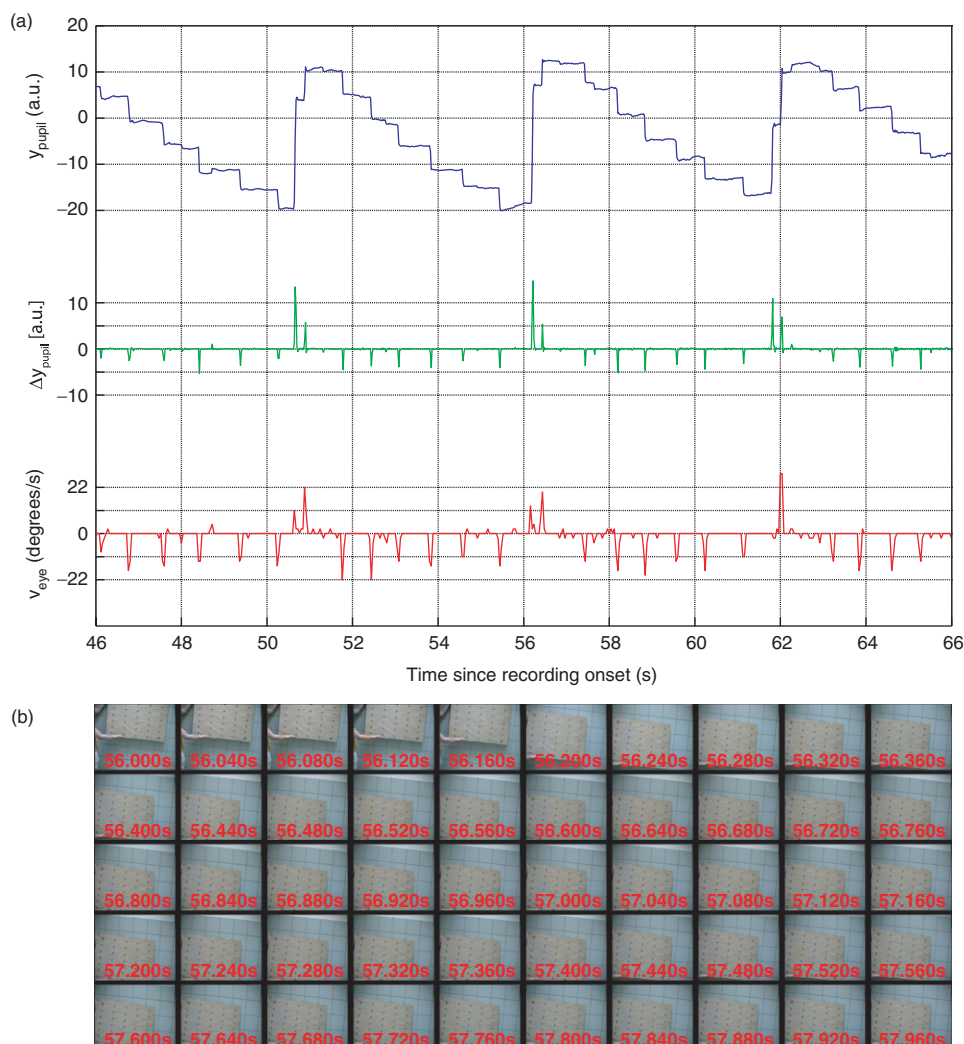


Figure 10. Eye-position and velocity from servo control. Recording of a validation sequence (Subject: A, environment: station) (a) Top (blue): Vertical eye-position as reconstructed from servo control signals. The mapping to visual angle is linear, but the scale and the zero point are arbitrary. Middle (green): Derivative of vertical eye-position. Bottom (red): $v_{\text{eye},y}$ as estimated from gazecamera and headcamera flow. (b) Two second excerpt (56–58 s) of the gaze camera during the validation sequence depicted in panel a. The laser-pointer, which is switched on only during validation, is directed parallel to the gaze camera and therefore remains at the center of the gaze camera image.

the time-course of horizontal velocities for the horizontal compensation experiment in Subject A. The velocity estimated from the flow fields of the head shows a sinusoidal modulation of horizontal head velocity. In contrast, the gaze camera shows only little flow. Note how static the gaze-centered camera remains on the fixated target (the central tree) despite the large and rapid movements of the head (bottom lines in Figure 11b). As expected from the design of the experiment,

