Perceiving a stable world during self-motion: measurement and modeling of visual-vestibular integration and segregation

Isabelle Teresa Garzorz

München 2018

Perceiving a stable world during self-motion: measurement and modeling of visual-vestibular integration and segregation

Isabelle Teresa Garzorz
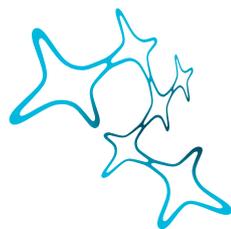
Graduate School of
Systemic Neurosciences

LMU Munich

Dissertation an der
Graduate School of Systemic Neurosciences
der Ludwig-Maximilians-Universität München

1$^{\text{st}}$ of August 2018

Supervisor & First Reviewer: Prof. Dr. Paul MacNeilage
Second Reviewer: Prof. Dr. Heiner Deubel
Third Reviewer: Prof. Dr. Wolfgang Becker

Date of Submission: 1st of August 2018
Date of Defense: 10th of December 2018

Abstract

Human spatial orientation - during both active self-motion, such as walking and running but also passive self-motion, such as sitting on a bus, train, or plane – comes very naturally to us and usually does not involve conscious control. This might easily lead us to underestimate its importance and complexity as a finely tuned multisensory interplay. The major goal of this interaction is to provide us with the percept of a stable world helping us preserve our equilibrium.

A basic mechanism allowing us to experience spatial constancy while walking, moving our head and eyes is multisensory cue integration, which can yield a more precise estimate of self-motion. However, signals should only be integrated when they provide information about the same event or object. The process of conflict detection between multisensory cues has been far less investigated and is the focus of this thesis.

In a virtual reality set-up comprising a 6-degree-of-freedom hexapod motion platform and a stereo screen, two psychophysical experiments have been conducted. The first experiment evaluated crossmodal discrimination as a model of conflict detection. Participants were seated on a motion platform and performed 2-alternative forced-choice tasks with either visual, vestibular or both visual and vestibular stimulation, the latter either being in accordance to evaluate multisensory integration or containing small conflicts to evaluate conflict detection performance.

Multisensory interactions such as integration or conflict detection require the brain to find a common reference frame for the contributing signals. These transformations can be incomplete, yielding sensory illusions such as the Aubert-Fleischl illusion in which objects appear to move more slowly when pursued with the eyes. The second experiment investigates a vestibular analogue of this phenomenon, i.e. objects that are physically pursued through passive whole-body rotations seem slower than non-pursued objects.

A third experiment investigates whether the findings about conflict detection from the first experiment are generalizable to more naturalistic self-motion contexts. It examines how eye movements influence conflict detection performance during active head turns, using a head-mounted display.

Finally, a link between the current findings on conflict detection and possible implications for applied topics, such as virtual reality devices and motion sickness research, will be provided. In addition, the Aubert-Fleischl illusion will be interpreted with respect to models on object motion perception during self-motion.

Overview

The first chapter of this thesis introduces the different sensory modalities supporting self-motion perception, with the most important being the visual and vestibular senses. It proceeds explaining multisensory cue interactions with a special focus on visual-vestibular integration, its Bayesian framework and the mechanisms underlying visual-vestibular conflict detection. Another section focuses on reference frame transformations, which are the basis of multisensory interactions between head, eye and visual motion. While those reference frame transformations (e.g. from retinocentric to head-centric) are necessary for comparing and integrating multisensory signals, they can sometimes be incomplete, yielding sensory biases, such as the Aubert-Fleischl illusion in which pursued targets seem to move more slowly compared to non-pursued ones. The first chapter concludes with the basics of psychophysical experiments to provide necessary background information on the experimental design of the reported studies.

The second chapter reports the first study which evaluated crossmodal discrimination as a model for conflict detection and pointed out the important role of eye movements for visual-vestibular interactions. This study has been published in the journal *Current Biology*.

The third chapter contains a manuscript that has meanwhile been published in the *Journal of Vision*. In this study, we present our findings about a vestibular analogue of the Aubert-Fleischl phenomenon and discuss possible explanations for the insufficient compensation for self-motion during perception of object speed (mismatch between retinal and vestibular estimates).

The fourth chapter reports a follow-up study that is currently in press in the journal *Displays*. It investigates whether the findings of our first study using highly controlled passive rotations are generalizable to active head turns.

The fifth chapter summarizes the findings and discusses them in a broader context. It also provides an outlook on future research questions and experiments to be conducted.

Contents

List of Figures

Chapter 1

General Introduction

When moving through the world, a continuous stream of sensory signals enables our brain to form a coherent percept of the environment and to estimate self-motion – a prerequisite for accomplishing challenging tasks such as navigation, obstacle avoidance and the control of gait and posture. There are several sources of motion that must be disentangled for estimating self-motion: objects moving relative to the observer, locomotion of the observer relative to the environment and both locomotion and object motion with respect to the environment. Although our body, head and eyes move relative to the environment at any time, we perceive the world as stable. The brain achieves this spatial constancy by integrating and comparing sensory information from visual, vestibular and somatosensory origin and motor information about the position and movement of the eyes and the head. Nevertheless, sensory information is noisy and sometimes ambiguous which can lead to sensory illusions and conflicts.

   This thesis aims at elucidating the processes underlying spatial constancy which involve both integration of multisensory information and detection of conflicts between sensory inputs. It also shows how physically congruent multisensory input can lead to illusory precepts of mismatching signals and provides possible explanations for these phenomena.

1.1   The visual system – Retinal cues

The presumably most informative and rich sense for self-motion is vision. When moving through the world, a stream of constantly changing reflections of the outside world is registered by the retina, the inner layer of our eye that detects changes in light intensity and sends this information to the visual cortex. When moving forward, the visual scene seems to pass by on both

Figure 1.1: Optic flow field for an observer moving forward.

sides of our eyes. The pattern of motion on the retina that results from our own movement relative to the environment is called optic flow (Gibson, 1950). The point in distance which we are heading to is the focus of expansion. It is the location from which the optic flow field emerges (see Fig. 1.1). The scene continuously expands from this point – unless we stop moving – until it passes the boundaries of our visual field and finally disappears. If we know the scale of the scene, optic flow enables us to know how fast we are moving and how far away objects are located from us (Gibson, 1954). Given the same setting, the faster we move, the higher the speed of optic flow. Objects in the far distance seem to move more slowly compared to objects that are closer to us, a phenomenon called motion parallax. Optic flow also reveals the direction of self-motion. While moving forward yields an expanding flow field, moving backward results in an optic flow that radially flows from the outside towards the point of contraction. Rotating about one's own axis causes an optic flow in the opposite direction of self-motion.

All previous examples involve real self-motion in space. However, optic flow can induce the percept of apparent self-motion without actual displacement. This can happen for example when we are sitting on the train while the train next to us slowly starts accelerating. The resulting retinal motion may induce the illusionary experience that our train started moving in opposite direction and immediately gets resolved when we start looking at a stationary reference point such as the roof of the train station. This visually induced apparent self-motion is also called vection and can appear during

any kind of large-field optic flow stimulation (Berthoz et al., 1975; Brandt et al., 1973). An experimental device inducing vection is the so-called optokinetic drum, a pattern of black and white vertical stripes which is painted on the walls of a rotating drum. An observer sitting on a stationary chair in the center of the drum will initially perceive the drum as rotating, but after a few seconds the illusory feeling of self-motion, here circular vection, will evolve and the pattern will be perceived as stationary.

Visually determined heading thresholds - a measure of the precision of heading judgments - have been shown to be 60% lower than those during walking in darkness, indicating the importance of our visual sense for orientation (Telford et al., 1995). However, retinal cues alone are ambiguous as expanding optic flow could be the result of our own forward self-motion or the effect of objects passing by or expanding. Resolving this ambiguity requires us to take into account both extraretinal cues to eye movement, as well as extraretinal cues to head movement. This information allows us to perceive the world as stable despite our eye movements and self-motion.

1.2 Extraretinal signals

1.2.1 Oculomotor signals

While retinal signals provide large-scale information about the environment and our self-motion via optic flow, eye movements enable us to shift our focus of attention in accordance with the task requirements of our actions (Hayhoe and Rothkopf, 2011). They help us direct the fovea, i.e. the point of highest visual acuity of our retina, towards the point of interest in the environment to see it clearly, with the highest resolution possible. Such voluntarily initiated gaze-shifting eye movements can either be very short and quick jumps, so-called saccades, or slow smooth pursuits. While saccades are intended to rapidly shift the fovea and the focus of attention towards a new target of interest, smooth pursuit eye movements follow a moving target and, interestingly, cannot be initiated and conducted without any target.

The question is how we are able to perceive the world as stable despite moving our eyes (von Helmholtz, 1867). Two options should be considered. Reafferent signals about the stretch of the extraocular muscles could help predict the change of the retinal image due to eye movements (Sherrington, 1918). However, when we passively move our eye by tapping on the eyeball, the retinal image jumps and the world no longer seems stable. A more plausible option is that the brain knows about the motor commands that have been sent to the eye muscles and predicts the retinal image based on them.
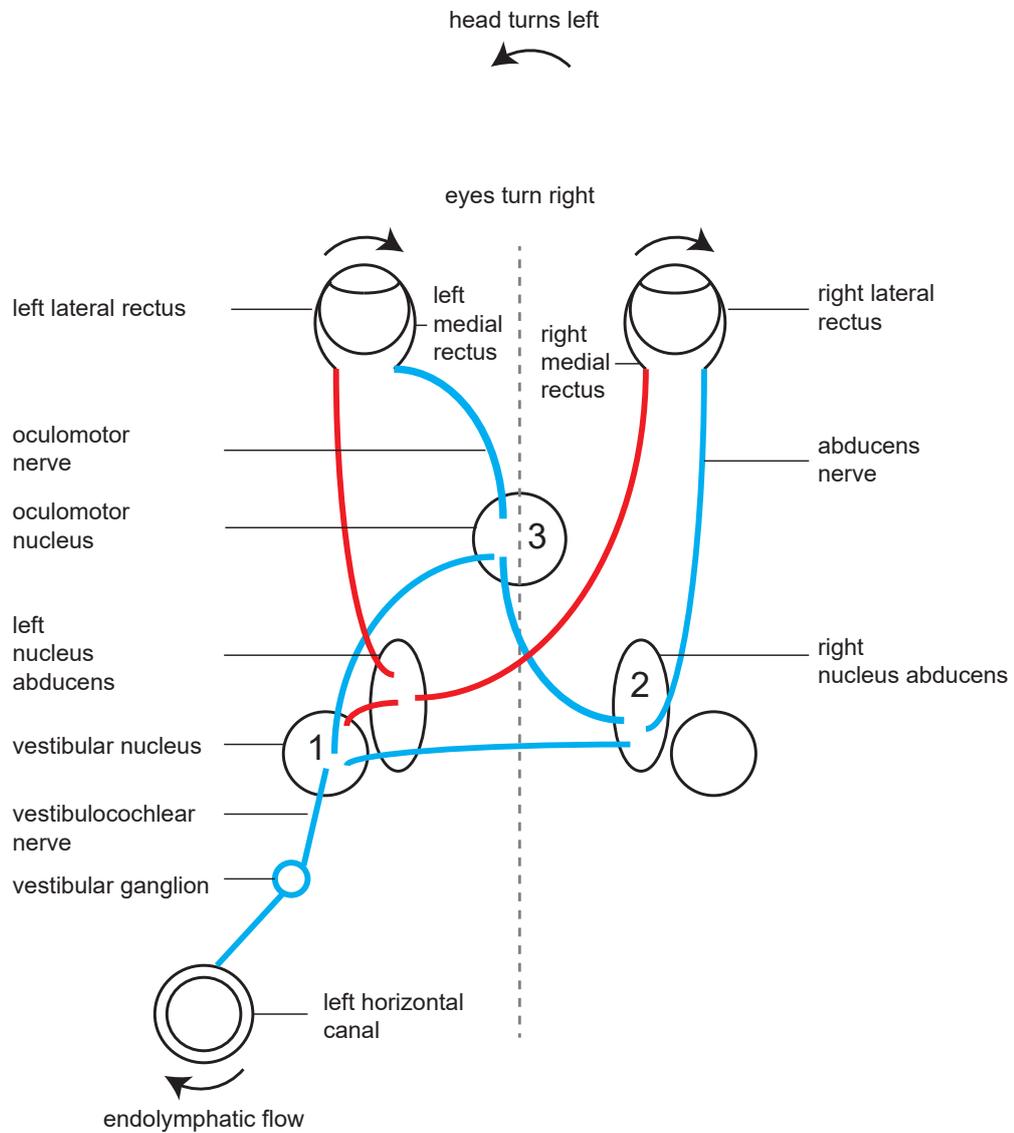
Figure 1.2: 3-neuron arc of the horizontal vestibulo-ocular reflex. Head turns to the left, inducing vestibular canal signals. Blue lines indicate the resulting excitatory pathway for compensatory eye movements to the right, which is supported by the inhibitory pathway (in red).

Corollary discharge or so-called efference copy signals of the motor commands (Sperry, 1950; von Holst and Mittelstaedt, 1950) have been proposed as the underlying signaling mechanism that "cancels" the retinal motion signals induced by eye movement. It implies that an apparent displacement of the external world can be avoided by subtracting the image motion generated by eye movements. A neural basis of efference copy signals has been found in the medial dorsal nucleus of the thalamus of monkeys where neurons convey such corollary discharge signals to the frontal eye field (Sommer and Wurtz, 2002, 2008; Wurtz and Sommer, 2004). Generally, efference copy signals help us distinguish sensory signals that are the result of our own movements from those that are externally elicited.

So far, we have considered saccades and smooth pursuits. In everyday life, we do not only saccade to targets of interest, but we also fixate them to stabilize the image on the fovea while moving. One such gaze-stabilizing eye movement that is very important for self-motion perception is the vestibulo-ocular reflex (VOR). The VOR compensates for head movements by counter-rotating the eyes opposite to the head movement. In contrast to visually driven compensatory eye movements (such as refixation saccades) which have a latency of about 200 ms (Carpenter, 1988), the VOR is much faster with a latency of only 9 ms (Collewijn and Smeets, 2000). Thus, whereas pursuing a moving target at high velocity can be very challenging, you can easily keep a spot on the wall fixated while quickly shaking your head. The VOR fundamentally relies on the vestibular system in our inner ear which serves as a trigger for a 3-neuron reflex (see Fig. 1.2). Whenever we move our head to the left, vestibular signals of the left canals are transferred via the left vestibulocochlear nerve through the left vestibular ganglion to the left vestibular nuclei of the brainstem. One part of the fibers projects ipsilaterally to the left oculomotor nucleus which innervates the left medial rectus muscle. The other part of the fibers in the vestibular nuclei crosses to the contralateral (right) nucleus abducens where they excite, firstly, motor neurons which project to the right abducens nerve and excite the right lateral rectus muscle and, secondly, interneurons that cross the midline to the left oculomotor nucleus where neurons that project to the left medial rectus muscle through the left oculomotor nerve get activated. This excitatory pathway elicits compensatory eye movements to the right, which is additionally supported by another inhibitory pathway. The signals of the left canals also excite neurons in the vestibular nuclei that inhibit neurons of the left abducens which project to the left lateral and the right medial rectus muscles. The reduced activity of these neurons prevents eye movements to the left. Both direct pathways (excitatory and inhibitory) complement each other in realizing the VOR but need an additional indirect pathway to hold the eyes at an eccen-

tric position once the head stops moving. This functionally essential part of the VOR is located in the brainstem and called the neural integrator which keeps track of head position by temporally integrating the velocity signals of the vestibular ganglion. Resulting position signals are then conveyed to the ocular motoneurons to compensate not only for the velocity but also the displacement of the head. One important point to be mentioned is that the VOR can also be actively suppressed, for example while tracking a moving target with combined head and eye movements.

1.2.2   The vestibular system

The vestibular system – the phylogenetically oldest part of the inner ear – provides us with a sense of balance and orientation (Carriot et al., 2014). The important contribution of vestibular signals to self-motion perception has long been investigated (Harris et al., 2000a,b; Young, 1970). Although sometimes underestimated as being widely inferior to vision (Telford et al., 1995), vestibular signals have been recognized to broadly complement visual cues during self-motion (Berthoz et al., 1975). The relevance of the vestibular sense to self-motion becomes easily apparent when we close our eyes while for example sitting in a car. We are still able to notice the car start moving, mainly due to our vestibular organs, our inertial measurement units inside the labyrinth of our inner ear, which detect head acceleration in space (DeAngelis and Angelaki, 2012). In this way, the vestibular system enables us to maintain balance and postural stability even in darkness during both self-induced and passively applied movements (Cullen, 2012). Together with other sensorimotor signals, vestibular cues are indispensable for perceiving a stable world during self-motion and action by controlling and stabilizing gaze (Medendorp and Selen, 2017). They help us navigating through and interacting with a highly complex environment.

The vestibular system comprises the two otolith organs (the vertical saccule and the horizontal utricle) that sense linear accelerations due to gravity and translation and the three semicircular canals (anterior, lateral and posterior) that register angular accelerations (Fig. 1.3).

The functional principle of these organs is the transduction of mechanical motion into electrochemical signals to be further processed and interpreted by the nervous system. To this end, the semicircular canals have a thickened region at the base of each canal, the ampulla, which contains a cone-shaped zone of epithelium, the ampullary crista. This epithelium contains receptor cells from which hair bundles, so-called cilia, protrude into the cupula, a gelatinous membrane, spanning the entire cross-section of the canal. When we move our head, the endolymph that fills the labyrinth lags behind the
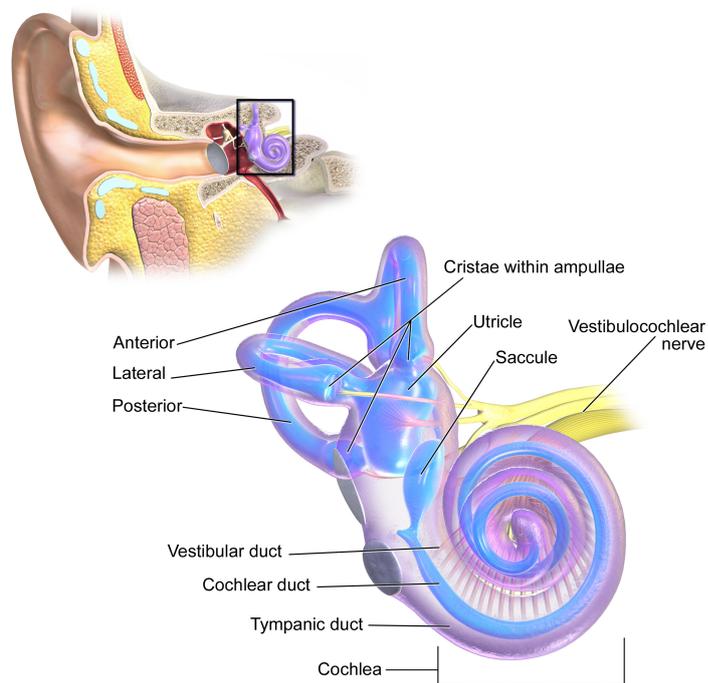
Figure 1.3: Vestibular organs of the inner ear. Image used under the license CC-BY-3.0 from Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). doi: 10.15347/wjm/2014.010. ISSN 2002-4436

membranous structure of the labyrinth due to inertial forces. Therefore, the motion of the fluid exerts pressure on the cupula and bends it. The resulting deflection of the cilia inside the cupula changes the membrane potential of the hair cells, eliciting action potentials and signal transfer via the vestibulocochlear nerve to the vestibular nuclei in the brain stem.

The hair cells inside the sensory epithelium of the otolithic organs, the macula, follow a similar mechanism. Their hair bundles extend into the gelatinous layer of the otolithic membrane that is covered by crystals of calcium carbonate, so-called otoconia (Greek for "ear dust"). When the head translates in the plane of the respective otolith organ, inertial forces cause the heavier otolithic membrane to lag behind the motion of the macula and thus produce a shear force upon the cilia of the hair cells that protrude into the gelatinous layer. Depending on whether the stereocilia of a hair bundle are bent towards or away from the longest and most important cilium, the kinocilium, the firing rate of the cell increases or decreases, thereby encoding

acceleration and deceleration.

The three semicircular canals, i.e. the anterior, posterior and horizontal canals, are mutually orthogonal so that all possible head rotations can be physically represented (Zupan et al., 2002). Besides, the canals of the left and right inner ear are bilaterally symmetrical and complementarily inhibit and excite their coplanar counterparts (push/pull principle), which means that stimulation of one canal elicits inhibition of the respective canal on the other side. In contrast, the otoliths work in two dimensions, i.e. the saccules fully encode the sagittal plane and the utricles the horizontal plane.

Taken together, the semicircular canals and otolithic organs cover the entire range of physical states, e.g. linear motion in all directions (forward/backward (surge), up/down (heave), left/right (sway)), knowing where up and down is (the gravitational vertical) and rotations in all degrees of freedom (yaw (vertical axis), pitch (horizontal axis), and roll (longitudinal axis)). Their signals project via the vestibular division of the vestibulocochlear nerve (VIII cranial nerve) to the vestibular nuclei from which efferent information is guided to the brain structures for the control of eye movements, posture and balance (Cullen, 2012). In contrast to other sensory inputs, the vestibular system acts as a "silent sense" that constantly sends a flow of information to the brain which cannot be "turned off" and which we are not consciously aware of (Day and Fitzpatrick, 2005). However, one can introspectively access vestibular sensations (Benson, 1990) or become aware of the vestibular sense during sensory conflicts (Benson et al., 1986) or whenever it is not working properly due to diseases (Brandt et al., 2005). In contrast to other senses, the vestibular sense becomes multisensory early in the processing stream (e.g. VOR in the brain stem) and thus helps us not only maintain stability but also navigate, voluntarily move and control our eye movements to stabilize gaze (Angelaki and Cullen, 2008).

Despite the complete representation of all head movements, the vestibular system faces two problems. First, while the cupulae and otolithic membrane encode brief accelerations, the vestibular system delivers a velocity signal with high-pass characteristics, meaning the response adapts during constant-velocity motion. Second, the vestibular system is subject to Einstein's equivalence principle, meaning the otoliths cannot physically distinguish linear acceleration due to translational movements from gravitational acceleration due to head tilts. Any accelerometer measures net gravito-intertial force which is the vector sum of gravitational and inertial forces (Angelaki and Cullen, 2008).

This ambiguity may cause severe misperceptions, such as the somatogravic illusion which challenges pilots during flight conditions of poor visibility (Gillingham and Previc, 1993). Whenever the aircraft strongly accelerates,

the pilot may misinterpret the vestibular cue as a tilt back, indicating that the aircraft is pitching up which he incorrectly compensates for by lowering the nose, thereby ultimately steeply diving the plane and crashing into the ground. However, during smaller accelerations, signals from the semi-circular canals can help disambiguating head tilt from translational movements. Internal models which resolve the ambiguity by integrating linear acceleration and angular velocity signals inside the cerebellum have been proposed (Green et al., 2005; Merfeld et al., 2005, 1999). Moreover, visual signals help disambiguating vestibular signals (MacNeilage et al., 2007).

1.2.3   The somatosensory system

The somatosensory system, which is especially important for our perception of active self-motion, comprises receptors for the detection of both external and internal signals.

Cutaneous receptors in the dermis of our skin provide us with information from the environment, i.e. forces acting on our body surfaces such as touch, pressure and vibration. Proprioceptive receptors such as muscle spindles, joint afferents and Golgi tendon organs, in contrast, respond to changes within the body. They transduce information about the length and tension of muscles, the sensation from ligaments, fascia, tendons and joints to the central nervous system, thereby allowing us to sense the position of our body segments with respect to each other (Cullen, 2011). Like the mechanoreceptors of the vestibular system, somatosensory receptors transduce mechanical forces into electrical signals. These are passed through a 3-neuron network from the periphery via the spinal cord, brainstem and thalamic relay nuclei to the sensory cortex in the parietal lobe (Kandel et al., 2000, p. 338-345).

Somatosensory information complements vision and together with vestibular signals helps us maintaining balance, even in darkness (Kars et al., 2009). Cutaneous mechanoreceptors in the feet, for example, are important for the control of upright stance (Kavounoudias et al., 1998). In addition, neck proprioceptors indicate head position with respect to the trunk and help the vestibular system distinguish pure head rotations in space from whole-body rotations (Mergner et al., 1991). Interestingly, some studies investigating vestibular-somatosensory interactions suggest that proprioception dominates over vestibular signals (Hlavacka et al., 1992) and that vibrotactile stimulation can impair the detection of near-threshold vestibular stimuli (Cabolis et al., 2018). Nevertheless, the vestibular organs are one of the main suppliers of sensory information to self-motion perception.

1.2.4   Further cues to self-motion perception

Apart from the sensory systems mentioned above, auditory signals contribute to self-motion perception. We are usually moving through a world full of sounds and noises which provide us with a constant auditory flow. The cochlear mechanoreceptors respond to sound waves, i.e. changes in air pressure, and send this auditory information to the brain. The most well-known example of an auditory contribution to self-motion perception is the phenomenon of auditory circular vection, which is induced by rotating auditory stimuli (Dodge, 1923; Lackner, 1977; Sakamoto et al., 2008). It has also been shown that the frequency of tones can help estimating the speed of self-motion in a distance reproduction task (von Hopffgarten and Bremmer, 2011).

Furthermore, it has been suggested that the kidney complements the graviceptive signals from our vestibular organs by providing additional afferent information to the brain. The mechanism behind it is not entirely understood but one option could be that mechanoreceptors inside the kidney register pressure changes that occur during self-motion due to density differences between the fluid inside the kidney and the surrounding adipose capsule (Mittelstaedt, 1992, 1997). Similarly, mechanoreceptors in the fundus of the stomach could potentially register the gravitational load of a full stomach on the surrounding gastric arteries (Trousselard et al., 2004).

1.3   Multisensory interactions during self-motion

The perception of a stable world during self-motion fundamentally relies on the interactions between the sensory systems described above. However, retinal signals and extra-retinal signals from the vestibular organs are the primary and most informative source of information to distinguish our own movements from movements of the environment. Thus, the thesis focuses on interactions between visual and vestibular cues that are additionally mediated by oculomotor signals.

1.3.1   Multimodal cue integration

As every sensory system has its own strengths and weaknesses, the brain combines information from various sensory modalities to form a coherent estimate of the property in question which is more reliable, accurate, and robust than the unisensory estimates. In case of self-motion perception, it

is beneficial to combine information from the visual and vestibular senses to overcome their individual weaknesses such as the inability of the vestibular system to encode constant velocity signals or the ambiguity of retinal signals causing illusory perception such as vection.

1.3.2 Maximum Likelihood Estimation

When we move through the world, both our visual and vestibular systems provide us with redundant information, for example about our heading direction. However, each cue is corrupted by noise from external sources, such as unclear vision during foggy weather, and internal sources, such as neuronal signaling noise. Noise compromises the precision of the sensory estimates which implies that they do not always represent the true value of the measured quality but a slightly different one. The noisier the estimate, the less certain we can be about a specific measurement. Thus, it is convenient to mathematically describe sensory estimates using a likelihood function. Assuming Gaussian sensory noise, the likelihood function is normally distributed with the maximum value centered on the most likely value for the property under consideration such as heading direction (H). The noisier the sensory estimate, the broader the distribution, i.e. the larger the variance and the less reliable it is. If the noise sources of the sensory estimates are conditionally independent, combining them according to Maximum Likelihood (ML) estimation yields the statistically optimal and most precise estimate. It implies a weighted average of the single cue estimates (Eq. 1.1) where the weight given to each cue is proportional to its reliability, i.e. the inverse of its variance (Eq. 1.2 and 1.3).

$$H_{comb} = H_{vis} \cdot w_{vis} + H_{vest} \cdot w_{vest} \tag{1.1}$$

with

$$w_{vis} = \frac{1/\sigma_{vis}^2}{1/\sigma_{vis}^2 + 1/\sigma_{vest}^2} \tag{1.2}$$

$$w_{vest} = \frac{1/\sigma_{vest}^2}{1/\sigma_{vest}^2 + 1/\sigma_{vest}^2} \tag{1.3}$$

and

$$w_{vis} + w_{vest} = 1 \tag{1.4}$$

Mathematically, such ML integration equals the normalized product of the two likelihood functions (Fig. 1.4). The resulting combined likelihood function is again normally distributed, has a reduced variance (Eq. 1.5)
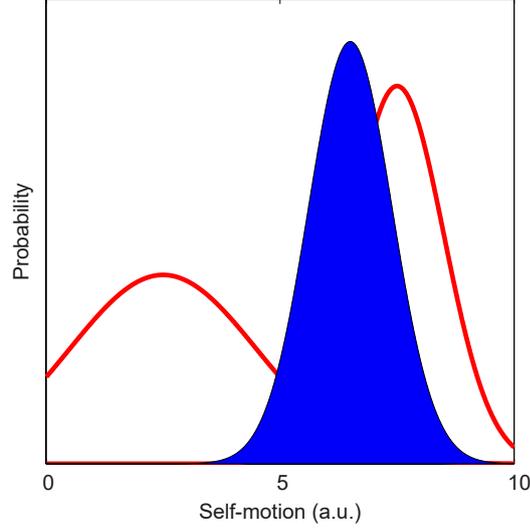
Figure 1.4: Multisensory Integration. The red distributions represent unisensory estimates, the blue distribution inidcates the combined estimate following ML integration.

compared to the single cue estimates and its mean is shifted relatively towards the more reliable sensory stimulus (Ernst and Banks, 2002).

$$\sigma^2_{comb} = \frac{\sigma^2_{vis} \cdot \sigma^2_{vest}}{\sigma^2_{vis} + \sigma^2_{vest}} \tag{1.5}$$

In fact, experiments have shown that weighted averaging of sensory information appears to be the mechanism by which the perceptual system decreases the effects of noise in sensory signals (Ernst and Banks, 2002; Hillis et al., 2002; van Beers et al., 1999). During self-motion, visual and vestibular signals get integrated following ML integration which has been shown in experiments using passive full body rotations or translations and optic flow (Butler et al., 2010; Fetsch et al., 2009; Gu et al., 2008; Jurgens and Becker, 2006; Prsa et al., 2012; ter Horst et al., 2015). Thus, our judgments of sensory estimates are not only based on the more reliable estimate while discarding the other one but both cues are considered, thereby reducing the overall variance. But how could the brain represent the reliability of sensory estimates to make such calculations possible? It could determine the variances on-line by using population coding. Individual neurons could selectively respond to one specific magnitude of the property in question, meaning they are maximally activated by this magnitude and respond less strongly to a different one. When presented with a specific magnitude of that quality, the neural population would then form a distribution with a clear peak and

some variance (Ernst and Bulthoff, 2004; Knill and Pouget, 2004; Ma et al., 2006).

1.3.3 Bayesian priors

In everyday life, we do not only rely on our senses but also take prior beliefs about the world into account to optimally interact with the environment. Combining perceptual prior beliefs with sensory information can be mathematically modeled in a Bayesian framework. The prior distribution represents the probability of an expected sensory estimate according to previous experience. Multiplying the likelihood function of the sensory estimate $P(B|A)$ with the prior distribution $P(A)$, using Bayes' rule, yields a posterior distribution $P(A|B)$ which represents the combined estimate:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \qquad (1.6)$$

where A is the property to be estimated, B is the sensory input and $P(B)$ is a normalization factor.

The stronger the prior, i.e. the smaller its variance, and the less reliable sensory information, the more the final estimate is influenced by the perceptual prior. Evidence supporting the assumption of statistically optimal integration using prior knowledge comes for example from a study showing that an *a priori* preference for slower velocities has a stronger influence on the estimate of perceived velocity when the visual target motion gets noisier (Weiss et al., 2002). ML integration and Bayesian inference are very closely related as ML integration represents a subset of the more general Bayesian approach in modeling human perception. ML integration coincides with Bayesian inference if we assume a uniform prior distribution, i.e. a prior that has no influence on the maximum a posteriori estimation (MAP) and can thus be neglected.

1.3.4 Causal inference

Multisensory integration is beneficial as it decreases the variance of the combined estimate, resulting in a more precise percept. Since our environment provides us with a constant stream of various sensations from different objects, our brain must decide which signals come from the same event or object and should thus be integrated, and which inputs have different sources and should be segregated. How the brain solves this correspondence problem is still widely unknown, but one solution could be that the brain uses statistics

about the co-occurrence of multisensory signals, i.e. prior knowledge about
the probability of signals occurring at the same time, which influence if sig-
nals get fused or not. Such co-occurrence statistics have been suggested by
Hillis et al. (2002) who have shown that disparity and texture gradients in
vision provide us with a very strong prior about their co-occurrence. This
means that these cues are highly spatially coupled in everyday perception as
we hardly use disparity cues without recognizing the texture of the object at
the same time. On the other hand, when asked to judge an object's property
by using multimodal sensory cues (e.g. visual and tactile), our underlying as-
sumptions about their co-occurrence might be less strong. Tactile and visual
senses are generally less coupled as we can for example look at a different ob-
ject than the one we are holding in our hands. These co-occurrence statistics
can be represented as Bayesian priors that influence whether sensory cues
are mandatorily or only partially fused. Ernst (2006) suggested a mathe-
matical model including a two-dimensional coupling prior which represents
the knowledge about the strength of the co-occurrence between the sensory
inputs. If the coupling prior is weak (i.e. it has a large variance), signals only
get partially fused whereas forced fusion occurs whenever the coupling prior
is strong (i.e. has a small variance). This model can also explain cases in
which we do not integrate signals despite a strong coupling prior, for example
when sensory signals are not redundant but highly discrepant.

An alternative model for solving the correspondence problem is called
Hierarchical Causal Inference. This model explicitly states the possible
causal structures underlying a multimodal percept (Fig. 1.5). Imagine an
experiment in which participants must judge the speed of a visual stimulus
while they experience passive whole-body rotations on a motion platform.
The visual and vestibular stimulation could be congruent, or in conflict with
each other. In other words, the multisensory stimulation could have a "com-
mon cause" or different causes. The probability of the visual and vestibular
signals having one cause (Fig. 1.5 A) can be calculated according to Bayes'
rule, taking into account the prior assumptions about the probability of a
common cause (depending on how strongly visual and vestibular signals are
correlated in general) multiplied by the joint likelihood function of the visual
and vestibular single cue estimates. If the visual and vestibular signals are
assumed to have different causes (Fig. 1.5 B), the optimal estimate of visual
speed is entirely determined by the visual stimulation, i.e. the mean of the
visual single cue distribution and the prior of independent sources. The ob-
server can now infer which scenario (i.e. common or independent source) is
more probable by forming a weighted average of the two optimal estimates
(Fig. 1.5 C, Shams and Beierholm, 2010; Shams, 2012).
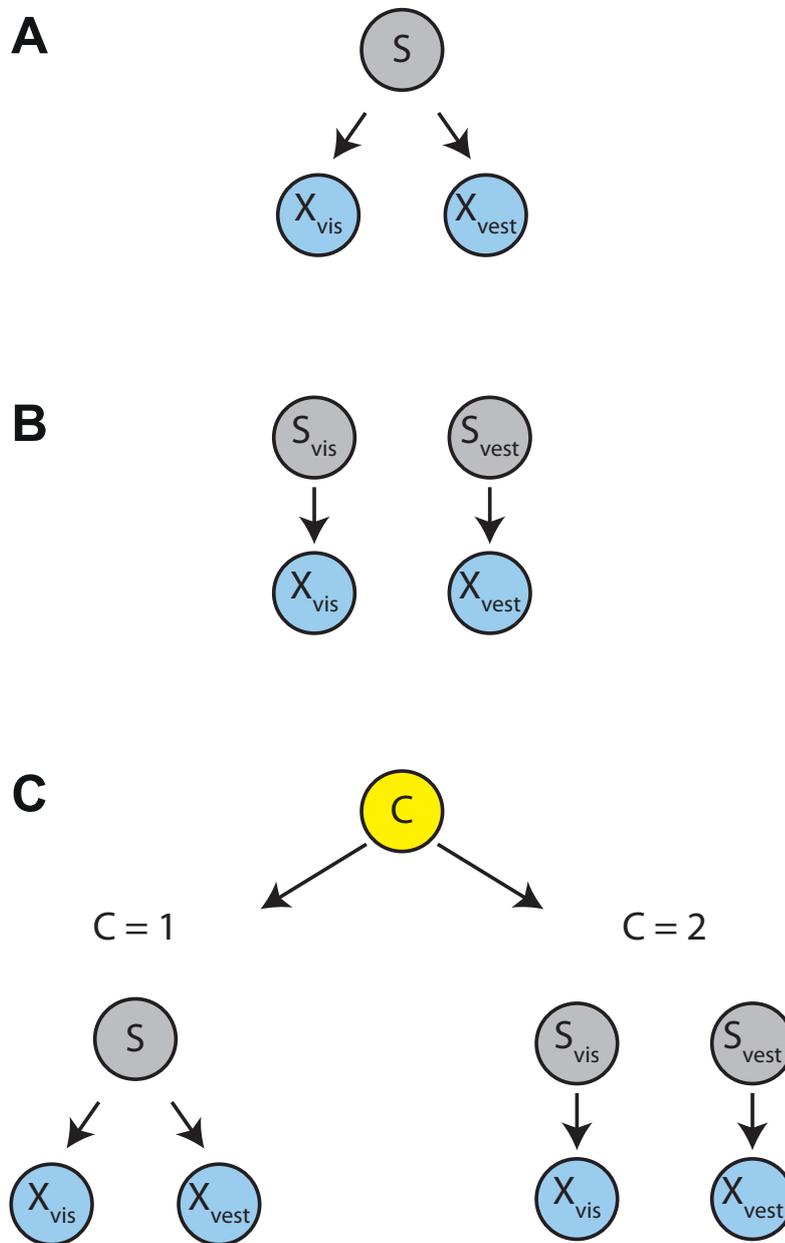
Figure 1.5: Hierarchical causal inference. Visual and vestibular signals are assumed to have (A) a common cause so they get integrated (B) different causes so the visual speed is determined by the visual estimate. (C) Causal structure gets inferred from the weighted average of both scenarios (A) and (B).

1.3.5   Cue comparison

The basic question underlying causal inference models is how the brain decides which signals come from the same event or object and should be integrated and which signals emerge from different sources and should therefore be kept apart. This question can be reformulated: How does the brain evaluate agreement and/or conflict (spatial or temporal) between different sensory inputs? To accomplish the task of conflict detection, signals must be compared. Subsequent processing (integration vs. segregation) depends on the outcome of this comparison.

The first study reported in this thesis (Chapter 2) focuses on visual-vestibular conflict detection, which is crucial to our everyday locomotor and navigation behavior. The nervous system constantly compares these signals, and when conflict arises it can have wide-ranging consequences including vertigo (Bronstein, 2004), inappropriate postural responses (Nashner et al., 1982), motion sickness (Bertolini and Straumann, 2016; Oman, 1990; Reason and Brand, 1975), cyber sickness (Hettinger et al., 1990; Kennedy et al., 1993; McCauley and Sharkey, 1992) and adaptation of reflexive eye movements (Colagiorgio et al., 2015; Mulavara et al., 2005; Solomon et al., 2003). Despite extensive prior research into the consequences of visual-vestibular conflict, few studies have investigated the principles that govern conflict detection. How does the brain evaluate the agreement between signals? The study shows that sensitivity to conflict depends on signal variabilities and is consistent with crossmodal discrimination which is based on simple Signal Dectection Theory (SDT). SDT characterizes how perceivers separate meaningful information from "noise", in other words, how we make decisions under conditions of uncertainty. It assumes that the observer's sensory estimates are graded signals which depend on the sensory evidence and which are corrupted by both external and internal noise. As noise varies from trial to trial, every single sensory estimate can be regarded as a draw from a normal distribution with the mean representing the real value of the stimulus and the variance depending on the amount of noise (Kingdom and Prins, 2010). Observers could for example be asked to compare a certain property of two stimuli, which means that they must subtract the two draws from the respective normal distributions of the stimuli (see Fig. 1.6). The difference distribution resulting from subtracting two normally distributed signals (here signals $S1$ and $S2$) is again normally distributed with mean equal to the difference of compared signal means, and variance equal to the sum of compared signal variances:

$$\sigma^2_{comb} = \sigma^2_{S1} + \sigma^2_{S2} \tag{1.7}$$
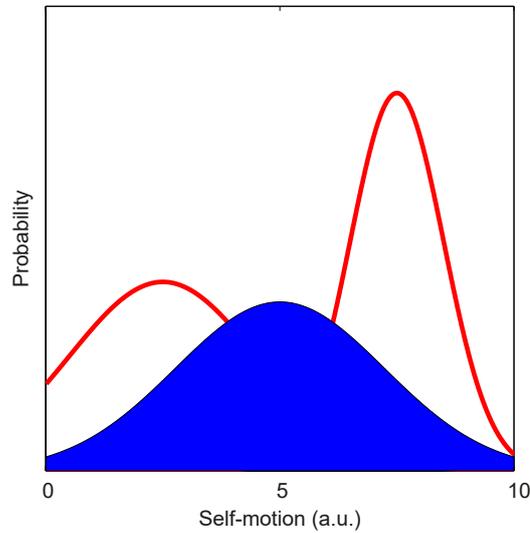
Figure 1.6: Multisensory Comparison. The red distributions represent unisensory estimates, the blue distribution inidcates the difference distribution following crossmodal discrimination.

The only assumption is that noise on the estimates is independent and normally distributed.

1.3.6 Reference frame transformations

Comparing sensory information from different sources requires the brain to find a common reference frame, i.e. a coordinate system enabling us to uniquely describe the location of any given object. Optic flow, for example, provides us with a very sensitive and valuable source for estimating the speed and direction of our self-motion with respect to the world. However, optic flow information is encoded in retinal coordinates and thus only tells about how the eye is moving in space. Without knowing where we are looking at and in which direction our head is pointing relative to the trunk, we do not have a general ego-centric measure of self-motion. Therefore, a crucial ability of the nervous system is to go back and forth between different reference frames to realize a unified representation. While retinal signals are encoded in eye-centered coordinates, vestibular signals are represented in head-centric coordinates as the vestibular organs are fixed within the head. Therefore, extra-retinal signals about eye movements are taken into account to transform retinal signals into a head-centric reference frame (Souman et al., 2006). Furthermore, proprioceptive information about the position of the head relative to all other body parts helps to align vestibular

signals with a body-centric reference frame. While reference frame transformations are often necessary for comparing and integrating multisensory signals, they can sometimes be incomplete, yielding sensory biases, i.e. an objectively perfect match between sensory signals is perceived as inconsistent. The Filehne illusion, for example, describes the phenomenon that stationary objects appear to slightly move in the direction opposite to eye movements during smooth pursuits (Filehne, 1922). This illusion has been explained by an insufficient compensation for retinal slip during eye movements, i.e. errors in reference frame transformations. Differences in perceptual estimates of retinal and oculomotor speed further induce another sensory illusion, the *Aubert-Fleischl (AF) phenomenon*, in which pursued targets are perceived to move more slowly than non-pursued ones (Aubert, 1886). The second study reported in this thesis (Chapter 3) replicated the classical AF phenomenon and investigated whether this illusion is generalizable to the vestibular system, meaning that an object tracked via a vestibular whole-body pursuit is also perceived as slower compared to non-pursued objects. Investigating the mechanisms behind these illusions can help understanding how perceptual stability is achieved during self-motion.

1.4   Psychophysics

While the previous chapters elucidated the importance of sensory noise and biases, this section will cover basic psychophysical methods used for estimating the underlying concepts of precision and accuracy. Further, the experimental setup of the studies reported in this thesis will be briefly explained.

1.4.1   Task

To investigate multisensory cue interactions, 2-interval forced-choice (2IFC) tasks are commonly used. The aim is to fit a psychometric function to the collected data, thereby inferring important parameters which describe the observed behavior, e.g. precision and accuracy. A potential task could be to compare the speed of two successively presented passive yaw rotations. While one interval contains the standard stimulus, which has a certain (constant) intensity/speed, the other interval contains the comparison stimulus, of which the speed varies from trial to trial. The task is to decide which interval contained the stimulus of higher speed. For large differences between the two intervals it is easy to correctly discriminate between the two intervals. For small differences, however, observers will make more errors as they must guess whenever they cannot decide on either of the two intervals.
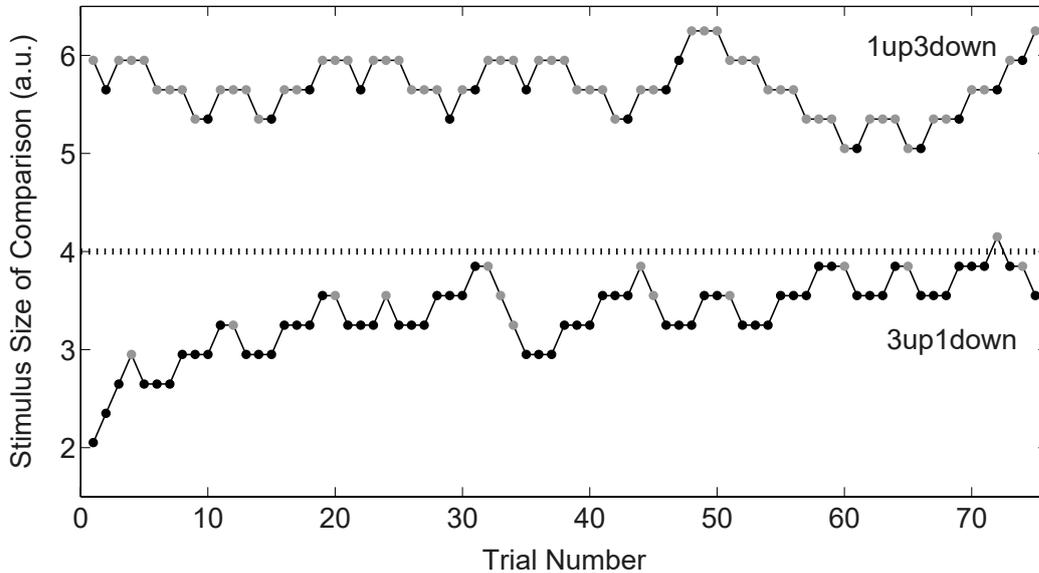
Figure 1.7: Staircase Procedure containing two interleaved staircases (*1up3down*, *3up1down*), light grey dots indicate the observer's answer "comparison perceived as more intense than standard", black dots represent the answer "comparison percevided as less intense than standard", the dotted line indicates the standard stimulus size.

1.4.2 Stimulus selection

In the studies reported in this thesis, an adaptive staircase procedure was used to determine the stimulus intensity of the comparison for the following trial. Staircases represent an efficient way of estimating the two main parameters, the accuracy and precision of sensory estimates, by adjusting the comparison stimulus size depending on responses to previous trials: After each response "comparison perceived as less intense than standard", stimulus intensity increases and after each response "comparison perceived as more intense than standard", stimulus intensity decreases. That way, the difficulty decreases and increases depending on the observer's performance to reach the pair of intensities that cannot be distinguished any more. Interleaving two staircases which start at easily discriminable stimulus intensities is an effective way to avoid response biases and keep up participants' attention during the experiment. Depending on the up/down rule, the staircase converges on different levels. For the *1up3down* and *3up1down* procedure shown in Figure 1.7, the staircases target 79% and 21%, respectively (Leek, 2001).
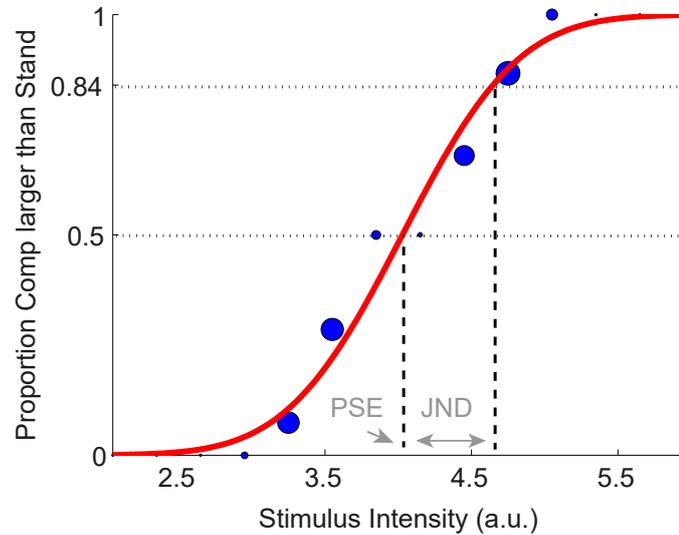
Figure 1.8: Psychometric fit. Data as a proportion of comparison perceived as larger than standard is plotted in blue with bigger size for more frequent stimulus levels. The 50% point corresponds to the PSE, the difference between the 84% and 50% points represents the JND, i.e. one standard deviation of the fitted curve (red).

1.4.3   Psychometric fit

After data collection, the proportion of correct responses (here: proportion of comparison trials perceived as faster than standard) for each comparison stimulus level is determined and a psychometric function is fit to that data using MATLAB (version R2012b) and the Palamedes toolbox (Prins and Kingdom, 2009; Kingdom and Prins, 2010) to determine important parameters that reflect the observer's performance (see Fig. 1.8). The fitting process used in the reported experiments is based on a Maximum Likelihood criterion which aims at finding the values for the free parameters of the psychometric function (out of all possible values) that maximizes the probability of getting the exact same data as the observer.

1.4.4   Parameters of psychometric function

A psychometric fit can be specified by four parameters, which are the location of the curve on the x-axis, its slope, the lapse rate (a lapse can happen when blinking during a visual detection task) and the guessing rate (in case of uncertainty of the judgment). The first two parameters are of primary interest and will thus be further explained with respect to 2IFC tasks.

The position of the psychometric curve on the x-axis represents the point of subjective equality (PSE), i.e. the stimulus intensity at which observers cannot distinguish between the standard and the comparison stimulus as they perceive them as equal in size. This point corresponds to the mean of the fitted curve representing the stimulus intensity that elicits 50% "proportion of comparison trials perceived as faster than standard" responses, meaning observers have to guess which interval contained the faster stimulus. While the PSE is usually close to the real value, it can happen that observers' perception is biased so that the PSE is shifted away from the real value, yielding low accuracy and low validity in psychological terms.

The second parameter of great importance is the slope of the psychometric curve which represents the level of sensory noise. It is also called the just noticeable difference (JND), i.e. the smallest difference in stimulus intensity that can still be noticed by an observer. The smaller this difference threshold the more precise the observer's estimate. This variability or noise measure corresponds to the psychological concept of reliability: the noisier and more variable an estimate, the less reliable it is.

For the reported experiments, a cumulative Gaussian distribution was fit to the data and the JND was defined as the difference between the PSE and the comparison stimulus intensity judged bigger 84% of the time (see Fig. 1.8). This corresponds to the standard deviation of the underlying Gaussian distribution.

1.4.5 Experimental setup

The first two experiments reported in this thesis (Chapter 2 and Chapter 3) were conducted in a virtual reality lab comprising a 6-degree-of-freedom hexapod motion platform (Moog$^{©}$ 6DOF2000E) and a 3D TV screen which allowed us to deliver well-controlled vestibular and visual self-motion stimuli respectively (see 1.9). Subjects were seated in a racing seat mounted on top of the platform and wore either passive-type circular polarizing filter glasses for visual and multisensory stimulation or blindfold goggles for tasks without visual stimulation such as vestibular-only tasks. During the third experiment (Chapter 4), participants wore head-mounted virtual reality goggles (Oculus Rift CV1 HMD) and were instructed to do active head turns instead of being passively rotated on the motion platform.

Figure 1.9: 6-degree-of-freedom hexapod motion platform

1.5   About this thesis

The previous chapter has shown how sensory cues from various modalities
interact to establish our highly complex sense of self-motion perception and
our experience of a stable world. This thesis has two major aims with re-
spect to understanding these achievements. On the one hand, it adds to the
theoretical framework of cue interactions by presenting a model for detec-
tion of conflicts between sensory cues – a step that precedes multisensory
cue integration, which has been widely studied. On the other hand, it high-
lights the crucial role of eye movements for self-motion perception, adding to
the existing but relatively small body of oculomotor literature in self-motion
perception.

The first study (Chapter 2) investigates whether visual-vestibular conflict
detection can be modeled as crossmodal discrimination and if eye movements
have an influence on the performance in such tasks. Whereas a lot of previous
work focused on visual-vestibular integration, this study takes a step back
and looks at the prerequisites for integration of sensory stimuli. Signals
should be integrated if they originate from the same external event or object,
i.e. when they are temporally and spatially congruent - otherwise, there
should be an awareness of incongruency. The process evaluating congruency
constitutes basic conflict detection and has been the center of the first study.

The second study (Chapter 3) approaches conflict detection from a differ-

ent perspective. Multisensory input can be objectively (physically) congruent and still provide illusory percepts of mismatching signals. One phenomenon is the so-called Aubert-Fleischl (AF) illusion, an underestimation of oculomotor speed with respect to retinal speed. The study presented here transfers the classical AF illusion to the vestibular domain and compares speed estimation based on vestibular whole-body pursuits to that based on retinal signals only.

The third study (Chapter 4) investigates whether our findings from our first study about conflict detection are subject to the specific and highly controlled setup or whether they are generalizable to more naturalistic self-motion contexts. Here, participants were presented visual stimuli through a head-mounted display (HMD) to allow for active head motion in contrast to the first study where participants experienced highly controlled passive whole-body yaw rotations on the motion platform, while the head was kept stationary with respect to the body.

The concluding chapter summarizes the findings and discusses them in a broader context, providing insights into possible implications of the research and suggestions for future research in this field.

Chapter 2

Visual-Vestibular Conflict Detection Depends on Fixation

Author contributions

I.T.G. and P.R.M. designed the study
I.T.G. collected the data
I.T.G. and P.R.M. analyzed the data, wrote and revised the paper

Current Biology

Visual-Vestibular Conflict Detection Depends on Fixation

Highlights

- Conflict detection can be simply modeled as crossmodal discrimination

- Visual-vestibular conflict detection is impaired relative to this benchmark

- Head-fixed fixation leads to impaired conflict detection but optimal integration

- Scene-fixed fixation leads to improved conflict detection but impaired integration

Authors

Isabelle T. Garzorz, Paul R. MacNeilage

Correspondence

pmacneilage@unr.edu

In Brief

Using a virtual reality motion simulator, Garzorz and MacNeilage measure how visual and vestibular signals are integrated and how they are compared to detect conflicts. Results reveal a tradeoff mediated by eye movements. Fixation of a head-fixed rather than scene-fixed target leads to optimal integration but highly impaired conflict detection.

CrossMark

CellPress

Visual-Vestibular Conflict Detection Depends on Fixation

Isabelle T. Garzorz[1,2] and Paul R. MacNeilage[1,2,3,4,*]
[1]German Center for Vertigo (DSGZ), University Hospital of Munich, Ludwig Maximilian University, 81377 Munich, Germany
[2]Graduate School of Systemic Neurosciences (GSN), Ludwig Maximilian University, 82152 Planegg-Martinsried, Germany
[3]Present address: Department of Psychology, Cognitive and Brain Sciences, University of Nevada, Reno, NV 89557, USA
[4]Lead Contact
*Correspondence: pmacneilage@unr.edu
http://dx.doi.org/10.1016/j.cub.2017.08.011

SUMMARY

Visual and vestibular signals are the primary sources of sensory information for self-motion. Conflict among these signals can be seriously debilitating, resulting in vertigo [1], inappropriate postural responses [2], and motion, simulator, or cyber sickness [3–8]. Despite this significance, the mechanisms mediating conflict detection are poorly understood. Here we model conflict detection simply as crossmodal discrimination with benchmark performance limited by variabilities of the signals being compared. In a series of psychophysical experiments conducted in a virtual reality motion simulator, we measure these variabilities and assess conflict detection relative to this benchmark. We also examine the impact of eye movements on visual-vestibular conflict detection. In one condition, observers fixate a point that is stationary in the simulated visual environment by rotating the eyes opposite head rotation, thereby nulling retinal image motion. In another condition, eye movement is artificially minimized via fixation of a head-fixed fixation point, thereby maximizing retinal image motion. Visual-vestibular integration performance is also measured, similar to previous studies [9–12]. We observe that there is a tradeoff between integration and conflict detection that is mediated by eye movements. Minimizing eye movements by fixating a head-fixed target leads to optimal integration but highly impaired conflict detection. Minimizing retinal motion by fixating a scene-fixed target improves conflict detection at the cost of impaired integration performance. The common tendency to fixate scene-fixed targets during self-motion [13] may indicate that conflict detection is typically a higher priority than the increase in precision of self-motion estimation that is obtained through integration.

RESULTS

Participants seated on a virtual reality motion simulator experienced two self-rotation stimuli (visual and/or vestibular) and indicated which rotation was larger (Figure 1). Difference in rotation magnitude was varied from trial to trial to find the differences that were just noticeable (the JNDs). Variability was quantified by fitting psychometric functions to the data (see STAR Methods).

All hypotheses were evaluated based on these variability measurements. To examine whether visual-vestibular conflict detection is well modeled as crossmodal discrimination, we first measured variability on visual and vestibular estimates in visual and vestibular conditions (Figure 1), then used the results to generate predictions according to the following equation:

$$\sigma_{Xmodal}^2 = \sigma_{vest}^2 + \sigma_{vis}^2 \qquad \text{(Equation 1)}$$

In line with standard signal detection theory [14], crossmodal discrimination performance should be limited by the sum of variabilities on visual and vestibular estimates if these estimates are conditionally independent.

This prediction was tested in two conditions, one in which visual and vestibular rotations were presented sequentially (Figure 1, sequential) and another in which they were presented simultaneously (Figure 1, simultaneous). Both conditions assessed crossmodal visual-vestibular discrimination, but only the simultaneous condition assessed detection of conflicts between visual and vestibular stimuli that were experienced at the same time, as would be required for conflict detection in natural settings.

In addition to conflict detection, we also evaluated visual-vestibular integration. Simultaneously presented visual and vestibular cues have previously been shown to be integrated in a maximum likelihood (ML) fashion [11, 12]:

$$\sigma_{comb}^2 = \frac{\sigma_{vest}^2 \sigma_{vis}^2}{\sigma_{vest}^2 + \sigma_{vis}^2} \qquad \text{(Equation 2)}$$

We sought to evaluate this prediction in the combined condition (Figure 1, combined), in which congruent visual and vestibular cues were presented in both intervals.

To examine the influence of eye movements, we ran all conditions with both head-fixed and scene-fixed fixation (Figure 2, left). During head-fixed fixation, observers fixated a point that was stationary relative to the head such that eye movement was minimal and optic flow was maximal, similar to prior visual-vestibular integration studies [11, 12]. During scene-fixed fixation, observers fixated a point that moved with the scene such that optic flow was minimal while eye movement was maximal, similar to natural gaze stabilization behavior [13].
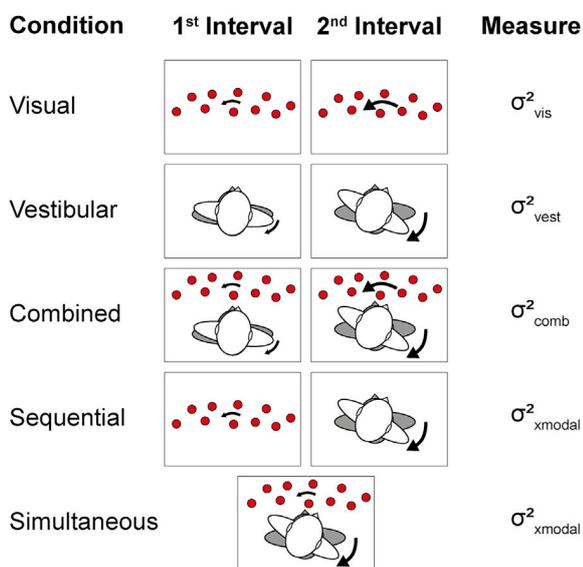
Figure 1. Experimental Conditions

In most conditions, a trial consisted of two consecutive passive self-rotations in yaw (1st interval, 2nd interval), and observers indicated which movement was perceived as larger. Each rotation had a raised cosine velocity profile with 0.8 s duration. Peak rotation velocity in the standard interval was always 10°/s (4° displacement), whereas velocity in the comparison interval varied (order was randomized). Gray and white icons depict top-down views of the observer on the motion platform and indicate that a physical rotation was presented. Red dots depict the simulated visual scene and indicate visual stimulus presentation. The size of the black arrows symbolizes speed of rotation (smaller indicates slower; larger indicates faster). In the visual condition, both intervals contained visually simulated self-rotation. In the vestibular condition, both intervals consisted of passive physical self-rotation in darkness. In the combined condition, both intervals consisted of synchronized physical and visually simulated self-rotation of equal speed. In the sequential condition, one rotation was visual and the other was physical (randomized order). The simultaneous condition consisted of a single interval with synchronized physical and visual rotations of different speed. Observers indicated whether the visual movement was too fast or too slow with respect to the physical rotation. All five conditions were tested with both head-fixed and scene-fixed fixation. Variabilities measured in each condition are listed in the rightmost column.


