Published in final edited form as:

Prog Brain Res. 2019 ; 248: 277–284. doi:10.1016/bs.pbr.2019.03.018.

Towards dynamic modeling of visual-vestibular conflict detection

Isabelle T. Garzorz^{a,b,*}, Paul R. MacNeilage^c

^aGerman Center for Vertigo and Balance Disorders, University Hospital of Munich, Munich, Germany

^bGraduate School of Systemic Neurosciences, Ludwig-Maximilian University, Munich, Germany

^cDepartment of Psychology, Cognitive and Brain Sciences, University of Nevada, Reno, NV, United States

Abstract

Visual-vestibular mismatch is a common occurrence, with causes ranging from vehicular travel, to vestibular dysfunction, to virtual reality displays. Behavioral and physiological consequences of this mismatch include adaptation of reflexive eye movements, oscillopsia, vertigo, and nausea. Despite this significance, we still do not have a good understanding of how the nervous system evaluates visual-vestibular conflict. Here we review research that quantifies perceptual sensitivity to visual-vestibular conflict and factors that mediate this sensitivity, such as noise on visual and vestibular sensory estimates. We emphasize that dynamic modeling methods are necessary to investigate how the nervous system monitors conflict between time-varying visual and vestibular signals, and we present a simple example of a drift-diffusion model for visual-vestibular conflict detection. The model makes predictions for detection of conflict arising from changes in both visual gain and latency. We conclude with discussion of topics for future research.

Keywords

Optic flow; Vestibular; Drift-diffusion; Decision making; Conflict; Perception; Psychophysics; Oculomotor; Oscillopsia; Perceptual stability

1 Background

Under normal circumstances, self-motion signaled by the vestibular system agrees with self-motion signaled by vision, and the result is perception of a stable visual environment. However, some common situations can result in visual-vestibular conflict and perceptual instability. For example, riding in a vehicle while fixating the interior of the vehicle leads to conflict because the vestibular system signals how the head is moving in space while the visual system provides information about head movement relative to the vehicle. Likewise, modern virtual reality devices can introduce visual self-motion cues that do not correspond

*Corresponding author: isabelle.garzorz@lrz.uni-muenchen.de.

to head movement signaled by the vestibular system. Finally, vestibular dysfunction leading to aberrant vestibular signals can also lead to visual-vestibular mismatch.

The perceptual instability that results from this conflict is known as oscillopsia; it seems as if the environment is not stable but rather in motion. This instability is governed by neural processes, likely circuits in the vestibular nuclei and cerebellum, that must compare visual and vestibular signals to evaluate their agreement. This comparison process underlies not only perceptual stability, but also calibration of reflexive eye movements, such as the vestibulo-ocular reflex, as well as generation of adaptive postural responses.

The research presented below is concerned with the question of how the nervous system evaluates the agreement between visual and vestibular signals. While the question may be addressed with physiological methods, such as measurement of reflexive eye movements or single-unit activity, we instead rely on psychophysical methods developed to evaluate perceptual instability in the face of visual-vestibular conflict. These methods have the advantage that they are relatively easy to undertake, and yet they provide a measure of conflict, or visual-vestibular mismatch error, at the end of every trial.

2 Perceptual measures of conflict detection

The first studies to investigate perceptual sensitivity to visual-vestibular conflict were conducted by Wallach and Kravitz (1965). A mismatch between head movement and visual scene motion was introduced through a mechanical device that was driven by yaw head rotation. The participant wore a harness on the head and head movement was transduced into movement of the visual scene through a series of gears with an adjustable gear ratio. Subjects made active head turns while fixating the visual scene and they indicated if they perceived the scene to be moving or not. Experimenters identified the ratio, or gain, of scene movement relative to head movement that was just noticeable. They referred to the intermediate range of gains that resulted in perception of a stable visual environment as the range of immobility. In various studies, they determined that this range of gains was ~ 0.97 – 1.03 (Wallach, 1987; Wallach and Kravitz, 1965).

Subsequent research by Harris and colleagues (Jaekl et al., 2005) used similar methods. They used a mechanical device to track active head movements and the visual scene motion was rendered on a head-mounted visual display (HMD). This study was significant because they measured perceptual stability for both translation and rotation of the head. The ranges of immobility that were reported were overall much larger than reported by Wallach and colleagues. This could be partially because they used a digital display with a motion-to-photons latency of 122 ms.