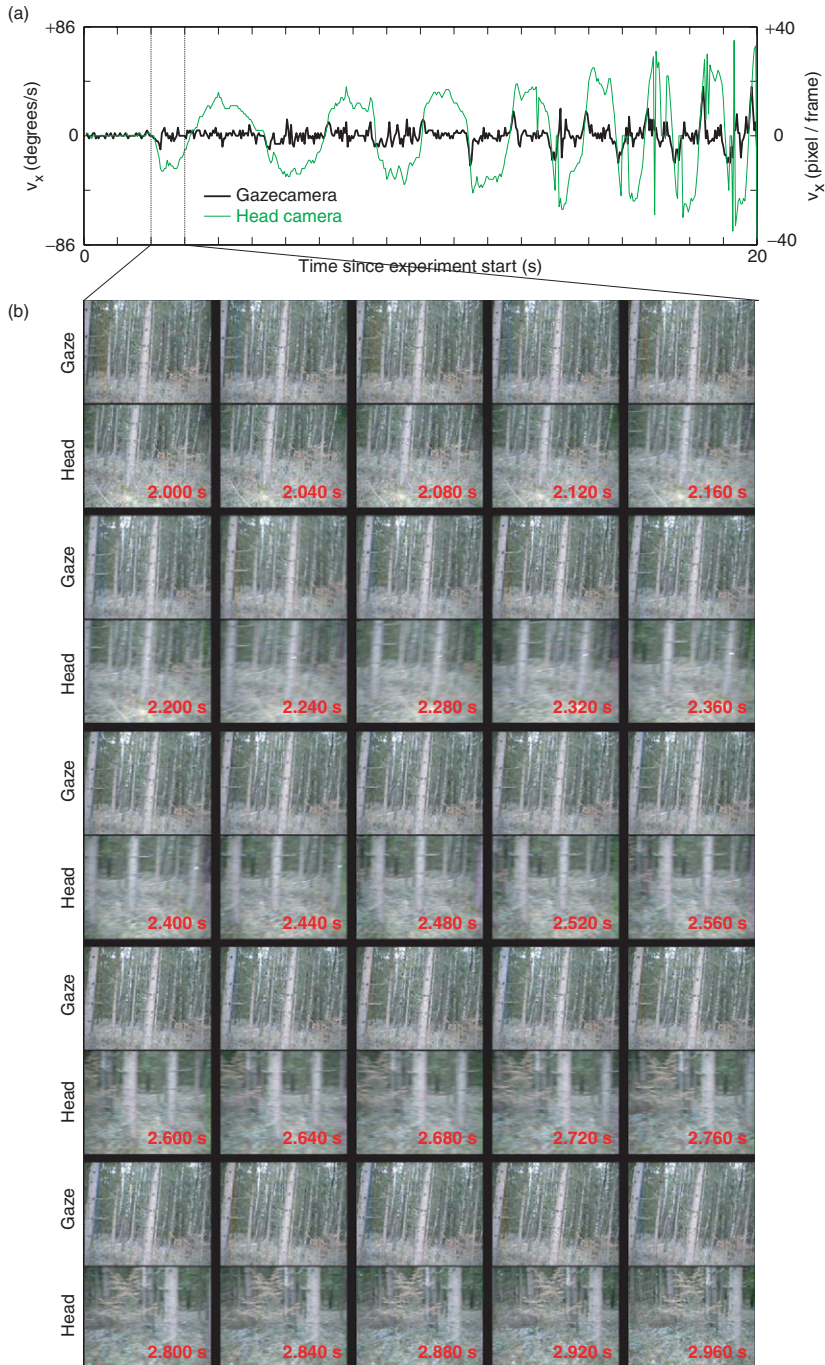


Figure 11. Control: Optimal compensation. Raw data for gaze, eye and head velocity for control experiment, in which subject A moved her head horizontally in the forest while fixating a tree. Markers as in Figure 3a. Note in the one second of the movie depicted in the bottom row (which is the third second of the experiment, as indicated by dotted lines), how little motion the gaze displays (the tree in the second row from the bottom remains centered), despite the massive movements of the head (bottom row).

combined eye and head movements account for most (80%) of the recorded time, and there were very few isolated head (3%) or eye movements (5%). If the movements were independent, these figures would reach 12 and 15%, respectively. Fixations (neither eye nor head movements) occur for 12% of the total time, most of which is accounted for by the first 2 s after experiment onset. If this period is excluded, fixations account for only 5% of the time. Similar results are obtained for vertical compensation in Subject A (4% isolated head, 9% isolated eye movement with 16 and 21%, respectively, independence predictions). In Subject K horizontal isolated movements occur with 3% (head) and 8% (eye); this is again about half of the independence prediction (9% and 14%). These numbers reach 3 and 16% for vertical compensation (independence prediction: 13 and 26%). Hence, the flow-field computation reliably estimates velocities of gaze and head and successfully returns the expected high fraction of combined eye and head movements. Does it also correctly reflect the primarily compensatory direction of eye and head movements? At 0 lag, nearly all movements are compensatory (A horizontal: 95.7%; A vertical: 98.1%; K horizontal: 94.8%; K vertical: 99.1%). This implies that less than 4% on average – and less than 6% in a worst case – of compensatory movements are misclassified. Note that this is a very conservative estimate, since imperfect compensation by the subject contributes to this number. Although we cannot objectively measure the relative contribution of the system noise as compared to the observer's imperfect stabilization, the measurement of Appendix B suggests that the system's contribution is not major. We, therefore, would assume that the average classification error is safely below 4%. This verifies that not only our velocity estimate faithfully reflects true movements, but that our measure of compensatory *versus* synergistic eye movements is also very robust. In summary, to the best of our knowledge the estimated velocities of eye and head and gaze obtained with our novel setup under free exploration conditions are sufficiently reliable to support the conclusions of the present study.

References

- Babcock J, Lipps M, Pelz JB. 2002 How people look at pictures before, during, and after image capture: Buswell revisited. In: Rogowitz B.E and Pappas T.N, editors. *Human Vision and Electronic Imaging VII*. Proceedings of SPIE. 466:34–47.
- Barnes GR. 1979. Vestibulo-ocular function during co-ordinated head and eye movements to acquire visual targets. *J Physiol* 287:127–147.
- Bell AJ, Sejnowski TJ. 1997. The 'independent components' of natural scenes are edge filters. *Vision Res* 37:3327–3338.
- Berkes P, Wiskott L. 2005. Slow feature analysis yields a rich repertoire of complex cell properties. *J Vis* 5(6):579–602.
- Betsch BY, Einhäuser W, Körding KP, König P. 2004. The world from a cat's perspective—statistics of natural videos. *Biological Cybernetics* 90:41–50.
- Bizzi E, Kalil RE, Tagliasco V. 1971. Eye-head coordination in monkeys: Evidence for centrally patterned organization. *Science* 173:452–454.
- Boltz RL, Harwerth RS, Smith EL. 1979. Orientation anisotropy of visual stimuli in rhesus monkey: A behavior study. *Science* 205:511–513.
- Brandt T, Glasauer S, Schneider E. 2006. A Third Eye for the Surgeon. *J Neurol Neurosurg Psychiatry* 77:278.
- Buswell GT. 1935. *How people look at pictures. A study of the psychology of perception in art*. Chicago, IL: The University of Chicago Press.

- Carpenter RHS. 1988. *Movements of the Eyes*. 2nd ed. London: Pion.
- Collewijn H. 1977. Eye- and head movements in freely moving rabbits. *J Physiol* 266:471–498.
- Collewijn H, Martins AJ, Steinman RM. 1983. Compensatory eye movements during active and passive head movements: Fast adaptation to changes in visual magnification. *J Physiol* 340:259–286.
- Collewijn H, Steinman RM, Erkelens CJ, Pizlo Z, Van der Steen. 1992. The effect of freeing the head on eye movement characteristics during 3-D shifts of gaze and tracking. In: Berthoz A, Vidal PP, Graf W, editors. *The Head-Neck Sensory Motor System*. Oxford: University Press, London. pp. 412–418.
- Coppola DM, White LE, Fitzpatrick D, Purves D. 1998a. Unequal representation of cardinal and oblique contours in ferret visual cortex. *Proc Natl Acad Sci, USA* 95:2621–2623.
- Dickinson S, Christensen H, Tsotsos J, Olofsson G. 1997. Active object recognition integrating attention and viewpoint control. *Comp Vis Image Und* 63:239–260.
- Dong DW. 1997. Spatiotemporal coupling and scaling of natural images and human visual sensitivities. In: Mozer MC, Jordan MI, Petsche T, editors. *Advances in Neural Information Processing Systems (NIPS 1996)*. Cambridge MA: MIT Press. pp. 869–865.
- Dong DW, Atick JJ. 1995a. Statistics of natural time varying images. *Network: Computation in Neural Systems* 6:345–358.
- Dong DW, Atick JJ. 1995b. Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems* 6:159–178.
- Einhäuser W, Hipp J, Eggert J, Körner E, König P. 2005. Learning viewpoint invariant object representations using a temporal coherence principle. *Biological Cybernetics* 93:79–90.
- Einhäuser W, Kayser C, Körding KP, König P. 2003. Learning distinct and complementary feature-selectivities from natural colour videos. *Rev Neurosci* 14:43–52.
- Einhäuser W, Kayser C, König P, Körding KP. 2002. Learning the invariance properties of complex cells from their responses to natural stimuli. *Eur J Neurosci* 15:475–486.
- Essock EA, DeFord JK, Hansen BC, Sinai MJ. 2003. Oblique stimuli are seen best (not worst!) in naturalistic broad band stimuli. *Vision Res* 43:1329–1335.
- Fairchild MD, Johnson GM, Babcock J, Pelz JB. 2001. Is Your Eye on the Ball? Eye Tracking Golfers while Putting. (<http://www.cis.rit.edu/people/faculty/fairchild/>).
- Freedman EG, Sparks DL. 1997. Eye-Head Coordination During Head-Unrestrained Gaze Shifts in Rhesus Monkeys. *J Neurophysiol* 77:2328–2348.
- Field DJ. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4:2379–2394.
- Földiák P. 1991. Learning Invariance from Transformation Sequences. *Neural Comput* 3:194–200.
- Fuller HJ. 1992. Head movement propensity. *Exper Br Res* 92:152–164.
- Guittot D, Volle M. 1987. Gaze control in humans: Eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *J Neurophysiol* 58:427–459.
- Guittot D, Douglas RM, Volle M. 1984. Eye-head coordination in cats. *J Neurophysiol* 52:1030–1050.
- Golz J, MacLeod DIA. 2002. Influence of scene statistics on colour constancy. *Nature* 415:637–640.
- Hancock PJB, Baddeley RJ, Smith LS. 1992. The principal components of natural images. *Network: Computation in Neural Systems* 3:61–70.
- Hansen BC, Essock EA, Zheng Y, DeFord JK. 2003. Perceptual anisotropies in visual processing and their relation to natural image statistics. *Network: Computation in Neural Systems* 14:501–526.
- Hashimoto W. 2003. Quadratic forms in natural images. *Network: Computation in Neural Systems* 14:765–788.
- Hayhoe M, Ballard D. 2005. Eye movements in natural behavior. *Trends Cogn Sci* 9(4):188–194.
- Hipp J, Einhäuser W, Conradt J, König P. 2005. Learning of somatosensory representations for texture discrimination using a temporal coherence principle. *Network: Computation in Neural Systems* 16:223–238.
- Howe CQ, Purves D. 2002. Range image statistics can explain the anomalous perception of length. *Proceedings of the National Academic Sciences, USA* 99:13184–13188.
- Koch C, Ullman S. 1985. Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum Neurobiol* 4:219–227.
- Körding KP, Kayser C, Einhäuser W, König P. 2004. How are complex cell properties adapted to the statistics of natural stimuli?. *J Neurophysiol* 91:206–212.
- Kowler E, Pizlo Z, Zhur GJ, Erkelens C, Steinman RM, Collewijn H. 1992. Coordination of head and eyes during the performance of natural (and unnatural) visual tasks. In: Berthoz A, Vidal PP, Graf W, editors. *The Head-Neck Sensory Motor System*. London: Oxford University Press. pp. 419–426.
- Land MF, Lee DN. 1994. Where we look when we steer. *Nature* 369:742–744.

- Land MF, Mennie N, Rusted J. 1999. The roles of vision and eye movements in the control of activities of daily living. *Perception* 28:1311–1328.
- Land MF, McLeod P. 2000. From eye movements to actions: How batsmen hit the ball. *Nat Neurosci* 3:1340–1345.
- Land MF, Hayhoe M. 2001. In what ways do eye movements contribute to everyday activities? *Vision Res* 41:3559–3565.
- Land MF, Tatler BW. 2001. Steering with the head. The visual strategy of a racing driver. *Curr Biol* 11:1215–1220.
- Li M, Clark JJ. 2004. A Temporal Stability Approach to Position and Attention-Shift-Invariant Recognition. *Neural Comput* 16:2293–2321.
- Long F, Purves D. 2003. Natural scene statistics as the universal basis of colour context effects. *Proceedings of National Academic Sciences, USA* 100:15190–15193.
- Marigold DS, Patla AE. 2007. Gaze fixation patterns for negotiating complex ground terrain. *Neuroscience* 144:302–313.
- Moeller GU, Einhäuser W, Conradt J, König P. 2003. Exploring the stability of the retinal image in the freely behaving cat. *Society for Neuroscience Annual Meeting Abstract* 386.5.
- Olshausen BA, Field DJ. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609.
- Patla AE, Vickers JN. 2003. How far ahead do we look when required to step on specific locations in the travel path during locomotion?. *Exp Brain Res* 148:133–138.
- Pelz JB, Hayhoe MM, Loeber R. 2001. The coordination of eye, head, and hand movements in a natural task. *Exper Br Res* 139:266–277.
- Pelz JB, Canosa R, Babcock J, Kucharczyk D. 2000. Portable eyetracking: A study of natural eye movements. In: *Proceedings of the SPIE, Vol. 3959: Human Vision and Electronic Imaging*. pp. 566–583. SPIE.
- Rapaport DH, Stone J. 1984. The area centralis of the retina in the cat and other mammals: Focal point for function and development of the visual system. *Neuroscience* 11:289–301.
- Rucci M, Iovin R, Poletti M, Santini F. 2007. Miniature eye movements enhance fine spatial detail. *Nature* 447(7146):851–854.
- Ruderman DL, Bialek W. 1994. Statistics of natural images: Scaling in the woods. *Phys Rev Lett* 73:814–817.
- Rutishauser U, Walther D, Koch C, Perona P. 2004. Is bottom-up attention useful for object recognition? *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* 2:37–44.
- Schneider E, Bartl K, Bardins S, Dera T, Boning G, Brandt T. 2005. Eye Movement Driven Head-Mounted Camera: It Looks Where the Eyes Look. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*. pp. 2437–2442.
- Schneider E, Bartl K, Dera T, Boning G, Wagner P, Brandt T. 2006. Documentation and teaching of surgery with an eye movement driven head-mounted camera: See what the surgeon sees and does. *Stud Health Technol Inform* 119:486–490.
- Stahl JS. 1999. Amplitude of human head movements associated with horizontal saccades. *Exper Br Res* 126:41–54.
- Stringer SM, Rolls ET. 2002. Invariant object recognition in the visual system with novel views of 3D objects. *Neural Comput* 14:2585–2596.
- van der Geest JN, Frens MA. 2002. Recording eye movements with video-oculography and scleral search coils: A direct comparison of two methods. *J Neurosci Meth* 114:185–195.
- van der Schaaf A, van Hateren JH van. 1996. Modelling the power spectra of Natural images. *Vision Res* 36:2759–2770.
- van Hateren JH, Ruderman DL. 1998. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc Biol Sci* 265:2315–2320.
- Vinje WE, Gallant JL. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287:1273–1276.
- Wagner P, Bartl K, Günthner W, Schneider E, Brandt T, Ulbrich H. 2006. A pivotable head mounted camera system that is aligned by three-dimensional eye movements. *Proceedings of Eye tracking research & applications 2006*. pp. 117–124.

- Wertheimer M. 1938. *Laws of Organizations in Perceptual Forms* London: Harcourt: Brace & Jovanowitch.
- Wyss R, König P, Verschure PF. 2006. A model of the ventral visual system based on temporal stability and local memory. *PLoS Biol* 4(5):e120.
- Yarbus AL. 1967. *Eye movements and vision* New York: Plenum Press.