Figure 2. Crossmodal Discrimination

JNDs observed during sequential (left) and simultaneous (right) conditions with head-fixed (top) and scene-fixed (bottom) fixation are plotted versus JNDs predicted by the signal-detection model of crossmodal discrimination (Equation 1). Predictions were generated using data from single-cue visual and vestibular conditions (Figure S1) and corrected for visual-to-vestibular gains bigger than 1 (Equation 6; Figure S3). The slope of the green line in each panel is the (geometric) mean of the observed-to-predicted ratio (Figure S2), and it indicates the degree to which variability deviated from the prediction. Gray shaded areas represent the SEM of the geometric mean (A: 1.34 [± 0.31], B: 2.18 [± 0.36], C: 1.27 [± 0.28], D: 1.46 (± 0.40]). Deviation was significant in all conditions (Figure S2). Comparing across conditions, deviation was significantly greater during head-fixed fixation (top; ANOVA, F = 5.75, p = 0.02) and during the simultaneous condition (right; ANOVA, F = 10.97, p = 0.001). Light and dark green dots indicate data from sub-conditions in which either the visual or vestibular interval (respectively) was the standard stimulus (see STAR Methods). There was no effect of standard type, and there were no interactions. Complete statistical and bootstrap analyses are reported in Figure S2.

Crossmodal discrimination was worse than predictions in all conditions (t test, p < 0.001). We speculate that this may be a consequence of mapping uncertainty. In other words, if observers are uncertain about visual-vestibular matching such that the gain perceived as matching varies from trial to trial, this would be reflected as increased variability in matching performance.

Crossmodal discrimination agreed more closely with predictions when fixating a scene-fixed target (Figure 2, bottom) than when fixating a head-fixed target (Figure 2, top), suggesting that gaze-stabilizing eye movements facilitate crossmodal discrimination (ANOVA factor fixation; F = 5.57, p = 0.02). This may be because retinal motion, i.e., retinal slip, provides a direct measure of conflict during scene-fixed fixation.

Crossmodal discrimination also depended on whether presentation was simultaneous (Figure 2, right) or sequential (Figure 2, left) (ANOVA factor condition; F = 10.97, p < 0.01). Prior

studies have reported variability for matching tasks similar to the simultaneous condition [15–17], referring to this as the "range of immobility" because within this range the environment is perceived as stationary [16]. However, no prior study has measured single-cue variabilities and tested results against the crossmodal discrimination benchmark.

The sequential condition served as a crossmodal discrimination control, closely mimicking the two-interval design used to measure visual and vestibular variabilities in visual and vestibular conditions. Additional variability in the simultaneous relative to the sequential condition is interpreted to be a result of conflict detection processes that are engaged specifically when visual and vestibular signals are synchronized. Efficiency in conflict detection was therefore calculated by comparing performance between the simultaneous and sequential conditions (Figures 3A and 3C). The null hypothesis was that performance would be the same in these conditions because the task in both cases amounts to crossmodal discrimination. Similar performance was observed during scene-fixed fixation (Figure 3C), suggesting

Conflict Detection
Hypothesis

Integration
Hypothesis

Head-Fixed

Scene-Fixed
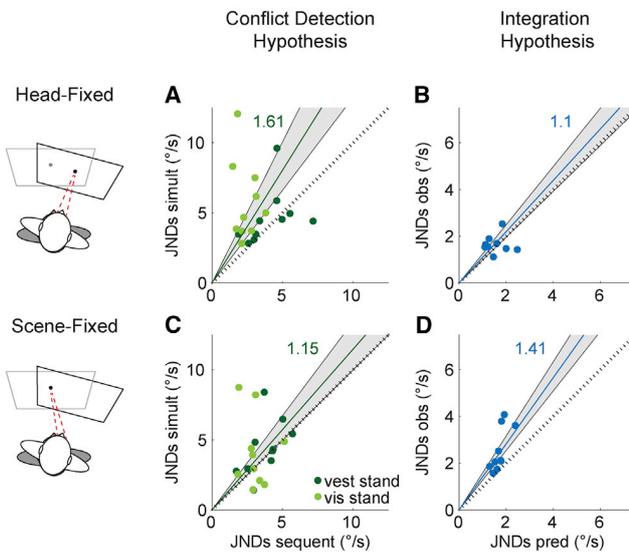
Figure 3. Conflict Detection and Integration

(A and C) Conflict detection in the simultaneous condition was hypothesized to match crossmodal discrimination measured in the sequential condition (see also Figure S1); the line with unity slope indicates hypothesized performance. The slope of the green line is the (geometric) mean of the simultaneous-to-sequential ratio (see also Figure S2) and indicates the average proportional increase in variability in the simultaneous compared to the sequential condition. Gray shaded areas represent the SEM of the geometric mean (A: 1.61 [± 0.40], C: 1.15 [± 0.40]). This increase was significant (t test, p < 0.01) during head-fixed (A), but not scene-fixed (C) fixation (t test, p = 0.30), suggesting that suppression of gaze-stabilizing eye movements compromised conflict detection (paired t test, p < 0.05).

(B and D) Integration was hypothesized to follow predictions of the ML integration model; the line with unity slope indicates hypothesized performance. The slope of the blue line is the (geometric) mean of the observed-to-predicted ratio (see also Figure S2) and indicates the degree to which variability exceeded the prediction. Gray shaded areas represent the SEM of the geometric mean (B: 1.10 [± 0.45], D: 1.41 [± 0.40]). The increase was significant (t test, p < 0.01) during scene-fixed (D) but not head-fixed (B) fixation (t test, p = 0.41), suggesting that gaze-stabilizing eye movements compromised ML integration (paired t test, p = 0.08). Complete statistical and bootstrap analyses are reported in Figure S2.

efficient conflict detection during naturalistic gaze stabilization. However, head-fixed fixation (Figure 3A) led to conflict detection that was significantly impaired relative to the sequential control (p < 0.01).

While naturalistic gaze stabilization (i.e., scene-fixed fixation) allowed for the best conflict detection, it led to integration performance (Figure 3D) that was significantly impaired (p < 0.01) relative to the ML integration model (Equation 2). Performance in the head-fixed condition (Figure 3B), on the other hand, resembled predictions of the ML model, despite the observed inefficiency in conflict detection.

In summary, eye movements modulated the effectiveness of the system with respect to conflict detection and integration. When eye movements were nulled via fixation of a head-fixed point, integration was consistent with optimal predictions but

conflict detection was impaired. When the eyes instead tracked a scene-fixed point, such that eye movement was approximately equal and opposite head movement, conflict detection improved but integration was suboptimal.

DISCUSSION

Conflict detection can be modeled as crossmodal discrimination with benchmark performance limited by variabilities on the signals being compared. Visual-vestibular conflict detection is generally impaired relative to this crossmodal discrimination benchmark, especially when participants are instructed to fixate a head-fixed target. When the eyes instead track a scene-fixed target, conflict detection is better but integration is impaired. We conclude that (1) there is a tradeoff whereby the system can be optimized for *either* conflict detection *or* integration and that (2) the priority placed on these operations is modulated depending on eye movements.

Forced Fusion Can Account for Impaired Conflict Detection

Forced-fusion models of cue integration have been proposed to explain integration of visual and vestibular signals [12], as well as integration in other contexts [18]. According to these models, when cues are integrated, the observer loses access to the contributing single-cue estimates, which could explain the association between optimal integration and impaired conflict detection observed during head-fixed fixation. The idea is illustrated in Figure 4 with a model that borrows from the work of Ernst and colleagues [19]. Integration and conflict detection is governed by the posterior distribution (Figures 4C and 4G), which is the product of the likelihood function (Figures 4A and 4E) and prior (Figures 4B and 4F) distribution. The likelihood represents information available from visual and vestibular modalities in isolation. The coupling prior represents the strength of expectation that visual and vestibular cues will be in agreement. To evaluate conflict, we marginalized the posterior distribution onto the visual and vestibular axes (blue curves) and compared these distributions (Figures 4D and 4H).

The model predicts that conflict detection depends on the coupling prior. When variability of the prior is small (Figure 4B), the posterior is pulled toward the diagonal (Figure 4C). Consequently, marginal distributions are similar, and conflict is difficult to detect (Figure 4D, green difference distribution not significantly different from zero). Conversely, when the coupling prior is weak (Figure 4F), the posterior (Figure 4G) resembles the likelihood, and conflict detection follows predictions of the crossmodal discrimination model (Figure 4H).

The only prior study to measure both visual-vestibular conflict detection and integration performance [17] also reports that these are related. In this study, conflict detection was best in subjects that gave less weight to the visual cue during integration. They propose a fusion-referenced detection (FRD) model to account for their results. Conflict detection is preceded by a fusion stage in which visual and vestibular signals are always integrated, similar to forced-fusion models described above. However, conflict detection is not evaluated through comparison of (marginalized) visual and vestibular estimates (i.e., crossmodal discrimination). Instead, the visual estimate is compared with

Likelihood | Prior | Posterior | Conflict Detection
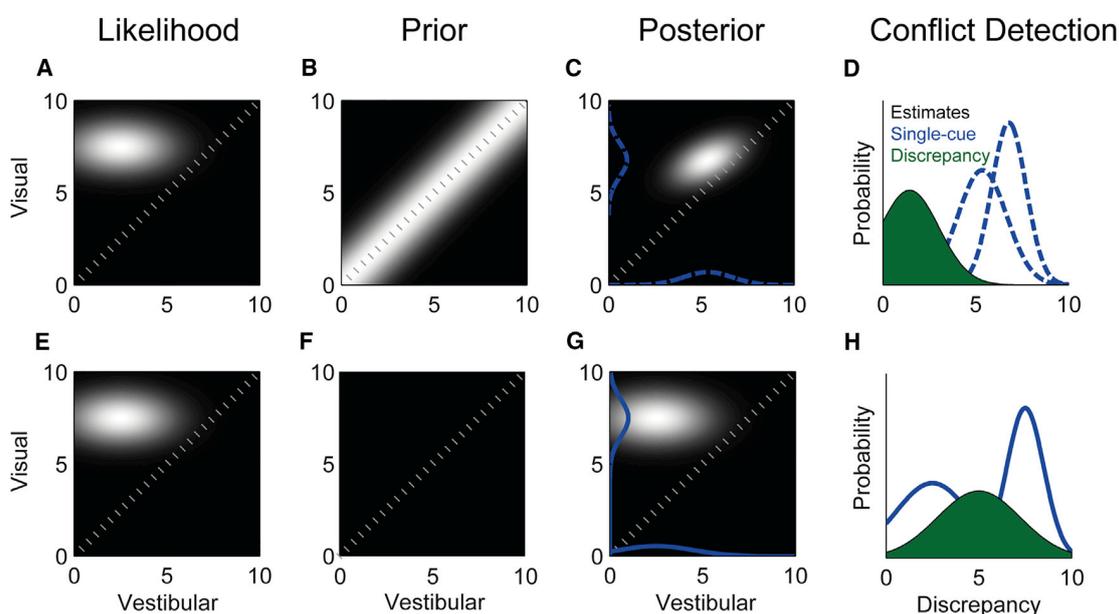
Figure 4. Forced-Fusion Model of Integration and Conflict Detection

Bayesian combination of visual and vestibular information with different prior distributions. The left column (A and E) shows the joint likelihood of the noisy visual and vestibular measurements in the form of a bivariate normal distribution. The middle column (B and F) represents a strong (B) and a uniform (F) coupling prior distribution. Applying Bayes rule, the normalized product of the likelihood functions and prior distributions results in the posterior distribution (C and G). Conflict detection is impaired if participants have a strong coupling prior indicated by the shift of the difference distribution toward zero (D, green). In the case of a uniform coupling prior, the posterior is identical to the joint likelihood, yielding a difference distribution with mean significantly different from zero (H, green). Note that the difference distribution obtained by marginalizing cues onto the axes (D and H, blue) and then subtracting (D and H, green) is equivalent to the distribution that would be obtained by marginalizing onto the negative diagonal.

the fused (i.e., integrated) visual-vestibular estimate. When visual weight on the fused estimate is high, it approaches the visual estimate, making conflict difficult to detect.

We did not measure visual weights, but we evaluated whether their model is generally consistent with our results by calculating visual weights predicted by the ML model [20] ($\widehat{w}_{vis} = \sigma^2_{vest}/(\sigma^2_{vis} + \sigma^2_{vest})$) and testing whether these are higher during head-fixed than scene-fixed fixation, i.e., higher when conflict detection is worse. Mean visual weight predicted based on the ML model is indeed significantly higher during head-fixed fixation (head-fixed = 0.61; scene-fixed = 0.49; one-sided paired t test, p = 0.03), consistent with predictions of the FRD model. This raises the possibility that the observed dependence of conflict detection on fixation may be a specific case of a more general dependence of conflict detection on visual weight.

Fixation Modulates Tradeoff between Integration and Conflict Detection

Integration and conflict detection support precision and accuracy on visual-vestibular estimates, respectively. Ideally, integration and conflict detection could proceed in parallel, but the apparent tradeoff observed here suggests a relatively inflexible system that is optimized for either one or the other. We observed that self-motion is processed differently depending on the percentage of the visual self-motion signal that is due to retinal versus oculomotor motion (i.e., head-fixed is 100% retinal; scene-fixed is 100% oculomotor), suggesting that vestibular sig-

nals are most effectively *compared* with oculomotor signals and most effectively *integrated* with retinal signals.

During locomotion, we most often fixate world-fixed targets [13] in order to collect information about the environment. Under these circumstances, dynamic visual acuity depends on the vestibulo-oculur reflex [21] which must be calibrated by an error signal, i.e., by detecting conflicts. Whereas gaze stabilization may be best served by conflict detection, precise postural and locomotor control are better served by integration, because it allows for greater precision in estimating self-motion. However, when one signal is much more variable than the other, integration is less advantageous, and perhaps unnecessary, because variability of the combined estimate approaches that of the less-variable signal (see Equation 2). Overall, the tendency to fixate scene-fixed objects may indicate that conflict detection is typically a higher priority than the sometimes small increase in precision of self-motion estimation that is obtained through ML integration.

Alternatively, fixation-dependent performance observed here may be specific to the context of the current experiment, in which subjects made perceptual judgments during slow passive head rotation, and may not be relevant during natural locomotion. Additional studies are needed to evaluate this possibility. Nevertheless, the current finding aligns with prior behavioral research [22–24] documenting differential processing that depends on whether the visual motion signal is predominantly retinal or oculomotor.

Physiological studies have also documented these dependencies [25–28]. Retinal, oculomotor, and vestibular signals are

known to converge in numerous brain areas involved in control of posture, eye movements, and perception [25, 29–31]. Perceptual judgments like those assessed here likely depend on processing in cortical areas including the medial superior temporal area (MST), parietoinsular vestibular cortex (PIVC), and ventral intraparietal area (VIP) [32–34]. Signals are represented in different reference frames across these areas, with some being more eye-centered [35] and others being more head- or body-centered [31]. Head-fixed fixation brings eye and head reference frames into alignment, which could facilitate optimal readout of eye-centered neural populations for ML integration. A potential neural substrate for ML visual-vestibular integration has been identified [36–38], but these studies were conducted with head-fixed fixation. The failure of integration during scene-fixed fixation brings into question the generality of these findings. Populations of neurons with opposite, rather than congruent, visual vestibular tuning [39, 40] are well-suited for detecting conflict between visual and vestibular signals. Moving forward, the complex interactions between retinal, oculomotor, and vestibular signals during naturalistic movement and their dependence on task demands remains a worthy topic for future research.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR RESOURCE SHARING
- SUBJECT DETAILS
- METHOD DETAILS
  - Equipment
  - Experimental Procedure and Conditions
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Fitting Psychometric Functions
  - Predictions
  - Statistics
  - Correction of Predicted JNDs for Crossmodal Discrimination
  - Eye Movement Recording and Analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.08.011.

AUTHOR CONTRIBUTIONS

I.T.G. and P.R.M. designed the study, analyzed the data, and wrote the paper. I.T.G. collected the data.

REFERENCES

1. Bronstein, A.M. (2004). Vision and vertigo: some visual aspects of vestibular disorders. J. Neurol. 251, 381–387.

2. Nashner, L.M., Black, F.O., and Wall, C., 3rd. (1982). Adaptation to altered support and visual conditions during stance: patients with vestibular deficits. J. Neurosci. 2, 536–544.

3. Bertolini, G., and Straumann, D. (2016). Moving in a moving world: a review on vestibular motion sickness. Front. Neurol. 7, 14.

4. Oman, C.M. (1990). Motion sickness: a synthesis and evaluation of the sensory conflict theory. Can. J. Physiol. Pharmacol. 68, 294–303.

5. Reason, J., and Brand, J.J. (1975). Motion Sickness (Academic Press).

6. Hettinger, L.J., Berbaum, K.S., Kennedy, R.S., Dunlap, W.P., and Nolan, M.D. (1990). Vection and simulator sickness. Mil. Psychol. 2, 171–181.

7. Kennedy, R.S., Lane, N.E., Berbaum, K.S., and Lilienthal, M.G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. 3, 203–220.

8. McCauley, M.E., and Sharkey, T.J. (1992). Cybersickness: perception of self-motion in virtual environments. Presence: Teleoper. Virtual Environ. 1, 311–318.

9. Fetsch, C.R., Turner, A.H., DeAngelis, G.C., and Angelaki, D.E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. J. Neurosci. 29, 15601–15612.

10. ter Horst, A.C., Koppen, M., Selen, L.P., and Medendorp, W.P. (2015). Reliability-based weighting of visual and vestibular cues in displacement estimation. PLoS ONE 10, e0145015.

11. Butler, J.S., Smith, S.T., Campos, J.L., and Bülthoff, H.H. (2010). Bayesian integration of visual and vestibular signals for heading. J. Vis. 10, 23.

12. Prsa, M., Gale, S., and Blanke, O. (2012). Self-motion leads to mandatory cue fusion across sensory modalities. J. Neurophysiol. 108, 2282–2291.

13. Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E., and König, P. (2007). Human eye-head co-ordination in natural exploration. Network 18, 267–297.

14. Wickens, T.D. (2002). Elementary Signal Detection Theory (Oxford University Press).

15. Jaekl, P.M., Jenkin, M.R., and Harris, L.R. (2005). Perceiving a stable world during active rotational and translational head movements. Exp. Brain Res. 163, 388–399.

16. Wallach, H. (1985). Perceiving a stable environment. Sci. Am. 252, 118–125.

17. Jürgens, R., and Becker, W. (2011). Human spatial orientation in non-stationary environments: relation between self-turning perception and detection of surround motion. Exp. Brain Res. 215, 327–344.

18. Hillis, J.M., Ernst, M.O., Banks, M.S., and Landy, M.S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. Science 298, 1627–1630.

19. Ernst, M., and Di Luca, M. (2011). Multisensory perception: from integration to remapping. In Sensory Cue Integration, J. Trommershäuser, M. Landy, and K. Körding, eds. (Oxford University Press).

20. Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415, 429–433.

21. Colagiorgio, P., Bertolini, G., Bockisch, C.J., Straumann, D., and Ramat, S. (2015). Multiple timescales in the adaptation of the rotational VOR. J. Neurophysiol. 113, 3130–3142.

22. Freeman, T.C.A., Champion, R.A., and Warren, P.A. (2010). A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. Curr. Biol. 20, 757–762.

23. Banks, M.S., Ehrlich, S.M., Backus, B.T., and Crowell, J.A. (1996). Estimating heading during real and simulated eye movements. Vision Res. 36, 431–443.

24. Clemens, I.A., Selen, L.P., Pomante, A., MacNeilage, P.R., and Medendorp, W.P. (2017). Eye movements in darkness modulate self-motion perception. eNeuro 4, ENEURO.0211-16.2016.

25. Bremmer, F. (2011). Multisensory space: from eye-movements to self-motion. J. Physiol. *589*, 815–823.

26. Brostek, L., Büttner, U., Mustari, M.J., and Glasauer, S. (2013). Neuronal variability of MSTd neurons changes differentially with eye movement and visually related variables. Cereb. Cortex *23*, 1774–1783.

27. Bremmer, F., Ilg, U.J., Thiele, A., Distler, C., and Hoffmann, K.P. (1997). Eye position effects in monkey cortex. I. Visual and pursuit-related activity in extrastriate areas MT and MST. J. Neurophysiol. *77*, 944–961.

28. Sunkara, A., DeAngelis, G.C., and Angelaki, D.E. (2016). Joint representation of translational and rotational components of optic flow in parietal cortex. Proc. Natl. Acad. Sci. USA *113*, 5077–5082.

29. Cullen, K.E. (2012). The vestibular system: multimodal integration and encoding of self-motion for motor control. Trends Neurosci. *35*, 185–196.

30. Gu, Y., Cheng, Z., Yang, L., DeAngelis, G.C., and Angelaki, D.E. (2016). Multisensory convergence of visual and vestibular heading cues in the pursuit area of the frontal eye field. Cereb. Cortex *26*, 3785–3801.

31. Chen, X., Deangelis, G.C., and Angelaki, D.E. (2013). Diverse spatial reference frames of vestibular signals in parietal cortex. Neuron *80*, 1310–1321.

32. Chen, A., Gu, Y., Liu, S., DeAngelis, G.C., and Angelaki, D.E. (2016). Evidence for a causal contribution of macaque vestibular, but not intraparietal, cortex to heading perception. J. Neurosci. *36*, 3789–3798.

33. Roberts, R.E., Ahmad, H., Arshad, Q., Patel, M., Dima, D., Leech, R., Seemungal, B.M., Sharp, D.J., and Bronstein, A.M. (2017). Functional neuroimaging of visuo-vestibular interaction. Brain Struct. Funct. *222*, 2329–2343.

34. Smith, A.T., Greenlee, M.W., DeAngelis, G.C., and Angelaki, D.E. (2017). Distributed visual–vestibular processing in the cerebral cortex of man and macaque. Multisens. Res. *30*, 91–120.

35. Chen, X., DeAngelis, G.C., and Angelaki, D.E. (2014). Eye-centered visual receptive fields in the ventral intraparietal area. J. Neurophysiol. *112*, 353–361.

36. Gu, Y., Angelaki, D.E., and Deangelis, G.C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. Nat. Neurosci. *11*, 1201–1210.

37. Angelaki, D.E., Gu, Y., and Deangelis, G.C. (2011). Visual and vestibular cue integration for heading perception in extrastriate visual cortex. J. Physiol. *589*, 825–833.

38. Fetsch, C.R., Pouget, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Neural correlates of reliability-based cue weighting during multisensory integration. Nat. Neurosci. *15*, 146–154.

39. Bremmer, F., Kubischik, M., Pekel, M., Lappe, M., and Hoffmann, K.P. (1999). Linear vestibular self-motion signals in monkey medial superior temporal area. Ann. N Y Acad. Sci. *871*, 272–281.

40. Gu, Y., Watkins, P.V., Angelaki, D.E., and DeAngelis, G.C. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. J. Neurosci. *26*, 73–85.

41. Prins, N., and Kingdom, F.A.A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. http://www.palamedestoolbox.org.

42. Watt, S.J., Akeley, K., Ernst, M.O., and Banks, M.S. (2005). Focus cues affect perceived depth. J. Vis. *5*, 834–862.

43. Grabherr, L., Nicoucar, K., Mast, F.W., and Merfeld, D.M. (2008). Vestibular thresholds for yaw rotation about an earth-vertical axis as a function of frequency. Exp. Brain Res. *186*, 677–681.

44. Correia Grácio, B.J., Bos, J.E., van Paassen, M.M., and Mulder, M. (2014). Perceptual scaling of visual and inertial cues: effects of field of view, image size, depth cues, and degree of freedom. Exp. Brain Res. *232*, 637–646.

45. Brenner, E. (1993). Judging an object's velocity when its distance changes due to ego-motion. Vision Res. *33*, 487–504.

46. Brenner, E. (1991). Judging object motion during smooth pursuit eye movements: the role of optic flow. Vision Res. *31*, 1893–1902.

47. Brenner, E., and van den Berg, A.V. (1994). Judging object velocity during smooth pursuit eye movements. Exp. Brain Res. *99*, 316–324.

48. van den Berg, A.V., and Beintema, J.A. (2000). The mechanism of interaction between visual flow and eye velocity signals for heading perception. Neuron *26*, 747–752.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
MATLAB R2010b (Data analysis)	MathWorks Inc.	https://www.mathworks.com
Psychtoolbox-3 (Visual stimulation)	N/A	http://psychtoolbox.org
Palamedes toolbox	[41] Palamedes: MATLAB routines for analyzing psychophysical data.	http://www.palamedestoolbox.org
Hexapod motion platform Moog 6DOF2000E	N/A	http://www.moog.com
Stereo screen JVC - GD-463D10	N/A	http://www.jvc.net
(EyeSeeCam)	Interacoustics	http://www.interacoustics.com/eyeseecam
OpenGL	N/A	https://www.opengl.org

CONTACT FOR RESOURCE SHARING

Further information and requests for resources should be directed to the Lead Contact, Paul R. MacNeilage (pmacneilage@unr.edu).

SUBJECT DETAILS

Ten healthy participants (seven females and three males), ranging in age from 25 to 41 (mean age = 28.4 years), with normal or corrected-to-normal vision, took part in the experiment. All but two were naive to the purpose of the study. Participants had no history of neurological, visual, or vestibular sensory disorders and had normal or corrected-to-normal vision. They were paid 8 Euros per hour for their participation.

Prior to the experiment, all participants gave informed consent to participate. The study was conducted in accordance with the ethical standards of the World Medical Council as laid down in the Declaration of Helsinki. All procedures were approved by the ethics committee of the University Hospital of Munich.

METHOD DETAILS

Equipment

The experiment was conducted on a 6-degree-of-freedom hexapod motion platform (Moog 6DOF2000E). Participants were seated in a racing seat mounted on top of the platform. In order to keep a stable head position during the experiment, the head was fixated with two padded mounts at the temples. White noise presented via noise-cancelling headphones masked the sounds of the active motion platform during the trials. The vestibular-only conditions were conducted in darkness with participants wearing opaque goggles. The visual stimuli were presented on a stereo screen (JVC - GD-463D10, refresh rate: 60 Hz) with dimensions 101.8 cm x 57.3 cm, located 33 cm in front of the eyes, yielding a 110° x 80° of visual angle field of view. Participants wore custom-made goggles consisting of a circular polarizing filter for stereo and a blurring film to weaken accommodative cues to screen distance [42]. The goggles also prevented observers from seeing the edges of the screen. The visual stimulus was rendered using Psychtoolbox and OpenGL. Responses in the experiment were collected using a response box with two buttons.

Experimental Procedure and Conditions

On each trial, participants performed a two-alternative-forced-choice (2AFC) task in which they indicated with a button press which of two yaw self-rotations was "bigger." Each rotation had a raised cosine velocity profile [43] of constant duration of 0.8 s. Because duration was fixed, displacement, velocity and acceleration scaled together, and we informed subjects that this was the case. We explicitly instructed them to respond which rotation was larger (displacement), faster (velocity), and stronger (acceleration). On each trial, one rotation was the standard stimulus which had peak velocity of 10°/s (4° displacement), and the other rotation was the comparison stimulus which had peak velocity that varied from trial to trial according to a staircase procedure (described below). Order of standard and comparison stimulus was randomized.

Depending on condition, rotations were either physical, i.e., delivered via rotation of the motion platform, visually-simulated, i.e., delivered via a visual motion stimulus on the display, or simultaneous visual and physical rotation. The visual scene consisted of a 3-dimensional volume (150 cm x 50 cm x 150 cm) of randomly placed red spheres (radius = 0.3 cm) at a density of 0.007 spheres/cm³ with an empty band 10 cm below and above the fixation point. During visually-simulated self-motion, motion of the spheres on the screen elicited perception of self-motion relative to this scene.

In most conditions, the two rotations to be judged were presented in two consecutive intervals, i.e., two-interval forced choice (2IFC). The direction of rotation (left versus right) was randomized across the experiment, but both rotations for a given trial were in the same direction. Rotation intervals were separated in time by a 0.5 s pause. After the trial was completed participants were prompted by a tone (0.2 s) to respond using the button box. Then the motion platform and/or visual scene (depending on condition) were rotated back to the initial position and the next trial began after a pause of 0.5 s with a black screen. For conditions with both visual and physical rotation, visual and vestibular stimulation were matching during the move back (i.e., no conflict).

There were five conditions (Figure 1). In the two single-cue conditions (visual and vestibular) both rotations were either visual or physical. These conditions were run to measure variabilities on visual and vestibular estimates. In the combined condition both rotations consisted of synchronized and matching visual and physical rotation. This condition was run to evaluate the ML integration. In the sequential condition, one rotation was physical and the other visual, with order randomized across trials. This condition was run to evaluate crossmodal discrimination. In the simultaneous condition, synchronized visual and physical rotation were presented during a single interval and subjects were instructed to respond whether the visual rotation was slower or faster than the physical rotation, or equivalently, whether the visual scene appeared to move with or against their own self-motion in world coordinates. This condition was run to evaluate how simultaneous presentation impacts crossmodal discrimination.

Each condition was run twice by each subject, once with head-fixed and once with scene-fixed fixation (Figure 2, left). During head-fixed fixation, a white fixation point was presented in the middle of the screen at screen depth and remained there throughout the trial, resulting in suppression of eye movements. In conditions with scene-fixed fixation, the fixation point moved at the same speed as the surrounding spheres, eliciting eye movement that was equal and opposite the self-rotation.

Each condition and fixation type was run separately. The size of the comparison yaw rotation on each trial was calculated using staircase procedures (Palamedes toolbox). For visual, vestibular and combined conditions, participants completed a total of 150 trials, consisting of two interleaved staircases (3up1down, 1up3down) of 75 trials each. There were two variants of the simultaneous and sequential conditions, one in which the visual rotation was the standard and the vestibular was the comparison (visual standard), and vice-versa (vestibular standard), allowing for a balanced design. Therefore, participants completed 300 trials total for these conditions, four interleaved staircases of 75 trials each. At the start of each condition participants completed 10 practice trials with verbal feedback to make sure that they had understood the task correctly. All conditions were divided into blocks of 50 trials with short breaks in between to maintain participants' attentiveness.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using MATLAB (version R2010b).

Fitting Psychometric Functions

Cumulative Gaussians were fit to the data for each participant and condition using the Palamedes toolbox [41] in order to estimate the point of subjective equality (PSE) and the just-noticeable difference (JND). The PSE is defined as the mean of the cumulative Gaussian fit, i.e., the stimulus intensity that elicits 50% "comparison bigger than standard" responses. The JND is the difference between the PSE and the comparison stimulus intensity judged bigger 84% of the time. This corresponds to the standard deviation of the cumulative Gaussian fit. Deviation of the PSE from the reference stimulus represents accuracy whereas JND represents precision or variability.

Predictions

According to standard signal detection theory, performance in a two-alternative-forced-choice discrimination task is limited by the sum of variabilities associated with the two alternative estimates assuming these are conditionally independent [14]: $\sigma^2_{disc} = \sigma^2_{alt1} + \sigma^2_{alt2}$. (Note, this is the generic form of Equation 1.) In the visual, vestibular and combined conditions $\sigma^2_{alt1} = \sigma^2_{alt2}$, so variability on these estimates (σ^2_{vest}, σ^2_{vis}, σ^2_{comb}) is equal to half the squared JND: $\sigma^2 = JND^2/2$. In the sequential and simultaneous conditions, on the other hand, variability on crossmodal discrimination performance is equal to the squared JND: $\sigma^2_{xmodal} = JND^2_{xmodal}$.

The crossmodal discrimination model predicts that discrimination performance will be limited by the sum of variabilities associated with single-cue visual and vestibular estimates (see Equation 1):

$$JN\widehat{D}_{xmodal} = \sqrt{\sigma^2_{vest} + \sigma^2_{vis}} \qquad \text{(Equation 3)}$$

The Maximum-Likelihood cue integration model predicts that the combined JND should also depend on variance associated with the single-cue visual and vestibular estimates (see Equation 2):

$$JN\widehat{D}_{comb} = \sqrt{2\,\frac{\sigma^2_{vest}\sigma^2_{vis}}{\sigma^2_{vest} + \sigma^2_{vis}}} \qquad \text{(Equation 4)}$$

Thus, JNDs measured in visual and vestibular conditions can be used to predict those observed in the sequential, simultaneous, and combined conditions according to Equations 3 and 4.

Deviation of observed from predicted JNDs was evaluated by taking their ratio and then calculating the log value:

$$D = log\left(\frac{JND_{obs}}{JND_{pred}}\right) \qquad \text{(Equation 5)}$$

The ratio allows for a normalized measure of deviation that does not depend on the absolute magnitudes of the JNDs, and the log transformation preserves symmetry of positive and negative deviations.

When evaluating the crossmodal discrimination model, observed JNDs were those measured in sequential and simultaneous conditions and predicted JNDs were calculated according to Equation 3 (Figures 2 and S2). When evaluating integration, observed JNDs were those measured in the combined condition and predicted JNDs were calculated according to Equation 4 (Figures 3B, 3D and S2D, S2H). When evaluating conflict detection, observed JNDs were those measured in the simultaneous condition and predicted JNDs were those observed in the sequential condition (Figures 3A, 3C and S2C, S2G).

Statistics

All statistical calculations and tests were performed on the deviation measures described above (Equation 5). Mean deviation values are illustrated by the horizontal lines in Figure S2. The exponentials of these mean values are the geometric means of the observed-to-predicted ratios; these values are illustrated by the slopes of the green lines in Figures 2 and 3.

T tests were performed on all deviation measures to determine if deviation was significantly different from zero, i.e., if predicted and observed JNDs were significantly different from one another (Figure S2). To compare deviation from the crossmodal discrimination hypothesis across conditions (Figures 2 and S2A, S2B, S2E, S2F) we performed a three-way ANOVA with factors fixation (head-fixed versus scene-fixed), trial type (sequential versus simultaneous) and standard (visual versus vestibular). Paired t tests were performed to evaluate whether deviation from both ML and conflict detection predictions (Figures 3 and S2C, S2D, S2G, S2H) depended on fixation (i.e., head-fixed versus scene-fixed). A significance level of 0.05 was used for all statistical tests.

In addition, bootstrap analyses were performed (n = 400) using the Palamedes toolbox to establish confidence intervals on the JNDs for each subject and condition (Figure S1). These bootstrapped JNDs were then passed through Equations 3-5 to calculate confidence intervals on the deviation measures. These confidence intervals are reported alongside results of statistical tests in the caption of Figure S2.

Correction of Predicted JNDs for Crossmodal Discrimination

For most participants, the visual stimulus in the crossmodal comparison had to move faster than the vestibular one in order for them to be perceived as matching (Figure S3), consistent with prior reports of visual-to-vestibular gains that are greater than 1 [15, 16, 44]. Higher matching visual speed resulted in higher variability, lower matching vestibular speed in lower variability (PSE and JND were correlated, Figures S3C and S3F). Based on the observation that noise on the single-cue estimate also scales with the magnitude of the stimulus (i.e., following Weber's law), crossmodal discrimination predictions were corrected based on the mean visual-to-vestibular gain g for each individual subject:

$$g = \left(\frac{stand}{PSE_{visstand}} + \frac{PSE_{veststand}}{stand}\right)\Big/ 2 \qquad \text{(Equation 6)}$$

where $stand$ is the 10°/s velocity of the standard stimulus and PSEs are for the sub-conditions where either the visual ($visstand$) or the vestibular ($veststand$) stimulus was the standard. These gain values are reported in the caption of Figure S3. When calculating crossmodal predictions (Equation 3), variability of the comparison stimulus was scaled according to this gain factor. Specifically, for the vestibular standard sub-condition, variability of the visual estimate was multiplied by this gain factor (faster matching visual speed → greater variability), and for the visual standard sub-condition, variability of the vestibular estimate was divided by this gain factor (slower matching vestibular speed → less variability). This correction did not significantly impact the resulting predictions or statistical analyses (ANOVA without correction: factor fixation, F = 3.89, p = 0.05; factor condition, F = 9.26, p < 0.01).

Eye Movement Recording and Analysis

It is not uncommon to forgo recording of eye movements when studying effects of fixation under minimally demanding conditions [45–48]. Eye movements were not recorded during data collection in this experiment because participants wore stereo goggles that were incompatible with our eye tracking equipment. However, the fixation task, whether head-fixed or scene-fixed, was not demanding (e.g., standard peak velocity 10°/s), and we are confident that subjects obeyed instructions and were able to fixate as instructed. To support this claim, after completion of the experiment, we removed the stereo goggles and recorded eye movements during an average of 22 trials with scene-fixed fixation and 28 trials with head-fixed fixation in the combined condition for four participants using an infrared eye tracking system (EyeSeeCam) at a sampling frequency of 60 Hz. Eye tracking data was filtered

by applying a fourth-order Butterworth filter with a low-pass cut-off frequency of 5 Hz. Eye velocity was calculated by numerical differentiation of the position data. Resulting mean eye velocity traces for the standard movement (peak velocity of 10°/s) are shown in Figure S4. Despite some lag, the traces clearly show that subjects were following instructions such that eye movements and visual optic flow signals were successfully manipulated across fixation conditions as intended.

DATA AND SOFTWARE AVAILABILITY

Data are available on request. Please contact the lead author.

Supplemental Information

Visual-Vestibular Conflict Detection

Depends on Fixation

Isabelle T. Garzorz and Paul R. MacNeilage
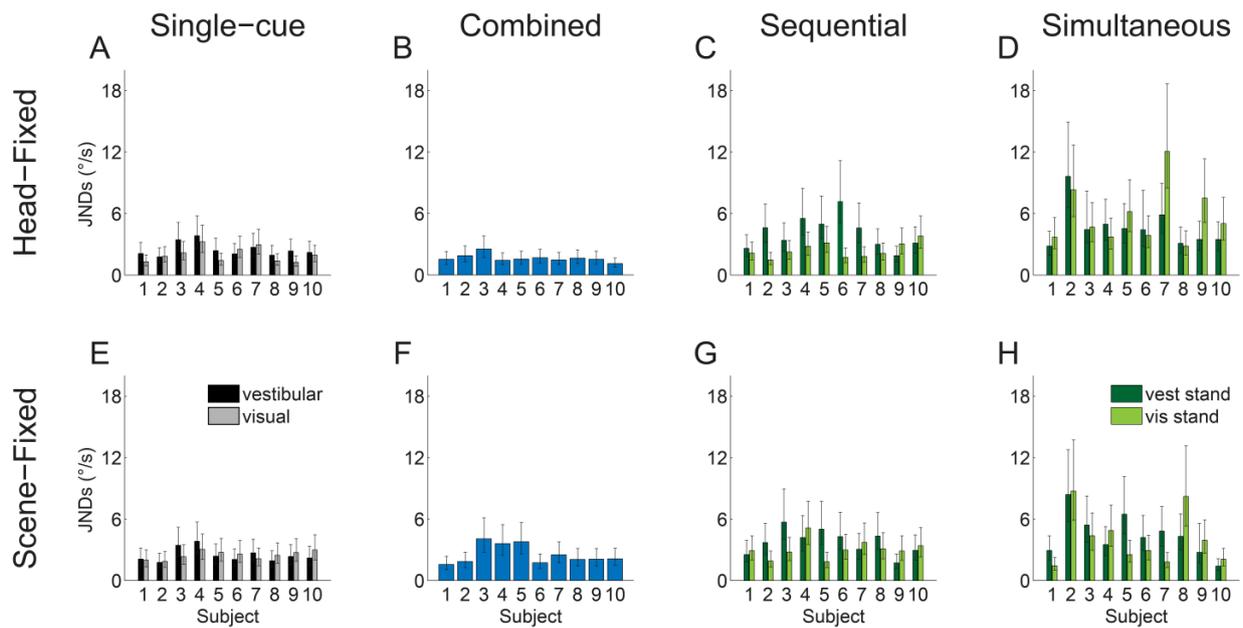
Figure S1. JNDs in all conditions for all subjects, Related to Figures 2 and 3 and STAR Methods.

(A-H) Observed JNDs for all subjects during single-cue (first column, vestibular (black) and visual (gray)), combined (second column), sequential (third column) and simultaneous (right column) conditions with either head-fixed (top row) or scene-fixed (bottom row) fixation. The last two columns depict the JNDs for both vestibular standard (dark green) and visual standard (light green) staircases. Error bars indicate bootstrapped 95% confidence intervals.
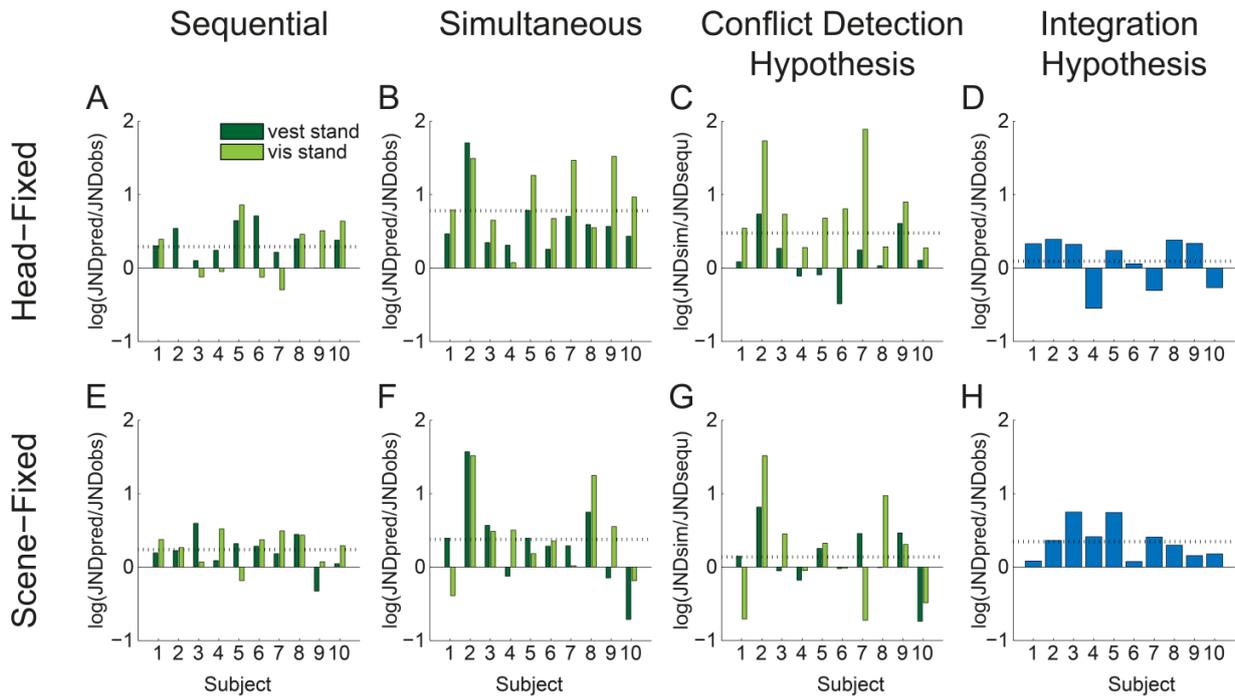
Figure S2. Deviation of observed from predicted JNDs, Related to Figures 2 and 3 and STAR Methods

Bars indicate deviation of observed from predicted JNDs, evaluated by their log ratio (Eq.5). All statistical evaluation was performed on these measures. Mean deviation value for each panel is illustrated by the black dotted line. T tests were used to asses if deviations were significantly different from zero. (A,B,E,F) Evaluation of the crossmodal discrimination model; deviation from zero was significant in all conditions (A: $\mu = 0.29$, $p < 0.001$; B: $\mu = 0.78$, $p < 0.001$; E: $\mu = 0.24$, $p < 0.001$; F: $\mu = 0.38$, $p < 0.01$). (C,G) Evaluation of Conflict Detection Hypothesis; deviation from zero was significant in the head-fixed condition (C: $\mu = 0.48$, $p < 0.01$) but not in the scene-fixed condition (G: $\mu = 0.14$, $p = 0.30$). Deviation was significantly larger during head-fixed (C) compared to scene-fixed (G) fixation (paired t test, $p < 0.05$). (D,H) Evaluation of integration hypothesis; deviation from zero was significant in the scene-fixed condition (H: $\mu = 0.35$, $p < 0.01$) but not in the head-fixed condition (D: $\mu = 0.09$, $p = 0.41$). While deviation was larger during scene-fixed (H) compared to head-fixed (D) fixation, the difference fell short of statistical significance (paired t-test, $p = 0.08$). Bootstrapped JNDs for each subject and condition were used to calculate confidence intervals on the deviation measures (95% CI [low up]; A: [0.20 0.33], B: [0.68 0.81], C: [0.38 0.55], D: [-0.01 0.17], E: [0.15 0.28], F: [0.29 0.41]), G: [0.04 0.23], H: [0.25 0.42]). These CIs support the statistical testing reported above. CIs for panels A, B, E, and F (crossmodal discrimination) do not include zero. CIs for panels C and G (Conflict Detection Hypothesis) do not overlap. CIs for panels D and H (Integration) do not overlap.
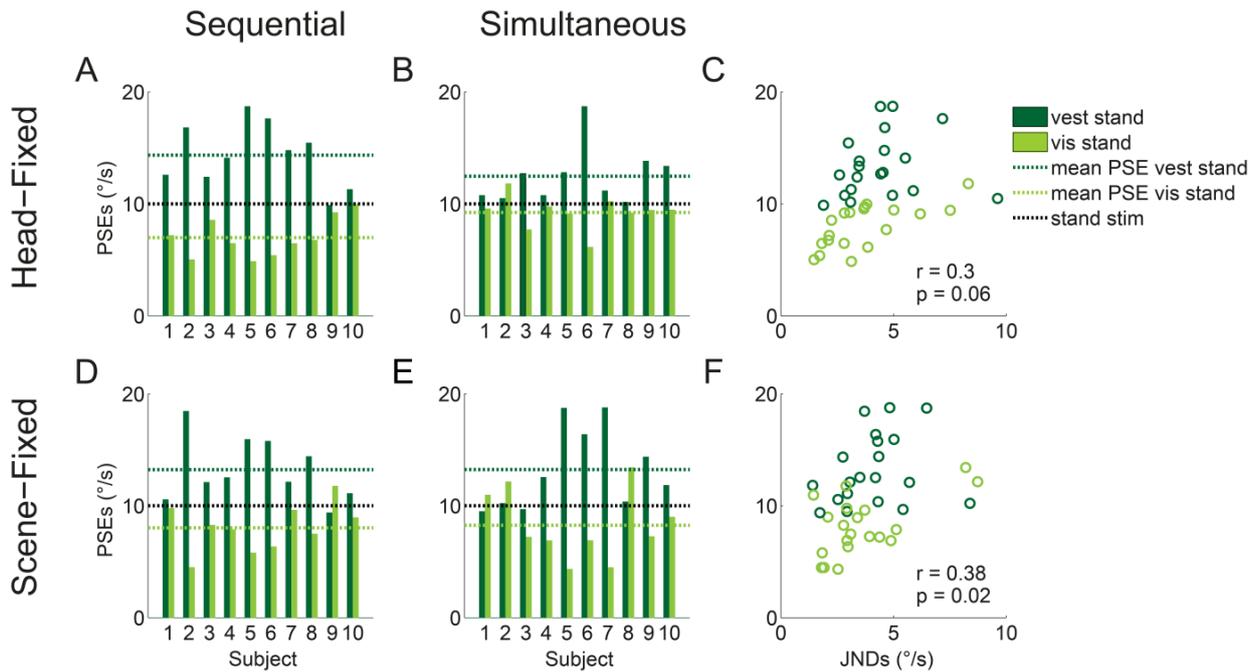
Figure S3. Weber's law, Related to Figure 2 and STAR Methods.

(A,B,D,E) Bars show PSEs for all subjects in sequential (first column) and simultaneous (second column) conditions with head-fixed (top row) and scene-fixed (bottom row) fixation for vestibular standard (dark green) and visual standard (light green) staircases. Black dotted lines indicate the size of the standard stimulus. Mean PSEs are shown by the green dotted lines. These mean values for the vestibular-standard staircases (dark green), and corresponding (geometric) means for gain (Eq. 6) are: (A) 14.36°/s (g=1.41), (B) 12.47°/s (g=1.23), (D) 13.23°/s (g=1.30), (E) 13.23°/s (g=1.28). These values for the visual-standard conditions (light green) are: (A) 6.98°/s (g=1.47), (B) 9.23°/s (g=1.10) (D) 8.03°/s (g=1.29), (E) 8.25°/s (g=1.29). (C and F) Similar to the current observations, prior studies have reported that visually induced self-motion must be faster than the physical head motion in order for these to be perceived as matching (i.e. visual-vestibular gain >1) [S1-S3]. Correlations between JNDs and PSEs for all subjects during head-fixed (C) and scene-fixed (F) fixation suggest that Weber's law holds.
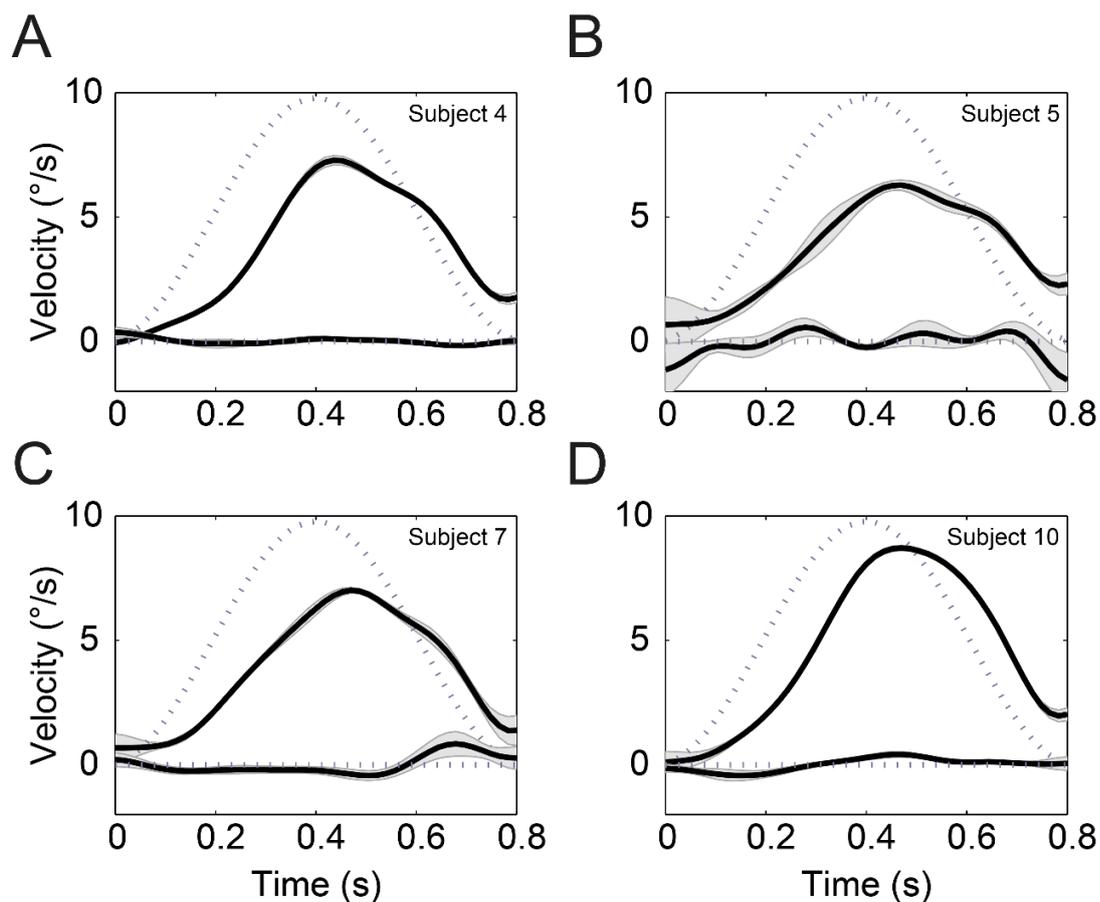
Figure S4. Example eye traces, Related to STAR Methods.

(A)-(D) Example eye velocity data for four subjects. Thick black lines show mean eye velocity for both head-fixed (flat curve) and scene-fixed fixations for both leftward and rightward rotation. Grey shaded area represents the standard error of the mean. The grey dotted line depicts fixation point velocity.

Chapter 3

Insufficient Compensation for Self-motion During Perception of Object Speed: The Vestibular Aubert-Fleischl Phenomenon

Author contributions

I.T.G. and P.R.M. designed the study
I.T.G. collected the data
I.T.G. and P.R.M. analyzed the data
I.T.G., P.R.M., T.C.F, and M.O.E. wrote the paper

Insufficient Compensation for Self-motion During Perception of Object Speed: The Vestibular Aubert-Fleischl Phenomenon

Garzorz, Isabelle T.[1,2], Freeman, Tom C.A.[3], Ernst, Marc O.[4], MacNeilage, Paul R.[1,2,5]

1 German Center for Vertigo (DSGZ), University Hospital of Munich, Ludwig Maximilian University, 81377 Munich, Germany

2 Graduate School of Systemic Neurosciences (GSN), Ludwig Maximilian University, 82152 Planegg-Martinsried, Germany

3 School of Psychology, Cardiff University, Cardiff, CF10 3AT, United Kingdom

4 Applied Cognitive Psychology, Faculty for Computer Science, Engineering, and Psychology, Ulm University, 89081 Ulm, Germany

5 Present address: Department of Psychology, Cognitive and Brain Sciences, University of Nevada, Reno, NV 89557, USA

Author Note

Abstract

To estimate object speed with respect to the self, retinal signals must be summed with extra-retinal signals that encode the speed of eye and head movement. Prior work has shown that differences in perceptual estimates of object speed based on retinal and oculomotor signals lead to biased percepts such as the Aubert-Fleischl phenomenon (AF), in which moving targets appear slower when pursued. During whole-body movement, additional extra-retinal signals, such as those from the vestibular system, may be used to transform object speed estimates from a head-centered to a body-centered reference frame. Here we demonstrate that whole-body pursuit in the form of passive yaw rotation, which stimulates the semi-circular canals of the vestibular system, leads to a slowing of perceived object speed similar to the classic oculomotor AF. We find that the magnitude of the vestibular and oculomotor AF is comparable across a range of speeds, despite the different types of input signal involved. This covariation might hint at a common modality-independent mechanism underlying the AF in both cases.

Keywords: Aubert-Fleischl, Speed Perception, Vestibular, Oculomotor, Retinal

Introduction

Accurately estimating the speed of moving objects and self-motion with respect to the world is an important task for the nervous system which supports safe locomotion and interaction with the environment. This crucial ability depends on transformations between different reference frames, including retinal, head, and world coordinate systems. Extra-retinal signals about eye movements must be taken into account to transform retinal signals into a head-centric reference frame and estimate object speed relative to the head (von Holst & Mittelstaedt, 1950). Similarly, when the head moves, vestibular and neck-muscle information must be used to transform signals from head-centred to body-centred reference frames.

The estimates resulting from these transformations can be biased, yielding phenomena such as the Filehne illusion in which stationary objects appear to move during smooth pursuit eye movements (Filehne, 1922). The Filehne illusion is thought to arise from differences in the perceptual estimates of retinal and oculomotor speed (Freeman, Champion, & Warren, 2010; Furman & Gur, 2012; Haarmeier & Thier, 1996; Souman, Hooge, & Wertheim, 2005; Souman, Hooge, & Wertheim, 2006; Wertheim, 1981, 1987). This mismatch leads also to a related bias in perceived speed called the Aubert-Fleischl (AF) phenomenon, in which pursued targets are perceived to move more slowly than non-pursued ones (Aubert, 1886). The underestimation of object speed during oculomotor pursuit was originally attributed to an erroneous estimate of eye velocity via an extra-retinal signal, while retinal motion estimates were hypothesized to be veridical (Mack & Herman, 1973; Raymond, Shapiro, & Rose, 1984). Subsequent work cast doubt on this hypothesis by showing that the AF and Filehne illusion can be reversed, i.e. retinal velocity may become underestimated compared to oculomotor velocity, depending on the spatial frequency of the stimulus (Freeman & Banks, 1998; Wertheim, 1987). In other words, the strength and direction of the AF phenomenon depends on the relationship between retinal and oculomotor speed estimates, with the former being a function of the stimulus. If both signals are linearly related to speed, then their ratio captures the behaviour of phenomena such as the AF and Filehne illusion (Freeman, 2001; Furman & Gur, 2012; Souman et al., 2006).

In addition to eye movement, head and body movement also lead to motion at the retina, so the question arises how other reference frame transformations, e.g. into body or world coordinates, influence the perception of object speed. When the head moves, retinocentric estimates can be transformed into a head-centric coordinate system using additional extra-retinal cues, i.e. signals from the vestibular system. These carry information about linear and angular accelerations and thereby allow for estimation of head motion.

In summary, estimation of object motion when the observers move their eyes, head, and body can be recovered by integrating the speed of the object on the retina, the speed of the eyes with respect to the head, and the movement of the head in space. Previous studies have focused on the perception of object speed in experimental conditions where the head was held still while the eyes were either fixating a stationary target or pursuing a moving target (Dichgans, Wist, Diener, & Brandt, 1975; Freeman, 2001; Freeman & Banks, 1998; Freeman et al., 2010; Powell, Meredith, McMillin, & Freeman, 2016; Raymond et al., 1984; Souman et al., 2006; Wertheim, 1987). The present study, in contrast, investigates the impact of vestibular signals on the perception of perceived object motion during passive whole-body rotations.

There are some previous studies that have investigated the impact of vestibular signals on perception of object motion (e.g. Dyde & Harris, 2008; Jaekl, Jenkin, & Harris, 2005). However, these studies were not conducted under conditions that allowed for direct comparison between vestibular and oculomotor compensation which was our intention here. To preview our findings, we observe a phenomenon analogous to the classical (oculomotor) AF, in which objects that are pursued with a whole-body rotation, appear to move more slowly than non-pursued objects. We call this the vestibular AF. Our results show that the oculomotor and vestibular AF effects are similar in magnitude and can be described by a simple linear model of the signals involved.

Methods

Participants

Nine observes (four male, five female) with normal or corrected-to-normal vision participated in the experiment. They were aged 21 to 27 years (average age = 23.2 years) and had no history of neurological, visual, or vestibular disorders. Participants gave informed consent before taking part in the study. All but two were naïve to the purpose of the study. The experiment was approved by the ethics committee of the University Hospital of Munich and conducted in accordance with the ethical standards of the Declaration of Helsinki.

Equipment

The experiment was conducted in a virtual reality setup consisting of a 6-degree-of-freedom hexapod motion platform (Moog© 6DOF2000E) and a stereo screen (JVC© - GD-463D10, Refresh rate: 60 Hz) with dimensions 101.8 cm x 57.3 cm. Participants were seated in a racing seat mounted on top of the platform at a viewing distance of 47 cm. They wore custom-made welding goggles, restricting the field of view (FOV) to prevent them from seeing the edges of the screen (effective FOV: ~ 90° x 60° visual angle). The goggles consisted of a circular polarizing filter enabling display of stereoscopic images and a blurring film to blend neighbouring pixels thereby weakening accommodative cues to screen distance. The visual scene rendered via OpenGL© and Psychtoolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) consisted of a volume of randomly placed red spheres (radius = 0.4 cm) at a density of 0.004 spheres/cm³ on a black background. To avoid spheres from obstructing the fixation point during rotation, only spheres located 4 cm above and below the white fixation point were visible. In addition, spheres located at a radial distance closer than 49 cm and further away than 69 cm from the centre of the head were not visible. Thus, the visual scene only contained spheres within the annulus of 10 cm in front of and behind screen depth (Fig. 1).

During the experiment, the head was kept fixated with two padded restraints at the temples to minimize any movement of the head. Participants wore noise-cancelling headphones through which white noise was played to mask the sounds of the active

platform. Responses in the experiment were collected using a response box with two buttons.
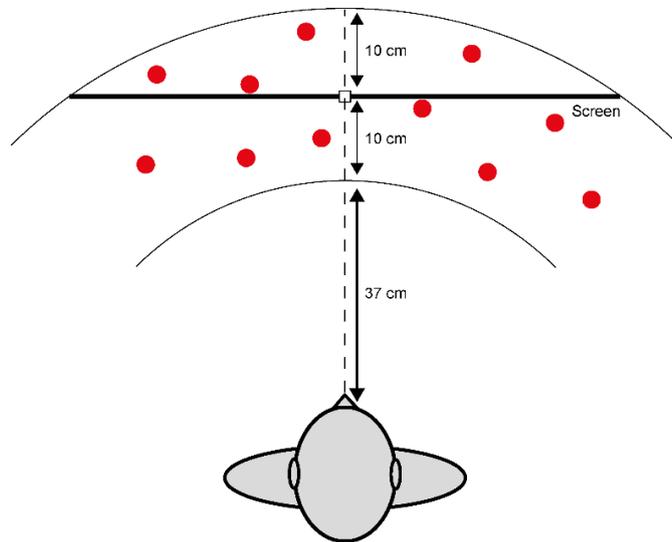

Figure 1. *Experimental Setup.* Participants seated at a viewing distance of 47 cm watched a 3-dimensional volume of randomly placed red spheres on a black background. Only spheres within the annulus of 10 cm in front of and behind screen depth were visible.

Experimental Procedures and Conditions

Participants performed a two-interval forced-choice (2IFC) task in which they indicated with a button press in which of the two intervals the annulus rotated more (displacement), faster (velocity), and stronger (acceleration) in the world (Fig. 2). Displacement, velocity and acceleration scaled together, and participants were explicitly informed that they could make judgements based on any of these. Anecdotally, subjects report having an intuition about the judgment without knowing whether they were judging displacement, velocity, or acceleration. For simplicity, we will only refer to velocity or speed in the following.

Depending on condition, different cues were available to estimate annulus speed. In retinal motion intervals (R), participants were instructed to fixate a white head-fixed fixation point at screen depth while the rotation of the annulus of red spheres induced optic flow. This condition nulled eye movements and maximized retinal motion. In the eye pursuit intervals (Fig. 2, left upper panel, E) the fixation point was moving at the same speed as the annulus. Here, eye movements were maximized and optic flow minimized.

In the vestibular pursuit intervals (Fig. 2, left lower panel, V), participants were passively rotated on the motion platform around the centre of the head while the fixation point and annulus were stationary on the screen. Here, the judgement about how fast the annulus had rotated in the world was dominated by vestibular signals.

Every trial consisted of a standard stimulus interval (either oculomotor or vestibular, Fig. 2, left column) and a retinal-motion comparison interval (Fig. 2, right column). The speed of the latter was adjusted from trial to trial according to a staircase procedure (Palamedes toolbox, Prins and Kingdom (2009)) consisting of two interleaved staircases (2up-1down, 1up-2down) with a step size of 0.2 natural logarithms. The order of standard and comparison stimulus was randomized. Physical and visually simulated yaw rotation was about a vertical axis passing through the midpoint of the inter-aural axis of the head. Each movement had a raised cosine velocity profile:

$$v(t) = \frac{D}{T}\left[1 - cos\left(\frac{2\pi t}{T}\right)\right]$$

51

with a duration (T) of 1 s, such that displacement (D), velocity (V) and acceleration scaled together.

The experiment consisted of two conditions that were run separately. In the (classical) oculomotor AF condition (E-R), the eye pursuit interval (E) was the standard and the retinal motion interval (R) the comparison. In the vestibular perceived speed condition (V-R), the oculomotor pursuit was replaced by a passive vestibular whole-body pursuit (V) while the retinal motion interval (R) still served as the comparison. The direction of rotation (left vs. right) across trials was randomized but both rotations within each trial were in the same direction and were separated in time by a pause of 0.5 s. At the end of each trial, a tone of 0.2 s served as a signal for participants to respond using the button box.

Both conditions were tested at three speeds (standard peak velocity V_{max} of 6, 12 or 18°/s) with 75 trials per staircase, condition and speed level, resulting in 900 trials in total. Ten practice trials with verbal feedback at the start of each condition ensured that participants had understood the task correctly. Each condition was divided into blocks of 75 trials with short breaks in-between to maintain participants' attentiveness.

Standard

Comparison

Oculomotor

Eye pursuit interval (E)

$R = 0$
$E \neq 0$
$V = 0$

Retinal motion interval (R)

$R \neq 0$
$E = 0$
$V = 0$

Vestibular

Vestibular pursuit interval (V)

$R = 0$
$E = 0$
$V \neq 0$

Retinal motion interval (R)

$R \neq 0$
$E = 0$
$V = 0$
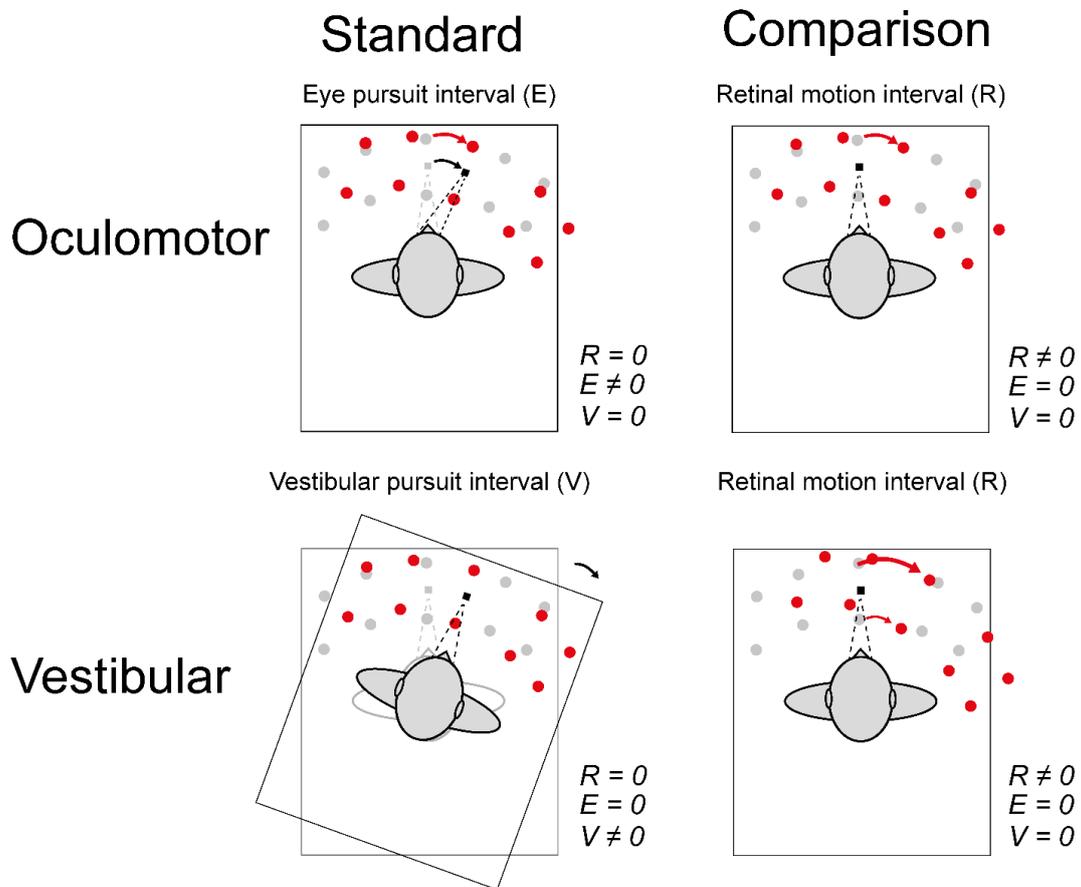
Figure 2. *Experimental Conditions.* On each trial, participants experienced a standard pursuit interval (left) and a comparison retinal motion interval (right) and indicated which rotation (around the center of the head) was larger, faster, and stronger (2IFC). The order of standard and comparison intervals was randomized across trials. The annulus elements are shown in red, the fixation point in black, the dashed lines indicate oculomotor fixation, and the rectangular outline indicates the motion platform. In the oculomotor condition (top), the pursuit was executed by following the fixation point with the eyes (dashed lines). In the vestibular condition (bottom), the pursuit was executed by maintaining fixation while the platform and annulus rotated together. These conditions were run in separate blocks. Equations to the lower right of each panel indicate if the stimulus velocity (R: retinal motion, E: eye pursuit, V: vestibular pursuit) was held constant (=0) or varied ($\neq 0$).

Statistical Analyses

Using MATLAB (version R2010b) and the Palamedes toolbox (Prins & Kingdom, 2009), each participant's data for every condition was fit by cumulative Gaussians using a GLM with a probit link and a Maximum Likelihood fitting routine. The mean of the cumulative Gaussian fit was taken as the point of subjective equality (PSE), i.e. the stimulus intensity that elicits 50% "comparison faster than standard" responses. Significant deviation of the PSE from the standard was interpreted as a bias, i.e. incomplete compensation for eye or body motion. The standard deviation of the cumulative Gaussian fit was taken as the just-noticeable difference (JND), i.e. the change in stimulus intensity relative to the PSE that results in 84% comparison faster judgments.

To estimate the size of the AF in the two conditions, we calculated the ratio of retinal speed at the PSE to oculomotor or vestibular speed of the standard stimulus, i.e. R/E or R/V respectively. We call this the 'gain ratio' in keeping with definitions in the literature (see Furman and Gur (2012) for review). If the perceived speed of retinal motion is greater than either the oculomotor or the vestibular standard, then the comparison interval would need to be slowed down to achieve the speed-match at the PSE. In this case, R/E or R/V < 1, as in the classic AF. If, on the other hand, oculomotor or vestibular standards appeared faster, then R/E or R/V > 1.

Results

The mean gain ratios shown in Fig. 3B suggest similar AF effects in the oculomotor and vestibular conditions. In fact, gain ratios from both conditions are significantly correlated ($rho = 0.65$, $p < 0.001$) and the confidence intervals of slope and intercept of a total-least squares fit contain 1 and 0, respectively (see Fig. 3A), supporting the high similarity between both AF effects.

A two-way repeated-measures ANOVA did not reveal any main effects of self-motion condition ($F(1,8) = 1.19$, $p = 0.31$), nor was there a main effect of peak standard speed ($F(2,16) = 2.67$, $p = 0.10$), which suggests that the AF effects can be described reasonably well by a linear model in which the signals estimating retinal, oculomotor and vestibular speed depend on the relevant physical speed of movement times some fixed gain factor (Freeman, 2001; Furman & Gur, 2012; Souman et al., 2006). In support of this hypothesis, the ANOVA did not reveal any interaction between factors ($F(2,16) = 1.01$, $p = 0.39$).

Since there was no significant main effect of speed, gain ratios from all speed levels were pooled to test both AF effects. A one-sample t-test showed the mean gain ratio in the oculomotor condition ($M = 0.77$, $SD = 0.24$) was significantly less than 1 ($t(26) = -4.90$, $p < 0.001$), which is consistent with the classic AF. Another one-sample t-test revealed the mean gain ratio in the vestibular condition ($M = 0.70$, $SD = 0.30$) was also significantly less than 1 ($t(26) = -5.28$, $p < 0.001$), which supports our hypothesis of a vestibular analogue of the classic AF.

Weber fractions, i.e. JNDs expressed as a fraction of the standard speed (Fig. 3C), for all conditions and speeds are highly correlated ($rho = 0.82$, $p < .001$). A two-way repeated-measures ANOVA did not show any main effect of self-motion condition ($F(1,8) = 0.95$, $p = 0.36$) but a significant main effect of speed ($F(2,16) = 52.4$, $p < 0.001$), with smaller Weber fractions at higher peak standard speeds, similar to previous findings by Freeman et al. (2010). The ANOVA did not show any interaction between speed and condition ($F(2,16) = 1.11$, $p = 0.35$).
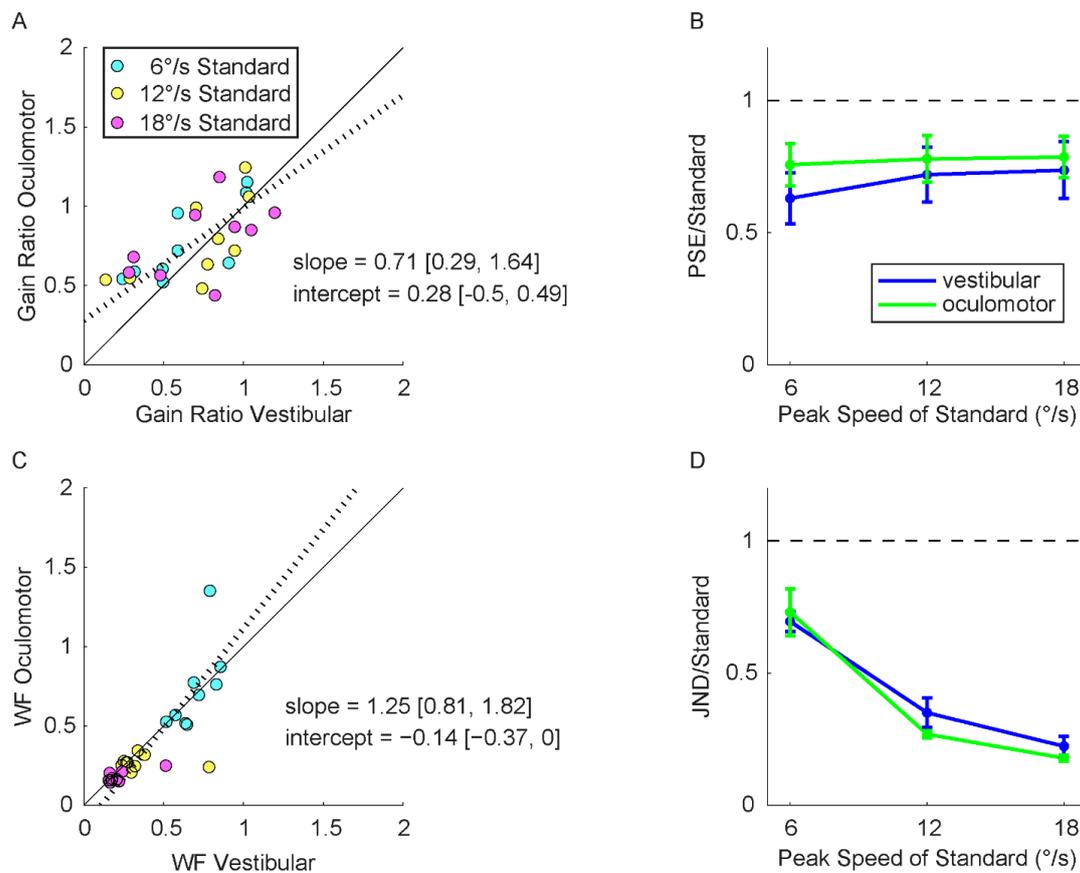
Figure 3. *Results* (A) Oculomotor and vestibular gain ratios of all participants for three peak speeds of standard (6, 12, and 18°/s); the black line represents unity slope, the black dotted line indicates a total least squares fit with confidence intervals of slope (0.71) and intercept (0.28) in brackets. (B) Mean ratios of retinal to oculomotor (R/E, green) and retinal to vestibular (R/V, blue) speed at PSE for the three peak speeds of standard are shown. Gain ratios smaller than 1 represent an underestimation of oculomotor or pursuit speed with respect to retinal speed. Error bars indicate the standard error of the mean. (C) Weber fractions (expressed as the ratio of JND to peak speed of standard) for all participants; the black line represents unity slope, the black dotted line indicates a total least squares fit with confidence intervals of slope (1.25) and intercept (-0.14) in brackets. (D) Mean Weber fractions for oculomotor (green) and vestibular (blue) conditions for the three peak speeds of standard. Error bars indicate the standard error of the mean.

56

Discussion

Targets are typically perceived to move more slowly when pursued by eye (Aubert, 1886; Dichgans et al., 1975; Freeman, 2001; Freeman & Banks, 1998; Freeman et al., 2010; Raymond et al., 1984; Souman et al., 2006). This so-called Aubert-Fleischl (AF) phenomenon reveals inconsistencies between signals encoding image motion and those encoding eye movement. Here, we found the same was true for extra-retinal signals that originate from the vestibular system. Objects that are physically pursued via whole-body rotations appear to move more slowly than non-pursued objects. The ratio of vestibular pursuit speed to the perceptually equivalent retinal speed during the non-pursuit interval was significantly smaller than 1, demonstrating a vestibular analogue of the AF.

Vestibular and Oculomotor AF Compared

It is remarkable that similar gain ratios were observed in the oculomotor and vestibular conditions (Fig. 3A and B). In both cases, the estimate of object speed based on the extra-retinal signal, whether oculomotor or vestibular, was reduced relative to the estimate of speed based on retinal motion with stationary eyes and head. We also found that the gain ratios did not depend on speed. Thus, a linear model in which the speed estimates defined by the underlying signals are a fixed fraction of speed (e.g. Freeman, 2001; Furman & Gur, 2012; Souman et al., 2006) can also characterize cases in which extra-retinal signals originate from the vestibular system.

We also found that the JNDs, converted into Weber fractions, become smaller at higher peak standard speeds, akin to an inverse power law. This is consistent with previous reports, which show that Weber fractions for speed discrimination decline more rapidly for slow compared to medium speeds (De Bruyn & Orban, 1988), irrespective of whether stimuli are fixated or pursued (Freeman et al., 2010). As with the linear speed estimates, the noises associated with oculomotor and vestibular signals appear to be remarkably similar.

In the two remaining sections of the discussion we relate our newly found vestibular AF effect to research on self-motion perception and we discuss a possible common explanation of both oculomotor and vestibular AF.

57

The AF during self-motion

While previous research investigating the AF effect primarily focused on the perception of object motion in stationary subjects, our study broadened the scope of the AF by asking observers to judge object motion during passive self-motion.

In contrast to previous studies using very sparse stimuli with a small field of view (Brenner & van den Berg, 1994; Freeman & Banks, 1998; Freeman et al., 2010; Raymond et al., 1984; Souman et al., 2005; Turano & Heidenreich, 1999), observers in our experiment viewed a stereoscopic visual scene with a relatively large field of view. The current results reveal a systematic underestimation of object speed, not only for stationary observers (classic AF), but also for passively moved observers (vestibular AF). The latter implies incomplete compensation for self-motion, which has previously been shown for estimation of object paths (Dokka, MacNeilage, DeAngelis, & Angelaki, 2015). A possible explanation may be an underestimation of self-motion based on vestibular signals during passive movement (Dyde & Harris, 2008) which appears to be reduced when movements are actively generated and efference copy and proprioception are available (Dupin & Wexler, 2013; Dyde & Harris, 2008; Genzel, Firzlaff, Wiegrebe, & MacNeilage, 2016). This suggests that the vestibular AF may be reduced when head movement is actively generated, but this idea remains to be tested.

Speed estimation under natural conditions

In real life, moving the eyes to pursue a moving object usually results in relative motion of the object compared to the stationary background. Also during self-motion, fixating a body-fixed target (e.g. looking at a spot on the car window) induces relative motion of the target compared to the background, i.e. optic flow with respect to the stationary fixation point (or spot on the window). Therefore, our prior experience generally predicts retinal motion of the stationary background in the direction opposite the pursuit movement while the target remains fixated. The effect of background motion on speed perception is demonstrated by the illusion of "induced motion", i.e. when we fixate an object, moving the background makes the object seem to move while the background appears stationary (Duncker, 1929).

In our experiments, the visual stimuli were kept as simple as possible to isolate pursuit from any influence of relative motion; there was no background motion. Here we raise the possibility that the AF could be partially due to the discrepancy between our prior knowledge about real world statistics of relative motion between fixated object and background and impoverished experimental stimuli where no relative motion signal is available. Revisiting previous studies that investigated the classical AF reveals a general lack of motion of the target relative to the background during pursuit intervals (Dichgans et al., 1975; Freeman et al., 2010; Raymond et al., 1984; Souman et al., 2006). Freeman et al. (2010) tried to rule out this explanation of the AF by testing whether relative motion has an influence on the perceived speed during stationary fixation. Speed perception was compared between conditions with and without a textured background, but no difference was observed. While this speaks against relative motion as an explanation of the AF effects, Brenner and van den Berg (1994) found that under natural conditions, perceived target velocity was accurate (i.e. no AF was observed) if the relative motion between a target and the background was maintained. Further experiments investigating the influence of background motion are needed to determine whether the AF can be partially ascribed to the lack of naturalistic relative motion signals during pursuit.

References

Aubert, H. (1886). Die Bewegungsempfindung. *Archiv für die gesamte Physiologie des Menschen und der Tiere, 39*(1), 347-370.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Brenner, E., & van den Berg, A. V. (1994). Judging object velocity during smooth pursuit eye movements. *Exp Brain Res, 99*(2), 316-324.

De Bruyn, B., & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Res, 28*(12), 1323-1335.

Dichgans, J., Wist, E., Diener, H. C., & Brandt, T. (1975). The Aubert-Fleischl phenomenon: a temporal frequency effect on perceived velocity in afferent motion perception. *Exp Brain Res, 23*(5), 529-533.

Dokka, K., MacNeilage, P. R., DeAngelis, G. C., & Angelaki, D. E. (2015). Multisensory self-motion compensation during object trajectory judgments. *Cereb Cortex, 25*(3), 619-630.

Duncker, K. (1929). Über induzierte Bewegung. *Psychologische Forschung, 12*(1), 180-259.

Dupin, L., & Wexler, M. (2013). Motion perception by a moving observer in a three-dimensional environment. *J Vis, 13*(2), 15.

Dyde, R. T., & Harris, L. R. (2008). The influence of retinal and extra-retinal motion cues on perceived object motion during self-motion. *J Vis, 8*(14), 5.1-10.

Filehne, W. (1922). Über das optische Wahrnehmen von Bewegungen. *Zeitschrift für Sinnesphysiologie, 53*, 134-144.

Freeman, T. C. (2001). Transducer models of head-centred motion perception. *Vision Res, 41*(21), 2741-2755.

Freeman, T. C., & Banks, M. S. (1998). Perceived head-centric speed is affected by both extra-retinal and retinal errors. *Vision Res, 38*(7), 941-945.

Freeman, T. C., Champion, R. A., & Warren, P. A. (2010). A Bayesian Model of Perceived Head-Centered Velocity during Smooth Pursuit Eye Movement. *Current Biology, 20*(8), 757-762.

Furman, M., & Gur, M. (2012). And yet it moves: perceptual illusions and neural mechanisms of pursuit compensation during smooth pursuit eye movements. *Neurosci Biobehav Rev, 36*(1), 143-151.

Genzel, D., Firzlaff, U., Wiegrebe, L., & MacNeilage, P. R. (2016). Dependence of auditory spatial updating on vestibular, proprioceptive, and efference copy signals. *J Neurophysiol, 116*(2), 765-775.

Haarmeier, T., & Thier, P. (1996). Modification of the filehne illusion by conditioning visual stimuli. *Vision Research, 36*(5), 741-750.

Jaekl, P. M., Jenkin, M. R., & Harris, L. R. (2005). Perceiving a stable world during active rotational and translational head movements. *Exp Brain Res, 163*(3), 388-399.

Kleiner, M., Brainard, D., & Pelli, D. (2007). *What's new in Psychtoolbox-3?*

Mack, A., & Herman, E. (1973). Position constancy during pursuit eye movement: an investigation of the Filehne illusion. *Q J Exp Psychol, 25*(1), 71-84.

Powell, G., Meredith, Z., McMillin, R., & Freeman, T. C. (2016). Bayesian Models of Individual Differences: Combining Autistic Traits and Sensory Thresholds to Predict Motion Perception. *Psychol Sci.*

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data.

Raymond, J. E., Shapiro, K. L., & Rose, D. J. (1984). Optokinetic backgrounds affect perceived velocity during ocular tracking. *Perception & Psychophysics, 36*(3), 221-224.

Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2005). Vertical object motion during horizontal ocular pursuit: compensation for eye movements increases with presentation duration. *Vision Research, 45*(7), 845-853.

Souman, J. L., Hooge, I. T. C., & Wertheim, A. H. (2006). Frame of reference transformations in motion perception during smooth pursuit eye movements. *Journal of Computational Neuroscience, 20*(1), 61-76.

Turano, K. A., & Heidenreich, S. M. (1999). Eye movements affect the perceived speed of visual motion. *Vision Research, 39*(6), 1177-1187.

61

von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip - (Wechselwirkungen Zwischen Zentralnervensystem Und Peripherie). *Naturwissenschaften, 37*, 464–476.

Wertheim, A. H. (1981). On the relativity of perceived motion. *Acta Psychologica, 48*(1–3), 97-110.

Wertheim, A. H. (1987). Retinal and extraretinal information in movement perception: how to invert the Filehne illusion. *Perception, 16*(3), 299-308.

Chapter 4

Sensitivity to Visual Gain Modulation in Head-Mounted Displays Depends on Fixation

Author contributions

M.M., I.T.G., E.F., and P.R.M. designed the study

M.M. collected the data

M.M. and P.R.M. analyzed the data

M.M., I.T.G., E.F., and P.R.M. wrote and revised the paper

Sensitivity to Visual Gain Modulation in Head-Mounted Displays Depends on Fixation

Matthew Moroz[a,*], Isabelle Garzorz[b], Eelke Folmer[c], Paul MacNeilage[a]

[a]*Department of Psychology, University of Nevada, Reno*
[b]*Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München*
[c]*Department of Computer Science, University of Nevada, Reno*

Abstract

A primary cause of simulator sickness in head-mounted displays (HMDs) is conflict between the visual scene displayed to the user and the visual scene expected by the brain when the user's head is in motion. Agreement between visual scene motion and head motion can be quantified based on their ratio which we refer to as visual gain. We suggest that it is useful to measure perceptual sensitivity to visual gain modulation in HMDs (i.e. deviation from gain=1) because conditions that minimize this sensitivity may prove less likely to elicit simulator sickness. In prior research, we measured sensitivity to visual gain modulation during slow, passive, full-body yaw rotations and observed that sensitivity was reduced when subjects fixated a head-fixed target compared with when they fixated a scene-fixed target. In the current study, we investigated whether this pattern of results persists when 1) movements are faster, active head turns, and 2) visual stimuli are presented on an HMD rather than on a monitor. Subjects wore an Oculus Rift CV1 HMD and viewed a 3D scene of white points on a black background. On each trial, subjects moved their head from a central position to face a 15° eccentric target. During the head movement they fixated a point that was either head-fixed or scene-fixed, depending on condition. They then reported if the gain applied to the visual scene motion was too fast or too slow. Gain on subsequent trials was modulated according to a staircase procedure to find the gain change that was just noticeable. Sensitivity to gain modulation during active head movement was reduced during head-fixed fixation, similar to what we observed during passive whole-body rotation. Additionally, conflict detection seems to be significantly improved with higher peak velocity of head rotation. We conclude that fixation of a head-fixed target is an effective way to reduce sensitivity to visual gain modulation in HMDs, and may also be an effective strategy to reduce susceptibility to simulator sickness.

Keywords: Motion sickness, vection, simulator sickness, visual-vestibular, conflict detection, passive vs. active, head rotation, gender difference, optic flow

1. Introduction

Head-mounted displays work by presenting a rendered view of a virtual environment that is updated based on the users head movement. Consequently, when the user turns the head, optic flow is presented on the HMD that is consistent with the users head movement. Disagreement between the head movement and the visual motion that is rendered is the most widely accepted explanation for the initiation of simulator sickness symptoms [1, 2, 3, 4, 5, 6, 7, 8]. However, user tolerance for this disagreement has not been extensively studied [9, 10]

---

*Corresponding author
Email addresses: `mmoroz@nevada.unr.edu` (Matthew Moroz), `isabelle.garzorz@campus.lmu.de` (Isabelle Garzorz), `efolmer@unr.edu` (Eelke Folmer), `pmacneilage@unr.edu` (Paul MacNeilage)

In prior work, we evaluated this tolerance by introducing conflicts between the physical head motion and the visual scene motion and measuring participants' ability to detect these conflicts [11]. We found that sensitivity to conflict depended on how participants moved their eyes, with the best sensitivity observed when participants moved their eyes to track scene-fixed targets. Head motion in these experiments was generated through passive full-body rotation with participants seated on a moving platform and with visual stimuli presented on a display mounted to the platform. Here we examine whether our previous findings generalize to the most common VR use-case, that is active turning of the head relative to the body with visual stimuli presented on an HMD.

Active and passive head movements differ in several important respects. Perhaps most importantly, during active head movements, additional non-visual cues to head motion are available, including proprioception and efference copies of motor commands. Head movements are also more variable from trial to trial; self-generated head movements lack the absolute consistency of velocity and duration afforded by the motion platform.

Using active head movements, we therefore address a slightly different set of questions in the current study. Similar to the previous study, we aimed to 1) measure sensitivity to conflict, 2) measure sensitivity to the visual optic flow stimulus, and 3) measure how these depend on fixation. However, in contrast with the previous study, we were not able to measure sensitivity to non-visual head motion cues during active movement, since we have not yet devised a method that would allow us to make this measurement. In addition, by varying the medium of display (HMD vs. 3D monitor) we are able to validate the usefulness of the HMD as a tool in psychophysical experimentation. We also sought to investigate whether an individuals sensitivity to conflict is predictive of their susceptibility to motion sickness, paying particular attention to possible effects of gender. While some studies claim that females are three times more susceptible [12, 9, 13], many others fail to replicate this finding [14, 15]. In this study, participant numbers were balanced with respect to gender allowing us to investigate whether gender is predictive of sensitivity to conflict, susceptibility to simulator sickness, or both.

Overall, the experimental design of the current study allows us to test the below hypotheses related to active head movement (Table 1).

Table 1: Hypotheses
H_1 : Conflict detection is improved during scene-fixed compared with head-fixed fixation.
H_2 : Fixation does not influence performance in a visual speed-discrimination task.
H_3 : Females exhibit more VR sickness than males.
H_4 : VR sickness ratings and conflict sensitivity are correlated.

2. Methods

2.1. Participants

Nineteen healthy participants (nine F, ten M), ranging in age from 20 to 41 (mean age = 26.5 years) completed the study. All possessed normal, or corrected-to-normal vision, unrestricted head/neck movement, and had no history of visual, or vestibular sensory disorders. All but two were naive to the hypothesis being tested; results of these subjects did not differ significantly from those of the other subjects. Three further participants began the study. Two were excluded due to an inability to consistently execute the trials, one dropped out due to time demands. The subjects in the current study were not the same as those that participated in our previously published study [11]. All procedures were approved by the Institutional Review Board of the University of Nevada, Reno and all participants provided informed consent.

2.2. Equipment

The experiment was conducted using an Oculus Rift (CV1) head mounted display (HMD) and Oculus-ready Alienware PC with NVIDIA GeForce GTX 1060 video card. Latency of this VR system was measured using the built in Oculus latency tester. Average latency measurements showed Flip to Mid-Photon as 7.23ms, Timewarp to Mid-Photon as 10.19ms, and App Tracking to Mid-Photon as 13.84ms. Participants were seated in a fixed, high-backed chair, to ensure head rotations originated from the neck while avoiding postural rotation. Participant responses were input using a standard keyboard. Textured Velcro tape was attached to the response keys to allow them to be identified haptically. All conditions were conducted in a quiet darkened room. Headphones in the HMD enabled delivery of

auditory beeps, tones and instructions that helped orient the subjects in the experiment. The virtual environment was programmed using C# within the Unity programming environment. Visual stimuli were rendered with a refresh rate of 90 Hz. Participants altered the HMDs interpupillary distance (IPD) themselves to a comfortable setting to account for individual differences in interpupillary distance. However, the IPD setting was not checked or recorded. We did not have the capability to record eye movements inside the HMD in order to verify that subjects were accurately following the fixation instructions. However, in a previous, similar study [11], we did record eye movements in a subset of subjects, and the data show that subjects successfully followed instructions (see Fig. S4 in [11]).
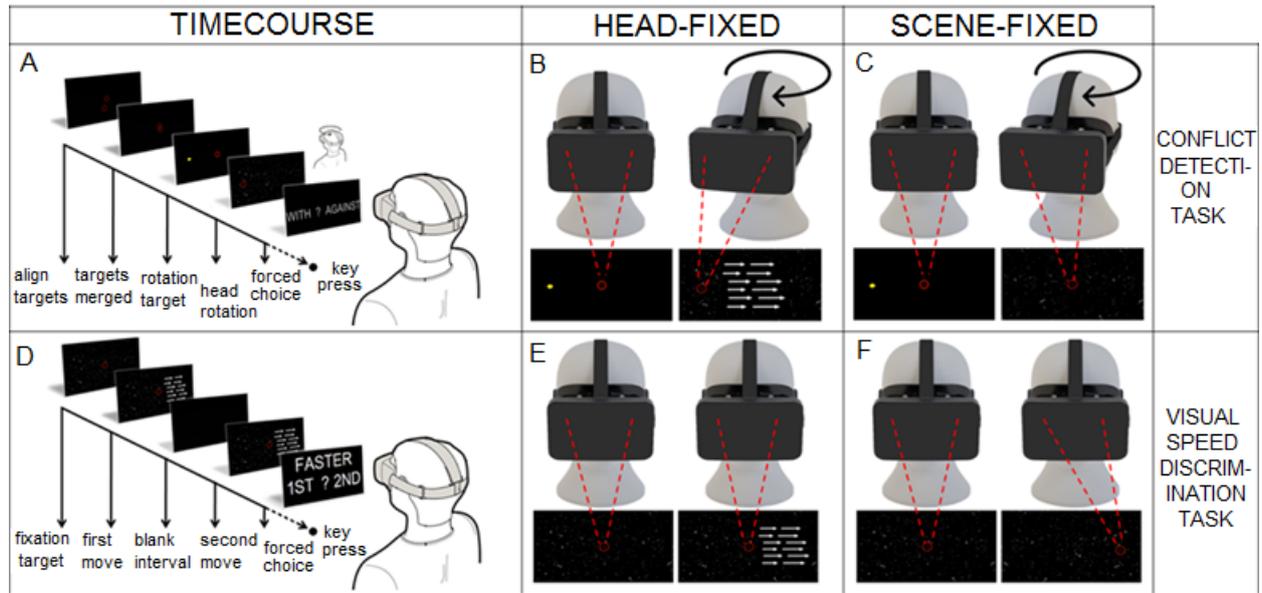

Figure 1. Experimental set up. Tasks focused on either visual speed discrimination (A,B,C) or conflict detection (D,E,F). A) illustrates the timecourse for the visual speed discrimination task. Participants focused on a red fixation target in the center of the scene and judged which of two yaw rotations of the background scene was faster (2AFC). Rotations were separated by a 0.5s inter-stimulus interval. B) and C) highlight the different conditions for this task. In B) participants eyes were fixed in their head while the visual stimulus moved across the retina. In C), the fixation target moved with the scene movement and thus by focusing on the target the image was stable on the retina. D) illustrates the timecourse for the conflict detection task. To ensure a consistent starting point, participants initiated each trial by aligning a red fixation target fixed in the scene with one attached to their virtual head. Once the targets were aligned, a yellow rotation target flashed (0.1s) at 15° eccentricity and participants rotated their heads to point towards this target. The visual gain was modulated to create conflict between the physical motion and displayed motion of the visual scene. Then participants answered whether the visual scene had moved too slowly or too quickly, i.e. with or against the direction of head rotation in world coordinates. E) and F) highlight the different conditions. In E), the fixation point moved with the head, so participants kept their eyes fixed in their heads. In F), the fixation point stayed fixed with respect to the scene, so participants had to counterrotate their eyes to maintain fixation.

2.3. Conflict Detection Task

During normal use of an HMD, the view on the visual environment is updated based on head movement, resulting in movement of the visual scene that is equal and opposite the head motion, consistent with a stationary, earth-fixed environment. However, in this task, the visual speed was modified to be either faster or slower than the head movement. Visual speed remained a function of the participant's head movement velocity and remained consistently slower or faster by a constant gain factor. Participants were asked to judge if the visual motion was too fast or too slow compared to their head movement. In other words, they were asked to detect the conflict between visual motion and head motion. The goal of the experiment was to measure the threshold for detection of this conflict.

The exact timecourse of an individual trial is illustrated in Fig. 1. Participants initiated each trial by adjusting their head position to the designated central position. This was achieved by making small head movements to align a head-fixed target (rendered 300cm in front of the cyclopean eye) with a central, scene-fixed target in an otherwise black, featureless environment. Once aligned, one of the red targets vanished, and a yellow target flashed (0.1s duration) within the participants left or right near-peripheral vision at an eccentricity of 15°. Then a randomly generated 3D
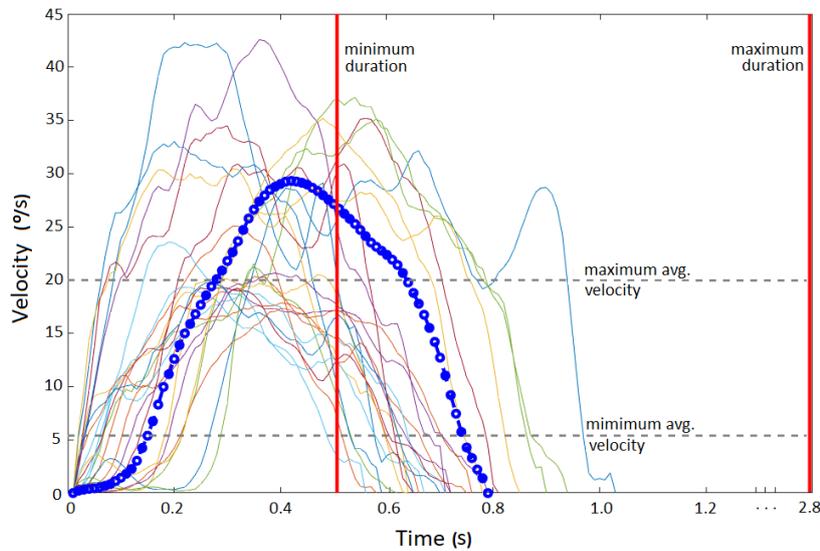
Figure 2. Example head movement traces. Thin colored lines indicate individual head movement traces. The thick blue line indicates the average over all of the individual traces. The dashed horizontal lines indicate the exclusion criteria for minimum and maximum average head velocities. The vertical solid red lines indicate the exclusion criteria for minimum and maximum duration of head movement.

starfield appeared and the participant performed a head rotation to point the head to where the yellow rotation target had flashed. The starfield consisted of 8000 randomly distributed white spheres (radius = 33cm), at a minimum distance of 1000cm, and average distance of 5000cm from the participants position. After the head rotation, the starfield disappeared, and an audible beep and text on the screen prompted the participant to indicate via keypress if they perceived the visual scene as moving with or against their head motion. If the visual scene moves too slowly, it is perceived to move with the head movement, and if it moves to quickly, it is perceived to move against the head movement. After the response, the next trial was initiated.

Because head movements were actively generated, care was taken to ensure consistency of head movements across trials. Specifically, the duration and speed of head movement was monitored and if duration was too short or long ($0.5s \leq dur \leq 2.78s$), or speed was too slow or fast ($5.4°/s \leq speed \leq 20°/s$), the trial was rejected and the participant received feedback in the form of a metal clang sound and a warning corresponding to the nature of the fault (e.g. too fast, or too slow). Parameters for acceptable head movements were identified in a pre-experiment pilot study. The movement profile was chosen as one that felt natural and was easy enough to repeat trial after trial. Example head movements from this pilot study along with exclusion criteria are illustrated in Figure 2

When a trial was rejected, to maintain consistency and encourage a smooth flow through the experiment, the starfield disappeared as before, and the participants were still presented with the same forced choice task. While responses on these trials were recorded, they did not affect the adaptive psychophysical procedure and were not used to calculate thresholds. There was an average of 7.8 unacceptable trials per block.

Each experimental block consisted of 150 trials with acceptable head movement profiles. The stimulus for a given trial was generated using two interleaved adaptive staircases (2up1down, 1up2down) of 75 trials each (Fig.3). New staircases were initiated in subsequent blocks.

There were two conditions, and participants completed three blocks for a total of 450 trials in each. The conditions were distinguished based on fixation behavior during the active head turn. In the head-fixed fixation condition, after head- and scene-fixed targets were aligned (Fig. 1), the scene-fixed target disappeared, and participants were left to fixate the target that moved with the head (Fig. 1). This behavior maximized retinal image motion (optic flow), while minimizing eye movement. In the scene-fixed fixation condition, the head-fixed target disappeared, and participants were left to fixate the scene-fixed target during the head movement (Fig. 1). This behavior maximized eye movement, while minimizing optic flow.

2.4. Visual Speed-discrimination Task

In this task, participants kept their heads still and were presented with full-field visual motion similar to the optic flow presented during the conflict detection task. Two consecutive motion intervals of 1.0 secs were presented on

each trial. One motion interval was the standard stimulus with an average speed of 12.8°/s and a peak velocity of 29.5°/s. Participants then responded which interval contained the faster movement (Fig. 1). Motions were in the same direction in both intervals, and participants completed 150 trails in each block. There were two conditions, head-fixed and scene-fixed, similar to those described above.

2.5. Training, order of experiments, and simulator sickness ratings

Experimentation was split over five sessions, each separated by 24 hours or more. These sessions always began with a thorough training session to ensure that participants understood the task. The training protocol explained and demonstrated each element of the required movement step-by-step. Participants performed each component part numerous times, before they finally executed 20 practice trials. The training was implemented to elicit consistent head-turning behavior within and across participants; training is often required when movements are generated actively rather than presented passively [16, 17]. After the training, participants completed a maximum of two experimental blocks of 150 trials each per session. All blocks for a given condition were completed in sequence, but the order of conditions was counterbalanced across subjects. To encourage focus/attentiveness within each session, mandatory breaks of 30 seconds were implemented after 50 and 100 trials. Longer breaks of 4 - 5 minutes were enforced between blocks. During this time participants removed the HMD and the light was turned on. Participants sat comfortably, stretched, or walked around before beginning the next block of trials. Each session was performed with an average running time of 25 minutes.

At the end of each condition, subjects provided a single rating of their feeling of simulator sickness on a 4-point scale, including ratings of none (0), slight (1), moderate (2), and severe (3). This scale is identical to that used in the more extensive Simulator Sickness Questionnaire (SSQ) [6] which asks participants to rate the severity of several specific symptoms.
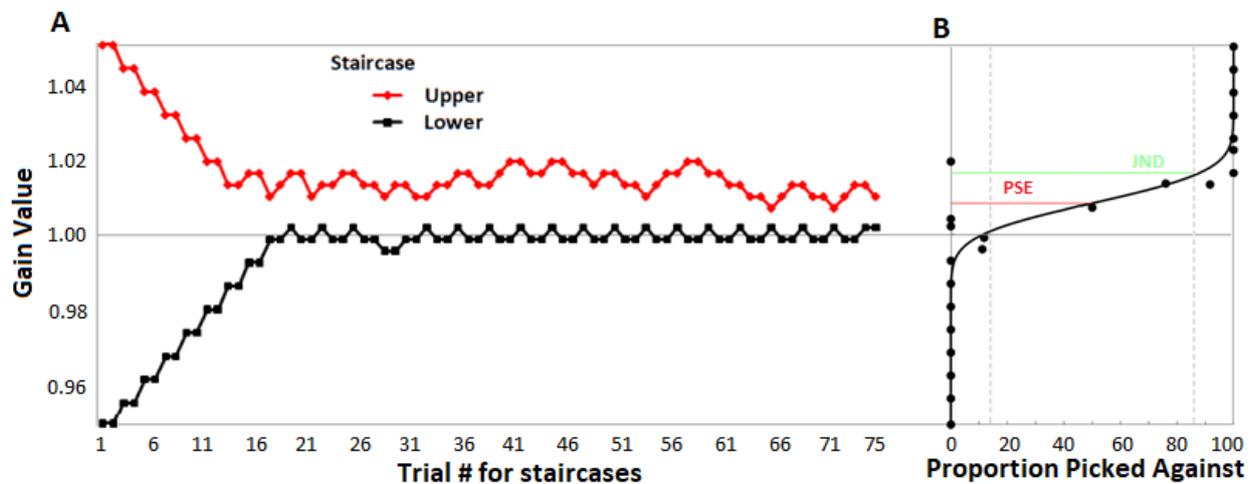
Figure 3. Psychophysics. Stimuli presented to a single participant in the conflict condition were modulated adaptively based on previous responses and a cumulative Gaussian psychometric function was fit to responses. A) illustrates two interleaved adaptive staircases (2 up 1 down and 2 down 1 up) which made up each block of 150 trials. B) illustrates how participants answers of with or against were used to fit a psychometric function. The parameters of interest are the mean or position of the curve (i.e. the point of subjective equality), and the standard deviation or steepness of the curve (i.e. the just noticeable difference).

2.6. Statistical Analysis

Analyses were conducted using Matlab R2016a together with the Palamedes toolbox package developed by Kingdom & Prins [18]. PAL PFML (Palamedes psychometric function: maximum likelihood) functions enabled us to fit a cumulative Gaussian to a participants response data (Fig.3). Parameters for lapse rate λ and guess rate γ were both set to zero. We refer to the mean parameter of the fit as the point of subjective equality (PSE), and standard deviation parameter as the just noticeable difference (JND).

Paired t-tests were carried out using participants JNDs across conditions. We additionally conducted one-sample t-tests to examine whether PSE values differed significantly from a gain value of 1 (1 represents the point where the visual stimuli has not been manipulated and matches the physical head motion). We analyzed VR sickness ratings using rank-based Wilcoxon tests to examine possible effects of Task (Conflict, Visual), Fixation (Head-fixed, Scene-fixed) and Gender (Male, Female). Additionally, the correlation between threshold and sickness ratings was analyzed using Spearmans rank correlation.

3. Results

3.1. Conflict Detection Task

Psychometric fits to the data from each individual subject and condition (e.g. Fig. 3B) provide a measure of the visual gain that is perceived to match the physical head motion (the PSE) as well as the threshold for the change in gain that leads to detection of conflict (the JND). Thresholds in the head-fixed condition (mean=0.032; SD=0.021) were significantly larger (t=2.828, p=0.011) than those in the scene-fixed condition (mean=0.022; SD=0.009), in support of hypothesis H1. These thresholds are plotted in Fig. 4A. The slope of the blue line is the log-average (i.e. geometric mean) of the ratio of the scene-fixed versus head-fixed JND (1.33) and the shaded area represents the standard error of this ratio (0.11). For comparison we have also plotted the thresholds measured in our previous study with passive head movements (Fig. 4C). As in the current study, thresholds in the head-fixed condition were larger than those in the scene-fixed condition. Again, the slope of the blue line is the log-average of the ratio of the scene-fixed versus head-fixed JND (1.27) and the shaded area represents the standard error of this ratio (0.22). The relationship between head-fixed and scene-fixed thresholds appears to be independent of whether head movement was active or passive.

The most conspicuous difference between results of current and former studies is that conflict detection thresholds were reduced by an order of magnitude when head movement was fast (peak vel.~29.5°/s) and active rather than slow (peak vel. 10°/s) and passive. Possible reasons for this increased sensitivity are explored in the discussion.

Results of the current study also differ from previous reports because the visual gains perceived as matching head motion (the PSEs) are much closer to unity in the present study. The gain perceived as matching in the head-fixed condition was 0.990 (0.018 SD) which is close to a gain of 1, but significantly different (t=-2,342, p=0.031). The gain perceived as matching in the scene-fixed condition was 0.995 (0.012 SD) which was not significantly different from 1. In contrast with the present results, prior studies consistently report that the visual gain perceived as matching is greater than 1 [19, 20, 21]. Methodological differences may be able to explain this difference in results (see Discussion).

3.2. Visual speed-discrimination task

In addition to measuring visual-vestibular conflict detection, we also measured visual speed discrimination thresholds. These thresholds were measured to examine whether the effect of fixation behavior on conflict detection could simply reflect an effect of fixation on visual speed estimation. For example, more variable visual speed estimation during head-fixed fixation could explain higher conflict detection thresholds. However, results do not support this hypothesis. Discrimination thresholds were approximately equal during head-fixed (mean=0.244, SD=0.119) and scene-fixed (mean=0.239, SD=0.109) fixation (t=0.297, p=0.770), in support of hypothesis H2, so differences in visual variability alone cannot explain the observed dependence of conflict detection on fixation. Thresholds in the two fixation conditions were not only approximately equal, they were also significantly correlated (r=0.846, p<0.001; Fig. 4B), likely reflecting their dependence on similar underlying visual motion processing mechanisms.

For comparison, we also examined visual speed discrimination thresholds from our prior study which used much slower movements (Fig. 4D). Again, thresholds were not significantly different during scene-fixed (mean=0.249, SD=0.040) compared with head-fixed (mean=0.200, SD=0.071) fixation (t=2.0541, p=0.070). Thresholds were not significantly correlated (r= 0.177, p= 0.626). The magnitude of the thresholds were comparable between the current (Fig. 4B) and previous study (Fig. 4D), unlike during the Conflict detection condition (Fig. 4A vs. Fig. 4C) where thresholds were found to be much lower in the current study.

Figure 4. Results. Plots compare performance between Scene-fixed and Head-fixed fixation in the current study (top row) and the previous study (bottom row) in the Conflict condition (left column) and the Visual condition (right column). All values represent JND as a proportion of the magnitude of the standard stimulus (JND/standard) in order to allow comparison between current and previous studies which used different standard speeds. We refer to this as the gain increment because it indicates the increment or decrement relative to unity gain that was required for the difference to be just noticeable. A) demonstrates that conflict detection was better (JNDs were lower) during scene-fixed compared to head-fixed fixation, a pattern that agrees with results of our previous study, shown in panel C). Note, however, that the gain increment needed for the difference to be just-noticeable was approximately an order of magnitude greater in the previous study (C) than in the current study (A). B) shows that the just-noticeable gain increment in the head-fixed and scene-fixed visual conditions in the current study were comparable and significantly correlated. Result of our previous study, shown in D), show that performance was better during head-fixed compared to scene-fixed fixation. Thus fixation dependent changes in visual variability (B,D) cannot explain fixation dependent changes in conflict detection (A,D).
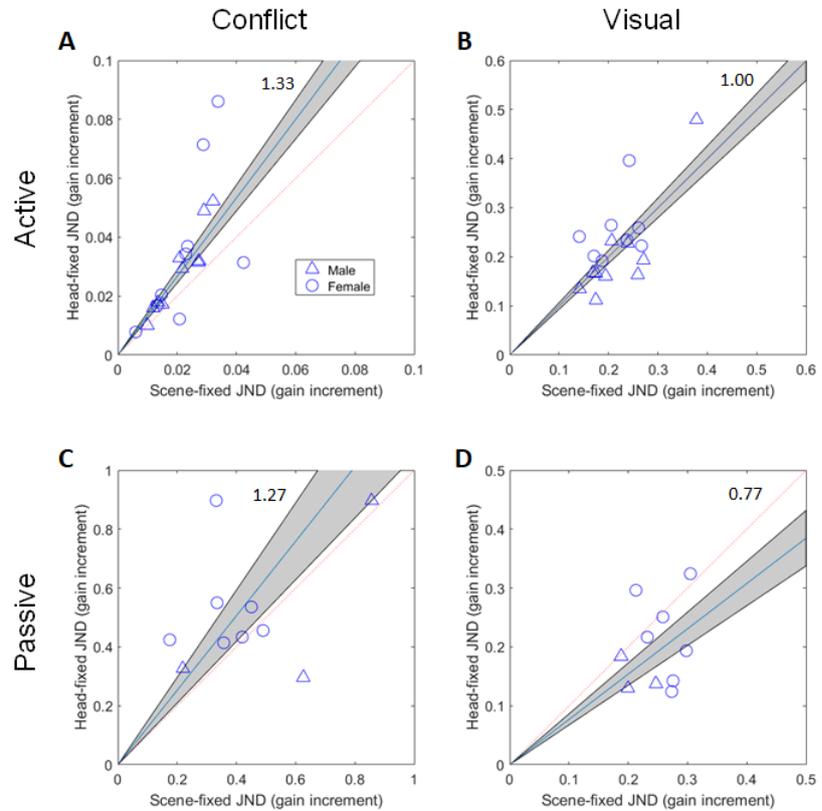
3.3. Simulator Sickness Questionnaire

At the end of each condition, participants rated their feeling of simulator sickness on a 4-point scale, similar to that used in the more comprehensive SSQ [6]. A summary of all responses is shown in Figure 5. Responses were not obtained from 2 subjects (1M, 1F). We hypothesized that greater sensitivity to conflict in the Conflict Detection task would be associated with more severe sickness ratings, but this association was not present in the data (Spearman rank correlation $r=0.15$, $p=0.35$; see also Fig. 5A). This does not support hypothesis H4. Surprisingly, we observed a significant correlation between performance in the visual speed-discrimination task and sickness with less sensitivity associated with more severe sickness (Spearman rank correlation $r=0.49$, $p=0.002$; see also Fig. 5B). Possible reasons for this association are explored in the discussion section.

Sickness ratings were additionally analyzed to examine the effects of Task (Conflict, Visual), Fixation (Head-fixed, Scene-fixed) and Gender (Male, Female). There was a significant effect of Task (Wilcoxon signrank test, $p<0.001$), with more adverse responses reported during the visual discrimination task. This makes sense because in this task, visual optic flow consistent with head motion was presented to stationary observers. We also observed a significant effect of Fixation (Wilcoxon signrank test, $p=0.02$), with more adverse responses reported during Scene-fixed fixation. Conflict detection thresholds were also lower during scene-fixed fixation; this association is consistent with the hypothesis that conflict detection mechanisms that underlie perceptual reports may be the same as those that can ultimately elicit simulator sickness. Finally, there was a significant effect of Gender (Wilcoxon ranksum test, $p<0.001$) with more sickness reported by female participants. This is consistent with gender-dependent effects reported previously, but the cause for these effects remains unknown.

4. Discussion

In the current study, we examined sensitivity to visual gain modulation during active head-on-body rotation using an HMD. This work builds on our prior study [11] in which we measured this sensitivity during slow, passive full-body rotation. We found that conflict detection depends similarly on fixation behavior, regardless of whether head
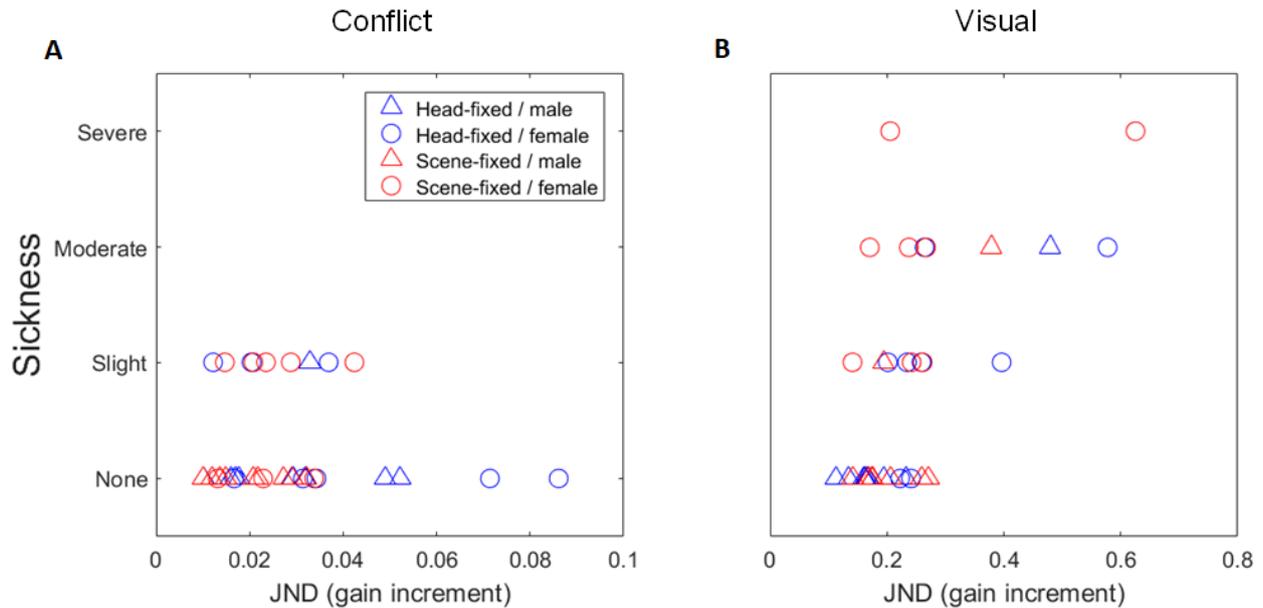
Figure 5. Sickness ratings by condition and gender. Head-fixed indicated by blue, and scene-fixed by red. Males indicated by triangles, females by circles. A) Ratings of sickness following the Conflict Detection conditions. B) Ratings of sickness following the Visual Speed Discrimination conditions.

movements were actively or passively generated. This effect of fixation cannot be explained based on the variability of the visual self-motion estimates because visual discrimination thresholds do not depend similarly on fixation behavior. Instead, we hypothesize that cue comparison mechanisms operate differently depending on oculomotor behavior, regardless of whether movements are active or passive. We also found that conflict detection thresholds were an order of magnitude lower during the current study compared to the previous study. Possible explanations for this large difference in thresholds are discussed below along with a discussion of the relation between conflict sensitivity and simulator sickness.

4.1. Visual-vestibular conflict detection during active head movements

While the experimental design and procedure used to measure conflict detection in the current study was very similar to our prior study, there were several important methodological differences. First, head movements were generated actively in the current study via yaw head turns, whereas the whole body was passively rotated in the previous study. Second, the speed of rotation was much slower in the previous study (~10°/s peak vel.) compared to the current study (~30°/s peak vel.). Third, the visual stimuli were presented on an HMD in the current study, compared with a 3D TV in the previous study; the HMD moved with the head while the screen moved with the platform. Finally, the characteristics of the visual scene were different. Both scenes were composed of a volume of random dots, but the volume, size, density and color of the dots differed.

Despite these differences, the effect of fixation was similar across studies (Fig. 4A, C). Conflict detection is facilitated when fixating a scene-fixed point compared with fixating a head-fixed point, but the reason for this scene-fixed advantage is unclear. We speculate that during scene-fixed fixation, natural gaze stabilization behaviors, including the vestibulo-ocular reflex (VOR), are allowed to operate more or less normally. Increased sensitivity to conflict under these conditions may reflect perceptual access to error signals, in the form of low-velocity retinal slip, that are normally used to drive calibration of the VOR. During head-fixed fixation, on the other hand, the VOR is suppressed resulting in high-velocity retinal slip as the eye moves relative to the scene. Conflict detection in this case depends on comparison of retinal slip and vestibular velocity, and this comparison process appears to operate less efficiently, perhaps because it is not integral to everyday reflexive stabilization behaviors [22].

Surprisingly, conflict detection was found to be an order of magnitude better during the current study with active head movements. An increase in visual speed on the order of ~3% was detected as conflicting in the current study

(Fig. 4A), whereas an increase on the order of ~30% was needed in the prior study (Fig. 4C). Note that this difference was not observed in the visual speed-discrimination condition (compare Fig. 4B and D), suggesting that differences in conflict detection performance across fixation conditions and across studies cannot be attributed to difference in noise on visual motion estimates.

We speculate that the conflict detection advantage in the current study may be a consequence of active head movement, which would allow generating a prediction of the visual scene motion based on efference copy signals [23]. Intuitively, extra information flows should reduce uncertainty and allow for a more accurate assessment of whether any conflict was present. Indeed, several previous studies [24, 25, 26, 27, 28] have shown greater precision of responses when all modalities (vestibular, proprioception, efference copy) signaled a rotation [17].

In addition to the role of active head movement, the difference in performance between current and prior studies could also be due to other methodological differences including speed of head movement (~10°/s vs. ~30°/s peak vel.), display type (HMD vs. screen), and differences in the visual scenes. Previous studies on the detection of latency in HMDs have revealed significant improvements in sensitivity when peak velocity of rotation is increased [29, 30]. The source of the conflict detection advantage in the current study will be investigated in future using identical methods (i.e. same procedure, display, and scene) to investigate conflict detection during identical active and passive head movements (i.e. same speed and trajectory).

4.2. Visual gain perceived as matching

Previous work, including our own, has generally demonstrated that the visual gain necessary to match a physical head rotation is significantly greater than 1 [20, 19, 21]. In this study, we instead observed very little difference from a gain of 1. While this disparity demands closer inspection we suspect that this can be explained by methodological differences. In particular, the comparison process is likely to operate differently during passive compared to active head movement. Direction and speed of head movement are also likely to play a role. Examining the previous literature in greater detail we observe that measurement of visual gain perceived as matching during yaw, rather than roll, pitch or linear head movement, tends to result in matching gains that deviate less from gain of unity [19, 21]. Size and depth cues also influence self-motion perception [31]. Previous studies have shown that manipulating size and distance attributes of the visual scene can effect the matching visual gain [32, 33, 34]. Due to these many methodological variations, prior studies often report that a wide spread of gains (0.8 to 1.4) are accepted as stable [19].

4.3. Sickness and relation to conflict detection

Perceptual measures of conflict detection, like those presented here, may prove to be valuable predictors of susceptibility to motion sickness. Stimulus conditions that lead to improved conflict detection may also be those that lead to increased incidence of simulator sickness. Likewise, individuals who are more sensitive to conflict may be those that are more likely to experience simulator sickness.

In addition to measurements of perceptual sensitivity, we obtained from each participant a 4-point sickness rating after each condition. Participants reported greater discomfort during scene-fixed compared to head-fixed fixation (Fig. 5, red vs. blue). Perceptual sensitivity to conflict was also greater (i.e. JNDs were lower) during scene-fixed compared to head-fixed fixation (Fig. 4A). Thus, in support of hypothesis H_4, greater perceptual sensitivity to conflict was associated with greater likelihood of discomfort in response to conflict. Despite this association, the correlation between conflict JND and sickness ratings (Fig. 5A) was not significant. This non-significant result may be the consequence of insufficient statistical power. There was little sickness reported (only ratings of none or slight reported). Future studies may seek to drive increased levels of sickness, or use a finer scale of reporting, in order to observe a significant correlation.

We observed significantly more severe sickness reports in the visual speed-discrimination task than in conflict detection (Fig 5A vs 5B). This is perhaps not surprising because visually-simulated self-motion in the absence of physical motion is known to be conducive to simulator sickness [35, 36, 37]. The reduced sickness in the conflict detection task is likely because non-visual cues to self-motion were approximately in agreement with the visual cues. The only difference was due to the manipulation of visual gain. The conflict detection task included active head rotations. Motor activation may mitigate the effects of conflict due to agency or control [38]. Previous studies have shown how postural instability [39, 40] and control and the resulting expectations [41, 38] mediate motion sickness. An active participant necessarily has greater control which may account for the reduction in symptoms.

Interestingly, in the visual speed-discrimination task, sickness rating and JND were positively correlated (Fig. 5B); greater sickness was reported by those subjects who were least sensitive. This result appears to contradict hypothesis H_4, but only if the following assumptions are valid: 1) performance on the visual discrimination task is an accurate measure of noise on the visual estimate, and 2) conflict detection is limited by this noise (i.e. signal detection model of conflict detection [42]). The observation that conflict JNDs are significantly lower than visual JNDs (compare Fig. 4A and 4B) suggests that these assumptions do not hold. Alternatively, it may be that participants who reported greater levels of sickness were simply less able to concentrate on the task at hand. This association deserves to be investigated more deeply in future experiments.

Finally, we observed an effect gender similar to previous studies, with more sever sickness reported by females than males (Fig. 5, triangles vs circles). With so many contradictory findings in existence, this additional result does little to swing the balance of evidence either way. Far more work is necessary in this arena before any definite conclusions can be made.

References

[1] S. Davis, K. Nesbitt, E. Nalivaiko, A systematic review of cybersickness, in: Proceedings of the 2014 Conference on Interactive Entertainment, ACM, 2014, pp. 1–9.

[2] G. Bertolini, D. Straumann, Moving in a moving world: a review on vestibular motion sickness, Frontiers in neurology 7.

[3] C. M. Oman, Motion sickness: a synthesis and evaluation of the sensory conflict theory, Canadian journal of physiology and pharmacology 68 (2) (1990) 294–303.

[4] J. T. Reason, J. J. Brand, Motion sickness., Academic press, 1975.

[5] L. J. Hettinger, K. S. Berbaum, R. S. Kennedy, W. P. Dunlap, M. D. Nolan, Vection and simulator sickness., Military Psychology 2 (3) (1990) 171.

[6] R. S. Kennedy, N. E. Lane, K. S. Berbaum, M. G. Lilienthal, Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness, The international journal of aviation psychology 3 (3) (1993) 203–220.

[7] M. E. McCauley, T. J. Sharkey, Cybersickness: Perception of self-motion in virtual environments, Presence: Teleoperators & Virtual Environments 1 (3) (1992) 311–318.

[8] D. Harm, Motion sickness neurophysiology, physiological correlates, and treatment, Handbook of virtual environments (2002) 637–661.

[9] J. Jerald, The VR book: Human-centered design for virtual reality, Morgan & Claypool, 2015.

[10] B. D. Lawson, Motion sickness symptomatology and origins. (2014).

[11] I. T. Garzorz, P. R. MacNeilage, Visual-vestibular conflict detection depends on fixation, Current Biology 27 (18) (2017) 2856–2861.

[12] K. M. Stanney, R. S. Kennedy, K. S. Hale, Virtual environment usage protocols. (2014).

[13] A. S. Fernandes, S. K. Feiner, Combating vr sickness through subtle dynamic field-of-view modification, in: 3D User Interfaces (3DUI), 2016 IEEE Symposium on, IEEE, 2016, pp. 201–210.

[14] J. Treleaven, J. Battershill, D. Cole, C. Fadelli, S. Freestone, K. Lang, H. Sarig-Bahat, Simulator sickness incidence and susceptibility during neck motion-controlled virtual reality tasks, Virtual Reality 19 (3-4) (2015) 267–275.

[15] S. Von Mammen, A. Knote, S. Edenhofer, Cyber sick but still having fun, in: Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, ACM, 2016, pp. 325–326.

[16] D. Genzel, P. MacNeilage, L. Wiegrebe, Updating and orientation in auditory space, The Journal of the Acoustical Society of America 137 (4) (2015) 2201–2201.

[17] D. Genzel, U. Firzlaff, L. Wiegrebe, P. R. MacNeilage, Dependence of auditory spatial updating on vestibular, proprioceptive, and efference copy signals, Journal of neurophysiology 116 (2) (2016) 765–775.

[18] N. Prins, Kingdom, faa (2009). palamedes: Matlab routines for analyzing psychophysical data (2014).

[19] P. Jaekl, M. Jenkin, L. R. Harris, Perceiving a stable world during active rotational and translational head movements, Experimental brain research 163 (3) (2005) 388–399.

[20] H. Wallach, Perceiving a stable environment, Scientific American 252 (5) (1985) 118–125.

[21] B. C. Grácio, J. Bos, M. Van Paassen, M. Mulder, Perceptual scaling of visual and inertial cues, Experimental Brain Research 232 (2) (2014) 637–646.

[22] A. Skavenski, R. Hansen, R. M. Steinman, B. J. Winterson, Quality of retinal image stabilization during small natural and artificial body rotations in man, Vision research 19 (6) (1979) 675–683.

[23] K. E. Cullen, The vestibular system: multimodal integration and encoding of self-motion for motor control, Trends in neurosciences 35 (3) (2012) 185–196.

[24] J. A. Crowell, M. S. Banks, K. V. Shenoy, R. A. Andersen, Visual self-motion perception during head turns, Nature neuroscience 1 (8) (1998) 732.

[25] K. E. Cullen, J. E. Roy, Signal processing in the vestibular system during active versus passive head movements, Journal of neurophysiology 91 (5) (2004) 1919–1933.

[26] T. Mergner, G. Nasios, C. Maurer, W. Becker, Visual object localisation in space, Experimental Brain Research 141 (1) (2001) 33–51.

[27] T. Nakamura, A. M. Bronstein, The perception of head and neck angular displacement in normal and labyrinthine-defective subjects: A quantitative study using a remembered saccadetechnique, Brain 118 (5) (1995) 1157–1168.

[28] J. E. Roy, K. E. Cullen, Dissociating self-generated from passively applied head motion: neural mechanisms in the vestibular nuclei, Journal of Neuroscience 24 (9) (2004) 2102–2111.

[29] J. J. Jerald, Scene-motion-and latency-perception thresholds for head-mounted displays, Ph.D. thesis, The University of North Carolina at Chapel Hill (2009).

[30] J. Jerald, M. Whitton, F. P. Brooks Jr, Scene-motion thresholds during head yaw for immersive virtual environments, ACM Transactions on Applied Perception (TAP) 9 (1) (2012) 4.

[31] P. R. MacNeilage, M. S. Banks, D. R. Berger, H. H. Bülthoff, A bayesian model of the disambiguation of gravitoinertial force by visual cues, Experimental Brain Research 179 (2) (2007) 263–290.

[32] S. Palmisano, Perceiving self-motion in depth: The role of stereoscopic motion and changing-size cues, Perception & Psychophysics 58 (8) (1996) 1168–1176.

[33] B. Rogers, M. Graham, Motion parallax as an independent cue for depth perception, Perception 8 (2) (1979) 125–134.

[34] J. Kim, S. Khuu, A new spin on vection in depth, Journal of Vision 14 (5) (2014) 5–5.

[35] L. J. Hettinger, G. E. Riccio, Visually induced motion sickness in virtual environments, Presence: Teleoperators & Virtual Environments 1 (3) (1992) 306–310.

[36] E. M. Kolasinski, Simulator sickness in virtual environments., Tech. rep., Army research inst for the behavioral and social sciences Alexandria VA (1995).

[37] S. Palmisano, R. Mursic, J. Kim, Vection and cybersickness generated by head-and-display motion in the oculus rift, Displays 46 (2017) 1–8.

[38] A. Rolnick, R. Lubow, Why is the driver rarely motion sick? the role of controllability in motion sickness, Ergonomics 34 (7) (1991) 867–879.

[39] G. E. Riccio, T. A. Stoffregen, An ecological theory of motion sickness and postural instability, Ecological psychology 3 (3) (1991) 195–240.

[40] J.-R. Chardonnet, M. A. Mirzaei, F. Merienne, Features of the postural sway signal as indicators to estimate and predict visually induced motion sickness in virtual reality, International Journal of Human–Computer Interaction 33 (10) (2017) 771–785.

[41] A. Rolnick, Exploring the active-passive distinction in movement control remote control of passive rotation can reduce vestibular after-effects and motion sickness, Proceedings of the 8th International symposium on posturography.

[42] T. D. Wickens, Elementary signal detection theory, Oxford University Press, USA, 2002.

Chapter 5

General Discussion

The projects presented in this thesis provide new insights into how the nervous system maintains the percept of a stable environment despite self-motion. On the one hand, a model for conflict detection has been evaluated, providing insights into possible mechanisms that underlie our ability to evaluate incongruencies between different sensory signals (Chapter 2). On the other hand, a vestibular analogue of the classical Aubert-Fleischl phenomenon has been presented, highlighting the role of reference frame transformations for perceiving a stable environment (Chapter 3). In addition, it has been shown that conflict detection performance is improved during active head turns (Chapter 4) compared to passive whole-body rotations (Chapter 2). Taken together, this work aims to further multi-sensory research in the field of self-motion perception – including visual and vestibular but also oculomotor and efference copy signals.

In the following sections, I compare the two reported studies on conflict detection (Chapter 2 and Chapter 4) and discuss potential implications for virtual reality applications and motion sickness. I also discuss the second study on the vestibular Aubert-Fleischl illusion with respect to models explaining perception of object motion during self-motion. The closing remarks summarize the findings of this thesis.

5.1   Conflict detection and implications

The first study of this thesis (Chapter 2) has shown that visual-vestibular conflict detection can be modeled as crossmodal discrimination, in agreement with standard signal detection theory. Interestingly, there is a trade-off between the integration and segregation of signals, which depends on oculomotor signals. Note that segregation is used as a conceptual term to contrast

with integration and should be understood in the context of conflict detection. It does not have a functional meaning implying separately accessible percepts of self-motion in inertial and visual space. Self-motion is always unitary with incongruence/conflict being perceived as a separate quality, namely a scene-in-space motion. Conflicts can be more easily detected when the eyes are moving whereas integration is optimal when the eyes are stationary and only optic flow is available. The critical role of eye movements for mediating this trade-off can be best understood by comparing various task requirements for natural self-motion. During locomotion, we most often fixate world-fixed targets (Einhauser et al., 2007) to collect information about those parts of the environment which are of greatest interest and importance to us. This is achieved by maintaining high dynamic visual acuity, i.e. by stabilizing gaze. During world-fixed fixation, retinal slip represents an indication that the compensation of head motion by combined pursuit and vestibulo-ocular reflex (VOR) systems is insufficient. In case of slow head movements, the smooth pursuit system would probably compensate almost perfectly for the head rotation. During fast movements, however, retinal slip could be an indicator for mismatches or conflict between visual and vestibular input. The degree of conflict, which is an error signal, could then be used to calibrate gaze stabilization via the VOR (Colagiorgio et al., 2015).

On the other hand, self-motion requires control of balance and heading direction – tasks which are relying on integration processes to achieve greater precision. However, integration is less beneficial when one signal is much more variable than the other. We suggest that the tendency to fixate world-fixed objects may indicate a higher priority of conflict detection than the increase in precision of self-motion estimation that is obtained through Maximum Likelihood integration. To investigate whether our findings are subject to our specific and highly controlled setup or whether they are generalizable to more naturalistic self-motion contexts, we conducted the follow-up study (Chapter 4), examining how eye movements influence conflict detection performance during active head turns.

In the first study (Chapter 2), participants experienced highly controlled passive whole-body yaw rotations on the motion platform, while the head was kept stationary with respect to the body. This setup aimed at comparing retinal, oculomotor and vestibular signals and minimizing proprioceptive cues and efference copy signals from neck muscles. In the follow-up study, in contrast, participants were instructed to actively turn their head to allow for additional proprioceptive feedback and efference copies from the neck. In addition, the visual input was delivered via a head-mounted display (HMD) instead of an external screen and head rotations were of higher peak velocity ($\sim 29.5°/s$) compared to the passive whole-body rotations (peak velocity of

10°/s for the standard).

Both studies showed that conflict detection performance is better during scene-fixed fixation compared to head-fixed fixation. However, the tenfold increase in precision of conflict detection performance during self-generated head rotations compared to passively generated whole-body rotations requires further explanation.

Assuming the difference in precision is not due to artifacts[1], possible reasons could be:

1. Since visual thresholds for both head-fixed and scene-fixed fixation were comparable to those of the previous study, visual variability as a reason for increased conflict sensitivity can be ruled out. The most obvious difference between the two studies is the additional input from efference copy and proprioceptive signals during active head rotations. Prior research has shown that precision increases when all available modalities provide information about rotation (Crowell et al., 1998; Cullen and Roy, 2004; Genzel et al., 2016). However, as put forward in the study of Chapter 2, the model of crossmodal discrimination limits conflict detection performance to the sum of the single cue variabilities. As visual discrimination thresholds are an order of magnitude larger than the conflict detection thresholds, crossmodal discrimination cannot readily explain our findings. Alternatively, as the current study does not simply investigate bimodal discrimination but a comparison between visual (retinal and oculomotor), vestibular, proprioceptive and neck efference copy signals, the underlying process might be different. Visual signals are possibly fused with efference copy signals in the first stage and the resulting combined and more precise estimate is then compared to the vestibular input. That way, visual single cue thresholds do not directly go into the crossmodal discrimination stage, but as part of the fused estimate.

2. Active head turns may involve different processing of stimuli compared to passive whole-body rotations. According to von Holst and Mittelstaedt (1950), passive movements yield sensory signals which represent changes in the world (exafferent signals), while active movements induce sensory changes resulting from our own movements (reafferent signals). It is crucial for us to differentiate between these two kinds of sensory signals to ensure proper navigation and a correct spatial

---

[1]To ensure that the increased sensitivity during active head turns is not simply due to the HMD or other artifacts, we propose a control experiment for the passive condition on the motion platform with the HMD.

representation of the world. Efference copy signals from our motor commands (see Chapter 1.2.1) which predict the sensory outcome of our movements get subtracted from the actual sensory signals to filter out that part of sensory stimulation which results from one's own movement (Sommer and Wurtz, 2002, 2008). Roy and Cullen (2004) have provided evidence for this assumption, showing that neurons in the vestibular nuclei (but not the primary vestibular afferents) react differently to active versus passive movements in the absence of visual input. During active head turns, a signal cancelling the vestibular stimulation is generated, but only if proprioceptive signals from the neck correspond to predictions from the neck motor efference copy. In daylight, however, the expected consequences of a head command as predicted by an internal model do not only include vestibular and proprioceptive but also visual and oculomotor signals. Therefore, a more general mechanism of comparing prediction and outcome could serve to cancel reafferent information and allow for detection of exafferent signals, such as the deviation of the visual scene from a stable world in our paradigm. Only the non-predicted bit of visual or vestibular input, which represents the conflict in our paradigm, remains and accumulates as an error signal over the course of the trial, allowing for improved conflict detection during active head turns. Such an accumulation of the error signal could be understood in the framework of so-called dynamic drift-diffusion models of decision making which explain decisions on the level of cognitive processing (Ratcliff and McKoon, 2008). According to these models, noisy sensory information is added up and maintained over time until one of two boundaries (representing the two alternatives of two-decision tasks) which represent a specific criterion is reached and the respective response (decision A versus decision B) is elicited. The criterion depends for example on the instructions and task requirements. In case of a reaction time task, the boundary would be lower than in a task maximizing accuracy. In the latter case, the boundary would be shifted further away, implying more sensory evidence has to be acquired. How fast information is accumulating (i.e. the *drift rate*) depends on the information quality. In the abovementioned case, the speed of error accumulation and decision making (conflict detection) would depend on the size of the conflict. The advantage of drift-diffusion models is that they can explain performance across individual trials in contrast to the analysis by means of psychometric functions. Research on rhesus monkeys (Roitman and Shadlen, 2002) has provided evidence for a neuronal correlate of drift-diffusion processes, such as neurons in the lateral intraparietal cortex

(for a review of physiological underpinnings of the drift-diffusion model: Smith and Ratcliff 2004).

3. One might argue that the noise measured in the visual single-cue conditions is different from the visual noise contribution in the conflict detection task so that single cue thresholds cannot be used to predict conflict detection performance. Oculomotor signals in the single cue condition consisted of smooth pursuit eye movements, whereas the respective scene-fixed conflict condition involved oculomotor signals enhanced by the VOR. Retinal signals in the single cue condition evolved from optic flow while the eyes were stationary, whereas retinal signals in the head-fixed conflict condition emerged from relative background motion due to the head-fixed fixation target during active head turns, cancelling the VOR.

The next section focuses on the implications of findings from the studies on conflict detection (Chapter 2 and Chapter 4) for the development of VR devices.

5.1.1 VR technology

Individual trials in the follow-up study (Chapter 4) are experimentally less controlled since active head turns yield small individual variations in the motion profile and movement duration. Despite reduced experimental control, this setup is ecologically more valid. Participants were exposed to virtual reality in a head-mounted display while moving their head on the body. This active head movement is more similar to common everyday experiences than passive whole-body rotations with the stimulus rendered on an external screen. In general, HMDs disentangle the naturally correlated sensory signals during self-motion by shutting out any real-world visual cues, replacing them with virtual reality stimuli and thereby allowing for active, large-scale movements through space (Campos and Bülthoff, 2012). HMDs such as the Oculus Rift used in this study, update the visual scene based on the participant's head movements. Whenever there is a mismatch between the rendered visual scene and the vestibular signals, simulator sickness may result (Oman, 1990; Reason and Brand, 1975; Hettinger et al., 1990; Kennedy et al., 1993; McCauley and Sharkey, 1992).

One hypothesis is that persons who are perceptually more sensitive to conflict have a higher risk of developing symptoms of motion sickness. This hypothesis assumes that the higher the response to sensory conflict on the neuronal level – in terms of conflict signals being linked to emetic regions

of the brain (Oman, 1991) – the higher conflict detection sensitivity on the perceptual level. In the follow-up experiment, we tried to test this hypothesis by assessing motion sickness after each condition. Higher perceptual sensitivity to conflict was generally associated with higher motion sickness ratings. However, we could not find a significant correlation between them, mainly because the motion sickness ratings were generally low and thus not suitable to sufficiently differentiate between participants. Future research may further investigate the hypothesized correlation between susceptibility to motion sickness and the perceptual sensitivity to conflict with means of a finer assessment of motion sickness (including subjective reports about everyday experience of motion sickness for example on a boat). Since motion sickness is typically experienced during passive self-motion (Bertolini and Straumann, 2016) it would also be interesting to assess its symptoms during passive whole-body rotations (compare with Chapter 2) with an HMD and during prolonged stimulation.

To summarize, investigation of conflict sensitivity, as in our studies, may be relevant to researchers using virtual reality devices and developers of HMDs. The following section will elaborate on the hypothesized mechanisms of motion sickness.

5.1.2   Motion sickness

"Treisman (1977) proposed that motion sickness evolved as an accidental byproduct of an early-warning system for detecting the effects of ingested neurotoxins. Whether or not this is the case, the fact remains that motion sickness is not an inevitable consequence of the human condition: if we had remained as self-propelled animals content to stay within our normal Earth gravity environment, the problem would not have arisen. To this extent, therefore, it is a self-inflicted malady." (Reason, 1978)

Trying to understand the mechanisms underlying motion sickness with its typical symptoms such as cold sweating, nausea, vomiting and fatigue has been a focus of applied vision research at least as early as the 1960s. Studies by Guedry (1964), Steele (1961), and Reason (1970) supported the hypothesis that motion sickness is caused by conflicts between unfamiliar sensory self-motion signals which observers are unable to compensate for. However, situations involving little sensory conflict such as swinging or being on a ship cause tremendous symptoms of motion sickness while highly conflicting stimuli such as tilting rooms (Witkin, 1949) reveal a low incidence of symptoms. Later, this hypothesis was rejected by some based on observations showing that motion sickness cannot sufficiently be explained by current conflicting

sensory stimulation since repeated exposure finally causes symptoms to disappear although the present stimulation is unchanged – a finding which can be best understood in a framework based on research by von Holst (1954). He postulated that our brain deploys a mechanism which distinguishes between re-afferent sensory feedback coming from our own movements and ex-afferent sensory signals resulting from passive movements to allow for appropriate locomotion (see Chapter 1.2.1).

Held (1961) adopted von Holst's basic summation model of re-afferent and ex-afferent signals but modified it to explicitly account for adaptation processes. He proposed that the efference copy of a motor command (for example sent to the extra-oculomotor muscles) activates the re-afferent trace in the so-called *correlation storage* which corresponds to the respective efferent signal, based on previously acquired co-occurrence statistics. Actual re-afferent signals arising from the movement are then compared to the revived re-afferent signal in the so-called *comparator stage*. A resulting mismatch, i.e. a conflict between the present sensory input and the one rooted in the past "exposure-history" is necessary for motion sickness to evolve. It then drives adaptation (here: *sensory rearrangement*) processes by adjusting the co-occurrence statistics of the correlation storage.

Following this model, Reason (1978) added another premise to the evolution of motion sickness, the indispensable contribution of the vestibular organs, even if they are only indirectly involved as in visually-induced motion sickness. This could explain why patients without an intact vestibular system are not susceptible to motion sickness and why changes in experienced velocity are necessary for motion sickness symptoms (Money, 1970).

The idea that mismatches between expectations and actual sensory outcomes yield motion sickness was further developed by Oman (1982, 1990, 1991) who proposed an *internal model*, analogous to Held's correlation storage. This internal model receives efferent input from motor commands and afferent input from sensory feedback to estimate head and body posture and to predict the sensory input resulting from the motor commands. Those are constantly compared and during pure active head movement, the sensory conflict between the two (prediction and actual sensory feedback) is small, implying almost complete cancellation of reafferent input. In contrast, passive motion or a systematic change in the normal relationship between body movement and sensory afference yield "sensory conflict" signals, providing an error signal and driving sensorimotor learning.

Oman (1991) not only transferred Held's and Reason's ideas into a quantitative model but also included a linkage between the conflict signals and emetic regions in the brain for the evolution of motion sickness. This linkage is based on a nonlinear leaky integrator, i.e. a differential equation which

transfers the integral of an input signal into an output signal, while gradually dropping some part of the initial signal. It reflects the typical characteristics of symptoms such as nausea which have slow latencies of up to several minutes, until finally vomiting occurs. Another important determinant of the leaky integrator is an output (as opposed to an input) threshold to account for the latency and the fact that strong, but brief sensory conflicts do not cause motion sickness. Only when the sensory conflict which gets integrated over time reaches the threshold, nausea appears. In addition, the leaky integrator model can also explain the slow latency of symptoms relief after conflicting signals have vanished.

Evidence supporting Oman's model based on reafference cancellation comes from studies on neurons in both cerebellum and the vestibular nuclei (Oman and Cullen, 2014). So-called "vestibular only" (VO) neurons respond to signals from the otolith and the semicircular canals (McCrea et al., 1999; Roy and Cullen, 2002, 2004; Cullen et al., 2009; Sadeghi et al., 2009; Carriot et al., 2013) and react differently to active versus passive movements. During active head turns, a signal cancelling the vestibular reafference is generated, but only if reafferent proprioceptive signals from the neck match predictions from the neck motor efference copy (Cullen et al., 2009). Thus, only vestibular signals stemming from the passive component of a movement generate activity of VO neurons. Following Brooks and Cullen (2013), rostral fastigial nucleus neurons in the cerebellum also play a key role in reafference cancellation by exhibiting decreased activity in response to self-generated movements.

In summary, permanent conflicts between motor and sensory processes, i.e. between predicted and actual stimulation have a high potential to elicit motion sickness. Therefore, they are of significant importance for virtual and augmented reality devices which artificially perturb the correlational relationship between visual and proprioceptive stimulation during active movements.

5.2   Object motion during self-motion: The vestibular Aubert-Fleischl phenomenon and its implications

All multisensory interactions such as integration or segregation require the brain to transform signals from different modalities into a common reference frame. The reference frame to which multisensory signals are converted to, depends not only on the contributing signals but also on their magnitudes (Fetsch et al., 2007). As pointed out in Chapter 1.3.6 those transformations

can be incorrect or incomplete, resulting in sensory illusions, i.e. objectively matching signals are perceived as containing conflicting information. An illusion which has been ascribed to such an erroneous reference frame transformation is the so-called *Aubert-Fleischl (AF) illusion*, which describes the quite common phenomenon, that objects seem to move more slowly when pursued with the eyes compared to when the object is moving in front of stationary eyes (Aubert, 1886). To make this phenomenon intuitively accessible, imagine you are a pedestrian waiting for the traffic lights to turn green. Let us consider two possible scenarios. First, your eyes keep fixating the lights when a car passes by. Your estimate of the speed of the car will then fully depend on the retinal motion of the car passing through the scene as your eyes remain stationary in their orbits. Alternatively, you could be interested in the traffic on the street and start pursuing the car when it passes by. In this case, your speed estimate will mainly depend on the oculomotor pursuit signals (for simplicity, neglecting any head movement which naturally occurs as a combined eye-head pursuit). What observers usually report is that the car appears to move more slowly during oculomotor pursuit compared to the first situation where the eyes were kept stationary.

While this classical AF illusion is caused by differences in retinal and oculomotor speed estimates, the vestibular AF illusion (Chapter 3) represents an analogous phenomenon, i.e. differences in perceptual estimates of retinal and vestibular speed. In an experimentally controlled task we replicated the classical AF illusion and we found the analogous vestibular phenomenon, in which objects that are physically pursued through passive whole-body rotations appear to move more slowly than non-pursued objects. The experiment shows that both extra-retinal signals from the vestibular system and from eye movements are not sufficiently compensated for when judging object speed during self-motion. The comparable sizes of the classical and the newly discovered AF effect suggest a common explanation for the underestimation of object speed in both conditions.

Furthermore, our experiment adds to previous research showing that object speed is often underestimated during self-motion. Dyde and Harris (2008) have shown that pursued targets must move in the same direction as the moving observer to be perceived as stationary in the world. They also found that the underestimation of self-motion is slightly more pronounced in darkness compared to light conditions and during passive compared to active motion. The average stationarity gain, i.e. the ratio between the actual and the geometrically required target motion (relative to the observer) reported in this study amounted to 0.54 for active and 0.42 for passive motion, similar to results from a study by Wexler (2003) who found an average gain of 0.62 for active and 0.43 for passive motion. While the study by Dyde and Har-

ris (2008) shows that self-motion is underestimated progressively as retinal (absence of optic flow) and extra-retinal signals (passive motion, darkness) are removed, the influence of eye movements versus retinal signals could not be determined since the visual stimulation contained no retinal motion (optic flow) but only fixating eye movements (a mixture of smooth pursuit and oculo-vestibular reflexes). In contrast, our study also disentangled the contributions of eye movement versus retinal and relative motion components.

In addition, the vestibular AF phenomenon, in which observers are no longer stationary, showed that self-motion is insufficiently compensated for during the perception of object speed. It exemplifies the direct interplay and interdependence between the perception of self-motion and the perception of object motion. On the one hand, self-motion perception requires correct interpretation of the optic flow components resulting from self-motion versus object motion (flow parsing, Warren and Rushton 2007). On the other hand, object motion perception requires correct estimation of how fast and in which direction we are moving.

All of these results are highly relevant for understanding the interaction of the various senses contributing to self-motion perception and their goal to establish a percept of a stable world. Interestingly, although during high speed and challenging activities, the gain between retinal and extra-retinal signals (such as oculomotor or vestibular signals) is supposedly not always 1 (Wertheim, 1994), the world seems stationary to us. Wertheim (1994) pointed at the possibility that these incongruencies between signals are masked to a high degree because all sensory estimates, especially vestibular ones, are noisy. In fact, prior studies that have investigated observers' ability to match visual self-motion to physically experienced motion have often reported high variability on matching judgments for yaw rotations for both active (Jaekl et al., 2005; Wallach, 1985) and passive (Jurgens and Becker, 2011) head rotation. Wallach (1985) referred to this variability as the "range of immobility" because within this range the environment is perceived as stationary. The notion of a "range of immobility" corresponds to the masking due to high JNDs. Wertheim (1994) also mentions that the tendency to perceive the world as stationary due to high thresholds, comes at the cost of perceiving moving objects as stationary or underestimating their speed – a situation resulting in phenomena such as the AF illusions.

5.2.1   Background motion in natural settings

One important aspect to be considered when interpreting the AF illusions is that both in our study as well as in previous studies investigating the classical AF illusion (Dichgans et al., 1975; Freeman et al., 2010; Raymond

et al., 1984; Souman et al., 2006) there was no background stimulus present across which the target was moving in the pursuit interval. This lack of relative motion signal contrasts with everyday life where we are surrounded by many stationary objects and backgrounds which move relative to our eye movements. Our prior experience predicts relative motion of the stationary world surrounding the target in the opposite direction of pursuit.

During the vestibular AF condition, there was also no relative background motion with the purpose to investigate the mere influence of vestibular signals on the perception of object motion. Participants were fixating a head-fixed target on a background which was stationary relative to the eyes. An analogous situation in real life such as sitting on a bus while looking at a spot on the window, however, induces optic flow, i.e. retinal slip due to relative motion of the background with respect to the spot on the window. This implies that both illusions could be ascribed to unnatural impoverished laboratory conditions which unwind the natural co-occurrence of retinal, oculomotor and vestibular signals.

There is a trade-off between the ecological validity and the experimental control of studies. However, experiments which artificially disentangle the various signals contributing to self-motion perception still help evaluating different models and hypotheses of how we perceive object and self-motion. Thus, it is worthwhile investigating even rare situations and transferring the results to make predictions about perception during real-life and less controlled situations.

The following sections will discuss two different models of perception of object speed during self-motion and their potential for explaining the AF phenomena.

5.2.2 Reference-signal models of perception of object motion during self-motion

The reference-signal model by Wertheim (1994) provides a theoretical framework to explain how we maintain the percept of a stable world despite our eye movements and our own self-motion. It contrasts with the theory of direct perception (Gibson, 1968) which assumes that both object motion and self-motion perception can simply be derived from afferent retinal signals. According to the theory of direct perception, extra-retinal signals are not required since our nervous system can sufficiently infer object motion and self-motion from the optic flow field. Wertheim's model also contrasts with classical inferential theories (e.g. von Holst and Mittelstaedt 1950) which assume that our nervous system achieves to perceive the world as stable due to

a continuous comparison between retinal and extra-retinal signals from the eyes and the vestibular system. According to these models, whenever retinal and extra-retinal signals agree, objects are perceived as stationary. As soon as a difference is registered, object motion is perceived.

A counter-example to these theories is vection, i.e. visually induced perception of self-motion, where only retinal signals are present (Berthoz et al., 1975; Brandt et al., 1973, compare Chapter 1.2.1). Although the comparison process between retinal and extra-retinal signals should induce the percept of object motion according to inferential theories, (illusory) self-motion is perceived after some time. The theory of direct perception, on the other hand, cannot explain vection since the visual stimulus remains the same although the percept changes from initial object motion to self-motion.

Wertheim's model is a modification of the classical inferential theories and can explain vection by considering visual afferents which result from optokinetic image flow, i.e. large visual patterns of low spatial frequency moving across the retina. According to Wertheim, presenting an observer with an optokinetic stimulus (e.g. an optokinetic drum) will initially (and correctly) make him perceive object motion since only retinal signals are present. Visual afferents will then gradually build up as they are gated through a low band-pass spatiotemporal filter until they equal the (unchanged) retinal signals. As soon as there is no difference between the retinal and the visual (optokinetic) afferents any more, the visual pattern appears stationary in space and the illusory percept of self-motion develops. To summarize Wertheim's idea, instead of comparing retinal with extra-retinal signals to decide between object motion and self-motion, he suggests a comparison between retinal and so-called *reference signals*, which include any combination of extra-retinal (efference copy and vestibular) signals *and* visual (optokinetic) afferents. In case of vection, the reference signal which gradually counterbalances the retinal signals consists only of visual afferents.

Another situation supporting Wertheim's idea of visual afferents contributing to a reference signal is our percept of a stable world during movements of constant velocity. As mentioned in the introduction (Chapter 1.2.2), the vestibular organs only respond to changes in velocity. Thus, the visual afferents take over during constant velocity and allow the reference signal to permanently equate retinal image motion, thereby rendering the world as stable while we are moving through it. A possible location for this interaction of vestibular and visual cues is the medial vestibular nucleus which receives inputs from retinal neurons via the optic tract and from vestibular afferents (Kandel et al., 2000, p. 812-813).

Wertheim's model (Wertheim, 1994) which underlines the tight relationship between the perception of object motion and self-motion, also provides

an explanation for the classical Aubert-Fleischl illusion. According to the model, perceived velocity depends on the difference between retinal (image motion across the retina) and reference signals (eye movement in space), minus the JND between both. Thus, whenever pursuing a target across a stationary background, both retinal and reference signals increase equally, meaning their difference stays the same. However, the JND between them increases with increasing velocity according to Weber's law, so perceived velocity is reduced. This implies that any smooth pursuit of a target against a stationary background is underestimated compared to when the target is moving across stationary eyes (Wertheim, 1994), yielding phenomena such as the AF illusion.

5.2.3 A Bayesian model of object speed perception

Freeman et al. (2010) provide an alternative explanation of the classical AF phenomenon with reference to Bayesian theory. Instead of assuming any inaccuracy in the sensorimotor estimates contributing to perceived object speed, their model postulates unbiased sensory estimates that only differ in precision. According to this idea, noisy sensory estimates are combined with prior expectations about object speed, following Bayes' law. In the first stage of their model, the so-called measurement stage, signals are kept separate, meaning both retinal and pursuit target motion signals have their independent noise sources, and are unbiased. In the estimation stage, each of the two likelihood functions is multiplied by a zero-motion prior. These two priors express the assumption that objects are generally at rest. The stronger the zero-motion prior and the noisier (less reliable) the sensory signal, the more the posterior distribution is shifted towards zero. By summing the two resulting estimated speeds, head-centric velocity is determined. As comparisons between pursuit and retinal motion in the AF experiment happen across intervals, head-centric velocity of the two intervals is compared according to standard signal detection theory, so that the probability of the decision of which interval was perceived as faster can be determined. As pursuit estimation was found to be generally noisier than retinal motion estimation, the zero-motion prior has bigger influence on the pursuit estimate, yielding an underestimation of the pursuit signals, which can explain the AF effect.

To summarize, according to Freeman et al. (2010), the size of the AF phenomenon is dependent on the precision of the contributing sensory estimates and the strength of the assumed zero motion prior with which they are combined. The advantage of this model is that it does not assume biased sensory estimates. The AF effect is explained simply by different degrees of precision and not due to a lack of accuracy of sensory estimates. This is

compelling since differences in the accuracy of visual speed estimation are less easy to be justified as the nervous system could calibrate the sensory systems via adaptive procedures.

However, further research is needed to elucidate whether this model also applies to our newly found vestibular AF phenomenon, meaning whether vestibular signals also lack in precision compared to retinal motion signals.

5.3 Concluding remarks

Understanding self-motion as a highly complex interaction of cues from various sensory modalities has entailed a lot of fruitful neuroscientific research. While psychophysical studies used to focus on multisensory cue integration as the basic mechanism underlying our percept of a stable world, the interest is slowly shifting towards the investigation of more complex situations where multisensory signals originate from different objects. The processing of conflicting information about self-motion has thus been the focus of the current thesis.

While it showed that visual-vestibular conflict detection during passive self-motion can be modeled as crossmodal discrimination, it also elucidated the decisive, but widely neglected role of eye movements for self-motion perception. Oculomotor information seems to facilitate conflict detection at the cost of impaired cue integration. It also showed that conflict detection performance is largely improved during active head turns, promoting further research on the influence of proprioceptive and efference copy signals for self-motion perception.

The second study of this thesis additionally showed that reference frame transformations required for the integration or segregation of multisensory input can sometimes be incomplete, thereby inducing illusions such as the oculomotor and vestibular AF effects. Their comparable size suggests common mechanisms, however models to explain these phenomena are needed.

Finally, the research presented in this thesis also relates to more applied topics, such as the origin of motion sickness and the development of virtual reality devices.

Bibliography

Angelaki, D. E. and Cullen, K. E. (2008). Vestibular system: the many facets of a multimodal sense. *Annual Review of Neuroscience*, 31:125–50.

Aubert, H. (1886). Die Bewegungsempfindung. *Archiv für die gesamte Physiologie des Menschen und der Tiere*, 39(1):347–70.

Benson, A. J. (1990). Sensory functions and limitations of the vestibular system. In Warren, R. and Wertheim, A. H., editors, *The Perception and Control of Self-Motion*, book section 8, pages 145–70. Lawrence Erlbaum Associates, Hillsdale, NJ.

Benson, A. J., Kass, J. R., and Vogel, H. (1986). European vestibular experiments on the spacelab-1 mission: 4. thresholds of perception of whole-body linear oscillation. *Experimental Brain Research*, 64(2):264–71.

Berthoz, A., Pavard, B., and Young, L. R. (1975). Perception of linear horizontal self-motion induced by peripheral vision (linearvection) basic characteristics and visual-vestibular interactions. *Experimental Brain Research*, 23(5):471–89.

Bertolini, G. and Straumann, D. (2016). Moving in a moving world: A review on vestibular motion sickness. *Frontiers in Neurology*, 7:1–11.

Brandt, T., Dichgans, J., and Koenig, E. (1973). Differential effects of central verses peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research*, 16(5):476–91.

Brandt, T., Dieterich, M., and Strupp, M. (2005). *Various Vertigo Syndromes*, book section 6, pages 123–38. Springer London.

Bronstein, A. M. (2004). Vision and vertigo: some visual aspects of vestibular disorders. *Journal of Neurology*, 251(4):381–7.

Brooks, J. X. and Cullen, K. E. (2013). The primate cerebellum selectively encodes unexpected self-motion. *Current Biology*, 23(11):947–55.

Butler, J. S., Smith, S. T., Campos, J. L., and Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision*, 10(11):1–13.

Cabolis, K., Steinberg, A., and Ferre, E. R. (2018). Somatosensory modulation of perceptual vestibular detection. *Experimental Brain Research*, 236(3):859–65.

Campos, J. L. and Bülthoff, H. H. (2012). Multimodal integration during self-motion in virtual reality. In Murray, M. and Wallace, M., editors, *The Neural Bases of Multisensory Processes*, book section 30. CRC Press/Taylor & Francis, Boca Raton (FL).

Carpenter, R. H. S. (1988). *Movements of the eyes*. Pion, London, 2nd edition.

Carriot, J., Brooks, J. X., and Cullen, K. E. (2013). Multimodal integration of self-motion cues in the vestibular system: Active versus passive translations. *Journal of Neuroscience*, 33(50):19555–66.

Carriot, J., Jamali, M., Chacron, M. J., and Cullen, K. E. (2014). Statistics of the vestibular input experienced during natural self-motion: implications for neural processing. *Journal of Neuroscience*, 34(24):8347–57.

Colagiorgio, P., Bertolini, G., Bockisch, C. J., Straumann, D., and Ramat, S. (2015). Multiple timescales in the adaptation of the rotational vor. *Journal of Neurophysiology*, 113(9):3130–42.

Collewijn, H. and Smeets, J. B. (2000). Early components of the human vestibulo-ocular response to head rotation: latency and gain. *Journal of Neurophysiology*, 84(1):376–89.

Crowell, J. A., Banks, M. S., Shenoy, K. V., and Andersen, R. A. (1998). Visual self-motion perception during head turns. *Nature Neuroscience*, 1(8):732–7.

Cullen, K. E. (2011). The neural encoding of self-motion. *Current Opinion in Neurobiology*, 21(4):587–95.

Cullen, K. E. (2012). The vestibular system: multimodal integration and encoding of self-motion for motor control. *Trends in Neurosciences*, 35(3):185–96.

Cullen, K. E., Brooks, J. X., and Sadeghi, S. G. (2009). How actions alter sensory processing: reafference in the vestibular system. *Annals of the New York Academy of Sciences*, 1164:29–36.

Cullen, K. E. and Roy, J. E. (2004). Signal processing in the vestibular system during active versus passive head movements. *Journal of Neurophysiology*, 91(5):1919–33.

Day, B. L. and Fitzpatrick, R. C. (2005). The vestibular system. *Current Biology*, 15(15):R583–R586.

DeAngelis, G. C. and Angelaki, D. E. (2012). Visual-vestibular integration for self-motion perception. In Murray, M. M. and Wallace, M. T., editors, *The Neural Bases of Multisensory Processes*, book section 31. CRC Press/Taylor & Francis, Boca Raton FL.

Dichgans, J., Wist, E., Diener, H. C., and Brandt, T. (1975). The aubert-fleischl phenomenon: a temporal frequency effect on perceived velocity in afferent motion perception. *Experimental Brain Research*, 23(5):529–33.

Dodge, R. (1923). Thresholds of rotation. *Journal of Experimental Psychology*, 6(2):107.

Dyde, R. T. and Harris, L. R. (2008). The influence of retinal and extra-retinal motion cues on perceived object motion during self-motion. *Journal of Vision*, 8(14):5, 1–10.

Einhauser, W., Schumann, F., Bardins, S., Bartl, K., Boning, G., Schneider, E., and Konig, P. (2007). Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, 18(3):267–97.

Ernst, M. (2006). A Bayesian view on multimodal cue integration. In Knoblich, G., Thornton, I. M., Grosjean, M., and Shiffrar, M., editors, *Human Body Perception From the Inside Out*. Oxford University Press.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–33.

Ernst, M. O. and Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4):162–9.

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., and Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *Journal of Neuroscience*, 29(49):15601–12.

Fetsch, C. R., Wang, S., Gu, Y., Deangelis, G. C., and Angelaki, D. E. (2007). Spatial reference frames of visual, vestibular, and multimodal heading signals in the dorsal subdivision of the medial superior temporal area. *Journal of Neuroscience*, 27(3):700–12.

Filehne, W. (1922). Über das optische wahrnehmen von bewegungen. *Zeitschrift für Sinnesphysiologie*, 53:134–44.

Freeman, T. C. A., Champion, R. A., and Warren, P. A. (2010). A bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Current Biology*, 20(8):757–62.

Genzel, D., Firzlaff, U., Wiegrebe, L., and MacNeilage, P. R. (2016). Dependence of auditory spatial updating on vestibular, proprioceptive, and efference copy signals. *Journal of Neurophysiology*, 116(2):765–75.

Gibson, J. J. (1950). *The perception of the visual world*. Houghton Mifflin, Oxford, England.

Gibson, J. J. (1954). The visual perception of objective motion and subjective movement. *Psychological Review*, 61(5):304–14.

Gibson, J. J. (1968). What gives rise to the perception of motion? *Psychological Review*, 75(4):335–46.

Gillingham, K. and Previc, F. (1993). *Spatial Orientation in Flight*. Defense Technical Information Center.

Green, A. M., Shaikh, A. G., and Angelaki, D. E. (2005). Sensory vestibular contributions to constructing internal models of self-motion. *Journal of Neural Engineering*, 2(3):S164–79.

Gu, Y., Angelaki, D. E., and DeAngelis, G. C. (2008). Neural correlates of multi-sensory cue integration in macaque area mstd. *Nature Neuroscience*, 11(10):1201–10.

Guedry, F. E., J. (1964). Visual control of habituation to complex vestibular stimulation in man. *Acta Oto-Laryngologica*, 58:377–89.

Harris, L. R., Jenkin, M., and Zikovitz, D. C. (2000a). Vestibular capture of the perceived distance of passive linear self motion. *Archives Italiennes de Biologie*, 138(1):63–72.

Harris, L. R., Jenkin, M., and Zikovitz, D. C. (2000b). Visual and non-visual cues in the perception of linear self-motion. *Experimental Brain Research*, 135(1):12–21.

Hayhoe, M. and Rothkopf, C. A. (2011). Vision in the natural world. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(2):158–66.

Held, R. (1961). Exposure-history as a factor in maintaining stability of perception and coordination. *The Journal of Nervous and Mental Disease*, 132:26–32.

Hettinger, L. J., Berbaum, K. S., Kennedy, R. S., Dunlap, W. P., and Nolan, M. D. (1990). Vection and simulator sickness. *Military Psychology*, 2(3):171–81.

Hillis, J. M., Ernst, M. O., Banks, M. S., and Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, 298(5598):1627–30.

Hlavacka, F., Mergner, T., and Schweigart, G. (1992). Interaction of vestibular and proprioceptive inputs for human self-motion perception. *Neuroscience Letters*, 138(1):161–4.

Jaekl, P. M., Jenkin, M. R., and Harris, L. R. (2005). Perceiving a stable world during active rotational and translational head movements. *Experimental Brain Research*, 163(3):388–99.

Jurgens, R. and Becker, W. (2006). Perception of angular displacement without landmarks: evidence for bayesian fusion of vestibular, optokinetic, podokinesthetic, and cognitive information. *Experimental Brain Research*, 174(3):528–43.

Jurgens, R. and Becker, W. (2011). Human spatial orientation in non-stationary environments: relation between self-turning perception and detection of surround motion. *Experimental Brain Research*, 215(3-4):327–44.

Kandel, E., Schwartz, J., and Jessell, T. (2000). *Principles of Neural Science, Fourth Edition*. McGraw-Hill Education.

Kars, H. J. J., Hijmans, J. M., Geertzen, J. H. B., and Zijlstra, W. (2009). The effect of reduced somatosensation on standing balance: A systematic review. *Journal of diabetes science and technology (Online)*, 3(4):931–43.

Kavounoudias, A., Roll, R., and Roll, J. P. (1998). The plantar sole is a 'dynamometric map' for human balance control. *Neuroreport*, 9(14):3247–52.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–20.

Kingdom, F. A. A. and Prins, N. (2010). *Psychophysics: A Practical Introduction*. Elsevier, Amsterdam [u.a.].

Knill, D. C. and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–9.

Lackner, J. R. (1977). Induction of illusory self-rotation and nystagmus by a rotating sound-field. *Aviation, Space, and Environmental Medicine*, 48(2):129–31.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63(8):1279–92.

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11):1432–8.

MacNeilage, P. R., Banks, M. S., Berger, D. R., and Bulthoff, H. H. (2007). A bayesian model of the disambiguation of gravitoinertial force by visual cues. *Experimental Brain Research*, 179(2):263–90.

McCauley, M. E. and Sharkey, T. J. (1992). Cybersickness: perception of self-motion in virtual environments. *Presence: Teleoperators and Virtual Environments*, 1(3):311–8.

McCrea, R. A., Gdowski, G. T., Boyle, R., and Belton, T. (1999). Firing behavior of vestibular neurons during active and passive head movements: vestibulo-spinal and other non-eye-movement related neurons. *Journal of Neurophysiology*, 82(1):416–28.

Medendorp, W. P. and Selen, L. J. P. (2017). Vestibular contributions to high-level sensorimotor functions. *Neuropsychologia*, 105:144–52.

Merfeld, D., Park, S., Gianna-Poulin, C., Black, F. O., and Wood, S. (2005). Vestibular perception and action employ qualitatively different mechanisms. ii. vor and perceptual responses during combined tilt&translation. *Journal of Neurophysiology*, 94(1):199–205.

Merfeld, D., Zupan, L. H., and Peterka, R. J. (1999). Humans use internal models to estimate gravity and linear acceleration. *Nature*, 398(6728):615–18.

Mergner, T., Siebold, C., Schweigart, G., and Becker, W. (1991). Human perception of horizontal trunk and head rotation in space during vestibular and neck stimulation. *Experimental Brain Research*, 85(2):389–404.

Mittelstaedt, H. (1992). Somatic versus vestibular gravity reception in man. *Annals of the New York Academy of Sciences*, 656:124–39.

Mittelstaedt, H. (1997). Interaction of eye-, head-, and trunk-bound information in spatial perception and control. *Journal of Vestibular Research*, 7(4):283–302.

Money, K. E. (1970). Motion sickness. *Physiological Reviews*, 50(1):1–39.

Mulavara, A. P., Houser, J., Miller, C., and Bloomberg, J. J. (2005). Full-body gaze control mechanisms elicited during locomotion: effects of vor adaptation. *Journal of Vestibular Research*, 15(5-6):279–89.

Nashner, L. M., Black, F. O., and Wall, C., r. (1982). Adaptation to altered support and visual conditions during stance: patients with vestibular deficits. *Journal of Neuroscience*, 2(5):536–44.

Oman, C. M. (1982). A heuristic mathematical model for the dynamics of sensory conflict and motion sickness. *Acta Oto-Laryngologica. Supplementum*, 392:1–44.

Oman, C. M. (1990). Motion sickness: a synthesis and evaluation of the sensory conflict theory. *Canadian Journal of Physiology and Pharmacology*, 68(2):294–303.

Oman, C. M. (1991). Sensory conflict in motion sickness: an observer theory approach. In Ellis, S., editor, *Pictorial communication in real and virtual environments*, pages 362–76. CRC Press/Taylor & Francis, London.

Oman, C. M. and Cullen, K. E. (2014). Brainstem processing of vestibular sensory exafference: implications for motion sickness etiology. *Experimental Brain Research*, 232(8):2483–92.

Prins, N. and Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data.

Prsa, M., Gale, S., and Blanke, O. (2012). Self-motion leads to mandatory cue fusion across sensory modalities. *Joural of Neurophysiology*, 108(8):2282–91.

Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922.

Raymond, J. E., Shapiro, K. L., and Rose, D. J. (1984). Optokinetic backgrounds affect perceived velocity during ocular tracking. *Perception & Psychophysics*, 36(3):221–4.

Reason, J. and Brand, J. J. (1975). *Motion sickness.* Academic Press, London.

Reason, J. T. (1970). Motion sickness: a special case of sensory rearrangement. *Advanced Science*, 26(130):386–93.

Reason, J. T. (1978). Motion sickness adaptation: a neural mismatch model. *Journal of the Royal Society of Medicine*, 71(11):819–29.

Roitman, J. D. and Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22(21):9475–89.

Roy, J. E. and Cullen, K. E. (2002). Vestibuloocular reflex signal modulation during voluntary and passive head movements. *Journal of Neurophysiology*, 87(5):2337–57.

Roy, J. E. and Cullen, K. E. (2004). Dissociating self-generated from passively applied head motion: neural mechanisms in the vestibular nuclei. *Journal of Neuroscience*, 24(9):2102–11.

Sadeghi, S. G., Mitchell, D. E., and Cullen, K. E. (2009). Different neural strategies for multimodal integration: comparison of two macaque monkey species. *Experimental Brain Research*, 195(1):45–57.

Sakamoto, S., Suzuki, F., Suzuki, Y., and Gyoba, J. (2008). The effect of linearly moving sound image on perceived self-motion with vestibular information. *Acoustical Science and Technology*, 29(6):391–93.

Shams, L. (2012). Early integration and bayesian causal inference in multisensory perception. In Murray, M. and Wallace, M., editors, *The Neural Bases of Multisensory Processes*, book section 12. CRC Press/Taylor & Francis, Boca Raton (FL).

Shams, L. and Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Science*, 14(9):425–32.

Sherrington, C. S. (1918). Observations on the sensual role of the proprioceptive nerve-supply of the extrinsic ocular muscles. *Brain: A Journal of Neurology*, 41:332–42.

Smith, P. L. and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3):161–8.

Solomon, D., Zee, D. S., and Straumann, D. (2003). Torsional and horizontal vestibular ocular reflex adaptation: three-dimensional eye movement analysis. *Experimental Brain Research*, 152(2):150–5.

Sommer, M. A. and Wurtz, R. H. (2002). A pathway in primate brain for internal monitoring of movements. *Science*, 296(5572):1480–2.

Sommer, M. A. and Wurtz, R. H. (2008). Brain circuits for the internal monitoring of movements. *Annual Review of Neuroscience*, 31:317–38.

Souman, J. L., Hooge, I. T. C., and Wertheim, A. H. (2006). Frame of reference transformations in motion perception during smooth pursuit eye movements. *Journal of Computational Neuroscience*, 20(1):61–76.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, 43(6):482–9.

Steele, J. E. (1961). Motion sickness and spatial perception. a theoretical study. *ASD Techical Report*, 61(530):1–23.

Telford, L., Howard, I. P., and Ohmi, M. (1995). Heading judgments during active and passive self-motion. *Experimental Brain Research*, 104(3):502–10.

ter Horst, A. C., Koppen, M., Selen, L. P., and Medendorp, W. P. (2015). Reliability-based weighting of visual and vestibular cues in displacement estimation. *PLoS One*, 10(12):1–15.

Treisman, M. (1977). Motion sickness: an evolutionary hypothesis. *Science*, 197(4302):493–5.

Trousselard, M., Barraud, P. A., Nougier, V., Raphel, C., and Cian, C. (2004). Contribution of tactile and interoceptive cues to the perception of the direction of gravity. *Brain Research. Cognitive Brain Research*, 20(3):355–62.

van Beers, R. J., Sittig, A. C., and Gon, J. J. D. v. d. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, 81(3):1355–64.

von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Voss, Leipzig.

von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *The British Journal of Animal Behaviour*, 2(3):89–94.

von Holst, E. and Mittelstaedt, H. . (1950). Das reafferenzprinzip - (wechselwirkungen zwischen zentralnervensystem und peripherie). *Naturwissenschaften*, 37:464–76.

von Hopffgarten, A. and Bremmer, F. (2011). Self-motion reproduction can be affected by associated auditory cues. *Seeing Perceiving*, 24(3):203–22.

Wallach, H. (1985). Perceiving a stable environment. *Scientific American*, 252(5):118–125.

Warren, P. A. and Rushton, S. K. (2007). Perception of object trajectory: Parsing retinal motion into self and object movement components. *Journal of Vision*, 7(11):1–11.

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6):598–604.

Wertheim, A. (1994). Motion perception during self-motion: The direct versus inferential controversy revisited. 17(2):293–311.

Wexler, M. (2003). Voluntary head movement and allocentric perception of space. *Psychological Science*, 14(4):340–6.

Witkin, H. A. (1949). Perception of body position and of the position of the visual field. *Psychological Monographs: General and Applied*, 63(7):1–46.

Wurtz, R. H. and Sommer, M. A. (2004). Identifying corollary discharges for movement in the primate brain. *Progress in Brain Research*, 144:47–60.

Young, L. (1970). On visual-vestibular interaction. In NASA, editor, *5th Symposium: The role of the vestibular organs in space exploration*, volume 314, pages 205–10. NASA, Pensacola.

Zupan, L. H., Merfeld, D. M., and Darlot, C. (2002). Using sensory weighting to model the influence of canal, otolith and visual cues on spatial orientation and eye movements. *Biological Cybernetics*, 86(3):209–30.

Acknowledgments

I am very grateful for the support and help I got during my PhD. First of all, I would like to thank my supervisor Prof. Dr. Paul MacNeilage for giving me the great opportunity to work in his lab on a broad variety of interesting projects. I greatly appreciate his scientific advice and guidance as well as the constructive discussions and collaboration throughout my PhD.

I would also like to thank the members of my thesis advisory committee Prof. Dr. Heiner Deubel and Prof. Dr. Stefan Glasauer for their time and support. Thanks a lot also to my collaborators Prof. Dr. Tom Freeman and Prof. Dr. Marc Ernst for the fruitful discussion on the second manuscript.

I would like to thank the GSN not only for the financial support during the many conferences and lab visits I attended but also for the broad scientific and cultural inspiration and most importantly the amazing support by the GSN team whenever needed.

Thanks a lot to the Forschungshaus people with whom it was a lot of fun to work together. Special thanks to Kathrin Kostorz, Judita Huber, Nisha Dalal and Dr. Alicia Costalago Meruelo who are always there to listen and give me emotional support. Thanks also to Dr. Vivian Holten for being a great office colleague and good collaborator. Thanks also to Alex Knorr for all his help and advice and the great collaboration.

I would like to thank Dr. Alex Loebel for his scientific support and deep conversations.

I would especially like to thank Annika Gawliczek and Marius Gebauer for being such amazing friends who always stand by me, also during challenging times.

Eidesstattliche Versicherung/Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation "Perceiving a stable world during self-motion: measurement and modeling of visual-vestibular integration and segregation" selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation "Perceiving a stable world during self-motion: measurement and modeling of visual-vestibular integration and segregation" is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München, 01.08.2018

Isabelle T. Garzorz

Author contributions

Garzorz, Isabelle T., MacNeilage Paul R. (2017). Visual-vestibular conflict detection depends on fixation. Current Biology, 27:2856-2861, doi: 10.1016/j.cub.2017.08.011
I.T.G. and P.R.M. designed the study; I.T.G. collected the data; I.T.G. and P.R.M. analyzed the data, wrote and revised the paper.

Garzorz, Isabelle T., Freeman, Tom C.A., Ernst, Marc O., MacNeilage, Paul R. (2018). Insufficient compensation for self-motion during perception of object speed: The vestibular Aubert-Fleischl phenomenon. Journal of Vision, 18(13):9, 1–9, doi: 10.1167/18.13.9
I.T.G. and P.R.M. designed the study; I.T.G. collected the data; I.T.G. and P.R.M. analyzed the data; I.T.G., P.R.M., T.C.F and M.O.E. wrote the paper.

Moroz, Matthew, Garzorz, Isabelle T., Folmer, Eelke, and MacNeilage, Paul R. (currently in press). Sensitivity to visual gain modulation in head-mounted displays depends on fixation. Displays, doi: 10.1016/j.displa.2018.09.001
M.M., I.T.G., E.F., and P.R.M. designed the study; M.M. collected the data; M.M. and P.R.M. analyzed the data; M.M., I.T.G., E.F., and P.R.M. wrote and revised the paper.