

LEVERHULME  
TRUST \_\_\_\_\_



# Perceptual Compensation For Self-Movement and Object Distance in Vision and Hearing

Joshua David Haynes

Thesis submitted in partial fulfilment of the requirements  
for the degree of Doctor of Philosophy

2023

Supervisors: Prof. Tom C.A. Freeman, Prof. John F. Culling, Dr. Maria Gallagher

# Thesis Summary

The overarching theme of this thesis is investigating perceptual compensation while determining object movements. The two forms of compensation highlighted here are compensation for self-movement, as self-movement creates reafferent motion in the image that must be interpreted when determining the movement of objects, and auditory speed constancy, or the compensation for object distance that we are able to perform when determining object movement. A new paradigm is introduced and used throughout most of the thesis to investigate the precision of the image and non-image signals that we use during compensation for self-controlled self-movements, with the complications of allowing participants to control the stimulus accounted for in a new psychometric model that includes an external source of variability not present in standard cumulative Gaussian fits. The precision of the non-image signals is a main focus throughout, with discussion surrounding the finding that non-image signal precision depends on the modality of the image, that the non-image signal follows Weber's law while image signals do not, and that the standard Bayesian model of movement perception is not relevant in the context of self-controlled head-rotations. These "anti-Bayesian" results are verified quantitatively with proof that a Bayesian model derived here, that can account for the stimuli being participant controlled, is a less good fit than the psychometric model mentioned above. The first investigation into auditory speed constancy is presented in this thesis, with results suggesting that individual differences in distance perception underly the incomplete speed constancy that is found. These findings are summarised in the context of using compensation to interpret the movement of ourselves and visual and auditory objects at different distances from us.

# Acknowledgements

“We can write a thesis”. This thesis would not have been possible without the help of so many incredible people. Firstly, I would like to thank my supervisors: Tom, John and Maria. Your belief, support and wisdom were invaluable in this process and I could not have done it without you. Thank you for pushing me when I needed to be pushed and correcting me when I needed to be corrected. Next, my parents and my sister. Thank you for your unconditional love and for being there for me whenever I needed a helping hand. Thank you also for instilling a “never give up” attitude in me from a young age. To the lovely PhD students and RAs I met along the way: firstly, Alice. Thank you for being the perfect next-desk-neighbour. Sorry for eating sugary cereal and not communicating well sometimes but I hope you know that I really would not have made it to this point without your unbreaking positivity towards me and my work. We will always share a desk. Charlie, Nathan, Lauren, Emma, Rebecca, Tyler, and so many more. Thank you for the pub trips and bowling games and holidays, you have all been so helpful in picking me up on days when I needed it and you have been a huge part of what has made this experience so special. I know that all of you will do brilliantly in your careers. To Archie, Liam and the staff at Tiny Rebel Cardiff. Thank you for helping me to be able to achieve this. I am so grateful for the opportunity that you gave me and I’m glad that I met and worked with each and every one of you. And last but not least, to my other friends: firstly, Jamie. Thank you for uprooting your life and moving to a different country (not a principality) to live with me in the middle of a global pandemic. I will always be grateful for the risk you took and I am so glad that it paid off for us both. Iestyn. Thank you for allowing me into your home when I had nowhere else to go. You supported me when I really needed it and I will never forget that. Isaac, Dale, Sumayah, Ben, Rhian, Hannah, Tom, Joe, Adam and so many more. Thank you for always being there for me if I needed to talk. I know that you all have my back and I hope that you know that I have each and every one of yours.

Every person in this list and so many others that I have not had time to name have done their part in writing this thesis. The words on these pages may be mine but make no mistake, together, we wrote this thesis.

# Contents

General Introduction.....	7
1.1 – Why is Interpreting Motion Important?.....	7
1.2 – Compensation For Object Distance.....	8
1.3 – Compensation During Self-Movement.....	10
1.4 – Compensation During Self-Movement is Imperfect .....	11
1.5 – Models of Compensation During Self-Movement .....	13
1.6 – Compensation in Hearing.....	18
1.7 – Using Precision to Explain Bias.....	19
1.8 – Measuring the Precision of Active Self-Movement .....	23
1.9 – Summary .....	25
The Precision of the Non-Image Signal.....	26
2.0 - Preface .....	26
2.1 – Introduction .....	26
2.2 – Methods.....	30
2.3 – Results.....	36
2.4 – Discussion.....	38
2.5 – Appendix .....	41
Investigating Weber’s Law in the Non-Image Signal.....	47
3.0 – Preface.....	47
3.1 – Introduction .....	47
3.2 – Methods.....	50
3.3 – Results.....	52
3.4 – Discussion.....	59

Can a Bayesian Model Explain Self-Movement Compensation Errors in the Context of Self-Controlled Head Rotations? .....	64
4.0 – Preface.....	64
4.1 – Introduction .....	65
4.2 – Methods .....	68
4.3 – Results .....	71
4.4 – Discussion.....	75
Derivation of a Bayesian Model Which Accounts for Trial-by-Trial Variation In Stimuli .....	81
5.0 – Preface.....	81
5.1 – Derivation.....	82
5.2 – Fitting Procedure .....	90
5.3 – Summary .....	94
5.4 – Appendix .....	94
Investigating Auditory Speed Constancy.....	97
6.0 – Preface.....	97
6.1 - Introduction .....	97
6.2 – Experiment 1 – Investigating Auditory Depth Perception.....	104
6.2.1 – Methods .....	104
6.2.2 – Results .....	107
6.3 – Experiment 2 – Investigating Auditory Speed Constancy .....	109
6.3.1 – Methods .....	109
6.3.2 – Results .....	112
6.4 – Discussion.....	114
General Discussion .....	117
7.1 – Summary of Findings.....	117
7.2 – Relation to Previous Work.....	121

7.3 – Future Directions.....	124
7.4 – Thesis Conclusions .....	125
Thesis Appendix .....	127
FitSMMModel.....	127
BayesianPh1Fit.....	131
Reference List.....	134

# General Introduction

## 1.1 – Why is Interpreting Motion Important?

Understanding motion in the images that we obtain, and movement of ourselves and objects in the world, is an integral component of our perceptual experience. Whether we are crossing a road, catching a ball, or threading a needle, being able to interpret the movements of objects in the world and how our own movement can interact with them, is an essential skill. We rarely observe movement passively with our perceptual systems, because our receptive organs (e.g., eyes and ears) also move. Our own movements create reafferent motion in the images captured by these receptors, which needs to be interpreted correctly in order for us to accurately and precisely judge the movement of objects. One method for accounting for our self-movement is to use 'non-image' signals such as proprioceptive, vestibular, and muscular cues that contain information about our self-movement, while it is also often possible to use the information residing in the reafferent motion itself. These methods will be introduced in more detail below, however, no matter the chosen method, understanding the consequences of our self-movements and compensating for them is necessary for our perception of movement.

Some researchers believe that self-movement is the only reason that the brain evolved, with the life cycle of a sea squirt often cited in support of this theory (Glenberg et al., 2007; Llinás, 2002). Sea squirts begin life with the ability to actively move and a functioning brain but, when they find a suitable lifelong home, they digest their brain matter and remain stationary for the rest of their lives. Clearly, then, self-movement is an important part of our existence as it necessitates the development of an organ with the intricacies of our brain. At least part of this necessity must come from the need to understand the consequences of our self-movement. This thesis will explore the solutions that we employ, involving the art of compensation, to measure and interpret both the movement of ourselves, and of objects.

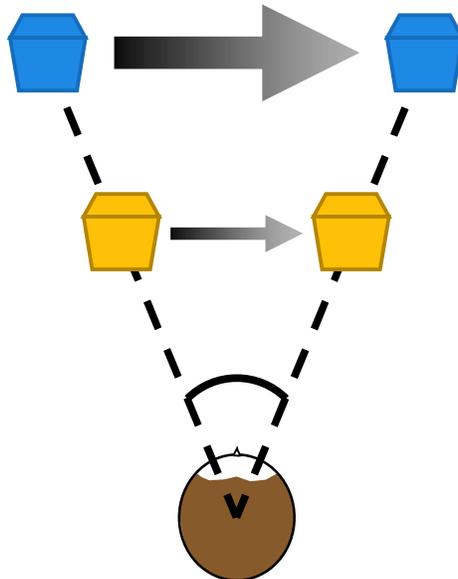
The majority of this thesis is centered around the perception of object movement during self-movement, with Chapter 2 introducing a new paradigm designed to allow us to measure the precision of the non-image signals that can be used during self-movement compensation to inform an observer about their own self-controlled self-movements. In

Chapter 3, the same paradigm will be used to investigate whether these signals adhere to Weber's law, in other words, whether their precision decreases with increasing magnitude. Not only will this provide a verification of the paradigm put forward in Chapter 2, it also facilitates a comparison between the non-image signals and those which encode image motion. This comparison will be pursued further in Chapter 4, where a noise manipulation will also be introduced to the stimuli to test the theory behind a standard Bayesian model. In Chapter 5, a Bayesian model will be derived that is the first of its kind to account for the between-trial variability of stimuli which are based on self-controlled self-movement, and a quantitative test of this Bayesian model will be presented. Transitioning away from the perception of object movement during self-movement, Chapter 6 will instead contain an investigation into a different type of compensation, namely speed constancy as a function of the distance of moving auditory objects from a stationary listener. Ultimately, the chapters of this thesis are connected by a single theme, compensation, and how we utilise compensation to perceive movement of ourselves and of objects.

## 1.2 – Compensation For Object Distance

Broadly speaking, compensation involves accounting for, or neutralising the effect of, something. As mentioned above, most of this thesis contains investigations into how self-movement is compensated for when trying to judge object movement, with close attention paid to both the inaccuracies that occur during this process, and the precisions of the signals that are used to complete it. It is not only our self-movement that is compensated for during movement perception, though. We also need to compensate for the effects of the distance between ourselves and objects when we interpret their movement. This is because the movement information contained in images is relative to the receptor and therefore corresponds to motion in angular form. For example, objects that are further from the observer need to move faster if they are to create the same angular change in the image, as shown in Figure 1.1. To recover the lateral speed of the object in 3D space, the distance to the object needs to be taken into account. This is known as speed, or velocity, constancy and has been well investigated in the vision literature (e.g., Brown, 1931; Distler et al., 2000; Epstein, 1973; McKee & Welch, 1989; Rock et al., 1968; Wallach, 1939; Zohary & Sittig, 1993). Speed constancy is thought to have close links to size constancy (Rock et al., 1968; Wallach, 1939), which is the compensation for object distance that we are able to achieve when interpreting

the sizes of objects. Here, using size constancy to predict distance and then this distance prediction to interpret movement is the strategy that is assumed during speed constancy. We do not solely use an object's size to determine how far away from us it is though, in reality, many depth cues are used and each cue gives additional information to our distance measurement. When these cues are removed systematically, there is a measurable decrease in our ability to perform speed constancy (Distler et al., 2000; McKee & Welch, 1989), which shows the importance of all of the available cues to depth perception when interpreting object movement. Further discussion of speed constancy is in Chapter 6, where it will be investigated whether speed constancy can also be found within the auditory system, with stationary observers.



*Figure 1.1: The object that is further from the observer has to move faster (larger arrow denotes faster movement) than the nearer object, in order to create the same angular change in the image.*

One way to think about the compensation for object distance that is integral to speed constancy is in terms of reference frames (e.g., Paillard, 1991). These set the spatial context that information is interpreted with respect to. Compensation can therefore be thought of as a process that transforms information from one reference frame to another. For example, in the context of speed constancy, the initial image motion is eye-centred and defines changes in visual angle. This can be converted into a world-centred reference frame (i.e. motion in 3D space) using distance information. Importantly, there are other reference frames that the

perceptual systems use to interpret image motion, including reference frames centred on the self (or ego). Compensation for our self-movement can also be thought of as a transformation of reference frames. For example, when using information from the visual image to interpret the movement of an object, the image is eye-centered, so we can use self-movement information, from the non-image signals described above, to transform the image motion into world-centred coordinates. This transformation of reference frames is a way of identifying and accounting for the motion in the image that is due to our self-movement.

### 1.3 – Compensation During Self-Movement

Measuring the movement of an object during self-movement involves not just compensating for the effect of object distance on image motion, but also the consequences of the self-movement itself. As mentioned above, if the observer is stationary, eye-centred and world-centred reference frames contain motion in different units (angular direction versus movement in 3D). However, when we move, there is another difference in the motion that the two frames contain. Take the example of an object that is stationary within a room. During eye-movement, the eye-centred image of the object contains motion due to the eye moving, however the object has not moved within a world-centred reference frame. If the object is the only information in the image, observers need to take into account the movement of their eye, to perceive the object as stationary. During eye and/or head rotation, the method of compensation for self-movement stems from the geometric relationship defined by Equation 1.1. Object movement is the sum of image motion and self-movement because the self-movement generates equal and opposite reafferent image motion. However, perceptual systems do not have direct access to image motion and self-movement, so these must be estimated from neural signals, as shown in Equation 1.2. Perceived object movement is therefore the sum of an image signal and a non-image signal.

$$\textit{Object Movement} = \textit{Image Motion} + \textit{Self Movement} \tag{1.1}$$

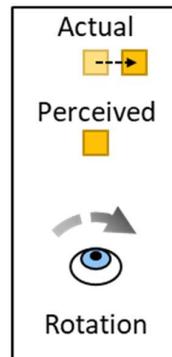
$$\textit{Perceived Object Movement} = \textit{Image Signal} + \textit{Non Image Signal} \tag{1.2}$$

Evidence that this form of compensation for self-movement occurs, comes from instances where it is absent. Haarmeier et al. (1997) describe the situation of a patient, R.W., who, when they pursued a stimulus with their eyes, perceived the world as moving at nearly the same velocity as their eye movements. This was demonstrated by measuring the degree to which a stationary object appeared stationary during a smooth pursuit eye movement. For the healthy controls, this was indeed the case. But for R.W., the object had to move at almost the same speed and in the same direction as the eyes in order for it to appear stationary. In other words, R.W. thought objects were stationary when image motion was 0. With respect to Equation 1.2, it appears that R.W. has no access to their non-image signal. Another example comes from Lindner et al. (2005), who found that schizophrenic patients with a specific symptom type, delusions of influence, could not compensate for smooth pursuit as well as healthy controls. Compensation was not completely absent this time, the object movement was somewhat slower than the self-movement, but in the same direction as it, when the object appeared stationary. This is a common effect in healthy participants known as the Filehne illusion (Filehne, 1922) and is discussed in more detail below. However, the error in compensation shown by the schizophrenic group was more than that shown by healthy controls. Lindner et al. (2005) explained their finding based on the idea that delusions of influence arise from an error in attributing the cause of perceptual events, specifically the delusion that self-movements are controlled by an external force. It was assumed that this meant that the non-image signal contained less motion for the schizophrenic participants than the non-image signal for the control participants during the same self-movement, which caused the perceived movement of the object to be more similar to the image motion estimate (in the opposite direction to the self-movement; see Equation 1.2), causing an increased Filehne illusion for the schizophrenic participants compared to the control participants. It could be the case that this increased Filehne illusion was caused by atypical image or non-image signals, or it could arise because the process that compares the two types of signal is at fault.

#### 1.4 – Compensation During Self-Movement is Imperfect

The experiments of Haarmeier et al. (1997) and Lindner et al. (2005) above make the point that compensation for self-movement is not always accurate. These studies are based on a simple stimulus situation in which the perceived movement of a single object is judged

during self-movement, specifically smooth pursuit eye movement. The rest of the scene is not visible. Wertheim carried out a series of experiments that showed that simple stimulus manipulations like changing spatial frequency (Wertheim, 1987), size (Wertheim, 1994) and eccentricity (De Graaf & Wertheim, 1988) altered perceived movement in these simplified contexts. In these experiments, the movement of stimuli was adjusted psychophysically until the target stimulus appeared stationary while making a smooth pursuit eye movement across it. Perfect compensation for the eye movement would mean that a stationary stimulus should appear as such, and no motion would need to be added. However, participants typically see stationary stimuli as moving against the eye movement, an effect first reported by Filehne (1922). In these simplified studies then, researchers need to add stimulus motion in the same direction as the eye movement in order to make the stimulus appear stationary, as shown in Figure 1.2.



*Figure 1.2: In order to counter the Filehne illusion (whereby a stationary object is perceived to move in the opposite direction to the self-movement), experimenters need to move objects slightly in the same direction as the self-movement for the object to be perceived as stationary.*

Many authors have measured the Filehne illusion in a variety of settings. Most notably, it has been found that the Filehne illusion is not solely a visual illusion, with analogous errors occurring in other modalities such as tactile motion perception during hand movements (Moscatelli et al., 2015), and auditory motion perception during head rotations (Freeman et al., 2017). Moscatelli et al. (2015) observed this effect when they asked participants to determine whether a surface was moving towards or away from them underneath their finger, either with their hand stationary or moving. It was found that during hand movement, the surface needed to move slightly in the same direction as the hand to appear stationary. With their hands stationary, participants could accurately determine when the surface was

stationary. Effects similar to the Filehne illusion were also observed in the auditory system by Freeman et al. (2017). In their experiments, it was also found that the extent of the Filehne illusion effect remained constant at different eccentricities, however precision of auditory motion perception deteriorated when the stimulus was presented at greater eccentricity.

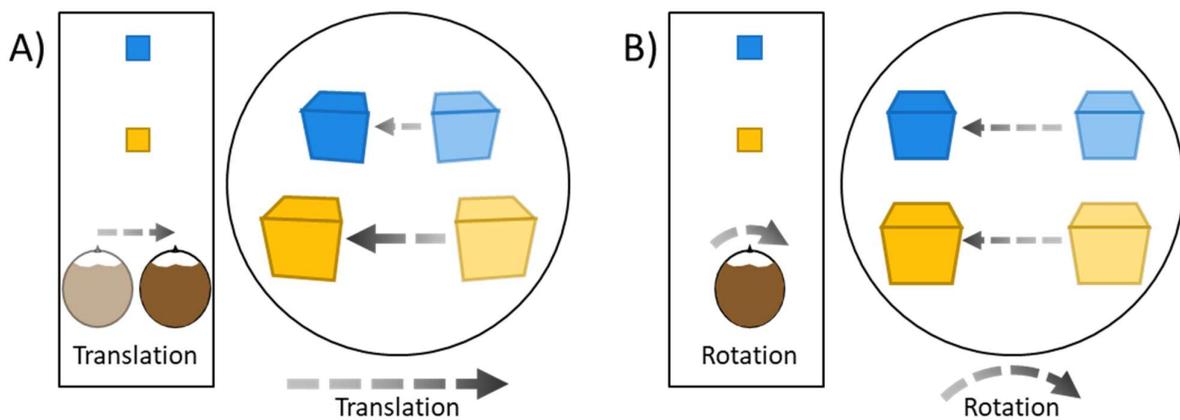
A perceptual error closely related to the Filehne illusion is the Aubert-Fleischl phenomenon (Aubert, 1887; Fleischl, 1882). This describes the fact that a visual object tracked by a smooth pursuit eye movement appears to move slower than the same object observed when the eyes are stationary. In this phenomenon, image and non-image signals are again compared, but unlike the Filehne illusion which occurs when the signals are simultaneous, the Aubert-Fleischl phenomenon contains motion in image and non-image signals at different times. As both of these errors are due to a mismatch between our estimates of eye movement, during eye pursuit, and object movement, utilising image motion when the eyes are stationary, they will be referred to as “self-movement compensation errors” throughout this thesis.

## 1.5 – Models of Compensation During Self-Movement

### Non-Image Signals versus Reference Signals

In a target article, Wertheim (1994) discussed the Filehne illusion in a much wider context, claiming that investigations into the Filehne illusion had revealed deficiencies in two general approaches to understanding perception, namely direct perception, as proposed by Gibson (1954), and inferential (or indirect) perception (e.g., Gregory, 1980). Wertheim argued that direct perception theory states that we measure the movement of ourselves based on the motion of retinal images alone. This was not quite the same as the direct perception theory that Gibson proposed. Instead, Gibson and others believed that the information contained in invariants in the scene was sufficient to determine the movements of objects (Gibson, 1968; Koenderink, 1986; Koenderink & van Doorn, 1987; Lee, 1980), and only more recently has it been shown that optic flow patterns can be used to determine the movements of ourselves and objects (Rushton & Warren, 2005; Warren & Rushton, 2008). It can be shown mathematically that the direction and speed of the motion in the image can allow us to distinguish whether we are moving or not, and also how we move along, and rotate about, each Cartesian axis (Bruss & Horn, 1983; Lappe et al., 1999; Longuet-Higgins & Prazdny, 1980).

Figure 1.3 displays some of the basic features identified by the mathematics. Translational and rotational movements of the observer generate uniquely different patterns of motion. For sideways translation, the motion of each point in the image has speed that is inversely proportional to the distance between the observer and the point, as demonstrated in Figure 1.3A. For rotation, however, the motion of each point in the image is the same irrespective of the distance between the observer and the point, as shown in Figure 1.3B. Different types of self-movement therefore produce different types of reafferent image motion that we can use to determine how we moved. According to the theory that Wertheim introduces to oppose the direct and indirect perception theories, the encoded image therefore becomes its own comparison as the image motion and the self-movement are both represented in the image signal (Wertheim refers to this as a ‘strange loop’). Wertheim then sets direct perception against inferential perception by arguing that ‘inferential’ compensation means that image signals are exclusively used in comparison against non-image signals. Crucially, in both cases, he uses versions of Equation 1.2, with non-image signals being used as an estimate of self-movement in the inferential theory, and reafferent motion in image signals being used as an estimate of self-movement in the direct perception theory. The issue therefore comes down to how we estimate our own movement in these simple stimulus contexts that are used to



probe the Filehne illusion.

*Figure 1.3: (A) the motion in the image is inversely proportional to the distance from the observer during horizontal translation of the observer. (B) the motion in the image is the same irrespective of the distance from the observer during rotation of the observer.*

In order to explain why perceived movement varies with various stimulus properties during pursuit, Wertheim suggested that the image-based cues to self-movement, used in

direct perception theories, are combined with the non-image cues put forward by inferential theory. In relation to the experiments Wertheim was focussing on, the issue becomes convoluted because the stimulus being judged is the only stimulus visible (apart from a small pursuit target to control the eye movement), and also because Wertheim was one among many who had assumed, albeit implicitly, that the image signal was accurate (e.g., Mack & Herman, 1973, 1978; Wertheim, 1987; Yasui & Young, 1975). Wertheim proposed that the self-movement from Equation 1.1 is measured with a combination of non-image and image information, rather than the non-image signal alone (in Equation 1.2). He called this a “reference signal”. Wertheim also proposed that the weight given to the image signals within the reference signal depended on a low band-pass spatiotemporal filter that allows stimuli which generate a sensation of vection to contribute to the reference signal. Small and fast stimuli presented for a short period of time consequently have no impact. This idea could therefore potentially explain why perceived movement during pursuit changes with spatial frequency, stimulus size, duration and other stimulus properties. As these stimulus properties change, so does the size of reference signals.

But as Freeman and Banks (1998) later pointed out, the inferential theory considered by Wertheim and the others assumed that image motion is somehow recovered veridically. In other words, they assume that the image signal in Equation 1.2 is immune to any stimulus manipulation, which it is not (e.g., changes in spatial frequency; Campbell & Maffei, 1981; Diener et al., 1976; Smith & Edgar, 1990). If this assumption is relaxed, Freeman and Banks (1998) showed that inferential theory based on Equation 1.2 can explain these findings without relying on a reference signal. In their first experiment, Freeman and Banks asked participants to adjust the speed of a stimulus until it appeared to match a standard stimulus. During some of the intervals, both stimuli were observed with stationary eyes to ensure that only image signals were present. Echoing previous findings (e.g., Campbell & Maffei, 1981; Diener et al., 1976), Freeman and Banks found that in these fixated conditions, the lower the spatial frequency, the slower the stimulus appeared to move. Note that there is some evidence that at high speeds and with high spatial frequency stimuli this effect is reversed (Smith & Edgar, 1990). Using this behavioural data, Freeman and Banks were then able to predict quantitatively how changing the spatial frequency of their stimuli would affect participants’ compensation for self-movement. In particular, they demonstrated that it was

possible to invert self-movement compensation errors by decreasing the spatial frequency of the stimulus. This was demonstrated for both the Filehne illusion and Aubert-Fleischl phenomenon: stationary objects could be made to appear to move in the same direction as the smooth pursuit eye movements, and pursued objects made to appear to move faster than objects observed with stationary eyes.

Nevertheless, it is difficult to tease these different models apart. Turano and Massof (2001) presented moving dots to participants during fixation and then pursuit, and used a staircase procedure to identify when participants perceived the speeds of the two sets of dots to match. Using this methodology, they found that the function that best described the relationship between perceptually equivalent image motion velocity and self-movement (eye) velocity was a non-linear model that introduced the modulation of the non-image signal by the image signal (an implementation of Wertheim's reference signal). However, in order to do this, they also included a non-linear relationship between input and output speed in their reference signal model, compared to a linear relationship in the model without the interaction term. Therefore, the two models had different numbers of free parameters that produced different relationships between input and output speed in the signals involved in the compensation. Around the same time, Freeman (2001) showed that including non-linear terms in a model without the interaction term fit the data better than a linear version for velocity matching, but he did not implement a reference-signal version. Souman et al. (2006) did, and compared all three models. They also devised a novel velocity-matching task in which both speed and direction 'around the clock' were assessed, unlike all the studies cited above, where stimulus and pursuit were always along the same horizontal axis. They found that Turano and Massof's model fit the data best, but noted that the improvement was not large. All of the models performed well.

### Flow-Parsing

One conclusion that can be drawn from the studies in the previous section is that using small, solitary stimuli might not be the best way to discover which cues are being used to perform self-movement compensation. As discussed above, according to Wertheim, the image motion serves two roles in these experiments: as an input to a reference signal, and as the input to an image signal that needs to be compared to the reference signal for compensation to occur. In a significant departure, Rushton and Warren (2005) argued that the

main role of image signals, that include reafferent motion (i.e. retinal flow), was to enable the observer to separate out the image motion that is related to object movement from the reafferent image motion that is related to self-movement. In a series of experiments, they have demonstrated that the visual system is able to identify differences between the motion of one part of the image (that contains an object) from the global motion in the rest of the image, and also between the motion of one part of the image and another, spatially similar, part of the image (Rushton et al., 2018; Rushton & Warren, 2005; Warren & Rushton, 2007, 2008, 2009). The motion that differs from its global or local counterpart can be used as an estimate of object movement. They termed this process ‘flow-parsing’ as the global or local flow is parsed out, leaving only the image motion due to the movement of the object remaining. Their experiments typically consist of a flow pattern presented on a screen with a target probe that moves in its own way. Participants are then asked about the movement of the probe, for instance, being asked to replicate its movement direction. In this case, the perceived movement direction would become biased by the global or local flow pattern. In the context of flow parsing, compensation is the identification and separation of motion due to object movement from reafferent motion due to self-movement, in the image.

Many of the experiments on flow-parsing concentrate exclusively on image motion. Participants are stationary, with head and eyes still. However, Warren and Rushton did not intend for non-image signals to be ignored – they viewed these signals as being combined somehow when available (Warren & Rushton, 2009). In fact, Warren and Rushton (2007) found that flow parsing was present during eye movements. Participants were instructed to fixate a moving probe and report its direction of movement while flow patterns were used that corresponded to rotation or translation of the eyes. The effect of the perceived distance of objects was different in these two cases (see Figure 1.3), suggesting that participants were parsing the flow patterns in order to determine how their eyes were moving with respect to the screen, despite the use of only rotational eye movements in the experiments. This idea is further supported by the work of Dupin and Wexler (2013) who found that participants used flow-parsing strategies alongside the use of non-image signals when performing self-movement compensation.

In the experiments presented in this thesis, the stimuli were individual, small, fast stimuli presented for a short period of time with no background context. The image signals

should therefore contain no flow patterns for participants to identify their own movement through flow-parsing, and there should be no influence of the image motion on the reference signal (due to the low band-pass spatiotemporal filter). For this reason, we reject the term reference signal and discount the method of flow-parsing in this context, instead referring to non-image signals throughout this thesis as our estimate of self-movement. This means that the image and non-image signals can be interpreted as separate entities without any interaction. These signals will be investigated throughout this thesis to test current theories that interpret self-movement compensation errors.

## 1.6 – Compensation in Hearing

Throughout this chapter so far, theories of visual motion perception have been discussed, but this is not the only modality of perception that is investigated in this thesis. In the auditory modality, movement perception is performed quite differently to vision. For a start, unlike eyes, ears cannot move independently from the head. This means that the geometry of the image signals that are produced is different. Another difference between the modalities is that visual image signals are inherently spatially organized whereas interpretation of auditory signals is needed to make the information spatially relevant. Movement of auditory objects is measured through constant monitoring of location cues including interaural time delay and level differences (ITDs and ILDs) which are the delay between a sound reaching each ear, and the difference in level of the sound when it reaches each ear. Spectral cues are also used during auditory localisation, which refer to the profile of the sound entering the ear. These are unique to the shape and size of our head and ears, and form what is known as a Head-Related Transfer Function (HRTF). These location cues enable us to measure the location of auditory objects with respect to the head, and it is through monitoring of these positions that we are able to determine distance travelled. Note that it is also possible to determine the speed of these objects by differentiating distance with respect to time, however, another difference between the movement perception of the modalities is that, while the visual system has a preference for measuring the speed of motion (Freeman et al., 2018; Reisbeck & Gegenfurtner, 1999), the auditory system has a preference for measuring the distance (Carlile & Best, 2002; Freeman et al., 2014).

Despite these differences between the movement perception of the auditory and visual systems, the small body of literature investigating compensation for self-movement in the auditory modality (Freeman et al., 2017; Genzel et al., 2016; Yost & Pastore, 2019) also finds that self-movement compensation is not complete. Genzel et al. (2016) presented stationary stimuli to their participants before and after active, passive, counteracted, or no self-movement. Their participants were instructed to determine whether a second stimulus was presented to the left or the right of an initial stimulus and were therefore judging location rather than movement. Between the stimuli, they: actively turned their head; were turned passively by a rotating chair; rotated their head actively while the rotating chair underneath them counteracted the rotation (resulting in their body turning underneath a world-stationary head); or remained stationary. It was found that the type of self-movement influenced the extent of self-movement compensation. In the active self-movement condition, the second stimulus had to be presented in the direction opposite the head turn to be perceived as in the same location as the first stimulus, whereas in the counteracted self-movement condition this effect was reversed. It is interesting to note that in Genzel et al.'s study, and in the situation of location perception during self-movement in the visual modality (e.g., Mita et al., 1950), during active self-movement, stationary objects appear to move in the same direction as the self-movement, which is contrary to self-movement compensation errors in movement perception where stationary objects appear to move in the opposite direction to the self-movement.

## 1.7 – Using Precision to Explain Bias

The previous explanations of self-movement compensation errors relied on errors in the underlying signals that we use in the compensation process. If these systematic inaccuracies are really the culprit, it would be sensible to think that they can be calibrated out by our perceptual systems over multiple presentations of stimuli. Similar calibration of visual signals occurs during prism adaptation (e.g., Facchin et al., 2019; Herlihey & Rushton, 2012; Redding & Wallace, 1985, 2006), where inaccurate sensory evidence is presented to an observer through a prism, and, by changing the mapping of visual or motor coordinates, or perhaps the relationship between the two, participants can account for the errors in the sensory evidence and use the visual information afforded to them effectively. Despite the evidence that perceptual systems are continually calibrating, self-movement compensation

errors persist. One potential explanation for this is offered by a different class of theory based on the Bayesian framework.

All perceptual signals contain some variability. There are many sources of this variability, including external sources such as other stimuli or background variation, and internal sources such as variability in neuronal firing. While precision and accuracy are distinct measurements, a trade-off between the two may be an important part of perception, with increasing the precision of our estimates often coming at the cost of biases, for example, in Bayesian models (e.g., Landy et al., 1995).

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

(1.3)

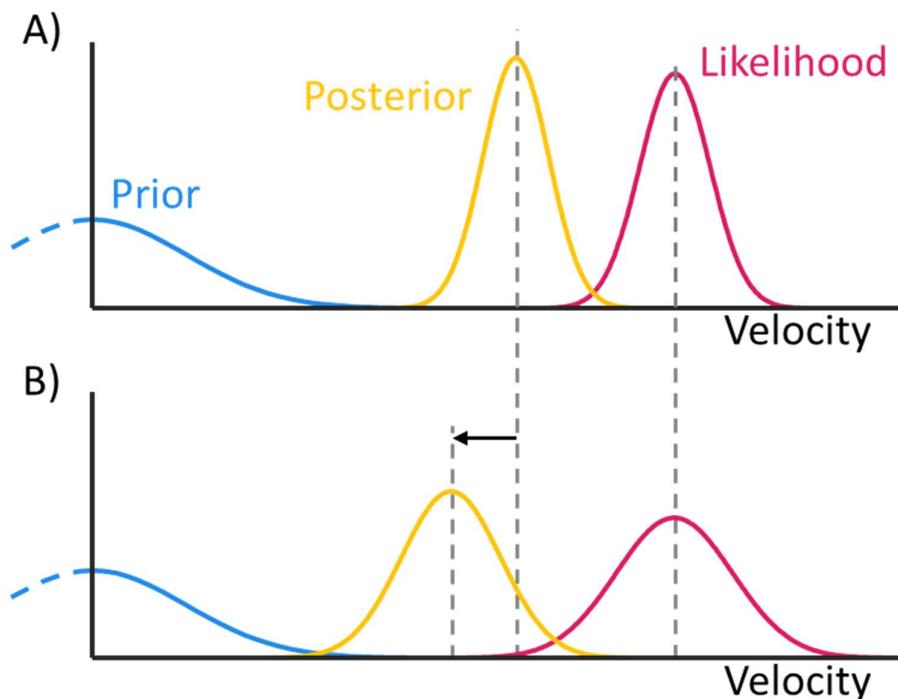


Figure 1.4: (A) a standard Bayesian model of motion perception. (B) the same Bayesian model but with more noise in the sensory evidence (wider likelihood distribution). This causes a shift in the posterior, towards the prior, as denoted by the arrow.

At the heart of Bayesian theories is the idea that perception is based on making statistical inferences about the state of world given that sensory measurements are variable. The probability of an event, A, happening, given that another event, B, is confirmed, can be

expressed as in Bayes' rule (Equation 1.3). This can also be visualised by presenting the probabilities as frequency distributions, as can be seen in Figure 1.4A. Applying Bayes' rule to a motion perception problem, for object movement,  $A$ , and low level sensory evidence,  $B$ ,  $P(A)$  (blue in Figure 1.4A) represents the prior, which is the probability that the object movement is equal to  $A$ , given all of the experience that the observer has had with previous object movements;  $P(B|A)/P(B)$  (pink in Figure 1.4A) represents the likelihood, which is the probability that the sensory evidence,  $B$ , provides support for the object movement being equivalent to  $A$ ; and  $P(A|B)$  (yellow in Figure 1.4A) represents the posterior, which is the resulting perception, after multiplication of the prior and likelihood. This can also be thought of as the probability that the object movement is equal to  $A$ , given that the sensory evidence is equal to  $B$ . Note that in Figure 1.4, the prior has a peak of 0 as it is a velocity distribution which is symmetrical about 0 (note also that this is a simplified distribution that only takes into account motion in two directions, this depiction in Figure 1.4 assumes that the X axis is oriented to the direction of the object movement. A more representative distribution would need to be 4-D, to account for motion in any direction in 3D while also displaying a probability). The slow-speed nature of the prior is due to the slow speeds of most objects that we interact with (Weiss et al., 2002). In Figure 1.4, the effect of changes to the precision of the sensory evidence is shown. In panel B, compared to panel A, the sensory evidence is less precise (has more variability), so the likelihood distribution is wider, and this results in a shift of the posterior so that it is closer to the prior. As the sensory evidence is assumed to be accurate in these Bayesian models, the shift in the posterior is a bias that makes the posterior more inaccurate. Note, however, that the likelihood distribution is slightly wider than the posterior distribution in Figure 1.4, evidencing that the use of the prior does allow us to utilise posterior measurements that are more precise than the original sensory evidence alone. It is this increase in precision of the posterior over the sensory evidence that is prioritized in these Bayesian models, over the accuracy of the posterior with respect to the sensory evidence, and it is this prioritization of precision over accuracy that causes the errors in our perception according to Bayesian models.

Bayesian models have been used to model the effects of different stimulus properties on perceived speed (Ascher & Grzywacz, 2000; Champion & Warren, 2017; Senna et al., 2015; Stocker & Simoncelli, 2006). One of these such effects is the contrast effect. It is well known

that at slow speeds, stimuli that have low contrast are perceived as moving slower than stimuli with high contrast, and at high speeds this effect is reversed (Thompson, 1982). As mentioned previously, there is a tendency for objects in the world to move slowly, leading to a slow-speed prior that is centered on no movement. The contrast effect at slow speeds can be explained by a Bayesian model that uses this slow-speed prior. As the contrast of a stimulus decreases, the precision of the sensory evidence decreases. As shown in Figure 1.4, if the sensory evidence is less precise, the posterior shifts more towards the prior. It follows that low contrast stimuli are shifted more towards the slow-speed prior, and are therefore perceived as slower than higher contrast stimuli. As the contrast effect changes at higher stimulus speeds, some researchers have disputed the appropriateness of Bayesian modelling in this context (Champion & Warren, 2017), while others have derived more complex Bayesian models which include filters that enable the explanation of the duality of the contrast effect (Ascher & Grzywacz, 2000).

Freeman et al. (2010) used a Bayesian model to explain the effects of self-movement on perceived speed (self-movement compensation errors). From Equation 1.2, we know that perceived object movement is based on the combination of a signal that denotes our own movement and a signal that denotes the motion in the image. Freeman et al. argued that these two signals are subjected to separate Bayesian operations but that they are based on the same slow-speed prior. A crucial prediction of the model was that the estimate of self-movement, in their case smooth pursuit eye movement, should be less precise than the estimate of image motion. In their behavioural experiments, Freeman et al. generated psychometric functions by asking participants to compare the speeds of objects presented while the eyes were stationary, and, separately, to compare the speeds of objects that they pursued. They found that speed judgements were less precise for the pursued objects, in agreement with their Bayesian model. Participants were also asked to compare the speeds of one object presented while the eyes were stationary, and another that they pursued, as in experiments investigating the Aubert-Fleischl phenomenon. The pursued stimuli appeared to move at around half the speed of the stimuli presented while the eyes were stationary. Taking these two results together, Freeman et al. proposed that self-movement compensation errors may be explained by having accurate image and non-image signals, but with a difference in precision, such that the non-image signal is less precise than the image signal. In a Bayesian

model, this would cause a greater shift towards the slow-speed prior for the self-movement estimate than the image motion estimate, making the self-movement estimate slower than the image motion estimate during equivalent motion. This is the hallmark of self-movement compensation errors. As these Bayesian models have highlighted, it is important to investigate not only the accuracy or inaccuracy of our perception but also the precision of perception and its underlying signals, as precision can have a marked effect on accuracy, and could cause the perceptual errors that we experience.

## 1.8 – Measuring the Precision of Active Self-Movement

In psychophysics, the precision of perceptual signals is assessed by measuring thresholds. These can be absolute, such as measuring the minimum stimulus level for detectability, or relative, such as measuring the smallest change needed to discriminate between two stimuli. Many methods have been developed to determine thresholds, including different types of staircase, where stimuli are manipulated depending on the perceptual decisions made by an observer in real time, and the method of constant stimuli, which will be used throughout this thesis. The latter method allows experimenters to pre-select a range of constant stimulus levels that they expect to encompass the threshold of interest. Participants are then asked to compare each of these constant stimulus levels to a standard stimulus that remains the same throughout the experiment. The responses of the participant can be used to generate a psychometric function, and a measurement of the threshold can be obtained from this function. This is a simple procedure for measuring the precision of signals, as long as the stimulus levels that are used are experimenter-controlled.

One of the main goals of this thesis, however, is to investigate the precision of perceptual signals that encode *participant-controlled* self-movement. The example of self-movement that will be investigated in this thesis is self-controlled head rotation. While training techniques can be used to promote the repeatability of these movements, as will be evidenced in Chapter 3, trial by trial variability still remains and ultimately the participant has control over their own movement. This contrasts with eye movements which can be more easily influenced by the experimenter, with repeatable pursuit targets or, in the case of saccades, flashed fixation points with known timing and location. The fact that head rotations are not easily experimentally controlled makes the use of procedures like the method of

constant stimuli non-trivial. In Chapter 2, a novel, two-phase paradigm will be developed to overcome these problems, and this paradigm will be used throughout Chapters 2-5 to investigate the precisions of the image and non-image signals, and self-movement compensation errors.

In the vision literature, the non-image signals that we use to measure our own movement are known as “extra-retinal” signals and eye movements are typically used as an example of self-movement. However, in this thesis, the term non-image signal is used, as one of the aims of this thesis to generalise this idea for other forms of self-movement (head rotations) and other modalities of perception (audition). It is important to note that non-image signals arise from multiple sources. In the case of head rotation, these signals include: vestibular measurements of the acceleration and velocity of the rotation (Cullen & Zobeiri, 2021; Israel & Warren, 2005; St George & Fitzpatrick, 2011); muscular efferent and afferent signals that instruct the neck muscles in how to move, and relay information about how the neck muscles did move respectively (Tuthill & Azim, 2018); and responses from skin receptors, for example, if long hair brushes against the observer’s neck during a rotation (Churan et al., 2017; Dallmann et al., 2015; Moscatelli et al., 2015). Throughout this thesis, it will be assumed that these different sources are interpreted as a single, combined non-image signal, with no attempt made to separate out their influence. As rotational self-movement induces motion in the visual and auditory images that is equal and opposite to the angular velocity of the self-movement, no matter the distance to the object (see Figure 1.3B), it can be assumed that the angular velocity of the object is equivalent to a summation of an image signal and a non-image signal (see Equation 1.2; e.g., Bridgeman, 2010; Epstein, 1973; Von Holst, 1954; Woodworth & Schlosberg, 1938). Angular velocity will be used as a measurement of the speed of objects in Chapters 2-5 while it is noted that, in order for participants to measure the 3D velocity with respect to the world, a compensation for the distance between the observer and the object (speed constancy; e.g., Brown, 1931; Distler et al., 2000; Epstein, 1973; McKee & Welch, 1989; Rock et al., 1968; Wallach, 1939; Zohary & Sittig, 1993) would be necessary.

As the head rotations investigated in this thesis are self-controlled, the non-image signal includes muscular efferent signals. These signals are copies of the instructions that are sent to the muscles telling them how to contract or relax and are only present during active, rather than passive, movement. The terms ‘active’ and ‘passive’ stem from the vestibular

literature, with one difference between the two types of movement being the amount of participant and experimenter influence over the movement. Many examples of this distinction can be found in literature pertaining to head movements in both monkeys (Klam & Graf, 2003; McCrea et al., 1999) and humans (Brooks & Cullen, 2019; Cullen, 2004), limb movements (London & Miller, 2013), and even touch (Chapman, 1994). A typical example of passive self-movement would be an experimenter controlling the movement of a participant through a rotating chair, however the non-image signal would then not contain muscular efferent signals.

## 1.9 – Summary

Throughout this chapter, compensation has been discussed in the context of depth perception, investigated further in Chapter 6, and self-movement, investigated throughout the rest of the thesis. Compensation for self-movement can be interpreted in many ways: as a transformation of reference frame; as a measurement of relative motion between a representation of an object and the rest of the scene in an image; or through the combination of non-image signals, that encode self-movement, and image motion signals. The latter is the only method that is relevant in the context of the experiments presented in this thesis as the stimuli are small, singular and presented for a short duration of time with no background information. There may be inaccuracies in the measurement of image motion or self-movement when using image and non-image signals, but it would make sense for these inaccuracies to be calibrated out over time. A Bayesian explanation of self-movement compensation errors that instead focuses on the relative precisions of these signals suggests that the prioritisation of precision over accuracy causes these errors, so the precisions of the image and non-image signals will be investigated throughout Chapters 2-5. In the next chapter, the two-phase paradigm mentioned above that enables the measurement of perceptual thresholds during participant-controlled self-movement will be explained, and an initial experiment investigating the precision of the non-image signal will be presented.

# The Precision of the Non-Image Signal

## 2.0- Preface

The main aim of this thesis is to investigate the signals that we use to compensate for our own self-movement. Presented here is an experiment that utilises a new paradigm, developed during this thesis, that enables the measurement of the precision of both the image and non-image signals that we use to measure object movement during self-movement. In this chapter, only the precision of the non-image signal will be investigated. The paradigm will be outlined throughout this chapter as an experiment is presented in which participants observed either visual or auditory stimuli either during self-movement or not. This chapter also introduces a novel psychometric function that is necessary in a scenario where stimuli with external variation (e.g., when they are based on self-controlled self-movements) are multiplied by gain factors that are then used as the independent variable in a psychophysical investigation. In the case presented here, participant head rotations vary across trials, violating the assumption of a single cumulative Gaussian in the psychometric function, and requiring the novel function presented here. This work is taken from a paper that is published on BioRxiv: <https://doi.org/10.1101/2023.09.20.558633>

## 2.1 – Introduction

Bodily movement is a key part of everyday life. Our eyes, head, limbs and torso are seldom at rest. Action therefore sets the backdrop in which perceptual systems normally operate, with many everyday tasks relying on information about current self-movement. This is derived from a number of perceptual signals, some based on images such as retinal flow, and some based on non-image sources including the vestibular and motor systems. Information from non-image sources also plays a role in interpreting images, allowing the observer to differentiate between self-generated movement and movements of external objects.

Success in these active tasks is constrained by two fundamental types of error, namely the precision and accuracy of the underlying perceptual signals. Precision is driven by internal and external noise and corresponds to the width of the distribution of the underlying perceptual signal as it varies across time. Accuracy, on the other hand, corresponds to the

distribution's average and is usually referred to as bias. While a lot is known about the precision and accuracy of image signals, especially in vision and hearing, much less is known about the errors accompanying non-image signals. This is especially the case when the self-movement is 'active' (i.e., self-controlled), partly because it is difficult to apply standard psychophysical techniques to actions that are under participant control. We therefore developed a new way to measure precision in these circumstances, using head rotation as an example of self-movement. The technique makes no attempt to differentiate between the various sources of non-image information that are used to encode active self-movement. Rather, it assumes that they are combined to provide a single non-image signal, and it is the precision of this composite signal that we measure. Our technique relies on using image signals as a comparison, which allowed us to compare the results when using vision or hearing.

In a typical precision-measuring task, participants are asked to compare a fixed standard stimulus with a range of test stimuli shown over a series of trials (e.g., Altman & Viskov, 1977; De Bruyn & Orban, 1988; Mallery et al., 2010). The values assigned to the test are usually controlled by a method of constant stimuli or a staircase procedure, both of which provide a measure of a just-noticeable-difference that can be used to estimate the precision of the underlying signal. Crucially, these methods rely on the ability to repeat a set of stimuli over trials. Self-controlled self-movements, however, are not repeatable: every instance of every action is unique. To measure non-image signal precision, this variability must be accounted for, both within and across trials.

One solution is to use 'passive' self-movement because the action is then controlled by the experimenter. The best-known examples come from vestibular research, where participants are moved on a chair or platform (Brooks & Cullen, 2019). Notable examples also come from studies of perceived stability during eye movement, where various contraptions and implements have been used to passively rotate the eye (Merton, 1964; Skavenski, 1972). But active non-image signals also include efferent sources, such as copies of motor commands (Von Holst, 1954), not just the vestibular, somatosensory and, in the case of passive rotation of the eye, the proprioceptive cues that passive stimulation generates (Cullen & Zobeiri, 2021; Israel & Warren, 2005; St George & Fitzpatrick, 2011; Tuthill & Azim, 2018). Passive stimulation therefore reduces the number of non-image sources by removing the effortful

efferent signals, potentially producing cue conflicts between the signals that contain motion and those which are silenced.

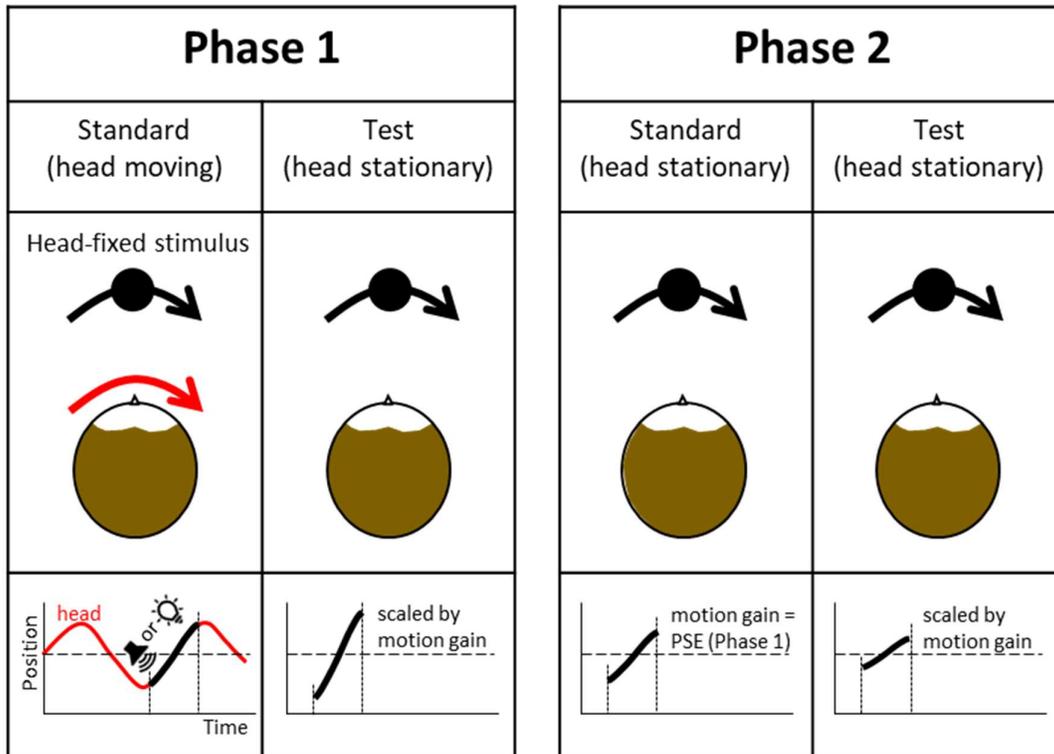


Figure 2.1: Schematic of the two-phase procedure for measuring the precision of non-image signals encoding active self-movement. We use head rotation as an example. Phase 1 consists of two intervals: a standard interval, in which the head moves and a head-fixed stimulus (visual or auditory) appears ‘on the nose’ in the 3rd sweep, and a test interval, in which the same movement of the stimulus, scaled by the motion gain, is replayed but with head stationary. Phase 2 also consists of two intervals, both with the head stationary. Here, the motion gain used to scale stimulus motion in the standard interval is set to the Point of Subjective Equality found in Phase 1. For both phases, a method of constant stimuli is used to manipulate motion gain across trials and construct psychometric functions. These are then used to determine the precision of the image signal in the head-stationary intervals, and the non-image signal in the head-moving interval, based on a model described in the Appendix of this chapter.

Instead, we focus on active self-movement, specifically head rotation, where non-image sources consist of vestibular cues, motor commands, and proprioceptive feedback, plus any number of somatosensory cues, such as the gliding of hair across the back of the neck. Our method for measuring non-image signal precision uses two experimental phases,

combined with a novel analysis that accounts for the variability of self-movement across trials. The paradigm is sketched in Figure 2.1. Based on two-interval forced-choice, the participant makes self-controlled left and right head rotations in the first interval of each trial of Phase 1, and an auditory or visual stimulus appears in the 3<sup>rd</sup> sweep (see bottom left panel). This stimulus is head-centred – it moves with the participant – and is used to mark which portion of the head movement to judge. This methodology was preferred to asking participants to pursue a moving target with the head as it is known that this type of self-movement is somewhat inaccurate (Chen et al., 2002; Collins & Barnes, 1999). We refer to this interval as the ‘standard’. In the second ‘test’ interval of Phase 1, shown in the 2<sup>nd</sup> column of the figure, an auditory or visual stimulus is again shown, but this time with the head stationary. The stimuli move with a trajectory defined by the recorded head movement from the first interval, but scaled up or down by a multiplicative factor we call ‘motion gain’. Hence, the pattern and duration of motion experienced in the two intervals is the same, apart from overall magnitude, and is encoded by different motion cues. In the head stationary interval, the motion cues depend on image signals. In the head-moving interval, they depend on non-image signals, including any extra-retinal contributions related to smooth compensatory eye movements like the vestibulo-ocular reflex (Barnes, 1988), or an inhibitory pursuit drive to keep the eye head-centred (Bedell et al., 1989). We note that the auditory and visual stimuli used to mark the 3<sup>rd</sup> head sweep do not provide any informative motion cue as they are head-fixed and presented in a dark and quiet lab. This remains the case even if their perceived positions shift due to audiogyral (Clark & Graybiel, 1949) and oculogyral illusions (Graybiel & Hupp, 1945). That said, it is doubtful that these illusory shifts in position occur over the time scales used in our experiments (Carriot et al., 2011).

Following the two intervals, participants indicate which interval appears to ‘move more’. We avoided the terminology ‘faster’ or ‘further’ because cue preference depends on modality: for vision, participants prefer speed rather than displacement and duration (Freeman et al., 2018; Reisbeck & Gegenfurtner, 1999), whereas for hearing the reverse is true (Carlile & Best, 2002; Freeman et al., 2014). Motion gain is manipulated across trials using a method of constant stimuli, resulting in a psychometric function that includes two sources of internal noise, one based on the image signal (e.g., visual or auditory motion) and one based on the non-image signal. To tease these two sources of noise apart, Phase 2 isolates

the internal noise of the image signal, using the same set of head movement recordings from Phase 1 to move the stimuli in the same trial-by-trial order, but with the head always stationary. Again, the second interval is a scaled version of the first. The precision of the non-image head rotation signal can then be recovered from Phase 1, with image precision now known.

In the experiment presented here, we used this novel technique to measure non-image precision accompanying head rotation, using either auditory or visual stimuli. The modality used to deliver image motion should not affect the precision of the non-image signal because the same image signal (with the same underlying noise) is present in all intervals apart from the first interval of Phase 1. This paradigm assumes that the eye movements are similar (but not necessarily absent) in the three head-stationary intervals. If this were not the case, Phase 2 could not be used to estimate the precision of the non-image signal in the head-stationary interval of Phase 1.

Note that Phase 1 of the paradigm also provides information about bias, specifically whether the magnitude of perceived object movement is the same with or without head movement. This is interesting in its own right, partly because it is well known that objects pursued by an eye movement appear slower (Aubert, 1887; Fleischl, 1882). In this case, the non-image ‘extra-retinal’ signal evidently provides a lower estimate of speed than the image signal. Analogous perceptual slowing has been demonstrated for passive head rotation (Garzorz et al., 2018) and active touch (Moscatelli et al., 2019), and has also been implicated for the auditory system during active head rotation (Freeman et al., 2017). But as far as we are aware, whether head rotation produces self-movement compensation errors whereby object motion during self-movement is perceived as slower than during fixation is currently not known for either vision or hearing.

## 2.2 – Methods

### Stimuli and Materials

Auditory stimuli were played over a 2.4m diameter ring of 48 Cambridge Audio Minx speakers as shown in Figure 2.2. The room was sound treated and completely dark. The speakers were controlled by two MoTU 24-channel sound cards, each linked to four 6-channel Auna amplifiers. Intensity was normalised across individual speakers. The stimuli consisted of

white noise spatially windowed by a Gaussian distribution ( $\sigma = 5.25^\circ$  in power, equivalent to 0.7 of the speaker spacing i.e.,  $\sigma = 7.5^\circ$  in amplitude). We have previously shown that this value avoids aliasing artifacts in our speaker system that could occur if the Gaussian distribution is undersampled, while at the same time avoiding the sound becoming too diffuse (Stevenson-Hoare et al., 2022). The noise was sampled at a rate of 48KHz with a peak level of 70dB. The position of the spatial Gaussian was refreshed at a rate of 240Hz, a rate set by the motion tracker described below. The result was a 'blob' of noise that could be moved smoothly across the speakers. The actual motion path taken was determined by the measured head movements, using the motion gain parameter to scale its magnitude.



*Figure 2.2: Laboratory set-up.*

Visual stimuli were presented to the participant using an AdaFruit NeoPixel strip of 342 LEDs driven by a single Arduino Uno microcontroller. The LED strip was positioned just below the speakers, as shown in Figure 2.2, and was driven at a framerate of 40Hz. The strip subtended  $128^\circ$  either side of straight ahead. This yielded an LED spacing of  $0.75^\circ$ . To ensure that the LEDs presented stimuli at a comfortable brightness, a single layer of 1.2f neutral density filter reduced the intensity of the display. As with the auditory stimuli, smoothly moving stimuli were created by using a Gaussian distribution that spatially windowed the LED

output for each display frame ( $\sigma = 1.05^\circ$ ). In order to prevent individual LEDs being visually resolved, the strip was placed in a curved enclosure with one open side that was covered by three layers of diffuser gel at a distance of 35mm, blurring the image. The overall size of the resulting blob was increased slightly by the diffuser ( $\sigma = 1.07^\circ$ ), which we confirmed using a Minolta LS100 photometer and an array of small apertures. The peak luminance of the blob was  $\sim 0.042\text{cd/m}^2$ .

### Head Tracking

Head movement was measured using a Polhemus Liberty tracker that sampled position at a rate of 240Hz. The tracker was mounted to a head band worn by the participant. For the head-moving interval of Phase 1, the head-tracking data were used to detect the 3<sup>rd</sup> sweep in real-time and keep the subsequent auditory or visual stimulus head-centred (i.e., motion gain = 1). To detect a change in head-movement direction, we convolved the head tracker samples with a finite difference filter to obtain a smoothed derivative. The filter was 13 samples long, meaning there was a 7-frame delay in detecting the head-turn ( $\sim 30\text{ms}$ ). An example waveform is shown in Figure 2.3A, with the detected 3<sup>rd</sup> sweep shown in black and blue.

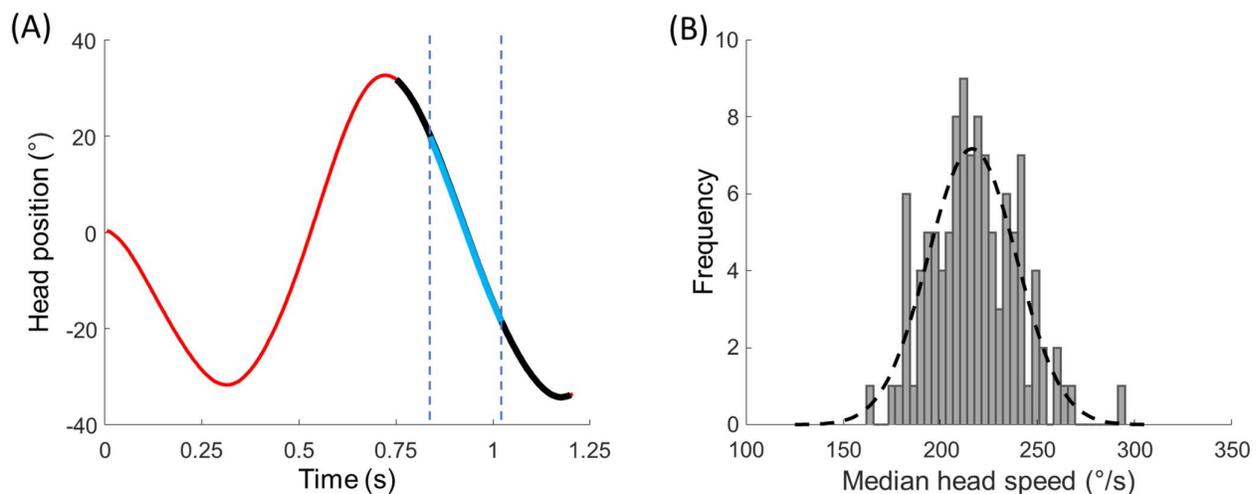


Figure 2.3: (A) Example head movement waveform. The black portion corresponds to the 3<sup>rd</sup> sweep as detected by the algorithm described in the text. The visual or auditory stimulus appeared during this time. The blue portion defines the region of interest over which median head speed was calculated for analysis. (B) An example distribution of median head speeds for a single repetition of Phase 1 (110 trials) for one participant. The dotted line shows the best fitting Gaussian, which was used to determine the mean and standard deviation of the distribution.

## Procedure

In Phase 1, each trial consisted of a 'head-moving' standard followed by a 'head-stationary' test. The start of the first interval was signalled by a short beep (0.25s) followed by a momentary pause to check the head was centred before the experiment moved on. 'Centred' was defined as 10 consecutive head-tracker samples within  $\pm 7.5^\circ$  of the centre of the LED/speaker array. Participants were then instructed to rotate their heads smoothly left and right, or vice versa, at a pace and amplitude that they felt comfortable with. While it was their free choice, we found some participants alternated the start direction from trial to trial, while others mostly started in the same direction. The auditory or visual stimulus appeared during the 3<sup>rd</sup> sweep and moved with the head. The start of the head-stationary test interval was signalled using a blue blob that appeared for 0.2s, with progress again paused to check the head was centred. Note that the motion in the test interval was based on the 3<sup>rd</sup> sweep recorded in the standard interval only: the dead-time created by the initial 2 sweeps was skipped. In Phase 2, the same beep and light were used to identify the start of each interval, with both intervals head-stationary and the initial 2 sweeps skipped. Each replication of Phase 1 and Phase 2 contained the same number of trials, based on the same head movement recordings, shown in the same order.

Psychometric functions were collected based on a method of constant stimuli using 11 motion gain values. For Phase 1, these ranged from 0.2 – 1.2 in 0.1 steps for the auditory condition, and 0.4 – 1.0 in 0.06 steps for the visual condition. The ranges were based on pilot experiments that showed the visual condition produced steeper psychometric functions than the auditory condition. Each motion gain was repeated 10 times, yielding 110 trials per session. For Phase 2, the same step sizes were used, but the range was centred on the Point of Subjective Equality (PSE) calculated from Phase 1. This ensured that the precision of the image-motion signal we estimated for each replication of Phase 1 and 2 were based on motion gains centred on a comparable value. The PSE was derived using the Palamedes toolbox (Prins & Kingdom, 2018; Wichmann & Hill, 2001; note the model described in the Appendix returns the same PSE as the toolbox).

Participants sat in the centre of the speaker/LED ring and wore the head tracking equipment. Head position was checked with a laser crosshair mounted above the centre of the ring, which enabled the interaural axis and speaker ring to be aligned. The head tracker was boresighted with the participant facing forwards and pointing their head towards the central speaker. Boresighting was repeated at the start of each replication of each phase of the experiment.

Each participant repeated three pairs of Phases 1 and 2 for each modality. Psychometric functions were fit to each replication separately. Three out of five participants carried out the auditory condition first.

### Head-Movement Analysis

To analyse the head movements after data collection, position samples were first smoothed using MatLab's 'lowpass' function with a passband of 8Hz. The temporal derivative was then taken and the median velocity calculated over a portion of the 3<sup>rd</sup> sweep that ranged from 20-60% of the sweep length (shown in blue in the example waveform of Figure 2.3A). This Region-Of-Interest (ROI) was adopted because it maximised the number of head-movement samples and goodness-of-fit of the psychometric function (see Appendix for evaluation). Figure 2.3B shows an example for one participant of the distribution of these median velocities for one run of Phase 1. For modelling purposes, the distribution of each 110-trial run was fit with a Gaussian (dotted line) to extract a mean and standard deviation.

### Psychophysical Analysis

The distribution shown in Figure 2.3B emphasises the fact that, as with other self-movements, head rotation varies across replications. Using motion gain therefore seems a good way of controlling for this variability because it links the related image motion to the ongoing self-movement in real-time. The patterns of motion are therefore identical, meaning the only difference between signal inputs is speed and displacement – duration is fixed. On the face of it, therefore, motion gain provides the experimenter with a repeatable parameter that can be used to define a psychometric function or drive a staircase. Examples are provided by Serafin et al. (2013) and Steinicke et al. (2009), who plot psychometric functions defined by changes in motion gain within acoustic and visual virtual reality set-ups, respectively. However, closer inspection of their figures suggests a consistent feature not accounted for by

fitting a standard cumulative Gaussian: on occasions, their data appear to asymptote more than a constrained lapse rate parameter would allow (e.g. <6%, as suggested by Wichmann & Hill, 2001). In the Appendix, we construct a model of the psychophysical task that shows why. The model emphasises that motion gain is not always a good shorthand for the actual stimulation experienced by the participant, namely the magnitude of motion (speed or displacement).

In keeping with standard Signal Detection Theory, the model assumes that participants base their judgement on a point estimate of stimulus magnitude (e.g., the peak speed of head and image movement, or average speed, or displacement). Crucially, the point estimates vary across trials due to the external noise introduced by variable self-movement, as well as the internal noise. The external noise produces some surprising effects (see Figure 2.A1). First, the function's true slope is steeper than the best-fitting single cumulative Gaussian. Second, as the variability of the self-movement increases, the function's asymptotes depart markedly from 0 and 100%, much further than a typical constrained lapse rate parameter of 6% would allow.

Following standard practice, we assume that internal and external noise is Gaussian distributed. The precision of a given signal is therefore defined by its standard deviation. If the self-movement did not vary at all, the precision of the non-image signal could be calculated by standard fitting of a cumulative Gaussian to the psychophysical data, and then applying the 'variances sum' law to both phases. Thus, for Phase 1,  $\sigma_{G_1}^2 = \sigma_i^2 + \sigma_h^2$ , where the subscripts correspond to the cumulative Gaussian fit to the data, the image signal (auditory or visual), and the non-image signal encoding head rotation. For Phase 2,  $\sigma_{G_2}^2 = 2\sigma_i^2$ ; hence the precision of the non-image signal ( $\sigma_h^2$ ) in Phase 1 can be found by substitution. But when self-movement varies, this standard approach is an approximation at best. Variable self-movement adds external noise that varies across the psychometric function because it is scaled by motion gain; hence the assumption of a single cumulative Gaussian is not correct. We develop the appropriate formulae in the Appendix and show how these can be used to extract the internal noise of the image signal ( $\sigma_i^2$ ) and non-image signal ( $\sigma_h^2$ ) from the two phases of our experiment. Similar formulae can be applied to a more typical motion gain scenario used in virtual reality set-ups, where both self-movement and image motion are

shown at the same time (e.g., Cherni et al., 2020; Nilsson et al., 2018; Serafin et al., 2013; Steinicke et al., 2009).

## Participants

All observers gave informed consent, and the experimental procedures were approved by the School of Psychology, Cardiff University Ethics Committee (EC.12.04.03.3123GRA2). Five participants took part in the experiment (2 female, 3 male). Two participants were naïve to the purposes of the experiment and three were experimenters. Participants wore spectacle correction if required.

The code used for fitting the model using the two-phase paradigm, together with the raw data and summary data, can be found here:

[https://osf.io/qcz7w/?view\\_only=2e2bb846820d4862bcb02a036d3ee815](https://osf.io/qcz7w/?view_only=2e2bb846820d4862bcb02a036d3ee815)

## 2.3 – Results

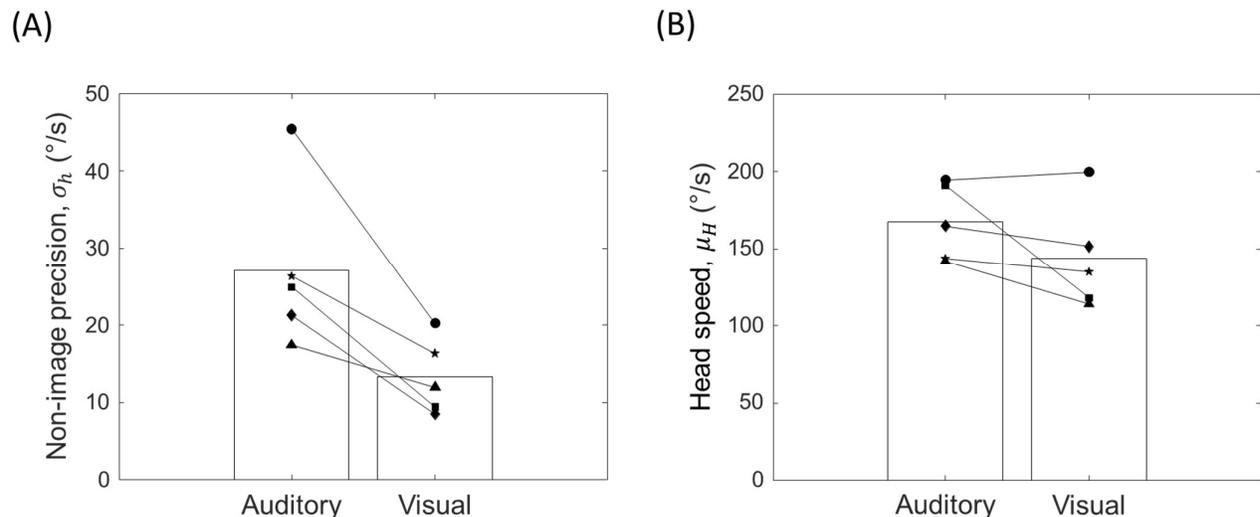


Figure 2.4: (A) Precision of the non-image signal encoding head rotation for the two stimulus conditions. Precision is defined as the standard deviation of the underlying signal distribution, as defined by the model in the Appendix. Note therefore that larger standard deviations correspond to less precise signals. Bars correspond to the mean of the individual data points shown as solid symbols. (B) Head speed using the same format.

Figure 2.4A shows the mean precision of the non-image head-rotation signal (bars) across the five participants together with their individual data (solid points). Non-image

signals were less precise in the auditory condition, producing an increase in the standard deviation of the underlying signal distribution as defined by the model described in the Appendix ( $t(4) = 4.19, p = 0.01$ ). Contrary to our prediction, therefore, modality appears to matter. Figure 2.4B shows that head speeds were slightly faster when an auditory target appeared in the 3<sup>rd</sup> head sweep compared to a visual one, however this difference was non-significant ( $t(4) = 1.76, p = 0.15, NS$ ). It is unlikely, therefore, that the marked difference in non-image signal precision measured using visual and auditory stimuli could be explained by a scaling of precision with magnitude (i.e. Weber's law). This point is explored further in Chapter 3, where an investigation of Weber's law on the non-image signal is presented.

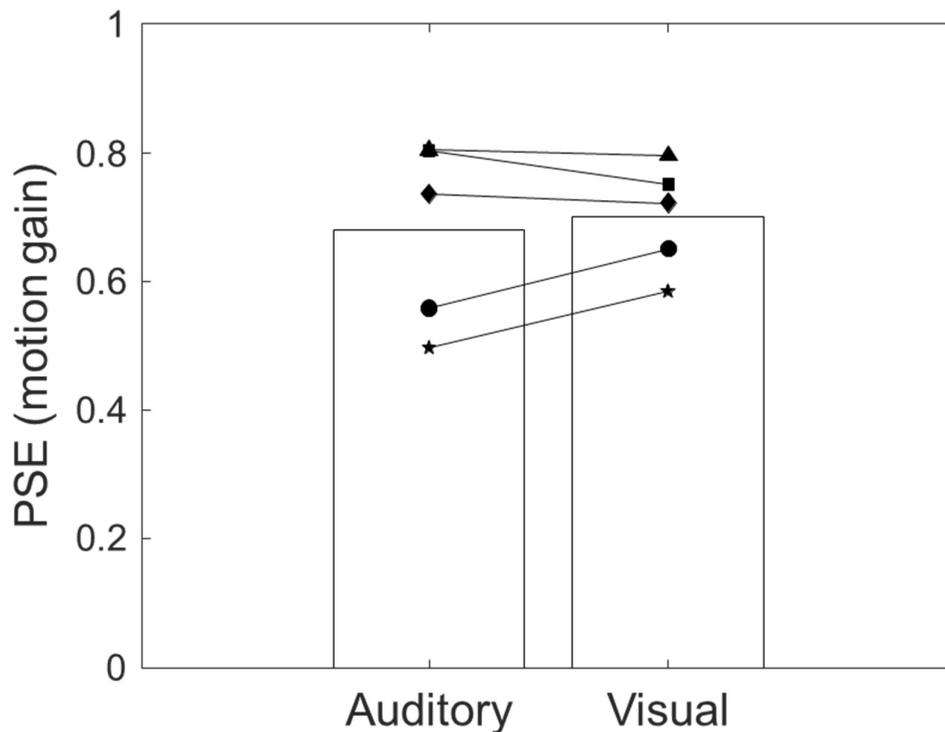


Figure 2.5: PSEs for the psychometric functions collected in Phase 1. These indicate the relative bias between image and non-image signals for both modalities. A value less than one means that head tracked stimuli appear slower. This is a self-movement compensation error.

Figure 2.5 shows the mean PSEs obtained from Phase 1. The PSEs are remarkably similar for vision and hearing ( $t(4) = -0.71, p = 0.52, NS$ ). The PSEs are around 0.7, meaning that stimuli had to be slowed by 30% in the head stationary interval, in order to be perceived

as the same speed as the head moving interval. For vision, this finding resembles the Aubert-Fleischl phenomenon (a self-movement compensation error; Aubert, 1887; Fleischl, 1882), where motion appears slower if stimuli are followed by a smooth eye pursuit. Our data show that perceived slowing occurs for active head movement too, albeit for auditory and visual stimuli that are linked directly to the self-movement as opposed to being pursued in a more typical closed-loop manner. The biases found are comparable to those during passive head rotation (Garzorz et al., 2018).

## 2.4 – Discussion

We have proposed a novel technique for measuring the combined precision of the non-image signals that encode active self-movement (e.g. vestibular cues, proprioception and motor commands). Traditional psychophysical techniques are difficult to use in these situations because the self-movement is under participant control. Stimulation is therefore not repeatable across trials. The new technique relies on three factors: (1) linking image motion to self-movement using a motion gain parameter that can be manipulated in a consistent fashion across trials; (2) the generation of two psychometric functions limited by identical sources of noise, apart from the internal noise related to non-image signals encoding active self-movement; (3) a model that yields the internal noise sources, while controlling for the external noise created by self-movement as it varies across trials. The technique could easily be adapted for other examples of active self-movement, such as walking and active touch, and situations where non-image signals and image signals are experienced simultaneously (e.g. virtual reality).

### Assumptions of the Model

The model assumes that eye movements made in all head-stationary test intervals are similar. If this were not the case, then the image noise in Phase 2 could be different from the noise in the test interval of Phase 1. For instance, if observers pursued the stimuli in the test interval of Phase 1, but not in Phase 2, then motor cues related to the pursuit system would be present in the first phase but not the second. However, eye movements were recorded in a subsequent experiment, presented in Chapter 3, and showed that fixation accuracy was similar in all intervals where the head was stationary, with the evidence suggesting that observers were able to keep their eyes stationary for a large percentage of the time. Fixation

accuracy dipped for the head-moving interval, but most samples were still near the auditory target. Hence the eyes were not moving significantly for much of the time, see Chapter 3 for more detail. While we did not measure eye movements in this experiment, it is unlikely that fixation accuracy would differ across intervals where the head was stationary in the two phases for vision and hearing. We say this for two reasons. First, we'd expect more variability in eye fixation to head-fixed auditory stimuli than visual stimuli, and yet fixation accuracy was good in the auditory-only experiment presented in Chapter 3. Second, in Chapter 4, fixation accuracy was similar across phases for both visual and auditory stimuli.

The model assumes that internal noise is fixed. While head speed was controlled by participants and therefore free to vary from session to session, we view the fixed noise assumption as a reasonable approximation for the range of stimuli contributing to any single psychometric function. Indeed, the fixed noise assumption is implicit when fitting a single cumulative Gaussian to the data, unless the logarithms of stimulus values are used.

### Measurement Noise

The psychometric functions derived from Phase 1 and 2 are based on the same measurements of head rotation. This controls for any effect of measurement noise introduced by the head tracker because the (head-stationary) test intervals are based on the same set of recordings, presented in identical order. The effect on the precision of the image signal in the two phases is therefore the same, such that any influence is cancelled out.

This does not mean that all of our conclusions are immune to the effect of measurement noise. However, the effect of measurement noise is likely small. The standard deviation of position samples output by our head tracker is  $\sim 0.55^\circ$ . This is considerably lower than the positional noise needed to produce significant changes in speed discrimination thresholds previously reported in vision (Bentvelzen et al., 2009; Rideaux & Welchman, 2020). For instance, in the 'high noise' condition of Bentvelzen et al, positional noise was added to their LED system with a standard deviation of  $7.4^\circ$  at an update rate of 25Hz. This produced thresholds that doubled compared to baseline. We would therefore expect visual image signal thresholds to change around 7.4% (i.e.  $100 * 0.55 / 7.4$ ). Bentvelzen et al used a two-interval technique, so this equates to a 5.25% change in the standard deviation of the underlying signal. This is quite a small change, moreover, spatial hearing is considerably less precise than

perifoveal vision, suggesting the effect would be even smaller still for these less precise auditory signals. While image signals were not investigated in this experiment, auditory image signals were investigated in the experiment presented in Chapter 3, and visual and auditory image signals were investigated in the experiment presented in Chapter 4. In those chapters, the effects of this 5.25% theoretical measurement noise will be noted.

### Vision versus Hearing

In the experiment presented here, non-image signal precision was lower when using auditory stimuli in the head-stationary test intervals. This finding was corroborated in a further experiment, presented in Chapter 4. Here we speculate that these differences stem from the need to convert perceptual signals into common units.

Suppose that the non-image signal is dominated by vestibular cues. Vestibular activity is based on acceleration, but as pointed out in the Introduction, vision prefers speed (Freeman et al., 2018; Reisbeck & Gegenfurtner, 1999) while hearing prefers displacement (Carlile & Best, 2002; Freeman et al., 2014). The motion signals therefore start in different units and must be transformed before they can be compared. One strategy is to integrate the vestibular cue once to get speed for vision, and twice to get displacement for hearing. Each transformation step adds noise. Therefore, non-image precision will be lower when auditory stimuli are used.

Alternatively, let's suppose that motor signals (proprioception and motor commands) dominate the non-image signal. Freeman et al. (2018) showed that observers prefer speed versus displacement and duration cues when judging the motion of a pursued target, and that the cue being used was extra-retinal (i.e. motor commands and/or proprioception). Assuming the same is true for head rotation, the motor signals in our experiments start out in speed units. No transformation is therefore needed when comparing motor signals to vision because vision prefers speed anyhow, but one transformation step is needed to get the preferred cue of displacement for hearing. Once again, the precision measured using auditory stimuli would be lower because it needs more transformation steps. The situation is more complex if the non-image signal consists of both vestibular and motor cues because they would need to be converted into common units before comparing with vision or hearing (known as cue promotion in the cue combination literature: Landy et al., 1995). Nevertheless, hearing will

always need an additional transformation step compared to vision, meaning that the precision of the non-image signal should decrease whenever auditory stimuli are used.

## Conclusions

We have presented a novel technique for the measurement of the precision of non-image signals encoding active self-movement. We used head rotation as an example of self-movement, and showed that the precision measured was different when using auditory versus visual stimuli, which may be caused by the additional transform that must take place for comparison between non-image and auditory image signals.

This paradigm and psychometric function will be used throughout Chapters 3-5 to further investigate the precisions of the signals encoding object movement perception during self-movement. Chapter 3 will further investigate some of the themes from this chapter, with an investigation into the effects of self-movement speed on the precision of the non-image signal, which also allows for an investigation of the comparison between the precisions of the image and non-image signals.

## 2.5 – Appendix

Phase 1 consists of a head-movement interval followed by an image-motion interval. In the first, there is no image motion as the object is spatially linked to the movement of the participant. Perceived motion therefore depends on a point estimate ( $h$ ) of the non-image signal encoding head rotation. We assume that  $h$  is corrupted by fixed additive Gaussian noise across trials. Using  $N(\mu, \sigma)$  to denote a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ , the non-image signal is therefore distributed as  $h = \mu_h + N(0, \sigma_h)$ . The mean  $\mu_h$  depends on the head movement magnitude ( $H$ ), which we also assume is normally distributed across trials (see Figure 2.3B in the main text). Perceived motion in interval 1 ( $M_1$ ) is therefore:

$$M_1 = bN(\mu_H, \sigma_H) + N(0, \sigma_h) \tag{2.A1}$$

where  $b$  is a linear bias term that sets the gain of the head-movement signal relative to its input i.e.  $h = bH$ . Note that either speed or displacement could be used to characterise the distributions of head rotation and signals (to reiterate a point made in the main text, displacement and speed are perfectly correlated when manipulating motion gain because duration is fixed). The model is ambivalent. Swapping between speed and displacement

changes the units but not the relative differences found for a chosen parameter across conditions.

In the second interval, image motion ( $I$ ) moves as a fixed proportion ( $g$ ) of the head movements recorded in interval 1:  $I(t) = gH(t)$ . We refer to  $g$  as the ‘motion gain’. As the head and eyes are stationary, sensed movement depends on an image signal ( $i$ ). Following similar logic to interval 1, the perceived motion in interval 2 is therefore:

$$M_2 = gN(\mu_H, \sigma_H) + N(0, \sigma_i) \quad (2.A2)$$

Note that Equation 2.A2 assumes that the image signal is unbiased. Hence  $b$  in Equation 1 defines the *relative* bias between  $h$  and  $i$ , such that  $b < 1$  means that the non-image signal registers a lower magnitude than the image signal. Following standard Signal Detection Theory (e.g. Jones, 2016), we assume observers base their choice on an internal decision variable ( $d$ ) that depends on the difference between the perceived motion in the two intervals:

$$d = M_2 - M_1 \quad (2.A3)$$

The choice ‘Interval 2 appears to move more’ corresponds to  $d > 0$ . From signal detection theory we define

$$d' = \frac{\mu_d}{\sigma_d} \quad (2.A4)$$

such that the probability of choosing Interval 2 is given by:

$$P = \frac{\lambda}{2} + (1 - \lambda)\Phi\left(\frac{d'}{\sqrt{2}}\right) \quad (2.A5)$$

where  $\lambda$  is the lapse rate and  $\Phi$  is the cumulative distribution function of the standard normal distribution.

Substituting (2.A1) and (2.A2) into (2.A3):

$$d = (g - b)N(\mu_H, \sigma_H) + N(0, \sigma_i) + N(0, \sigma_h)$$

(2.A6)

By inspection:

$$\mu_d = (g - b)\mu_H$$

(2.A7)

Note that the PSE occurs when  $\mu_d = 0$ . At this point  $g = b$ ; hence the relative bias between  $h$  and  $i$  can be read directly from the psychometric function. If the bias  $b < 1$ , then the PSE occurs when image motion is slower than head-movement. This is analogous to the Aubert-Fleischl phenomenon (Aubert, 1887; Fleischl, 1882), in which moving objects appear slower when pursued. Conversely, if  $b > 1$ , then image motion must be faster to achieve the PSE.

To obtain  $\sigma_d$ , we sum the variances of the three distributions defined by Equation 2.A6 and take their square root:

$$\sigma_d = \sqrt{(g - b)^2\sigma_H^2 + \sigma_h^2 + \sigma_i^2}$$

(2.A8)

If the head movement did not vary across trials ( $\sigma_H^2 = 0$ ), then the square root of the sum  $\sigma_h^2 + \sigma_i^2$  is the slope of the best-fitting cumulative Gaussian. The precision of the non-image signal ( $\sigma_h$ ) could then be obtained by measuring  $\sigma_i^2$  in Phase 2 and subtracting it from the sum. However,  $\sigma_H^2 \neq 0$ . Variable head movements make the recovery of  $\sigma_h$  more complicated because they act as an external source of noise that varies with motion gain across the psychometric function.

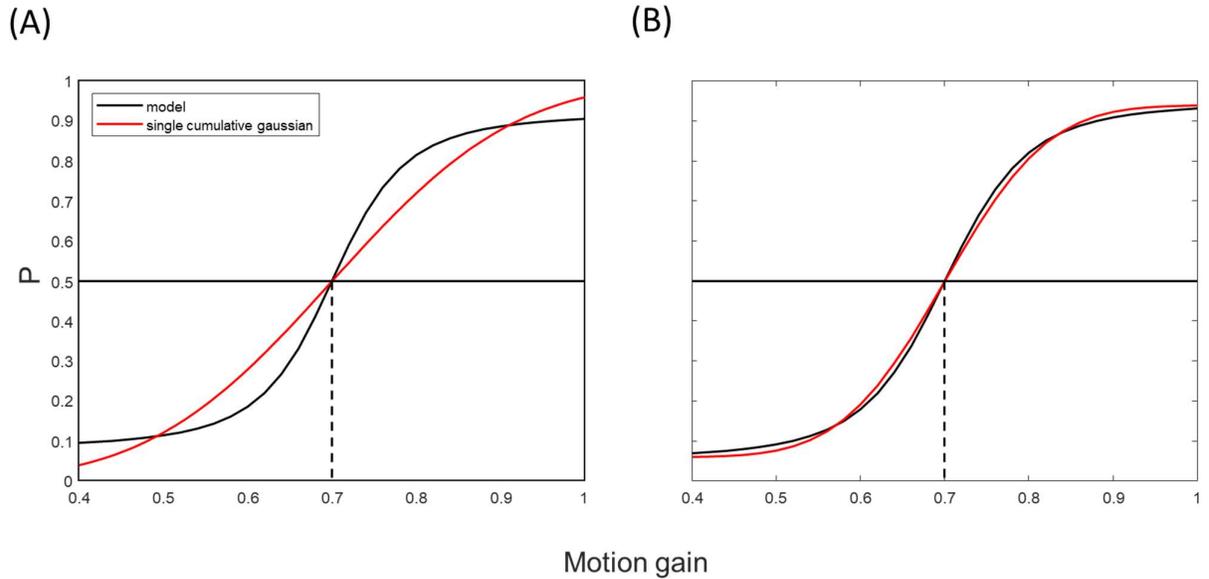


Figure 2.A1: (A) The black curve shows a psychometric function based on gain-dependent noise with  $[\mu_H, \sigma_H^2, \sigma_h^2, \sigma_i^2, b, \lambda] = [20, 100, 2, 1, 0.7, 0]$ . The red curve is the best fitting cumulative Gaussian as determined by the Palamedes toolbox, with lapse rate  $\lambda = 0$ . (B) Same curves but with lapse-rate  $\lambda = 0.06$  for the gain-dependent noise psychometric function, and free to vary for the single cumulative Gaussian ( $\lambda \leq 0.06$ , a standard constraint suggested by Wichmann and Hill (2001)).

Figure 2.A1 shows that fitting a single cumulative Gaussian is an approximation at best. The black curves show example psychometric functions based on the formulae above (see legend for parameter values) while the red curves show the best-fitting single cumulative Gaussian. The difference between the two panels is whether a lapse-rate is included or not. The external noise has two effects: (1) the asymptotes of the psychometric function move away from  $P=0$  and 1; (2) the slope becomes steeper and is not well fit by a single cumulative Gaussian. The degree to which the external noise causes substantial departures from the standard fit depends on the relationship between the values of  $\mu_H, \sigma_H^2, \sigma_h^2, \sigma_i^2, b$  and whether lapse-rate is allowed to vary in the standard fit.

### Fitting Procedure

We fit psychometric functions to our data based on the formulae above, using the measured head movements to estimate the mean and standard deviation of  $H$ . A MatLab function for doing this can be found in the Thesis Appendix (“fitSMmodel”). Phase 2 data were fit first, with  $\sigma_i^2$  and  $\lambda$  free to vary, and  $\mu_H$  and  $\sigma_H^2$  fixed, using the Gaussian distributions that

we fit to the obtained head movement speeds (see Figure 2.3B in the main text). Phase 1 was then fit, with  $\sigma_h^2$ ,  $b$  and  $\lambda$  free to vary and  $\sigma_i^2$ ,  $\mu_H$  and  $\sigma_H^2$  fixed. To avoid local minima in the fit, each parameter was cycled through a search space of 20 values and the best fit chosen. This yielded  $20^n$  separate cycles of the fitting routine, where  $n$  is the number of free parameters which was different for the two phases (400 cycles for Phase 2 and 8000 cycles for Phase 1).

We did not find much difference between fitting the new psychometric function and fitting a single cumulative Gaussian. One likely explanation for this similarity was that the head movements were relatively consistent ( $\sigma_H^2$  low) given the repetitive nature of the task. It may also be the case that including a constrained lapse-rate parameter soaked up a proportion of the asymptotic effect of the external noise. This can be seen by comparing Figure 2.A1A (no lapse rate) with Figure 2.A1B (constrained lapse rate  $\leq 6\%$ ). The lapse rate mimics the asymptotic behaviour produced by the external noise.

#### Region-Of-Interest for Calculating Head Rotation Speed

The analysis depends on mean and variance of the head movements made in interval 1. The mean and variance were estimated from histograms of average speeds in the 3<sup>rd</sup> sweep as described in the main text. To determine the region-of-interest (ROI), we compared the goodness-of-fit of psychometric functions from three ROIs: 20-80%, 20-60% or 40-60% of the sweep length. The psychometric functions were fit using MLE, so the appropriate measure of goodness-of-fit is the deviance (Wichmann & Hill, 2001). Figure 2.A2B shows that deviance did not change with the different ROIs used (the deviance has been averaged across conditions, phases and participants). However, Figure 2.A2A shows that an ROI of 20-80% produced a slower estimate of head movement speed than the other two ROIs, which was also more variable due to the inclusion of salient periods of acceleration and deceleration. We therefore opted for an ROI of 20-60%.

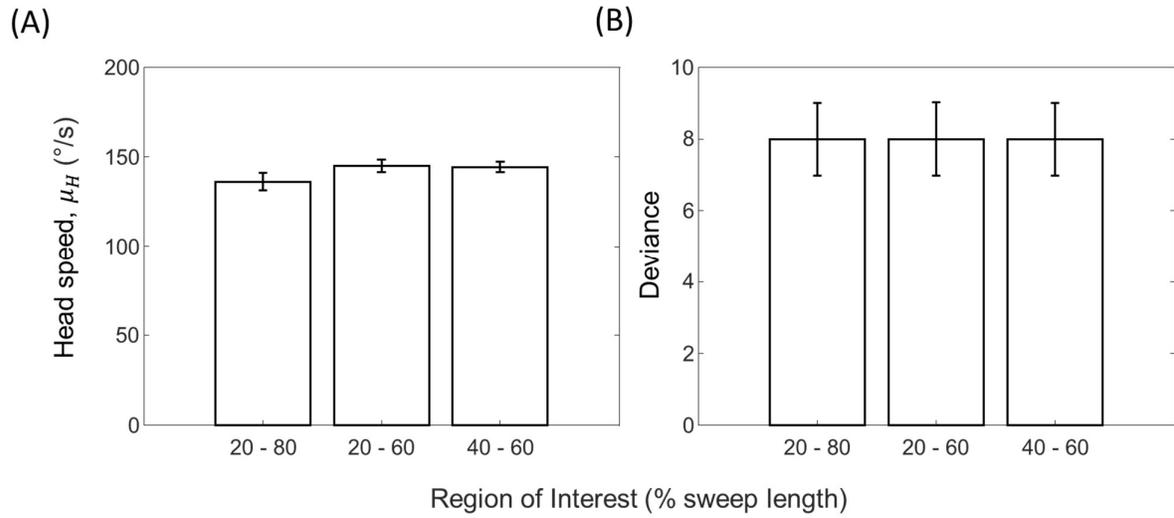


Figure 2.A2: (A) Head speed and (B) model goodness-of-fit for each ROI used to analyse the head movements. Error bars represent  $\pm 1SE$ .

# Investigating Weber's Law in the Non-Image Signal

## 3.0 – Preface

The previous chapter outlined the paradigm that will be used throughout this chapter and Chapters 4 and 5 to investigate the precisions of the signals involved in perceiving object movement during self-movement. This chapter focuses on how the precisions of these signals change as the speeds of the self-movement and object movement change. A classic assumption within perceptual science is that discrimination thresholds scale proportionally with stimulus magnitude (Weber's law; e.g., Laming, 2009). This assumption will be investigated within the context of object movement perception during self-movement in this chapter. This will also allow for a comparison between the precisions of the image and non-image signals as distributions of precision across different speeds will be generated.

## 3.1 – Introduction

As noted by Algom (2021), previously when discussing Weber's Law there have been issues of misunderstanding so, for clarity, the term "Weber's Law" will be used in this thesis to refer to the observation that the just noticeable difference (JND) of a signal is proportional to the signal's magnitude. This means that the smallest physical change in a stimulus that is detectable by the observer grows proportionally as the stimulus becomes more intense. This idea was first proposed by Ernst Heinrich Weber (1795–1878) and is not to be confused with what some researchers call Weber-Fechner Law which highlights the change in *sensation* that stems from changes to stimulus intensity. Fechner's Law assumes that each JND is perceived as equal to the last, despite Weber's law suggesting that the intensity of the JND is proportional to the stimulus intensity. Fechner's law goes on to suggest that if JNDs that are different in intensity are perceived as the same, stimuli can be judged by measuring how many JNDs they are above the detection threshold. Weber and Gustav Theodor Fechner (1801–1887) are seen as the creators of psychophysics as early advocates for a more quantitative approach to psychological research.

Weber's Law has been investigated in a multitude of different contexts with varying results. An in-depth review of such studies was produced by Masin (2009) where it can be seen that despite some studies finding "acceptable" evidence for Weber's Law (Leshowitz et al., 1968; Masin, 2009), most others have found that this concept breaks down at low and/or high stimulus magnitudes, despite the authors often claiming that Weber's law held. Throughout this thesis, the signals encoding the perception of object movement and self-movement are pertinent. Some research that investigated Weber's law in object movement perception includes the work of De Bruyn and Orban (1988) who looked at Weber's Law in visual motion perception. The first experiment they present focuses on visual velocity discrimination. In this experiment, the stimuli were random dot patterns that were moved by turning a mirror at rates ranging from 0.5 to 256°/s. De Bruyn and Orban found that the Weber fractions, which are a proportional measurement of the JND with respect to the stimulus intensity and should remain constant if Weber's law holds, decreased substantially between 0.5 and 4°/s and then shallowed until around 64°/s where they increased substantially again to form a U-shaped distribution. These findings suggest that Weber's law breaks down during visual motion perception at slow speeds and fast speeds. In hearing, Altman and Viskov (1977) investigated Weber's law in the context of implied auditory motion. They achieved this by presenting auditory tones over headphones with differing spatial properties and a delay between them. This method allowed for implied motion in the fused auditory image which enabled Altman and Viskov to investigate the precision of auditory motion perception. Their Figure 3B shows the Weber fractions at different implied motion velocities which show a distinct decrease between velocities of 20 to 40°/s, shallowing at higher speeds until the Weber fraction became constant at around 100°/s. This implies that Weber's law breaks down at slow speeds in auditory motion perception, but there was no evidence here of a breakdown at fast speeds. This investigation only obtained speeds of 140°/s which may explain the difference between the distributions of precision obtained in the investigations of visual and auditory motion perception. We would therefore predict that Weber's law should break down for the image motion signal at slow object movement speeds, with the possibility of a further breakdown at high object movement speeds, creating a U-shaped distribution if the movement speeds that we investigate are high enough.

Other relevant experimentation was completed by Mallery et al. (2010) who investigated Weber's law in the context of the vestibular system. They achieved this by turning participants using a rotating chair and found that Weber's law predicted their data relatively well, however, they found that a power function with an exponent of around 0.4 produced a more successful fit. When converted to Weber fractions, this data appears to decrease as the speed of the movement increases but is shallowing throughout, suggesting that Weber's law may not hold over this range of speeds. The maximum rotational velocity used in Mallery et al's study was  $150^{\circ}/s$ . Similar findings were obtained by Nouri and Karmali (2018) who investigated the vestibulo-ocular reflex. When converted to Weber fractions, their data appears to follow a similar trend to Mallery et al's data, in fact, the power function that Mallery et al obtained fits Nouri and Karmali's data quite well. If we draw comparisons from vestibular passive self-movement perception, it seems likely that our investigation of the non-image signal encoding self-controlled self-movements should find that Weber's law fits the data relatively well, but that the data could potentially be better described by a power function.

In order to investigate Weber's law in the context of self-controlled self-movements, we selected rotational head turns as an example of self-movement and trained participants to turn their heads at one of five angular velocities (20, 50, 80, 110,  $140^{\circ}/s$ ). We then used the same novel paradigm that was introduced in Chapter 2 to measure the precision of the combined non-image signal that encodes our perception of self-movement. When using this paradigm to obtain a value for the precision of the non-image signal, a measurement of the precision of the image signal encoding object movement perception is also generated. It is important to note that the precision of the image signal was not presented in Chapter 2 because the image signal, whose precision can be measured with Phase 2 of the paradigm outlined in Chapter 2, represented slower object movement than the self-movement in Phase 1. This was due to the self-movement compensation error, shown by the bias that was found in Chapter 2. In this chapter, the values for the precisions of the image signals will be relevant, as we can construct distributions of both the non-image and image signal precisions at different signal speeds, to compare them. This experiment was conducted with auditory stimuli, which, as seen in Chapter 2, may affect the obtained value for the precision of the

non-image signal. The effect of this stimulus selection on this investigation into Weber's law will be taken up in the General Discussion chapter.

## 3.2 – Methods

### Stimuli and Procedure

The aim of this experiment was to determine signal precision as a function of head rotation and stimulus speed. Before each replication of the main experiment, a training session was therefore run to help participants rotate their heads at one of the five target speeds. The training stimuli were audiovisual, consisting of visual and auditory blobs as used in the experiment presented in Chapter 2. These moved in synchrony. During the main experiment, however, only auditory stimuli were used. For the training, audiovisual stimuli were used in place of auditory-only stimuli as pursuit of auditory-only stimuli is poor (Leung et al., 2016). The procedure for the main experiment was the same as in Chapter 2, including the auditory and visual signals denoting the start of each interval. The experiment consisted of two experimental phases linked by the set of head movements recorded in the first. Instead of the 11 gain values used during the method of constant stimuli in Chapter 2, here only 7 gain values were used. The same ranges were employed, meaning that the gain values used for this auditory-only experiment ranged from 0.2-1.2 in steps of one sixth. The head tracking was also completed in the same way as in Chapter 2. Each trained head speed was investigated by completing training and main experiment pairs. This process was repeated three times, yielding 15 training and main experiment pairs presented in a random order.

### Head Speed Training Sessions

Training sessions consisted of a two-stage process that was run ahead of each replication of the main experiment. In Stage 1, participants were asked to track an audiovisual stimulus with their head. The stimulus moved independently along a sinusoidal path at a frequency of 1Hz at one of five amplitudes: 5, 12.5, 20, 27.5, or 35°. These correspond to median target speeds from 30.0°/s to 209°/s for the ROI defined in Experiment 1. Five and three-quarter periods were shown for each speed to generate 12 sweeps. In Stage 2, participants attempted to reproduce the trained head speed, this time using a head-stationary audiovisual fixation target moving with the nose as a guide (motion gain = 1). Again, they completed five and three-quarter periods, determined by recording the number of head

direction reversals detected in the head tracking as described in Chapter 2. The accuracy of head rotation was assessed by calculating the median head speed for each of the final 10 head sweeps as described in Chapter 2. If 7/10 sweeps had a median within  $5^\circ/\text{s} \pm 5\%$  of the desired training speed, performance on that training run was deemed sufficiently accurate. If not, the participant was given feedback on how many sweeps were accurate, and how many were too fast and/or slow, and the run repeated. Participants had to complete at least three training runs, with at least one successful run before progressing to each replication of the main experiment.

### Psychophysical Analysis

The psychometric function derived in Chapter 2 was used again in this experiment to obtain values for the variability of the image and non-image signals while also returning measurements for the Point of Subjective Equality (PSE) that are equivalent to Palamedes fits.

### Eye Tracking and Analysis

Eye movements were tracked using a Pupil Labs Pupil Core head mounted eye tracker. The tracker had a 120Hz sampling frequency and a front-facing world camera. The camera was used for calibration by having participants look at a 3 by 2 array of calibration points that can be seen in Figure 2.2 in the previous chapter. These were used to convert the eye tracker's normalised units into degrees. Two forms of analysis were performed on the eye movement data. First, the proportion of the eye position data that showed that participants were looking straight ahead, with a spatial ROI of  $\pm 3^\circ$ , was calculated. This was done by first excluding all samples with less than 0.6 confidence as defined by the Pupil Labs software, and then drift correcting trial by trial. During the head-moving interval, this corresponded to the participant focusing their eyes on the position of the auditory blob. During the head-stationary intervals, this corresponded to the participant avoiding pursuit of the auditory blob with their eyes. For reference, the outcome of this measure if participants were to perfectly counter-rotate to maintain fixation during the head-moving intervals was calculated.

In a second, more standard, analysis, after removing samples with less than 0.6 confidence, the gaps were filled using linear interpolation (e.g., Halow et al., 2023). If the waveform had 50% or greater dropped samples it was excluded from further analysis. Gaps were more frequent for the head-moving condition. The resulting waveform was then

smoothed using a Gaussian filter ( $\sigma = 16\text{Hz}$  in the frequency domain) and the 1st, 2nd and 3rd derivatives taken numerically, corresponding to velocity, acceleration and jerk. Saccades were detected using Wyatt's jerk analysis, with a jerk threshold of  $20,0000^\circ/\text{s}^3$  (Wyatt, 1998). Saccadic samples were removed from the analysis, along with 4 samples either side of each saccade detected, as well as the initial 20 samples at the start and beginning of each waveform. Mean velocity and speed were then calculated for the 3rd sweep for each head-movement interval, and the single sweep constituting the head-stationary intervals. As a comparison, the VOR needed to maintain stable fixation on a point at the same distance as the speakers was calculated, using an approximation given by Leigh and Zee (1999, p274):  $E = -H \left(1 + \frac{R}{D}\right)$ , where H is the head velocity, R = 0.1m, the distance from eye to centre of head rotation, and D = 1.2m, the distance of the speakers from the participant. Note that we assume the eyes were fixating at this distance because a visible fixation point appeared before the 2nd interval of each trial (for details, see the Procedure section of Chapter 2).

### Participants

All observers gave informed consent, and the experimental procedures were approved by the School of Psychology, Cardiff University Ethics Committee (EC.12.04.03.3123GRA2). Three experimenters and seven participants studying psychology at Cardiff University (8 female, 2 male) took part in the experiment. The student group were unaware of the purposes of the experiment. Participants completed at least two replications of the experiment, with eight participants completing three. Eye movements were recorded for nine of the participants. Only one of these normally wore spectacles, which were removed to allow the eye tracker to operate.

The code used for fitting the model using the two-phase paradigm, together with the raw data and summary data, can be found here:

[https://osf.io/qcz7w/?view\\_only=2e2bb846820d4862bcb02a036d3ee815](https://osf.io/qcz7w/?view_only=2e2bb846820d4862bcb02a036d3ee815)

### 3.3 – Results

Figure 3.1 plots the mean head rotation speed that participants executed in the main experiment. The dotted line indicates perfect performance with respect to the head speeds they were trained on in the head training session prior to data collection. Head movements

were reasonably accurate in the main experiment, producing a well separated set of rotation speeds that covered a wide range ( $F(4,45) = 16.89, p < 0.001$ ).

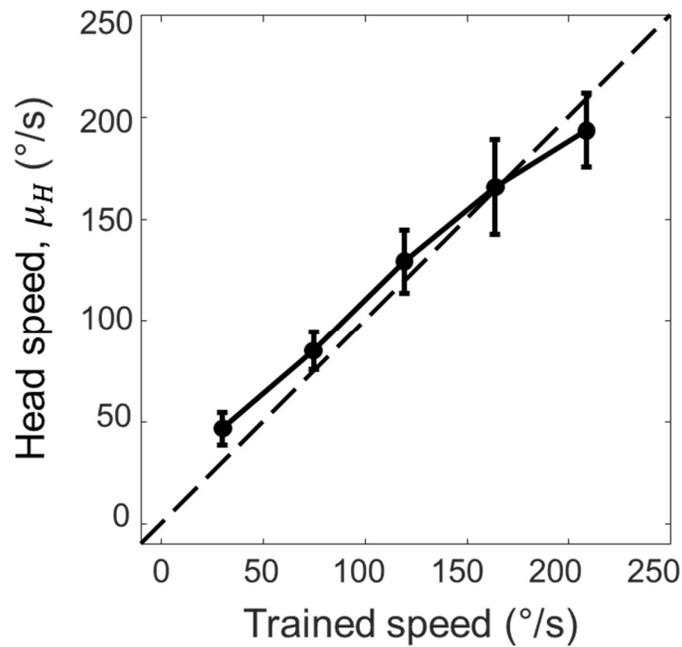


Figure 3.1: mean head velocity for 10 participants over the ROI defined in Figure 2.3A in Chapter 2. The dotted line indicates perfect performance with respect to the head speeds they were trained on in the head training session prior to data collection. Error bars represent  $\pm 1SE$ .

Figure 3.2A plots the percentage of eye position samples within a spatial ROI of  $\pm 3^\circ$ , as a function of trained head speed of 9/10 of the participants. As can be seen, all head-stationary intervals (open points) produced a similar level of performance. Accuracy was lower for the head-moving interval but was still quite high. As a reference, the dashed line

plots the same measure assuming participants had perfectly stabilised eye position via counter-rotation so as to fixate a world-stationary point.

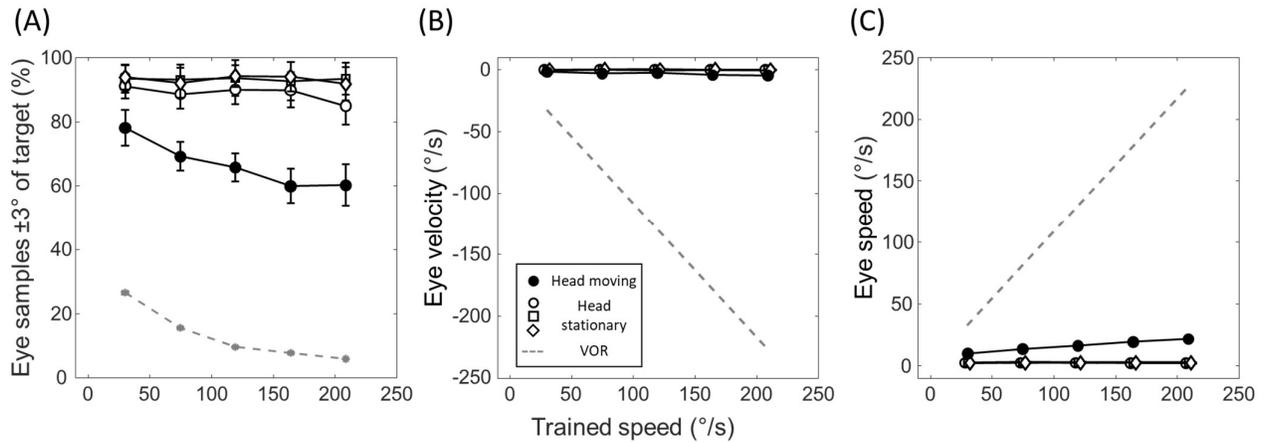


Figure 3.2: (A) Eye fixation accuracy expressed as % of samples within  $\pm 3^\circ$  of head-centred target. Open symbols correspond to the three head-stationary intervals (one in Phase 1, two in Phase 2) and closed circles the head-moving interval. The stars and dashed line correspond to the equivalent measure if participants had perfectly counterrotated the eye to remain fixed in world-centred coordinates. (B) Mean eye velocity over 9 participants for the two intervals in each phase. Open symbols correspond to the three head-stationary intervals (one in Phase 1, two in Phase 2) and closed circles the head-moving interval. The dashed line defines the VOR needed to maintain stationary gaze on a world-centred point at the same distance as the speakers. Negative velocities correspond to eye movement against the head rotation. (C) Mean eye speed (unsigned average). Format is the same as (B). Error bars represent  $\pm 1SE$  and can be smaller than symbol size.

Figure 3.2B plots the mean eye velocity for 9/10 of the participants, together with the predicted VOR needed to maintain fixation on a virtual point at the same distance as the speakers (see Methods for calculation). Figure 3.2C plots the mean speed. During the head-stationary intervals (open symbols), both mean velocity and speed were close to 0. In the head-moving interval, participants made small counter-rotations against the head velocity (ranging from  $-1.59^\circ/s$  for the slowest head speed to  $4.57^\circ/s$  at the fastest). When an ANOVA was conducted on the eye velocities, there was found to be a significant difference based on the interval ( $F(3,16) = 35.764$ ,  $p < 0.001$ ) and during post-hoc tests it was found that the significant difference was between the head-moving interval and each of the head-stationary intervals (see Table 3.1). The same analysis was performed on the eye speeds, and again a significant difference based on the interval was found ( $F(3,16) = 42.101$ ,  $p < 0.001$ ). Similarly,

during post-hoc tests it was found that the significant difference was between the head-moving interval and each of the head-stationary intervals (see Table 3.2).

	Phase 1 Interval 1	Phase 1 Interval 2	Phase 2 Interval 1
Phase 1 Interval 2	<0.001**		
Phase 2 Interval 1	<0.001**	0.934	
Phase 2 Interval 2	<0.001**	0.992	0.990

Table 3.1 shows the significance of the Tukey post-hoc tests performed after the ANOVA on eye velocity (\* denotes significance at the <0.05 level and \*\* denotes significance at the <0.01 level).

	Phase 1 Interval 1	Phase 1 Interval 2	Phase 2 Interval 1
Phase 1 Interval 2	<0.001**		
Phase 2 Interval 1	<0.001**	0.992	
Phase 2 Interval 2	<0.001**	1.000	0.997

Table 3.2 shows the significance of the Tukey post-hoc tests performed after the ANOVA on eye speed (\* denotes significance at the <0.05 level and \*\* denotes significance at the <0.01 level).

Crucially, there was no significant difference found between the velocity of any of the head-stationary intervals, confirming the assumption that the eye movements in all of the head-stationary intervals are similar. Also, despite being significantly different from the eye-stationary intervals, the mean eye velocity and speed in the head-moving interval were many

orders of magnitude smaller than the VOR necessary to counter the head rotation perfectly (see Figure 3.2).

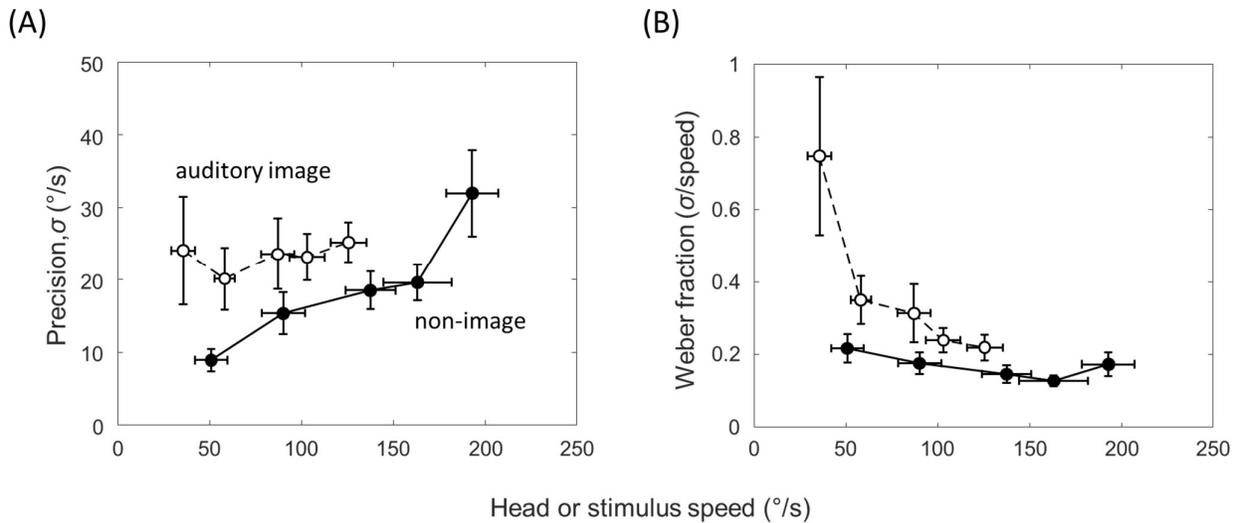


Figure 3.3: (A) Precision as a function of stimulus speed for the auditory image signal (open symbols) and non-image signals (closed symbols). Precision is defined as the standard deviation of the underlying signal distribution defined by the model in the Appendix. Note therefore that larger standard deviations correspond to less precise signals. (B) The same data expressed as Weber fractions i.e. standard deviation / speed. Error bars represent  $\pm 1SE$ .

Figure 3.3A plots the non-image signal precision (filled circles) and auditory image signal precision (open circles) as a function of the mean head or stimulus speed, respectively. The latter compresses horizontally because the speeds are set by the PSE obtained from Phase 2 of the main experiment. This corresponds to a motion gain of around 0.7 (see Figure 3.4 for the PSEs at each training speed). The horizontal compression is therefore around 30% compared to the closed circles.

For the auditory image signal, precision did not vary with stimulus speed ( $F(4,45) = 0.154$ ,  $p = 0.96$ , NS). However, for the non-image signal, precision decreased with head speed, such that the standard deviation of the underlying signal distribution increased ( $F(4,45) = 6.035$ ,  $p < 0.001$ ). Also evident is the fact that the auditory image signal is less precise than the non-image signal over the range of stimulus speeds tested. Figure 3.3B plots the same data as Weber fractions i.e., standard deviation divided by head or stimulus speed. For both types of signal, precision adheres Weber's law for medium to high speeds. Thus the Weber fractions are approximately constant over much of the range of speeds tested. At lower

speeds, however, the two functions differ, where Weber fractions start to rise steeply for the auditory image signal. This rise is reminiscent of other studies of Weber's law in the perception of auditory motion (Altman & Viskov, 1977). The same is not true for the non-image signal, where Weber's law appears to hold reasonably well across all head speeds investigated. This finding is similar to previous work using passive stimulation of the vestibular system (Mallery et al., 2010).

Figure 3.4 plots the mean PSEs from Phase 1 as a function of mean head speed. They appear similar for all head speeds experienced ( $F(4,45) = 0.57$ ,  $p = 0.69$ , NS). Hence the same proportional reduction in image speed was needed to match the perceived motion in the head movement interval. This value was around 0.7, replicating the findings in Chapter 2. Over a wide range of head speeds, therefore, moving auditory stimuli appear slower during head movement, akin to the Aubert-Fleischl phenomenon in vision. However, unlike vision and pursuit eye movement (Freeman et al., 2010; Powell et al., 2016), the non-image signal appears more precise than the auditory image signal, which could have important implications for the interpretation of the bias shown in Figure 3.4. This point is taken up in more detail in the Discussion section of this chapter.

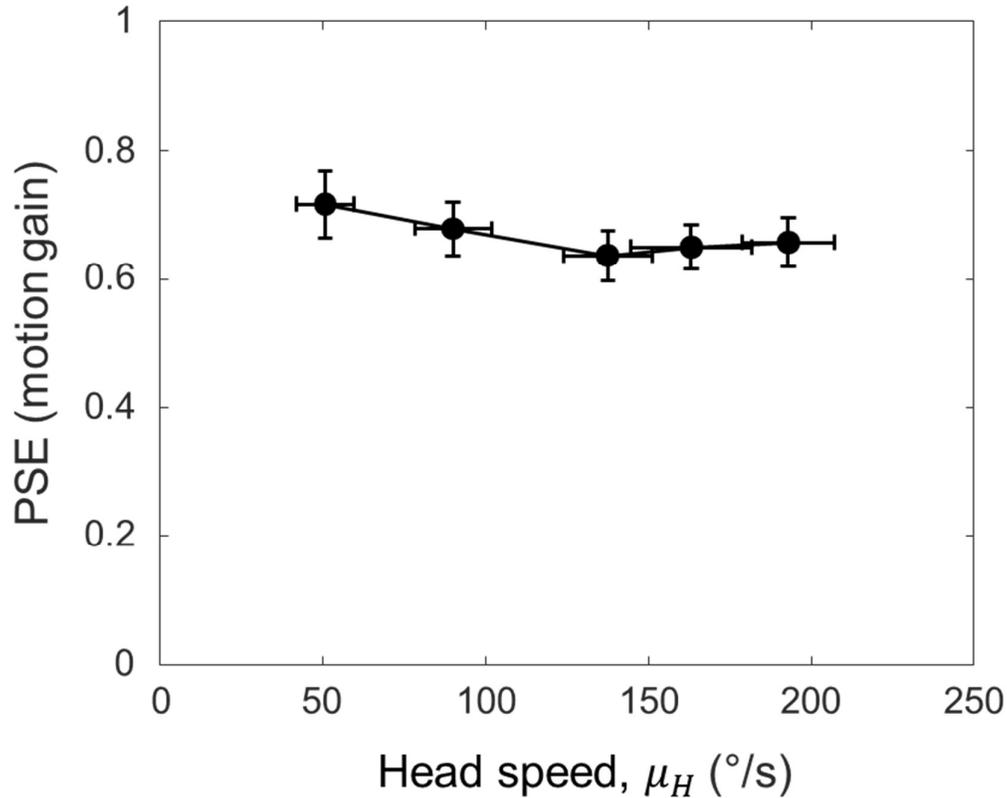


Figure 3.4: PSEs from Phase 1 as a function of head speed. Error bars represent  $\pm 1SE$ .

In Chapter 2, we found that non-image signal precision was higher using visual stimuli compared to auditory stimuli. To investigate further, we fit a regression line to the non-image precisions in Figure 3.3A using Deming's technique, a procedure that is used when both X and Y values are dependent measures with error (see Harrison et al., 2015). The result is shown in Figure 3.5, together with the two non-image precision values found in Chapter 2. The regression analysis shows good agreement between the experiments presented in Chapter 2 and here for auditory stimuli. The precision value from Chapter 2 (open circle) falls very close to the regression line determined by the experiment presented here. At the same time, however, the analysis casts further doubt on whether Weber's law can explain the better non-image precision found using visual stimuli (open triangle). If Weber's law were to account for the discrepancy in precision, the head speeds in the visual condition of Chapter 2 would need to be halved in order to shift the open triangle horizontally onto the regression line. Taken together with the evidence from Chapter 2, it appears that the modality of a stimulus affects

the precision of the non-image signal; one reason that this may be the case was discussed in Chapter 2.

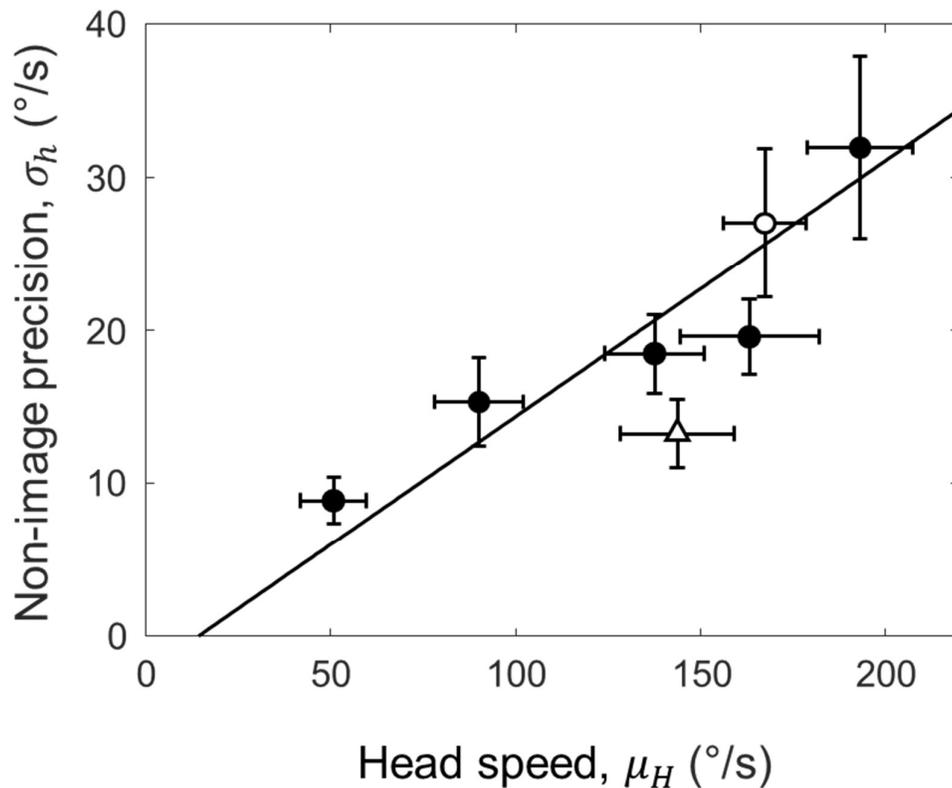


Figure 3.5: Deming regression (solid line) for the non-image signal precision data from this experiment (filled circles). The open circle is the non-image signal precision from the auditory condition of Chapter 2, and the open triangle is the visual condition of Chapter 2. Error bars represent  $\pm 1SE$ .

### 3.4 – Discussion

Results from this experiment suggest that Weber’s law describes the precision of both image and non-image signals for medium to high speeds. However, at low speeds Weber fractions for the auditory image signal rose steeply, unlike those for non-image precision. Both findings echo previous reports in the literature. For auditory motion based on ITDs, Altman and Viskov (1977) found Weber fractions were roughly constant from around 60-140 $^{\circ}/s$  but rose steeply at lower speeds. For vision, the same rise at slower speeds is found but matched by a similar rise at faster speeds (De Bruyn & Orban, 1988). For passive vestibular

stimulation, the trend is similar to the Weber fractions we found for the non-image signal, although further analysis by Mallery et al. (2010) showed that a power law with an exponent around 0.4 is better description of the raw thresholds than the straight line predicted by Weber's law. Similar behaviour has been reported for the variability in the vestibulo-ocular reflex, an eye movement controlled by the vestibular system (Nouri & Karmali, 2018). One implication is that the non-image signal we measured is dominated by the vestibular system. This assumes that the precision of motor signals behaves differently but we are unaware of any studies that have tried to measure the precision of motor signals on their own.

Further evidence that supports the findings of Chapter 2 can also be found in this experiment. Here we present more evidence that perceived speed slows during self-controlled head rotations, with this finding being extended to different rotational velocities, and we present further evidence that there is a difference between the precision of the non-image signal when it is compared to a visual or an auditory stimulus. A Deming regression showed that the difference between these modalities in Chapter 2 is greater than the difference that would be expected due to Weber's law, as the non-image signal precision value from the visual version of the experiment in Chapter 2 falls below the Deming regression line based on the precisions of the non-image signals at different head rotation speeds obtained in this experiment.

### Auditory Stimuli

Only auditory stimuli were used in this experiment for participants to compare to their self-movement. As seen in Chapter 2, this stimulus selection may return values for the precision of the non-image signal that are lower than if visual stimuli were used. We assume that this stimulus selection did not change the outcome of the investigation into Weber's law, however, as we assume that the difference in the absolute precision values found in Chapter 2 is due to an extra conversion step needed to convert the non-image signal into relevant units for comparison to auditory, over visual, stimuli (for more detail see the discussion section of Chapter 2). These conversions should have a constant additive effect on the variability measurements obtained at different speeds, such that the distribution of precision values should have a consistent shape, no matter the modality of the stimuli. We predict that, if this experiment was conducted with visual stimuli, a similar pattern of non-image signal precision would be obtained with slightly lower Weber fractions across the board. We would

also expect that the distribution of image signal precisions would be a U-shaped distribution in that case, in line with the evidence from De Bruyn and Orban (1988).

### Measurement Noise

As mentioned in Chapter 2, according to the work of Bentvelzen et al. (2009), when added to a visual stimulus, positional noise with a standard deviation of  $7.4^\circ$  was enough to double the threshold when compared to baseline. Using our measurement that the standard deviation of the position samples from our head tracker is  $\sim 0.55^\circ$ , we would expect that the thresholds obtained in a visual version of this experiment to change by 7.4%. As auditory spatial perception is considerably less precise than vision, we would expect that the effect of this noise would be even smaller than 7.4% for this auditory version of the experiment, and yet the difference between the precision of the auditory image signal and non-image signal is much greater than 7.4% across the board.

### Bayesian models of motion perception

In both Chapter 2 and this chapter, we found that perceived speed was lower when the head rotated. The bias was very consistent across modalities (Chapter 2) and stimulus speed (this chapter), adding to a large body of evidence showing that non-image signals based on eye rotation, head rotation, and hand/arm movement typically provide lower estimates of motion magnitude than signals encoding image motion in vision, hearing and touch (see below). On the face of it, the bias between non-image signals and image signals is puzzling because one might expect this type of constant error to be calibrated out by the perceptual system. One possible explanation, as mentioned in the General Introduction, is that the bias results from a Bayesian observer trying to optimise precision. According to the Bayesian hypothesis, the fact that early signals are noisy means that perception needs to infer the state of the world by combining imprecise measurements with prior expectations about the world state. The result is a posterior distribution that has greater precision than the original measurements, but not necessarily greater accuracy. As signals become noisier, the position of the posterior is increasingly pulled towards the prior distribution such that accuracy shifts. For motion, the claim is that the prior peaks at 0 because most objects are at rest (Weiss et al., 2002). Hence, as signals become noisier, speed estimates reduce.

The Bayesian framework has been used to explain why perceived visual speed slows at low contrast (Stocker & Simoncelli, 2006); why pursued objects appear slower (Freeman et al., 2010; Powell et al., 2016); why moving sounds appear slower when presented against background noise (Senna et al., 2015); and why tactile stimuli appear slower when made noisier or ‘pursued’ by an hand/arm movement (Moscatelli et al., 2019). It can also be used to account for individual differences in motion perception (Powell et al., 2016). Nevertheless, the overarching theory is not without its detractors (Hammett et al., 2007; Hassan & Hammett, 2015; Thompson et al., 2006). One simple test is to correlate measures of precision (e.g. thresholds) with bias – the Bayesian hypothesis predicts that as precision declines, perceived speed should slow. Many of the papers cited above show this to be case. However, there are a growing number of reports that this isn’t always true. Some recent studies in vision, hearing and vestibular research have shown changes in bias with little change in precision (Freeman et al., 2017; Freeman & Powell, 2022; Hassan & Hammett, 2015), and vice versa (Rideaux & Welchman, 2020). The findings of this experiment add to these seemingly ‘non-Bayesian’ set of results. They show that auditory image signals are less precise than non-image signals, even though the latter produce substantially lower estimates of motion magnitude.

## Conclusions

In agreement with current literature, we found that the non-image signal obeys Weber’s law over a wide range of stimulus speeds, unlike its image-based counterpart. Taking the results from this and the previous chapter, we also found that the magnitude of perceived motion is reduced during head movement for both vision and hearing. This finding is difficult to explain within a Bayesian framework because image precision was not greater than non-image precision over the wide range of stimulus speeds investigated.

Alongside Chapter 2, the results obtained from this chapter suggest that a standard Bayesian model may struggle to explain the finding that perceived speed slows during self-movement (self-movement compensation error). For this reason, it is important to test this standard Bayesian model in the situation that the paradigm that is central to this thesis generates. Chapter 4 will therefore investigate whether a Bayesian model can explain the self-movement compensation error, by attempting to replicate the findings, here, that the auditory image signal is less precise than the non-image signal and extend them to the visual

modality, and then by adding an external source of noise to investigate whether this yields Bayesian effects. The latter is a critical test of Bayesian modelling for the explanation of self-movement compensation errors.

# Can a Bayesian Model Explain Self-Movement Compensation Errors in the Context of Self-Controlled Head Rotations?

## 4.0 – Preface

In the previous chapter, it was found that our estimate of auditory velocity is less precise than our estimate of self-movement speed. There, it was noted that this evidence adds to a body of seemingly ‘non-Bayesian’ results, as a Bayesian model using the standard slow-speed prior would predict that, if this were the case, we would perceive objects as moving faster when we are stationary rather than when we are moving with the object (the opposite effect to the Aubert-Fleischl phenomenon and the results of the previous two chapters which show evidence of self-movement compensation errors).

However, the experiment in Chapter 3 utilised only auditory stimuli, so this chapter will also investigate whether similar ‘non-Bayesian’ results occur with visual stimuli. Also in this chapter, one of the critical assumptions of Bayesian modelling will be tested as a source of external noise will be added to the visual and auditory stimuli. Assuming that this causes a change in the precision of motion perception, it should also cause a change in bias, due to the

inherent link between precision and bias in Bayesian modelling (see Figure 4.1, a copy of Figure 1.4).

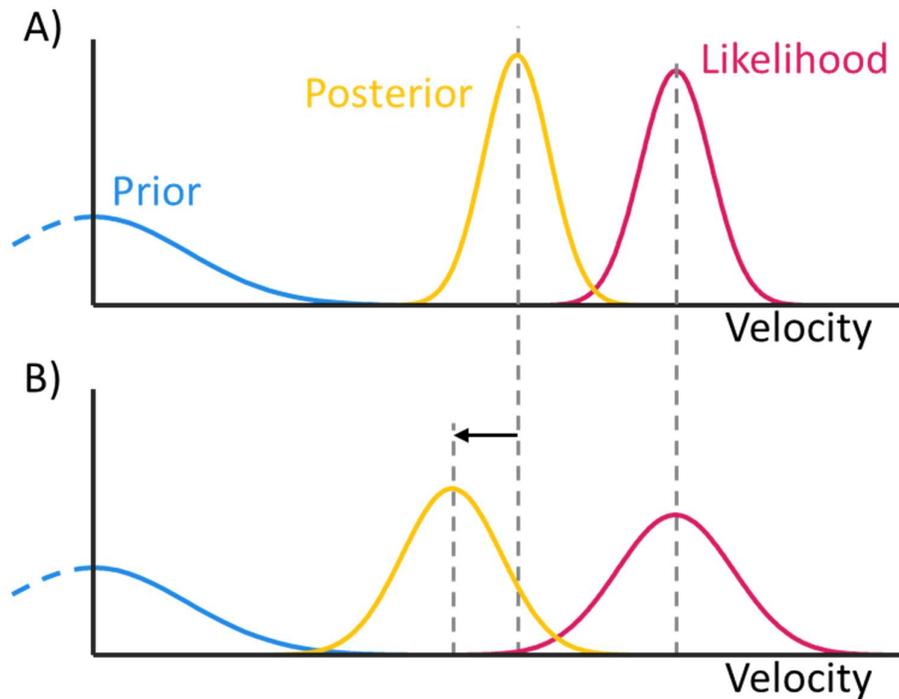


Figure 4.1: (A) a standard Bayesian model of motion perception. (B) the same Bayesian model but with more noise in the sensory evidence (wider likelihood distribution). This causes a shift in the posterior, towards the prior, as denoted by the arrow.

## 4.1 – Introduction

As explained in the General Introduction, the standard Bayesian approach assumes that the sensory evidence is, on average, accurate. Bayesian theorists point out that sensory evidence is noisy, so the observer has to make best guesses to determine what the sensory evidence represents in terms of the state of the world. One route to a best guess is to combine current sensory evidence with prior expectations about the world, and weight this according to the precision of the evidence. Conceptually, the prior and the likelihood (the distribution of sensory evidence) are thought of as distributions, which when multiplied together yield a posterior distribution, the peak of which represents the ‘best guess’ (Figure 4.1). The location of this peak is shifted away from the accurate peak of the likelihood as accuracy is sacrificed for extra precision (the posterior is more precise or reliable, and less variable or contains more noise, than the sensory evidence, which is demonstrated by the likelihood being wider than

the posterior, see Figure 4.1) The multiplication, by definition, places more weight on the likelihood if it is more precise. Hence, the precision of sensory measurements has a direct impact on perceptual bias because it determines the extent to which the peak of the posterior differs from the peak of the likelihood (which is assumed to be accurate).

Bayesian models have been used to understand motion perception in a variety of contexts. There is a prior expectation of slow movement in visual speed perception due to the tendency of objects to move slowly in the world (Weiss et al., 2002). Accordingly, less precise sensory evidence should result in a slower perception of movement because more weight is given to the slow-speed prior (see Figure 4.1). We can manipulate the variability of sensory evidence by changing stimulus properties. Changing variability like this has been shown to have a direct effect on perceived speed. For instance, low contrast stimuli appear to move slower than higher contrast stimuli at slow speeds (Thompson, 1982) and this effect has been predicted by Bayesian models based on the logic presented above (Ascher & Grzywacz, 2000; Weiss & Adelson, 1998). A similar effect occurs for auditory stimuli presented in quiet or loud backgrounds, with sounds that have low signal-to-noise ratios (those presented in louder backgrounds, for example) appearing to move slower than those with higher signal-to-noise ratios. This effect can also be predicted by a Bayesian model (Senna et al., 2015). When it comes to self-movement compensation errors, objects are perceived as moving faster when their movement is interpreted by a stationary observer than when the same movement is interpreted by an observer that has equivalent self-movement (with the eyes or head) to the movement of the object. Our estimate of object movement should therefore be faster than our estimate of self-movement when the two are equivalent. In a Bayesian model with a slow-speed prior, this would occur if the sensory evidence pertaining to self-movement (in this thesis, a non-image signal) is less precise than the sensory evidence pertaining to object movement (an image signal). This appears to be the case for smooth pursuit eye movement, as objects appear to move slower, and speed discrimination is worse, when objects are followed by eye movements compared to with the eyes stationary (Freeman et al., 2010). A similar effect has been observed during finger movement, where the movement of a surface appears slower, and the precision of speed discrimination is worse, when participants move their fingers with a platform compared to with their finger stationary (Moscatelli et al., 2019). Moscatelli et al. (2019) also included an experiment where they manipulated the texture of

the surface and found that the effect was more prominent with a textured surface and disappeared with a smooth surface (this is a method of adding variability to the sensory evidence).

However, in Chapter 3, it was shown that similar slowing of perceived auditory speed during head rotation is not accompanied by more precision when the head is stationary. One explanation for this may be that auditory movement perception is not Bayesian. Another is that the standard Bayesian framework does not apply; perhaps the assumption that image and non-image signals are accurate is violated for auditory movement perception, or perhaps hearing has a prior for fast movement (after all, sounds are moving waves which are typically generated through movement). For these reasons, the experiment in this chapter compared vision and hearing. Also included in this experiment was a manipulation that added external noise to manipulate the precision of the sensory evidence when the head was stationary. This manipulation was used as, according to a Bayesian model, it should cause a change in the bias, irrespective of whether there is a slow-speed or fast-speed prior and irrespective of the accuracy of the sensory evidence. This manipulation, then, will be a critical test of the Bayesian nature of auditory movement perception.

The aim of the two-phase paradigm used in the previous chapters was to measure the precision of the non-image signal that we use to encode our self-movement when there is no relevant movement information in the image. To do this, a measurement of the precision of the image signal was obtained, from Phase 2, and used to measure the precision of the non-image signal from Phase 1. This was necessary because Phase 1 contained internal noise from both the image and non-image signals. In order to isolate the noise due to the non-image signal from Phase 1, a measurement for the precision of the image signal was obtained in Phase 2, with the stimuli in all of the head stationary intervals (the second interval of Phase 1 and both intervals of Phase 2) moving at the same speed. While it was shown in the previous chapter that the auditory image signal is relatively constant in its precision at a wide range of different speeds, the same cannot be assumed for vision, for example, the results of De Bruyn and Orban (1988) suggest a more pronounced change in precision as a function of visual speed. To match the speeds of the image signals in Phases 1 and 2, we set the motion gain of the standard in Phase 2 to be equivalent to the Point of Subjective Equality (PSE). This meant that the average perceived speed of the objects during the head-moving interval and each

interval of Phase 2 were equivalent, however, the average objective speeds were different (motion gain of 1 for the head-moving interval in Phase 1, and average motion gain equal to the PSE for both intervals in Phase 2). A consequence of this is that we were unable to directly compare our measurements of the precisions of the image and non-image signals. In the experiment presented in this chapter, we added a third phase to the paradigm that allowed us to obtain a measurement of the precision of the image signal at the same average objective speed as the head-moving interval of Phase 1, by replicating Phase 2 except with stimuli having average motion gain equal to 1, rather than the PSE.

The key manipulation in this experiment is the addition of external stimulus noise to produce a change in the precision of the image signals. To manipulate stimulus noise, we included dynamic changes in stimulus width. This is different from adding positional (or speed) noise that has been used before (Bentvelzen et al., 2009; Rideaux & Welchman, 2020). However, the reason for the inclusion of this noise manipulation is that, in a pilot study, we found that dynamic changes in stimulus width produced salient changes in precision for both hearing and vision, whereas positional/speed noise only produced sizeable changes for vision.

## 4.2 – Methods

### Stimuli and Materials

The stimuli and materials used in this experiment were similar to those used in the previous chapters. The only difference between the stimuli used here and in Chapter 2, was the inclusion of a dynamic stimulus width jitter in the ‘jitter’ condition. The jitter was added to all of the 5 intervals where the participant was stationary (the second interval of Phase 1 and both intervals in Phases 2 and 3), and this was achieved by updating the standard deviation of the Gaussian distribution that spatially windowed the visual and auditory stimuli. The standard deviation was updated every 0.2s, with one of four standard deviations randomly chosen from a set of predetermined values ( $\sigma = 7.5, 15, 22.5,$  and  $30^\circ$  for the auditory stimuli and  $\sigma = 1.05, 2.1, 4.2,$  and  $6.3^\circ$  for the visual stimuli). An example of the effect of the dynamic width jitter on the standard deviation of the stimulus can be seen in Figure 4.2. Technical note: due to a mistake in coding, the 2 sets of jitters used resulted in a mean stimulus width greater than the non-jittered stimuli (mean stimulus width of the non-jittered auditory stimulus was  $7.5^\circ$  vs theoretical mean width of  $18.75^\circ$  for the jittered auditory

stimulus, and mean stimulus width of the non-jittered visual stimulus was  $1.05^\circ$  vs theoretical mean width of  $3.41^\circ$  for the jittered visual stimulus). However, this does not affect the predictions because an increase in average stimulus width would likely have the same effect (a decrease) on image signal precision as the dynamic stimulus width jitter, or perhaps no effect at all.

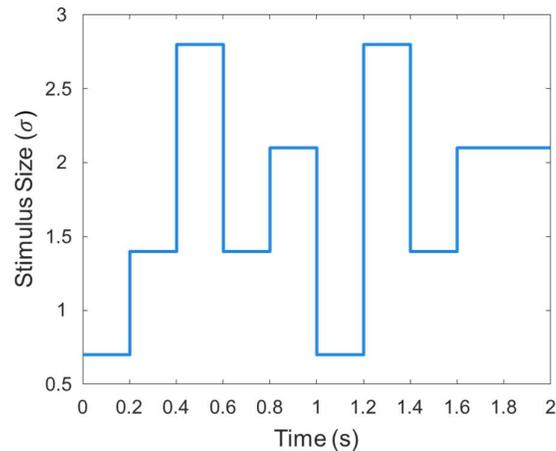


Figure 4.2: an example of the standard deviation of the Gaussian distribution that windowed the jittered stimulus across the duration of the presentation.

## Procedure

The procedure was also similar to previous experiments. Phase 1 still consisted of a ‘head-moving’ standard and a ‘head-stationary’ test, pre-empted by a short beep and a blue blob respectively. Each interval only started once the participant had their head ‘centred’ as defined in Chapter 2 (10 consecutive head-tracker samples within  $\pm 7.5^\circ$  of the centre of the LED/speaker array). The ‘head-moving’ standard appeared after the participant began rotating their head, on the third sweep of the head rotation, and moved with the head. The ‘head-stationary’ test was based on the motion of this third sweep, with a multiplicative gain factor applied. In Phases 2 and 3, the same beep and light were used to signal the start of each interval, with all of the intervals head-stationary. The stimuli were again based on the head recordings from the ‘head-moving’ intervals from Phase 1 with gain factors applied. All three phases consisted of the same number of trials presented in the same order. The presentation order of Phase 2 and 3 was randomised, however Phase 1 was always presented first.

As mentioned in the introduction, the difference between Phases 2 and 3 was the gain factor that was applied to the stimuli. In Phase 2, the standard was multiplied by a gain factor

equivalent to the PSE from the first phase. In Phase 3, the standard was multiplied by a gain factor of 1, so that the stimulus moved in the exact same way as the participants' head movements from the first interval of Phase 1.

As in the previous chapters, the method of constant stimuli was used, here with 7 gain values. The ranges of the gain values was consistent with the previous chapters with the visual version utilising gain values from 0.4-1 in steps of 0.1 and the auditory version utilizing gain values from 0.2-1.2 in steps of one sixth. Each gain value was repeated 10 times giving 70 trials per phase, and the three-phase paradigm was completed three times by each participant for the visual and auditory versions in both the jittered and non-jittered conditions. Head and eye tracking was completed in the same manner as Chapter 3. Participants were not trained to move at any particular speed and were instead instructed to move at whatever reasonable speed they felt comfortable at.

#### Psychophysical analysis

The same analysis as detailed in the appendix of Chapter 2 was used to measure the PSE and the precision of the image and non-image signals.

#### Head and Eye Movements

The head and eye movements of participants were recorded using the same Polhemus Liberty head tracker and Pupil Labs Pupil Core eye tracker as in Chapter 3. Only the position analysis used in Chapter 3 was used here as similar conclusions were made by all types of eye movement analysis in Chapter 3.

#### Participants

All observers gave informed consent, and the experimental procedures were approved by the School of Psychology, Cardiff University Ethics Committee (EC.12.04.03.3123GRA2). Six participants took part in the experiment (2 female, 4 male), including three of the experimenters and three participants who were naïve to the purposes of the experiment. Eye movements were recorded for five of the participants. Two of these usually wore spectacles, which were removed to allow for the eye tracking.

## 4.3 – Results

### Non-Jittered Condition Only

Figure 4.3 shows the PSE in terms of the motion gain applied to the test stimulus in the auditory and visual conditions. It is clear that the PSE in each modality is less than 1 (Auditory:  $t(5) = -11.550$ ,  $p < 0.001$ ; Visual:  $t(5) = -15.769$ ,  $p < 0.001$ ), with the object moving around 30% slower when the participants perceived the object movement in the head-stationary condition (test) as equivalent to the head-moving condition (standard). This is the same self-movement compensation error effect that has been shown in the previous two chapters and it is again the case that this effect is consistent across both the auditory and visual modalities. For a Bayesian model to explain this effect, it would need to be the case that the precision of the image signal was greater than the precision of the non-image signal.

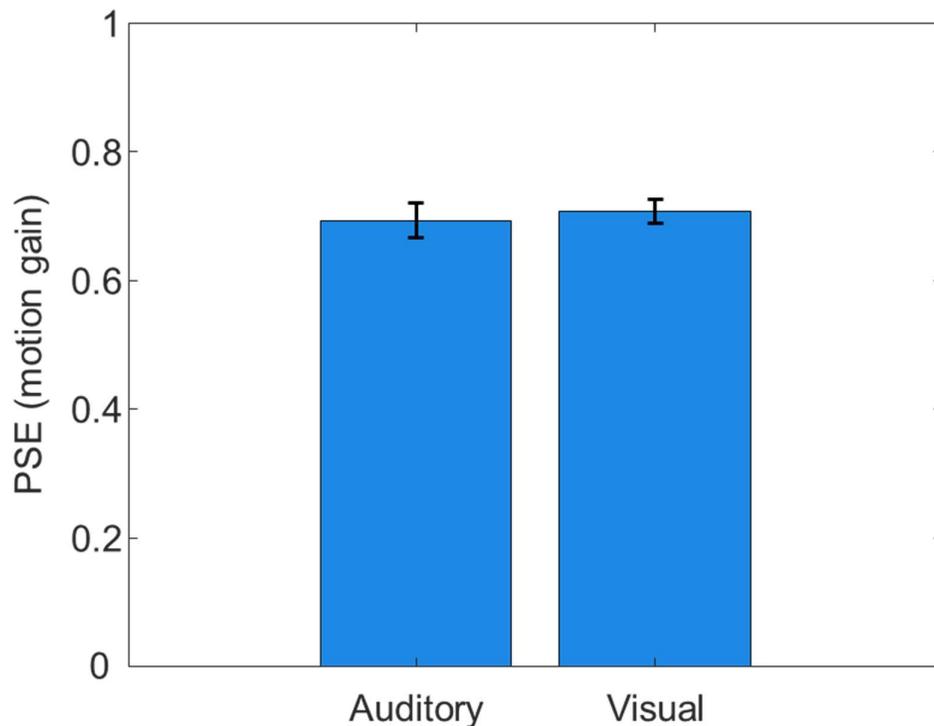


Figure 4.3: Point of Subjective Equality (in terms of the motion gain of the test stimulus) for the auditory and visual versions of Phase 1. Error bars represent  $\pm 1SE$ .

In Figure 4.4, the precisions of the non-image (pink bars) and image (blue bars) signals are plotted, and it is clear that the image signals are not more precise than the non-image signals. This conclusion is supported by an ANOVA that showed no significant interaction

between modality and signal type ( $F(1,20) = 0.198$ ,  $p = 0.661$ , NS) and no significant main effect of modality or signal type (Modality:  $p = 0.154$ , NS; Signal type:  $p = 0.505$ , NS). With these results, there is evidence that the standard Bayesian model is not relevant here. However, this evidence could point to a Bayesian model that includes biased signals, similar to the model suggested by Freeman and Powell (2022).

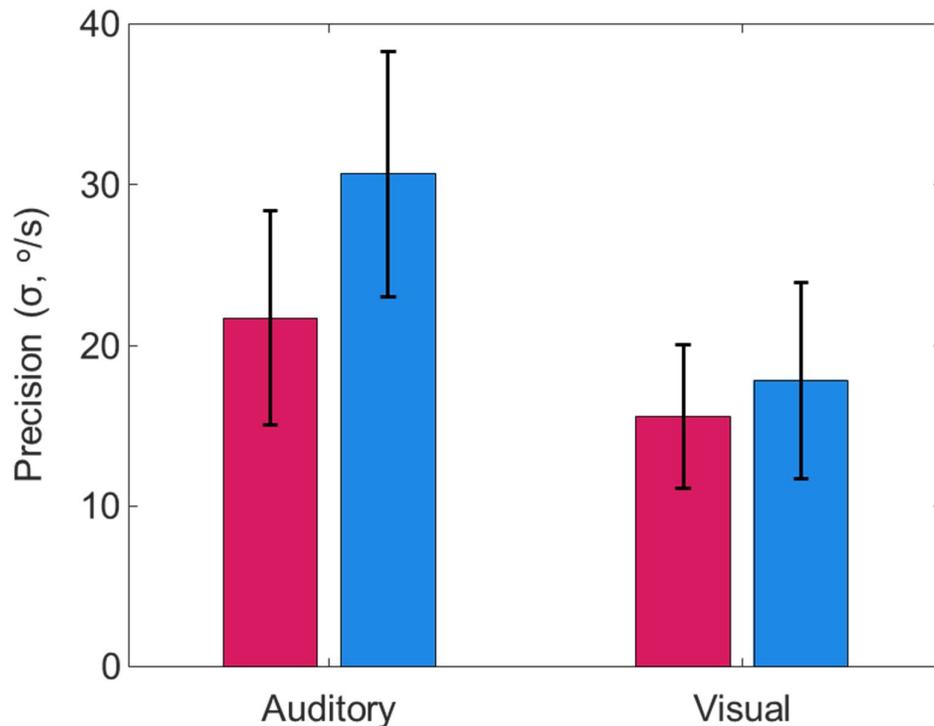


Figure 4.4: Precisions of the non-image (pink) and image (blue) signals in the auditory and visual versions of the experiment. Lower bars mean more precision. Error bars represent  $\pm 1SE$ .

#### The Effect of Dynamic Size Jitter

The critical test of the Bayesian model in this experiment is the effect of the dynamic size jitter. The jitter should cause the image signal for the jittered stimulus to be less precise than the image signal for the non-jittered stimulus. Figure 4.5 shows the precisions of the image signal in the non-jittered (blue bars) and jittered (yellow bars) conditions. Here, it appears as though the jitter successfully decreased the precision of the visual image signal but did not decrease the precision of the auditory image signal. To verify this, an ANOVA was conducted which, instead, showed no significant interaction between modality and jitter condition ( $F(1,20) = 0.143$ ,  $p = 0.709$ , NS) and no significant main effect of modality or jitter

condition (Modality:  $p = 0.141$ , NS; Jitter:  $p = 0.516$ , NS). These results suggest that the jitter manipulation was not successful at decreasing the precision significantly in either modality.

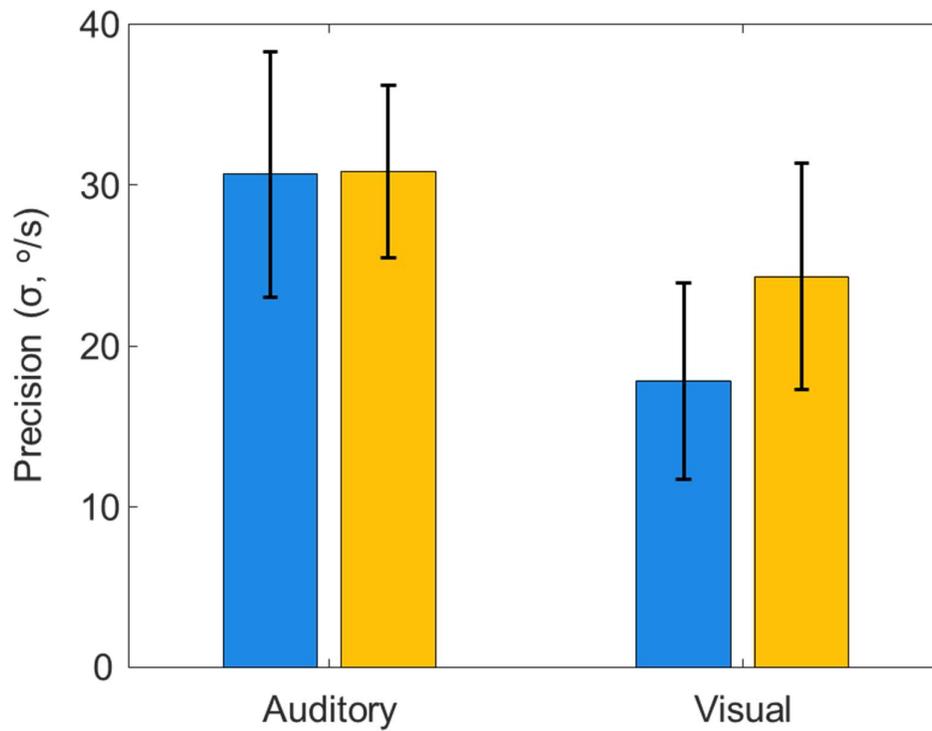


Figure 4.5: Precisions of the image signals in the non-jittered (blue) and jittered (yellow) conditions, in the auditory and visual versions of the experiment. Lower bars mean more precision. Error bars represent  $\pm 1SE$ .

As shown above, there is no statistically significant difference between the precisions of the image signals in the non-jittered and jittered conditions. It should be expected, then, that the PSE in the jittered conditions should not differ from the non-jittered conditions. As Figure 4.7 shows, this is the case as the PSEs are consistent throughout the auditory and visual, non-jittered (blue bars) and jittered (yellow bars) conditions. This conclusion is supported by an ANOVA that showed no significant interaction between modality and jitter condition ( $F(1,20) = 1.008$ ,  $p = 0.327$ , NS) and no significant main effect of modality or jitter condition (Modality:  $p = 0.794$ , NS; Jitter:  $p = 0.583$ , NS).

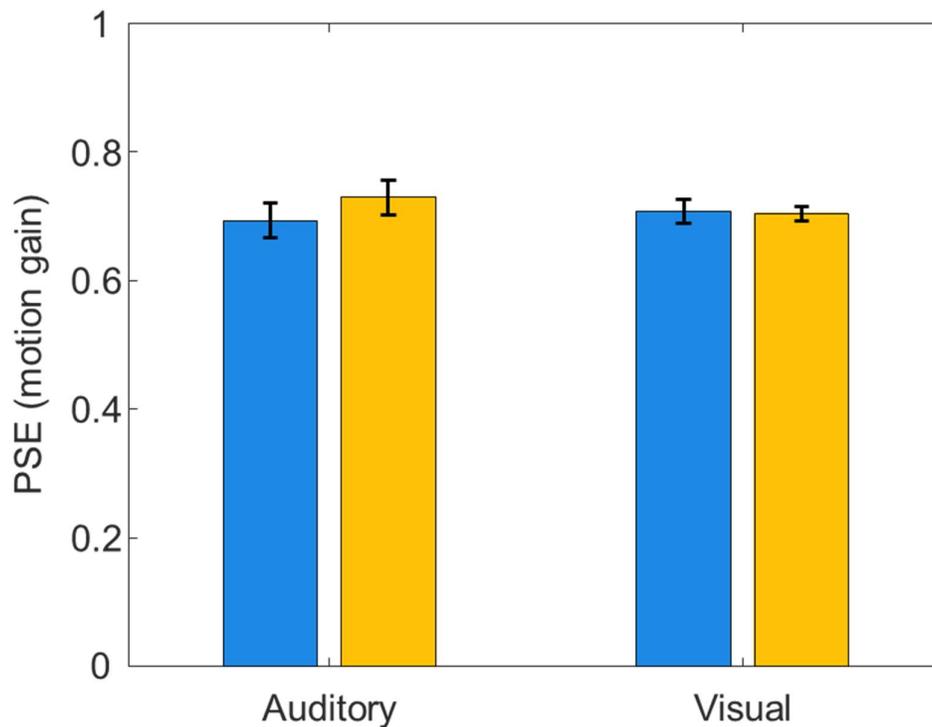


Figure 4.7: Point of Subjective Equality (in terms of the motion gain of the test stimulus) for the auditory and visual versions, both without (blue) and with (yellow) dynamic size jitter, of Phase 1. Error bars represent  $\pm 1SE$ .

### Head and Eye Movements

Figure 4.8A plots the mean head speed of participants in the head moving intervals of both the auditory and visual versions of the experiment in the non-jittered (blue bars) and jittered (yellow bars) conditions. The head movement speeds appear to be similar across all of the conditions. This conclusion is supported by an ANOVA that showed no significant interaction between modality and jitter condition ( $F(1,20) = 0.055$ ,  $p = 0.818$ , NS) and no significant main effect of modality or jitter condition (Modality:  $p=0.862$ , NS; Jitter:  $p=0.956$ , NS).

Figure 4.8B plots the percentage of eye samples within a spatial ROI of  $\pm 3^\circ$  for each condition, the same style of analysis can be found in Chapter 3. In Chapter 3 this analysis was followed with a more standardised eye velocity and speed analysis but similar conclusions were made. Here, it can be seen that the head stationary intervals (open symbols) have a similar level of fixation. Accuracy is lower in the head moving interval (closed symbols), like in

Chapter 3, but still much higher accuracy than if participants performed counter-rotation (dashed line).

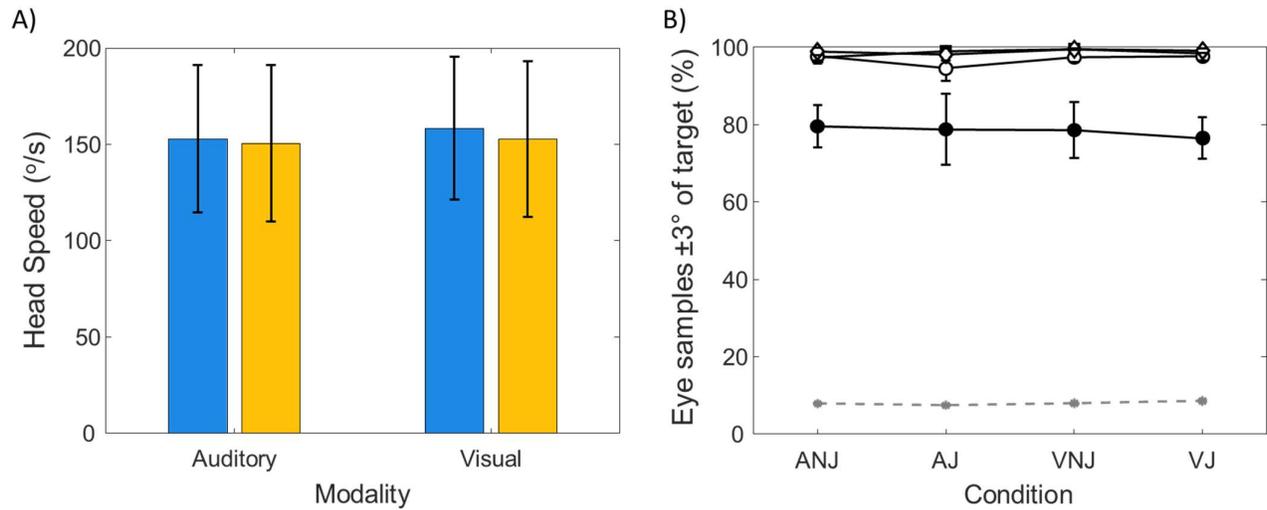


Figure 4.8: (A) head rotation speeds for participants in the auditory and visual versions, both without (blue) and with (yellow) dynamic size jitter, of interval 1 of phase 1. (B) Eye fixation accuracy expressed as % of samples within  $\pm 3^\circ$  of head-centred target. Open symbols correspond to the three head-stationary intervals (one in Phase 1, two in Phase 2) and closed circles the head-moving interval. The stars and dashed line correspond to the equivalent measure if participants had perfectly counterrotated the eye to remain fixed in world-centred coordinates. Error bars represent  $\pm 1SE$ .

#### 4.4 – Discussion

In the experiment presented in this chapter, the precisions of the image and non-image signals were measured with and without external noise consisting of dynamically changing stimulus width. This noise manipulation was unsuccessful in adding variability to the either the visual or the auditory image signal. It is unclear why this occurred, because pilot experiments showed a reasonable change in precision.

With the self-movement compensation error found throughout this thesis, a standard Bayesian model of movement perception, as detailed in Chapter 1, predicts that the non-image signal should be less precise than the image signal. However, this was not the case for either vision or hearing. Without jitter, the precision of the image and non-image signals were similar. In the previous chapter, it was found that the auditory image signal was less precise than the non-image signal. A similar effect was not significant here but can be seen by

observation of the data. These results are difficult to explain using the standard Bayesian model.

Unfortunately, the key manipulation for a test of Bayesian models when it comes to explaining self-movement compensation errors, the dynamic size jitter, was unsuccessful. This unsuccessful manipulation lead to no change in bias, which lends some support to a non-standard Bayesian model, as similar signal precisions caused similar biases.

### The Assumptions of the Bayesian Model

It is an integral assumption of standard Bayesian modelling that the sensory evidence from each signal is accurate. In the case of this experiment, that also means that the image and non-image signals we measure contain sensory evidence that is equivalent in its average speed (as the stimuli that we use to determine the precisions of the signals are matched in their speed; motion gain = 1). Another assumption of the Bayesian models of movement perception is that the priors for the motion in the image and non-image signals are equivalent. If the priors are equivalent and the average speed of the sensory evidence is equivalent, the only way to produce a relative difference between the average of the posteriors is for the sensory evidence to have different precisions. The experiment presented here showed that this is not the case in vision or hearing. I will now inspect the assumptions of the Bayesian model to attempt to address these findings.

First, I would like to address the assumption that the image and non-image signals share the same prior. While this may not appear intuitive, the slow-speed prior for visual object movement perception has been widely accepted in the literature with stationary observers (e.g., Ascher & Grzywacz, 2000; Hürlimann et al., 2002; Stocker & Simoncelli, 2006; Weiss & Adelson, 1998; Weiss et al., 2002; Welchman et al., 2008). In these papers, as participants remain stationary, it is difficult to know whether the prior is for slow object movement or slow motion in the images that our receptors obtain, as these are equivalent. However, it is typically assumed that the prior is for slow object movement. If there is no object movement, as this prior suggests is typical, and assuming that the sensory evidence from the image and non-image signals is accurate, self-movement generates equal and opposite motion in the image and non-image signals. This inherent link between image and non-image motion means that, whatever the prior for motion in the non-image signal is

during self-movement, a reflected prior should be expected for motion in the image signal, and vice versa (it should also be assumed that the prior is symmetrical as stimuli are not likely to tend to move rightwards more than leftwards, for example).

As mentioned earlier, it is typically assumed that the slow-speed prior found in visual object movement perception relates to object movement rather than the motion in the image, so the question remains whether image motion and non-image motion also have slow-speed priors. To address this assumption, Figure 4.10 shows that if the prior was not centered on no motion, and was instead an expectation for motion that was equivalent to the motion in the sensory evidence, it would be impossible to create any kind of bias. Assuming that the prior is for slow-speed motion gives the Bayesian model its best chance of generating an appropriate bias in order to explain self-movement compensation errors.

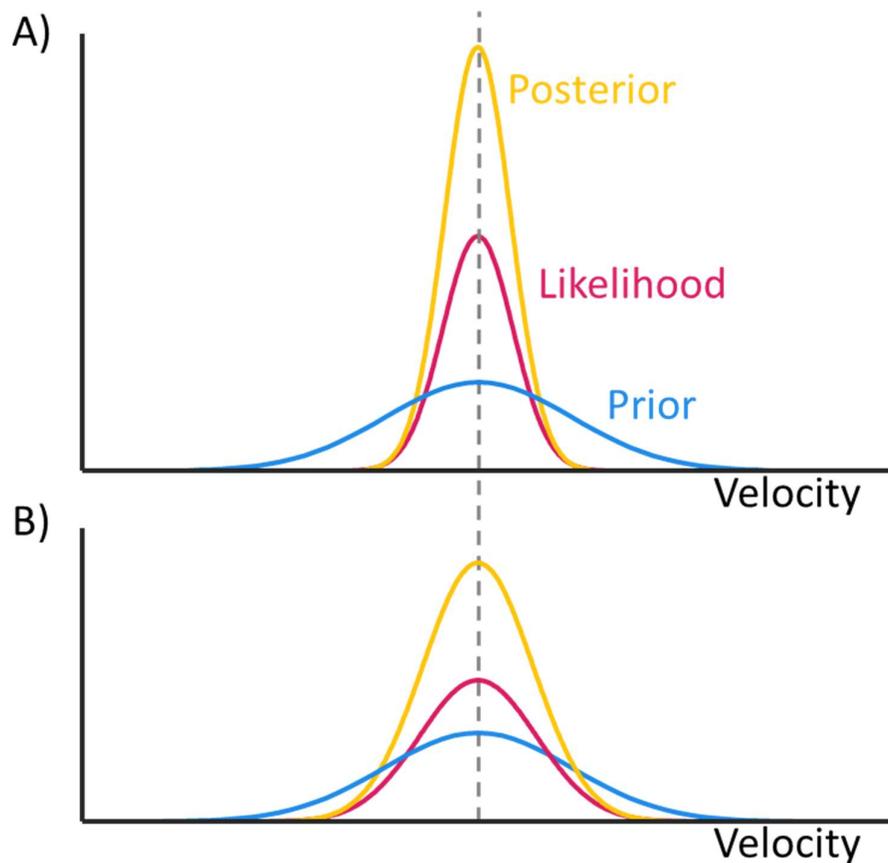


Figure 4.10: (A) a Bayesian model for movement perception that, instead of having a slow-speed prior, has a prior that is equivalent to the motion in the sensory evidence. (B) the same Bayesian model but with more noise in the sensory evidence (wider likelihood distribution). This does not cause a shift in the posterior, unlike the standard Bayesian model in Figure 4.1.

Now, to address the assumption that the sensory evidence from each signal is accurate. Introducing a bias term so that one or both of the signals could contain inaccurate sensory evidence would allow for the finding that object movement appears slower during self-movement. Relaxing this assumption could be thought of as a hybrid model, because the original theories about inaccurate signals (e.g., Dyde & Harris, 2008; Freeman & Banks, 1998; Mack & Herman, 1973; Wertheim, 1981) that the Bayesian models appeared to usurp, would be introduced back into a Bayesian framework. It would also allow for differences in the priors, as the biased sensory evidence over time would generate different priors for image and non-image motion. Here, it is important to note that without constraints, a Bayesian model would be able to fit any dataset. It is with constraints that we can limit the free parameters and find a realistic model of our perception. Current Bayesian models apply the assumption of accurate sensory evidence because it is thought that if signals were consistently inaccurate, over time we would learn to accurately compensate for these errors, whereas differences in the precisions of the signals would be more challenging to equate over time.

#### “Anti-Bayesian” Evidence

This is not the first experiment to find evidence that standard Bayesian accounts cannot explain. For example, researchers have investigated the perceived speed of objects with different luminance and found non-Bayesian results (Freeman & Powell, 2022; Hassan & Hammett, 2015). Freeman and Powell conducted four experiments with mixed support for the standard Bayesian model before concluding that the assumptions of the model need to be changed. It was found by these researchers that decreasing the luminance of a moving stimulus increased its perceived speed. Their participants were more precise at judging the speed of their lower luminance stimuli, which lends support to a standard Bayesian model, as the additional precision at low luminance would decrease the bias towards the slow motion prior. However, in a further experiment, Freeman and Powell accounted for contrast changes when manipulating luminance. With their “equicontrast” stimuli, the bias such that low luminance stimuli appeared faster still occurred, however, the difference in sensory precision no longer remained. This result cannot be explained by a standard Bayesian model. Later, Freeman and Powell introduced external noise by adding luminance jitter which generated results that appeared to restore Bayesian behaviour. Freeman and Powell concluded that the

standard Bayesian assumption that the sensory information is accurate needs to be relaxed, as their behavioural results fit with a Bayesian observer with biased inputs. Unlike with the luminance jitter added by Freeman and Powell (2022), our dynamic size jitter did not produce significant changes to the precisions of the image signals, however the jitter also did not produce changes in the bias. Overall, our results, along with those of Freeman and Powell, suggest that a hybrid Bayesian model with biased sensory evidence explains behaviour well.

Our manipulation of precision did not cause significant changes in the precisions of the image signals. Other methods of manipulating the precision of sensory evidence are, of course, available, including changing the speed, position, luminance, or constant size of the stimulus to name just a few. Rideaux and Welchman (2020) investigated whether natural scene statistics could explain the effect of contrast on visual speed perception, whereby lower contrast stimuli appear to move slower. They compared this effect to manipulating precision by adding speed noise to the stimulus trajectory. They found that while low contrast stimuli did indeed appear to move slower than high contrast stimuli, high contrast stimuli with speed noise did not appear to move slower than during monotonic movement. This suggests that certain types of additional variability may create bias changes like those suggested by Bayesian models, while other forms of additional variability may not. That said, their findings are made more puzzling by the fact that the contrast manipulation produced no change in precision, in contrast to studies by Powell et al (2016) and Warren & Champion (2017).

Nevertheless, Rideaux and Welchman, in their analysis of scene statistics, found that contrast and speed are linked in nature and assumed that this means that we expect stimuli with low contrast to move slower, due to this being commonly found in natural scenes. This suggests that the manipulations of contrast and speed jitter are different, as a manipulation to contrast directly affects the overall perceived speed in the scene, whereas the overall perceived speed in the scene is not affected by their speed jitter, as it causes no change to the average speed of the stimulus. It appears to be the case that selecting the right manipulation for the external variance of the image is an important task when investigating how the precisions of signals change with external variance. A further study with multiple different precision manipulations may find results that provide evidence that even a hybrid Bayesian model is inappropriate in the context of movement perception.

## Conclusions

In this chapter, I have presented an experiment that allowed for an investigation into the Bayesian nature of the perception of object movement during self-controlled head rotations. This experiment consisted of a three-phase version of the paradigm used throughout this thesis so far, with the third phase being used to provide image and non-image signal precision measurements for stimuli that moved at the same speed. A similar bias to the previous chapters was found such that the perceived speed of an object slows during self-movement (termed a self-movement compensation error in this thesis), however the image signal was not significantly more precise than the non-image signal in either the auditory or visual modality. This replicated the result shown for the auditory modality in the previous chapter.

Adding a dynamic stimulus width jitter did not cause significant changes to the precisions of the image signals, meaning that a hybrid Bayesian model explains our data well. It may, however, be the case that a different method of adding external noise may create a situation where even a hybrid Bayesian model is inappropriate.

# Derivation of a Bayesian Model Which Accounts for Trial-by-Trial Variation In Stimuli

## 5.0 – Preface

So far in this thesis, a method has been introduced that can measure the precisions of the image and non-image signals that we use to interpret the speed of object movement during self-movement. In Chapter 2, a new psychometric function was derived that allowed for the fact that the stimuli used in these experiments are fully dependent on the self-movements of participants which are variable from trial to trial. In Chapter 4, results found with the paradigm were used to critique the standard Bayesian model which, it appears, is not appropriate in the context of perceiving object movement during self-controlled head rotations. It remains an interesting challenge, however, to attempt to derive a standard Bayesian model that allows for the external variability that is introduced when stimuli are based on the self-movements of participants. This derivation is completed in this chapter, with the equations then being used to prove quantitatively that the standard Bayesian model produces psychometric equations that are a less good fit to the behavioural data from Chapter 4, than the psychometric functions that were derived in Chapter 2.

As outlined in the General Introduction, one method that can be used to convert sensory inputs to an optimally precise percept, is to combine the low-level sensory input, or likelihood, to a prior expectation of stimuli in the world. This combination is a Bayesian combination whereby the influence of the prior and the likelihood are mediated by their precisions. This is most clearly demonstrated when we represent the probabilities of the prior, likelihood, and posterior as frequency distributions (see Figure 5.1, a copy of Figure 1.4 and Figure 4.1). From Figure 5.1, we can see that the width of the frequency distribution of the likelihood is greater than that of the posterior, and also that changes to the widths of the prior and likelihood affect not just the precision of the posterior, but also its mean (Figure 5.1B). Given that the posterior is the basis for perceptual decisions, the change in mean implies a change in perceptual accuracy. This second point is key for the model that we will derive in this chapter. As discussed in the General Introduction, in Bayesian models, it is differences in

the precisions of the sensory inputs that explain the changes in bias that underlie perceptual errors, including self-movement compensation errors.

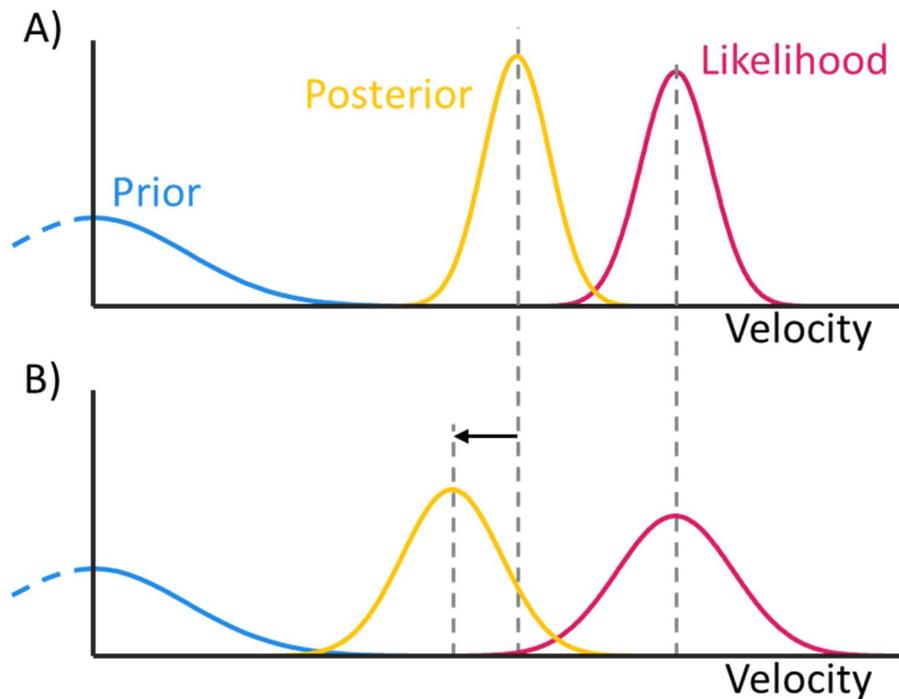


Figure 5.1: (A) a standard Bayesian model of motion perception. (B) the same Bayesian model but with more noise in the sensory evidence (wider likelihood distribution). This causes a shift in the posterior, towards the prior, as denoted by the arrow.

## 5.1 – Derivation

This derivation takes many of its equations and ideas from the work of Freeman et al. (2010). They provided a derivation of a Bayesian model that they used to explain the Aubert-Fleischl phenomenon (Aubert, 1887; Fleischl, 1882), an example of a self-movement compensation error. They fit the Bayesian model to their behavioural data which not only confirmed the existence of the Aubert-Fleischl phenomenon in the context of their stimuli, but also showed that participants' behavioural data were more precise when interpreting the movement of objects that were presented while their eyes were stationary, than objects that they pursued with their eyes. Freeman et al. used a standard Bayesian framework whereby differences in the precision of the sensory inputs explain changes in perceived speed (e.g., Ascher & Grzywacz, 2000; Moscatelli et al., 2019; Powell et al., 2016; Senna et al., 2015; Weiss & Adelson, 1998). These changes in perceived speed are thought to be caused by the existence

of a universal prior for object movement that is a slow-speed prior (Stocker & Simoncelli, 2006; Weiss & Adelson, 1998), due to all of our experience with slow or not moving objects (Weiss et al., 2002). This slow-speed prior biases any sensory inputs that are noisy, so that there is a greater shift towards slow-motion for these inputs. In the context of Bayesian models of movement perception: the less precise the stimulus, the slower it appears to move (Figure 5.1). In Freeman et al.'s Bayesian model, two Bayesian estimates of motion were used. These measurements were of the relative motion between the stimulus and a background, when the eyes were stationary, and the motion of the pursuit target, when the participants pursued it. In their experiment, with accurate eye movements in each condition, these relative and pursuit target motions were equivalent to the motion in the perceptual image and the self-movement of the eye respectively. These separate Bayesian measurements can explain self-movement compensation errors, as the less precise sensory information during self-movement would lead to an underestimation of pursuit speed, compared to the estimate of relative motion. Here, a similar Bayesian model will be derived but for our two-phase paradigm. This model will be tested later in the chapter by fitting it to the data from Chapter 4 and comparing the goodness of fit to the psychometric functions derived in Chapter 2.

Each phase of the paradigm used throughout this thesis is made up of multiple two-interval forced choice (2IFC) trials and each interval contains a participants' measure of either their own movement, or the movement of a stimulus. In Bayesian models it is assumed that each of these measurements must be a combination of sensory evidence denoting the speed of the stimulus and a prior expectation based on all previous perception. This Bayesian combination is simply a multiplication of two distributions, so it is possible to calculate the mean and variance of the result (S) from the mean and variance of the likelihood, (L) and prior (P). Adapting the equations from Ma et al. (2006), the mean and variance of S can be written as:

$$\mu_S = \frac{\mu_L \sigma_P^2 + \mu_P \sigma_L^2}{\sigma_L^2 + \sigma_P^2}$$

(5.1)

$$\sigma_S^2 = \frac{\sigma_L^2 \sigma_P^2}{\sigma_L^2 + \sigma_P^2} \quad (5.2)$$

As mentioned above, objects are typically at rest, so  $\mu_p = 0$  (Stocker & Simoncelli, 2006; Weiss & Adelson, 1998; Weiss et al., 2002). Hence Equation 5.1 reduces to:

$$\mu_S = \frac{\mu_L \sigma_P^2}{\sigma_L^2 + \sigma_P^2} \quad (5.3)$$

Equations 5.2 and 5.3 are relevant for one instance of a Bayesian measurement (e.g., measuring the motion in the non-image signal), however throughout our paradigm, participants are simultaneously expected to measure the motion in the image and non-image signal, and they are expected to do this across two separate intervals in each trial, and multiple trials per phase. Across these multiple trials, the sensory evidence will vary stochastically due to internal and external noise. The variance of the likelihood,  $\sigma_L^2$ , can be thought of as an internal and an external component,  $\sigma_{int}^2 + \sigma_{ext}^2$ . The internal component represents the intrinsic variability of the observer's measurement of the stimulus, including the noise in the sensory information and also the noise in the observer's interpretation of that information (e.g., noise in the neuron firing rates), whereas the external component represents the variability of the sensory evidence across trials. Even if the same stimulus is presented on every trial, so the sensory evidence is constant across trials, the likelihood varies stochastically across trials due to the observer's interpretation of the information, meaning that, instead of one single likelihood with a mean  $\mu_L$  and variance  $\sigma_{int}^2 + \sigma_{ext}^2$ , the likelihood across multiple trials becomes a Gaussian distribution of likelihoods, with mean  $\mu_L$  and variance  $\sigma_{int}^2 + \sigma_{ext}^2$ . Note that, in this case, because the stimuli are the same across trials,  $\sigma_{ext}^2$  remains constant. Now we can describe  $\mu_S$  with mean and variance as in equations 5.4 and 5.5 (Freeman et al., 2010; Stocker & Simoncelli, 2006). This mean and variance come from the fundamental principles of expected value and variance. When a distribution is multiplied by a constant, its expected value is also multiplied by the constant ( $E(rX) = rE(X)$ ), and its variance is multiplied by the square of the constant ( $Var(rX) = r^2Var(X)$ ). Using this, and the other fundamental principle that variance is unaffected by the addition of a constant

( $Var(X + a) = Var(X)$ ), we can calculate the expected value for the mean and the variance as:

$$\mu_S = \frac{\mu_L \sigma_P^2}{\sigma_{int}^2 + \sigma_{ext}^2 + \sigma_P^2} \quad (5.4)$$

$$\sigma_S^2 = \sigma_{int}^2 \left( \frac{\sigma_P^2}{\sigma_{int}^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 \quad (5.5)$$

We utilise a typical assumption alongside the method of constant stimuli, that the internal variability is constant across each of the levels of our stimulus.

In our paradigm, every interval is made up of two Bayesian measurements performed simultaneously, one measurement of the non-image signal (which describes the head movement), and one measurement of the image motion. In each interval, one of these measurements is dependent on the head movement that is generated for that trial. These head movements vary across trials, meaning that the stimulus is not the same on every trial. Instead, the stimuli that are presented in each trial form a Gaussian distribution, with mean  $\mu_V$  and variance  $\sigma_V^2$ . This causes an additional source of noise not accounted for by other psychophysical methodologies that scale stimuli based on head movements (Serafin et al., 2013; Steinicke et al., 2009). This source of noise was discussed in more detail in Chapter 2. Assuming that the observer has no bias,  $\mu_L = \mu_V$ , therefore:

$$\mu_S = \frac{\mu_V \sigma_P^2}{\sigma_{int}^2 + \sigma_{ext}^2 + \sigma_P^2} \quad (5.6)$$

As this Gaussian distribution is independent of the distribution of internal variability across time, the variances of these distributions will add over multiple presentations of that interval, leaving the mean of the posterior the same, but changing its variance. Again, this mean and variance come from the fundamental principles.

$$\sigma_S^2 = (\sigma_{int}'^2 + \sigma_V^2) \left( \frac{\sigma_P^2}{\sigma_{int}'^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 \quad (5.7)$$

As each interval of our paradigm is a simultaneous combination of a non-image signal and an image signal measurement, we need to combine two posteriors together to evaluate the interval as a whole. While one of the measurements in each interval is dependent on the head movement for that trial, the other measurement has  $\mu_L = 0$ . This is because the image is matched to the head rotation during the head-moving intervals, and the head is stationary during the head-stationary intervals. From Weber's law, we assume that, as the magnitude of a measurement approaches 0, its variance also approaches 0, so we discount the measurement with  $\mu_L = 0$  entirely when interpreting each interval of our paradigm. Note that it was found in Chapter 3 that the auditory image signal did not follow Weber's law. It is noted in Chapter 2, however, that there may be contributing components to the information in the head-moving interval that we are able to obtain a precision measurement of that are not non-image signals, including information due to the vestibulo-ocular reflex (Barnes, 1988) and inhibitory pursuit signals (Bedell et al., 1989). In the same way, variability of the image signal, despite its magnitude being zero, is another potential component of the combined measurement of precision that we are able to obtain that we term the "non-image" signal in this thesis. In this derivation, this variance is contained within the subsequent  $\sigma_h^2$  term.

For the first interval of Phase 1 (1.1), the stimulus speeds are equivalent to the head movement speeds, such that  $\mu_V = \mu_H$  and  $\sigma_V^2 = \sigma_H^2$ . As this is a head movement interval, the measurement that has  $\mu_L \neq 0$  is of a non-image signal, so  $\sigma_{int}'^2$  is replaced with  $\sigma_h^2$ .

$$\mu_{(1,1)} = \frac{\mu_H \sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \quad (5.8)$$

$$\sigma_{(1,1)}^2 = (\sigma_h^2 + \sigma_H^2) \left( \frac{\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 \quad (5.9)$$

For the second interval of Phase 1 (1.2), the stimulus speeds are equivalent to the head movement speeds scaled by a multiplicative gain factor,  $g$ , such that  $\mu_V = g\mu_H$ . As each individual stimulus was multiplied by the same gain factor, the variance of this set of factors is scaled by  $g^2$ , such that  $\sigma_V^2 = g^2\sigma_H^2$ . This is a head stationary interval, so  $\sigma_{int}'^2$  is replaced with  $\sigma_i^2$ .

$$\mu_{(1,2)} = \frac{g\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \quad (5.10)$$

$$\sigma_{(1,2)}^2 = (\sigma_i^2 + g^2\sigma_H^2) \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 \quad (5.11)$$

For the first interval of Phase 2 (2.1), the stimulus speeds are equivalent to the head movement speeds scaled by the bias, or Point of Subjective Equality, from Phase 1,  $b$ , multiplied by a small scaling factor,  $\alpha$ , to account for the bias related to the PSE, such that  $\mu_V = \alpha b\mu_H$  and  $\sigma_V^2 = b^2\sigma_H^2$ . This is another head stationary interval, so  $\sigma_{int}'^2$  is replaced with  $\sigma_i^2$ .

$$\mu_{(2,1)} = \frac{\alpha b\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \quad (5.12)$$

$$\sigma_{(2,1)}^2 = (\sigma_i^2 + b^2\sigma_H^2) \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 \quad (5.13)$$

For the second interval of Phase 2 (2.2), the stimulus speeds are again equivalent to the head movement speeds scaled by a gain factor,  $g$ , such that  $\mu_V = g\mu_H$  and  $\sigma_V^2 = g^2\sigma_H^2$ . This is another head stationary interval, so  $\sigma_{int}'^2$  is replaced with  $\sigma_i^2$ .

$$\mu_{(2,2)} = \frac{g\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \quad (5.14)$$

$$\sigma_{(2,2)}^2 = (\sigma_i^2 + g^2 \sigma_H^2) \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 \quad (5.15)$$

From standard signal detection theory (e.g., Jones, 2016), we assume that during our 2IFC task, participants utilise an internal decision variable,  $d$ , to determine which of the intervals moved more. This decision variable depends on the difference between the perceived motion in each interval and can therefore be thought of as a subtraction of the second interval from the first. We can calculate the mean and variance of the decision variable for each of the phases as follows. Due to the scaling of the head movements, certain components of the variance are perfectly correlated (as they have  $\sigma_H^2$  as a factor). For these components, the standard deviations of these components subtract ( $Var(X - Y) = (\sqrt{Var(X)} - \sqrt{Var(Y)})^2$ ). The other, independent components of the variance do add ( $Var(X - Y) = Var(X) + Var(Y)$ ) and the mean of the resultant distribution is equal to the difference between the means of the intervals ( $E(X - Y) = E(X) - E(Y)$ ). Using these principles, the mean and variance of the decision variable for Phase 1 can be written as:

$$\mu_d = \frac{g\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} - \frac{\mu_H\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \quad (5.16)$$

$$\begin{aligned} \sigma_d^2 = & \left( \sqrt{g^2\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} - \sqrt{\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} \right)^2 + \sigma_h^2 \left( \frac{\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 \\ & + \sigma_i^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 \end{aligned} \quad (5.17)$$

And for Phase 2:

$$\mu_d = \frac{g\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} - \frac{\alpha b\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \quad (5.18)$$

$$\sigma_d^2 = \left( \sqrt{g^2 \sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} - \sqrt{b^2 \sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} \right)^2 + 2\sigma_i^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 \quad (5.19)$$

In order to model the proportion of trials where a participant will perceive the second interval as faster than the first, we can look at the proportion of trials where  $d > 0$ . From standard signal detection theory, we state the equation for  $d'$ , which then allows us to calculate the proportion of trials where a participant perceives  $\mu_2$ , the mean of the second interval, as faster than  $\mu_1$ .

$$d' = \frac{\mu_d}{\sigma_d} \quad (5.20)$$

$$P(\mu_2 > \mu_1) = \Phi(d') \quad (5.21)$$

These equations give us models that we can fit to our psychometric data for each phase where the multiplicative gain value of the second interval is represented along the x axis, with the proportion of times that the second interval was selected as faster than the first, along the y axis. Using our data, we can fix the parameters  $\mu_H$  and  $\sigma_H^2$ , by fitting a normal distribution to the head speeds that were used by the participant. We can also fix the parameter  $b$ , by finding the PSE of the psychometric function of the first phase. This leaves us with the free parameters of  $\sigma_h^2$ ,  $\sigma_i^2$ ,  $\sigma_{ext}^2$  and  $\sigma_P^2$ . The full equations for Phases 1 and 2 can be found in the Appendix of this chapter, with versions that are most similar to the ones in this text and that are fully simplified. Note that Phase 3 is a replica of Phase 2 but with the value of  $b$  set to 1. Utilising our paradigm, we can measure the precision of the non-image signal ( $\sigma_h^2$ ), and by including the third phase mentioned in Chapter 4, we are able to measure the precision of the image motion signal ( $\sigma_i^2$ ). These values can also be fitted with this Bayesian model, so a comparison between the outputs of the paradigm and this model can be made. In each method, psychometric functions are fit to behavioural data, which enables us to compare the suitability of each model, by comparing goodness of fit measures. These comparisons will be made in the next section.

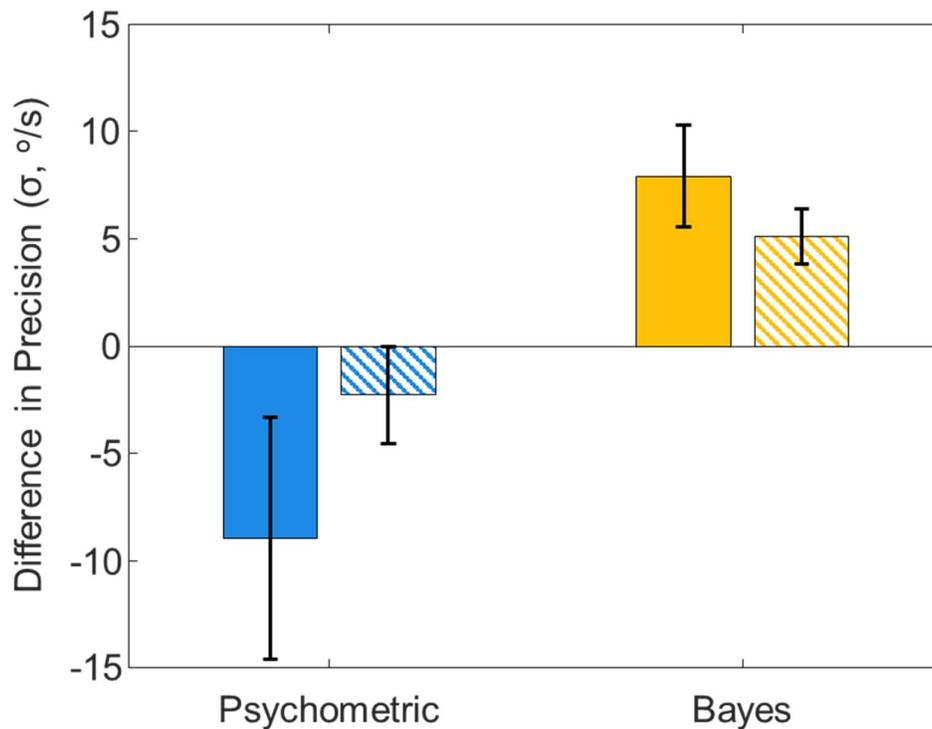
## 5.2 – Fitting Procedure

Like in the Appendix of Chapter 2, I will fit psychometric functions to our behavioural data based on the formulae in the Appendix of this chapter. The equation derived for Phase 2 (and 3) using the Bayesian model is equivalent to the equation used in Chapter 2 (once simplified, see Appendix), because the influence of the prior is the same in the two intervals and, so, cancels out (a formal demonstration of this is provided by Powell et al., 2016). The motion estimates are only based on image signals in Phases 2 and 3, meaning that these phases are standard speed discrimination experiments. As this is the case, the value of  $\sigma_i^2$  from the previous Phase 2 analysis was used, alongside the data for the head movements which defined  $\mu_H$  and  $\sigma_H^2$ . The remaining variables ( $\sigma_h^2, \sigma_P^2, \sigma_{ext}^2$ ) were free to vary in the fitting routine which now only features the equation for Phase 1. To avoid fitting local minima, each parameter was cycled through a search space of 20 start values and the best fit was chosen.

This yielded  $20^3$  (8000) separate cycles of the fitting routine. A MatLab function for this fitting procedure can be found in the Thesis Appendix (“BayesianPh1Fit”).

### Non-Image Signal Variance

Using this fitting procedure on the behavioural data from Chapter 4 allows direct comparison between the values for the precision of the non-image signal that are produced by the Bayesian model, and the psychometric model developed in Chapter 2. In Figure 5.2, the difference between the non-image and image signal precisions (non-image signal precision measurement minus image signal precision measurement) is plotted for both the psychometric model that was developed in Chapter 2 (blue bars) and the Bayesian model that was developed in this chapter (yellow bars). This analysis was completed for the auditory (solid bars) and visual (striped bars) versions. An ANOVA was conducted on this difference data which showed no significant interaction between modality and model ( $F(1,20) = 2.049$ ,  $p = 0.168$ , NS) and no significant main effect of modality ( $p = 0.564$ , NS), but did show a significant main effect of model ( $p = 0.02$ ).



*Figure 5.2: the difference between the precision of the non-image and image signal (positive value means that the non-image signal is less precise than the image signal) for the auditory (solid bars) and visual (striped bars) versions of the experiment presented in Chapter 4. Presented are the two analyses, the psychometric model that was proposed in Chapter 2 (blue) and the Bayesian model that was proposed in this chapter (yellow). Error bars represent  $\pm 1SE$ .*

It is clearly the case that the Bayesian analysis produces signal precisions such that the non-image signal is less precise than the image signal in both modalities (positive difference in Figure 5.2), while the psychometric model produces differences in the other direction. It would be impossible for the Bayesian model to generate the appropriate bias that has been seen in the previous chapters (such that object movement appears to be slower during self-movement) without the difference that is shown here. As discussed previously, if the non-image signal was less precise than the image signal it would be influenced more by the slow-speed prior and be perceived as slower, explaining the bias. While it was found with the psychometric model that this was not the case in previous chapters, providing evidence against the Bayesian model, here the Bayesian model returns values that appear to show that the non-image signal is less precise than the image signal.

### Goodness Of Fit

Figure 5.3 shows the mean and standard error of the Deviance of the fits. Deviance is the recommended goodness-of-fit measure for psychometric functions (Wichmann & Hill, 2001). It is clear that the deviance of the Bayesian model is greater than the deviance of the psychometric model. This is a statistically significant difference ( $t(67) = -3.754, p < 0.001$ ) and it is likely that this additional deviance is due to the non-image signal being less precise than the image signal in the Bayesian fits. The additional noise that is added to the non-image signal in the Bayes model allows the model to fit the right bias for the dataset, however it also has the effect of adding additional noise to the psychometric function, causing a shallower function that is a less good fit to the behavioural data (e.g., Figure 5.4). As the fits to the behavioural data are significantly less good for the Bayesian model, it is unlikely that the value for the precision of the non-image signal that the model returns is a good indicator of the true precision of the non-image signal.

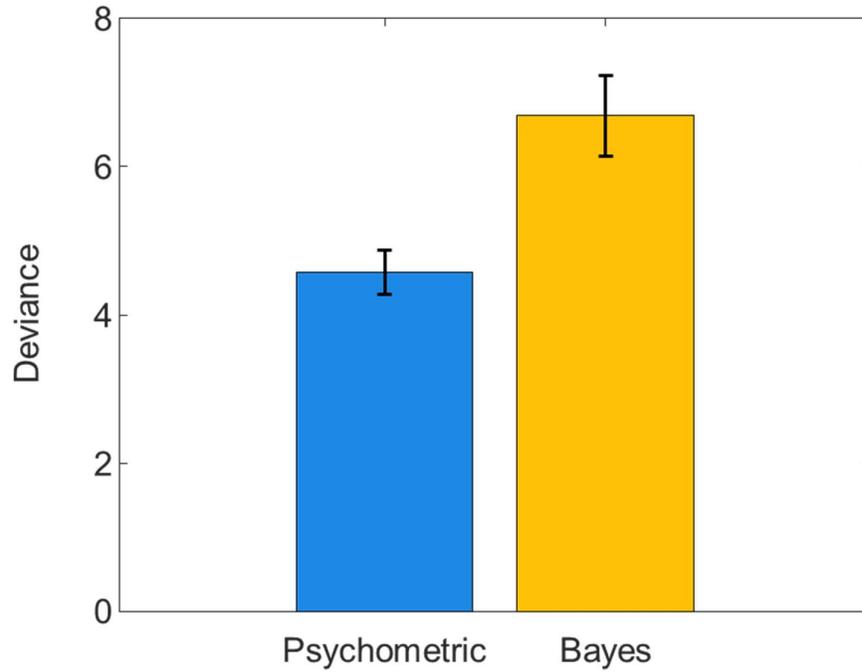


Figure 5.3: the mean deviance of the psychometric functions fitted to the behavioural data for the psychometric (blue) and Bayes (yellow) models. Greater deviance means a worse fit. Error bars represent  $\pm 1SE$ .

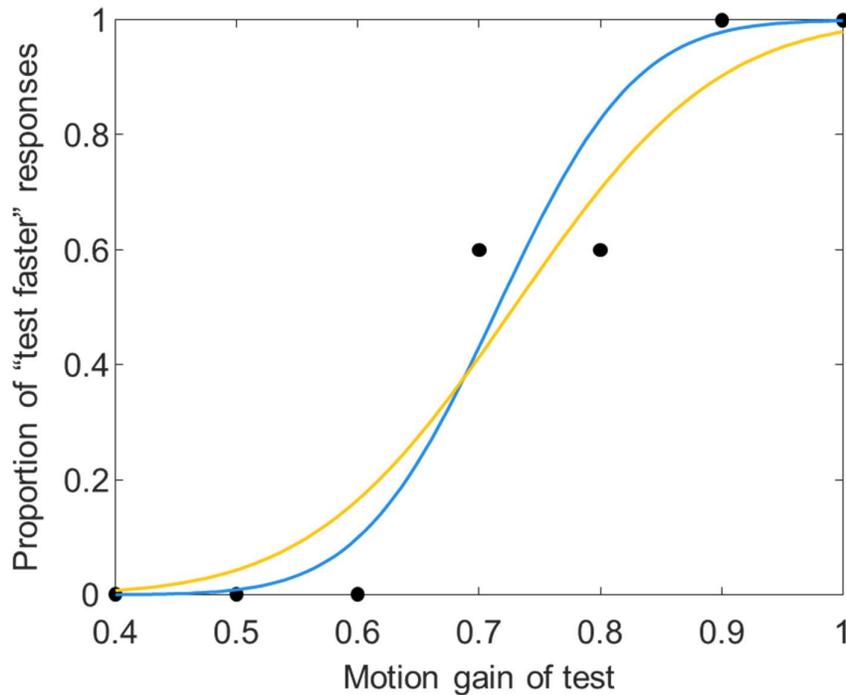


Figure 5.4: an example of behavioural data (circles) and the fits of the psychometric (blue) and Bayes (yellow) models. The Bayes model fit is shallower and has more deviance (Psychometric model fit deviance = 3.75; Bayes model fit deviance = 4.52).

### 5.3 – Summary

This chapter has shown that it is possible to derive a standard Bayesian model that accounts for the external variance that is introduced when stimuli are based on participant movements that are variable across trials. It was also shown here, though, that this Bayesian model is not as good at fitting the behavioural data in this thesis as the psychometric model that was derived in Chapter 2. This is likely to be due to the requirement of the standard Bayesian model for the non-image signal to be less precise than the image signal if self-movement compensation errors are to be explained, and the fact that this is not likely to be the case in reality, as shown in Chapters 3 and 4.

The next chapter will contain an investigation into a different form of compensation, namely speed constancy as a function of the distance between an observer and an object. This has been investigated many times in vision but, to the author's knowledge, has not yet been investigated in the auditory modality. Speed constancy is an important form of compensation for the experiments already presented in this thesis, though it has not been the focus, as all of the stimuli have been presented at the same physical distance from the observer in each experiment.

### 5.4 – Appendix

Most similar to equations 5.16-5.19 in the main text:

Phase 1:

$$\mu_d = \frac{g\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} - \frac{\mu_H\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2}$$

$$\sigma_d^2 = \left( \sqrt{g^2\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} - \sqrt{\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} \right)^2 + \sigma_h^2 \left( \frac{\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2$$

$$+ \sigma_i^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2$$

$$P(\mu_2 > \mu_1) = \Phi \left( \frac{\frac{g\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} - \frac{\mu_H\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2}}{\sqrt{\left( \sqrt{g^2\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} - \sqrt{\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} \right)^2 + \sigma_h^2 \left( \frac{\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 + \sigma_i^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2}} \right)$$

Phase 2:

$$\mu_d = \frac{g\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} - \frac{\alpha b\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2}$$

$$\sigma_d^2 = \left( \sqrt{g^2\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} - \sqrt{b^2\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} \right)^2 + 2\sigma_i^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2$$

$$P(\mu_2 > \mu_1) = \Phi \left( \frac{\frac{g\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} - \frac{\alpha b\mu_H\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2}}{\sqrt{\left( \sqrt{g^2\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} - \sqrt{b^2\sigma_H^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2} \right)^2 + 2\sigma_i^2 \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2}} \right)$$

Simplified versions:

Phase 1:

$$\mu_d = \mu_H\sigma_P^2 \left( \frac{g}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} - \frac{1}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)$$

$$\sigma_d^2 = (\sigma_H^2 g^2 + \sigma_i^2) \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 + (\sigma_H^2 + \sigma_h^2) \left( \frac{\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 - \frac{2\sigma_H^2 g \sigma_P^4}{(\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2)(\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2)}$$

$$P(\mu_2 > \mu_1) = \Phi \left( \frac{\mu_H\sigma_P^2 \left( \frac{g}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} - \frac{1}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)}{\sqrt{(\sigma_H^2 g^2 + \sigma_i^2) \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 + (\sigma_H^2 + \sigma_h^2) \left( \frac{\sigma_P^2}{\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2 - \frac{2\sigma_H^2 g \sigma_P^4}{(\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2)(\sigma_h^2 + \sigma_{ext}^2 + \sigma_P^2)}}} \right)$$

Phase 2:

$$\mu_d = (g - \alpha b)\mu_H \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2}$$

$$\sigma_d^2 = ((g - b)^2\sigma_H^2 + 2\sigma_i^2) \left( \frac{\sigma_P^2}{\sigma_i^2 + \sigma_{ext}^2 + \sigma_P^2} \right)^2$$

$$P(\mu_2 > \mu_1) = \Phi\left(\frac{(g - \alpha b)\mu_H}{\sqrt{(g - b)^2\sigma_H^2 + 2\sigma_t^2}}\right)$$

# Investigating Auditory Speed Constancy

## 6.0 – Preface

The experiments presented in this chapter were conducted as part of an Experimental Psychology Society (EPS) Study Visit Grant. I attended Ulm University under the supervision of Prof. Marc Ernst for a 12-week placement. I would like to extend my thanks to the incredibly supportive team that I found there.

This chapter contains a departure from the investigations in Chapters 2-5. Whereas those chapters investigated the compensation for self-movement that we are able to perform when interpreting object movement, this chapter will instead focus on the compensation for distance that we are able to perform when interpreting object movement. The experiments presented in this chapter were conducted with stationary participants, so no compensation for self-movement was needed. This investigation was completed with auditory stimuli only so it focuses on distance perception in audition and provides a first look at compensation for distance during auditory object movement perception.

## 6.1- Introduction

As highlighted throughout this thesis, movement perception can be completed through the summation of an image signal and a non-image signal that represent the motion in the image and our self-movement. This could potentially mean that interpreting image motion in the absence of self-movement is made easier because one of these components is not present. However, the image motion is in an angular form because the motion is relative to the receptor. If an observer relied on image motion alone, they would find it difficult to differentiate between a fast moving object that is further away and a slower moving object that is nearer. In order to determine the 3D movement of an object in space, then, the observer needs a measurement of the distance between themselves and the object. In keeping with the overarching theme of this thesis, the use of this distance information to interpret the image motion is a form of compensation, as we compensate for the distance between us and objects in the world. In the vision literature, this compensation has been referred to as speed constancy (e.g., Brown, 1931; Epstein, 1973; Rock et al., 1968; Wallach,

1939) or velocity constancy (e.g., Distler et al., 2000; McKee & Welch, 1989; Zohary & Sittig, 1993).

### Visual Depth Cues

Measuring the distance between ourselves and objects is fundamental to a number of perceptual judgements. Alongside interpreting the movement of objects, we use these distance measurements when estimating the size, and location, of objects. For the visual system, this can be achieved using many different depth cues, which fall into three main categories: monocular, binocular, and active cues. Monocular cues are present in single two-dimensional images, and include occlusion, height in the visual field, image size, texture gradients and image blur. Combining two 2-D images that are highly overlapping creates our binocular vision. The main binocular cue to depth is binocular disparity which is where the images from each eye differ slightly. Due to this, objects that are nearer to us have a greater visual angle than objects that are further away, meaning that the images for each eye become more disparate for nearer objects and are more similar for further objects. Binocular disparity is often thought of as the main cue to depth, however the fact that depth perception is still possible with only one eye is evidence that it is certainly not the only depth cue. These monocular and binocular depth cues are passively available, whereas others involve active movement. While these movements can be as small as changing the tension of the muscles controlling the lens of the eye, they also include movements of the eye or head. Accommodation is the ability to control the tension of the muscles that control the lens in order to focus an image onto the retina (Watt et al., 2005). The amount of accommodation that is necessary depends on the distance to an object, hence this active tensing of muscles can be used as a depth cue. As noted in the General Introduction, self-movement causes unique reafferent motion profiles in the images that our eyes obtain. In Figure 6.1, a copy of Figure 1.3, it is clear that translational self-movement causes a different motion profile than rotational self-movement. During translational self-movement, stationary objects that are at different distances from an observer will generate different motion in the resulting image (Figure 6.1A). Differences in the motion of objects in the image during translational self-movement can be used as a cue to the distance to those objects. This cue is known as motion parallax.

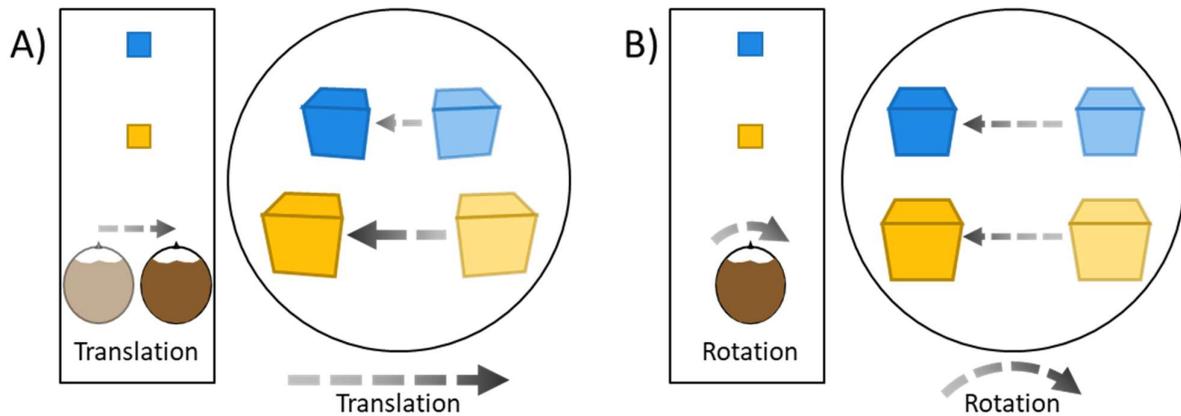


Figure 6.1: (A) the motion in the image is inversely proportional to the distance from the observer during horizontal translation of the observer. (B) the motion in the image is the same irrespective of the distance from the observer during rotation of the observer

While it is important to note the many different depth cues, distance measurements will be thought of as one composite signal that combines all of the available depth cues in this chapter. This is analogous to the way that the non-image signal was used to refer to a combination of all of the cues that we use to determine our self-movement, excluding image cues, throughout the rest of this thesis.

### Visual Speed Constancy

As mentioned previously, speed constancy is the ability to compensate for the distance between ourselves and an object when we interpret the motion of that object in 3-D space. Speed constancy is almost complete in the visual system (Brown, 1931; Epstein, 1973; Rock et al., 1968), however, the process used to achieve speed constancy has been debated. Throughout this thesis, I have investigated non-image and image signals as it has been the case that a summation of these signals is the only appropriate method for self-movement compensation in the experiments that I have presented so far. In a similar way, it may be the case that a measurement of image motion and a measurement of object distance are both used in an equation to return perceived object movement. Because of the geometry of this form of compensation, the relevant equation is presented in Equation 6.1. It may not be the case, though, that this is the only appropriate method for speed constancy in the experiments presented here.

$$\text{Object Movement} = \tan(\text{Image Motion}) * \text{Object Distance}$$

(6.1)

In a commentary, Wallach (1939) proposed that the speed constancy described earlier by Brown (1931) was similar to another effect that Brown also described: the transposition effect. The transposition principle states that when two moving systems (containing multiple objects) of different physical sizes are presented to a participant at the same distance, the larger system appears to move slower than the smaller one. Brown's experiments presenting this principle involved participants matching the speed of a test system to a standard system, and found that the larger test system moved at twice the speed of the smaller standard system at the Point of Subjective Equality (PSE). Wallach pointed out that there is no difference between the retinal size of a system that is larger and also further from the observer than a smaller, closer system. He claimed that it follows that, no matter whether it is the physical size or the distance to the systems that differs, the perceived speed would differ, because the retinal size is different. Wallach's claim was that Brown's transposition principle is not confined to systems that change in physical size, rather, the difference in perceived speed occurs due to differences in retinal size.

If differences in retinal size, instead of distance to the object, are used to determine the speed of an object, it may also be the case that retinal size is used to determine other properties of an object that are otherwise conflated by distance. As pointed out by Wallach, the retinal size of a larger object that is farther from the observer, and a smaller object that is closer to the observer, are equivalent. Here, the physical size of the objects differ, but this is difficult to know without measurement of the distance to the objects. Interpreting the sizes of these objects correctly is an analogous form of compensation to speed constancy, named size constancy, and it has been shown by Rock et al. (1968) that size and speed constancy both depart from reality by the same amount and are predictable by one another. This appears to link both speed and size constancy through retinal size changes.

Other researchers have argued that size and speed constancy may not be as linked as Rock et al. proposed. McKee and Welch (1989) showed that speed constancy may not be as complete as size constancy. The stimuli in their experiment were lines presented on two CRT monitors, with one image presented to each eye. This allowed participants to use binocular

disparity as a depth cue, and allowed the experimenters to manipulate this cue. Participants could use this cue to estimate distance reasonably well, despite there being no differences in retinal size across their stimuli. This suggests that retinal size is not used alone to perform both speed and size constancy.

McKee and Welch (1989) went on to suggest that the transposition effect may have a more direct link to speed constancy. This idea goes back to the transposition principle and states that it is key that the difference between the sizes of the systems is quantifiable. It was suggested that the size of each system with respect to an object of known (or assumed) size was the reason for the differences in their perceived speeds. In this case, large objects relative to the object of known size appear slower than smaller objects on the same scale. This explanation was supported by a further experiment where the participants' physical distance to the stimulus was altered, and the accuracy of speed constancy improved. Similar findings were reported by Zohary and Sittig (1993) who take terminology from Epstein (1978) and discuss two algorithms, distance calibration and relative displacement. They state that distance calibration (the compensation procedure outlined in Equation 6.1) can lead to velocity constancy if distance measurements are accurate, but that distance measurements may be unnecessary, as the relative sizes of objects are enough for velocity constancy, through the transposition effect. In naturally occurring situations, there should be no difference between the outcomes of the two processes.

Another explanation for the difference in speed constancy in the studies by McKee and Welch (1989) and Zohary and Sittig (1993), and previous investigations into speed constancy, is the number and quality of the available depth cues. McKee and Welch admit that the only cue to distance in their experiments was binocular disparity, while Zohary and Sittig's only quantification of the depth perception of their participants shows that the perceived distances were not close to veridical. In experiments by Distler et al. (2000), a VR (Virtual Reality) environment was used to measure speed constancy with different numbers of depth cues available. The cues included or excluded were: perspective, size, texture, viewing height, disparity, and motion parallax. It was found that speed constancy was apparent when all of these cues were present but became less apparent for situations where fewer cues were available. It could be the case then, that some of these depth cues, despite giving a sufficient sensation of distance, may not be enough to support speed constancy,

where additional cues are needed. In naturally occurring scenarios, these depth cues are all available and speed constancy is almost complete.

### Auditory depth cues

In the auditory system, information about the distance to a sound source is limited to the intensity, reverberation and spectral content of sounds, along with an auditory motion parallax cue, during self-movement. Other localisation cues such as the interaural time delay (ITD) and interaural level difference (ILD) have been identified as possible distance cues (Coleman, 1963; Hirsch, 1968), however, there is some debate as to whether these cues can be used, as they only vary by very small amounts for sound sources at different distances but the same angular localisation. Greene (1968) mathematically simulated almost optimal listening conditions and still found that depth measurements relying on these cues alone would be likely to be incorrect by a factor of nearly half. These theoretical limitations were confirmed by Molino (1973) who disputed the arithmetic of Greene, but also found that their own calculations did not support participants' use of the ITD and ILD cues for depth perception. They presented stationary sounds at different distances from participants in a straight line, calibrated with inverse-square law so that their intensities matched. Their participants were unable to reliably determine the distance to each sound, suggesting that the ITD and ILD cues are not sufficient to determine the distance between a listener and a sound source.

A much studied auditory cue to depth is intensity, which decreases as the distance between a listener and a sound source increases (e.g., Ashmead et al., 1990; Blauert, 1997; Gamble, 1909; Steinhauser, 1879; Stevens & Guirao, 1962; Thompson, 1882; von Békésy, 1949; Zahorik et al., 2005). The inverse-square law is often cited to describe the decrease in intensity as distance increases (Equation 6.2, where  $I$  is the intensity of the stimulus and  $R$  is the distance between the observer and the sound source). This is because the power of a stimulus follows inverse square law, meaning that if the stimulus intensity is measured in decibels (dB; a unit representing an intensity ratio), the inverse square law holds. Note that if

stimulus intensity is measured as a pressure change, an inverse first power loss is a better descriptor (Coleman, 1963).

$$I = \frac{1}{R^2}$$

(6.2)

In reverberant areas, intensity remains a depth cue, but is enhanced by an additional cue based on reverberations. The more the primary, or direct, sound dominates reverberation in terms of its intensity, the nearer the sound source is perceived (e.g., Bronkhorst & Houtgast, 1999; Eargle, 1960; Maxfield, 1931; Steinberg & Snow, 1934). The direct-to-reverberant energy ratio provides absolute depth information (Mershon & Bowers, 1979; Mershon & King, 1975; Zahorik et al., 2005), in contrast to stimulus intensity alone, which provides relative cues to distance (e.g., which sound source is closer) without generating absolute distance measurements (Mershon & King, 1975). This is not to say that the direct-to-reverberant energy cue is a direct replacement for the stimulus intensity cue, instead it appears that the sensitivity of the direct-to-reverberant energy ratio is quite low, with just-noticeable differences that correspond to a more than doubling of the auditory distance (Zahorik, 2002; Zahorik et al., 2005).

The spectral content of an auditory stimulus can have one of two effects. For stimuli beyond around 15m in distance, the effects of frequency-dependent attenuation impact the stimulus, attenuating the lower frequency components of the sound by more than the higher frequency components. This leads to lower frequency stimuli being perceived as nearer to the observer. For nearby stimuli, within around 3m, the spectral content of the stimulus can again be used as a cue, however here the change in spectral content may be linked somewhat to the intensity of the stimulus (Blauert, 1997).

We will not focus much on the auditory version of motion parallax here as participants remained stationary on a chinrest throughout the experiments presented in this chapter, however all other auditory depth cues were intentionally present, as the laboratory was not sound treated. As noted before, the distance measurements in this chapter will be thought of as one composite signal including all of the depth cues available.

## Does Hearing Exhibit Speed Constancy?

While speed constancy has been investigated extensively in the visual system, to the knowledge of the author, no investigation has yet been made into speed constancy in the auditory system. The findings of Distler et al. (2000) indicate that environments rich in depth cues are more likely to support speed constancy than those with reduced cues. For this reason, the experiments in this chapter were carried out in a typical indoor environment that was not sound treated. A moveable speaker played a broad band sound which provided good binaural and monaural cues, while the use of room that was not sound treated enhanced reverberation cues.

In Experiment 1, it needed to be verified that participants were able to make use of these cues to judge the distance to the sound source. A cross-modal matching task was used, where participants first listened to a static sound source whilst wearing a blindfold. They then removed the blindfold and moved the visible, but now silent, speaker to the location they thought the source had been played from. In Experiment 2, speed constancy was tested by having blindfolded participants compare the perceived speed of two sound sources moving at a constant speed along a frontoparallel plane. They were presented sequentially at two different depths. Perfect speed constancy would mean that the two stimuli should appear to move at the same speed when the linear speed in 3D space matched, as opposed to the angular speed relative to the participant. If speed constancy failed and the depth cues were ignored, then the perceived speed match would occur when the angular speed was the same, not the linear speed.

## 6.2 – Experiment 1 – Investigating Auditory Depth Perception

### 6.2.1 – Methods

#### Stimuli & Materials

The stimuli in this experiment were bursts of white noise that were presented via a speaker attached to a cable robot that moved with six degrees of freedom (see Figure 6.2). The stimuli were stationary and presented at different distances from the participants, who were instructed to place their chin on a chin rest to reduce movement. The robot could move the speaker 1m in any direction from a central point while participants were positioned with

the centre of their head 2.21m from that central speaker position. The laboratory was not sound treated which allowed reflections and reverberations to be used as depth cues, as they would be in a natural setting. The cable robot was made up of eight motors that controlled eight cables that attached to a speaker housing unit. The cable runs were designed such that the motors could be positioned in a separate room to the speaker, in order to minimise the impact of the sound of the motors. A lightweight aluminium frame was bolted to the floor to house the cable robot and reduce any extraneous movement (see Figure 6.2).



*Figure 6.2: Laboratory set-up.*

## Procedure

During a familiarisation phase, participants were instructed to pay attention to the distance between themselves and the stimulus that was about to be presented, this will subsequently be referred to as the stimulus distance. The speaker was repositioned to the nearest and furthest stimulus distances that were to be used in the experimental phase, and the auditory stimulus, a burst of white noise, was played. This ensured that the stimuli in the subsequent experiments were not novel to the participants, as it has been shown by Coleman (1962) that depth perception is poor for novel sounds, and improves when sounds have previously been presented to participants. During pilot testing for this experiment,

participants commented that they were unaware of the minimum and maximum distances that the speaker could move to, and this was represented in the results, where participants showed a strong tendency towards the centre of the movement area of the speaker. Providing the minimum and maximum distance information in the familiarisation phase served to provide a scale for participants to determine the distance to the sound source. This links to the work of McKee and Welch (1989) and Zohary and Sittig (1993) who claimed that the transposition effect also required a scale.

Participants were then blindfolded before the experimental phase began. In each trial in the experimental phase, the speaker moved from a starting position to a stimulus position, where a short sound burst was played, and then to an ending position. The participants then removed their blindfold and repositioned the speaker, by pressing the left and right buttons on a mouse to move the speaker towards or away from them in steps of 0.05m, to where they perceived the sound burst to have been played. They then locked in their answer by clicking in the scroll wheel of the mouse and refitted their blindfold before the speaker was moved to its starting position for the next trial. The stimuli were all presented along the intercept of the sagittal and transverse planes of the head and the distances to the stimuli were 1.51, 1.71, 1.91, 2.11, 2.31, 2.51, 2.71, and 2.91m. The starting and ending positions were randomised to ensure that participants could not make a comparative judgement between the stimulus distance and the starting or ending position. These positions were always within 0.2m of the central position, which was 2.21m from the participant, and along the same axis as the stimuli (see Figure 6.3A). Each of the 8 depths was presented to the participant 10 times, making 80 trials that were presented in a random order.

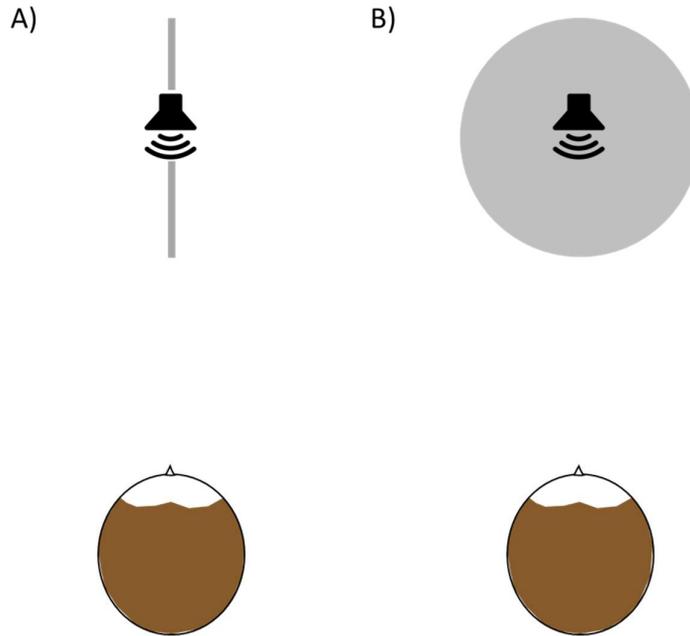


Figure 6.3: the starting and ending positions of the speaker were within 0.2m of the central position. (A) In experiment 1, along the intercept of the sagittal and transverse planes of the head, (B) and in experiment 2, in any direction along the transverse plane.

## Participants

The nine participants were students or staff at Ulm University (4 female, 5 male). Seven participants were naïve to the purposes of the experiment and two were experimenters. All participants self-reported normal hearing and vision.

## 6.2.2 – Results

For each participant, a speaker position setting, corresponding to a perceived stimulus distance, was calculated for each repetition of the experiment. These were then averaged across participants as shown in Figure 6.4. There is clear effect of stimulus distance ( $F(7,64) = 27.05$ ,  $p < 0.001$ ), with multiple comparison tests revealing 19 of the 28 pairwise comparisons as significant (Table 6.1). Highlighted in grey in Table 6.1 is the comparison between the two distances that were selected as the stimulus distances in the next experiment.

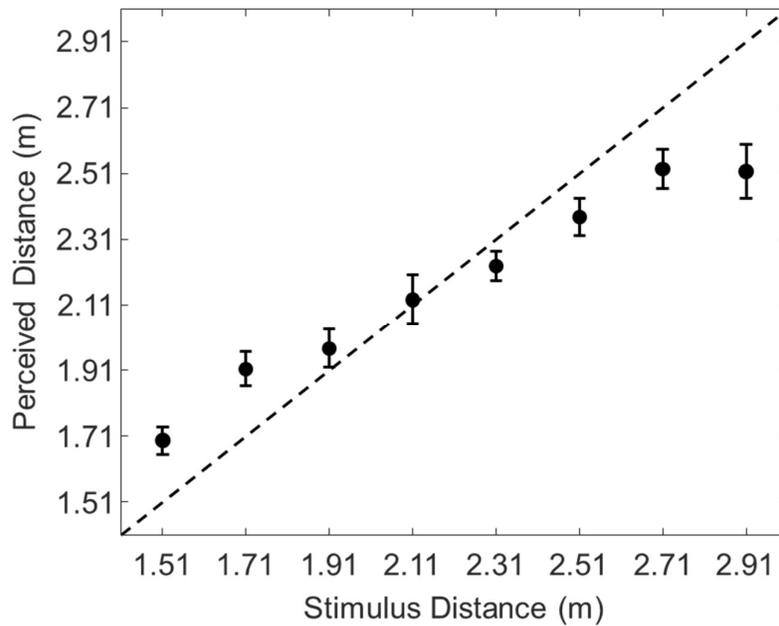


Figure 6.4: mean perceived stimulus distance as a function of actual stimulus distance. Error bars represent  $\pm 1SE$ .

Stimulus Distance	1.51	1.71	1.91	2.11	2.31	2.51	2.71
1.71	0.172						
1.91	0.037*	0.998					
2.11	<0.001**	0.213	0.582				
2.31	<0.001**	0.006**	0.040*	0.872			
2.51	<0.001**	<0.001**	<0.001**	0.041*	0.589		
2.71	<0.001**	<0.001**	<0.001**	<0.001**	0.013*	0.646	
2.91	<0.001**	<0.001**	<0.001**	<0.001**	0.011*	0.615	1.000

Table 6.1 shows the significance of the Tukey post-hoc tests performed after the ANOVA on stimulus distance (\* denotes significance at the  $< 0.05$  level and \*\* denotes significance at the  $< 0.01$  level). Highlighted in grey is the stimulus distances that were selected for Experiment 2, which have a significant difference in perceived distance at the  $< 0.01$  level.

The data also show a slight tendency towards the centre of the distances, with underestimation of distance at greater stimulus distance and overestimation of distance at shorter stimulus distance. Biases like these have been reported before, for example Gogel (1969) showed that participants tended towards a perceived stimulus distance of 2m. Gogel

called this the specific distance tendency, however, the stimuli in their study were visual. Similar findings have been obtained with auditory stimuli (Anderson & Zahorik, 2014; Mershon & King, 1975; Zahorik et al., 2005), where the specific distance tendency has been shown to depend on the reverberation characteristics of the room, however a specific distance tendency of 1.9m was found on average by Zahorik et al. (2005). This aligns with our data, which suggest a specific distance tendency somewhere in the region of 2-2.1m (perceived distance matches stimulus distance; Figure 6.4).

The results of Experiment 1 show that the stimulus set-up produced auditory cues that were sufficient to support reasonably good depth perception. This allowed us to investigate whether auditory movement perception exhibits speed constancy in Experiment 2. This was investigated by placing moving sound sources at different distances and having observers judge their speed. Perfect speed constancy predicts that the perceived speed of the sound sources should be the same when movement is expressed as a linear speed in 3D space as opposed to an angular speed subtended at the listener.

## 6.3 – Experiment 2 – Investigating Auditory Speed Constancy

### 6.3.1 – Methods

#### Stimuli & Materials

Experiment 2 was completed with the same materials as Experiment 1, however the participants remained blindfolded throughout the entire experimental phase. Also differently from Experiment 1, the stimuli in Experiment 2 were not stationary. Instead they were presented whilst the speaker moved in straight lines parallel to the transverse plane of the participants, as shown in Figure 6.5. The stimuli were presented at two distances, a near and a far distance (1.71 and 2.71m from the participant). Previously, in Experiment 1, it was clear that there was a statistically significant difference between participants' depth perception at these two stimulus distances (grey highlighted box in Table 6.1). The stimulus directions (left or right) were randomised.

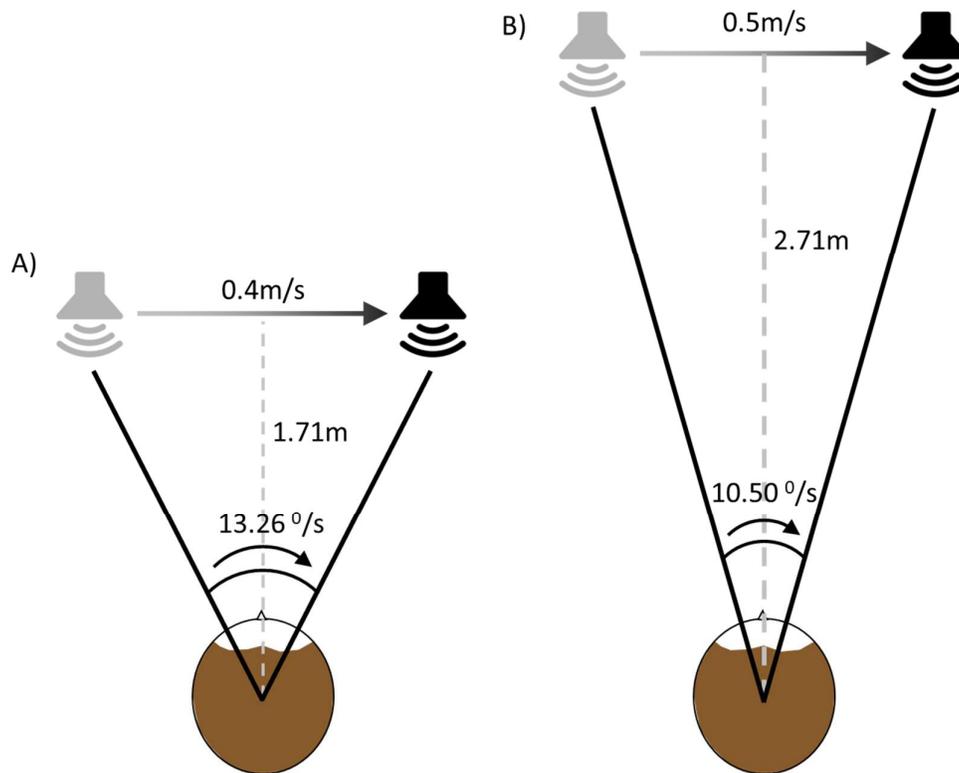


Figure 6.5: schematic of the movement of the standard stimuli in Experiment 2. (A) in the standard near condition, the standard stimulus was presented along a straight line 1.71m from the participant and parallel to the transverse plane of the participant. This standard stimulus moved with a velocity of 0.4m/s, leading to a mean angular velocity (the distance of the motion path was randomized) of 13.26°/s. (B) in the standard far condition, the standard stimulus was presented along a straight line 2.71m from the participant and parallel to the transverse plane of the participant. This standard stimulus moved with a velocity of 0.5m/s, leading to a mean angular velocity of 10.50°/s.

## Procedure

As in Experiment 1, Experiment 2 also began with a familiarisation phase. In this phase, the fastest (0.75m/s) and slowest (0.15m/s) moving stimuli were presented at both the near (1.71m) and far (2.71m) distances, along with the corresponding standard stimuli (explained later in this section), so that the participants had a scale for the movement speeds of the stimuli in the subsequent experimental phase.

Participants were then blindfolded before the experimental phase began. Like in Experiment 1, starting and ending positions were determined for the speaker that were within 0.2m of the central position, however in this experiment this random variation was in any direction on the transverse plane (see Figure 6.3B). As Experiment 2 employed a two-interval forced choice procedure, an intermediary position was selected for the speaker to move to in

between the two stimulus presentations. For each stimulus presentation, a speed was selected for the stimulus (either the standard speed or one of the test speeds) and a random duration was also selected. This speed and duration were used to calculate a distance and therefore stimulus onset and offset positions were generated. If it was possible to use these positions, they were selected, else a new random duration was selected.

Each trial started when the speaker moved to its start location and then to the first stimulus onset position. The stimulus was presented to the participant during the movement of the speaker to the first stimulus offset position, then the speaker moved to its intermediary position. Next, the speaker moved to the second stimulus onset position and the stimulus was presented during the movement of the speaker to the second stimulus offset position. Finally, the speaker moved to its ending position and the participant made a judgement about whether the first or second stimulus moved faster. The Point of Subjective Equality (PSE), where the test and standard stimulus are perceived to move at the same speed is the statistic of interest in this experiment, as perfect speed constancy would be achieved if the PSE occurs when the standard and test have the same linear speed, whereas no speed constancy would be achieved if the PSE occurs when the standard and test have the same angular speed.

Crucially, the two stimuli were presented one at either of the near (1.71m) and far (2.71m) distances from the participant, meaning that stimuli with the same linear speed had different angular speeds and vice versa. As the standard stimulus was presented at either of these distances, two different speeds of standard stimulus were used. This ensured that the linear and angular speeds of the standard stimuli fit within the range of the linear and angular speeds of the test stimuli. When the standard was presented at the near distance, it moved at 0.4m/s (angular velocity of  $13.26^{\circ}/s$ , see Figure 6.5A), and when it was presented at the far distance, it moved at 0.5m/s (angular velocity of  $10.50^{\circ}/s$ , see Figure 6.5B). The test stimulus always moved at one of 9 velocities (0.15, 0.225, 0.3, 0.375, 0.45, 0.525, 0.6, 0.675, and 0.75m/s) no matter whether it was presented at the near or far distance (the test and standard were always presented at different distances). The order of presentation of the test and standard stimuli was randomised. Each of the nine test stimulus speeds was presented 10 times for each of the two stimulus depths, making 180 trials that were presented in a random order.

The randomisation of the starting, intermediary, and ending positions, along with the randomisation of duration and, subsequently, the distance travelled by the stimulus, ensured

that none of duration, distance, or start and end locations could be measured as a proxy for speed. It is to be expected that in this scenario, participants will use the velocity cue to determine the motion of the stimulus, however the precision of the motion judgement may be decreased (Freeman et al., 2014).

### Participants

Seven of the participants that completed Experiment 1 also completed Experiment 2 (4 female, 3 male). One author was a participant in this experiment. All participants self-reported normal hearing and vision.

### 6.3.2 – Results

To assess speed constancy, psychometric functions were fit to the data using the Palamedes tool box (Prins & Kingdom, 2018) and the PSE was obtained as a measure of the speed of the test stimulus that the participants perceived to be equivalent to the speed of the standard. Both linear velocity and angular velocity could be used to assess speed constancy, here angular velocity was selected for analysis. In the case of angular velocity, a lack of speed constancy would be evident if the PSE occurred at the point when the test stimulus had the same angular velocity as the standard stimulus. This corresponds to the blue bars in Figure 6.6A. The angular velocity at the PSE that would represent perfect speed constancy was calculated and corresponds to the yellow bars in Figure 6.6A.

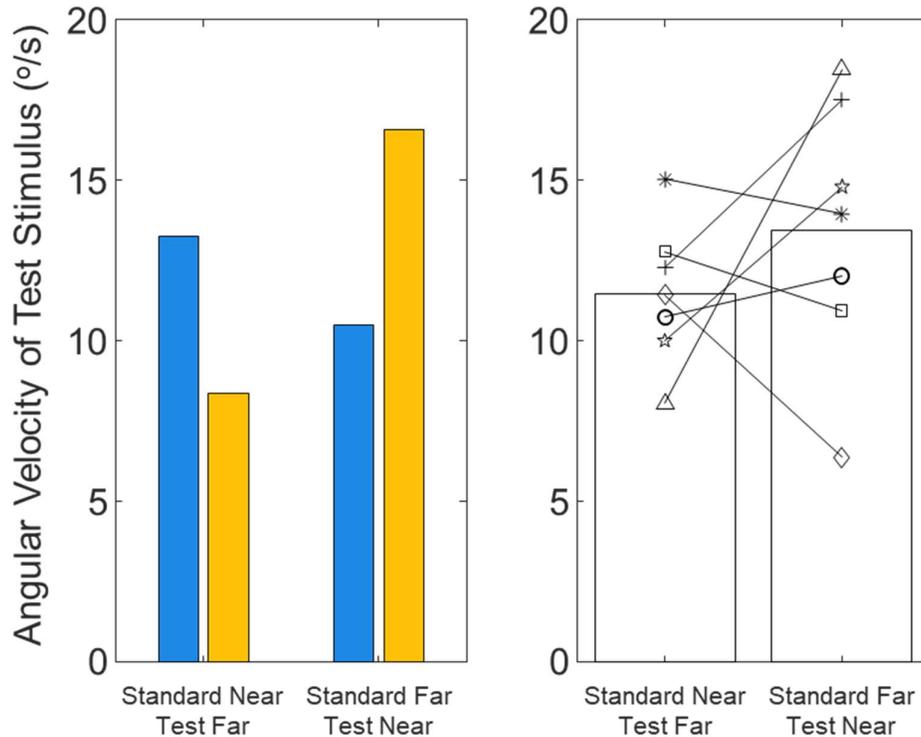


Figure 6.6: (A) predictions of the angular velocity of the test stimulus at the PSE for no speed constancy (blue) and perfect speed constancy (yellow) in the standard near and standard far conditions. (B) behavioural data showing the angular velocity of the test stimulus at the PSE for each participant in each of the standard near and standard far conditions. Each participant is denoted by a different symbol and pairs of symbols are connected by lines. Bars represent the mean PSE across participants for each condition.

As can be seen in Figure 6.6B, some participants (denoted by a triangle, a plus sign and a five-point star) show near perfect speed constancy, while others (denoted by a diamond, a square and an asterisk) show almost no speed constancy. In order to test for speed constancy, the angular velocities at the PSE were compared between the standard near and standard far conditions, as it should be the case that if speed constancy occurs, the angular velocity of the PSE in the standard far condition should be greater than in the standard near condition, while the opposite should be true if no speed constancy occurs. There was found to be no significant difference between the angular velocities at the PSEs in the conditions ( $t(6) = 0.996$ ,  $p = 0.358$ , NS), suggesting that speed constancy is somewhere between non-existent and perfect. From Figure 6.6B, it is likely that this outcome is due to some participants exhibiting near perfect speed constancy and some participants exhibiting no speed constancy. Clearly individual differences are important in determining whether speed constancy is present in the auditory modality.

As individuals appeared to exhibit different amounts of speed constancy, the data for two individuals from Experiment 1 is presented in Figure 6.7. The symbols refer to the same participants as the equivalent symbols in Figure 6.6B, where it can be seen that the triangle participant performed almost perfect speed constancy, while the diamond participant performed almost no speed constancy. In Figure 6.7, it can be seen that the triangle participant showed a more veridical distance perception, while the diamond participant strayed from the veridical at the further stimulus distances. It could be the case that this difference in the distance perception of these participants caused the difference between their speed constancy.

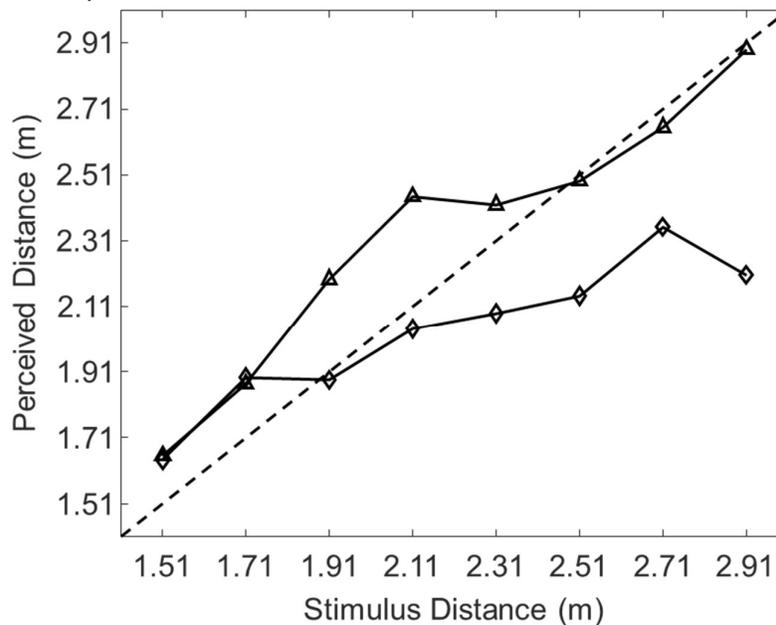


Figure 6.7: perceived stimulus distance as a function of actual stimulus distance from Experiment 1 for the two participants selected after the analysis of Experiment 2, see text for details.

## 6.4 – Discussion

This chapter has been an investigation into the presence and extent of auditory speed constancy. In the first experiment presented here, participants were able to determine the distance between themselves and an auditory stimulus, albeit with some amount of departure from veridical. This inaccuracy at the nearer and farther distances aligns with the specific distance tendency observed by previous researchers (e.g., Anderson & Zahorik, 2014; Gogel, 1969; Gogel & Tietz, 1973; Mershon & King, 1975; Zahorik et al., 2005). Distances that were perceived as being statistically significantly different were then used in Experiment 2 to

present moving stimuli in order to test the extent of participants auditory speed constancy. Results suggest that some participants perform almost perfect auditory speed constancy while other participants perform no auditory speed constancy. Overall, it appears that some auditory speed constancy is present but that the compensation is not perfect.

Looking at the literature outlined earlier, along with the conclusions of Experiment 2, it appears that the speed constancy in the auditory system, if it exists, may be incomplete. Previously, it has been shown that without all of the cues that we use to determine the distance between ourselves and visual objects, visual speed constancy is incomplete (Distler et al., 2000; McKee & Welch, 1989). While there was no attempt to remove any of the auditory cues to distance in the experiments presented here, we did find that participants' perception of distance strayed from the veridical. The accuracy of perceived distance was investigated for both the visual and auditory modalities by Anderson and Zahorik (2014). They found that auditory distance perception was less accurate and less precise than visual distance perception. It could therefore be the case that the near perfect speed constancy that occurs in the visual modality (Brown, 1931; Epstein, 1973; Rock et al., 1968) requires a highly accurate perception of distance that is generated by multiple cues in the visual system, and this may not be possible to the same extent in the auditory system. This would explain the less-than-perfect speed constancy found in Experiment 2.

An interesting finding from Experiment 2 was that some participants showed near perfect speed constancy while other participants showed nearly no speed constancy. The data from Experiment 1 for two participants, one that performed near perfect speed constancy and one that performed almost no speed constancy, were presented in Figure 6.7, showing that the participant who performed near perfect speed constancy had a more veridical distance perception than the participant who performed almost no speed constancy. This reinforces the idea that veridical distance perception is a necessary requirement of speed constancy. The increased precision of distance perception in vision, compared to audition (Anderson & Zahorik, 2014), across participants, is likely to be the cause of the finding that speed constancy is consistently nearly perfect in vision but that the same across participant consistency is not present in auditory speed constancy. Further investigations into auditory speed constancy should take into account the accuracy of individual participants' distance perception when interpreting auditory speed constancy.

## Summary

This chapter contains a departure from the method of compensation investigated in Chapters 2-5. Here, speed constancy, or the compensation for distance that we are able to perform during perception of object movement, was investigated. Investigations into visual speed constancy are plentiful, however this is not the case for audition. Experiment 1 showed that participants were able to determine the distance between themselves and auditory objects, while Experiment 2 showed that some participants employed auditory speed constancy while others did not. Interestingly, it may be the case that the participants that were more able to determine the physical distance of the auditory objects were also more able to perform speed constancy though more investigation into this is necessary. The next chapter will summarize the important discussion points from Chapters 2-5 and this chapter in a discussion of what has been gleaned about perceptual compensation for self-movement and object distance in vision and hearing throughout this thesis.

# General Discussion

## 7.1 – Summary of Findings

Throughout this thesis, experiments have been conducted to investigate perceptual compensation for self-movement and object distance in vision and hearing. Starting with the investigation into compensation for self-movement, in Chapter 2, a novel paradigm was introduced that enables the measurement of the precision of a signal encoding self-movement that combines a number of ‘non-image’ cues but excludes reafferent image motion (retinal flow). This combined signal can be used to compensate for the reafferent motion that is introduced into the image signal due to self-movement. The paradigm made the movement of the objects presented to participants dependent on their own self-movement via a ‘motion gain’ parameter. This revealed a new source of external noise that is not accounted for in standard psychometric function fitting and is based on the trial to trial variation in the self-movement. For this reason, we developed a new psychometric function fitting procedure to separate the external noise from the variability (i.e. internal noise) of the non-image signal itself. Both visual and auditory stimuli were investigated in this experiment, with participant-controlled head rotations used as an example of self-movement. Chapter 2 showed that the non-image signal appears to be less precise when the stimuli are auditory than when they are visual. As was speculated in Chapter 2, this may be due to the different motion units that auditory, visual, and self-movement perception operate with. The measurement of the displacement of an object is typically used in auditory movement perception (Carlile & Best, 2002; Freeman et al., 2014), whereas a speed measurement is typically used by the visual system (Freeman et al., 2018; Reisbeck & Gegenfurtner, 1999). It was speculated that non-image signals either originate in acceleration units (in the vestibular system) or speed units (from motor/proprioceptive cues), meaning that there are fewer transformations needed to get the image and non-image signals into the same units in the visual version of the experiment than the auditory version. This experiment also revealed another finding. It appears that there is a consistent bias in compensation for self-movement across both the visual and auditory modalities, such that objects appear slower when they are linked to self-movement than when the same stimulus passes by a stationary observer. This was referred to as a self-movement compensation error throughout this thesis. Other

notable example of errors of this type include the Filehne illusion (Filehne, 1922) and the Aubert-Fleischl phenomenon (Aubert, 1887; Fleischl, 1882).

In Chapter 3, the same paradigm was used to investigate whether non-image signals obey Weber's law. This required participants to rotate their heads at different average speeds, which was achieved using a head rotation speed training session. The paradigm, developed in Chapter 2, also enables the measurement of the precision of the image signal encoding object movement with no self-movement present. This was used in Chapter 3 to compare the precisions of the image and non-image signals by measuring the precision of each signal at different speeds. Only auditory stimuli were used in this experiment. Participants learned to rotate their heads at roughly equivalent speeds to the target speeds of the training sessions. The precisions of the image and non-image signals were calculated for each of the average head speeds investigated. Results showed that the auditory image signal precision was not affected by head rotation speed, whereas the non-image signal was affected by head rotations speed, with faster head rotations causing less precise non-image signals. The findings of this experiment corroborate with other relevant investigations into Weber's law in motion perception. It was also clear that the auditory image signal was less precise than the non-image signal at all the measured speeds, a finding that cast some doubt on the current Bayesian explanations of self-movement compensation errors. As mentioned throughout this thesis, a standard Bayesian model needs the difference in precision to be the other way round to predict the self-movement compensation error that is found throughout Chapters 2-4. Again, the self-movement compensation error appeared to be quite constant in Chapter 3, this time within the same modality (audition) but across different speeds.

Chapter 4 further investigated the finding from Chapter 3 that seems to contradict the predictions of the standard Bayesian model by replicating this result, this time across both visual and auditory modalities. This was done by introducing a third phase to the paradigm that allowed the speeds of the image and non-image signals to match when their precision was measured (this is important due to the effects of Weber's law on the non-image signal). It was found in Chapter 4 that the precisions of the image and non-image signals were similar and this was the case for both the auditory and visual versions of the experiment. This is not a finding that can be explained by a standard Bayesian model. Then, a noise manipulation was introduced to create a key test of the Bayesian model in this context. In Bayesian modelling,

there is an inherent link between the precision and accuracy of perception, with less precise sensory evidence causing greater shifts toward the prior. Introducing a dynamic stimulus width jitter to the intervals measuring image signal precision should therefore have caused the image signal to be less precise and the object movement to be perceived as slower than without the jitter. This would, in turn, decrease the self-movement compensation error. The dynamic stimulus width jitter was unsuccessful at decreasing the precision of the image signal significantly in either the auditory or visual versions of the experiment and the self-movement compensation error remained constant when the jitter was added. Overall, Chapter 4 provided evidence that suggested that the standard Bayesian model needs to be updated to include biased sensory evidence, as the non-image signal is not less precise than the image signal.

While it appears, from Chapter 4, that the standard Bayesian model may not be appropriate in the experiments presented throughout this thesis, it is still useful to derive a standard Bayesian model that can account for the additional external variability, identified in Chapter 2, due to the trial-by-trial variability of self-movement. This Bayesian model was derived in Chapter 5. The implementation created an opportunity to validate the qualitative arguments against the standard Bayesian model that were discussed in previous chapters with some quantitative evidence, by comparing the measurements of the precisions of the image and non-image signals in the psychometric model used in Chapters 2, 3 and 4, and the Bayesian model derived in Chapter 5. This comparison showed that the Bayesian model produced measurements of the image and non-image signal precisions which suggested that the non-image signal was less precise than the image signal, while the psychometric model had previously demonstrated that the image signal was less precise than the non-image signal. The goodness of fit of the psychometric functions generated in each case were compared and it was clear that the Bayesian model produced a significantly less good fit to the behavioural data than the psychometric model. It was speculated that this difference in the goodness of fit was due to the difference in the precisions of the signals, with the Bayesian model inflating the variability of the non-image signal in order to model the bias in the data. This caused the psychometric functions from the Bayesian model to be shallower than the ones from the psychometric model, which caused the difference in goodness of fit.

Finally, Chapter 6 contained an investigation into auditory speed constancy, where we compensate for the distance between ourselves and an object when interpreting its movement. While speed constancy is well-documented in the visual literature, it had not been investigated in the auditory modality. An initial experiment showed that participants were able to distinguish between sounds presented at two distances using only auditory cues. Then, moving auditory stimuli were presented at two distinguishable distances with their movement being compared by participants. If the Point of Subjective Equality (PSE), where the participants would perceive two objects to be moving at the same speed, occurred when the angular velocities of the stimuli were equivalent, that would suggest that participants performed no speed constancy. In contrast, if the linear velocities of the stimuli matched at the PSE, then participants were correctly able to compensate for the difference between the presentation distances of the two stimuli when interpreting their movement. Overall, the data suggested that participants were able to perform some compensation but that the compensation was not complete, however, when individual participants were compared, it was clear that some participants performed near perfect speed constancy, while others performed almost no speed constancy. A link was suggested between the extent of the speed constancy performed and the ability of the participants to veridically perceive the distance between themselves and the auditory objects in the initial experiment.

These experiments are all connected by the theme of compensation during movement perception. The main finding that trends throughout this thesis is that we make errors during compensation. The self-movement compensation error in Chapters 2, 3, and 4 is consistent across modality (auditory and visual), across different self-movement speeds, and both with and without a dynamic stimulus width jitter, while compensation is not complete during auditory speed constancy, in Chapter 6, either. Explaining why these compensation errors occur is more complicated than finding them, with evidence from this thesis suggesting that the current Bayesian model needs adaptation before it is appropriate in the situation of compensating for motion due to self-controlled head rotation. The findings of this thesis also create a focus on the external variance that is present during the presentation of stimuli based on variable self-movements, with an adapted psychometric function and Bayesian model derived to account for this additional variability.

## 7.2 – Relation to Previous Work

It is not a novel finding that self-movement compensation errors exist. These errors include the Filehne illusion (Filehne, 1922) and the Aubert-Fleischl phenomenon (Aubert, 1887; Fleischl, 1882), for example, and have been researched in many different contexts including eye movements (e.g., Aubert, 1887; Filehne, 1922; Fleischl, 1882); hand movements (Moscatelli et al., 2015); hearing during head rotations (Freeman et al., 2017); and passive observer rotation (Garzorz et al., 2018), and translation (Dyde & Harris, 2008). What had not been investigated before this thesis is the precisions of the signals used during compensation for self-controlled head rotation. Allowing participants to control their own movement decreases the amount of experimenter control over the stimulus, inviting the use of multiplicative gain values to allow some experimenter control over the movement of the stimuli. This method of stimulus manipulation is also not novel as it has been used in the Virtual Reality literature (Serafin et al., 2013; Steinicke et al., 2009). Where this thesis expanded on that previous work is in the recognition of an external variance source that is not accounted for in the standard psychometric function fits of previous studies. This external variability, due to the inconsistency of participant movements, should be taken into account during all psychometric function fits using stimuli that are based on participant movement. This includes when modelling behavioural data with Bayesian models.

In an investigation into whether the image and non-image signals used to measure object movement during self-movement follow Weber's law, it was found that the non-image signal does follow Weber's law, while the auditory image signal does not. Both of these results echo previous findings that suggest that Weber's law breaks down during auditory motion perception at slow speeds (Altman & Viskov, 1977), but holds during passive vestibular stimulation (Mallery et al., 2010) and the vestibulo-ocular reflex (Nouri & Karmali, 2018). The relevance of these previous works relies on the dependence of the non-image signal on vestibular cues, which may or may not be the case.

In this thesis, all of the experiments were completed in a light-treated laboratory or with participants wearing blindfolds, meaning that the participants were unable to use reafferent image motion, such as retinal flow, to estimate their self-movement. Reafferent information like this is critical to some models of our perception of object movement during self-movement, such as the flow parsing hypothesis (e.g., Rushton & Warren, 2005; Warren

& Rushton, 2008). Also, during head rotations, the only stimuli that were presented to the participants moved with their head rotation, further eliminating any image-based cue to head rotation. These experiment therefore isolated non-image signals encoding self-movement, including vestibular, motor and proprioceptive cues. The paradigm that was developed in Chapter 2 enabled the measurement of the precision of the combined non-image signal and the image signal during no self-movement. Measuring the precisions of these signals enabled the investigation of Bayesian models which allow for the interpretation of biases by assuming that accuracy is sacrificed for the sake of precision (e.g., Landy et al., 1995).

A piece of research central to this thesis is the work of Freeman et al. (2010) who investigated the precision and bias of the image and non-image signals during fixation and smooth pursuit eye movements. They proposed a Bayesian model that was adapted, in Chapter 5, to account for the paradigm that was used in this thesis. The results of their behavioural experiments supported the Bayesian model. They found more noise in the non-image eye pursuit signal compared to the image signal, and consequent slowing of perceived speed during pursuit. While similar slowing of perceived object movement during head rotation (a self-movement compensation error) was found throughout this thesis, this was not accompanied by a non-image signal that was less precise than the image signal. This suggests that the Bayesian model proposed by Freeman et al does not generalise to other forms of self-movement. Our research is not the first to investigate the standard Bayesian model of motion perception and find a distinct lack of support. Investigations into the perceived speed of objects with differing luminance have found evidence that a standard Bayesian model with unbiased sensory evidence is unable to explain (Freeman & Powell, 2022; Hassan & Hammett, 2015). The experiments in Chapters 3 and 4, and the quantitative test of the standard Bayesian model at the end of Chapter 5, add to this body of work.

Finally, when interpreting the results of Chapter 6, to the author's knowledge, the first investigation into auditory speed constancy, it is important to start with the initial investigation that determined that participants were able to perceived the difference between the distances of auditory objects. The results from this initial experiment matched those from previous studies showing a specific distance tendency (e.g., Anderson & Zahorik, 2014; Gogel, 1969; Gogel & Tietz, 1973; Mershon & King, 1975; Zahorik et al., 2005). This specific distance tendency was at an object distance of around 2-2.1m where participants'

distance perception was nearly veridical. At closer and farther object distances, the perceived distance tended towards 2-2.1m. This finding has been shown to be more prominent in auditory distance perception than visual distance perception (Anderson & Zahorik, 2014), which may have interesting consequences for the experiments in Chapter 2-4. In the experiments in those chapters, the stimuli, auditory and visual, were presented at a distance of 1.2m from the participants. Despite these objects being presented at the same physical distance from the participants, it is likely that the perceived object distances of the auditory and visual objects were not the same. This speculation is supported by the findings of Anderson and Zahorik (2014), who fit functions of the perceived distance with respect to the target distance for the auditory, visual and audiovisual versions of their experiment. They include values for the fitted parameters, which allows us to calculate that at an object distance of 1.2m, the participants perceived the auditory stimulus to be at 1.61m ( $1.44 * 1.2^{0.62} = 1.61$ ) and the visual stimulus to be at 1.12m ( $0.94 * 1.2^{0.95} = 1.12$ ). Assuming that this was the case in our experiments, this difference in apparent object distance could explain the difference between the precisions of the non-image signals in the auditory and visual modalities.

It was found in Chapter 2, and replicated in Chapter 4, that the non-image signal was less precise in the auditory version of the experiment, than the visual. It was speculated that this may be due to the number of conversions that need to be implemented to allow non-image, and visual or auditory image signals to be compared. However, this potential difference in object distance perception offers another interesting explanation. Assuming that the precision of our motion perception is affected by perceived speed (rather than actual speed), which has been the finding of studies investigated adapted visual motion perception (Bex et al., 1999; Clifford & Wenderoth, 1999), it could be the case that the difference in precision between the non-image signals in the auditory and visual versions of the experiment is due to a difference in the perceived speeds of the auditory and visual objects in the experiments (due to Weber's law, which was found to be relevant for non-image signals in Chapter 3). This difference in perceived speed could arise from the difference in perceived distance, due to speed constancy. If an auditory object is perceived to be further away from an observer than a visual object, and if the movement (and image motion) of the two objects is equivalent, the perceived 3D speed in the world of the auditory object would be greater

than the perceived 3D speed of the visual object. Here, the finding of, albeit incomplete, auditory speed constancy in Chapter 6, along with the findings of Anderson and Zahorik (2014), with the auditory version replicated in Chapter 6, offer a potential explanation for the finding in Chapter 2 (replicated in Chapter 4) that the non-image signal is less precise in the auditory version of the paradigm than the visual version.

The results of the second experiment in Chapter 6 suggest that some auditory speed constancy is present, with differences between the speed constancy employed by certain participants, and those differences being linked with the accuracy of perceived object distance. This is in contrast with visual speed constancy, that is almost complete (Brown, 1931; Epstein, 1973; Rock et al., 1968). It appears likely that the completeness of speed constancy is dependent on the accuracy of distance perception, with marked changes in visual speed constancy when depth cues are limited (Distler et al., 2000; McKee & Welch, 1989), alongside the suggestion in Chapter 6 that participants who had more accurate distance perception performed more complete speed constancy. This may explain why auditory velocity constancy is not as complete, or as consistent, as visual velocity constancy as visual distance perception is more accurate and precise than auditory distance perception (Anderson & Zahorik, 2014).

### 7.3 – Future Directions

Potential investigations that may stem from the work presented here on the precision of image and non-image signals during compensation for self-movement include investigating different types of self-movement. As mentioned earlier, the extent of self-movement compensation errors and the precisions of the signals involved have been investigated in difference movement contexts already (e.g., Aubert, 1887; Dyde & Harris, 2008; Filehne, 1922; Fleischl, 1882; Freeman et al., 2017; Garzorz et al., 2018; Moscatelli et al., 2015), however, the paradigm used in Chapter 2-4 of this thesis gives a new opportunity to investigate the precision of our non-image and image signals during self-controlled self-movements, allowing for future research to utilise different types of self-movement (e.g., walking). This paradigm and, importantly, the psychometric model that comes with it, can be used in conjunction with VR to investigate the perception of object and self-movement in even more different contexts. From the results presented in this thesis, predictions for these

experiments would include: the modality of the stimuli used may have an effect on the measurement of the precision of the non-image signal; the non-image signal should follow Weber's law, meaning that its precision decreases with increasing magnitude; and standard Bayesian models may not be able to account for the biases and precisions found.

In the Bayesian model that was derived in this thesis, the external variability due to stimuli being driven by the variable self-movements of participants was accounted for. While the Bayesian model provided a less good fit than the psychometric model (derived in Chapter 2), its derivation may prove useful for future research investigating Bayesian behaviour and utilising stimuli that move at speeds that are determined by participants' self-movements. It would be interesting to see if a manipulation that has been shown to produce Bayesian behaviour, like luminance jitter (e.g., in experiment 4 of Freeman & Powell, 2022) also produces Bayesian behaviour in a situation where the stimuli are dependent on variable self-movements.

Another interesting direction for future research would be to further investigate auditory speed constancy. This could include investigating a wider range of object distances to see how auditory speed constancy differs as a function of the separation of the objects; investigating which auditory depth cues are the most important in determining the extent of auditory speed constancy; or further investigating the variation in speed constancy across individual participants that was evident in Chapter 6. There are many potential avenues of research in auditory speed constancy that have not yet been investigated.

## 7.4 – Thesis Conclusions

In this thesis, I have presented experiments that investigate perceptual compensation for self-movement and object distance in vision and hearing. In Chapters 2-4, the precisions of the non-image and image signals encoding self-movement and image motion were investigated. These investigations spanned three contexts. In Chapter 2, the paradigm that was used throughout these chapters was introduced and comparisons were made between the precisions of the non-image signals when visual or auditory stimuli were used. It was found that the non-image signal was less precise when auditory stimuli were used which, it was speculated, is due to the number of coordinate transformations that are necessary when comparing non-image signals to visual or auditory image signals. This chapter also introduced

the concept that when stimuli are based on variable self-movements, there is an external source of variability that is not accounted for in standard psychometric function fitting routines, and a new routine that took this into account was derived. In Chapter 3, a test of Weber's law was presented which showed that the precision of the non-image signal is dependent on the speed of self-movement, while the precision of the auditory image signal is not. In Chapter 4, a Bayesian model was evaluated and it was suggested that this model was not relevant in the context of object movement perception during self-controlled head rotations, as it could not explain the self-movement compensation error that persisted throughout these three chapters while simultaneously modelling the finding that the non-image signal does not contain more noise than the image signal in either modality. This qualitative finding was quantitatively tested in Chapter 5, where a Bayesian model was derived, the first of its kind to account for the external variability due to stimuli being based on self-movements. The Bayesian model provided a significantly worse fit than the psychometric fitting routine developed in Chapter 2. The focus of Chapter 6 was then shifted to distance perception with an investigation into auditory speed constancy, the undocumented counterpart to visual speed constancy. While there was variation between participants, some who performed speed constancy and some who did not, overall, the results appeared to show that this kind of compensation is present but incomplete in the auditory system. Throughout Chapters 2-4, the auditory and visual stimuli were presented at the same distance from the participant, however the biases found in auditory depth perception in Chapter 6 could point to differences in the perceived distance of the stimuli, which may underlie the differences between the precisions of the non-image signal when compared to a visual, or an auditory, stimulus. When we perform self-movement, we must compensate for both the reafferent motion that is induced in our image signals and for the distance between us and objects so that we can interpret the world veridically. Investigation into how this is completed is ongoing, however this thesis provides methodology for further investigation, and insight into how compensation is performed.

# Thesis Appendix

## FitSMModel

```
function [best_model, pse2, gof1, gof2] = fitSMmodel(gain1, gain2, P1, P2, pse1,
n_per, H_mn, H_var, max_lapse, error_type, n_search, is_plot)
% Inputs
% gain1,gain2 = motion gains used in experiment for Phase 1 and 2
% P1,P2 = probability choosing interval 2 moved more
% pse1 = pse from Phase 1 used to set standard in Phase 2, in motion gain units
% n_per = number of trials per motion gain level in pmf, assumed all the same
% H_mn = mean measured head velocity% H_var = standard deviation
% max_lapse = typically < 0.06, as per wichman & hill 2001
% error_type = error to minimise on; least-squares 'ls' or 'mle'
%
% Outputs:
% best_model = fitted model params:
% (1) = bias as a proportion of head signal, latter in units set by H_* (=
degree of aubert, <1 classic, >1 reversed)
% (2) = h_var, head signal variance, again in units set by H_*
% (3) = i_var, image signal variance, ditto
% (4) = lapse_rate for phase 1
% (5) = lapse_rate for phase 2
% pse2 = pse of best fitting pmf in phase 2, in motion gain units
% gof1, gof2 = goodness-of-fit measures for Phase 1 and 2 [rms, deviance]
%
% Strategy:
% Fit Phase 2 data first with pse, i_var and lapse_rate free to vary
% Fit Phase 1 with bias, h_var and lapse_rate free to vary, but i_var now fixed by
previous step.
%
% Avoid local minima using n_search to set granularity of search space per
% free variable, with ranges based on educated guesses as defined below
%
% Tom Freeman, c.June 2023

global max_lapse
% from fitpmf: when error_type = 'mle', this get empirical Ps equal to 0 or 1 away
from floor or
% ceiling so log likelihood can be calculated without returning NaN
tweak = 0.001;
P1(P1==0) = tweak;
P1(P1==1) = P1(P1==1)-tweak;
P2(P2==0) = tweak;
P2(P2==1) = P2(P2==1)-tweak;
options = optimset('Display', 'off', 'FunValCheck', 'on');
try_var = logspace(0, 2, n_search); % either image or head signal
try_lapse = linspace(0, max_lapse, n_search);
try_bias = linspace(0.5, 1.5, n_search);
% fit phase 2 first to get i_var
err_1 = inf;
params_phase2 = [];
iii = 0; n = n_search^3; z = 4; zz = num2str(z); % merely for text output to
command line to show how far its got
mssg = 'phase 2: ';
fprintf('%10s',mssg)
for i = 1 : n_search
```

```

for r = 1 : n_search
for b = 1 : n_search
iii=iii+1; fprintf(['%zz 'i'],n-iii+1)
start = [try_bias(b) try_var(i) try_lapse(r)];
[current_best, ~, exitflag] = fminsearch(@fitfunc_phase2, start, options, gain2,
P2, n_per, pse1, H_mn, H_var, error_type);
current_best = current_best.^2; % see fitfuncs
err = error_metric(P2, n_per, pmf_phase2(current_best, gain2, pse1, H_mn, H_var),
error_type);
if exitflag && (err < err_1)
err_1 = err;
params_phase2 = current_best;
end
for s = 1 : z, fprintf('\b'), end
end
end
end
for s = 1 : 10, fprintf('\b'), end
pse2 = params_phase2(1)*pse1; % i.e. convert into motion gain units, see
pmf_phase2: this is for output and plotting only
% Now fit phase 1 to get bias and h_var with i_var fixed by above
gof1 = NaN;
gof2 = NaN;
best_model = NaN;
mssg = 'phase 1: ';
fprintf('%10s',mssg)
if ~isempty(params_phase2)
i_var = params_phase2(2);
params_phase1 = [];
err_1 = inf;
iii = 0;
for i = 1 : n_search
for r = 1 : n_search
for b = 1 : n_search
iii=iii+1; fprintf(['%'zz 'i'],n-iii+1)
start = [try_bias(b) try_var(i) try_lapse(r)];
[current_best, ~, exitflag] = fminsearch(@fitfunc_phase1, start, options, gain1,
P1, n_per, H_mn, H_var, i_var, error_type);
current_best = current_best.^2;err = error_metric(P1, n_per,
pmf_phase1(current_best, gain1, H_mn, H_var, i_var), error_type);
if exitflag && (err < err_1)
err_1 = err;
params_phase1 = current_best;
end
for s = 1 : z, fprintf('\b'), end
end
end
end
for s = 1 : 10, fprintf('\b'), end
gof1(1) = error_metric(P1, n_per, pmf_phase1(params_phase1, gain1, H_mn, H_var,
i_var), 'rms');
gof2(1) = error_metric(P2, n_per, pmf_phase2(params_phase2, gain2, pse1, H_mn,
H_var), 'rms');
gof1(2) = error_metric(P1, n_per, pmf_phase1(params_phase1, gain1, H_mn, H_var,
i_var), 'dev');
gof2(2) = error_metric(P2, n_per, pmf_phase2(params_phase2, gain2, pse1, H_mn,
H_var), 'dev');
end

```

```

if is_plot
sz = 30;
x0 = min([gain1 gain2]);
x1 = max([gain1 gain2]);
savefig = gcf;
figure(101), clf
subplot(211), hold on
x = linspace(x0, x1, 100);
pse = params_phase1(1);
if isnan(params_phase1)
plot(gain1, P1, 'rx')
else
plot(x, pmf_phase1(params_phase1, x, H_mn, H_var, i_var), 'k-')
plot(gain1, P1, 'k.', 'markersize', sz)
line([x0 pse], [0.5 0.5], 'linestyle', ':', 'color', 'k')
line([pse pse], [0.5 0], 'linestyle', ':', 'color', 'k')
end
title('Phase 1')
ylabel('P')
ylim([0 1])
xlim([x0 x1])
subplot(212), hold on
if isnan(params_phase2)
plot(gain2, P2, 'rx')
else
pp = pmf_phase2(params_phase2, x, pse1, H_mn, H_var);
plot(x, pp, 'k-')
plot(gain2, P2, 'k.', 'markersize', sz)
line([x0 pse2], [0.5 0.5], 'linestyle', ':', 'color', 'k')
line([pse2 pse2], [0.5 0], 'linestyle', ':', 'color', 'k')
end
title('Phase 2')
xlabel('gain')
ylabel('P')
ylim([0 1])
xlim([x0 x1])
figure(savefig.Number)
end
% reordered output for backwards compatibility with older local analyses scripts
% pse2 returned separately as isn't really part of model
best_model = [params_phase1(1:2) i_var params_phase1(3) params_phase2(3)];
% bias, h_var, i_var, lapse_rate1, lapse_rate2

function err = fitfunc_phase1(params, gain, P_data, n_per, H_mn, H_var, i_var,
error_type)
global max_lapse
params = params.^2; % force values to be positive, but needs to square best_fit
params outside to return correct values
P_model = pmf_phase1(params, gain, H_mn, H_var, i_var);
err = error_metric(P_data, n_per, P_model, error_type);
if params(3)>max_lapse.^2+0.0001, err = inf; end% gets max_lapse = 0 working

function P = pmf_phase1(params, gain, H_mn, H_var, i_var)
bias = params(1);
h_var = params(2);
lapse_rate = params(3);
mn_d = (gain-bias)*H_mn;
sd_d = sqrt( (gain-bias).^2*H_var + h_var + i_var );
sd_d(sd_d==0) = 0.0001;% avoids /0 on next line

```

```

z = mn_d./sd_d;
P = lapse_rate/2 + (1-lapse_rate)*((1 + erf(z./sqrt(2)))/2);

function err = fitfunc_phase2(params, gain, P_data, n_per, pse1, H_mn, H_var,
error_type)
global max_lapse
params = params.^2;
P_model = pmf_phase2(params, gain, pse1, H_mn, H_var);
err = error_metric(P_data, n_per, P_model, error_type);
if params(3)>max_lapse.^2+0.0001, err = inf; end% gets max_lapse = 0 working

function P = pmf_phase2(params, gain, pse1, H_mn, H_var)
alpha = params(1); % the pse for the pmf, as a proportion, see mn_d below; takes
account of sampling error, button bias etc
i_var = params(2);
lapse_rate = params(3);
mn_d = (gain-alpha*pse1)*H_mn;
sd_d = sqrt( (gain-pse1).^2*H_var + 2*i_var );
sd_d(sd_d==0) = 0.0001;% avoids /0 on next line
z = mn_d./sd_d;
P = lapse_rate/2 + (1-lapse_rate)*((1 + erf(z./sqrt(2)))/2);

% deviance calculation from psignifit, Wichman & Hill
function dev = deviance(y, n, p);
n = round(n);
r = round(n .* y);
w = n -r;
y = r ./ n;
residuals = 2 * (xlogy(r, y) + xlogy(w, 1-y) -xlogy(r, p) -xlogy(w, 1-p));
residuals(residuals < 0) = 0;
dev = sum(residuals, 2);

function a = xlogy(x, y)
k = (y==0);
y(k) = nan;
y(x==0 & k) = 1;
a = x.*log(y);
k = isnan(y);
a(k) = -sign(x(k)) * inf;

function err = error_metric(data, n, model, error_type)
% log likelihood
if strcmp(error_type, 'mle')
err = sum( n.*data.*log(data./model) + n.*(1-data).*log( (1-data)./(1-model)) ); %
from Klein (2001)
% least-squares
elseif strcmp(error_type, 'ls')
err = sum( (data-model).^2 )
elseif strcmp(error_type, 'rms')
err = sqrt(mean((data-model).^2));
elseif strcmp(error_type, 'dev')
err = deviance(data, n, model);
else
error('**fitSMmodel: unknown error_type')
end

```

## BayesianPh1Fit

```
function [BestModel, best_val, gof1] = BayesianPh1Fit(GainValues, Headspeed,
HeadspeedVariance, ImageSigVariance, SearchGridNonImageSigVariance,
SearchGridPriorVariance, SearchGridExternalVariance, Data, plot1)
% Inputs
% GainValues - the values for the motion gains of the test stimuli
% Headspeed - the mean headspeed calculated from the behavioural data
% HeadspeedVariance - the variance of the headspeed calculated from the
%behavioural data
% ImageSigVariance - the variance of the image signal calculated from the
%Phase 2 analysis
% SearchGridNonImageSigVariance - the search grid for the variance of the
%non-image signal % [1 100]
% Same for Prior and External variance
% Data - the behavioural data (number of test chosen)
% plot1 - whether or not you want to plot (0 or 1)
%
% Outputs
% BestModel - best fitting values at the positions of the num_chosen
%values for comparison
% best_val - best fitting values for the non-image signal variance, prior
%variance, and external variance respectively
% gof1 - root mean square error and deviance respectively

H = Headspeed;
sdH = HeadspeedVariance;
sdi = ImageSigVariance;
g = GainValues;
n_per = 10;
Data = Data/n_per;
error_type = 'mle';
iterationmax = 20;
iteration = 1;
SearchGridNonImageSigVariance =
linspace(SearchGridNonImageSigVariance(1),SearchGridNonImageSigVariance(end),itera
tionmax);
SearchGridPriorVariance =
linspace(SearchGridPriorVariance(1),SearchGridPriorVariance(end),iterationmax);
SearchGridExternalVariance =
linspace(SearchGridExternalVariance(1),SearchGridExternalVariance(end),iterationma
x);
options = optimset('Display', 'off');
err1 = inf;
for a = 1:iterationmax
    for b = 1:iterationmax
        for c = 1:iterationmax
            start = [SearchGridNonImageSigVariance(a) SearchGridPriorVariance(b)
SearchGridExternalVariance(c)];
            [current_val,~,exitflag] = fminsearchbnd(@fitfunc_phase,start,[sdi 0
0],[,],options,g,H,sdH,sdi>Data,n_per,error_type);
            err = fitfunc_phase(current_val,g,H,sdH,sdi>Data,n_per,error_type);
            errors(iteration) = err;
            iteration = iteration + 1;
            if err<err1, err1 = err; best_val = current_val; end
        end
    end
end
sdh = best_val(1);
```

```

sdp = best_val(2);
sde = best_val(3);
for j = 1:size(g)
    mu_d(j) = H*sdp*(g(j)/(sdi+sde+sdp)-1/(sdh+sde+sdp));
    sd_d(j) =
sqrt((sdH*g(j)^2+sdi)*(sdp/(sdi+sde+sdp))^2+(sdH+sdh)*(sdp/(sdh+sde+sdp))^2-
2*sdH*g(j)*sdp^2/((sdi+sde+sdp)*(sdh+sde+sdp)));
    BestModel(j) = mu_d(j)/(sqrt(2)*sd_d(j));
end
BestModel = (1/2)*(1+erf(BestModel));

gof1(1) = error_metric(BestModel, Data, n_per, 'rms');
gof1(2) = error_metric(BestModel, Data, n_per, 'dev');

if plot1 == 1
    gfine = linspace(min(g),max(g),100);
    Phase1 = Bayes_model(current_val,gfine,H,sdH,sdi);
    figure
    plot(gfine,Phase1)
    hold on
    plot(g,Data, '.')
    savefig('C:\Users\Josh\OneDrive\Documents\Combined
Fred\Experiments\Analysis\NonPMFanalysis\Noisy\figs\Chapter5Figs\BayesExample.fig'
)
end
end

function err = fitfunc_phase(params,g,H,sdH,sdi,Data,n_per,error_type)
sdh = params(1);
sdp = params(2);
sde = params(3);
for i = 1:size(g)
    mu_d(i) = H*sdp*(g(i)/(sdi+sde+sdp)-1/(sdh+sde+sdp));
    sd_d(i) =
sqrt((sdH*g(i)^2+sdi)*(sdp/(sdi+sde+sdp))^2+(sdH+sdh)*(sdp/(sdh+sde+sdp))^2-
2*sdH*g(i)*sdp^2/((sdi+sde+sdp)*(sdh+sde+sdp)));
    Phase1(i) = mu_d(i)/(sqrt(2)*sd_d(i));
end
Phase1 = (1/2)*(1+erf(Phase1));
err = error_metric(Phase1,Data,n_per,error_type);
if any(params > sdi*2.5), err = err^2; end
end

function Phase1 = Bayes_model(params,g,H,sdH,sdi)
sdh = params(1);
sdp = params(2);
sde = params(3);
for i = 1:100
    mu_d(i) = H*sdp*(g(i)/(sdi+sde+sdp)-1/(sdh+sde+sdp));
    sd_d(i) =
sqrt((sdH*g(i)^2+sdi)*(sdp/(sdi+sde+sdp))^2+(sdH+sdh)*(sdp/(sdh+sde+sdp))^2-
2*sdH*g(i)*sdp^2/((sdi+sde+sdp)*(sdh+sde+sdp)));
    Phase1(i) = mu_d(i)/(sqrt(2)*sd_d(i));
end
Phase1 = (1/2)*(1+erf(Phase1));
end

function dev = deviance(y, n, p)
n = round(n);

```

```

r = round(n .* y);
w = n - r;
y = r ./ n;
residuals = 2 * (xlogy(r, y) + xlogy(w, 1-y) - xlogy(r, p) - xlogy(w, 1-p));
residuals(residuals < 0) = 0; % can go negative due to (im)precision errors
dev = sum(residuals, 2);
end
function a = xlogy(x, y)
k = (y==0);
y(k) = nan;
y(x==0 & k) = 1;
a = x.*log(y);
k = isnan(y);
a(k) = -sign(x(k)) * inf;
end

function err = error_metric(model, data, n_per, error_type)
% log likelihood
if strcmp(error_type, 'mle')
    data(data==0) = 0.0001;
    data(data==1) = 0.9999;
    err = sum( n_per.*data.*log(data./model) + n_per.*(1-data).*log( (1-data)./(1-
model)) ); % from Klein (2001)
% least-squares
elseif strcmp(error_type, 'ls')
    err = sum( (data-model).^2 );
elseif strcmp(error_type, 'rms')
    err = sqrt(mean((data-model).^2));
elseif strcmp(error_type, 'dev')
    err = deviance(data, n_per, model);
else
    error('**fitSMmodel: unknown error_type')
end
end

```

# Reference List

- Algom, D. (2021). The Weber–Fechner law: A misnomer that persists but that should go away. *Psychological review*, 128(4), 757.
- Altman, J., & Viskov, O. (1977). Discrimination of perceived movement velocity for fused auditory image in dichotic stimulation. *The Journal of the Acoustical Society of America*, 61(3), 816-819.
- Anderson, P. W., & Zahorik, P. (2014). Auditory/visual distance estimation: accuracy and variability. *Frontiers in psychology*, 5, 1097.
- Ascher, D., & Grzywacz, N. M. (2000). A Bayesian model for the measurement of visual velocity. *Vision Research*, 40(24), 3427-3434.
- Ashmead, D. H., Leroy, D., & Odom, R. D. (1990). Perception of the relative distances of nearby sound sources. *Perception & Psychophysics*, 47, 326-331.
- Aubert, H. (1887). Die Bewegungsempfindung: Zweite Mittheilung. *Archiv für die gesamte Physiologie des Menschen und der Tiere*, 40(1), 459-480.
- Barnes, G. (1988). Head-eye co-ordination: visual and nonvisual mechanisms of vestibulo-ocular reflex slow-phase modification. *Progress in brain research*, 76, 319-328.
- Bedell, H. E., Klopfenstein, J. F., & Yuan, N. (1989). Extraretinal information about eye position during involuntary eye movement: Optokinetic afternystagmus. *Perception & Psychophysics*, 46(6), 579-586.
- Bentvelzen, A., Leung, J., & Alais, D. (2009). Discriminating audiovisual speed: optimal integration of speed defaults to probability summation when component reliabilities diverge. *Perception*, 38(7), 966-987.
- Bex, P. J., Bedingham, S., & Hammett, S. T. (1999). Apparent speed and speed sensitivity during adaptation to motion. *JOSA A*, 16(12), 2817-2824.
- Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*.
- Bridgeman, B. (2010). How the brain makes the world appear stable. *i-Perception*, 1(2), 69-72.
- Bronkhorst, A. W., & Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 397(6719), 517-520.
- Brooks, J. X., & Cullen, K. E. (2019). Predictive sensing: the role of motor signals in sensory processing. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(9), 842-850.
- Brown, J. F. (1931). The visual perception of velocity. *Psychologische Forschung*, 14(1), 199-232.
- Bruss, A. R., & Horn, B. K. (1983). Passive navigation. *Computer Vision, Graphics, and Image Processing*, 21(1), 3-20.
- Campbell, F., & Maffei, L. (1981). The influence of spatial frequency and contrast on the perception of moving patterns. *Vision Research*, 21(5), 713-721.
- Carlile, S., & Best, V. (2002). Discrimination of sound source velocity in human listeners. *The Journal of the Acoustical Society of America*, 111(2), 1026-1035.
- Carriot, J., Bryan, A., DiZio, P., & Lackner, J. (2011). The oculogyral illusion: retinal and oculomotor factors. *Experimental brain research*, 209, 415-423.
- Champion, R. A., & Warren, P. A. (2017). Contrast effects on speed perception for linear and radial motion. *Vision Research*, 140, 66-72.
- Chapman, C. E. (1994). Active versus passive touch: factors influencing the transmission of somatosensory signals to primary somatosensory cortex. *Canadian journal of physiology and pharmacology*, 72(5), 558-570.
- Chen, K. J., Keshner, E., Peterson, B., & Hain, T. (2002). Modeling head tracking of visual targets. *Journal of Vestibular Research*, 12(1), 25-33.
- Cherni, H., Métayer, N., & Souliman, N. (2020). Literature review of locomotion techniques in virtual reality. *International Journal of Virtual Reality*.

- Churan, J., Paul, J., Klingenhoefer, S., & Bremmer, F. (2017). Integration of visual and tactile information in reproduction of traveled distance. *Journal of Neurophysiology*, *118*(3), 1650-1663.
- Clark, B., & Graybiel, A. (1949). The effect of angular acceleration on sound localization: The audiogyral illusion. *Journal of Psychology*, *27*, 235.
- Clifford, C. W., & Wenderoth, P. (1999). Adaptation to temporal modulation can enhance differential speed sensitivity. *Vision Research*, *39*(26), 4324-4331.
- Coleman, P. D. (1962). Failure to localize the source distance of an unfamiliar sound. *The Journal of the Acoustical Society of America*, *34*(3), 345-346.
- Coleman, P. D. (1963). An analysis of cues to auditory depth perception in free space. *Psychological bulletin*, *60*(3), 302.
- Collins, C., & Barnes, G. (1999). Independent control of head and gaze movements during head-free pursuit in humans. *The Journal of physiology*, *515*(Pt 1), 299.
- Cullen, K. E. (2004). Sensory signals during active versus passive movement. *Current opinion in neurobiology*, *14*(6), 698-706.
- Cullen, K. E., & Zobeiri, O. A. (2021). Proprioception and the predictive sensing of active self-motion. *Current opinion in physiology*, *20*, 29-38.
- Dallmann, C. J., Ernst, M. O., & Moscatelli, A. (2015). The role of vibration in tactile speed perception. *Journal of Neurophysiology*, *114*(6), 3131-3139.
- De Bruyn, B., & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research*, *28*(12), 1323-1335.
- De Graaf, B., & Wertheim, A. H. (1988). The perception of object motion during smooth pursuit eye movements: Adjacency is not a factor contributing to the Filehne illusion. *Vision Research*, *28*(4), 497-502.
- Diener, H., Wist, E., Dichgans, J., & Brandt, T. (1976). The spatial frequency effect on perceived velocity. *Vision Research*, *16*(2), 169-IN167.
- Distler, H. K., Gegenfurtner, K. R., Van Veen, H. A., & Hawken, M. J. (2000). Velocity constancy in a virtual reality environment. *Perception*, *29*(12), 1423-1435.
- Dupin, L., & Wexler, M. (2013). Motion perception by a moving observer in a three-dimensional environment. *Journal of Vision*, *13*(2), 15-15.
- Dyde, R. T., & 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), 5-5.
- Eargle, J. (1960). Stereophonic localization: an analysis of listener reactions to current techniques. *IRE Transactions on Audio*(5), 174-178.
- Epstein, W. (1973). The process of 'taking-into-account' in visual perception. *Perception*, *2*(3), 267-285.
- Epstein, W. (1978). Two factors in the perception of velocity at a distance. *Perception & Psychophysics*, *24*(2), 105-114.
- Facchin, A., Folegatti, A., Rossetti, Y., & Farnè, A. (2019). The half of the story we did not know about prism adaptation. *Cortex*, *119*, 141-157.
- Filehne, W. (1922). Über das optische Wahrnehmen von Bewegungen. *Zeitschrift für Sinnesphysiologie*, *53*, 134.
- Fleischl, v. E. (1882). Physiologisch-optische Notizen [Physiological-optical notes], 2. Mitteilung. *Sitzung Wiener Bereich der Akademie der Wissenschaften*, *3*, 7.
- Freeman, T. C. (2001). Transducer models of head-centred motion perception. *Vision Research*, *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 Research*, *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.
- Freeman, T. C., Cucu, M. O., & Smith, L. (2018). A preference for visual speed during smooth pursuit eye movement. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(10), 1629.

- Freeman, T. C., Culling, J. F., Akeroyd, M. A., & Brimijoin, W. O. (2017). Auditory compensation for head rotation is incomplete. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 371.
- Freeman, T. C., Leung, J., Wufong, E., Orchard-Mills, E., Carlile, S., & Alais, D. (2014). Discrimination contours for moving sounds reveal duration and distance cues dominate auditory speed perception. *PLoS one*, *9*(7), e102864.
- Freeman, T. C., & Powell, G. (2022). Perceived speed at low luminance: Lights out for the Bayesian observer? *Vision Research*, *201*, 108124.
- Gamble, E. A. (1909). Minor studies from the psychological laboratory of Wellesley College: Intensity as a criterion in estimating the distance of sounds. *Psychological review*, *16*(6), 416.
- Garzorz, I. T., Freeman, T. C., Ernst, M. O., & MacNeilage, P. R. (2018). Insufficient compensation for self-motion during perception of object speed: The vestibular Aubert-Fleischl phenomenon. *Journal of Vision*, *18*(13), 9-9.
- Genzel, D., Firzlaff, U., Wiegrebe, L., & MacNeilage, P. R. (2016). Dependence of auditory spatial updating on vestibular, proprioceptive, and efference copy signals. *Journal of Neurophysiology*, *116*(2), 765-775.
- Gibson, J. J. (1954). The visual perception of objective motion and subjective movement. *Psychological review*, *61*(5), 304.
- Gibson, J. J. (1968). What gives rise to the perception of motion? *Psychological review*, *75*(4), 335.
- Glenberg, A. M., Jaworski, B., Rischal, M., & Levin, J. R. (2007). What brains are for: Action, meaning, and reading comprehension. *Reading comprehension strategies: Theories, interventions, and technologies*, *2*.
- Gogel, W. C. (1969). The sensing of retinal size. *Vision Research*, *9*(9), 1079-1094.
- Gogel, W. C., & Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception & Psychophysics*, *13*(2), 284-292.
- Graybiel, A., & Hupp, D. (1945). The oculo-gyral illusion; a form of apparent motion which may be observed following stimulation of the semicircular canals. *Navy. Sch. Aviation Med., US Navy. Air Train. Base., Pensacola, Florida (Nov.)*.
- Greene, D. C. (1968). Comments on "Perception of the Range of a Sound Source of Unknown Strength"[HR Hirsch, J. Acoust. Soc. Am. *43*, 373-374 (1968)]. *The Journal of the Acoustical Society of America*, *44*(2), 634-634.
- Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 181-197.
- Haarmeier, T., Thier, P., Repnow, M., & Petersen, D. (1997). False perception of motion in a patient who cannot compensate for eye movements. *Nature*, *389*(6653), 849-852.
- Halow, S., Liu, J., Folmer, E., & MacNeilage, P. R. (2023). Motor Signals Mediate Stationarity Perception. *Multisensory Research*, *36*(7), 703-724.
- Hammett, S. T., Champion, R. A., Thompson, P. G., & Morland, A. B. (2007). Perceptual distortions of speed at low luminance: Evidence inconsistent with a Bayesian account of speed encoding. *Vision Research*, *47*(4), 564-568.
- Harrison, J. J., Freeman, T. C., & Sumner, P. (2015). Saccadic compensation for reflexive optokinetic nystagmus just as good as compensation for volitional pursuit. *Journal of Vision*, *15*(1), 24-24.
- Hassan, O., & Hammett, S. T. (2015). Perceptual biases are inconsistent with Bayesian encoding of speed in the human visual system. *Journal of Vision*, *15*(2), 9-9.
- Herlihey, T. A., & Rushton, S. K. (2012). The role of discrepant retinal motion during walking in the realignment of egocentric space. *Journal of Vision*, *12*(3), 4-4.
- Hirsch, H. R. (1968). Perception of the range of a sound source of unknown strength. *The Journal of the Acoustical Society of America*, *43*(2), 373-374.
- Hürlimann, F., Kiper, D. C., & Carandini, M. (2002). Testing the Bayesian model of perceived speed. *Vision Research*, *42*(19), 2253-2257.

- Israel, I., & Warren, W. H. (2005). Vestibular, proprioceptive, and visual influences on the perception of orientation and self-motion in humans. *Head direction cells and the neural mechanisms of spatial orientation*, 347-381.
- Jones, P. R. (2016). A tutorial on cue combination and Signal Detection Theory: Using changes in sensitivity to evaluate how observers integrate sensory information. *Journal of Mathematical Psychology*, 73, 117-139.
- Klam, F., & Graf, W. (2003). Vestibular signals of posterior parietal cortex neurons during active and passive head movements in macaque monkeys. *Annals of the New York Academy of Sciences*, 1004(1), 271-282.
- Koenderink, J. J. (1986). Optic flow. *Vision Research*, 26(1), 161-179.
- Koenderink, J. J., & van Doorn, A. J. (1987). Facts on optic flow. *Biological cybernetics*, 56(4), 247-254.
- Laming, D. (2009). Weber's law. In *Insid. Psychol. a Sci. over 50 years*. Oxford University Press.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, 35(3), 389-412.
- Lappe, M., Bremmer, F., & van den Berg, A. V. (1999). Perception of self-motion from visual flow. *Trends in cognitive sciences*, 3(9), 329-336.
- Lee, D. N. (1980). The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 290(1038), 169-179.
- Leigh, R. J., & Zee, D. S. (1999). *The Neurology of Eye Movements: Text and CD-ROM* (Vol. 3rd ed.). Oxford University Press, USA.
- Leshowitz, B., Tavb, H. B., Raab, D. H., & College, B. (1968). Visual detection of signals in the presence of continuous and pulsed backgrounds. *Perception & Psychophysics*, 4(4), 207-213.
- Leung, J., Wei, V., Burgess, M., & Carlile, S. (2016). Head tracking of auditory, visual, and audio-visual targets. *Frontiers in neuroscience*, 9, 169200.
- Lindner, A., Thier, P., Kircher, T. T., Haarmeier, T., & Leube, D. T. (2005). Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Current Biology*, 15(12), 1119-1124.
- Llinás, R. R. (2002). *I of the vortex: From neurons to self*. MIT press.
- London, B. M., & Miller, L. E. (2013). Responses of somatosensory area 2 neurons to actively and passively generated limb movements. *Journal of Neurophysiology*, 109(6), 1505-1513.
- Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 208(1173), 385-397.
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, 9(11), 1432-1438.
- Mack, A., & Herman, E. (1973). Position constancy during pursuit eye movement: An investigation of the Filehne illusion. *The Quarterly Journal of Experimental Psychology*, 25(1), 71-84.
- Mack, A., & Herman, E. (1978). The loss of position constancy during pursuit eye movements. *Vision Research*.
- Mallery, R. M., Olomu, O. U., Uchanski, R. M., Militchin, V. A., & Hullar, T. E. (2010). Human discrimination of rotational velocities. *Experimental brain research*, 204, 11-20.
- Masin, S. C. (2009). The (Weber's) law that never was. *Proceedings of fechner day*, 25, 441-446.
- Maxfield, J. P. (1931). Some physical factors affecting the illusion in sound motion pictures. *The Journal of the Acoustical Society of America*, 3(1A), 69-80.
- McCrea, R. A., Gdowski, G. T., Boyle, R., & 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), 416-428.
- McKee, S. P., & Welch, L. (1989). Is there a constancy for velocity? *Vision Research*, 29(5), 553-561.
- Mershon, D. H., & Bowers, J. N. (1979). Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8(3), 311-322.
- Mershon, D. H., & King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 18, 409-415.

- Merton, P. (1964). Human position sense and sense of effort. *Symposia of the Society for Experimental Biology*,
- Mita, T., Hironaka, K., & Koike, I. (1950). The influence of retinal adaptation and location on the "Empfindungszeit." *The Tohoku journal of experimental medicine*, 52(3-4), 397-405.
- Molino, J. (1973). Perceiving the range of a sound source when the direction is known. *The Journal of the Acoustical Society of America*, 53(5), 1301-1304.
- Moscattelli, A., Hayward, V., Wexler, M., & Ernst, M. O. (2015). Illusory tactile motion perception: An analog of the visual Filehne illusion. *Scientific reports*, 5(1), 1-12.
- Moscattelli, A., Scotto, C. R., & Ernst, M. O. (2019). Illusory changes in the perceived speed of motion derived from proprioception and touch. *Journal of Neurophysiology*.
- Nilsson, N. C., Peck, T., Bruder, G., Hodgson, E., Serafin, S., Whitton, M., Steinicke, F., & Rosenberg, E. S. (2018). 15 years of research on redirected walking in immersive virtual environments. *IEEE computer graphics and applications*, 38(2), 44-56.
- Nouri, S., & Karmali, F. (2018). Variability in the vestibulo-ocular reflex and vestibular perception. *Neuroscience*, 393, 350-365.
- Paillard, J. (1991). Motor and representational framing of space. *Brain and space*, 163-182.
- 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. *Psychological science*, 27(12), 1562-1572.
- Prins, N., & Kingdom, F. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers in psychology*, 9, 1250.
- Redding, G. M., & Wallace, B. (1985). Cognitive interference in prism adaptation. *Perception & Psychophysics*, 37(3), 225-230.
- Redding, G. M., & Wallace, B. (2006). Generalization of prism adaptation. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 1006.
- Reisbeck, T. E., & Gegenfurtner, K. R. (1999). Velocity tuned mechanisms in human motion processing. *Vision Research*, 39(19), 3267-3286.
- Rideaux, R., & Welchman, A. E. (2020). But still it moves: static image statistics underlie how we see motion. *Journal of neuroscience*, 40(12), 2538-2552.
- Rock, I., Hill, A. L., & Fineman, M. (1968). Speed constancy as a function of size constancy. *Perception & Psychophysics*, 4, 37-40.
- Rushton, S. K., Niehorster, D. C., Warren, P. A., & Li, L. (2018). The primary role of flow processing in the identification of scene-relative object movement. *Journal of neuroscience*, 38(7), 1737-1743.
- Rushton, S. K., & Warren, P. A. (2005). Moving observers, relative retinal motion and the detection of object movement. *Current Biology*, 15(14), R542-R543.
- Senna, I., Parise, C. V., & Ernst, M. O. (2015). Hearing in slow-motion: Humans underestimate the speed of moving sounds. *Scientific reports*, 5(1), 14054.
- Serafin, S., Nilsson, N. C., Sikstrom, E., De Goetzen, A., & Nordahl, R. (2013). Estimation of detection thresholds for acoustic based redirected walking techniques. 2013 IEEE Virtual Reality (VR),
- Skavenski, A. A. (1972). Inflow as a source of extraretinal eye position information. *Vision Research*, 12(2), 221-IN227.
- Smith, A., & Edgar, G. (1990). The influence of spatial frequency on perceived temporal frequency and perceived speed. *Vision Research*, 30(10), 1467-1474.
- 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, 61-76.
- St George, R. J., & Fitzpatrick, R. C. (2011). The sense of self-motion, orientation and balance explored by vestibular stimulation. *The Journal of physiology*, 589(4), 807-813.
- Steinberg, J., & Snow, W. (1934). Physical factors. *Bell System Technical Journal*, 13(2), 245-258.

- Steinhauser, A. (1879). XXXIII. The theory of binaural audition. A contribution to the theory of sound. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 7(42), 181-197.
- Steinicke, F., Bruder, G., Jerald, J., Frenz, H., & Lappe, M. (2009). Estimation of detection thresholds for redirected walking techniques. *IEEE transactions on visualization and computer graphics*, 16(1), 17-27.
- Stevens, S. S., & Guirao, M. (1962). Loudness, reciprocity, and partition scales. *The Journal of the Acoustical Society of America*, 34(9B), 1466-1471.
- Stevenson-Hoare, J. O., Freeman, T. C., & Culling, J. F. (2022). The pinna enhances angular discrimination in the frontal hemifield. *The Journal of the Acoustical Society of America*, 152(4), 2140-2149.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, 9(4), 578-585.
- Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, 22(3), 377-380.
- Thompson, P., Brooks, K., & Hammett, S. T. (2006). Speed can go up as well as down at low contrast: Implications for models of motion perception. *Vision Research*, 46(6-7), 782-786.
- Thompson, S. P. (1882). LI. On the function of the two ears in the perception of space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(83), 406-416.
- Turano, K. A., & Massof, R. W. (2001). Nonlinear contribution of eye velocity to motion perception. *Vision Research*, 41(3), 385-395.
- Tuthill, J. C., & Azim, E. (2018). Proprioception. *Current Biology*, 28(5), R194-R203.
- von Békésy, G. (1949). The moon illusion and similar auditory phenomena. *The American journal of psychology*, 62(4), 540-552.
- Von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behaviour*.
- Wallach, H. (1939). On constancy of visual speed. *Psychological review*, 46(6), 541.
- Warren, P. A., & Rushton, S. K. (2007). Perception of object trajectory: Parsing retinal motion into self and object movement components. *Journal of Vision*, 7(11), 2-2.
- Warren, P. A., & Rushton, S. K. (2008). Evidence for flow-parsing in radial flow displays. *Vision Research*, 48(5), 655-663.
- Warren, P. A., & Rushton, S. K. (2009). Optic flow processing for the assessment of object movement during ego movement. *Current Biology*, 19(18), 1555-1560.
- Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, 5(10), 7-7.
- Weiss, Y., & Adelson, E. H. (1998). Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature neuroscience*, 5(6), 598-604.
- Welchman, A. E., Lam, J. M., & Bühlhoff, H. H. (2008). Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences*, 105(33), 12087-12092.
- 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.
- Wertheim, A. H. (1994). Motion perception during selfmotion: The direct versus inferential controversy revisited. *Behavioral and Brain Sciences*, 17(2), 293-311.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8), 1293-1313.
- Woodworth, R. S., & Schlosberg, H. (1938). Experimental Psychology. New York: Henry Holt & Co. In: Inc.

- Wyatt, H. J. (1998). Detecting saccades with jerk. *Vision Research*, 38(14), 2147-2153.
- Yasui, S., & Young, L. R. (1975). Perceived visual motion as effective stimulus to pursuit eye movement system. *Science*, 190(4217), 906-908.
- Yost, W. A., & Pastore, M. T. (2019). Sound-source localization when listeners and sound sources rotate: The Auditory Filehne Illusion. *Proceedings of Meetings on Acoustics*,
- Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4), 1832-1846.
- Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *ACTA Acustica united with Acustica*, 91(3), 409-420.
- Zohary, E., & Sittig, A. C. (1993). Mechanisms of velocity constancy. *Vision Research*, 33(17), 2467-2478.