A Framework for the Study of Vision in Active Observers

Carlo Nicolini[a], Carlo Fantoni[a,b], Giovanni Mancuso[a,c], Robert Volcic[a] and Fulvio Domini[a,d]

[a] Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy; [b] Department of Life Sciences, Psychology Unit "Gaetano Kanizsa", University of Trieste, Trieste, Italy; [c] CIMeC, University of Trento, Rovereto, Italy; [d] Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA

ABSTRACT

We present a framework for the study of active vision, i.e., the functioning of the visual system during actively self-generated body movements. In laboratory settings, human vision is usually studied with a static observer looking at static or, at best, dynamic stimuli. In the real world, however, humans constantly move within dynamic environments. The resulting visual inputs are thus an intertwined mixture of self- and externally-generated movements. To fill this gap, we developed a virtual environment integrated with a head-tracking system in which the influence of self- and externally-generated movements can be manipulated independently. As a proof of principle, we studied perceptual stationarity of the visual world during lateral translation or rotation of the head. The movement of the visual stimulus was thus parametrically tethered to self-generated movements. We found that estimates of object stationarity were less biased and more precise during head rotation than translation. In both cases the visual stimulus had to partially follow the head movement to be perceived as immobile. We discuss a range of possibilities for our setup among which the study of shape perception in active and passive conditions, where the same optic flow is replayed to stationary observers.

Keywords: Active vision, head tracking, virtual reality, optic flow, 3D, shape/depth/motion perception, perception and action, structure from motion

1. INTRODUCTION

In our every-day life, our brain faces a variety of external stimulations and is continuously asked to interact with an ever-changing, dynamical world. We unconsciously cope with multisensorial inputs that allow us to produce adequate responses to the environment. In fact, vision is one of the senses that permits such efficient responses, but in order for it to be effective it has to be integrated with other inputs from the body.

Typical laboratory experiments, designed to study visual mechanisms, do not fully replicate the complexity of a real-life environment, since the active and dynamic nature of an observer is often ignored. Instead, vision is usually studied in a context in which static observers look at static or, at best, dynamic stimuli. However, many important perceptual phenomena, like the perception of a stable world, are the result of the integration of visual and vestibular signals.

Our goal was to create a set of tools which can be used to study active vision, in contexts that fit the active nature of the human observer. In particular, we are interested in the functioning of the visual system during

---

Corresponding author: Carlo Nicolini: E-mail: carlo.nicolini@iit.it, Telephone: +39 0464808711
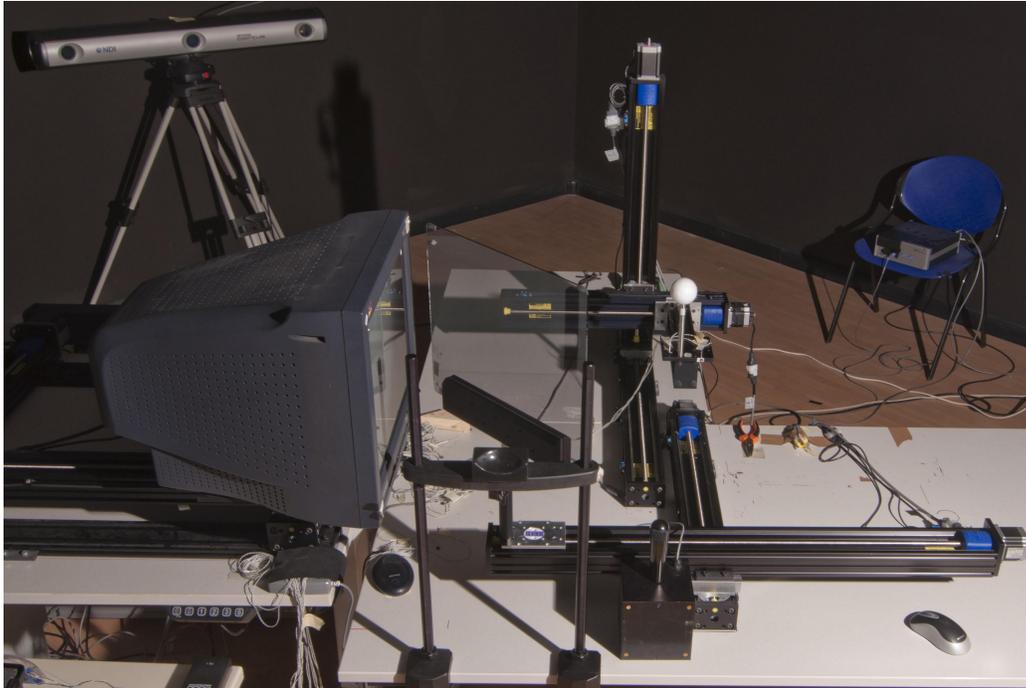
Figure 1. Picture of the active vision lab. The monitor plane is reflected on the slanted mirror. A chinrest is positioned in front of the reflected image to keep the subject fixed and aligned during passive vision. During active vision, the chinrest is removed to allow free movements of the head. Two actuators move the monitor on x and z axes, two other actuators can displace the mirror on x and z axes, while a system of three linear motors allows to move an object (e.g., the white sphere) along the three axes to provide tactile feedback. In the back, one of the two Optotrak position sensors is shown.

self-generated movements. We improved on previous methods[1,2] by developing a framework that is both simpler in its architecture and unnoticeable to the observer.

In our system the head is tracked with an infrared tracking system that communicates in quasi real time the spatial position of the observer's vantage point. The virtual world projection is accordingly updated on a high-resolution monitor, precisely located, through a system of motors, at the desired distance from the observer. The rendered images, the tracking system, as well as the motors are all coordinated by a dedicated software, specifically designed for psychophysical experiments.

Here we describe an example of how this system can be used for studying the problem of perceptual stability: how do we perceive a stable world in spite of constantly changing retinal projections? Specifically, how can observers tell whether a shift in retinal projection is produced by their own movement or by a dynamic change of the external environment?

In order to achieve perceptual stability of a static environment the brain must correctly integrate vestibular signals, which encode the observer's self-motion, with retinal signals.[3–5] Only when the changes in retinal projections are accurately accounted for by the sensed ego-motion information a stable environment can be perceived as such. How and to what extent the brain integrates visual and vestibular information is still under debate. In order to answer this question, we studied whether an isolated object is perceived as static (or moving) in the world while the observer is undergoing two types of head movements: linear (lateral translation) or angular (vertical rotation).

2. VIRTUAL REALITY SETUP

2.1 Hardware

The experimental setup (Figure 1) consists of a system of linear actuators (Velmex Inc.) that fully adapts the experimental environment to any required viewing geometry, a 19-inch CRT monitor (ViewSonic 9613, 19W) connected to a Dell T3400 workstation, a high-quality front-silvered mirror and an Optotrak Certus motion tracker system with two position sensors (Northern Digital Inc.) with a set of 20 infrared emitting diodes. A pair of orthogonal actuators can displace the monitor along x and z axis (Figure 2a). A triplet of actuators is mounted on the experimental table to freely move a support platform that can provide tactile feedback in reaching and grasping experiments.[6, 7] Another couple of linear actuators controls the position of the front-silvered mirror on x and z axes. This configuration allows the experimenter to change the effective distance from the pupil to the center of the projection screen (focal distance) in a range from 35 cm to 250 cm and to move physical objects in a large range (Figure 2a). Our setup, differently from other mechanical systems[1, 2] does not limit the freedom of movements of the participant in any way, since the participant can freely move his head in a large range and exploit all six degrees of freedom without negatively affecting the stereoscopic perception,[8] nor being constrained by mechanical joints.

In traditional devices for virtual reality, such as head mounted displays or various types of stereoscopes, a visual conflict is introduced between the viewing distance of the rendered 3D scene and the viewing distance specified by vergence of the eyes and accommodation. This artifact is shown to cause systematic distortions of 3D vision,[9] favoring a flattened 3D representation as well as increased visual discomfort.[10] Our setup is built to allow dynamic adjustments of the projection screen distance during experiments with sub-millimeter precision. Stimuli are viewed reflected by the mirror placed in front of the observer's central viewing position and slanted 45° away from the monitor and the observer's inter-ocular axis.

Real-time tracking of the head position as well as other body parts is done through an high-speed PCI interface with the Optotrak motion tracker APIs (Application Programming Interfaces) that provides a continuous stream of high precision position data. The Optotrak frame rate is made to correspond to the monitor refresh rate at 100 Hz. The advantages of having a 100 Hz refresh rate are a negligible time lag[11, 12] between user movements and visual feedback, an optimal integration with shutter glasses frame rate with almost total absence of flickering effects in artificially lit environments and, last, sufficiently large data collections for detailed and precise kinematics reconstruction and analysis. When measured with microsecond accuracy clock timer, the average frame interval during a typical experiment is 10.00 ± 0.15 ms.

For experiments that need stereoscopic vision, we adopted the quad-buffering technique thus halving the frame rate. The synchronization of the active shutter glasses (liquid crystal Cambridge Research System FE-1 goggles), to the sequential left and right eye frames is accomplished using a VESA-standard 3-pin stereo connector on a NVidia Quadro FX 4600 GPU. The high refresh rate achieved by our system maximally reduces the crosstalk of stereo goggles and minimizes retinal persistence.