A few years later, additional studies were conducted by Jerald et al. (2012), but with optical head tracking and with visual stimuli rendered on an external projection screen, but viewed through a head-mounted aperture to match the field of view in a head-mounted display. They report that people were more sensitive to conflict when (1) visual scene motion was against, rather than with, head motion, and (2) head movement was slow. More recently, a study was conducted in which the participants were moved passively, and only linear movements were

presented (Correia Gracio et al., 2014). They report large ranges of immobility and found that gains much greater than 1 were required for scenes to be perceived as stationary. This could be due to underestimation of depth in the virtual scene, leading to underestimation of speed from optic flow.

Other studies by Ellis et al. (2004) and Mania et al. (2004) investigated visual vestibular conflict in head-mounted displays, but focus specifically on the question of sensitivity to latency, or delay, of the visual scene motion relative to head motion. They asked subjects to indicate if there was a difference between two consecutive intervals, one with a constant pedestal latency, and another with a variable increment in latency on top of that pedestal. They found that a latency increment on the order of 25 ms leads to a 50% detection rate and that this threshold stays constant regardless of the pedestal latency. They suggest detection depends on magnitude of retinal slip.

While this prior research has quantified sensitivity to visual scene motion under various conditions, there has been little or no attempt to model the factors mediating conflict detection.

3 A static model of conflict detection

We model conflict detection as crossmodal discrimination such that the noise limiting discrimination performance (σ_{discr}), and thus the range of immobility, is dependent on the noise on both visual (σ_{vis}) and vestibular (σ_{vest}) motion estimates; $[\sigma_{discr}^2 = \sigma_{vis}^2 + \sigma_{vest}^2]$ (Fig. 1A). We test the model using methods similar to those described above. In addition, we examine the dependence of conflict detection on eye movements. We quickly summarize our methods and findings here; the reader is referred to published manuscripts (Garzorz and MacNeilage, 2017; Moroz et al., 2018) for further detail.

In order to test the crossmodal discrimination model, it is necessary to measure the noise on visual and vestibular estimates. These noises were measured in two interval-forced-choice (2IFC) experiments in which participants were presented with two consecutive yaw rotations (0.8 s duration each) and asked to judge which was larger (displacement), faster (Velocity), and stronger (acceleration). Displacement velocity and acceleration scale together, so discrimination could be based on any of these. In the vestibular condition, participants seated on a motion platform were physically rotated about their own axis two consecutive times while blindfolded. In the visual condition, subjects were physically stationary, but viewed two consecutive epochs of optic flow consistent with yaw rotation. Displacement for one rotation was always 4 degrees (~10 degrees/s peak velocity). Displacement in the other interval was varied from trial to trial according to a staircase procedure to find the increment/decrement that was just noticeable. A cumulative Gaussian was fit to the psychometric data and the standard deviation of the fit provided a measure of visual or vestibular noise (Fig. 1B and C).

The crossmodal discrimination model was tested in two conditions in which visual and vestibular rotation stimuli had to be compared with one another, but they were presented either sequentially or simultaneously. The sequential task was a control condition and was

identical to the 21FC task in the visual and vestibular condition. The simultaneous task, on the other hand, was the critical condition for evaluating the crossmodal discrimination model. It required subjects to judge if the simultaneously presented visual scene motion was too fast, in which case it should appear to move against the observer motion, or too slow, in which case it should appear to move with the observer. This is a task similar to that employed by Wallach and Kravitz (1965) and others and allows us to measure the range of immobility. As above, the standard deviation of cumulative Gaussian fits (Fig. 1D and E) provided a measure of noise on crossmodal discrimination (σ_{discr}) that could be compared with the sum of visual and vestibular noises ($\sigma_{vis}^2 + \sigma_{vest}^2$) measured in the visual and vestibular conditions.

Results suggest that crossmodal discrimination is a reasonable model for visual-vestibular conflict detection. In other words, the range of immobility depends on visual and vestibular noise. But this was only true when participants tracked the visual scene with their eyes. When participants instead fixated a point that moved with the head, in effect nulling eye movements, they were more insensitive to conflict. This suggests that oculomotor behavior plays a crucial role in conflict detection such that conflict detection may be mediated by retinal slip, i.e., the mismatch between compensatory eye movements and head motion (Garzorz and MacNeilage, 2017). This advantage for conflict detection during scene-fixed compared to head-fixed fixation was also confirmed during active, rather than passive, head movement (Moroz et al., 2018).

4 Dynamic modeling of visual-vestibular conflict detection

The simple model presented above is static in the sense that it depends on singular values for noise on visual and vestibular estimates for a specific trajectory of head movement. However, given that these are time-varying signals, a dynamic model of conflict detection is warranted. A popular model for decision making in dynamic contexts is the drift-diffusion model (Ratcliff, 1978; Ratcliff and McKoon, 2008; Smith and Ratcliff, 2004), in which evidence in favor of two alternative decisions is accumulated over time until a decision boundary is reached. Below we illustrate how such a model could be adapted to the context of visual-vestibular conflict detection, and we continue to discuss the advantages of such an approach over the previously proposed static model.

The model is built on the same decision framework described for the static model; the observer must decide whether the visual scene motion is too slow or too fast, or equivalently, whether the visual scene is moving with or against their own motion. As above, visual and vestibular estimates (Fig. 2A and B) must be compared, but here the comparison occurs continually over time. An error signal is calculated at each time step (Fig. 2C and D), and this error is accumulated over time (Fig. 2E and F). If the accumulated error becomes large enough, it crosses a bound (dashed line), at which time a decision can be made in favor of that alternative.

The model can be considered an iterative version of the static crossmodal discrimination model in that noise on the error signal at each time step depends on noise on the visual and vestibular estimates at that time step. The effect of noise can be seen in Fig. 2E and F where

the rate of accumulation to the bound varies somewhat depending on moment-to-moment noise on both visual and vestibular estimates. As a consequence of this noise, there is variability on the decision that is reached, even following presentation of identical stimuli. On some trials, the accumulated evidence crosses the bound (dashed line), resulting in a decision in favor of conflict (i.e., visual stimulus too slow or lagging), whereas on other trials insufficient evidence is accumulated and the conflict remains below the detection threshold. This captures natural variability in conflict detection, even in response to identical visual and vestibular stimuli.

The model is versatile in that it can be used to model mismatch introduced by visual gain modulation (Fig. 2, left), and also mismatch resulting from delayed visual motion (i.e., phase difference), which is common in virtual reality systems (Fig. 2, right). Interestingly, this dynamic model makes different predictions about when the conflict will be most noticeable in these two scenarios. For visual gain modulation, the most evidence is accumulated by the end of the trajectory, whereas for delayed visual motion, the most evidence is accumulated in the middle. These predictions are unique to the dynamic conflict detection model and could be tested experimentally, for example, by blanking the visual stimulus at various timepoints and measuring the impact on conflict detection performance.

5 Discussion

The dynamic conflict detection model presented above is one rather simple example of a broad range of models that could be constructed to investigate how the nervous system evaluates visual-vestibular conflict. Numerous other approaches have been proposed to model dynamic perceptual decision making in other contexts, and significant experimental work has attempted to differentiate among these models (see Merfeld et al., 2016 for review).

Notably, the only dynamic element of the model proposed here is the accumulation/integration of evidence. Since the error signal is velocity-based, the decision variable is proportional to positional error (i.e., integrated velocity). In this way, the model is equivalent to one that detects conflict whenever positional retinal slip exceeds a given bound, depicted here as 2 degrees (Fig. 2E and F). It remains an open question to what extent perception of visual-vestibular conflict depends on a position- versus velocity-based error.

This question about the nature of the error signal used to evaluate visualvestibular conflict has been addressed in the context of VOR adaptation. One recent study (Shin et al., 2014) suggests that VOR adaptation is best-predicted by eye and head velocity and their sum or difference and is not well-predicted by retinal slip per se. Similar studies are needed to evaluate the error signals driving perception. Given the known dependence of both oculomotor (e.g., Voogd and Barmack, 2006) and perceptual (e.g., Baumann et al., 2015) processes on cerebellar regions where visual, oculomotor, and vestibular signals are known to converge, it is parsimonious to hypothesize that the same or similar error signals are driving both adaptation of reflexive eye movements as well as perception, but this hypothesis remains to be tested.