Finally, in order to control for participant position, pose and current trial information, during the experiments, a stream of experimental data coming from the main computer is shown on a secondary monitor via TCP/IP transmission protocol.
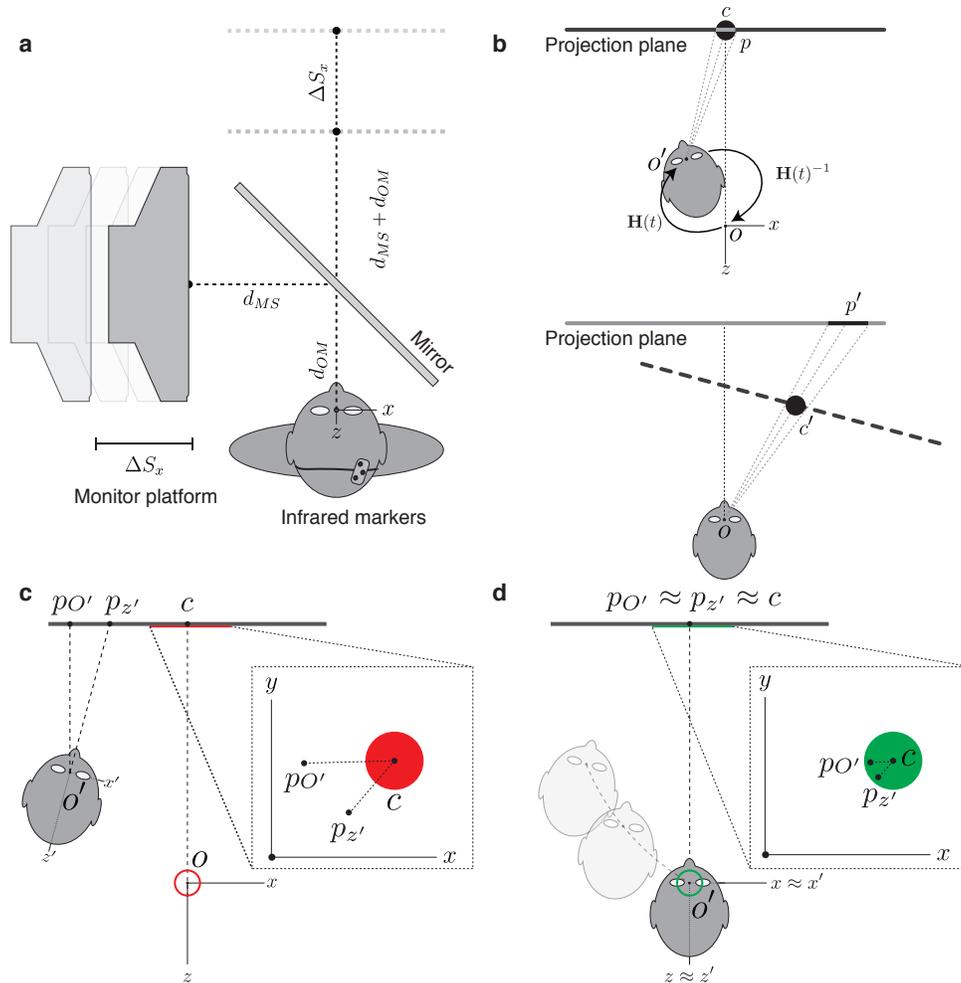
Figure 2. **a**, a scheme of the experimental setup: simulated objects are visualized in the space behind the mirror and the observers can freely move their head. Observers' head orientation and position are tracked with three infrared markers worn on the back of the head. The focal distance is computed as the sum of the origin-mirror distance d_{OM} and the mirror-screen distance d_{MS} and can be changed by ΔS_x by moving the monitor platform. **b**, active (top) and passive (bottom) observers. When the observer moves from O to O' (head translation and rotation described by $\mathbf{H}$), an object positioned in c generates the projection p on the projection plane. Conversely, in the passive case, a moving projection p' is generated to the immobile observer positioned at O reproducing the same visual input as in the active case. The black dashed line through c' represents the virtual projection plane of the corresponding active vision condition shown on top. Note that $||\overline{O'c}||$ always equals $||\overline{Oc'}||$ since $c' = \mathbf{H}^{-1}c$ and $O' = \mathbf{H}O$. As a consequence $||O - c'|| = ||\mathbf{H}^{-1}(O' - c)||$ given that affine transformations preserve lengths. **c,d**, the alignment procedure. Participants have to move from O' to O within a tolerance volume (red circle). Both the intersection of their optical axis with the projection plane ($\mathbf{p}_{z'}$) and the parallel projection of their cyclopean eye $\mathbf{p}_{O'}$ must lie inside a circle centered in c as shown in the two insets that represent the XY view of projection plane. After this procedure the head pose matrix $\mathbf{H}(t)$ is aligned to the tracker space $Oxyz$.

2.2 Software

We developed a specific cross-platform, open-source library written in C++, CNCSVision*, that is used: (1) to generate static and dynamic visual stimuli; (2) to interface with the Optotrak motion tracker; (3) to remotely

---

*We release CNCSVision under GPLv3 license. `https://github.com/CarloNicolini/CNCSVision`

command the linear actuators through serial ports; (4) to replay to an immobile observer the visual stimulation that a moving observer generates; (5) to store different types of behavioural responses (i.e., body movements kinematics, keyboard key presses). Our library is based on FreeGLUT[†] as the OpenGL layer for windowing and mouse/keyboard input manager. Modern OpenGL features (display lists and vertex buffer objects) and GLSL shading language are employed in order to minimize the computational load of the CPU and increase the responsiveness of the system, moving the computation and display of visual stimuli from the CPU to the GPU.

2.3 Off-Axis Generalized Perspective

The projection stage of our head tracking system has been modified with respect to the usual perspective projection technique implemented in OpenGL in order to avoid distortions due to an off-axis view of a moving observer relative to the projection plane center. Our implementation described in Algorithm 1 is based on the generalized perspective projection method of Kooima[13] that computes the projection matrix centering the camera model on the user's eye position.

The resulting perspective projection matrix is a more general $\mathbf{P}' = \mathbf{P} \cdot \mathbf{M}^T \cdot \mathbf{T}$, where $\mathbf{M}$ and $\mathbf{T}$ are respectively a rotation matrix and a translation matrix describing the position of the projection plane in tracker space coordinates. This projection model is generalized to achieve correct binocular stereopsis by interlacing frames correspondent to two alternating centers of projection.

We reinterpreted the method of Kooima[13] and extended it to account for the passive viewing of the visual motion generated by a moving/active observer. In order to do that we computed the correct off-axis perspective projection matrix for both moving/active and immobile/passive observers (presented with a replay of the visual motion that they themselves generated during active viewing), by introducing an additional affine transformation $\mathbf{A}(t) = [\mathbf{R}(t), \mathbf{x}(t)]$ that describes the motion of the projection screen corners in tracker space coordinates. Such a generalization of the method is necessary to disentangle the contribution of retinal and extra-retinal signals resulting from head movements in perception of 3D shape: the only difference between the active and the passive viewing of the same 3D shape consists indeed in the presence/absence of egomotion.

For active observers, the projection plane is aligned on XY plane in the tracker space $Oxyz$ (Figure 2b, top), so that $\mathbf{A}(t) \equiv \mathbf{I}$, i.e., no transformation is applied to it during observers' head movements. For passive observers (Figure 2b, bottom), $\mathbf{A}(t)$ corresponds to the current head pose $\mathbf{H}(t) = [\mathbf{R}_h(t), \mathbf{x}_h(t)]$, where $\mathbf{R}_h(t)$ is the 3×3 rotation matrix of the head with respect to a frame aligned to tracker space and $\mathbf{x}_h(t)$ is the current translation of the head from the tracker space origin.

2.4 Tracking Head and Arm Movements

Given that placing a physical marker on a particular body segment is frequently impossible due to both ergonomic and technical factors, we adopted a method to track arbitrary points on a rigid-body that is based on the initial relation between these points and a set of reference markers. We use this method for two main purposes: first, for the extraction of the eyes nodal points coordinates and the head pose, and second, for the extraction of finger tips coordinates.

For the first purpose, the participant interpupillary distance is measured by the experimenter and then the nodal points are computed as the points that lay equidistant from the center of the segment connecting two markers on the opposite sides of a pair of goggles worn by the participant. Those points are then updated in real

---

[†]http://freeglut.sourceforge.net/

Algorithm 1: Generation of off-axis perspective projection matrix for active or passive observers.

Input: Screen corners $\mathbf{p}_a, \mathbf{p}_b, \mathbf{p}_c$ in tracker space coordinates, center of projection (observer's eye) $\mathbf{p}_e$, an affine transformation matrix $\mathbf{A}(t) = [\mathbf{R}(t), \mathbf{x}(t)]$ where $\mathbf{R}$ is a 3×3 rotation matrix, $\mathbf{x}(t)$ is a translation, near and far plane distances n, f respectively.

Output: OpenGL Perspective projection matrix $\mathbf{P}'$ for active or passive observers.

Apply affine transformation $\mathbf{A}(t) = [\mathbf{R}(t), \mathbf{x}(t)]$ to monitor vertices $\mathbf{p}_i$ to obtain moving virtual vertices $\mathbf{p}'_i$:

$$\mathbf{p}'_i(t) = \mathbf{R}(t) \cdot \mathbf{p}_i + \mathbf{x}(t) \qquad \forall i = \{a, b, c\}$$

Compute an orthonormal base $(\mathbf{v}_r, \mathbf{v}_u, \mathbf{v}_n)$ for the virtual moving monitor plane:

$$\mathbf{v}_r = \frac{\mathbf{p}'_b(t) - \mathbf{p}'_a(t)}{||\mathbf{p}'_b(t) - \mathbf{p}'_a(t)||} \quad \mathbf{v}_u = \frac{\mathbf{p}'_c(t) - \mathbf{p}'_a}{||\mathbf{p}'_c(t) - \mathbf{p}'_a||} \quad \mathbf{v}_n = \frac{\mathbf{v}_r \times \mathbf{v}_u}{||\mathbf{v}_r \times \mathbf{v}_u||}$$

Determine off-axis frustum extent (l, r, t, b), where $d = -\mathbf{v}_n \cdot (\mathbf{p}'_a - \mathbf{p}_e)$:

$$l = \mathbf{v}_r \cdot (\mathbf{p}'_a - \mathbf{p}_e)n/d \qquad r = \mathbf{v}_r \cdot (\mathbf{p}'_b - \mathbf{p}_e)n/d \qquad b = \mathbf{v}_u \cdot (\mathbf{p}'_a - \mathbf{p}_e)n/d \qquad t = \mathbf{v}_u \cdot (\mathbf{p}'_c - \mathbf{p}_e)n/d$$

Obtain the off-axis projection matrix by matrix multiplication, $\mathbf{P}'(t) = \mathbf{P}(t) \cdot \mathbf{M}^T(t) \cdot \mathbf{T}(t)$ where $\mathbf{P}$ is the standard OpenGL perspective projection matrix, $\mathbf{M}$ and $\mathbf{T}$ are two 4×4 matrices to map onto the screen space coordinate system:

$$\mathbf{P} = \begin{bmatrix} \frac{2n}{r-l} & 0 & \frac{r+l}{r-l} & 0 \\ 0 & \frac{2n}{t-b} & \frac{t+b}{t-b} & 0 \\ 0 & 0 & -\frac{f+n}{f-n} & -\frac{2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad \mathbf{M} = \begin{bmatrix} v_{rx} & v_{ux} & v_{nx} & 0 \\ v_{ry} & v_{uy} & v_{ny} & 0 \\ v_{rz} & v_{uz} & v_{nz} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{T} = \begin{bmatrix} 0 & 0 & 0 & -p_{ex} \\ 0 & 0 & 0 & -p_{ey} \\ 0 & 0 & 0 & -p_{ez} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

---

time by applying the current pose transformation $\mathbf{H}(t)$ of the observer. We applied the same method to extract coordinates of other body parts where finger tips coordinates are needed but infrared markers can't be applied directly on them because of optical occlusion.[7] We use the Umeyama[14] method to compute rigid-body affine transformations with the best numerical stability. All matrix algorithms are implemented with the Eigen C++ library.[15]

2.5 Alignment of the Observer

When absolute orientation of the observer's head with respect to the world-centered reference frame $Oxyz$ (as shown in 2a) is needed, we run a first calibration phase (Figure 2c,d) to obtain the axis-aligned pose matrix $\mathbf{H}(t)$. In this phase, the observer is asked to align the orientation of the sagittal plane of their head (extracted relatively to a triplet of markers attached to the back of their head) so to be perpendicular to the projection plane. This process is accomplished by means of an interactive alignment procedure during which the participants must align the orthogonal projection of their cyclopean eye and a point representing their gaze direction inside a central reference circle displayed on the virtual projection plane. Through this alignment procedure, instantaneous absolute measures of head rotation and translations are obtained.

Having an axis-aligned pose matrix $\mathbf{H}(t)$ makes possible to systematically manipulate the stimuli position and orientation by means of an affine transformation $\mathbf{A}_s = [\mathbf{R}_s, \mathbf{x}_s]$. A stimulus aligned with XY plane with a constant

distance $\tilde{z}$ from the observer is described by $\mathbf{A}_s = [\mathbf{I}, \mathbf{x}_h - (0, 0, \tilde{z})^T]$, while a stimulus which is always aligned and centered to the observer optical axis at constant distance $\tilde{z}$ is described by $\mathbf{A}_s = [\mathbf{R}_h, \mathbf{x}_h - \mathbf{R}_h \cdot (0, 0, \tilde{z})^T]$. A stimulus centered and fixed at focal distance f_z, which is rotating around its y axis in a way proportional to the visual angle spanned by the observer translation about the x axis is described by $\mathbf{A}_s = [\mathbf{R}_y(\theta), (0, 0, f_z)^T]$ where $\theta = \arctan(x_{hx}/f_z)$ and $\mathbf{R}_y$ is the standard rotation matrix around y axis.

In several studies in our lab we have used this tethering technique by systematically manipulating the orientation and position of distal objects, to selectively disentangle the contribution of the different translational and rotational components of the optic flow on the perception of planar surface orientation and motion.[4, 16–19]

3. EXPERIMENT

3.1 Motivations

Vision and body motion are tightly linked during our everyday interaction with the environment. As we move our body, the projections of objects we are looking at continuously change on our retinae. Despite these continuous changes we perceive a stable 3D world. How this perceived stability is achieved, has been a matter of debate and ground for theoretical works.[3, 20, 21] In laboratory conditions, because of technological drawbacks, researchers have often studied the static observer neglecting the contribution of extra-retinal information resulting from body movements. Our system instead enables us to bring into the lab real world conditions characterized by a high degree of sensorimotor coherence.

We studied how and to what extent the brain compensates for retinal motion signals induced by body movement in order to achieve a stable representation of the external world. We consider the case of a moving observer (head translations or rotations) looking at a single point of light glowing in the dark at eyes height. When the point is stationary in the world its projection on the retinae is moving at a velocity proportional to head translation velocity and in the opposite direction of the head movement. To veridically perceive the point as static the visual system has thus to fully compensate the retinal motion signal through extra-retinal signals resulting from head motion: afferent signals (or an efferent copy of motor commands) and visual motion signals should cancel each other out and have the same intensity. If extra-retinal signals only partially compensate for the retinal motion signals, the point will be perceived as moving in the opposite direction of head motion; vice versa, if extra-retinal signals overcompensate for the retinal motion signals, the point will be perceived as moving in the same direction of head motion.

The way in which the brain uses retinal and extraretinal signals has been a longstanding and unsolved problem in the field of vision sciences. To effectively interact in a complex environment is crucial to correctly judge the motion of surrounding objects. In fact, at every moment in time, the visual input might result from two major sources of motion:[22] a source resulting from self-motion and a source resulting from the movement of objects relative to the observer. Therefore, the brain faces a source-separation problem as it has to parse the visual input into components due to self and objects motion. In our experiments we consider two sources of extra-retinal signals resulting from different receptors in the peripheral vestibular systems: the otolith organs stimulated by head translation, and the semicircular canals stimulated by head rotation.[23]

Detection experiments have indeed established that humans are able to sense rotation on the horizontal plane, but not translations.[2, 24] Given these findings we hypothesized a different sensitivity for visual motion during angular and linear head movements on the horizontal plane.

Visual stability was tested in passive and active vision conditions. In the active vision condition we measured the amount by which the point has to shift (in an allocentric reference frame) to induce a sensation of perceptual stability when the observers were laterally translating or laterally rotating the head, by keeping the visual stimulation constant in the two conditions. In the passive vision condition, immobile participants were presented with a replay of the visual motion that they themselves generated during the active vision session. By keeping the visual stimulation the same and removing the extra-retinal information available to the active observers, we were thus able to separate the contribution of retinal and extra-retinal signals for the detection of visual motion.

3.2 Methods

At the beginning of each active vision trial, a red fixation mark, corresponding to the projection of the cyclopean eye on the screen, was shown on the monitor and the observers were required to move their head rightward. Viewing was monocular. Depending on the head motion condition the participants were either translating laterally or rotating the head. In the head translation condition, the observer was instructed to reverse the direction of head motion after hearing a beep signaling a head shift of 50 mm (to the right) relative to the center of the screen and after hearing a beep at −50 mm (to the left) signaling a shift in the opposite direction. After two cycles of head movement, when the observer's head passed through the extreme left position moving rightward, the fixation mark was replaced by the point-light stimulus. The stimulus remained visible until the head passed the extreme right position and then was replaced by a black screen. After stopping the motion of the head, participants were asked to classify, with a button press, the direction of the point shift as left or right. For the head rotation condition the participant was instructed to rotate the head by about $10°$ (from $−5°$ to $5°$). This value was chosen because, at a distance of 568.5 mm (the focal distance during the experiment), the projection of the cyclopean eye on the monitor covers exactly the same amount of space (100 mm) spanned in the translation condition. The ordering of the rotation and translation conditions was randomly intermixed among the 8 participants.

The amount of point shift was varied through a staircase procedure devised to find the Point of Subjective Stability (PSS), i.e., the value of gain corresponding to chance level performance and the Just Noticeable Difference (JND), i.e., the smallest gain variation from the PSS which gives rise to a non-stable percept. The gain was introduced between the horizontal motion of the head in space and the horizontal motion of the visual stimulus on the screen. The motion of the stimulus was linearly related to the observers' motion with the equation $s = gT_x$ where s represents the horizontal position of the stimulus center on the projection plane, the gain g is a multiplicative term and T_x represents the x coordinate of the participants' right eye (in the head translation condition), or the position of the intersection between the observers' optical axis and the projection screen (in the head rotation condition). Therefore, if $g = 1$, the point translated on the screen in the same direction as the head by the same amount of head motion (i.e., 100 mm). If $g = 0$, the point was stationary in the world (thus appearing always at the center of the screen). If $g = −1$, the point translated on the screen by the same amount of head motion but in the opposite direction (see boxes below the x-axis of Figure 3).

If the brain fully compensates for the head displacement, then the staircase procedure should converge to a value of gain that is close to zero: any deviation from this value represents a systematic error and is informative about the reliability of head translation and rotation signals. Positive deviation would indicate a partial compensation of retinal motion signals, while a negative deviation would indicate an overcompensation (with extraretinal signals weighted more heavily than retinal signals). Note that when replayed to the immobile observer different gain values generate, a stationary point $(g = 1)$, a point moving leftward of about 100 mm $(g = 0)$, and a point moving leftward of about 200 mm $(g = −1)$ (see boxes below the x-axis of Figure 3).
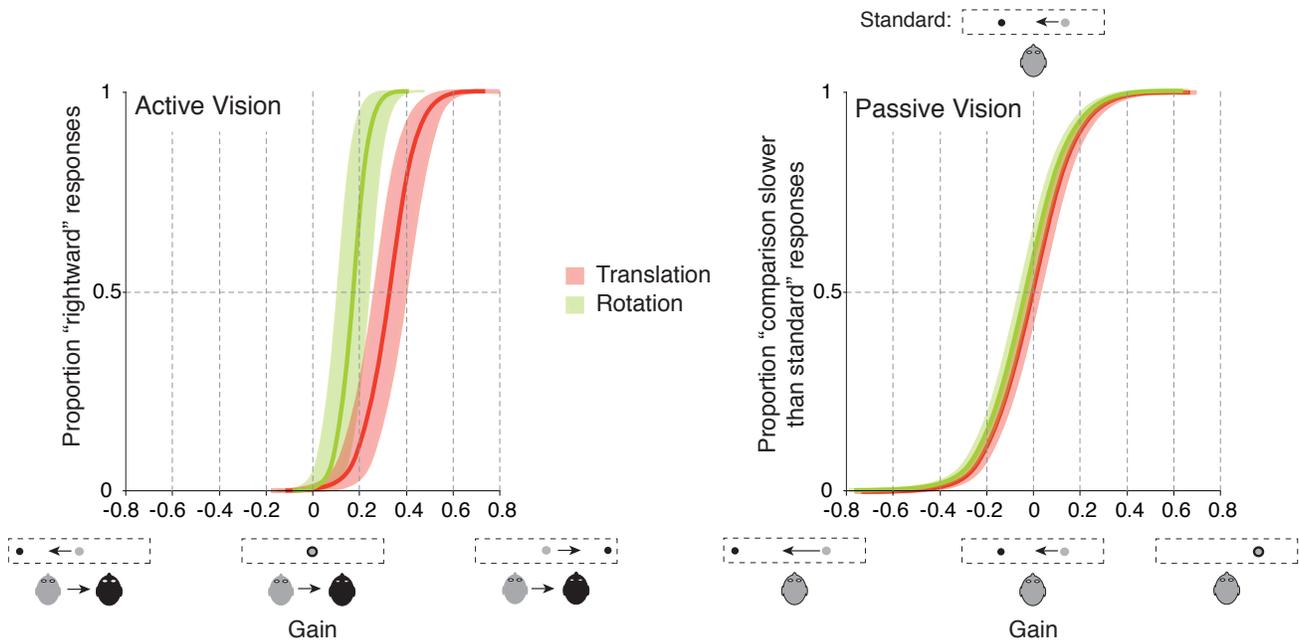
Figure 3. Average cumulative gaussian fits as a function of gain g, for the translation (red) and rotation (green) conditions, in the active (left) and passive (right) viewing conditions. Shaded bands indicate $\pm\,1$ s.e.m. Below each x-axis a top view of the stimulus presentation (referring to the head translation condition only), for $g=-1$, $g=0$ and $g=1$ (from left to right) is represented. Bottom left, the head and the point move in the direction of the black arrow: from the onset (grey head and point) to the end (black head and point) of stimulus presentation. Participants indicated whether the visual stimulus moved to the left or to the right. Bottom right, the head is immobile and the point moves in the direction of the black arrow: from the onset (grey point) to the end (black point) of stimulus presentation. Participants indicated in which of the two sequences (standard, shown on top of the panel, or comparison, shown below the x-axis) the visual stimulus moved faster.

Therefore, the experimental task was slightly different in the passive vision condition, in which we opted for a 2IFC task: a judgment about the speed difference between two successive sequences of motion was asked. Each trial consisted of two sequences: in one sequence the exact same trajectory covered by the participant head movement was replayed (standard), in the other sequence the trajectory was modulated by the corresponding gain value (comparison). The task of the participants was to indicate with a button press, in which of the two sequences the stimulus moved faster. The order of presentation of the standard and the comparison was random. Also, in order to avoid that the participants based their decisions on the total length of the trajectory covered on the screen, the starting position of the stimuli followed a random uniform distribution (bounded between 0 and 2.5 cm).

3.3 Results and Discussion

Classification performance was calculated by fitting a psychometric curve to individual proportions of "rightward" (in the active vision condition) or "the comparison is moving faster than the standard" responses (in the passive vision condition) as a function of gain. A gaussian model with parameters PSS and JND was estimated from the data using the constrained maximum likelihood and bootstrap inference method implemented by the psignifit software.[25, 26]

Figure 3 illustrates the fitted proportions of responses as a function of the rotation gain and the type of head movement conditions (translation in red, rotation in green), for active (left) and passive (right) observers. Analysis of average PSS demonstrates a partial compensation of retinal motion signals. Both the head translation and the head rotation conditions led to a biased estimate of visual motion in the direction of self-movement (positive gain values): in order to be perceived as stationary, the visual stimulus had to move in the same direction of motion of the observer (translation, $t(7) = 5.7324$, $p < 0.001$; rotation, $t(7) = 3.389$, $p < 0.05$). However, as shown in Figure 3, responses were more biased and less precise (i.e., larger JND) in the head translation condition (PSS=0.33 ± 0.07; JND=0.10 ± 0.019) than in the head rotation condition (PSS= 0.17 ± 0.07; JND= 0.06 ± 0.006), as revealed by the Welch t-test for unequal-variance samples on PSSs ($t(7) = 6.523$, $p < 0.0001$), and on JNDs ($t(7) = 2.6183$, $p < 0.05$). This difference may reflect a different sensitivity of the receptors in the peripheral vestibular system encoding head rotations and translations, leading to an almost negligible compensation of retinal motion signal during head translation, but not during head rotation. Nevertheless, there is also the possibility that the rotation and translation conditions differed in terms of the visual input. However, the results in the passive viewing condition exclude this latter possibility, since no significant difference was found between the translation and rotation conditions for neither PSSs (-0.033 vs. -0.006: $t(7) = 0.86$,$p = 0.41$) nor JNDs (0.16 vs. 0.15: $t(7) = 0.3841$,$p = 0.999$).

In summary, our results demonstrate that our perception of stability is biased and that this bias might result from a compensation of retinal motion signals that is only partial. Linear ego-motion signals were found to be poorly integrated with visual inputs, whereas angular motion signals were found to be almost correctly estimated and employed for a near full compensation of retinal motion signals. These findings are consistent and extend previous results on the perception of motion of full-field stimuli.[2]

REFERENCES

[1] Panerai, F., Ehrette, M., and Leboucher, P., "A VE framework to study visual perception and action," *Virtual Real.* **6**, 21–32 (2002).

[2] Jaekl, P., Jenkin, M., and Harris, L., "Perceiving a stable world during active rotational and translational head movements," *Exp. Brain Res.* **163**, 388–399 (2004).

[3] Wallach, H., "Perceiving a stable environment when one moves," *Annu. Rev. Psychol.* **38**, 1–27 (1987).

[4] Fantoni, C., Caudek, C., and Domini, F., "Systematic distortions of perceived planar surface motion in active vision," *J. Vis.* **10**, 1–20 (2010).

[5] Wexler, M., "Voluntary head movement and allocentric perception of space," *Psychol. Sci.* **14**, 340–346 (2003).

[6] Gerbino, W., Fantoni, C., Nicolini, C., Volcic, R., and Domini, F., "Active multisensory perception tool: Bus experience and action comfort," in [*Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference*], De Waard, D., Brookhuis, K., Wiczorek, R., Di Nocera, F., Barham, P., Weikert, C., Kluge, A., Gerbino, W., and Toffetti, A., eds. (2013).

[7] Volcic, R., Fantoni, C., Caudek, C., and Domini, F., "Visuomotor adaptation changes the processing of binocular disparities and tactile discrimination," *J. Neurosci.* **33**, 17081–17088 (2013).

[8] Kane, D., Held, R., and Banks, M., "Visual discomfort with stereo 3d displays when the head is not upright," *Proc. SPIE* **8288**, 828814 (2012).

[9] Hoffman, D., Girshick, A., Akeley, K., and Banks, M., "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.* **8**, 33 (2008).

[10] Shibata, T., Kim, J., Hoffman, D., and Banks, M., "The zone of comfort: Predicting visual discomfort with stereo displays.," *J. Vis.* **11**, 1–29 (2011).

[11] Swindells, C., Dill, J., and Booth, K., "System lag tests for augmented and virtual environments," in [*Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*], 161–170, ACM (2000).

[12] Steed, A., "A simple method for estimating the latency of interactive, real-time graphics simulations," in [*Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*], *VRST '08*, 123–129, ACM, New York, NY, USA (2008).

[13] Kooima, R., "Generalized perspective projection," `http://csc.lsu.edu/~kooima/articles/genperspective/` (2008).

[14] Umeyama, S., "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 376–380 (1991).

[15] Guennebaud, G., Jacob, B., et al., "Eigen v3." http://eigen.tuxfamily.org (2010).

[16] Fantoni, C., Caudek, C., Mancuso, G., and Domini, F., "Linear egomotion signals are mostly ignored in the interpretation of the self-generated optic flow," *J. Vis.* **12**, 1046 (2012).

[17] Mancuso, G., Fantoni, C., Caudek, C., and Domini, F., "Non-informative components of retinal and extra-retinal signals affect perceived surface orientation from optic flow," *J. Vis.* **12**, 241 (2012).

[18] Fantoni, C., Caudek, C., and Domini, F., "Perceived surface slant is systematically biased in the actively-generated optic flow," *PloS ONE* **7**, e33911 (2012).

[19] Caudek, C., Fantoni, C., and Domini, F., "Bayesian modeling of perceived surface slant from actively-generated and passively-observed optic flow," *PloS ONE* **6**, e18731 (2011).

[20] Wexler, M., Panerai, F., Lamouret, I., and Droulez, J., "Self-motion and the perception of stationary objects," *Nature* **409**, 85–88 (2001).

[21] Wexler, M. and Boxtel, J. J. V., "Depth perception by the active observer," *Trends Cogn. Sci.* **9**, 431–438 (2005).

[22] DeAngelis, G. and Angelaki, D., "Visual-vestibular integration for self-motion perception," in [*The Neural Bases of Multisensory Processes*], Boca Raton (FL): CRC Press (2012).

[23] Kandel, E., Schwartz, J., and Jessell, T. in [*Principles of Neural Science*], London: Prentice-Hall (2000).

[24] MacNeilage, P., Banks, M., DeAngelis, G., and Angelaki, D., "Vestibular heading discrimination and sensitivity to linear acceleration in head and world coordinates," *J. Neurosci.* **30**, 9084–9094 (2010).

[25] Wichmann, F. and Hill, N., "The psychometric function: I. fitting, sampling and goodness-of-fit," *Percept. Psychophys.* **63**, 1293–1313 (2001).

[26] Wichmann, F. and Hill, N., "The psychometric function: II. bootstrap-based confidence intervals and sampling," *Percept. Psychophys.* **63**, 1314–1329 (2001).