Acknowledgment

This research was supported by NIH P20GM103650 and BMBF 01 EO 1401.

References

- Baumann O, Borra RJ, Bower JM, Cullen KE, Habas C, Ivry RB, Leggio M, Mattingley JB, Molinari M, Moulton EA, Paulin MG, Pavlova MA, Schmähmann JD, Sokolov AA, 2015 Consensus paper: the role of the cerebellum in perceptual processes. *Cerebellum* 14, 197–220. [PubMed: 25479821]
- Correia Gracio BJ, Bos JE, Van Paassen MM, 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. [PubMed: 24292492]
- Elli SR, Mani K, Adelstein BD, Hil MI, 2004. Generalizability of latency detection in a variety of virtual environments. *Proc. Hum. Factors Ergon. Soc. Annu. Meet* 48, 2632–2636.
- Garzorz IT, MacNeilage PR, 2017 Visual-vestibular conflict detection depends on fixation. *Curr. Biol.* 27, 2856–2861.E4. [PubMed: 28889981]
- Jaekl PM, Jenkin MR, Harris LR, 2005 Perceiving a stable world during active rotational and translational head movements. *Exp. Brain Res.* 163, 388–399. [PubMed: 15856212]
- Jerald J, Whitton M, Brooks FP, 2012 Scene-motion thresholds during head yaw for immersive virtual environments. *ACM Trans. Appl. Percept* 9, 4. [PubMed: 25705137]
- Mania K, Adelstein B, Ellis S, Hill IM, 2004 Perceptual Sensitivity to Head Tracking Latency in Virtual Environments With Varying Degrees of Scene Complexity.
- Merfeld DM, Clark TK, Lu YM, Karmali F, 2016 Dynamics of individual perceptual decisions. *J. Neurophysiol.* 115, 39–59.
- Moroz M, Garzorz I, Folmer E, MacNeilage P, 2018 Sensitivity to visual gain modulation in head-mounted displays depends on fixation. In: *Displays*.
- Ratcliff R, 1978 A theory of memory retrieval. *Psychol. Rev.* 85, 59–108.
- Ratcliff R, McKoon G, 2008 The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20, 873–922. [PubMed: 18085991]
- Shin SL, Zhao GQ, Raymond JL, 2014 Signals and learning rules guiding oculomotor plasticity. *J. Neurosci.* 34, 10635–10644. [PubMed: 25100597]
- Smith PL, Ratcliff R, 2004 Psychology and neurobiology of simple decisions. *Trends Neurosci.* 27, 161–168. [PubMed: 15036882]
- Voogd J, Barmack NH, 2006 Oculomotor cerebellum. *Prog. Brain Res.* 151, 231–268. [PubMed: 16221591]
- Wallach H, 1987 Perceiving a stable environment when one moves. *Annu. Rev. Psychol.* 38, 1–27. [PubMed: 3548572]
- Wallach H, Kravitz JH, 1965 The measurement of the constancy of visual direction and of its adaptation. *Psychol. Sci.* 2, 217–218.

FIG. 1.

Static model of visual-vestibular conflict detection. (A) Conflict detection is modeled as crossmodal discrimination. It depends on the estimate of the discrepancy (red) between visual (blue) and vestibular (green) signals, which has variability (σ_{discr}) equal to the sum of variability on the visual (σ_{vis}) and vestibular (σ_{vest}) estimates. (B–E) Examples of data and cumulative Gaussian psychometric fits for an individual subject in the visual (B), vestibular (C), sequential (D), and simultaneous (E) conditions from Garzorz and MacNeilage (2017). The standard deviation of each fit provides a measure of variability that

can be used to test the model. Inset icons illustrate the experimental paradigm for each condition.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIG. 2.

Dynamic model of visual-vestibular conflict detection. The figure illustrates the cases of conflict introduced by manipulating the gain (left) and latency (right) of the visual stimulus. (A,B) Examples of time-varying estimates of visual and vestibular velocity. These estimates will be noisy, but noise is not shown. (C,D) The error signal is calculated by subtracting visual and vestibular estimates at each time step. Noise from visual and vestibular signals will propagate; noise is not shown. (E,F) The accumulated decision variable, calculated by

integrating the error signal. Shown are simulated responses for 20 trials with noise. Dashed horizontal line depicts the decision bound.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript