UNIVERSITY OF GENOA

PHD PROGRAM IN BIOENGINEERING AND ROBOTICS

Perceptual compasses: spatial navigation in multisensory environments

by

Silvia Zanchi

Thesis submitted for the degree of *Doctor of Philosophy* (35° cycle)

June 2023

Dr. Monica Gori	Supervisor
Prof. Giulio Sandini	Supervisor
Dr. Luigi F. Cuturi	Co-supervisor
Prof. Paolo Massobrio	Head of the PhD program

Thesis Jury:

Prof. Laurence R. Harris, *York University*	External examiner
Prof. Paul MacNeilage, *University of Nevada*	External examiner
Prof. Silvio P. Sabatini, *University of Genoa*	Internal examiner

Dibris

Department of Informatics, Bioengineering, Robotics and Systems Engineering

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

<div align="right">

Silvia Zanchi

June 2023

</div>

Abstract

Moving through space is a crucial activity in daily human life. The main objective of my Ph.D. project consisted of investigating how people exploit the multisensory sources of information available (vestibular, visual, auditory) to efficiently navigate. Specifically, my Ph.D. aimed at i) examining the multisensory integration mechanisms underlying spatial navigation; ii) establishing the crucial role of vestibular signals in spatial encoding and processing, and its interaction with environmental landmarks; iii) providing the neuroscientific basis to develop tailored assessment protocols and rehabilitation procedures to enhance orientation and mobility based on the integration of different sensory modalities, especially addressed to improve the compromised navigational performance of visually impaired (VI) people.

To achieve these aims, we conducted behavioral experiments on adult participants, including psychophysics procedures, galvanic stimulation, and modeling. In particular, the experiments involved active spatial navigation tasks with audio-visual landmarks and self-motion discrimination tasks with and without acoustic landmarks using a motion platform (Rotational-Translational Chair) and an acoustic virtual reality tool. Besides, we applied Galvanic Vestibular Stimulation to directly modulate signals coming from the vestibular system during behavioral tasks that involved interaction with audio-visual landmarks. In addition, when appropriate, we compared the obtained results with predictions coming from the Maximum Likelihood Estimation model, to verify the potential optimal integration between the available multisensory cues.

i) Results on multisensory navigation showed a sub-group of integrators and another of non-integrators, revealing inter-individual differences in audio-visual processing while moving through the environment. Finding these idiosyncrasies in a homogeneous sample of adults emphasizes the role of individual perceptual characteristics in multisensory perception, highlighting how important it is to plan tailored rehabilitation protocols considering each individual's perceptual preferences and experiences. ii) We also found a robust inherent overestimation bias when estimating passive self-motion stimuli. This finding shed new light on how our brain processes and elaborates the available cues building a more functional representation of the world. We also demonstrated a novel impact of the vestibular signals on

the encoding of visual environmental cues without actual self-motion information. The role that vestibular inputs play in visual cues perception, and space encoding has multiple consequences on humans' ability to functionally navigate in space and interact with environmental objects, especially when vestibular signals are impaired due to intrinsic (vestibular disorders) or environmental conditions (altered gravity, e.g. spaceflight missions). Finally, iii) the combination of the Rotational-Translational Chair and the acoustic virtual reality tool revealed a slight improvement in self-motion perception for VI people when exploiting acoustic cues. This approach shows to be a successful technique for evaluating audio-vestibular perception and improving spatial representation abilities of VI people, providing the basis to develop new rehabilitation procedures focused on multisensory perception.

Overall, the findings resulting from my Ph.D. project broaden the scientific knowledge about spatial navigation in multisensory environments, yielding new insights into the exploration of the brain mechanisms associated with mobility, orientation, and locomotion abilities.

Key-words: locomotion, vestibular system, landmarks, multisensory, visual impairment, idiosyncrasies

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Unless you are Christopher Columbus, orient and navigate poorly cannot lead you to anything good. In daily life, efficient spatial navigation strategies are indeed crucial for survival in both animals and humans. Spatial navigation is defined as the ability of planning and following a path, in order to reach or avoid specific locations or objects in the environment. Critically, we live in a redundant world in which multiple cues provide congruent spatial information useful to effectively explore the environment. We can rely on inertial signals coming from our proprioceptive and vestibular systems, that detect respectively the location and movements of our body and the acceleration of our head in space. To process spatial information, we strongly rely also on the visual system: visual cues provide highly accurate and precise information about spatial relationship of the elements in the environment (Alais and Burr, 2004). Although with less accuracy, also auditory system helps us detect environmental obstacles or objects, providing their distance and location, especially in situations where the visual information is not reliable (e.g., at night or in presence of fog).

The general aim of this doctoral thesis is to deepen into how multiple sources of information determine spatial navigation and orientation processes, both in sighted and visually impaired people. After presenting the theoretical background in Chapter 1, Chapter 2 of the thesis explores the multisensory integration between acoustic and visual landmarks during an inferential navigation task. Chapter 3 of the thesis explains the role of vestibular signals in self-motion perception and how they interact with virtually spatialized acoustic landmarks in both sighted and visually impaired populations. Subsequently, Chapter 4 of the thesis extends the knowledge about how artificial vestibular stimulation can help understand the contribution of vestibular signals to the interaction with environmental cues during spatial tasks. Chapter 5 of the thesis discusses the presented findings and highlights their novelty and their scientific and clinical implications.

1.1 Multiple sensory cues in spatial navigation

1.1.1 Navigating with sensory landmarks

Critical to achieving navigational aims are the environmental points of reference that are available in the outer world, called landmarks. Every element in the environment can act as a landmark if sufficiently salient for orientation. Usually, salience is determined by color, size, location, and stability (Chan et al., 2012). Generally, the more informative object's features are, the more likely it is that it will be used as a landmark. Landmarks play different roles in enhancing navigation. At its simplest, they are used as beacons (Chan et al., 2012), i.e. an object that indicates or is a target location (e.g., The Lighthouse of Genoa or the De Ferrari fountain in Genoa). In such a case, they provide an egocentric representation of the world, namely a representation based on the position of self in space: navigators can simply estimate the distance between their own position and the beacon's location and then reach it (see left panel of Figure 1.1). According to the literature, the perception of egocentric (self-centered) distance is quite accurate (Loomis et al., 1996). This kind of navigation gives rise to the "route" space representation. Traditionally, route navigation is investigated with tasks requiring either to reach a target by walking, to point at a target, or to ask to verbally estimate its distance (Kalia et al., 2013; Loomis et al., 1993, 1996).

On a more complex level, landmarks allow navigators to build an allocentric representation of space, i.e., based on the spatial relationship existing among the elements that compose the environment regardless of the navigator's position, and this gives rise to the so-called "survey" representation. This spatial representation requires a metric construction of space and inferential processes, thus building a map. To properly investigate the development of this ability, researchers usually attempt to use tasks in which self-motion and dynamic visual cues (i.e., optic flow) are excluded or made unreliable. These tasks require navigators to spatially orient from a viewpoint totally different from their initial position, thus relying only on environmental cues to reorient (Negen et al., 2018). For example, imagine standing in the center of a square: on the right, you can see a parking lot; on the left, you see a monument; and in front of you, there is a fountain (see right panel of Figure 1.1). After learning this spatial configuration, imagine being teleported on the right of the parking lot. Without self-motion information or optic flow, the only way to orient yourself would be to process and remember the spatial relationship between the landmarks. In this case, the orientation performance can be attributed to pure allocentric reasoning because navigators have been prevented from being aided by previously learned self-motion information (coming either from proprioceptive, vestibular, or visual cues). To achieve this aim in experimental setups,
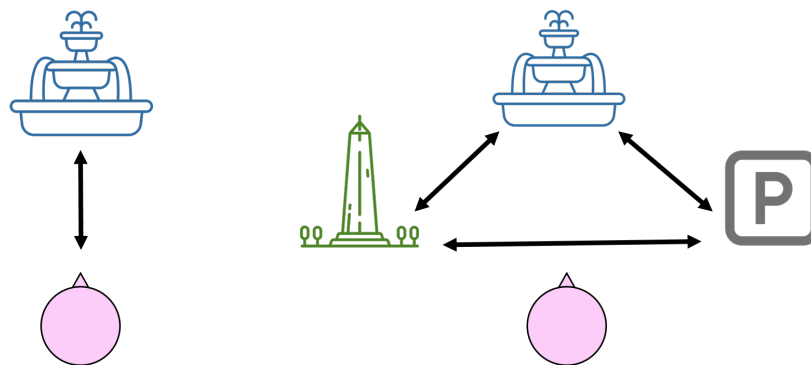
Figure 1.1 On the left, an example of egocentric spatial perspective. What matters is the relationship between the observer and the landmark (the fountain). On the right, an example of allocentric spatial perspective. Here, the observer must represent the spatial relationship between the landmarks available (the fountain, the parking lot and the monument), regardless of their position. *The landmarks icons are downloaded from thenounproject.com.*

researchers commonly use virtual reality or disorientation of participants before responses to disrupt the potential collected self-motion information.

So far, I described landmarks' role in helping spatial navigation according to which kind of spatial information they provide. I would bet that the reader imagined the landmarks as *visual* points of reference: landmarks are indeed mainly visual stimuli. The visual system conveys the most accurate information in spatial tasks due to its high spatial acuity (Alais and Burr, 2004). Besides, visual landmarks are often associated with sounds (e.g., the De Ferrari fountain I mentioned before is both a visual and an auditory cue, thanks to the sound of the gushing water in it). Even in the absence of vision, it has been shown that spatialized auditory cues can help improve orientation and navigation, both in real environments (Jetzschke et al., 2017; Nardi et al., 2020; Riecke, 2016; Shayman et al., 2020) and in acoustic virtual reality (Viaud-Delmon and Warusfel, 2014). For example, it is possible to explore an environment and successfully find a hidden target object by orienting exclusively with auditory landmarks (Viaud-Delmon and Warusfel, 2014). Interestingly, despite lower spatial acuity, auditory cues have few advantages relative to visual ones: first, they provide greater information regarding far space; second, they allow the navigator to identify their location even when not directly facing the source of the sound, e.g., when the navigator is turned or there is an occluding obstacle. Auditory cues are also essential for visually impaired people, who need to rely on non-visual points of reference available in the environment to build a complete representation of space. Spatialized auditory cues can also lead to self-motion illusions as well, especially when they consist of semantic sounds (Riecke, 2016), confirming that they convey spatial information.

In summary, landmarks play a crucial role in spatial orientation and navigation, providing navigators with greatly valuable information about spatial relationships of the environment's components. Critically, landmarks as spatial cues alone are evidence of how much our world is immersed in multisensory perception.

1.1.2   The contribution of the vestibular system to spatial information

In situations in which external signals, e.g. landmarks, are not available or not reliable, it is crucial to use an orientation and navigation strategy based on self-motion perception, called path integration. Path integration is defined as the ability to update own spatial position and movements from a starting point by taking into account internal cues, such as acceleration, direction, turns, and gravity, regardless of external cues in the environment (Loomis, 1999). This navigation strategy has been found both in animals and in humans (Etienne and Jeffery, 2004).

When we move, we continuously collect self-motion information coming from the vestibular system, an elaborate set of sensory transducer organs in the inner ears that is crucial for eye movement, movement coordination, postural control, balance, and orientation. Specifically, it includes three orthogonal semicircular canals (horizontal, anterior, and posterior) that constantly detect angular acceleration of the head in three-dimensional space, around the yaw, roll, and pitch planes, and two otolith organs (the saccule and utricle) that detect translational acceleration, indicating the orientation of the head relative to gravity (graviception) (Figure 1.2).

Crucially, differently from the other sensory systems, the vestibular system is morphologically developed and functioning before birth (Ronca et al., 2008). In addition, it can never be "turned off", constantly providing spatial information about the position and movement of our head during all one's actions and interactions with the external environment. The main role of vestibular signals is to precisely quantify the ability to control balance and movements (Karmali et al., 2021). In relation to this, it has been shown that after 40 years of life the ability to maintain balance decreases and the chance to fall increases and this is associated with less precise vestibular thresholds (Karmali et al., 2021), i.e. sensitivity in detecting or discriminating movements relying only on vestibular signaling. Moreover, vestibular signals are crucial to low-level automatic responses, such as the vestibulo-ocular reflex, which allows for stabilizing the gaze during head movements and thus moving efficiently.

Notably, vestibular signals do not project to a unimodal primary sensory cortex. Signals originating from the inner ears reach the vestibular nerve and travel to the vestibular nuclear
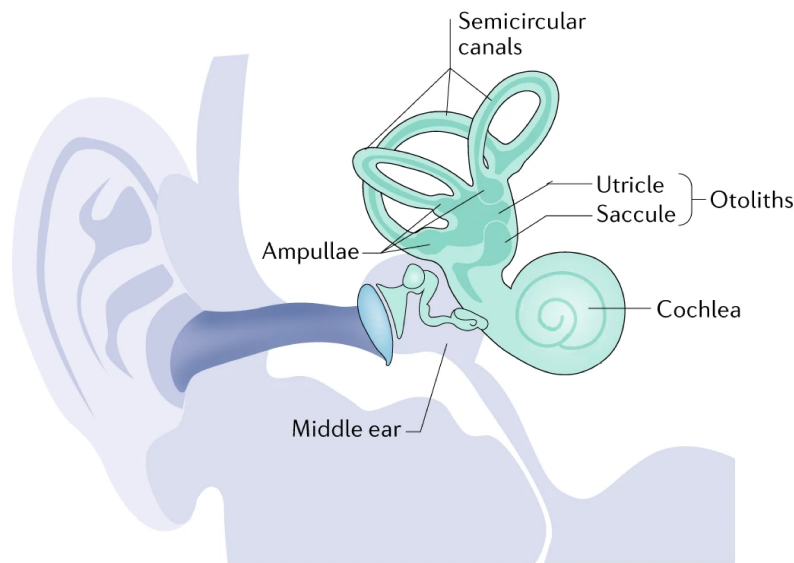
Figure 1.2 The vestibular periphery in the inner ear includes five organs: the three semicircular canals, roughly orthogonal to each other, and the two otoliths (the utricle and the saccule). *This figure has been adapted from Cullen (2019), Nat Rev Neurosci 20, 346–363, reproduced with permission from Springer Nature.*

complex in the brainstem and the cerebellum. Afterward, they reach a distributed multisensory cortical network, involving the temporoparietal junction (TPJ), the posterior and anterior insula, the somatosensory cortex, the posterior parietal cortex, and lateral and medial frontal cortices (Lopez and Blanke, 2011). These areas are defined as the "vestibular cortex" (Guldin and Grüsser, 1998) and are highly multisensory, integrating information coming from visual, auditory, haptic, and proprioceptive systems (Angelaki and Cullen, 2008).

The vestibular system plays an essential role in self-motion perception, the perceptual basis to accomplish efficient spatial navigation. Self-motion perception is defined as the ability to perceive the movements of own body and to discriminate between different duration, directions, and types of motions (i.e., rotations vs translations). Vestibular inputs provide every movement of our head in three-dimensional space with high precision and signal heading direction. Heading is another crucial activity that efficiently allows navigating in space, and corresponds to detecting and maintaining a specific direction of movement. Being accurate in determining heading direction entails effectively achieving our navigational goals. However, previous findings show that the heading perception based on the sole vestibular system is not flawless (Crane, 2012a; Cuturi, 2022; Cuturi and MacNeilage, 2013). It was indeed found that vestibular cues lead to an overestimation of the deviation from the most common straight-ahead direction. This apparently inaccurate signaling is actually

commonly interpreted as the result of how vestibular cues are processed to build a functional representation of the environment (Crane, 2012a; Cuturi, 2022; Cuturi and MacNeilage, 2013), in which our sensory system makes us highly sensitive to directions that are perceived as uncommon to optimally preserve the most useful ones. Thus, the study of the vestibular system provides crucial insights into how we optimize spatial processing and perceive the world in which we navigate.

Critically, recent research is showing that the vestibular system is also involved in a large amount of higher cognitive functions: this is actually unsurprising considering the complex cortical network involved in vestibular processing. For instance, previous animal research showed that vestibular lesions disrupt the ability to form an allocentric cognitive map and the ability to use this representation to effectively achieve navigational goals (Dallal et al., 2015). Similarly, patients with vestibular loss showed worse performance in using ego-centered cues to orient themselves, demonstrating that impaired vestibular signals lead to compromised performances in efficient goal-oriented movements (Xie et al., 2017). In a recent study, it has been shown that artificial stimulation on the vestibular nerve (galvanic vestibular stimulation, GVS) could influence visuospatial memory performance, modulating the ability to remember the position of target objects relative to landmarks virtually located in an environment (Hilliard et al., 2019). Moreover, GVS showed to modulate the representation of horizontal space, determining a spatial bias in the specific direction of the self-motion sensation induced by the stimulation (Ferrè et al., 2013).

Overall, this literature supports the pivotal role that the vestibular system plays in spatial processing and cognition, likely allowing for effective spatial navigation and orientation in our complex environment.

1.1.3 Multisensory interaction during self-motion and locomotion

While we move, we perceive our surrounding environment using all of our sensory systems simultaneously, building a multisensory representation of space. Conveniently, this multisensory spatial representation helps us navigate the environment and interact with the external world. Even before actually moving through the surroundings, visual, proprioceptive, vestibular, somatosensory cues offer pivotal information to perceive the orientation of our body relative to gravity, specifying the relationship between the egocentric (i.e., based on the body) and the external reference frameworks (Harris, 2022).

As briefly mentioned in the previous Section (1.1.2), at the basis of efficient spatial navigation, there is self-motion perception. In everyday life, self-motion perception involves

the integration of sensory signals coming from the visual, auditory, proprioceptive, and vestibular systems. Among all possible pairs of sensory cues, several studies highlight the crucial interaction between visual and vestibular spatial cues. As mentioned earlier (in Section 1.1.1), the visual system is extremely accurate in providing spatial information. During navigation, an optic flow is generated from the relation between the observer and the retinal information generated by their self-motion. Optic flow alone can trigger self-motion sensations, demonstrating the high relevance of visual information in self-motion perception. In ecological situations, we usually take advantage of both visual and inertial cues. It has been shown how primates improved their ability to discriminate the direction of heading motions when provided with additional congruent visual optic flow (Fetsch et al., 2009; Gu et al., 2008). The same results were found in humans (Butler et al., 2010; Kaliuzhna et al., 2018; Prsa et al., 2012).

Similar findings are found in measuring balance ability. Considering balance as a prerequisite for locomotion, the primary goal of multisensory navigation is exploiting multiple cues to maintain balance (Nardini & Cowie, 2012). Since infancy, the strict coupling between visual information and proprioceptive and vestibular cues is crucial to obtain an efficient balance. Yet, it is essential to notice that not only visual cues have an impact on balance. Stevens et al. (2017) demonstrated that external auditory signals could reduce the postural sway of a standing subject in the dark. Moreover, auditory cues added to vestibular signals can increase precision in discriminating left vs. right passive movements, suggesting its importance in tasks requiring self-orientation (Shayman et al., 2020) and supporting previous findings revealing the mechanisms underlying the contribution of hearing to balance (Karim et al., 2018; Lubetzky et al., 2020; Seiwerth et al., 2018). A massive amount of previous research showed that not only do we successfully exploit the interaction between different sensory cues, but we are also able to integrate them in an optimal fashion, improving precision in performance if compared with using a single modality, according to the Bayesian Maximum Likelihood Estimation (MLE) model (Ernst and Banks, 2002). Specifically, the MLE model claims that multiple cues are optimally integrated when both are perceived as equally reliable. This ability has been demonstrated for different sensory systems: haptics and vision (Ernst and Banks, 2002), vision and hearing (Alais and Burr, 2004; Gori et al., 2012), self-motion (vestibular and proprioceptive) and vision (Chen et al., 2017; Nardini et al., 2008; Petrini et al., 2016; Sjolund et al., 2018). The latter combination is the most investigated in spatial navigation. Prior research has extensively shown that the integration between visual landmarks and self-motion information based on vestibular and proprioceptive

cues leads to improved navigation performance in adult navigators (Bates and Wolbers, 2014; Butler et al., 2010; Chen et al., 2017; Frissen et al., 2011; Nardini et al., 2008).

So far, I described situations in which cues originating from different sensory sources are available to the navigator at the same time. However, this is not always the case. In spatial context, another sensory phenomenon that demonstrates the interaction between multiple sensory signals consists of cross-modal aftereffects. Cross-modal aftereffects occur when the adaptation to a stimulus in one modality triggers a subsequent effect in a different sensory modality. In this sense, they have the potential to shed light on multisensory interactions and the underlying neural mechanisms (Greenlee et al., 2016). For example, visual-auditory interactions are shown to occur during self-motion perception. Previous studies already demonstrated that adapting to the motion of external visual objects led to an auditory aftereffect, showing a strong interaction between visual and auditory cues in spatial perception (Ehrenstein and Reinhardt-Rutland, 1996; Kitagawa and Ichihara, 2002). However, recent research extended previous results showing how spatial localization ability of auditory stimuli can be also modulated by visually induced self-motion sensations (Honda et al., 2022). Moreover, Cuturi and MacNeilage (2014) showed that a long exposition (15 seconds) to an optic flow altered non-visual self-motion perception, causing aftereffects in estimating heading direction while on a motion platform in darkness. In this sense, crossmodal aftereffects provide several insights about how cortical representations of spatial stimuli may rely more on shared processes than on the involved sensory modalities themselves (Konkle and Moore, 2009).

Overall, in this Section, it was highlighted how we are constantly immersed in a multisensory environment and how different sensory cues available simultaneously or in subsequent moments communicate to make us accomplish efficient and effective spatial orientation and navigation in daily life.

1.1.3.1   When one sense is missing: navigation of visually impaired people

Earlier in the chapter, we explored both the unisensory contribution to spatial perception and how multiple cues are integrated together to improve spatial navigation. What happens when one sense is missing, for example, in visual impairments? Since vision plays a critical role in space representation, asking which are the consequences of its systematic absence is a key question (Ricciardi et al., 2020). Yet, the scientific literature reports contrasting findings about the spatial competencies of visually impaired people. As we mentioned earlier (in Section 1.1.1), navigation can be categorized into route navigation, which is based on the navigator perspective, and survey navigation, which is based on a map-like representation of

the environment and on taking into account the spatial relationship between the objects in the surroundings. Literature shows that blind people usually accomplish route navigation without difficulties (Thinus-blanc and Gaunet, 1997). Path integration, as we defined it before, is also well-preserved in people with visual disabilities. To corroborate these findings, also Seemungal and colleagues (2007) showed that early blindness does not impair the ability to retrace previously experienced paths based on vestibular and proprioceptive cues. Considering past studies, it seems then that individuals with visual impairment show an efficient representation of space based on egocentric coordinates.

Blind individuals might compensate for their lacking of vision exploiting the remaining auditory sense to localize themselves and external objects. In this sense, auditory information is crucial to people with visual disabilities, allowing them to access the representation of near and far space (Nardi et al., 2020). External auditory landmarks are pivotal in allowing visually impaired people to build a representation of space according to allocentric coordinates and infer novel paths never actively explored (Finocchietti et al., 2017; Gori et al., 2017), which are abilities usually defective, especially for congenitally blind individuals (Pasqualotto et al., 2013; Seemungal et al., 2007). According to the cross-sensory calibration theory (Gori, 2015), in the early stages of life, vision plays the crucial role of calibrating the other senses for spatial processing. This partially explains why people who acquired a visual disability later in life show spatial skills similar to sighted people (Finocchietti et al., 2015), even if it is not always the case.

The scientific advances reached so far have fortunately increased technology that can help improve spatial navigation and orientation in people with visual disabilities. The developed devices and tools are becoming increasingly accessible. Besides, the literature claims that the level of acceptance of the existing complex devices is low (Cuturi et al., 2016). For this reason, it is important to focus on developing simple technological solutions and testing their usability to effectively improve the navigation ability and the safety of visually impaired people. It becomes necessary to evaluate the efficiency of the assessment and rehabilitation procedures for the target users to increase the chances of responding to their real needs and necessities.

1.2   Objectives of the thesis

The goal of my Ph.D. project has been to attain a deeper understanding of how the brain processes mobility and locomotion skills. I focused on exploring the brain mechanisms associated with mobility, orientation, and locomotion abilities in adults, investigating how

people exploit the multisensory sources of information available, including auditory, visual, and vestibular cues. In addition, I investigated how different sensory systems cooperate during navigation to provide useful information about sensory deficits, such as visual impairments.

The main scientific objectives of my Ph.D. project were to:

1. Investigate the multisensory integration mechanisms underlying spatial navigation, taking into account individual characteristics of sensory perception and navigation strategies;

2. Establish the crucial role of vestibular signals in spatial encoding and processing and its interaction with environmental cues.

The main technological objectives of my Ph.D. project were to:

1. Provide the neuroscientific basis to develop tailored rehabilitation procedures to enhance orientation and mobility based on the integration of different sensory modalities, especially addressed to improve the compromised navigational performance of visually impaired people;

2. Develop novel assessment protocols to investigate multisensory perception and navigation abilities in healthy and clinical populations.

Along these lines, the present doctoral thesis aims at illustrating different and various facets of spatial navigation in multisensory environments, specifically considering the auditory, visual, and vestibular systems. Therefore, the projects concerning the abovementioned topics are reported in the following chapters and discussed in the final chapter.

Specifically, the first project aims at investigating the audio-visual landmarks integration during an active spatial navigation task, performed by healthy adults. Thus, behavioral measurements of the re-allocating performance of targets in the navigational environment are compared with the predictions of the Maximum Likelihood Estimation model.

The second project explores the opportunities and implications stemming from the use of a novel combination of two simulation tools: a motion platform, the Rotational-Translational Chair (RT-Chair; Cuturi et al. 2020), and an acoustic virtual reality tool. This work investigates self-motion perception abilities both in visually impaired and sighted adult participants.

The third project focuses on studying the interaction between the vestibular system and visual and acoustic cues as environmental cues in spatial contexts, employing vestibular artificial stimulation to interfere with the natural vestibular signaling.

The projects briefly described here are respectively presented in Chapter 2, Chapter 3, and Chapter 4. Chapter 5 discusses the findings described in previous chapters and their implications.

Chapter 2

Navigating with audio-visual cues

2.1   Introduction

To achieve optimal spatial navigation, we need to exploit multiple internal and external cues that are available in the surroundings. Previous research showed that optimal integration of self-motion and visual cues occurs during spatial navigation (Bates and Wolbers, 2014; Butler et al., 2010; Chen et al., 2017; Nardini et al., 2008). During the exploration of the environment, perceiving and processing multisensory cues should enhance navigational performance, reducing variability and improving accuracy compared to navigation performance based on a single cue (Bates and Wolbers, 2014; Nardini et al., 2008). As predicted by the Bayesian Maximum Likelihood Estimation (MLE) model, the greatest performance improvement occurs when the available cues have the same weight, i.e. reliability (Ernst and Banks, 2002). Conversely, when one cue weighs more than another, the first dominates over the second, modulating the behavior accordingly (Bates and Wolbers, 2014; Ernst and Banks, 2002; Gori et al., 2012; Nardini et al., 2008). Prior research has found clear improvements in precision but one-cue domination for accuracy (Zhao and Warren, 2015b). Other studies have observed behaviors consistent with a cue-competition model, which claims that multiple cues compete with each other, resulting in no difference between multisensory and unisensory situations (Petrini et al., 2016). Even though there is substantial agreement on how multisensory cues lead to enhanced performance, the heterogeneity in the context of spatial navigation suggests the possibility of idiosyncratic factors that may affect the performance of the tasks. A previous work introduced the concept of perceived subjective discrepancy among multiple cues (Cheng et al., 2007): in this sense, if navigators perceive the available sensory cues as greatly discrepant from each other, multisensory integration may not occur. Accordingly, previous research found that the absence of cue integration was associated with

the perceived discrepancy between visual and self-motion cues (Petrini et al., 2016). The contrasting findings in multisensory navigation may be then explained by the differences in methodologies and by the subjective discrepancy between cues perceived by participants.

In daily life, visual landmarks, i.e. visual points of reference in the environment, are often associated with sounds. As mentioned earlier in this thesis, when vision is not reliable or absent, auditory information provides crucial spatial cues for orientation and navigation. While the role of visual information and its interaction with self-motion inputs is well studied, the integration between visual and auditory cues in spatial navigation is not clear. One recent study has found that auditory landmarks can be used to reorient and navigate through space in a homing task (Jetzschke et al., 2017). Auditory cues alone can also help reorientation as they are used as geometric cues, providing navigators with information about the distance between auditory landmarks (Nardi et al., 2020). Besides, auditory inputs in space allow for successfully orienting in an auditory equivalent of the Morris water maze (Viaud-Delmon and Warusfel, 2014), i.e. an experimental paradigm used to assess spatial learning and memory in animals. Considering the abovementioned literature, it is then reasonable to believe that visual and auditory cues may enhance navigation task performance when integrated. Investigating whether external audio-visual information is integrated would aid our understanding of navigation strategies in a real environment. Few studies have examined the processing of audio-visual landmarks during navigation (Gröhn et al., 2005; Karimpur and Hamburger, 2016; Werkhoven et al., 2014). It remains unclear whether the integration between auditory and visual information is achieved according to Bayesian principles, which lead to better performances in the presence of multisensory sources of information.

The study presented in this chapter aimed to investigate whether audio-visual landmarks are optimally integrated and how their potential integration may have an impact on navigation performance. To achieve this aim, participants had to perform an object relocation task in a dark room (Nardini et al., 2008). Their goal was to relocate a target object to its original location, orienting themselves by exploiting the sole spatial information provided by landmarks that could be visual-only (animal-shaped arrays of lights), auditory-only (animal sounds), or audio-visual (animal-shaped lights and animal sounds presented simultaneously). We also ran a conflict condition that presented a spatial discrepancy between the visual and auditory landmarks by displacing the loudspeakers from the visual landmarks in order to assess how participants weighted each cue.

We hypothesized two potential outcomes. First, in case of similar weightings of the auditory and visual inputs, we expect to find optimal multisensory integration (reduced variability) consistently with the prediction of the MLE model (Ernst and Banks, 2002).

Secondly, in the case of different weights associated with the two cues, we expected to observe non-optimal integration, resulting in the predominance of the most weighted cue. To test our hypotheses, we performed two experiments in which we varied the reliability of visual landmarks. Specifically, in Experiment 1, visual landmarks had high reliability, while in Experiment 2, they had lower reliability.

2.2 Experiment 1: Testing optimal integration of audio-visual landmarks

2.2.1 Participants

Twenty-one participants (mean age $\pm$ SD = 29.2 $\pm$ 3.2 years old) took part in Experiment 1. All participants had a normal or corrected-to-normal vision and absent hearing impairments. The study has been approved by the ethics committee of the local health service (Ethical Committee, ASL 3, Genova, Italy), and it was conducted adhering to the Declaration of Helsinki. All the subjects gave informed written consent.

2.2.2 Apparatus and stimuli

The experiment took place in a 300 cm X 500 cm dark real-world environment. The goal of the participants consisted of exploring the area and picking up three bright objects which lay on the floor (Figure 2.1, panel A). Afterward, they had to reposition the object they collected first in its original location by orienting themselves with the available landmarks. Specifically, the visual landmarks consisted of three hand-made Light Emitting Diod (LED) arrays. Each of them was fixed on a black sheet of paper that measured 21 cm X 29.7 cm. The LED arrays were arranged to resemble the shapes of animal faces, specifically a frog, a dog, and a duck (Figure 2.1, panel A). In turn, the paper sheets with the LED arrays were fixed on adjustable height stands. As depicted in panel B of Figure 2.1, from the initial position viewpoint, the central landmark (the frog) was straight ahead at a distance of 400 cm and at a height of 150 cm. The two landmarks on the sides (the dog and the duck) were at 45° to the left and the right of the initial position at a distance of 208 cm and at a height of 50 cm. The "frog" was made of yellow LED lights, while the "dog" and the "duck" were made of white LED lights. In this way, the straight-ahead direction was always well detectable from the initial position due to the landmarks' colors.
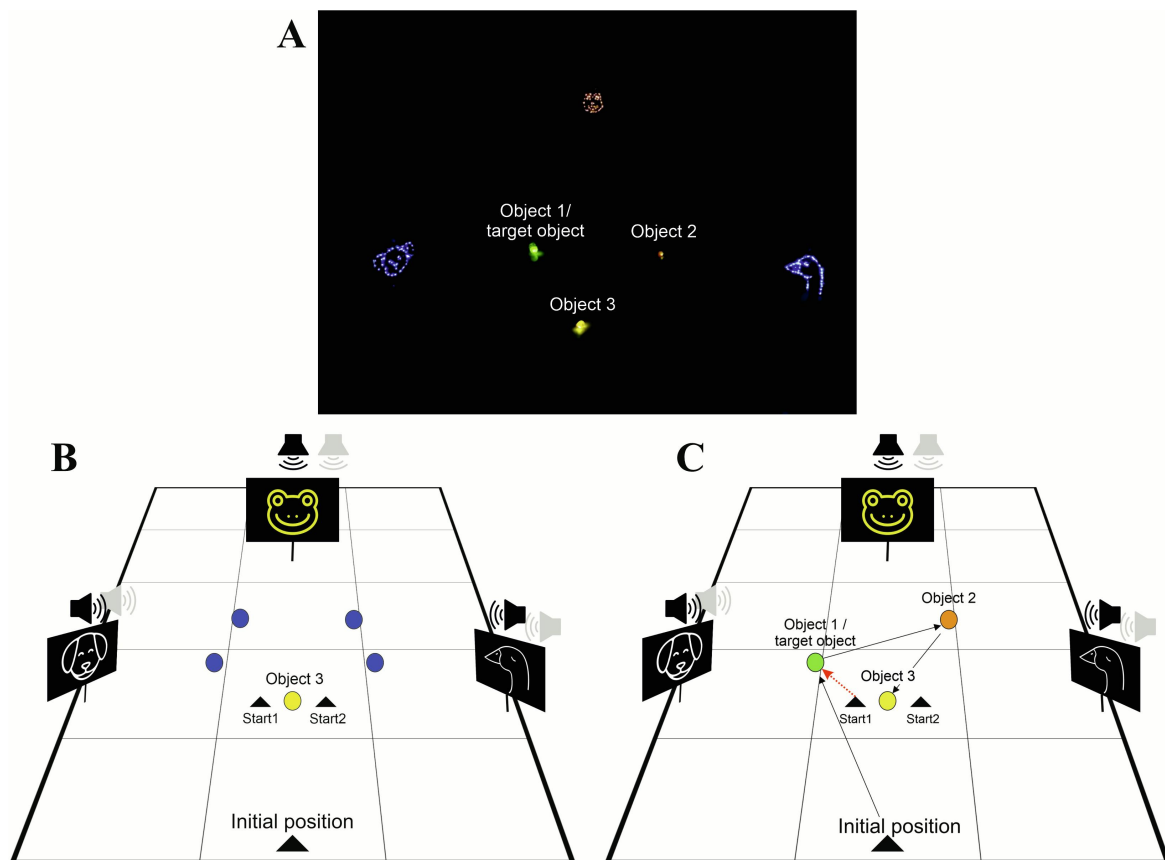
Figure 2.1 Panel A: View of the setup from the initial position: In the dark room, three illuminated landmarks (a dog, a frog, a duck) and three illuminated objects (the green apple—Object 1/target object, the orange tangerine—Object 2, and the yellow pear—Object 3) were visible. Panel B: Experimental setup: the big central triangle indicates the initial position. The circles on the right and the left represent the possible positions of Objects 1 (the target object) and 2, while the central circle represents the position of Object 3. Near Object 3, small triangles represent the two start positions (Start1 and Start2) from which participants attempted to relocate the target object in the response phase. Above each visual landmark, speakers are represented: the darker ones are the aligned auditory landmarks; the lighter ones are the misaligned auditory landmarks, which create a spatial conflict (in the example, right direction of conflict) between auditory and visual information in the conflict condition. Panel C: Example of an experimental trial. From the initial position, participants were instructed to reach the target object, then Objects 2 and 3. Once positioned randomly at one of two starting positions (Start1 in the example), participants needed to infer and travel the route from the start to the target object location (dashed red arrow). *Copyright © 2022 by American Psychological Association. Reproduced with permission. Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). Interindividual differences influence multisensory processing during spatial navigation. Journal of Experimental Psychology: Human Perception and Performance, 48(2), 174.*

The auditory landmarks consisted of three loudspeakers (Stilgut YB202STGD, 6.6 X 6.6 X 6 cm) that were aligned with the three visual landmarks. They were also mounted on the same stands, and oriented toward the exploration environment. The speakers were connected with jack cables to a computer (Dell Latitude 3340 with Intel Core i5- 4200U central processor at 1.60 GHz; 64-bit Windows 10 Enterprise Version 1809) through an external sound card (Xonar U5, Asus); notably, each speaker was connected to a separate channel. Crucially, each auditory landmark played a sound that semantically corresponded to the image represented by the visual landmarks; thus, the central speaker played a croaking frog, the left one played a barking dog, and the right one played a quacking duck. We downloaded the three sounds from a royalty-free sound web archive (https://freesound.org/). We picked animal sounds to make the landmarks distinguishable from one another and semantically congruent to the associated visual cues. Before beginning the experiment, the three loudspeakers were calibrated to ensure they had equal output volume (80 dB measured at a distance of 40 cm). During the condition with the spatial discrepancy between visual and auditory landmarks, we used an additional triplet of loudspeakers. Each of them was shifted by 29 cm to the left or the right of the corresponding central auditory landmark. We chose to use 29-cm shifts after piloting. Eleven participants experienced the left direction of conflict; the other 10 experienced the right direction of conflict. The experimenter controlled via Matlab (R2019b, The MathWorks, United States) when both auditory and visual landmarks would start and end by pressing a wireless mouse key.

The three objects to be picked up consisted of fruit toys made of semitransparent plastic (a green apple—Object 1, an orange tangerine —Object 2, a yellow pear—Object 3). We added small LED lights in the objects to make them glow in the dark (Figure 2.1, panel A). The objects could be placed in different positions. In particular, the positions of Objects 1 and 2 varied across trials, whilst Object 3 was kept in the same position for the whole experiment, at 173 cm straight ahead of the initial position. Object 1 and Object 2 could have been positioned in four different locations: 40° or 30° to the left or the right relative to the position of Object 3, at a distance of 140 or 103 cm. To ensure participants explored both sides of the area, Object 2 was always placed on the opposite side of the room relative to Object 1. Participants performed four trials in each condition. Each of the four possible positions of Object 1 (henceforth, the target object) was presented once in all conditions. Notably, there were two different starting points from which participants attempted to return the objects: 30 cm to the left (Start1) and 30 cm to the right (Start2) of the Object 3 position (Figure 2.1, panel B). The two starts required participants to use only the landmark information to orient themselves and to make the Object 3 position unreliable to relocate the objects successfully.

In addition to being in a dark space, the participants were given covering sunglasses that had increased opacity due to the addition of nylon filters over the lenses. The participants wore the sunglasses before entering the experimental room and for the entire duration of the experiment so that they would see only the switched-on visual landmarks and the bright objects; thus, they could not see the stands, the loudspeakers, the floor, the walls, or the ceiling.

2.2.3   Experimental design and procedure

As mentioned earlier in the Introduction of this chapter, the experiment involved four conditions, according to the typology of landmark available: a) visual-only; b) auditory-only; c) visual and auditory combined; d) combined but with a spatial conflict between the two types of sensory landmarks. We selected an experimental design that allowed us to rule out any potential learning effect or influence among the conditions. Thus, the experiment took place over two distinct sessions, which lasted 45 min each and were programmed at different moments of the day or on different days. Unimodal (a, b) and multimodal (c, d) conditions were completed during different sessions. The order of the sessions was counterbalanced across participants. Specifically, 11 participants experienced the unimodal conditions first and then the multimodal conditions, while the other 10 experienced the opposite order of the sessions. Within each session, the conditions were pseudorandomized across trials.

Two different phases constituted each trial. In the exploration phase, the subjects collected the objects lying on the floor and attempted to remember the target object's location relative to the landmarks' configuration. In the response phase, the participants had to place the objects in the original location of the target object. Between the two phases, a disorientation phase disrupted the use of gathered self-motion information that could have been attained during the response phase. The experimental procedure thus consisted of an inferential spatial navigation task in which participants had to infer new routes based on previously learned spatial relationships among landmarks (see Figure 2.1, panel C). This task allowed us to systematically investigate participants' ability to travel novel paths by basing their responses exclusively on the spatial information provided by the landmarks.

The procedure occurred as follows. After reading the experimental instructions, participants were equipped with sunglasses and guided through the dark environment until they reached a swivel chair that was already placed at the initial position (see Figure 2.1, panel B). Afterward, they were given passive noise-canceling headphones, and the experimenter rotated the chair to make the participants face the wall in front of the central landmark. Once

the experimenter positioned the objects on each trial, the participants were told on which side of the room the target object was located (right or left) to ensure they would clearly recognize it. Then, the experimenter switched on the landmarks and asked the participants to take off the headphones and begin the exploration phase. First, their goal was to pick up a transparent plastic bottle with the target object inside and to remember its original location considering the landmarks' configuration. They then had to pick up Objects 2 and 3 and put them inside the bottle. Afterward, the participants waited for the experimenter, facing back toward the initial position, and the landmarks were switched off. Participants again sat down on the swivel chair, which was now located near the Object 3 position, and they were disorientated for ~ 20 s by being spun around in the chair. During the disorientation, participants closed their eyes and wore headphones to prevent them from detecting any additional external cues. After disorientation, the experimenter positioned the participants (still sitting on the chair) randomly in one of the two starts (Start1 or Start2). Subsequently, they were asked to take off the headphones and open their eyes. Once the landmarks were switched on, the experimenter encouraged the subjects to use all three landmarks to correctly relocate the bottle with all the objects inside to the original target object location (response phase). Except for the typology of the provided landmarks, the procedure was the same for all trials. In the exploration phase of the conflict condition, the loudspeakers that worked as auditory landmarks were spatially aligned with the center of the visual landmarks, as they were in the auditory and combined conditions. Conversely, in the response phase of the conflict condition, sounds were played by the three misaligned loudspeakers to create a spatial conflict between the visual and auditory cues. At the end of the second session, the experimenter asked all the participants whether they noticed anything unusual during the experiment: no participant was explicitly aware of the spatial conflict.

2.2.4   Data analysis

Behavioral analysis. We measured the distance between the participant's response and the correct location of the objects for each trial (in centimeters). The individual trials were filtered to remove errors that were extreme outliers in the distribution of all the errors recorded for that condition. Extreme outliers were defined as values greater than the third quartile plus 3 times the interquartile range. Seven trials were removed as extreme outliers; one more trial was removed due to a technical issue (2.19%; one from the auditory condition, two from the visual condition, one from the combined condition, and three from the conflict condition). For the variability measures, no responses met the outlier definition. In addition,

we removed one participant from the analyses because of multiple outlying responses (n = 8 out of 16 trials) that prevented the calculation of the variables. Therefore, 20 participants' data constituted the final dataset for the analyses.

For each participant and each condition, we computed the mean response error, i.e., the constant error (CE). The variability of errors was taken as an index of precision (the lower the variability, the higher the precision), which was measured as the standard deviation (SD) of the responses. Shapiro-Wilk tests were used to verify the normality distribution of the variables in each condition. We then performed a repeated-measures ANOVA on each dependent variable, considering the condition as the within factor. Generalized eta squared (η_G^2) was calculated as effect size. When appropriate, we performed post-hoc comparisons with pairwise t-tests, calculating Cohen's d as the effect size for paired comparisons with the cohen.d function in the eff-size package (Torchiano, 2016) in RStudio (version 3.6.2, 2019). In case of violation of the assumption of normality, we performed non-parametric Friedman's test (Kendall's W value was calculated as effect size) and pairwise Wilcoxon signed-rank tests as post-hoc comparisons. The probabilities were considered significant when lower than .05 after the Bonferroni correction.

Modeling analysis and Bayesian predictions. The four experimental conditions we used determined whether people could integrate multisensory cues during a navigation task (Sjolund, 2016). Two unimodal conditions (auditory and visual conditions) measured the variability associated with each sensory cue. The combined condition measured the actual variability when in the presence of audio-visual cues. Finally, the conflict condition was administered to reveal participants' relative reliance on each of the two cues (Nardini et al., 2008). The undetectable spatial shift in the conflict condition created a mismatch between the correct target location based on the auditory cue and the correct target location based on the visual cue. The degree to which participants relied on each cue was given by the relative proximity of their responses to each of the single-cue-based target locations (Bates and Wolbers, 2014; Nardini et al., 2008). In the conflict condition, the relative proximity to the location given by the auditory cues (*rproxA*) was calculated as follows:

$$rproxA = (1/d_A)/(1/d_V + 1/d_A) = d_V/(d_A + d_V) \qquad (2.1)$$

in which *dA* and *dV* are the predicted response distances based on (shifted) auditory-only and visual-only cues, respectively (Bates and Wolbers, 2014; Nardini et al., 2008; Sjolund, 2014). We compared two different models of cue combination: an integration model and an alternation model. When multiple cues are available, the integration model predicts that

the variances of the two cues are integrated with a weighted average. For this model, the predicted variance σ_{V+A}^2 is calculated as:

$$\sigma_{V+A}^2 = w_V^2 \sigma_V^2 + w_A^2 \sigma_A^2 \qquad (2.2)$$

where σ_V^2 and σ_A^2 are the response variances in the visual and auditory conditions while w_V and w_A are the empirical weights given to the visual and auditory cues, respectively, and the sum to unity. In this model, if the participants combined the sensory cues using a weighted average, the relative proximities in Equation (2.1) would correspond to the empirical weights given to the cues. In contrast, the alternation model predicts that the participants do not integrate cues, but they alternate between them. In this case, since the cues would be used as separate sources of information, the variance would increase (Nardini et al., 2008). In this model, the predicted variance σ_{V+A}^2 is calculated as follows:

$$\sigma_{V+A}^2 = p_V(\mu_V^2 + \sigma_V^2) + p_A(\mu_A^2 + \sigma_A^2) - (p_V \mu^2 + p_A \mu_A)^2 \qquad (2.3)$$

in which p_V and p_A are the probabilities of following either cue and sum to unity; again, σ_V^2 and σ_A^2 are the response variances in the visual and auditory conditions. In this model, $\mu_V = 0$ and $\mu_A = 29$ because the auditory landmarks in the conflict condition were shifted by 29 cm to the left or right, relative to the center of the corresponding visual landmarks. To compare the observed variability with the predictions from the models, we performed paired t-tests. In addition, to evaluate the strength of the evidence in favor of the null versus the alternative hypothesis, we calculated Bayes factors (BF_{01}; Rouder et al. 2009) using the *ttestBF* function from the BayesFactor package (Morey and Rouder, 2018) in RStudio. BF_{01} shows evidence in favor of the null hypothesis, and it is directly interpretable as an odds ratio; for example, a BF_{01} greater than 3 means that the null hypothesis is more than 3 times as likely as the alternative hypothesis. In our cases, this would show evidence in favor of the null hypothesis that the observed data do not differ from the predictions of the model. We adopted a scale r on the effect size of .707. In line with a previous study (Chen et al., 2017), in case the p-value showed to be greater than .05 and the BF_{01} was greater than 3, we stated that cues are integrated optimally. In case the p-value showed to be greater than .05 and BF_{01} was between 1 and 3, which still favors the null hypothesis, cue integration was considered nearly optimal. BF_{01} lower than 1 was considered in favor of the alternative hypothesis. If participants switched from one cue to another during the conflict trials, the relative proximities (Equation 2.1) would correspond to the probabilities p_V and p_A. According to the MLE prediction,

when there are multiple sensory cues, the multisensory variance is smaller when the cues are weighted according to their reliabilities, which are inversely proportional to their variances (Ernst and Banks, 2002). We computed the predicted optimal weighting for the auditory landmarks, w_A, and the visual landmarks, w_V, as follows:

$$w_A = (1/\sigma_A^2)/(1/\sigma_V^2 + 1/\sigma_A^2) = \sigma_V^2/\sigma_A^2 + \sigma_V^2 \qquad (2.4)$$

$$w_V = (1/\sigma_V^2)/(1/\sigma_A^2 + 1/\sigma_V^2) = \sigma_A^2/\sigma_V^2 + \sigma_A^2 \qquad (2.5)$$

where σ_A^2 and σ_V^2 are the response variances in the auditory and visual conditions, respectively. To test whether the participants optimally weighted the cues during the task, we performed t-tests and computed Bayes factors to compare the empirical (Equation 2.1) and predicted weights (Equations 2.4 and 2.5). If the optimal weights are given to the cues, the MLE predicts that the variability in the combined condition is reduced optimally compared with the unimodal variabilities. In such a case, the optimal variance in the combined condition predicted by the model would be:

$$\sigma_{V+A}^2 = \sigma_A^2 * \sigma_V^2/(\sigma_A^2 + \sigma_V^2) \qquad (2.6)$$

To test whether the observed variance in the combined condition was statistically comparable to the one predicted, we performed paired t-tests in the case of a normal distribution of data; otherwise, we performed nonparametric Wilcoxon signed-rank tests. In the latter case, the r effect size would be calculated (interpretation: from .1 to .3 = small effect, from .3 to .5 = moderate effect, .5 and greater = large effect). All analyses were conducted using Matlab (R2019b, The MathWorks, United States) and RStudio (version 3.6.2, 2019).

2.2.5   Results

The repeated-measures ANOVA on CE showed there was a main effect of condition, $F(2, 19)$ = 7.532, p = .005, η_G^2 = .17, 95% CI [0, .36]. In particular, as shown in panel A of Figure 2.2, this effect was greater in the auditory condition compared with the visual, $t(19) = 4.11$, p adjusted (adj) = .002, Cohen's d = .89, 95% CI [.37, 1.41], and the combined conditions, $t(19) = 2.74$, p adj = .039, Cohen's d = .86, [.12, 1.60]. The visual and combined CE did not differ, $t(19) = .29$, p adj = 1, Cohen's d = .07, [-.44, .59]. Regarding the variability (Figure 2.2, panel B), the results suggested there was a main effect of condition on the SD

as well, F(2, 19) = 4.835, p = .014, η_G^2 = .15, 95% CI [0, .34]. Specifically, the SD was higher, meaning that there was lower precision in the auditory condition compared with the combined condition, t(19) = 3.7, p adj = .005, Cohen's d = 1.12, [.34, 1.91]; however, it did not significantly differ from the visual condition, t(19) = 1.78, p adj = .275, Cohen's d = .55, [-.12, 1.21]. No difference was found between the visual and combined SD, t(19) = 1.16, p adj = .786, Cohen's d = .41, [.33, 1.14]. These results did not show any improvement when both the auditory and visual cues were available compared with the visual-only condition. In turn, this suggested that the participants did not integrate the cues to accomplish the task. Comparison between observed and predicted combined variance with the Wilcoxon test revealed that the observed variance (95.3 ± 21.8) was not significantly different from the one that had been predicted by the model (MLE prediction = 66.35; V = 121, p = .57, r = .133, 95% CI [-.53, .32], BF_{01} = 2.45). This suggests that there was a near-optimal reduction of variability when both cues were available (Figure 2.2, panel C). In addition, we compared the observed behavior in the conflict condition with the predictions made using the integration and the alternation models (Figure Figure 2.2, panel D). A paired t-test showed that there was a significant difference between the conflict condition and the prediction made using the integration model, t(19) = 2.11, p = .048, Cohen's d = .52, [-.01, 1.05], BF_{01} = .7. Specifically, the observed SD (7.62 ± .7) was even lower than the prediction (9.25 ± .7). In contrast, with consideration of the prediction using the alternation model (16.73 ± 1.3), the observed data showed a statistically significant lower SD, which indicated that there was greater precision, t(19) = 15.228, p < .0001, Cohen's d = 4.03, [2.41, 5.66], BF_{01} < .001. These results showed that the variability in the conflict condition was significantly lower than the alternation prediction, which suggests that the participants did not perceive the cues as two different sources of information, as the variability would have increased when both the audio-visual cues were available. Interestingly, the variability in the conflict condition was lower than the integration prediction. The conflict condition allowed us to verify the weights associated with each unisensory cue. We compared the empirical weights that were assigned to the auditory and visual cues (computed using Equation 2.1 above) and the weights predicted by the optimal integration (Equation 2.4 and 2.5). In the conflict trials, the relative weighting for the auditory and visual landmarks were .42 ± .02 and .58 ± .02, respectively, and they did not statistically differ from the weighting expectations (model predictions: w_A = .38, w_V = .62), t(19) = -.72, p = .48, Cohen's d = .26, 95% CI [-.48, 1], BF_{01} = 3.44. These results are in line with the finding that the optimal variance predicted by the MLE did not differ from the observed one in the combined condition. This shows that the participants used the optimal weights to combine the cues. In addition, we observed that

the two empirical weights given to cues significantly differed from each other, t(19) = 4.68, p < .001, Cohen's d = 2.10, [.48, 3.71]; in particular, the auditory cues were weighted less than the visual ones.

2.2.6   Interim discussion

Generally, we observed results that did not clearly evidence multisensory integration between audio-visual landmarks. On the one hand, the modeling analysis showed that the variance in the combined condition was consistent with the MLE prediction and that the SD in the conflict condition was even lower than the integration model prediction. Along the same lines, the empirical weights associated with each cue were similar to the predictions. On the other hand, the behavioral results revealed comparable performance in visual-only and combined conditions, suggesting no advantage in the combined condition. However, multisensory integration predicts that there would be a reduction of variability, namely an increase in precision when multiple cues are available at the same time. It is plausible that, in this experiment, the response variability in the visual condition already reached a plateau; if it were the case, this would have prevented a multisensory improvement, in line with previous findings (Chen et al., 2017). In addition, the visual cues were weighted significantly more than the auditory cues were, demonstrating to be more reliable. Indeed, the predicted decreased variability in the combined condition occurs when the unimodal cues had the same level of reliability (Fetsch et al., 2009), while the dominance of the most reliable cue occurs when there is a discrepancy between cues' weights (Ernst and Banks, 2002).

For these reasons, we performed a second experiment (Experiment 2), in which we attempted to reduce the visual cues' weights and make them comparable with the auditory ones. To achieve this parity, we relocated all the landmarks farther from the target object locations, since previous studies showed lower reliability for far visual landmarks than closer ones (Chen et al., 2017; Zhao and Warren, 2015b).

2.3   Experiment 2: Reducing visual cue reliability

2.3.1   Participants

Experiment 2 was greatly similar to Experiment 1. The only feature we changed was that we lowered the reliability of the visual landmarks by moving them further from the exploration zone. Twenty-one participants took part in Experiment 2 (mean age $\pm$ SD = 26.4 $\pm$ 4.9 years old), and none of them participated in Experiment 1.
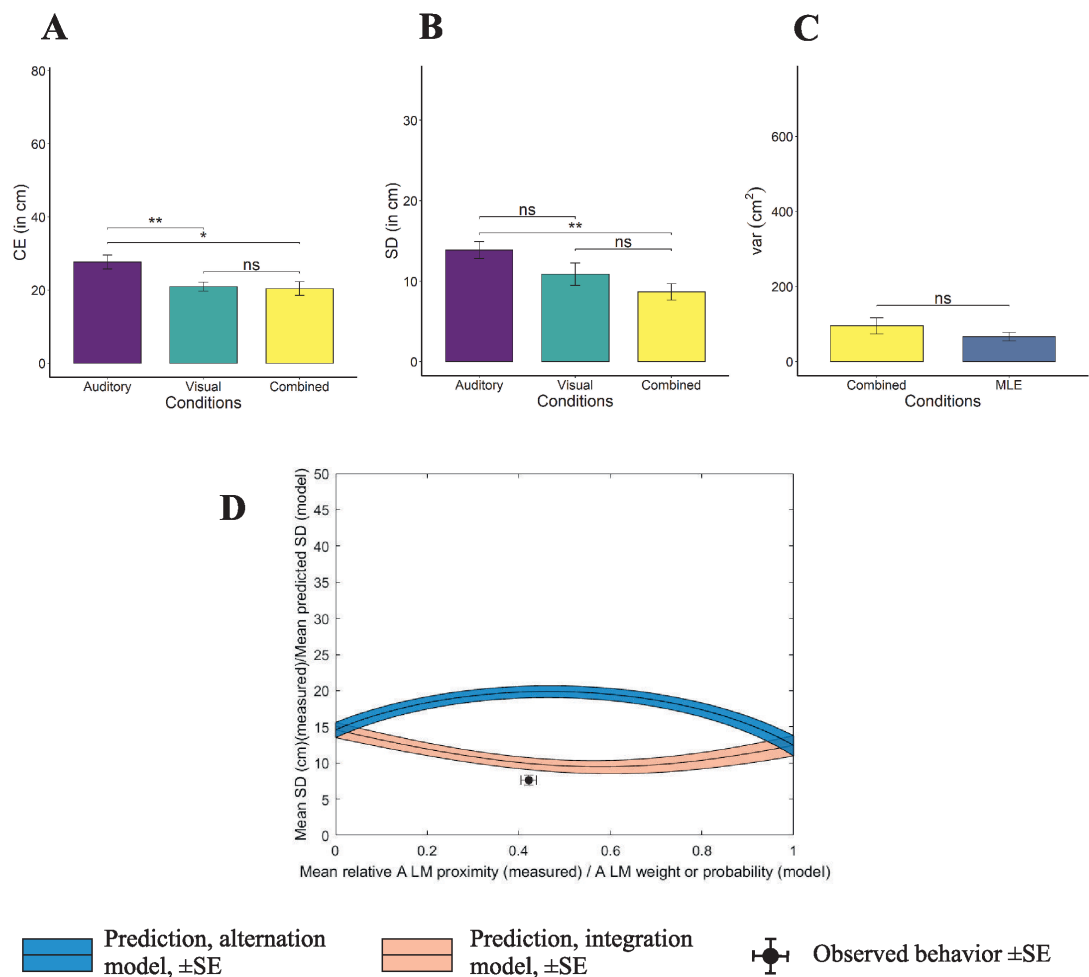
Figure 2.2 Panel A: The constant error (CE) in the auditory condition was significantly higher than in the other two conditions, which did not differ from each other. Panel B: Standard deviation (SD) was higher in the auditory condition if compared with the combined condition, while the other comparisons showed no differences. Panel C: Variance in the combined condition did not significantly differ from maximum likelihood estimation (MLE) prediction. Error bars are standard errors. Panel D: The curves represent the means of functions predicting mean SD from different auditory landmark weights (integration model, orange) or probabilities (alternation model, blue). The x-axes correspond to progressively greater reliance on auditory landmarks from 0 to 1. The points represent the observed mean SD in the conflict condition and mean relative proximities to the locations consistent with auditory landmarks, interpreted as empirical auditory landmarks weights (integration model) or auditory landmarks probabilities (alternation model). A LM = auditory landmarks; SE = standard error; ns = not significant; * = $p < .05$; ** = $p < .01$. *Copyright © 2022 by American Psychological Association. Adapted with permission. Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). Interindividual differences influence multisensory processing during spatial navigation. Journal of Experimental Psychology: Human Perception and Performance, 48(2), 174.*

2.3.2   Apparatus and stimuli

The three visual and auditory landmarks were the same as were used in Experiment 1; however, for this experiment, they were moved farther away relative to their locations in Experiment 1 by 97 cm. As such, they covered a total of 494 cm X 597 cm in the environment to be explored. The position of Object 3 was the same as was used in Experiment 1, while the positions of Objects 1 and 2 were moved forward by 40 cm relative to the initial position in order to be centered with the new landmarks' configuration. Notably, the distances between them remained equal to those in Experiment 1. Since the landmarks were moved farther from the target object, an increase in conflict shift was also needed. Therefore, in the response phase of the conflict condition, the three loudspeakers that were used as landmarks were shifted 36 cm to the left or the right of the visual landmarks. We chose the 36-cm shift as it was proportionally relative to the shift used in the Experiment 1 conflict; as such, it was also unnoticeable. Eleven participants experienced the left direction of conflict; the other 10 experienced the right direction of conflict.

2.3.3   Procedure

The procedure was the same as used in Experiment 1. The order of the sessions was again counterbalanced across participants: 10 participants first experienced the unimodal conditions and then multimodal conditions, while the other 11 experienced the opposite order of the sessions. Like in Experiment 1, none of the participants were aware of the spatial conflict.

2.3.4   Data analysis

As done for Experiment 1, before any calculation of mean CE, individual responses were filtered to remove extreme outliers in the distribution of all errors recorded for that condition. In Experiment 2, none of the trials met the definition of being outliers, and no participant was excluded. One trial from the combined condition was removed (0.3%) because the participant reported having forgotten the target object location, making that response unreliable. Twenty-one participants' data were included in the final analyses. Then, we performed the same behavioral and modeling analyses conducted on Experiment 1 variables for Experiment 2. For the computation of alternation model variables, the considered μ_A was 36 cm, according to the newly selected conflict shift.

2.3.5 Results

The repeated-measures ANOVA showed a main effect of the factor condition on CE, $F(2, 20) = 18.373$, $p < .0001$, $\eta_G^2 = .23$, 95% CI [.03, .43]. Specifically, as depicted in panel A of Figure 2.3, CE was greater in the auditory condition when compared with the visual, $t(20) = 3.92$, p adj = .003, Cohen's d = .88, 95% CI [.35, 1.42], and combined conditions, $t(20) = 5.34$, p adj < .0001, Cohen's d = 1.11, [.58, 1.64]; in turn, the last two conditions were statistically similar, $t(20) = 1.71$, p adj = .309, Cohen's d = .31, [-.07, .7]. For this experiment, the accuracy showed the same pattern of results as in Experiment 1. In terms of variability (Figure 2.3, panel B), the nonparametric Friedman's test suggested that the type of condition had a main effect on the SD, $\chi^2(2) = 7.52$, p = .023, Kendall's W value = .18, 95% CI [.06, .41]. In particular, auditory SD was significantly greater than combined SD (V= 204, p adj = .004, r = .67, [.36, .86]), while it did not differ from the visual one (V= 147, p adj = .864, r = .24, [-.21, .61]). No difference was found between visual and combined SD (V= 156, p adj = .504, r = .31, [-.15, .68]). As shown by the Wilcoxon test, the variance in the combined condition (347.7 $\pm$ 89.3) was statistically similar to the one predicted by the MLE model (MLE prediction = 218.8; V = 145, p = .32, r = -.225, 95% CI [-.59, .24], BF_{01} = 1.78). This suggests a near-optimal reduction of variability when both cues were available (Figure 2.3, panel C). To compare the model predictions and the behavior in the conflict condition (Figure 2.3, panel D), a paired t-test was implemented; it showed that there was no difference between the observed data and the prediction made by the integration model, $t(20) = -1.12$, p = .274, Cohen's d = -.23, [-.65, .19], BF_{01} = 2.52. When compared with the alternation model, a paired t-test revealed that the observer data showed a significantly lower variability, which reflected higher precision, $t(20) = 6.36$, p < .0001, Cohen's d = 1.34, [.75, 1.92], BF_{01} < .001, as also occurred in Experiment 1. Thus, as we did in Experiment 1, we tested the empirical weights assigned to the auditory and visual cues (Equation 2.1) and the weights predicted by the optimal integration (Equation 2.4 and 2.5). Here, in the conflict condition, the auditory and visual cue weights were .48 $\pm$ .02 and .52 $\pm$ .02. This weighting did not statistically differ from the expectations (model prediction: w_A = .43, w_V = .57), $t(20) = -.85$, p = .41, Cohen's d = -.28, [-.40, .95], BF_{01} = 3.19. Contrary to what we observed in the weighting analysis for Experiment 1, the empirical weights given to the auditory and visual cues did not differ from each other, $t(20) = -1.04$, p = .31, Cohen's d = .45, [-.47, 1.38]. This suggests that the information provided by auditory and visual landmarks did not differ in reliability.

Experiment 2
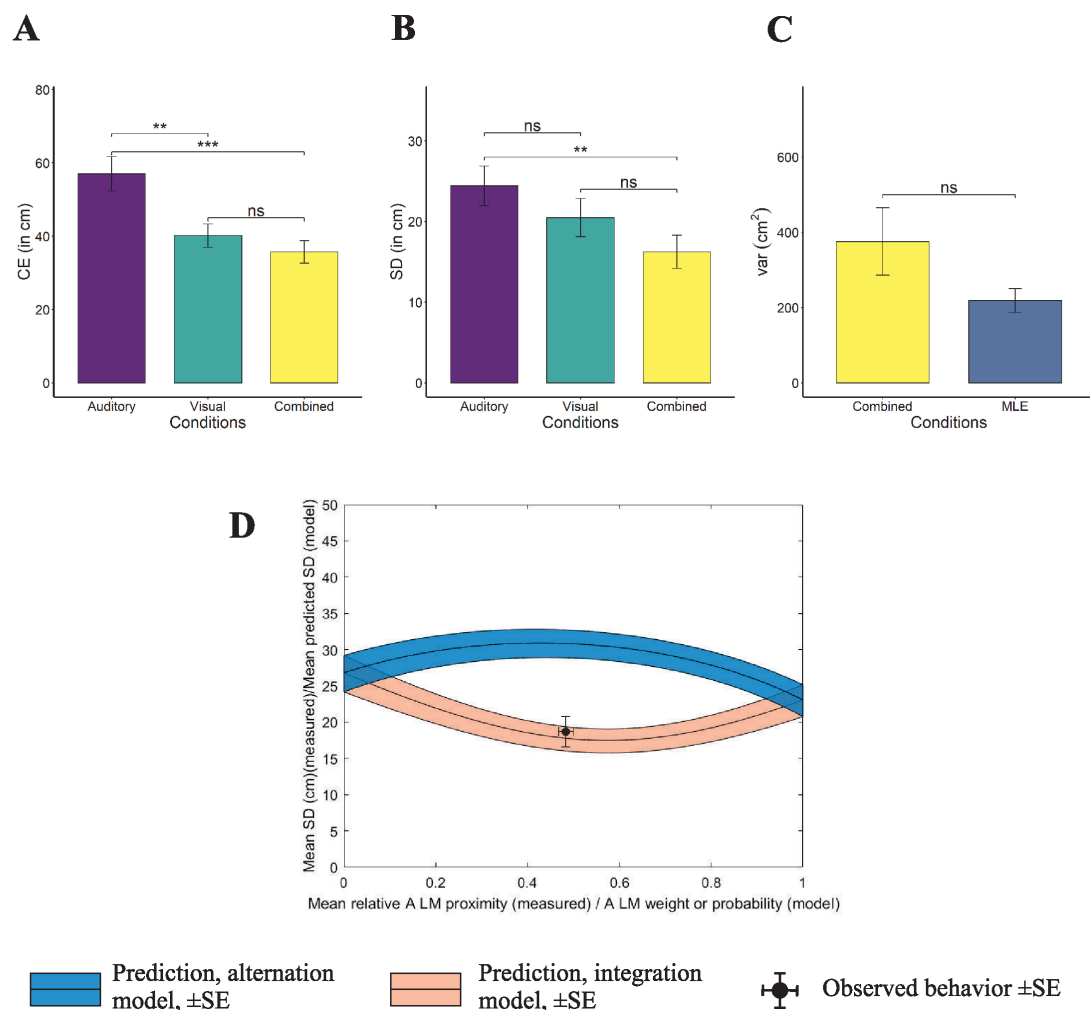
Figure 2.3 Panel A: Similarly to what was observed in Experiment 1, CE in the auditory condition was significantly higher than in the other two conditions, which did not differ from each other. Panel B: Likewise, SD was higher in the auditory condition if compared with the combined condition, while the other comparisons showed no differences. Panel C: Once again, variance in the combined condition did not significantly differ from maximum likelihood estimation (MLE) prediction. Error bars are standard errors. Panel D: The integration model is depicted in orange, and the alternation model is in blue. Note the poorer performance relative to Experiment 1, Figure 2.2, panel D. A LM = auditory landmarks; SE = standard error; ns = not significant; ** = p < .01; *** = p < .001. See the online article for the color version of this figure. *Copyright © 2022 by American Psychological Association. Adapted with permission. Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). Interindividual differences influence multisensory processing during spatial navigation. Journal of Experimental Psychology: Human Perception and Performance, 48(2), 174.*

2.3.6   Interim discussion

The rationale of Experiment 2 was to lower the reliability of visual landmarks relative to Experiment 1 and to increase the possibility of optimally integrating them with the auditory ones. However, the results from this experiment were greatly similar to the ones found in Experiment 1. Notably, moving the landmarks farther from the target object location generally increased the difficulty of the spatial navigation task, resulting in a global worsening of performance for all participants on all the dependent variables. Once more, the absence of a significant improvement in precision when the audio-visual cues were presented together suggested that, in Experiment 2, multisensory behavioral integration did not seem to occur. Conversely, the comparison between observed data and the model predictions and weights suggested again optimal multisensory integration.

2.4   Modeling individual differences in sensory weighting

To summarize, taking together both performed experiments, the behavioral results indicate that, in the presence of audio-visual cues, the participants followed the visual cue, i.e., the most accurate cue to relocate the objects. In fact, no difference was found between the visual and combined conditions. Conversely, the comparisons between the observed behavior and the models' predictions indicate that the participants' performance was in accordance with Bayes's optimal prediction. Thus, the lack of significant difference between the visual and combined variability prevented us from ruling out the overall predominance of visual cues over the performance in the navigation task. Indeed, it is known that the MLE predicts a greater bimodal precision relative to unimodal precision when the unimodal estimates are equally precise. In our experiments, we observed that the visual and auditory precisions did not differ from each other. This suggests that, somehow, the two sensory modalities might be integrated. One possible explanation for these apparently contradictory results is the potential individual difference in perception of each cue. We decided to investigate this aspect. Similar to the research by (Bentvelzen et al., 2009), we split the participants of both experiments into two groups, depending on how dissimilar the unimodal weights they associated with each cue were. Thus, in each group, we conducted an analysis on the weights given to the unimodal cues, which we computed using Equation 2.4 and Equation 2.5, in an effort to verify whether the participants who weighted the two unimodal cues more similarly showed a reduction of variability in the combined condition compatible with optimal multisensory integration. Conversely, we expected that the participants who associated the disparate weights with the

visual and auditory cues would not show a reduction of multisensory variability. In this case, the variability in the audio-visual trials would be similar to the most precise unisensory one (visual). Accordingly, we hypothesized that the multisensory variance for them would be significantly greater than the MLE prediction. We divided the participants into two subgroups according to the values of their predicted weighting (Equation 2.4 and Equation 2.5), following the procedure used in the research by Bentvelzen and colleagues (Bentvelzen et al., 2009). Participants whose data returned similar unimodal weights (the difference between weights was less than .5; |wV - wA| < .5) constituted the first group, which we called "similar weights" (SW; n = 21, 12 from Experiment 1, nine from Experiment 2). Those that returned dissimilar unimodal weights (difference between weights higher than .5; |wV - wA| > .5) constituted the second group, called "dissimilar weights" (DW; n = 20, eight from Experiment 1, 12 from Experiment 2).

As previously noticed, we observed different magnitudes of responses in the two different setups of the experiments. To allow for comparison among responses of participants belonging to the two experiments' samples, we standardized the response data transforming them into the corresponding z scores. Then, we performed the same analyses that we conducted for each individual experiment on the standardized CE and SD on both SW and DW groups.

For the SW group, the analysis of the standardized CE revealed a significant main effect of condition, $F(2, 20) = 8.68$, $p = .001$, $\eta_G^2 = .14$, 95% CI [0, .33]. As shown in panel A of Figure 2.4, the CE in the auditory condition was greater than those for the visual, $t(20) = 6.25$, p adj < .0001, Cohen's d = .94, 95% CI [.58, 1.31], and the combined conditions, $t(20) = 2.74$, p adj = .013, Cohen's d = .77, [.12, 1.43]; notably, the visual and combined conditions did not differ, $t(20) = -.09$, p adj = 1, Cohen's d = .02, [-.46, .43]. Thus, the results of the accuracy measure again suggested a visual predominance.

Concerning the variability (Figure 2.4, panel B), the repeated-measures ANOVA on the standardized SD showed a significant main effect of the factor condition, $F(2, 20) = 7.543$, p = .005, $\eta_G^2 = .17$, 95% CI [0, .36]. The SD in the auditory and visual conditions did not differ from each other, $t(20) = -.257$, p adj = 1, Cohen's d = .007, Cohen's d = .99, [.29, 1.69], and the visual conditions, -.05, [-.43, .33], while both the auditory, $t(20) = 3.49$, p adj = $t(20) = 2.73$, p adj = .039, Cohen's d = .92, [.11, 1.74], differed from the combined condition. To compare the standardized variance in the combined condition (z score variance = -.453 ± .09) with the MLE prediction (z score variance = -.335 ± .11), a Wilcoxon signed-ranks test was conducted due to the failing of normality assumption; in turn, this revealed that no difference occurred (V = 88, p = .355, r = .21, [-.23, .60], $BF_{01} = 3.16$; 2.4, panel C). Notably, the value of BF_{01} increased if compared with the same comparisons done in both Experiment 1 (BF_{01}

=2.45) and Experiment 2 (BF_{01} = 1.78), suggesting more robust evidence of integration in a Bayesian fashion. The analyses of the variability measures suggested that for the SW group, multisensory integration occurred in an optimal manner.

Regarding the DW group, the repeated-measures ANOVA on the standardized CE showed a main effect on it, F(2, 19) = 16, p < .001, η_G^2 = .27, 95% CI [.04, .46]; specifically, as shown in panel D of Figure 2.4, the CE was greater in the auditory condition than in both the visual, t(19) = 3.39, p adj = .009, Cohen's d = .90, 95% CI [.26, 1.54], and combined conditions, t(19) = 5.02, p adj = .0001, Cohen's d = 1.19, [.56, 1.81]. In turn, the visual and combined CE did not significantly differ, t(19) = 2.21, p adj = .12, Cohen's d = .48, [.02, .94]. Concerning variability (Figure 2.4, panel E), the repeated-measures ANOVA on standardized SD showed that there was a significant main effect of the factor condition, F(2, 19) = 6.69, p = .006, η_G^2 = .17, 95% CI [0, .37]. The auditory SD was significantly different from the visual, t(19) = 2.69, p adj = .044, Cohen's d = .92, [.09, 1.75], and the combined SD, t(19) = 3.88, p adj = .003, Cohen's d = .93, [.35, 1.51]; in turn, the visual SD did not differ from the combined SD, t(19) = -.201, p adj = 1, Cohen's d = -.06, [-.66, .54]. Importantly, it can be observed that the combined SD in the DW group is higher than the combined SD in the SW group (panels B and E in Figure 2.4), demonstrating the substantially lower precision of the DW group in the presence of audio-visual landmarks. To compare the standardized variance in the combined condition (z score variance = -.05 ± .2) and the MLE prediction (z score variance = -.463 ± .06), a Wilcoxon signed-ranks test showed that there was a significant difference between the two (V = 173, p = .009, r = -.568, [-.82, -.20], BF_{01} = .45; Figure 2.4, panel F). For the DW group, the overall results suggested that the participants followed only the visual cues, even in the presence of auditory cues. This indicated that the predominance of visual information prevented them from reaching the optimal performance predicted by MLE.

2.5   Conclusion

In the present study, we investigated whether the presence of audio-visual landmarks could enhance performance during a navigation task and whether the two unisensory cues could be integrated in an optimal manner. Performance was determined by the ability of participants to relocate an object orienting with landmarks. In Experiment 1, we observed that participants' accuracy in relocating the target object was similar comparing the visual-only and the audio-visual conditions. Their precision using audio-visual cues was comparable to the visual-only performance. Nonetheless, the participants' precision with both audio-visual landmarks
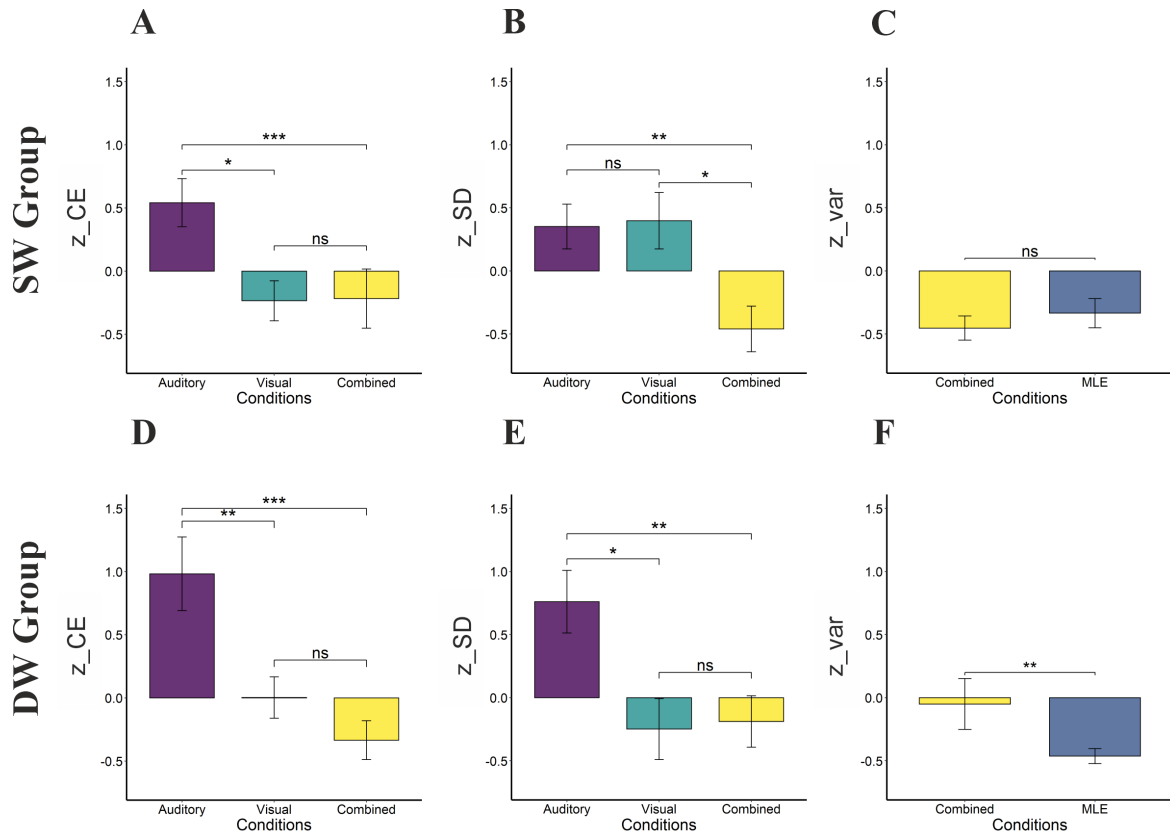
Figure 2.4 Standardized CE (z_CE) was lower in the auditory than in the visual and combined conditions in both groups (Panels A, D). However, the similar weights (SW) group showed similar standardized SD (z_SD) for the two unimodal conditions, which were significantly higher than z_SD in the combined condition (Panel B). Moreover, standardized combined variance (z_var) did not differ from maximum likelihood estimation (MLE) prediction (Panel C). Conversely, in the dissimilar weights (DW) group, the z_SD in the combined condition was not significantly lower than the best unimodal one (visual; Panel E), and the z_var was significantly higher than the MLE prediction (Panel F). Error bars are standard errors. ns = not significant; * = $p < .05$; ** = $p < .01$; *** = $p < .001$. *Copyright © 2022 by American Psychological Association. Reproduced with permission. Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). Interindividual differences influence multisensory processing during spatial navigation. Journal of Experimental Psychology: Human Perception and Performance, 48(2), 174.*

was statistically similar to the precision predicted by the MLE model. In Experiment 2, after lowering the reliability of the visual cues, we observed a pattern of results that were comparable to Experiment 1. To explain our apparently contrasting findings, we then decided to thoroughly examine the patterns of the results in the data sets of both experiments, considering the interindividual differences in perception, similar to a previous study about the integration of audio-visual cues for speed discrimination (Bentvelzen et al., 2009). Thus, we merged the data of both Experiments 1 and 2. We then divided the participants into two subgroups according to whether their predicted weights of the two unimodal cues were similar or highly discrepant (Similar weights (SW) vs. Dissimilar weights (DW)). As predicted, only SW group participants optimally integrated the auditory and visual cues, thus showing increased precision in the audio-visual condition relative to both unisensory conditions. Conversely, the DW group participants did not benefit from the presence of auditory landmarks being added to the visual ones; they accomplished their best performance with the visual-only cues.

Individual Differences in Multisensory Precision

Spatial navigation ability benefits in precision by combining similarly weighted audio-visual cues. Although we tested a homogeneous group of adult participants, we found two subgroups that behaved differently according to their ability to integrate multisensory cues: only one of them benefited from the multisensory environment. As previously mentioned, starting from the data sets of Experiments 1 and 2, we identified a subgroup of participants that behaved as integrators (the SW group) and a subgroup that behaved as non-integrators (the DW group). Therefore, the SW group showed a greater precision for the audio-visual trials than it was for either of the single-cue conditions; moreover, it was statistically undistinguishable from the MLE prediction. In contrast, in the DW group, the precision in audio-visual trials was equal to the visual-only condition, and it did not show the substantial improvement that had been predicted by the MLE model. In detail, the DW group perceived a large difference between visual and auditory reliability and relied on the best sensory information they had available (the visual landmarks), reducing their possibility of exploiting the advantage of multimodal trials. The origin of these differences among individuals is not clear yet. Many elements can indeed influence cues interaction. People's navigation ability is the result of their perceptual and cognitive processing (Wolbers and Hegarty, 2010). It can be speculated that the differences found among participants may be related to their natural navigation preferences, different navigation strategies, or previous experiences lived since childhood. Individual features and experiences may have modulated

the ability to optimize navigation performance by exploiting the audio-visual cues available without reliable self-motion concomitant cues. These findings reveal the importance of considering individual perceptual preferences when probing multisensory integration during spatial navigation. These aspects underlying interindividual variability are deeper discussed in Chapter 5.

Sensory Weighting in Conflict Condition

The condition with a spatial conflict between the auditory and visual cues suggested that participants integrate — rather than alternate — between cues. Performing the same modeling analyses of Nardini and colleagues (Nardini et al., 2008), we compared the precision of the conflict condition to the performances predicted by the integration and alternation models. In both experiments, the precision of the conflict condition was greater than the prediction made using the alternation model, thus suggesting that participants did not switch between auditory and visual cues, in line with previous outcomes (Bates and Wolbers, 2014; Nardini et al., 2008). Intriguingly, in Experiment 1, the observed precision was significantly better than the integration model prediction. This suggests that participants were more precise in the conflict condition than predicted based on their empirical unimodal weights. The integration model uses weights that originate from the relative proximity (Equation 2.1); this corresponds to the predicted distance of responses relative to the visual and conflicting auditory landmarks, reflecting a measure of accuracy. The observed precision in the conflict condition was better than predicted, probably because accuracy was not taken into account to perform precisely. In line with previous literature (Zhao and Warren, 2015b), accuracy and precision may have followed different principles of cue interaction, showing to be independent measures of behavior. Indeed, supporting this view, our participants were able to achieve high precision regardless of their accuracy, meaning that they could integrate the audio-visual cues without being affected by the spatial conflict between the cues. Even if no difference was found between the empirical and predicted weights, our observed results on precision are better explained by the MLE model, which considers the weights predicted by the inverse of the unisensory variances (Equation 2.4 and 2.5).

Visual domination for accuracy

Results on participants' accuracy suggest that visual information dominates spatial navigation. Indeed, accuracy was equal in conditions with and without auditory landmarks, thus indicating that the processing of the visual-only information was sufficient for participants to achieve their most accurate performance.

Moreover, in line with the prior literature (Alais and Burr, 2004; Gori et al., 2012), our results show greater accuracy in performing spatial tasks based on visual rather than auditory sensory information. In our experiments, we chose steady auditory landmarks, as done in previous studies (Jetzschke et al., 2017; Nardi et al., 2020; Viaud-Delmon and Warusfel, 2014), because we meant to provide constantly active sound stimuli that may be more likely to be detected and used to spatially navigate. Besides, in spatial contexts, vision is the most accurate perceptual sense because it specifies location more reliably than hearing (Gori, 2015). In addition, according to the cross-sensory calibration theory, during childhood, vision is the sense that calibrates the others, dominating spatial processing (Gori, 2015; Gori et al., 2008). In the present study, the DW group (non-integrators) behaved as if vision was still the dominating sense, preventing them from benefiting from the redundancy of the multisensory environment. Therefore, the participants who were already highly precise in the visual-only condition did not improve their performance when the visual and auditory landmarks were available simultaneously. The visual dominance for navigational accuracy is furtherly discussed in Chapter 5.

This chapter has been partially extracted and adapted from:

Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). Interindividual differences influence multisensory processing during spatial navigation. Journal of Experimental Psychology: Human Perception and Performance, 48(2), 174.

Chapter 3

Self-motion and auditory cues in spatial orientation

3.1 Self-motion cues in spatial orientation

Self-motion perception is essential in maintaining orientation and performing efficient spatial navigation. While navigating, we constantly update our position and orientation, yielding an estimation of movement direction, traveled distance, and trajectory. Self-motion perception originates from the combination of multiple cues. Among all, this chapter focuses on the contribution of vestibular cues. As described in Chapter 1, the peripheral vestibular system is particularly essential for self-motion perception because it provides constant and crucial information to sense the linear and angular acceleration of the head thanks to the two otoliths and the three semicircular canals in the inner ears, respectively. Therefore, all vestibular organs offer the basis to build a representation of three-dimensional space (Lackner and DiZio, 2005), taking into account three translational and three rotational axes. Besides, the vestibular system gives information about heading direction (Crane, 2012a; Cuturi and MacNeilage, 2013) and helps disambiguate object motion from self-motion (MacNeilage et al., 2012). It has been shown that healthy human subjects can discriminate the direction of rotational whole-body motion at velocities that are below 1 deg/s in darkness (thus without any visual hints). As further evidence of the essential vestibular contribution to self-motion perception, previous research found that monkeys show a substantial decrease in the ability to discriminate the direction of a translational heading following bilateral labyrinthectomy (Gu et al., 2007). Similarly, the threshold velocities that can be discriminated by patients with complete bilateral vestibular loss are higher than in healthy subjects (Valko et al., 2012).

In addition to the crucial role that the vestibular system plays in maintaining balance (via low-level reflexes) and sensing motion, recent research is increasingly providing evidence that vestibular signals significantly contribute to higher cognitive functions, such as spatial memory (Hilliard et al., 2019), spatial orientation and navigation (Dallal et al., 2015; Gu, 2018; Karn and Cinelli, 2019; Xie et al., 2017) on both humans and animals. Specifically, the processing of vestibular inputs creates an egocentric frame of self-motion, guiding the navigator through the external world. From the basic analysis of motion, vestibular processing leads to estimating traveled distances relative to a starting point in the environment and the body position relative to the outer world. Indeed, it has been demonstrated a dissociation between the contribution of vestibular inputs to perceiving motion and their contribution to understanding the temporal and spatial features of the passive movements experienced by the navigator (Kaski et al., 2016).

The vestibular system is behaviorally and functionally connected with the visual system. Prior research has indeed shown that optic flow alone can give rise to vestibular-based self-motion perceptual aftereffects (Cuturi and MacNeilage, 2014), suggesting a shared "hardware" to process spatial motion stimuli. Moreover, at a functional level, visual and vestibular cortical areas inhibit each other, suggesting a robust interaction between the spatial information that these two sensory systems provide to the brain (Brandt et al., 2002). Therefore, it becomes natural to ask what happens to vestibular functioning when vision is unavailable, such as in visual impairment. Indeed, studying sensory impairments allow for a deeper understanding of the fundamental role played by the impaired sensory system, during development and in daily life. Vision is the most accurate sense to process spatial information due to its high spatial acuity (Alais and Burr, 2004). For this reason, according to cross-sensory calibration theory (Gori, 2015), the visual system has the function of calibrating the other senses and guiding them to process spatial information during development. Therefore, the lack of vision, especially since the early stages of life, impacts the function of any other sense (e.g., hearing) that relies on it for calibration, e.g., for spatial perception.

Few previous studies have actually investigated the features of vestibular functioning in visually impaired people. They showed perceptual thresholds in interaural translational and rotational yaw (along the Earth-vertical axis) planes similar to sighted, but better sensitivity in the roll tilt plane in correspondence of an otoliths-semicircular canals interaction, suggesting a compensation to be more sensitive to head position changes to gravity in absence of vision (Moser et al., 2015). Even more interestingly, visually impaired people showed an intact ability to process basic self-motion information (e.g., reproduce a previously experienced passive movement), but a worse ability to take into account more complex spatial processing

and an absolute representation of space (e.g., infer a novel trajectory) (Seemungal et al., 2007). However, this finding is strongly affected by interindividual performance, suggesting that previous idiosyncratic spatial expertise could influence spatial processing based mainly on vestibular signals (Seemungal et al., 2007).

3.2 Auditory cues in spatial orientation

Traditionally, studies on spatial navigation and orientation have been dominated by vision, i.e., the aptest sensory system to process spatial information (Alais and Burr, 2004). However, as mentioned in Chapter 1, the environment in which we navigate is substantially multisensory, providing, among others, auditory landmarks to re-orient. Although hearing is less accurate than vision for spatial processing, auditory cues not only provide relevant spatial information to navigate, but they also have some advantages over visual ones. For example, sound stimuli convey fast information about stimuli located far from the navigator's position; they also allow the navigators to perceive multiple and simultaneous sources regardless of their position (Long and Giudice, 2010; Nardi et al., 2020). In Chapter 2, we have shown how, for a subgroup of navigators, visual and auditory landmarks can be combined to improve navigational performances. In addition, it has been shown that providing only auditory landmarks in the environment improves balance performance (Karim et al., 2018) and active vestibulospinal control of movements in healthy subjects (Seiwerth et al., 2018). Shayman and colleagues (Shayman et al., 2020) demonstrated that adding a sound source while passively moving can enhance the perceptual sensitivity in discriminating movements' directions. In the absence of vision, visually impaired people need to rely on the available acoustic elements to build a coherent spatial representation of the environment in which they travel. Working together with inertial and proprioceptive cues, auditory cues indeed help visually impaired people have access to far space and to a mental representation of space that does not require them to move. Recently, as occurred with visual stimuli, the research on the spatial information provided by auditory cues has been shifted from real environment settings to virtual reality tools. The flexible nature of virtual reality allows the creation of highly controllable auditory cues in terms of number, distance, and position. As one may easily imagine, the implementation of this kind of technology has a strong potential to build novel assessment tools and rehabilitation procedures that can help improve mobility skills for people with visual impairment.

3.3   General method and simulation devices

In the following sections, I am going to present two studies that examined the investigation of the auditory-vestibular interaction in visually impaired people (Study 1) and the vestibular-based accuracy of self-motion estimation in healthy subjects (Study 2). In particular, Study 1 focused on testing the feasibility of a novel setup composed of a motion platform, which delivers movements that elicit vestibular responses, and an acoustic virtual reality tool, which delivers spatialized auditory landmarks. Here, we have tested this new combination of simulation tools on sighted and visually impaired participants. Subsequently, Study 2 used the same setup to selectively investigate the self-motion estimation ability of healthy subjects with and without auditory landmarks.

Before moving forward with the details of these two studies, the following sections thoroughly describe the two mentioned simulation tools.

3.3.1   The Rotational-Translational Chair

Researchers often develop and use motion platforms for scientific and clinical purposes to deliver controllable rotations or translations and measure the consequent responses of the vestibular system. As reported in the literature (Cuturi et al., 2020), many types of existing devices can deliver rotational and/or translational motions, and they vary according to different sizes and degrees of freedom. For example, one of the most popular devices for these studies is the Barany or rotatory chair, a motorized chair capable of delivering 360 degrees endless rotations. For translations, motorized trolleys are used to deliver linear translations either on the inter-aural (left-right) or the naso-occipital axis (forward or backward). More complex systems with six degrees of freedom (e.g., the MOOG 6DOF2000E) are able to deliver a combination of rotatory and linear movements.

At the Italian Institute of Technology (Genoa, IT), Cuturi and colleagues (2020) have conceived and developed a novel motion platform that could combine the possibility of delivering both angular and linear movements, and that requires a small space to be functioning. This device is called the Rotational-Translational Chair or RT-Chair (left panel of Figure 3.2). This 2-degrees-of-freedom motion platform can provide smooth and vibration-free movements generating a highly controlled motion path centered on the head of the subject seated on the chair, thus eliciting responses from all the vestibular organs. The RT-Chair is composed of a racing car seat, with a four-point belt, to guarantee comfort and safety to subjects and the system. The translations and the rotations are driven by two brushless

motors located on trolleys. The generated motion stimuli follow a minimum jerk motion
profile, described by the following equation:

$$X_p(t) = X_0 + S_0 t + [10(X - X_0) - 6S_0]t^3 + [-15(X - X_0) + 8S_0]t^4 + [6(X - X_0) - 3S_0]t^5$$

(3.1)

where, $Xp(t)$ is the produced trajectory, t is time, X is the target position, X_0 is the initial
position and S_0 is the initial speed (usually zero). This kind of motion profile has been
implemented due to mechanical constraints but also to resemble the sinusoidal acceleration
profile, frequently used in experiments using motion platforms (e.g., in Grabherr et al. (2008);
Moser et al. (2015)). An example of how the minimum jerk profile looks is shown in Figure
3.1.
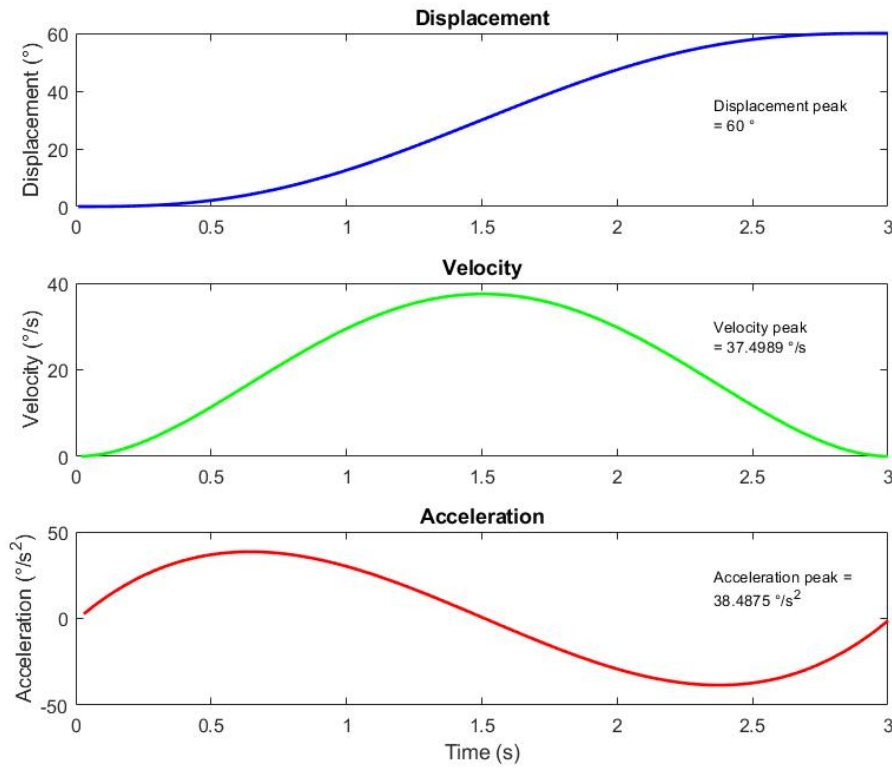
Figure 3.1 Here, it is shown the minimum jerk motion profile for a 3-second yaw rotation
(0.33 Hz) of 60° of displacement (upper panel). In the middle panel, the velocity profile
is displayed; in this case, the reached peak velocity is 37.5°/s. In the lower panel, the
acceleration profile is depicted (negative values on the y-axis represent deceleration); the
reached peak acceleration is 38.5°/s^2.
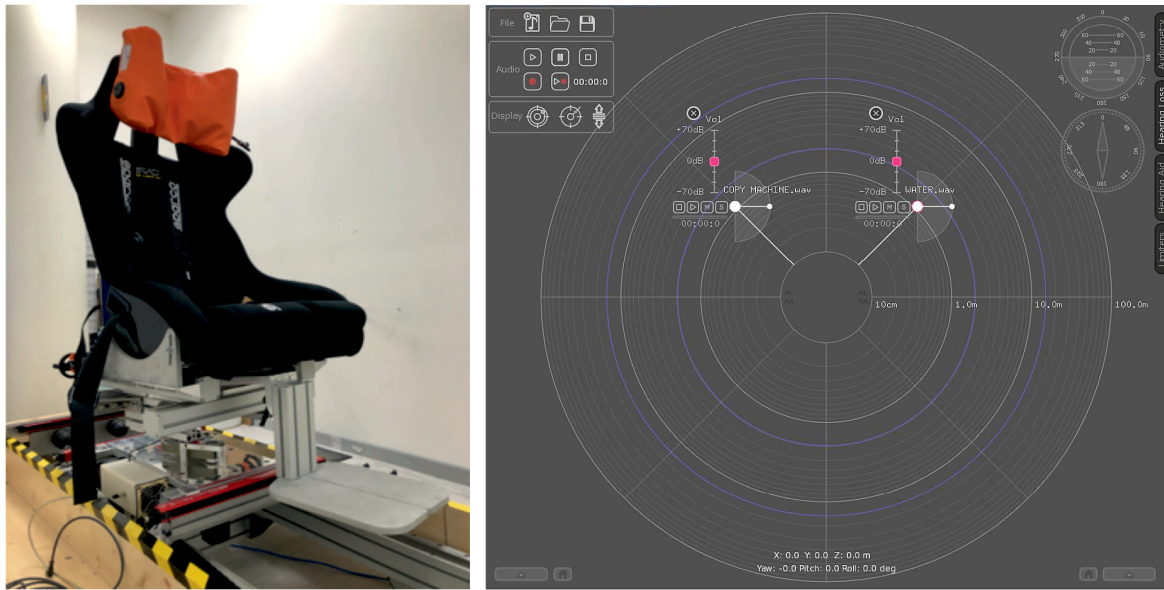
Figure 3.2 On the left panel, a picture of the RT-Chair; on the right panel, the 3DTI Toolkit user's interface with two virtually spatialized sounds as examples. © *2021 IEEE. Reprinted, with permission, from Zanchi et al., Evaluation of a Motion Platform Combined with an Acoustic Virtual Reality Tool: a Spatial Orientation Test in Sighted and Visually Impaired People, November 2021.*

In the two studies described below, we chose to use this newly developed motion platform to deliver only rotational movements, which trigger semicircular canals' responses. In particular, in both studies, the motion stimuli consisted of 3-second yaw rotations (0.33 Hz) along the Earth-vertical axis. The rotation angles ranged from 10° to 80° clockwise and -80° to -10° counterclockwise. Peak velocities ranged from 6.25°/s to 50°/s; peak accelerations ranged from $6.41°/s^2$ to $51.32°/s^2$.

3.3.2 The 3D Tune-In Toolkit

The 3D Tune-In Toolkit (3DTI Toolkit) is an open-source standard C++ library for audio spatialization and simulation of hearing loss and hearing aids, developed by researchers from Imperial College London and the University of Málaga (Cuevas-Rodríguez et al., 2019). This tool allows the building of three-dimensional (3D) soundscapes, thanks to real-time binaural spatialization. In this way, it is possible to provide the listener with the perception that sound is located at a specific point in the three-dimensional space, all through a pair of standard headphones used as a playback device. The binaural spatialization is based on convolving a monaural signal with head-related transfer functions (HRTFs), which model the directional

filtering of the incoming signal due to the characteristics of the ears and body of the listener. Regarding the simulation of distance, the 3DTI Toolkit implements the inverse square law (which states that the intensity of sound decreases by approximately 6 dB for each doubling of distance from the sound source) but also allows for customization of the attenuation.

For more details about the toolkit, please see (Cuevas-Rodríguez et al., 2019).

In our studies, the 3DTI Toolkit has been used to build an acoustic virtual environment that provided spatialized acoustic landmarks. To associate the 3DTI Toolkit with Matlab (The Mathworks, United States), we used a communication protocol previously implemented by our lab (Setti et al., 2021). The 3DTI Toolkit's test application receives instructions through Open Sound Control (Wright, 2005), based on wireless (WLAN) communication, a protocol for data exchange among devices. It is based on User Datagram Protocol (UDP), which uses the IP addresses and the port of the communication units. The UDP communication was established in Matlab by creating a UDP object (*udp()* function), while the *oscsend()* Matlab function allowed positioning the sounds where needed with instructions compatible with the toolkit. The right panel of Figure 3.2 shows the user's interface of the application. The sounds are simply delivered over binaural headphones (Sennheiser HD-650), used as a playback device.

3.4   Study 1: Audio-vestibular interaction in sighted and visually impaired people

In the present study, I tested a novel experimental procedure based on the combined use of the two simulation devices described above. The combination of the RT-Chair and the 3DTI Toolkit allowed us to test how vestibular self-motion cues and auditory landmarks may interact. Moreover, considering how visually impaired people rely on auditory information for orientation, we aimed to compare the use of virtual auditory landmarks between sighted and visually impaired participants in a spatial orientation task. Here we focused on participants' precision in a self-motion discrimination task in the presence or absence of virtual auditory landmarks.

3.4.1   Participants

To validate the combination of the simulation tools, five visually impaired (mean age $\pm$ SD = 48 $\pm$ 7 y.o.) and nine sighted participants (mean age $\pm$ SD = 28 $\pm$ 6.2 y.o.) took part in this study after giving written consent. Table 3.1 displays the clinical details of visually

impaired participants. The study was approved by the ethics committee of the local health service (Ethical Committee, ASL 3, Genova, Italy) and conducted in accordance with the Declaration of Helsinki.

Table 3.1 Clinical details of visually impaired participants.

Participants	Age (y.o.)	Pathology	Impairment onset (y.o.)	Residual vision
B01	55	Nystagmus and Retinis pigmentosa	30-40	Lights and shadows
B02	44	Loss of retina	18	Lights and shadows
B03	44	Congenital cone dystrophy	Birth	1/50 both eyes
B04	42	Retinis pigmentosa	22	Right eye: 3-5° visual field; Left eye: lights and shadows
B05	57	Retinis pigmentosa	7	Lights and shadows

3.4.2   Apparatus and stimuli

As briefly mentioned earlier in the chapter, we delivered yaw rotations of 0.33-Hz frequency with the RT-Chair (Cuturi et al., 2020). Regarding the auditory landmarks, they consisted of two semantic sounds (1 s of duration each). These sounds were: a working copy machine and water being poured. We selected semantic sounds because of the more significant impact they have on self-motion perception compared with non-semantic sounds (Riecke, 2016). We downloaded all sounds from a royalty-free website (https://freesound.org/). Relative to the starting position at azimuth 0° (in line with participants' nose), the copy machine sound was spatialized at azimuth -45° (90° on the left), the water sound at azimuth 45° (on the right) at a distance of 1.1 m (see Figure 3.3). To test participants' ability to effectively localize these sounds, we performed a brief control experiment, in which we asked them to discriminate sounds' position ("Is this sound on the right or left relative to the 0°?").

3.4.3   Experimental design and procedure

Participants sat on the padded racing seat of the RT-Chair. The head of participants was aligned with RT-Chair's rotation axis and leaned against a vacuum pillow, shaped according to the head, with their forehead held to the chair with a padded strap to reduce the potential use of neck proprioceptive cues as sources of information for orientation.

The experimenter described the task and gave participants a pair of binaural headphones and a wireless numeric keypad, whose keys to be used were made well recognizable by touch by applying a thick layer of tape on them. During the experiment, the room was

darkened; in addition, sighted participants had their eyes closed and covered by an eye mask to prevent any use of the irrelevant visual information available in the experimental room. Participants performed a self-motion discrimination task, in which they were rotated on the Earth-vertical axis and had to estimate the amplitude of each experienced rotation. Specifically, for clockwise rotations, after each rotation, they had to report verbally whether they felt to be closer to the point of reference at azimuth 0° (the starting position, in line with participants' noses) or to the point of reference at azimuth 90°. For counterclockwise rotations, they reported whether they felt to be closer to azimuth 0° or azimuth -90° (90° to the left). To make clear the ±90° points of reference to participants, before the experimental session, they experienced four rotations with an amplitude of 90°, one for each level of the experimental design explained below.

We tested two conditions (Vestibular-only vs. Multisensory, i.e. vestibular and auditory cues in the same trial) and two movement directions (counterclockwise vs. clockwise rotations). In Vestibular-only trials, participants had to estimate their movement's amplitude by only relying on the vestibular cue from clockwise or counterclockwise rotations. In the Multisensory condition, participants could rely on both vestibular and auditory cues to estimate the rotations' amplitude. Specifically, when rotating clockwise, the auditory cue consisted of the water sound, placed at azimuth 45°; when rotating counterclockwise, the cue was the copy machine sound, placed at azimuth -45°. Thus, the auditory cues worked as landmarks that helped recognize the middle amplitude between the 0° and ±90° reference points. Notably, considering the well known difficulty of localizing acoustic cues, we chose to inform all participants about the position of the auditory landmarks. Moreover, we tested participants' ability to localize the semantic auditory landmarks without any vestibular input. Before each trial, as a "GO" signal, a voice through headphones suggested the type of condition: the word "Rotazione" (Italian for rotation) indicated a vestibular-only trial; the word "Audio" (Italian for audio) indicated a multisensory trial. To start each trial, participants pressed one keypad button. In the Vestibular-only condition, right after the pressure of the keypad button, a rotation began. In the Multisensory condition, the trial was composed as follows: after the pressure of the keypad button, the auditory landmark was played; right after listening to the sound, the rotation began; when it stopped, the auditory landmark was presented again. In these trials, participants were explicitly instructed to use both vestibular and auditory cues to estimate their rotation. To keep the auditory landmark located at the same positions at ±45° after the rotation, the sound was virtually rotated with an amplitude equal to the presented rotational movement but in the opposite direction to simulate a sound
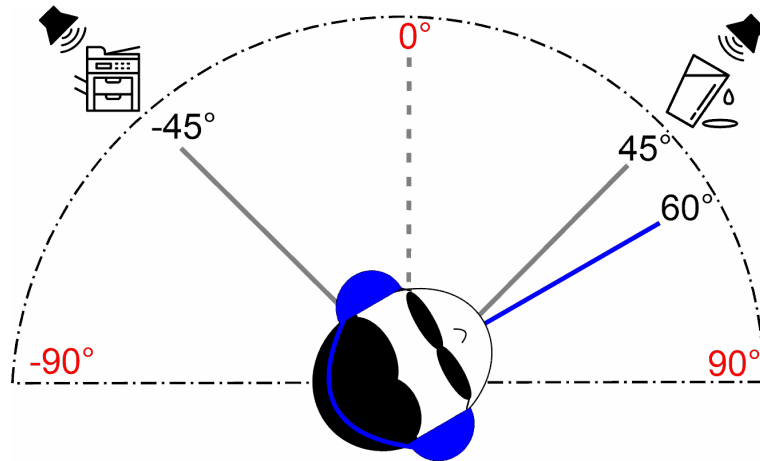
Figure 3.3 Experimental setup. At azimuth -45°, it is represented the sound of the copy machine landmark; at azimuth 45°, the sound of the water landmark. The considered points of reference at azimuth -90°, 0°, and 90° are highlighted in red. In the depicted example, the participant is rotated 60° clockwise, thus closer to azimuth 90°. © *2021 IEEE. Reprinted, with permission, from Zanchi et al., Evaluation of a Motion Platform Combined with an Acoustic Virtual Reality Tool: a Spatial Orientation Test in Sighted and Visually Impaired People, November 2021.*

fixed in the environment (e.g., after a 30° clockwise rotation, the sound would be rotated 30° counterclockwise).

During all RT-Chair rotations, a white noise sound was played through headphones to mask the sounds elicited by the device. For each trial, the experimenter collected the verbal response ("I feel closer to azimuth 0°" or "I feel closer to azimuth 90°"), and the participant was brought back to the starting point with a movement with reduced frequency (0.25 Hz). To prevent any potential aftereffects between two consecutive movements (Crane, 2012b), we presented a 3-second time window between the experimental motion stimuli. The order of conditions was randomized across trials for all participants. In each condition, 54 trials were tested, of which the first four were training trials with fixed movement magnitude. For the other trials, rotation amplitude was determined by the Psi adaptive procedure (Kontsevich and Tyler, 1999) implemented using the PAL_AMPM routine from the Palamedes toolbox (Prins and Kingdom, 2018) in Matlab (total n trials = 216). The experiment lasted 1 h and 30 min. To prevent fatigue, participants were encouraged to take two breaks at one-third and two-thirds of the experiment.

3.4.4 Data analysis

One sighted participant was excluded from the analysis due to a technical error. For clockwise rotations, we plotted the percentage of responses "I felt closer to azimuth 90°" as a function of the delivered stimulus displacement. In parallel, we plotted the percentage of responses "I felt closer to azimuth 0°" for counterclockwise rotations. For each participant and condition, we fitted a cumulative Gaussian to the data using the PAL_PFML_ Fit routine from the Palamedes toolbox (Prins and Kingdom, 2018), which finds the best fit in a maximum likelihood sense (Figure 3.4). We took the standard deviation of the distribution (the just noticeable difference, JND) as a measure of precision. The estimates' error was calculated by performing a non-parametric bootstrap analysis, running the function PAL_PFML BootstrapNonParametric from the Palamedes toolbox, which generated 400 simulated data sets (Prins, 2016). For each subject, we obtained a JND value for each condition and direction of rotation. Considering the violation of normality distribution for JND values, we conducted a non-parametric permutation ANOVA (using the function *ezPerm* from the *ez* package in RStudio, version 3.6.2, 2019) with Condition as a within factor and Group as a between factor. Generalized eta squared (η_G^2) is reported as effect size. We pooled the results of both clockwise and counterclockwise movements because we did not have reason to expect any difference in the direction of rotations. In addition, for the control experiment, we fitted a a cumulative Gaussian to the data using the PAL_PFML_ Fit routine from the Palamedes toolbox (Prins and Kingdom, 2018) and we took the point of subjective equality (PSE) as a measure of accuracy and JND for precision.

3.4.5 Results

To test the combination of a virtual acoustic environment and a motion platform, we compared both sighted and visually impaired people's precision in a self-motion discrimination task. Importantly, in the control experiment, participants oh both groups showed no significant bias in localizing the presented auditory landmarks (PSEs did not significantly differ from 0° for each group and sound) and good precision. Statistical analysis on JND showed neither a main effect of the Condition (Perms = 5000; p = .306, η_G^2 = 0.02, 95% CI [0.00, 0.34]) nor the Group (p = .395, η_G^2 = 0.06, [0.00, 0.42]) nor interaction (p = .411, η_G^2 = 0.01, [0.00, 0.32]). The absence of a statistically significant effect of the Condition indicates that the level of precision in estimating the rotations' amplitude relative to external points of reference is comparable between the condition in which only vestibular cues are available and the condition in which the virtual auditory landmark is provided. The lack of worsening
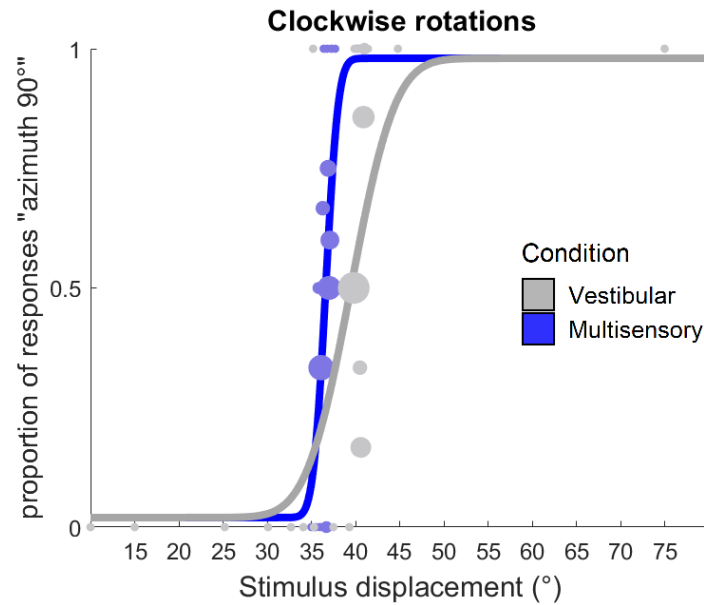
Figure 3.4 Example psychometric fits. Individual data of a visually impaired participant from the Vestibular-only (gray line and shaded gray dots) and Multisensory (blue line and shaded blue dots) conditions for clockwise rotations. The size of the dots is proportional to the number of presentations for that delivered stimulus displacement. © *2021 IEEE. Reprinted, with permission, from Zanchi et al., Evaluation of a Motion Platform Combined with an Acoustic Virtual Reality Tool: a Spatial Orientation Test in Sighted and Visually Impaired People, November 2021.*

in the precision in the self-motion discrimination task suggests that the combination of the 3DTI Toolkit and the RT-Chair tools in an experimental setting is feasible for both sighted and visually impaired populations. Interestingly, despite the large inter-individual variability, Figure 3.5 shows that for the visually impaired group, there is a trend to decrease the variability in the Multisensory condition, i.e., to increase the precision, becoming more similar to the sighted group's performance.

3.4.6 Conclusion

The present study allowed us to test a novel experimental procedure that involved a tool for acoustic virtual reality, the 3DTI Toolkit, and a novel motion platform, the RT-Chair, for both sighted and visually impaired participants. The 3DTI Toolkit allowed for building a flexible and immersive three-dimensional soundscape with manageable numbers, distance, and position of the sound sources, using simple binaural headphones as a playback device. The RT-Chair allowed for investigating a broad range of movements with a user-friendly Matlab interface. We observed a trend of improvement in visually impaired participants'
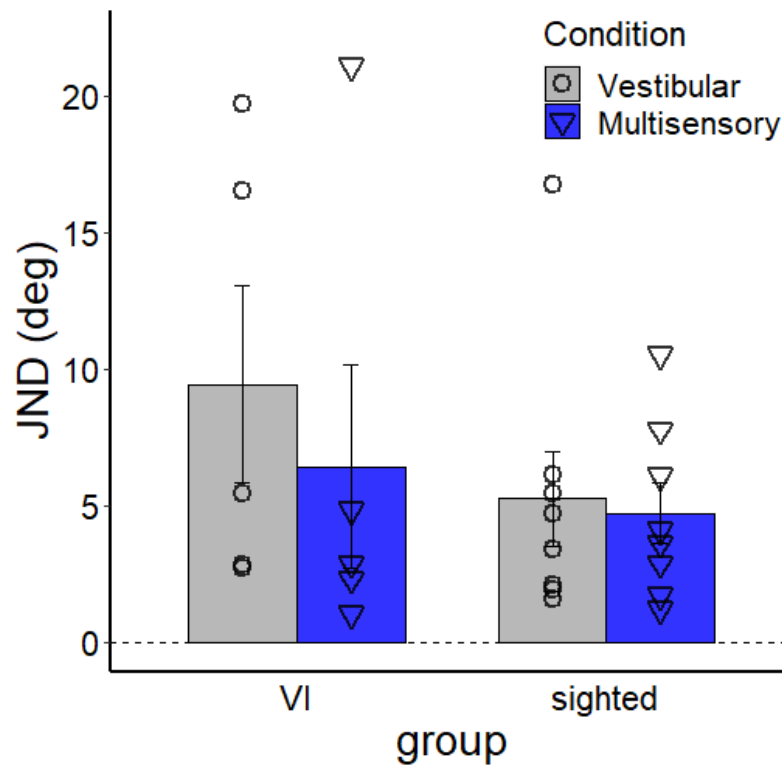
Figure 3.5 JND values for visually impaired (VI) and sighted participants in Vestibular-only (grey) and Multisensory conditions (blue). Error bars represent the standard error of the means. Individual data are represented by circles and inverted triangles for Vestibular-only and Multisensory conditions, respectively. Notice that the higher the JND bar, the lower the precision. *© 2021 IEEE. Reprinted, with permission, from Zanchi et al., Evaluation of a Motion Platform Combined with an Acoustic Virtual Reality Tool: a Spatial Orientation Test in Sighted and Visually Impaired People, November 2021.*

precision when in the presence of virtual auditory landmarks, making their performance in precision more similar to the sighted one. However, as a limitation of the study, it has to be noticed that we tested the performance of a heterogeneous and small sample of participants. Overall, this study confirms the successful use of a novel combination of the two simulation tools to investigate self-motion perception in healthy and clinical populations.

3.5 Study 2: Robust overestimation bias in self-rotation perception

The vestibular system provides essential insights into how we perceive the world in which we move. When perceiving the available interoceptive and exteroceptive stimuli in the

environment, we first represent their sensory readout; second, we attempt to interpret that representation (Seriès and Stocker, 2009; Wei and Stocker, 2015). World perception is not always accurate, leading to systematic errors known as biases. Perceptual biases can stem from morphological features of our sensory systems (Francl and McDermott, 2022; Gillingham and Previc, 1993; Li and Durgin, 2016) and also from prior knowledge (e.g., previous experience) about the world. According to the Bayesian perspective, the final stimuli of perceptual representation are the result of the combination between prior knowledge and perception of physical stimuli. In this context, perceptual biases due to previous experience allow one to respond to environmental stimuli more efficiently than one might if perception were flawless. For example, in vestibular perception, the Aubert effect is a well-established bias (Aubert, 1861), which leads to estimating the verticality towards the direction of the body tilt, likely due to an underestimation of the body tilt itself. A Bayesian model interprets this bias as a prior set at the most common position of the head, which is $0°$ in the roll plane (i.e., not tilted) (De Vrijer et al., 2009). Along the same lines, other functional vestibular biases are found in heading perception: when estimating heading direction, lateral deviations from the straight-ahead position are over-represented to signal any change from the most common direction of movement, which is the straight-ahead (Crane, 2012a; Cuturi and MacNeilage, 2013). These studies propose that vestibular cues are processed to obtain a functional representation of how people move through the environment, enhancing the discriminability of motion stimuli despite the resulting inaccuracy. Still, it remains unclear whether spatial representation biases occurred when estimating rotational displacements in the yaw plane. Rotations in the yaw plane, which are the rotations along the Earth-vertical axis, are essential for perceiving deviations from the straight-ahead direction during walking (e.g., sensing the veering). Prior research which investigated the accuracy of healthy participants in the estimation of passive rotational displacements has found discrepant findings. For example, some studies reported an underestimation of rotations (Blouin et al., 1995; Mergner et al., 1991). Other investigators found no bias in the estimates (Siegler et al., 2000). Finally, other studies reported an overestimation of passive rotational displacements (Israël et al., 1995; Ivanenko et al., 1997; Mackrous and Simoneau, 2011; Marlinsky, 1999). These contrasting findings likely stem from differences across tasks and inter-individual variability in spatial perception (Bruggeman et al. 2009; see also Chapter 2). Since signals coming from the semicircular canals play an important role in perceiving complex path perception, such as curvilinear motions (Cheng and Gu, 2018), and thus in everyday activities, further research is needed to better understand rotations perception.

Interestingly, the vestibular system interacts extensively with other sensory systems, such as visual, proprioceptive, and motor signals along the vestibular central pathway (Angelaki and Cullen, 2008). The information coming from vestibular organs is combined with external cues in the environment to build an efficient representation of the surroundings. Previous literature showed that vestibular and visual information are optimally integrated, leading to enhanced precision while perceiving heading direction (e.g., Butler et al. 2014; Fetsch et al. 2009; Gu et al. 2008). Recent studies are showing that vestibular signals also interact with spatialized auditory information, contributing to balance (Rumalla et al., 2015), improving ambulation (Karim et al., 2018), and self-motion perception (Shayman et al., 2020). Indeed, spatialized sounds can aid spatial orientation when vision is unavailable, such as in visual impairments, or unreliable, like in the presence of fog or at night. Several studies (see Väljamäe 2009, for a review on this topic) have proved that similar to optic flow, also moving sounds prompt vection, which consists of the illusion of self-motion in the presence of external moving cues without any true acceleration cue signaled by the vestibular system. Even though auditory vection is usually weaker than the corresponding visual illusion, studies suggest that one can perceive it as rotational and translational self-motion (Riecke, 2016). Overall, these findings suggest that auditory landmarks (i.e. external points of reference) may interact with vestibular information during self-motion. Remarkably, vestibular information is peripherally and centrally integrated with auditory processing (Smith, 2012).

It still needs to be investigated as to what extent rotational information contributes to the functional representations of space and whether audio-vestibular interaction might influence spatial perception's accuracy. To answer these scientific questions, in the present study, we aimed to investigate the perception of rotational displacements and the interaction between vestibular and auditory cues using a self-motion discrimination task. Similarly to what was described earlier in this chapter (in Section 3.4), we evaluated participants' ability to estimate the amplitude of passive rotations in the earth-vertical yaw plane, both with and without virtual auditory landmarks. Besides, with this study, we mainly focused on unveiling accuracy performance in self-rotation estimation.

Specifically, for Study 2, we performed two experiments. In Experiment 1, we asked participants to estimate the amplitudes of their rotations in a discrimination task wherein vestibular and audio-vestibular trials alternated randomly. The results of Experiment 1 did not clarify how participants actually used the auditory information that we made available spontaneously before and after rotations. Indeed, when a stimulus is available in the environment, our sensory systems likely process it differently according to whether we generate the stimulus or it is caused externally (Blakemore et al., 2000). Thus, in Experiment 2, we

investigated how administering actively generated sounds may influence the processing of the auditory cues themselves. We manipulated how the auditory landmarks were administered so that they were presented before and after the rotations as a consequence of a voluntary keypad button press. Here, we focused on revealing whether self-generated auditory cues would modulate the perceptual bias that we expected in the self-motion perception task.

We hypothesized to find: (i) an overestimation bias in self-rotation perception similar to heading perception biases observed in prior literature; (ii) an interaction between vestibular and auditory cues so that the available auditory landmarks would have influenced displacement estimations.

3.5.1 Experiment 1: Self-motion estimation with audio-vestibular cues

3.5.1.1 Participants

Sixteen healthy subjects (mean age $\pm$ SD: 27.1 ± 4.1 years old) participated in Experiment 1. None of them were aware of the study's aims. Participants did not report a history of neurological, acoustic, or vestibular sensory disorders and had normal or corrected-to-normal vision. The ethics committee of the local health service (Ethics Committee, ASL 3, Genoa, Italy) approved our study. It was conducted following the guidelines of the Declaration of Helsinki. All participants gave written informed consent.

3.5.1.2 Apparatus and stimuli

Regarding the motion stimuli, we delivered yaw rotations of 0.33-Hz frequency with the RT-Chair (Cuturi et al., 2020) because previous research (Shayman et al., 2020) showed the integration between auditory and vestibular cues for low-frequency stimuli (below 0.5 Hz). Regarding the auditory landmarks, they consisted of two semantic sounds (1 s of duration each). Similar to the auditory landmarks used in Study 1, these sounds were: a working copy machine, water being poured, and typing on a computer keyboard. Again, we chose semantic sounds considering the more significant impact they have on self-motion perception compared with non-semantic sounds (Riecke, 2016). We downloaded all sounds from a royalty-free website (https://freesound.org/). Relative to the starting position at azimuth 0° (in line with participants' nose), the copy machine sound was spatialized at azimuth -90° (90° on the left), the water sound at azimuth 0°, and the water sound at azimuth 90° at a distance of 1.1 m (Figure 3.6).
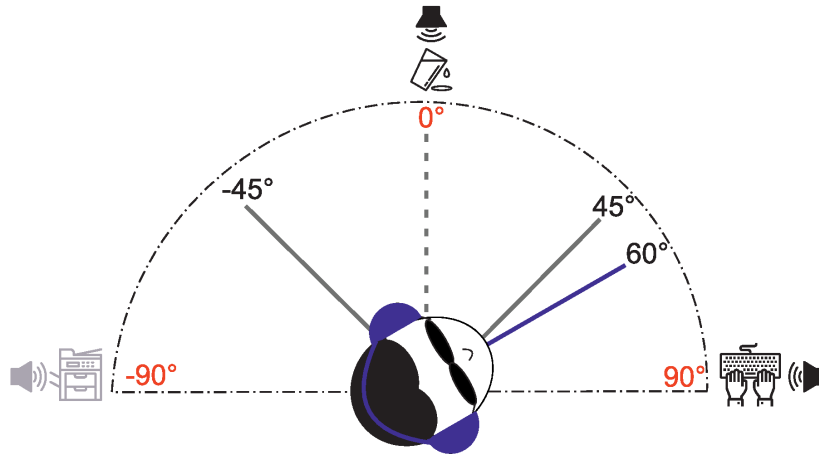
Figure 3.6 Experimental setup. The considered points of reference at azimuth -90°, 0°, and 90° are highlighted in red. In the represented example, the participant is rotated 60° clockwise, thus closer to azimuth 90°. With clockwise rotation, the delivered auditory landmarks were the sound of the water at azimuth 0° and the sound of the keyboard at azimuth 90°, black in the picture. In grey, the copy machine sound located at azimuth -90°, not active for clockwise rotations. *Reproduced with permission by Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). How much I moved: Robust biases in self-rotation perception. Attention, Perception, and Psychophysics, 1-14.*

3.5.1.3 Experimental design and procedure

The experimental procedure was similar to the one described in Study 1. Participants were seated on the padded racing seat of the RT-Chair (Cuturi et al., 2020). The experimenter explained the task and gave the participants the headphones through which the auditory landmarks would have been presented. Each participant's head was aligned with the RT-Chair's rotation axis and leaned on a pillow, molded according to the shape of the participant's head. A padded strap kept their forehead still to the chair to reduce the use of neck proprioceptive cues as sources of information for orientation. During the experiment, the room was darkened; in addition, participants kept their eyes closed and covered by an eye mask to prevent any use of the irrelevant room's available visual information. The task was similar to the one described in Study 1. After clockwise rotations, using the numeric keypad, participants had to report whether they felt closer to the point of reference at azimuth 0° or to the point of reference at azimuth 90°. Likewise, after counterclockwise rotations, they had to report whether they felt closer to azimuth 0° or -90°. Participants' responses were therefore interpreted as the perceived middle amplitude between azimuth 0° and 90°, i.e., a yaw rotation of 45°. For example, if participants felt closer to azimuth 90° after a clockwise rotation, it meant they perceived a rotation wider than the middle physical amplitude between azimuth

0° and 90° (perceived rotation > 45°). Conversely, if participants felt closer to azimuth 0° after a clockwise rotation, it would mean that they perceived a rotation smaller than the middle physical amplitude between azimuth 0° and 90° (perceived rotation < 45°). To let participants experience the extreme points of reference at ±90°, before the experimental session, they were rotated with an amplitude of 90° four times, one for each level of the experimental design described in the following text.

The experimental design involved testing two conditions (Vestibular-only and Multisensory, in which vestibular and auditory cues were available in the same trial) and two movement directions (clockwise and counterclockwise). On each trial, before the movement, a brief high-pitch tone through the headphones worked as a "GO" signal and was lateralized according to the direction of the next rotation (e.g., a high-pitch tone in the left ear for counterclockwise rotations). After the "GO" signal sound, participants triggered the motion stimulus by pressing the start button on the keypad. In the Vestibular-only condition, right after the pressure of the start button, participants experienced a 3-s yaw rotation, and upon finishing, they used the buttons on the keypad to report their amplitude estimations as instructed. In the Multisensory condition, participants could rely on vestibular and auditory cues to estimate the rotations' amplitude. To ensure that participants were fully aware of the positions of the sounds, we showed them a visual outline of the spatial configuration of auditory landmarks (similar to Figure 3.6) before the experimental session. In this condition, after the pressure of the start button, two auditory landmarks were presented automatically and sequentially (the first sound was always the one at azimuth 0°); right after both sounds, the rotation began (Figure 3.7). When the rotation was completed, the auditory landmarks were delivered automatically again. In these trials, participants were explicitly instructed to use both vestibular and auditory cues to estimate their rotation. To simulate the fixed position of landmarks in the environment, the sounds were virtually rotated with the same amplitude as the just presented rotational stimulus but in the opposite direction. To mask the sounds elicited by the RT-Chair rotations, an additional white noise sound was played through headphones. For all conditions, right after the response, participants were brought back to the start position at azimuth 0° with a reduced frequency of the just-presented stimulus (0.25 Hz). To avoid any potential aftereffects between two consecutive movements (Crane, 2012b), a 3-s time window interleaved experimental motion stimuli. The conditions and rotation directions were randomized across trials for all participants.

For each level of the experimental design, we assessed 54 trials, of which the first four were training trials with fixed movement magnitude. For the remaining trials, we used the Psi adaptive procedure (Kontsevich and Tyler, 1999) to determine rotations amplitude,
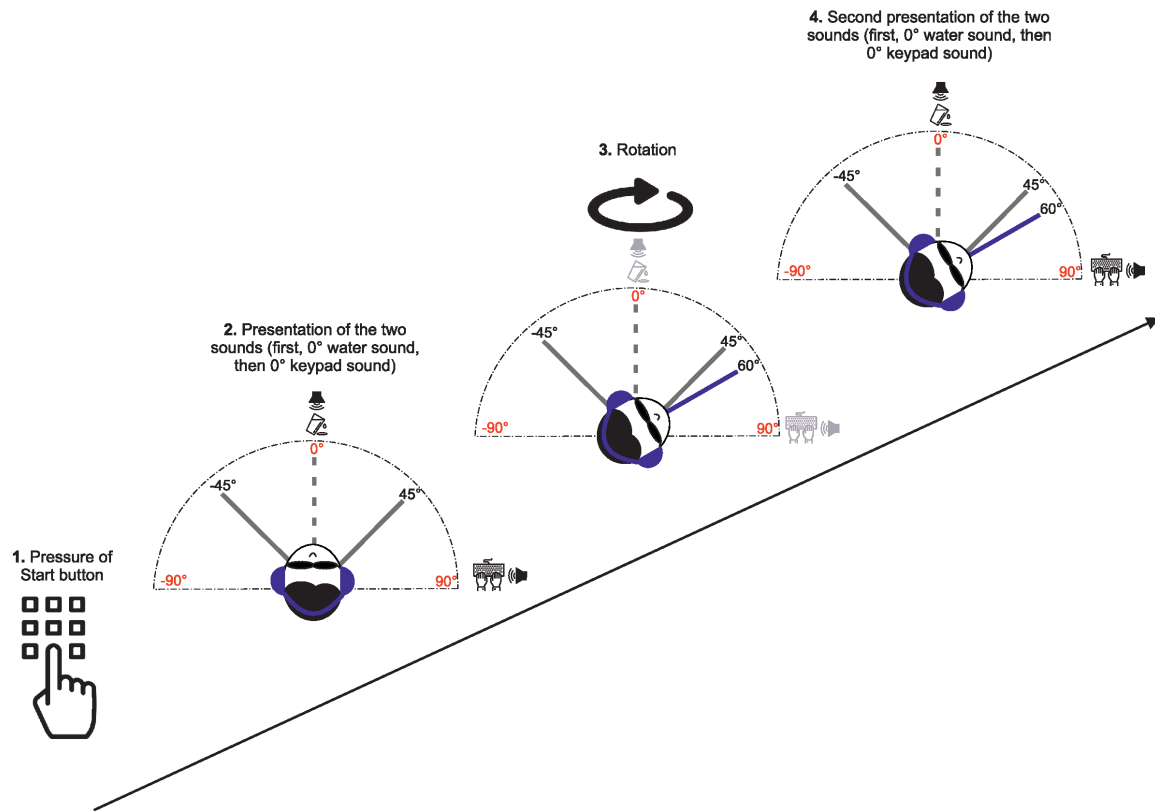
Figure 3.7 The experimental procedure in a Multisensory trial in Experiment 1. First, the participant pressed the start button (1); then, the two auditory landmarks were presented (2); afterward, the participant was rotated (3) and finally, the two sounds were played a second time (4). *Reproduced with permission by Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). How much I moved: Robust biases in self-rotation perception. Attention, Perception, and Psychophysics, 1-14.*

implementing the PAL_AMPM routine from the Palamedes toolbox (Prins and Kingdom, 2018) in Matlab (total number of trials = 216). Figure 3.8 shows an example of the Psi procedure for one participant and one condition (Vestibular-only, clockwise rotation). The whole experiment lasted 1 h and 30 min. Participants took two breaks at one-third and two-thirds of the experiment to prevent drowsiness.

3.5.1.4   Data analysis

Data analysis is similar to the one used in Study 1 as well. For clockwise rotations, we plotted the percentage of responses "I felt closer to azimuth 90°" as a function of the administered stimulus displacement (Figure 3.9, panel A). For counterclockwise rotations, we plotted the percentage of responses "I felt closer to azimuth 0°". For each participant, condition, and

direction of movement, we fitted a cumulative Gaussian to the data using the PAL_PFML_Fit routine from the Palamedes toolbox (Prins and Kingdom, 2018), which finds the best fit in a maximum likelihood sense (guess and lapse rate were fixed at 0.02). The mean provided a measure of the movement perceived as a 45° rotation (the middle amplitude between azimuth 0° and ±90°), which constituted the point of subjective equality (PSE). The standard deviation of the distribution represented a measurement of variability (the just noticeable difference (JND)) and a measure of the reliability of cues (the inverse of the variability corresponds to the reliability of each cue). We calculated the error of the estimates using a non-parametric bootstrap analysis, running the function PAL_PFML_BootstrapNonParametric, generating 400 simulated data sets (Prins, 2016). We then calculated the goodness of fit by using the PAL_PFML_GoodnessOfFit function in Matlab (Prins and Kingdom, 2018). In addition, to obtain a measure of the potential bias in amplitude estimations of rotations, we computed the difference between the unbiased amplitude of 45° and the absolute values of PSEs of each participant (bias = 45-|PSE|). Since the PSE is a measure of the movement perceived as a 45° rotation, a PSE smaller than 45° meant an overestimation of 45°, while a PSE greater than 45° meant an underestimation of 45°. Thus, we interpreted positive bias as overestimation bias (see panel B of Figure 3.9 for an example) and negative values as underestimation bias.

Before statistical analyses, we looked for outliers for each condition and the direction of rotation on the bias and the JND. The outliers were defined as the values above the third quartile plus 1.5 times the interquartile range, and below the first quartile, minus 1.5 times the interquartile range. We excluded participants whose variable values met this definition. Specifically, regarding bias, this experiment excluded one participant as an outlier and thus included 15 subjects in the final analyses. Regarding JND, in this experiment, three subjects were excluded as outliers (13 subjects were included in the final analyses). We verified the normality of the distribution of the variables in each condition and direction with the Shapiro-Wilk tests. We performed multiple one-sample t-tests for each condition and direction to confirm whether the bias differed significantly from zero, correcting multiple tests using a Bonferroni correction. To look for differences among conditions, we conducted a repeated-measures ANOVA (using the function *ezANOVA* from the *ez* package in RStudio, version 3.6.2, 2019), with Condition and Direction as within variables (Vestibular-only vs. Multisensory, counterclockwise vs. clockwise). We reported generalized eta squared (η_G^2) as effect size. Probabilities were considered significant when lower than .05. When in the presence of a violation of the normality assumption, we conducted the corresponding permutation tests (*ezPerm* from the *ez* package).
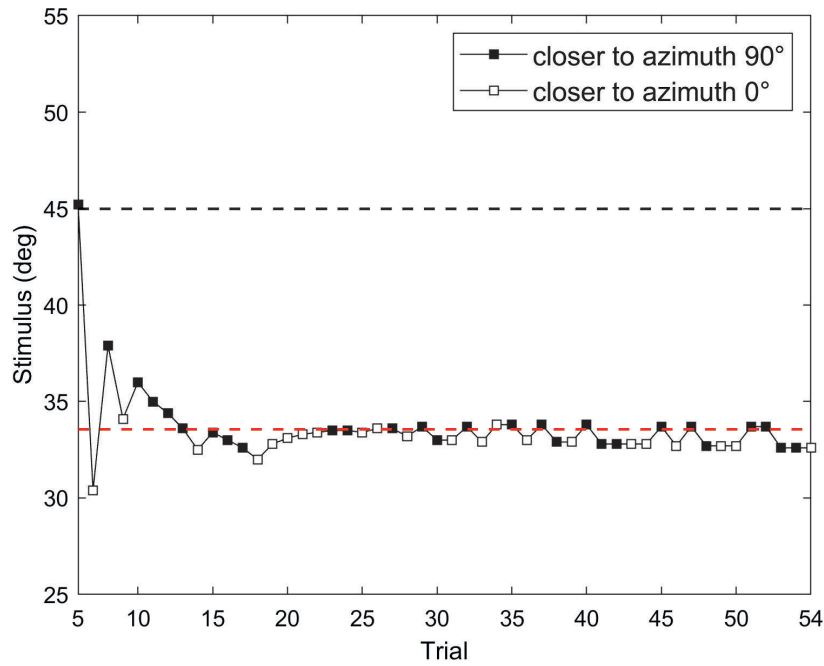
Figure 3.8 Trial history for the Psi adaptive procedure from one participant (Vestibular-only condition, clockwise rotation). Since the first four trials were training trials, only experimental ones are represented (trials from 5 to 54). The upper dashed grey line represents the investigated rotational amplitude at 45°. The lower dashed red line represents the mean of delivered motion stimuli following the Psi Procedure that approximates the point of subjective equality (PSE). White squares indicate the "closer to azimuth 0°" responses; black squares indicate the "closer to azimuth 90°" responses. *Reproduced with permission by Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). How much I moved: Robust biases in self-rotation perception. Attention, Perception, and Psychophysics, 1-14.*

3.5.1.5 Results

Panel A of Figure 3.10 represents the bias means in each condition. One-sample t-tests showed that bias significantly differed from zero in each condition and direction of movement, as shown in Table 3.2. In particular, in each condition and direction, participants overestimated the rotations (bias > 0), suggesting that they perceived rotations as wider than actuality. The repeated-measures ANOVA on bias values revealed no main effect of the factor Condition ($F(1,14) = 2.511$, $p = .135$, $\eta_G^2 = .008$, 95% CI [0.00, 0.26]), no main effect of the factor Direction ($F(1,14) = 0.326$, $p = .577$, $\eta_G^2 = .002$, [0.00, 0.20]), and no interaction ($F(1,14) = 1.155$, $p = .301$, $\eta_G^2 = .003$, [0.00, 0.21]). Panel B of Figure 3.10 shows the JND means. Given the violation of the normality assumption for JND, we conducted a permutation ANOVA on this variable, with 5000 permutations. It showed no main effect of
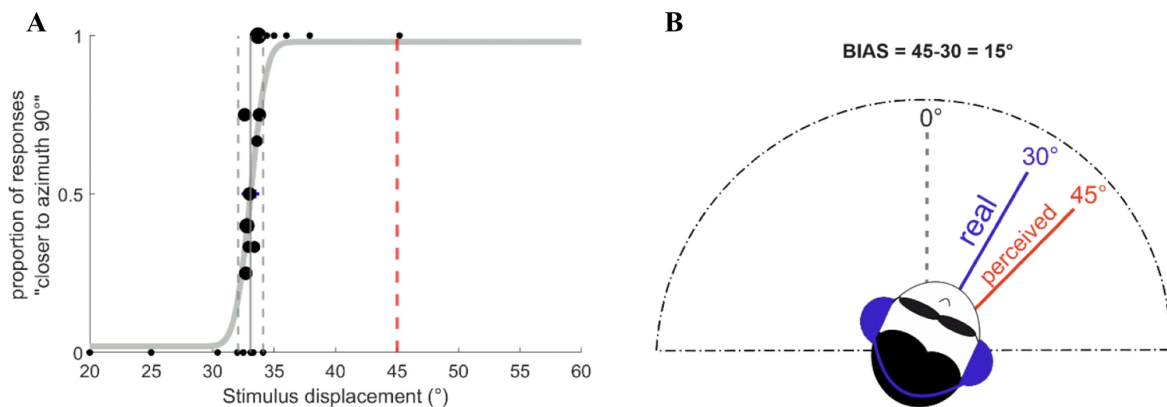
Figure 3.9 (A) Example of psychometric fit. Individual subject data from the Vestibular-only condition with clockwise rotations are represented. The vertical and dashed red line on the right indicates the unbiased estimate at 45°; the vertical solid grey line on the left indicates the point of subjective equality (PSE) (33°); the two vertical and grey dashed lines on either side of the solid one represent PSE ± the just noticeable difference (JND); black points are datapoints whose size is proportional to the number of presentations for that particular stimulus displacement. (B) Example of an overestimation bias. Here, the participant experiences a 30° clockwise rotation but perceives it as a 45° rotation, showing an overestimation bias of 15°. *Reproduced with permission by Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). How much I moved: Robust biases in self-rotation perception. Attention, Perception, and Psychophysics, 1-14.*

the factor Condition (p = .139, η_G^2 = 0.05, [0.00, 0.39]) and Direction (p = .907, η_G^2 < 0.001, [0.00, 0.04]), and no interaction (p = .997, η_G^2 < 0.001, [0.00, 0.00]).

Table 3.2 Experiment 1 one-sample t-tests on biases for each condition and direction of rotations.

Condition	T values	Degrees of Freedom	P values	Effect size (Cohen's d)
Vestibular, left	5.92	14	.000075	1.529
Vestibular, right	4.647	14	.000755	1.2
Multisensory, left	6.602	14	.000024	1.705
Multisensory, right	6.712	14	.000020	1.733

3.5.1.6　Interim discussion

Overall, the results on accuracy revealed a robust overestimation bias in participants' estimation of amplitude rotations, regardless of the type of condition or the direction of rotations. Variability, measured by JNDs, was similar in all conditions and directions.
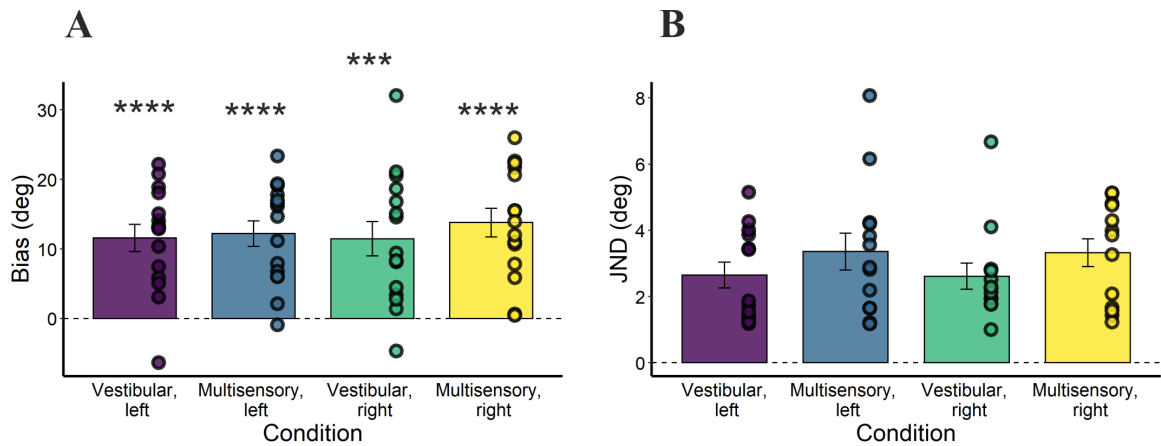
Figure 3.10 Bias (A) and just noticeable difference (JND) (B) of all participants in Experiment 1. As depicted, participants showed a strong overestimation bias (bias > 0) in all conditions, and their variability did not differ among conditions and rotation directions. Data points represent individual biases and JNDs; error bars are standard errors. **** = $p < 0.0001$; *** = $p < 0.001$. *Reproduced with permission by Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). How much I moved: Robust biases in self-rotation perception. Attention, Perception, and Psychophysics, 1-14.*

In our multisensory conditions, the auditory landmarks were presented automatically before and after rotations. It is plausible that the instant presentation of externally caused auditory cues hindered participants from grasping the spatial information conveyed by the auditory cues. Previous research showed improved processing of self-generated auditory cues than the processing of externally generated sounds (Myers et al., 2020). Likewise, actively manipulating auditory stimuli might help build a spatial map of auditory cues (Setti et al., 2018); according to a sensorimotor approach, the experience of the sensory consequences of voluntary actions aid in learning the spatial location of any sound source (Aytekin et al., 2008).

Thus, we performed a second experiment (Experiment 2), in which participants actively self-generated the auditory landmarks before and after the rotations. In this way, we attempted to exclude that the absence of interaction between acoustic and vestibular cues in Experiment 1 was due to the inability to actively explore the acoustic space around them and properly use auditory information for orientation.

3.5.2 Experiment 2: Self-generated auditory cues for self-motion perception

3.5.2.1 Participants

Sixteen healthy subjects (mean age $\pm$ SD: 25.6 $\pm$ 5.9 years) who did not participate in Experiment 1 participated in Experiment 2.

3.5.2.2 Apparatus and stimuli

The vestibular stimuli delivered were the same as in Experiment 1. Since we found no difference between directions of rotation in Experiment 1, we administered only clockwise rotations in Experiment 2. Still, to disrupt any potential habituation to clockwise rotations, we added ten counterclockwise rotations for each condition as catch trials, randomizing them across trials. Their amplitude was fixed, and we chose it from among the following values: -20°,-35°,-40°,-43°,-44°,-45°,-46°,-49°,-54°, and - 69°. During catch trials, participants performed the same task requested for the experimental trials. In the case of catch trials in the Multisensory condition, we presented the same landmarks as we did in counterclockwise rotations in Experiment 1: water sound at azimuth 0° and copy machine sound at azimuth -90°.

3.5.2.3 Procedure

As stated in the Apparatus and stimuli section (3.5.2.2), here we delivered only clockwise experimental motion stimuli. The same discrimination task of Experiment 1 to estimate the amplitude of rotations was used. The study tested two different conditions: Vestibular-only condition, in which participants estimated their movement's amplitude by exclusively relying on the vestibular cue, and Multisensory condition, in which they pressed a button on a keypad to generate and listened to the two auditory landmarks before and after the rotation. Specifically, before each trial, a registered voice indicated the type of condition through headphones: the Italian word for "Rotation" (Rotazione) indicated a Vestibular-only trial, while the Italian word for "Auditory" (Audio) suggested a Multisensory trial. In this way, participants knew when to press the button to explore the acoustic environment and when to press the button to only trigger the rotation. After listening to both sounds in the Multisensory condition, the rotation occurred automatically. After the rotation, participants needed to press the button again to listen to the auditory landmarks. Similar to Experiment 1, the landmarks were virtually rotated after the rotation with an amplitude equal to the presented movement

stimulus but in the opposite direction. Like in Experiment 1, participants completed 54 trials in each condition, of which the first four were training trials with fixed movement magnitude. The amplitude of other trials' rotations was determined using the Psi adaptive procedure (Kontsevich and Tyler, 1999) by means of the PAL_AMPM routine from the Palamedes toolbox (Prins and Kingdom, 2018) in Matlab (R2019b, The Mathworks, United States). Responses for the ten catch trials were not included in the analyses (total number of trials = 128) since they only acted to prevent habituation to clockwise movements. This experiment lasted about 1 hour.

3.5.2.4 Data analysis

We performed the same fitting and analyses done in Experiment 1. Considering that there was only one direction of rotation, we conducted paired t-tests on bias and JND values to compare Vestibular-only and Multisensory conditions. For the data analysis of this experiment, we excluded one subject from the dataset because of a technical issue with the 3DTI toolkit during the experiment. Besides, the study excluded one subject owing to an extremely poor fit in one condition (successful simulations = 204 out of 400), and another because of an apparent change of response strategy during the experimental session that severely influenced PSE and JND variables. We then defined outliers as done for Experiment 1. For bias values, two further participants were excluded as outliers; 11 subjects were included in the final analyses. Regarding JND values, one subject was excluded as an outlier (12 subjects were included in the final analyses).

3.5.2.5 Results

Panel A of Figure 3.11 displays the biases. One-sample t-tests showed that bias was again significantly different from zero in each condition (see Table 3.3). In each condition, participants overestimated the rotations (bias > 0). In other words, they perceived rotations as wider than they were. The paired t-test on bias revealed no significant difference between Vestibular-only and Multisensory conditions ($t(10) = -2.04$, $p = .07$, Cohen's d = 0.62, 95% CI [0.05, 1.74]). As depicted in Panel B of Figure 3.11, the paired t-test on JND values revealed no significant difference between Vestibular-only and Multisensory conditions ($t(11) = 1.19$, $p = .26$, Cohen's d = -0.35, [-1.24, 0.22]).

Table 3.3 Experiment 2 one-sample t-tests on biases for each condition.

Condition	T values	Degrees of Freedom	P values	Effect size (Cohen's d)
Vestibular, right	7.782	10	.00000749	2.346
Multisensory, right	12.071	10	.00000028	3.64
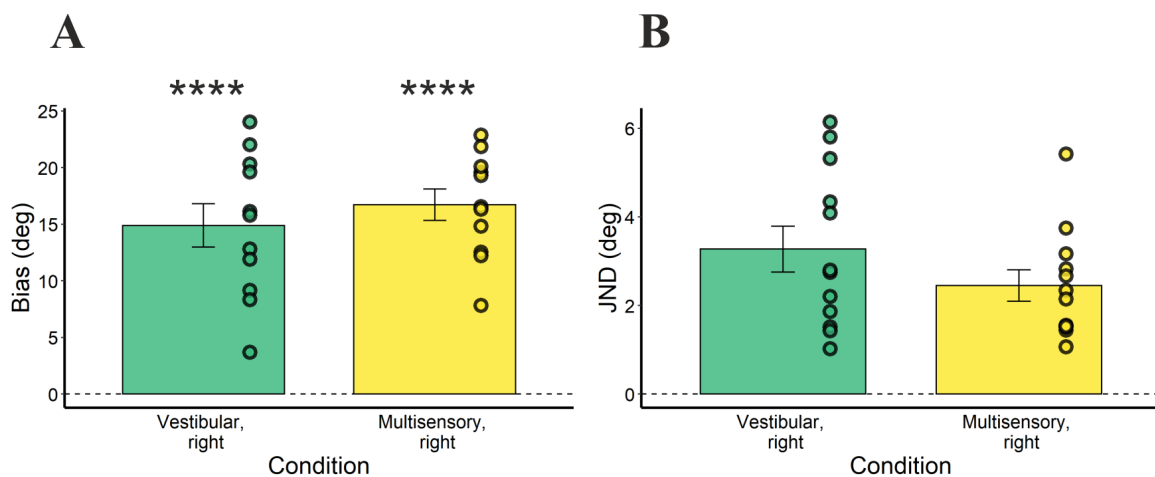
Figure 3.11 Bias (A) and just noticeable difference (JND) (B) of all participants in Experiment 2. As depicted, participants showed a strong overestimation bias (bias > 0) in all conditions, and their variability did not differ between conditions. Data points represent individual biases and JNDs; error bars are standard errors. **** = p < .0001. *Reproduced with permission by Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022). How much I moved: Robust biases in self-rotation perception. Attention, Perception, and Psychophysics, 1-14.*

3.5.2.6   Interim discussion

Taken together, the results of Experiment 2 again showed a strong overestimation bias in participants' estimation of the amplitude of the rotations, both with and without auditory landmarks available. Again, variability measured with JNDs did not differ between conditions.

3.5.3   Conclusion

The results in Experiments 1 and 2 revealed a robust overestimation bias in rotational displacement perception. Participants indeed perceived rotations wider than they were. The bias was resistant to the influence of spatialized auditory cues in both experiments, showing no difference between the use of self- or externally generated auditory landmarks.

This study had the aim to investigate the estimation ability of the amplitude of passive rotational self-displacement by relying only on vestibular information or on vestibular and additional auditory cues, which could be automatically presented or self-generated by participants. Specifically, the goal of Experiment 1 was to probe the human ability of self-motion estimation, with and without auditory cues. The purpose of Experiment 2 was to verify whether the expected perceptual bias in self-motion perception would be influenced by the self-generated auditory cues if compared with Experiment 1's spontaneous sound administration. Both experiments showed that participants had a significant overestimation bias, meaning that they perceived rotations to be wider than actuality. The amplitude and the presence of the bias were similar regardless of the availability of auditory cues. In addition, in Experiment 2, the results suggested that participants' manipulation of the auditory landmarks had no influence on using the spatial information provided.

Overestimaton bias

Our results revealed a robust bias in estimating the amplitude of experienced whole-body yaw rotations without visual information. A theoretic explanation of the observed bias could stem from the literature regarding heading estimation, which highlights overestimation biases in the estimation of heading directions. This apparently dysfunctioning perception is actually interpreted as the way in which the brain responds to deviations from the most common movements, likely preparing the organism to react optimally to changes. This topic will be addressed more extensively in Chapter 5. Importantly, in our study, we delivered 0.33-Hz rotational stimuli that are perceived with lower sensitivity by the vestibular system than higher-frequency movements (Grabherr et al., 2008). In other words, precision in discriminating low-frequency stimuli is poor. According to the Bayesian perspective, the perception of a stimulus is based on the combination of prior information (previous knowledge and experiences) and the available sensory information encoded by our sensors (in our case, the rotational self-motion sensations detected by the semicircular canals in the vestibular system). When sensory information is noisy and thus poorly reliable, the corresponding likelihood distribution will be more variable; thus, the resultant posterior probability will be pulled toward the prior distribution. Therefore, if, in our case, the prior is a preference for the straight-ahead direction, it is plausible to predict that the poorer the sensory signal in response to the rotational stimulus, the more the perceptual bias should increase. We can then speculate that, by increasing the frequency of the rotations, thus increasing their reliability (Grabherr et al., 2008; Valko et al., 2012), the bias should decrease or even disappear.

Vestibular dominance over hearing

We found that static auditory cues fixed in the environment did not impact self-motion perception accuracy. In the context of audio-vestibular integration (Shayman et al., 2020), vestibular and auditory information exhibit comparable reliability when the frequency of movements is between 0.2 and 0.5 Hz – the range of frequency of yaw rotations we provided in our current experiments (0.33 Hz). Nonetheless, our results on precision (i.e., the measured JNDs) showed no precision improvement in multisensory conditions. This suggests that vestibular information may have been more reliable than auditory, leading to its predominance. Importantly, the observed vestibular dominance and the difference between our study and the previous results found by Shayman et al. (2020) can be explained by one important feature of our experimental design. In both our experiments during Multisensory conditions, auditory landmarks were presented as static cues before and after the rotational stimuli but never simultaneously with the vestibular inputs. Conversely, in Shayman and colleagues' study (2020), the auditory cues were provided during the whole duration of the displacements. Arguably, not allowing the simultaneous experience of auditory and vestibular cues reduced the possibility for these sensory systems to interact. Considering what has been reported in previous literature (Shayman et al., 2020), it is plausible that keeping the sound throughout the movement could increase the relevance of the spatial information conveyed by the continuous sounds. This modulation is likely to increase the use of the external environment to update the spatial representation of self-motion and thus reduce the vestibular-based overestimation bias. Still, considering the robustness of the presented bias, it is unlikely that continuous and dynamic sounds could completely eliminate it. Interestingly, in our studies we chose semantic sounds that have more impact on self-motion perception (Riecke, 2016). In a previous control experiment (done in Study 3.4) investigating sound localization performance, we found that the chosen semantic auditory cues were localized with high precision and no bias by all participants, suggesting that the features of the selected sounds did not negatively influence self-motion perception. Nonetheless, future studies may also be addressed to evaluate the potential effect of the spectral composition of sounds on the self-motion perception based on audio-vestibular inputs (i.e., using broadband noises).

In addition, we found a similar magnitude of bias between Experiments 1 and 2, indicating that the observed rotational biases dominated the presence of auditory landmarks. Previous literature reported that participants experienced improved processing of the sensory consequences of voluntary actions (Desantis et al., 2014; Gozli et al., 2016; Salomon et al., 2011). Still, the robust overestimation bias we have found persists regardless of the modality of the presentation of auditory stimuli. Allegedly, the spatial information conveyed by the

vestibular system has been perceived as more relevant than the spatialized auditory cues, demonstrating its essential role in functional self-motion perception.

This chapter is partially extracted and adapted from:

Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2021, November). Evaluation of a Motion Platform Combined with an Acoustic Virtual Reality Tool: A Spatial Orientation Test in Sighted and Visually Impaired People. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 6078-6081). IEEE.

and

Zanchi, S., Cuturi, L. F., Sandini, G., and Gori, M. (2022b). How much I moved: Robust biases in self-rotation perception. Attention, Perception and Psychophysics, , 84(8), 2670-2683.

Chapter 4

Vestibular interaction with environmental spatial cues

Humans employ many sources of information to interact successfully with the environment. On the one hand, through our sensory systems, we can efficiently perceive the characteristics of external objects that compose the surroundings (e.g., their size, their position, their salience); on the other, we constantly collect information originating from our body traveling through space. Along these lines, an extensive amount of research acknowledges the fundamental role of the vestibular system in perceiving self-motion. As mentioned in the Introduction of this thesis, the vestibular system is composed of a set of sensory transducer organs that constantly detect the accelerations of our head in space. Crucially, vestibular signaling never shuts down, providing a reliable source of information for our brain. For example, otoliths in the inner ear continuously detect linear acceleration of our head, including gravity, the constant force on Earth under which humans live daily, which shapes our perception. In this sense, through graviception (i.e., the detection of gravity), the vestibular system allows the brain to keep perception anchored to our world.

Interestingly, the vestibular system contributes to several cognitive functions, such as spatial navigation, visuospatial memory, and spatial perception (Bigelow and Agrawal, 2015; Deshpande and Patla, 2005; Ferrè and Haggard, 2020; Ferrè et al., 2013; Hilliard et al., 2019). Indeed, several studies on humans and animals have demonstrated that any damage to the vestibular system leads to the inability to effectively explore the environment (Dallal et al., 2015; Xie et al., 2017). Bolstering the high-level functioning of the vestibular system, vestibular inputs are also strongly integrated with signals from other sensory modalities, such as vision, haptics, audition, and proprioception (Angelaki and Cullen, 2008). The convergence of multiple sensory cues has been described in almost all vestibular relays,

including the vestibular nuclei, the thalamus, and several areas in the cerebral cortex (Lopez and Blanke, 2011).

How the processing of vestibular signals and the processing of environmental cues specifically influence each other is still unclear. To answer this scientific question, in the present chapter, I specifically investigated with two studies how artificially stimulated vestibular signals may interfere with the perception of external cues and how external cues may, in turn, influence vestibular perception. Differently from studies described in Chapter 3, where I made use of a motion platform to passively move participants, requiring them to sit still in the motion simulator seat, here I exploited an artificial technique that directly stimulates the vestibular nerve, i.e., the galvanic vestibular stimulation (GVS). Specifically, GVS is defined as a simple yet effective and safe specific way to prompt vestibular reflexes (Fitzpatrick and Day, 2004), delivering a pure and selective disturbance at the receptor level. Briefly, GVS consists of applying direct current, delivered by a stimulator, by attaching two electrodes of different polarities (one anode and one cathode) to the mastoids behind the ears. The electrical current causes the person to perceive an illusory self-motion sensation, usually in the roll plane, and elicits cortical activation in the areas identified as the vestibular cortex (Bense et al., 2000; Brandt and Danek, 1994; Lobel et al., 1998). This technique reveals to be an advantageous resource to selectively focus on vestibular inputs during complex spatial navigation activities.

4.1   Study 1: Vestibular contribution to spatial encoding

Consider the first example of this thesis and put ourselves in Christopher Columbus and his crew's viewpoint. Imagine guiding the ship and finally seeing the so-much-yearned land. To dock successfully, the first step of your many-faceted goal would be to precisely compute the distance between your own position and the land.

Estimating the spatial relationship between objects and the position of our body in the environment is crucial for survival. Especially in unknown environments, it is essential to have a clear comprehension of the spatial relationship between our position and the objects in the surroundings, composed of either useful targets to be reached or potentially threatening targets to avoid. Goal-directed locomotion task, i.e., reaching a target by walking, requires first the successful encoding of the egocentric distance between the goal target and our starting position and then the updating of our position in space based on self-motion. In the present study, we focus on the first step of this complex activity, namely, the distance encoding of environmental objects.

Distance estimates are primarily based on visual cues (Sun et al., 2004). This might not be surprising, given that the high spatial acuity of the visual system is particularly important in representing spatial information (Alais and Burr, 2004). Therefore, while estimating objects' distance, navigators can represent the surroundings using static visual cues (i.e., visual cues available even when in the absence of body movements, such as convergence and binocular disparity) and dynamic visual cues (i.e., optic flow). Nonetheless, also other sensory modalities provide information about distance. In case of degraded or absent vision, auditory cues can signal essential distance information about environmental objects outside the navigator's perspective, even though less accurately than visual ones do (Kolarik et al., 2016; Nardi et al., 2020; Zahorik et al., 2005). Notably, to represent spatial information, visual and auditory signals constantly interact with each other and with self-motion cues, which can originate from proprioceptive and vestibular systems (Chen et al., 2017; Nardini et al., 2008; Werkhoven et al., 2014) (see also Chapter 2). However, it is still unclear how inertial cues may aid the perception of environmental objects, providing information about the distance between them and the body.

As previously mentioned, several studies are increasingly showing the involving of vestibular signals to higher cognitive functions that require complex spatial processing, such as complex spatial navigation activities (Xie et al., 2017; Yoder and Taube, 2014). Remarkably, the vestibular signals never stop providing information about the position of the head relative to gravity and the environment, and constantly communicate with auditory, visual, proprioceptive, and haptic cues (Angelaki and Cullen, 2008). This last aspect makes the vestibular system a key sensory system to investigate the interaction among multiple sensory cues. In the context of object distance perception, for instance, previous research demonstrated that additional vestibular cues could help process static auditory cues placed at different distances from the listener (Genzel et al., 2018). For these reasons, it is plausible to believe that the vestibular system may also play a role in the efficient perceptual estimation of the distance of environmental objects.

In the present study, we investigated whether vestibular cues contribute to the encoding of the egocentric distance between the body and an environmental object by using a spatial navigation task that requires participants to walk to the position of a previously learned target. To achieve this aim, we manipulated participants' vestibular signals during the distance encoding of targets using stochastic Galvanic Vestibular Stimulation (sGVS), which is known to interfere with vestibular signaling (Fraser et al., 2015; MacDougall et al., 2006; Moore et al., 2006). Further, we systematically explored whether the potential vestibular contribution to spatial encoding is selective for visual targets (Experiment 1) and whether it can also

influence acoustic target encoding (Experiment 2). Therefore, our study aims to understand how vestibular information helps efficient goal-directed spatial navigation and to identify how vestibular inputs play a role in the cognitive processes that allow for generating a coherent spatial estimation of the external environment.

4.1.1   Experiment 1: Spatial encoding of visual cues

4.1.1.1   Participants

Twenty healthy participants were recruited (mean age $\pm$ SD: 29.5 $\pm$ 5.84 years old) for Experiment 1. Nineteen participants were right-handed, as assessed using the Edinburgh handedness inventory (Oldfield, 1971). None of the participants had a history of neurological, vestibular, or psychiatric disorders. Informed consent was obtained before participation in the experiment. The research ethics committee of Birkbeck University of London (United Kingdom) approved the experimental protocol. The study adhered to the ethical standards of the Declaration of Helsinki.

4.1.1.2   Apparatus and stimuli

Stochastic galvanic vestibular stimulation. The vestibular stimulation was delivered using a commercial stimulator (Good Vibrations Engineering Ltd., Nobleton, Ontario, Canada) controlled by Matlab (R2019b, The MathWorks, United States). Carbon rubber electrodes (area 16 cm^2) covered with electrode gel were placed over the participants' mastoid processes and fixed with adhesive tape. The application area was first cleaned, then electrode gel was applied to reduce the impedance. Vestibular stimulation consisted of an alternating sum-of-sines voltage with dominant frequencies at 0.16, 0.32, 0.43, and 0.61 Hz, generating a stochastic stimulation (stochastic Galvanic Vestibular Stimulation, henceforth sGVS). The maximum intensity was 1mA, and the total duration was 10 seconds. These parameters were chosen to maximize the disruption of the vestibular signals (Fraser et al., 2015; MacDougall et al., 2006; Moore et al., 2006). Sham stimulation was also used to control for non-vestibular effects. Similarly to previous studies (Ferrè et al., 2013; Lopez et al., 2010), the electrodes for sham stimulation were placed on the left and right shoulders at the base of the neck, about 5 cm below the GVS electrodes. Notably, sham stimulation entails administering identical electrical stimulation to the GVS, yet it does not induce any vestibular reactions as it does not target the vestibular nerve.

   Visual targets. The visual targets consisted of three LED lights, placed on the floor along one of the two longest sides of the experimental room (2.31 X 3.91 m^2). From the start

position, the closest light was at a distance of 2.26 m, the middle one was at 2.53 m, and the furthest one was at 2.87 m. To reduce light reflection on the walls, each light was dimmed by an opaque black piece of nylon. The experimenter controlled when and which visual target would be switched on and off by pressing a key on the experimental computer (Dell Latitude 5300 with Intel Core i5- 8265U central processor at 1.60 GHz; 64-bit Windows 10 Enterprise Version 18362) keyboard, which remotely controlled the light sources via Matlab (R2019b, The MathWorks, United States).

4.1.1.3   Experimental design and procedure

The experimental procedure is depicted in Figure 4.1. Verbal instructions about the task were given to participants before the experimental session. Participants wore a backpack where the vestibular stimulator was placed, and they were allowed to briefly familiarize themselves with sGVS-induced sensation before the actual task. The experimental room was darkened. Participants stood behind a marked line on the floor along one of the two longest sides of the room (starting position). To reduce irrelevant environmental visual cues, they wore sunglasses for the whole experiment. While standing still at the starting position, participants received 10 s of sGVS or sham stimulation. During the last 500 ms of the stimulation, one of the visual targets was switched on. When both the visual target and the stimulation went off, participants were incited to reach the position of the just learned target by walking. Participants were instructed to align their heels and back with the target's remembered position. The experimenter measured their distance with a laser meter (RS PRO RSLDM-50H, precision ±1.5 mm, Class 2, 1268821, RS PRO, RSLDM-50H). After the measure, participants were invited to go back to the starting position for the next trial. The experimental session was divided into two blocks, each with 15 trials. In each block, only sGVS trials or sham trials were administered. The order of stimulation type was counterbalanced across participants, while the distance of the targets on each trial was randomized across each block (5 repetitions for each distance). The whole experiment lasted 45 minutes.

4.1.1.4   Data analysis

For each participant and each trial, we computed the difference between the actual distances of targets and the walked distance. First, we checked that participants actually perceived and processed the three different distances, simply measuring their walked distances for each target. Then, we calculated the mean error and the standard deviation (SD) of their

Figure 4.1 Experimental procedure. In the encoding phase, participants were instructed to focus on the switched-on target while stimulated with sGVS or SHAM. In the walking phase, participants had to reach the estimated position of the target (stimulation off). Then, the experimenter measured their performance in target localization. *The avatar representing the participants has been rendered in DAZ Studio 4.21 (DAZ productions, http://www.daz3d.com/).*

responses. Specifically, the error consists of the difference between the actual distances of targets and the measured distances walked by participants (absolute values). The mean error represents, therefore, a measure of accuracy (higher error means lower accuracy), and the SD is a measure of variability (higher variability means lower precision). Individual trials were filtered to remove errors that were extreme outliers in the distribution of all the

errors recorded for that condition. Extreme outliers were defined as values greater than the third quartile plus 3 times the interquartile range. Shapiro-Wilk tests were used to verify the normality distribution of the variables in each condition. We performed one-tailed paired t-tests to compare error and SD between the stimulation conditions (sGVS and SHAM), expecting higher error and SD for the interference of sGVS stimulation. Cohens'd was reported as the effect size measure. Moreover, we conducted repeated-measures ANOVA on each dependent variable considering the type of stimulation and target distances as within factors to understand whether a potential difference between conditions may be modulated by the targets' distances. In the case of normality assumption violation, we performed the corresponding permutation tests with *ezPerm* function from the *ez* package and *perm.t.test* from the *RVAideMemoire* package in RStudio (version 3.6.2, 2019).

4.1.1.5 Results

No individual trials met the given outlier definition; thus, no subject was excluded from the analysis. As expected, participants' walked distances increased as the target distance increased (Mean ± SD: Short = 1.93±0.21, Medium = 2.19±0.23; Long = 2.54±0.23), showing that they actually distinguished the three target distances. Regarding the error, given the violation of the normality assumption, we conducted a one-tailed permutation paired t-test on this variable, with 3000 permutations, comparing sGVS and SHAM. Results showed greater error in sGVS than in SHAM condition (p = .04, Cohens'd = 0.4, 95% CI [-0.03, 0.93]), (Figure 4.2, left panel). Regarding SD, the one-tailed paired t-test showed significantly higher variability in sGVS than in SHAM condition (t(19) = 2.209, p = .02, Cohens'd = 0.49, [0.11, 0.93]) (Figure 4.2, right panel). Generally, these results indicate worse accuracy and precision when participants' performance was disrupted with sGVS to encode visual targets. Permutation repeated-measures ANOVA on the error revealed a trend in the main effect of stimulation (p = .06, η^2_G = .02, 95% CI [0.00, 0.06]), but no main effect of target distance (p = .73, η^2_G = 0.001, [0.00, 0.00]) nor the interaction between the two (p = .60, η^2_G = 0.001, [0.00, 0.00]), thus indicating that the distance of targets did not influence the effect of stimulation on the accuracy performance. Conversely, regarding SD, we observed a main effect of stimulation (p = .04, η^2_G = .04, 95% CI [0.00, 0.07]), a main effect of target distance (p = .04, η^2_G = .05, [0.00, 0.05]), and an interaction between the two factors (p = .03, η^2_G = .03, [0.00, 0.07]). Further exploring this interaction, post-hoc tests with Bonferroni correction revealed that only in the longest distance of the target, sGVS variability was higher than in the SHAM condition (t(19) = 2.69, p adj = .04, Cohen's d = 0.60, 95% CI [0.26, 1.03]) (see Figure 4.3).

Exp. 1: Visual targets
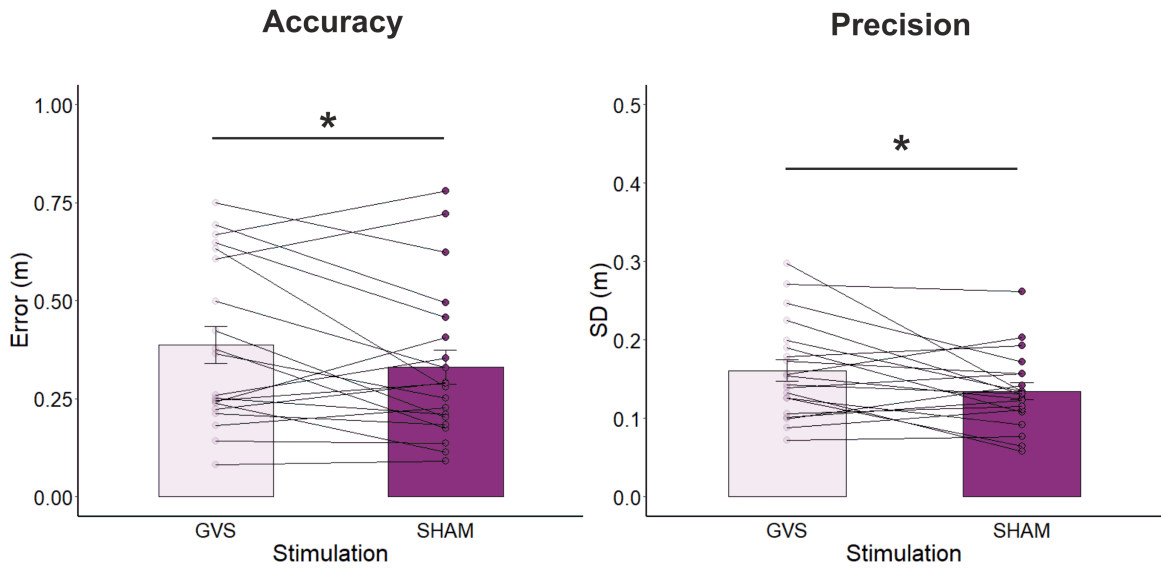
Figure 4.2 Results on accuracy (error, left panel) and precision (SD, right panel). Error and variability show to be systematically higher (meaning lower accuracy and precision) in sGVS condition than in SHAM condition. Individual points represent individual error and SD. The lines connect each participant's performance in the two stimulation conditions in each panel. Error bars are standard errors. * = p < .05.

Looking at the individual points shown in Figure 4.2, we observed high inter-subject variability. We decided to investigate it in an exploratory manner, taking into consideration the potential influence of the order of blocks experienced by each subject. As described in the experimental design and procedure section, we counterbalanced the order of blocks so that half participants began with the sGVS block ("GVS first" group) and the other half with the SHAM block ("SHAM first" group). We then attempted to look into sGVS and SHAM differences for the two groups taken separately. Given the violation of the normality assumption, a permutation mixed ANOVA with the order of blocks as a between factor and stimulation as a within factor was performed on error and SD. Regarding the error, the ANOVA revealed a main effect of stimulation (p = .04, $\eta_G^2 = 0.02$, 95% CI [0.00, 0.17]) and interaction between the order of blocks and the stimulation (p = .004, $\eta_G^2 = 0.04$, [0.00, 0.27]). Regarding the SD, the ANOVA revealed a main effect of the order (p = .03, $\eta_G^2 = 0.24$, 95% CI [0.04, 0.46]), a main effect of the stimulation (p = .02, $\eta_G^2 = 0.08$, [0.00, 0.17]), and a significant interaction between the two (p = .003, $\eta_G^2 = 0.07$, [0.00, 0.27]). Thus, we conducted permutation t-tests to disentangle the observed interactions with Bonferroni correction and computed Hedges's g as effect size due to the small sample size of each

Exp. 1: Visual targets
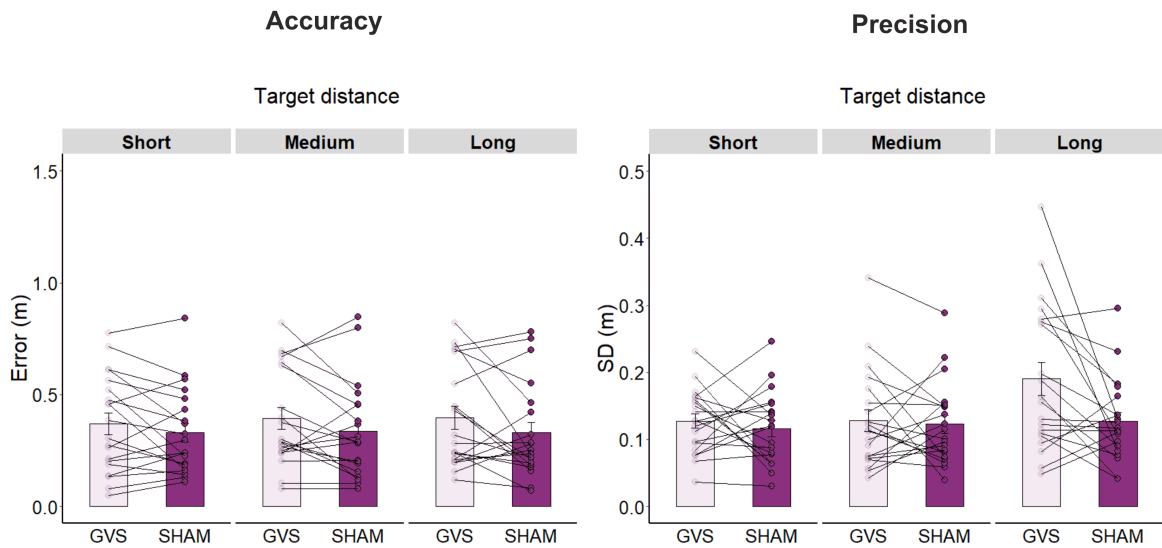
Figure 4.3 Results on accuracy (error, left panel) and precision (SD, right panel) for each target distance. Error shows a trend to be systematically higher (meaning lower accuracy) in sGVS condition than in SHAM condition. Variability is higher in sGVS than in SHAM only in the "long" target distance. Individual points represent individual error and SD. The lines connect each participant's performance in the two stimulation conditions in each panel. Error bars are standard errors.

group (n = 10). Specifically, we conducted planned comparisons between sGVS and SHAM stimulations. Permutation t-test on error comparing sGVS and SHAM in the "GVS first" group showed a significantly higher error in sGVS than in SHAM (t = 3.47, p adj = .01, Hedges' g = 0.63, 95% CI [0.21, 1.04]). At the same time, no difference was found in the "SHAM first" group (t = -0.84, p adj = .84, Hedges' g = -0.12, [-0.43, 0.18]). Regarding SD, permutation t-tests revealed a trend of higher SD in sGVS than in SHAM in the "GVS first" group (t = 2.65, p adj = .05, Hedges' g = 0.77, [0.07, 1.46]); no difference was found between the stimulations in the "SHAM first" group (t = 0.19, p adj = .85, Hedges' g = 0.06, [-0.54, 0.66]). This analysis suggests a potential influence of the order of received stimulation on the ability to encode visual targets so that receiving sGVS first might lead to worse performance with sGVS than SHAM (see Figure 4.4). However, the exploratory nature of this data analysis and the poor sample size of each group do not allow us to support this effect on the stimulation strongly. Further studies are needed to investigate this aspect more deeply.

Exp. 1: Visual targets
Stimulation and order of blocks
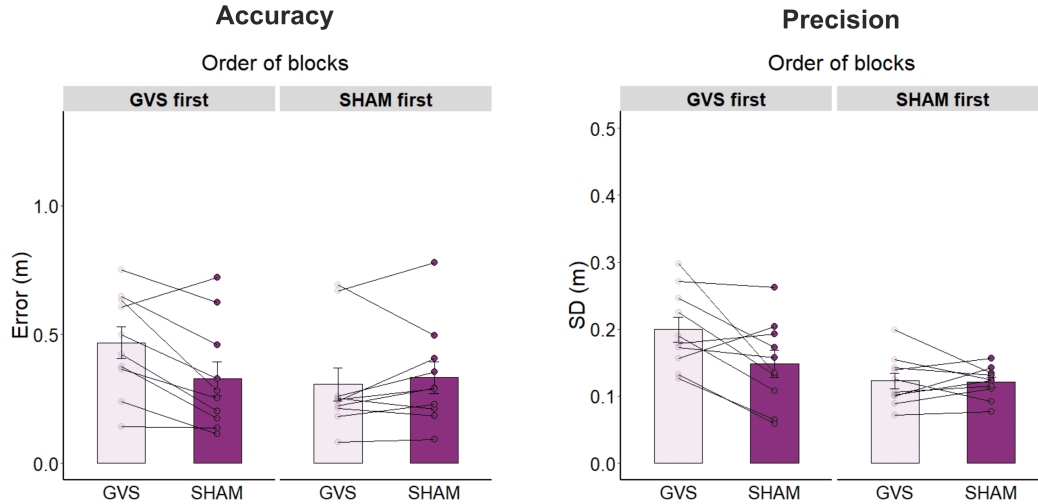
Figure 4.4 Results on accuracy (error, left panel) and precision (SD, right panel). The accuracy performance is split into "GVS first" and "SHAM first" groups. In "GVS first" group, error is greater in GVS than in SHAM stimulation conditions. No difference was found between conditions in the "SHAM first" group. Similarly, the precision is split into the two groups. Numerically, we observed higher SD in sGVS condition than in SHAM, even though the difference was only a trend. Points are individual errors and SDs. Error bars are standard errors.

4.1.1.6 Interim discussion

The disruption of vestibular signals on the space encoding of environmental visual targets resulted in the worsening of the participants' navigational performance. Specifically, we observed a systematically lower accuracy when sGVS impacted vestibular signals. Besides, variability remained similar across conditions, with the exception of the longest distance of trials, likely the most difficult condition. Therefore, the obtained results demonstrate a contribution of the vestibular system in localizing visual targets traveling in an unexplored environment, showing a functional interaction between vestibular and visual inputs during the initial encoding of spatial cues. Still, it remains unclear whether the vestibular system plays a role in the general estimation of the spatial distance of targets or whether this contribution is sensory-specific for visually encoded ones. For this reason, we decided to test this vestibular contribution in a spatial domain different from the visual one. Thus, we performed a second experiment (Experiment 2) in which vestibular signals were manipulated during the spatial encoding of acoustic targets.

4.1.2   Experiment 2: Spatial encoding of auditory cues

4.1.2.1   Participants

For Experiment 2, another group of twenty healthy participants was recruited (mean age $\pm$ SD: 29 ± 6.7 years old). Seventeen participants were right-handed as assessed using the Edinburgh handedness inventory (Oldfield, 1971). The inclusion criteria were the same as in Experiment 1. Informed consent was obtained before participation in the experiment. The research ethics committee of Birkbeck University of London (United Kingdom) approved the experimental protocol. The study adhered to the ethical standards of the Declaration of Helsinki.

4.1.2.2   Apparatus and stimuli

Stochastic galvanic vestibular stimulation. The sGVS parameters were identical to the ones used in Experiment 1.

   Acoustic targets. The acoustic targets of Experiment 2 consisted of three loudspeakers (Stilgut YB202STGD, 6.6 X 6.6 X 6 cm), lying on the floor along one of the two longest sides of the experimental room. The speakers were connected with jack cables to the experimental computer through an external sound card (Xonar U5, Asus); specifically, each speaker was connected to a separate channel. According to the procedure, each loudspeaker played a 500 ms pink noise sound. From the start position, the closest loudspeaker was at a distance of 1.96 m, the middle one was at 2.53 m, and the furthest one was at 3.24 m. Notably, since localizing auditory stimuli is more challenging than localizing visual ones (Zahorik et al., 2005), we moved the loudspeakers apart relative to visual targets in Experiment 1, to facilitate the discrimination of the three distances. Similarly to Experiment 1, the experimenter controlled when and which auditory target would be switched on and off by pressing a key on the experimental computer keyboard, which remotely controlled the speakers via Matlab (R2019b, The MathWorks, United States).

4.1.2.3   Experimental design and procedure

The experimental setup and procedure were similar to Experiment 1. Thus, after receiving instructions and familiarization with the stimulations, participants received 10 seconds of sGVS or sham stimulation while standing still at the starting position. During the last 500 ms of the stimulation, one of the acoustic targets played. When both the acoustic target and the stimulation went off, participants were incited to reach the position of the just learned

target by walking. Participants were instructed to align their heels and back with the target's remembered position. Data were analyzed as in Experiment 1.

4.1.2.4 Results

No individual trials met the given outlier definition, thus no subject was excluded from the analysis. Even though distinguishing different sound sources in depth is a highly difficult task, participants' walked distances increased as the target distance increased (Mean $\pm$ SD: Short = 2.17$\pm$0.35, Medium = 2.31$\pm$0.41; Long = 2.56$\pm$0.35), suggesting that they perceived and processed the three different target distances. Regarding the error, given the violation of the normality assumption, we conducted a one-tailed permutation paired t-test on this variable, with 3000 permutations, comparing sGVS and SHAM. Results on error showed no significant difference between the two stimulations (p = .84, Cohen's d = -0.23, 95% CI [-0.78, 0.21]) (see Figure 4.5, left panel). Similarly, regarding SD, the permutation paired t-test showed no difference between the two stimulations (p = 0.38, Cohen's d = 0.07, [-0.41, 0.49]) (Figure 4.5, right panel). Generally, these results indicate similar accuracy and precision in both stimulation conditions, most likely indicating no disruption of the encoding of acoustic targets when in the presence of sGVS. Permutation repeated-measures ANOVA on the error revealed no main effect of the stimulation (p = .34, η_G^2 = 0.001, 95% CI [0.00, 0.05]), nor the interaction between the two (p = .22, η_G^2 = 0.01, 95% CI [0.00, 0.04]), but we observed a main effect of the target distance (p = 0.003, η_G^2 = 0.16, 95% CI [0.00, 0.09]). Similarly, regarding SD, we observed no main effect of stimulation (p = .20, η_G^2 = 0.02, 95% CI [0.00, 0.01]), and no interaction (p = .29, η_G^2 = 0.01, 95% CI [0.00, 0.02]), but we observed a main effect of target distance (p = .02, η_G^2 = 0.05, 95% CI [0.00, 0.06]). Pairwise permutation paired t-tests (with Bonferroni correction for multiple comparisons) indicated that, regardless of the stimulation conditions, error was higher for the long distance of target if compared with short (t = -3.62, p = .006, Cohen's d = -0.57, 95% CI [-0.93, -0.27]) and medium distances (t = -5.56, p = .006, Cohen's d = -0.88, [-1.34, -0.52]). Likewise, we observed a higher SD in the long distances when compared to the medium distance (t = -8.62, p = .002, Cohen's d = -0.58, [-0.85, -0.4]) (see Figure 4.6).

Similarly to what was done in Experiment 1 analyses, we attempted to look into the inter-subject differences in performance, conducting the comparisons between sGVS and SHAM separately on "GVS first" and "SHAM first" groups. Given the violation of the normality assumption, a permutation mixed ANOVA with the order of blocks as a between factor and stimulation as a within factor was performed on error and SD. Regarding the error, the ANOVA revealed no main effect of stimulation (p = .32, η_G^2 = 0.003, 95% CI [0.00,

Exp. 2: Acoustic targets
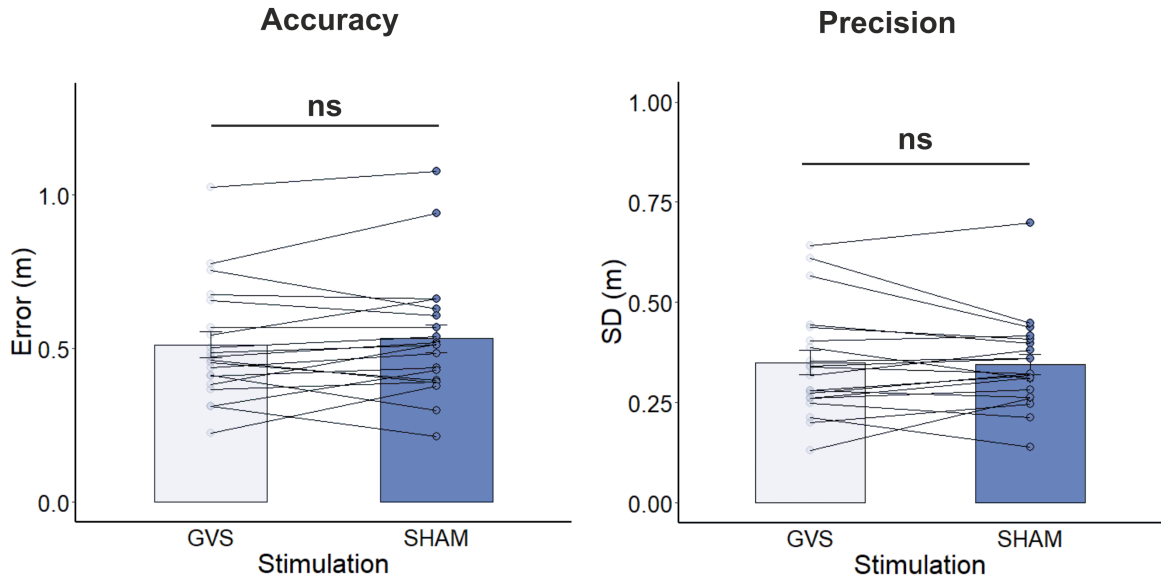
Figure 4.5 Results on accuracy (error, left panel) and precision (SD, right panel). Both error and SD did not differ between sGVS and SHAM conditions. Individual points represent individual error and SD. The lines connect each participant's performance in the two stimulation conditions in each panel. Error bars are standard errors. ns = not significant.

0.08]), no main effect of the order (p = .97, $\eta_G^2 < 0.001$, [0.00, 0.10]) and no significant interaction between the order of blocks and the stimulation (p = .32, $\eta_G^2 = 0.003$, [0.00, 0.13]). The same pattern of results was found for the SD. The ANOVA revealed indeed no main effect of stimulation (p = .23, $\eta_G^2 = 0.001$, [0.00, 0.08]), no main effect of the order (p = .49, $\eta_G^2 < 0.001$, [0.00, 0.10]) and no significant interaction between the two factors (p = .64, $\eta_G^2 = 0.004$, [0.00, 0.13]). Therefore, in this experiment, sGVS seemed to have no impact on participants performance in encoding and localizing acoustic targets.

To further investigate whether the disruption of the vestibular inputs on spatial encoding is sensory-specific, we computed an index of the impact of sGVS, namely the difference between sGVS and sham conditions in visual vs. auditory error and variability. Index computed for error yielded a significant difference between acoustic and visual targets' reaching performance (t(38) = 2.07, p = .04, Cohen's d = 0.65, 95% CI [0.06, 1.33]; left panel of Figure 4.7). Conversely, the index computed for variability showed a similar trend numerically but no significant difference between the two experiments (t(38) = 1.10, p = .28, Cohen's d = 0.35, [-0.27, 1.03]; right panel of Figure 4.7). Likely, the high variability found in Experiment 2 may explain the lack of significance between the two indexes on precision.

Exp. 2: Acoustic targets

Accuracy   **Precision**
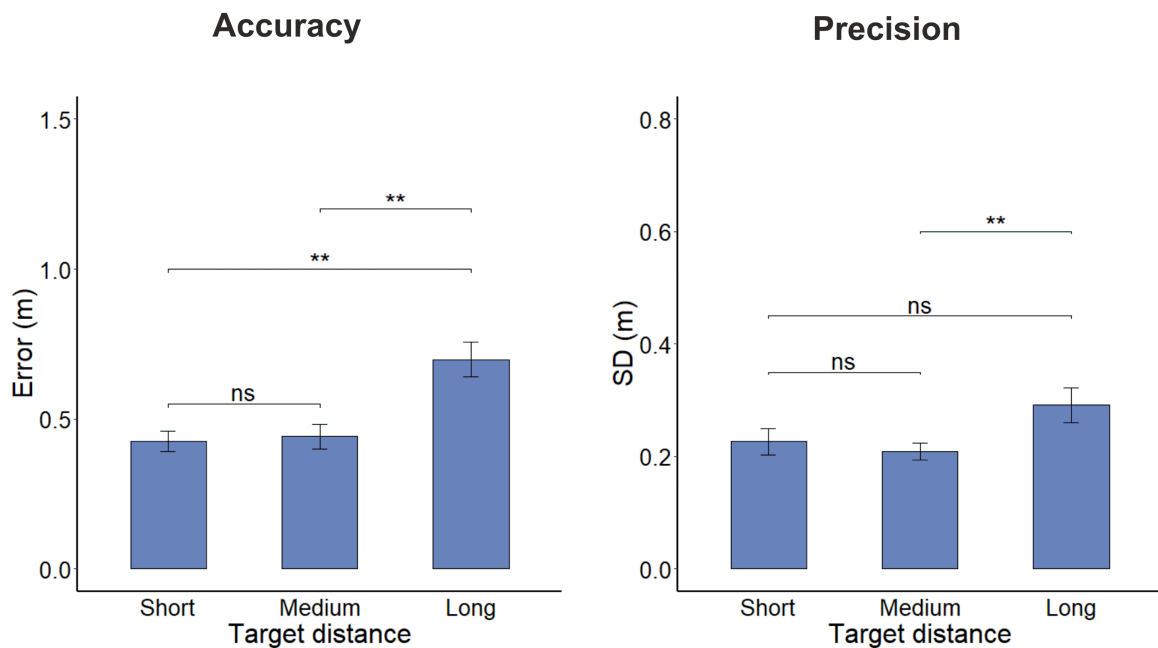
Figure 4.6 Results on accuracy (error, left panel) and precision (SD, right panel) in each target distance, regardless of the stimulation condition. Error bars are standard errors. ** = p < .01, ns = not significant.

4.1.2.5   Interim discussion

Globally, in contrast to the pattern of results observed in the localization of visual targets, the results from Experiment 2 showed no significant effect of vestibular disturbance on the encoding of the spatial location of auditory cues. This finding likely suggests a potential a selective effect of sGVS on visual encoding.

4.1.3   Conclusion

The spatial encoding of environmental features is characterized by the interaction between different sensory cues. In this study, we investigated the contribution of vestibular signals to the encoding of visual and auditory targets. Our results demonstrated the role played by the vestibular inputs in the encoding of visual (Experiment 1) but not auditory (Experiment 2) targets to be reached by walking.

Index of GVS impact on localization
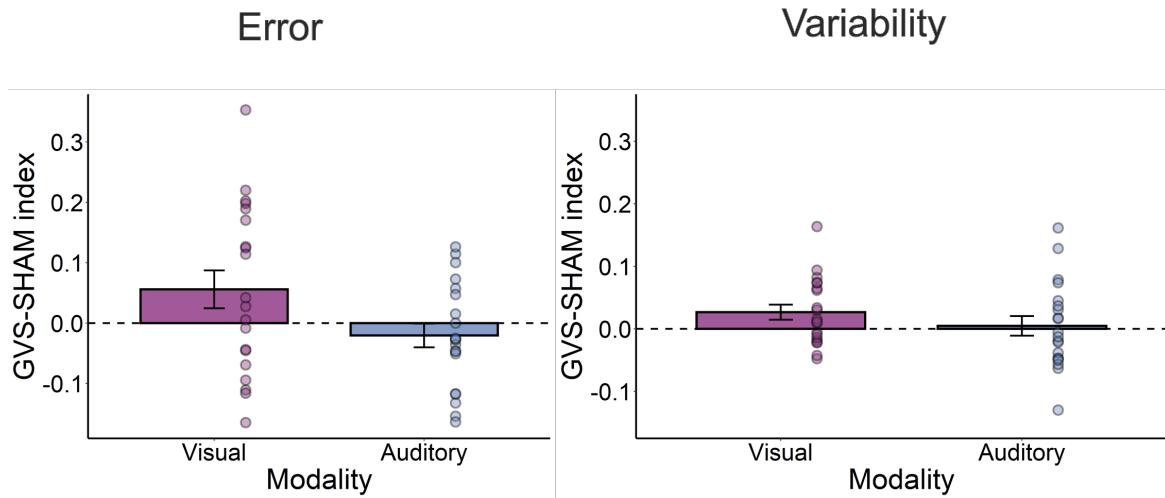
Figure 4.7 Indexes of GVS impact on localization measured for both experiments on accuracy (error, on the left) and precision (variability, on the right). Points represent individual indexes for each sensory modality. Error bars are standard errors.

sGVS interference on distance estimation

Our study highlights an essential contribution of the vestibular system to the preparation phase of goal-directed locomotion in the absence of inertial self-motion cues, namely the encoding of visually-encoded targets. Walking toward a previously observed target is a common behavior in our daily life. However, the performance in this task is influenced by several factors, such as visual perception, self-motion, calibration of locomotion and vision, and working memory (Rieser et al., 1990). Interestingly, in the present study, sGVS was applied in the "visual perception" phase, i.e. the moment in which navigators had to estimate the egocentric distance between their own position and the location of targets before actually walking. Put differently, no self-motion cues were strictly needed to compute the distance estimates. Nonetheless, we observed systematically worse accuracy (greater error) when participants were stimulated with sGVS than sham. Variability was conversely modulated only in the most challenging case, i.e., when the visual target was placed in the furthest position relative to the initial position of participants. This result highlights how vestibular inputs intervene in the spatial estimation of target locations in novel environments. Globally, the present finding suggests that even in the absence of actual movements, the vestibular system provides information about the spatial relationship between the head, the gravitational force, and the environment.

Visuo-vestibular vs. audio-vestibular interactions

In addition, our results indicate a sensory-specific interaction between visual and vestibular signals in spatial encoding. The interaction between visual and vestibular inputs has been extensively investigated both functionally and behaviourally. Previous research demonstrated that path integration navigation strategy (i.e., based on self-motion) is combined with remembered visual landmarks while navigating without visual information (Kalia et al., 2013). In addition, visual and vestibular cues are dynamically integrated during goal directed locomotion (Deshpande and Patla, 2005). Visual and vestibular cortical areas are shown to inhibit each other for spatial orientation and motion perception, representing a functional connection between the two sensory systems (Brandt et al., 2002). This cortical interaction is interpreted as the tentative to reduce or prevent interference in case of visual-vestibular mismatch, shifting to the most reliable sensory modality (Brandt et al., 2002). More specifically, in spatial contexts, dynamic visual information can cause aftereffects on self-motion estimation based on vestibular cues (Cuturi and MacNeilage, 2014), demonstrating the shared substrate of visual and vestibular processing (Gu et al., 2008). Here we showed an additional specific visual-vestibular interaction in the static encoding of environmental visual targets contributing to our ability to reach or avoid objects in space.

One might argue that the observed interaction between vestibular and visual inputs may be the consequence of the effect of sGVS on eye movements. Previous studies delivering a boxcar pulse of 1mA reported static ocular torsions in the direction of the anode (e.g., Zink et al. (1998)). Critically, a torsional eye movements effect would have impacted localization precision but not accuracy: conversely, we observed a systematic disruption of performance accuracy, while the precision was modulated only in the most difficult condition. Further, there is no published evidence that stochastic galvanic vestibular stimulation, as the stimulation used in this study, elicits ocular responses. However, future studies may directly investigate this aspect by measuring eye movements during sGVS and targets encoding.

No effect of sGVS was found on spatially encoding acoustic targets. Experiment 2 aimed to investigate the generalizability of the vestibular contribution to spatial encoding in a spatial domain different from the visual one. Auditory information is known to be less precise than visual cues for spatial tasks, especially in distance perception (Zahorik et al., 2005). However, auditory cues provide spatial information about objects outside the field of view when visual information is absent or less reliable (Kolarik et al., 2016; Nardi et al., 2020; Zahorik et al., 2005), allowing an egocentric representation of space of the body and the environment even in the absence of vision. Interestingly, our present results show that no interaction between

the two systems occurs in the context of spatial encoding, despite the peripheral and central connection between auditory and vestibular information (Smith, 2012).

4.2 Study 2: Vestibular-landmark interaction

Spatial navigation requires us to precisely perceive our position and the spatial relationships between our and environmental objects' location in space. During locomotion, self-motion perception is computed via the spatial relationship between environmental landmarks, mainly signalled by vision and hearing, and our body. These multiple concurrent sensory cues convey information that helps build a complex and coherent spatial representation of the world through which we move. In navigation and orientation contexts, prior studies have highlighted how different cues may be optimally integrated (Bates and Wolbers, 2014; Chen et al., 2017; Fetsch et al., 2009; Kaliuzhna et al., 2018; Nardini et al., 2008; Ramkhalawansingh et al., 2018; Sjolund et al., 2018; Zhao and Warren, 2015b) and influence each other, e.g., leading to cross-modal aftereffects (Cuturi and MacNeilage, 2014). Most studies focused on how vestibular and/or proprioceptive signals modulate the processing of external visual or acoustic cues. For instance, previous research showed that the ability to perceive self-motion by an optic flow might be modulated by concomitant vestibular inputs (Edwards et al., 2010; MacNeilage et al., 2012; Shirai and Ichihara, 2012), even though the extent to which this influence occurs is not homogeneous among studies. In addition, previous research found that vestibular cues improve navigation performance based on encoding proximal visual landmarks (Jabbari et al., 2021). Regarding acoustic cues, prior literature mainly showed a vestibular influence on central processes of auditory spatial coding (Genzel et al., 2018; Kim et al., 2013; Lewald and Karnath, 2000).

Less clear is how external cues may affect vestibular processing. It has been demonstrated that a long-lasting optic flow can influence vestibular cue processing in a self-motion discrimination task, leading to a robust heading bias (Cuturi and MacNeilage, 2014). This finding suggests that self-motion cues share a common neural substrate, regardless of the modality of the available cues. Another study (Gallagher et al., 2020) showed that the sensitivity to vestibular signals was reduced in the presence of optic flow stimuli delivered through Virtual Reality (VR), specifically when both visual and vestibular cues conveyed information on the same plane of self-motion (roll plane). This is interpreted as the reduction of the sensory weight associated with vestibular cues after exposure to spatially congruent visual self-motion stimuli. Put differently, the perception of the information conveyed by optic flow suggests that the felt self-motion sensations originate from visual cues rather than from

vestibular signals, thus leading to down-weighting vestibular inputs and then decreasing the sensitivity to them. This finding further supports the interaction between the two sensory systems in self-motion perception. Other than optic flow, we have access to fixed elements in the environment that represent stable points of reference on which rely while orienting, i.e., landmarks (Section 1.1.1). Landmarks provide direct information about the spatial features of the surroundings, allowing the navigators to build a metric representation of the world. While previous research has shown the advantageous influence of vestibular cues on the processing of landmarks (Jabbari et al., 2021), it is not clear yet the inverse relationship, i.e., whether landmarks can influence the processing of vestibular cues. We hypothesized that steady points of reference fixed in the environment might actually increase the sensitivity to vestibular inputs, helping disentangle the motion information coming from the body's movement and the spatial cues coming from the external world. Considering the crucial role of vestibular processing for efficient spatial orientation and navigation, exploring the potential influence of environmental (e.g., visual and auditory) inputs on vestibular inputs can shed new light on how self-motion cues are processed, taking into account the information coming from the outer world.

In the present study, we directly investigate the perceptual interaction between inertial cues and environmental landmarks, focusing on how external points of reference may modulate vestibular-related sensations. To achieve this goal, we asked participants to perform a detection task while stimulated with Galvanic Vestibular Stimulation (GVS). Critically, as mentioned earlier in this chapter, GVS activates the peripheral vestibular organs, i.e., the otoliths and semicircular canal afferents, eliciting a self-motion sensation (a roll tilt sensation). To test whether self-motion sensitivity could be aided by the environmental cues, participants performed the detection task with or without external visual or acoustic landmarks placed in front of them, in different blocks of trials. Participants' ability to detect virtual vestibular-induced self-motion sensation with and without a landmark was measured using a signal detection approach.

4.2.1   Participants

Twenty-six healthy participants were recruited from an online subject pool (mean age $\pm$ SD: 27.84 $\pm$ 6.98 years old). Twenty-two participants were right-handed, as assessed using the Edinburgh handedness inventory (Oldfield, 1971). None of the participants had a history of neurological or psychiatric disorders. Informed consent was obtained before participation in the experiment. All participants had normal or corrected-to-normal vision. The experimental

protocol was approved by the ethics committee of Birkbeck University of London (United Kingdom). The study was conducted following the ethical standards of the Declaration of Helsinki.

4.2.2 Apparatus and stimuli

Galvanic Vestibular Stimulation. Based on previous studies (Cabolis et al., 2018; Gallagher et al., 2020), bipolar GVS was applied to deliver a boxcar pulse of 0.7 mA with 250 ms duration using a commercial stimulator (Good Vibrations Engineering Ltd., Nobleton, Ontario, Canada). Considering that individual thresholds for GVS-induced roll-rotation sensations range from 0.4 to 1.5 mA (Oppenländer et al., 2015), we chose GVS parameters that could induce a relatively weak virtual sensation of roll rotation toward the anodal side but hardly noticeable (Day et al., 1997).

Electrodes (approximately 4 cm^2) were covered by NaCl gel and fixed to each of the mastoid processes. We administered left-anodal/right-cathodal stimulation (L-GVS). The opposite direction was not administered to avoid increasing stimulation noise. Moreover, previous research that used the same vestibular detection task (Gallagher et al., 2020) did not find any differences between gvs polarities. L-GVS is supposed to stimulate the right brain hemisphere, which is the dominant hemisphere for vestibular processing (Eguchi et al., 2019; Ferrè et al., 2021; Fink et al., 2003; Raiser et al., 2020). A sham stimulation was also used. In this case, two electrodes were placed on the shoulders at the base of the neck, approximately 5 cm below the upper electrodes, using left-anodal/right-cathodal configuration (L-SHAM) (see Figure 4.8). No sensations of self-motion were experienced during this type of stimulation. GVS and sham stimulation waveforms were conveyed to a commercial stimulator (Good Vibrations Engineering Ltd., Nobleton, ON, Canada) over a serial port connected to the experimental computer (Dell Latitude 5300 with Intel Core i5-8265U central processor at 1.60 GHz; 64-bit Windows 10 Enterprise Version 18362).

Landmarks. On different blocks of our experimental design, we presented two types of landmarks: one visual and one acoustic. The visual landmark consisted of a red LED light generated by a laser mounted on a chinrest and projected to a blackboard. The acoustic landmark consisted of a pink noise sound (65 dB) played by a loudspeaker. Both landmarks were placed in front of the participants' faces at a 30 cm distance (see Figure 4.9). When the sound played, the blackboard was removed to avoid obstructing the sound diffusion.
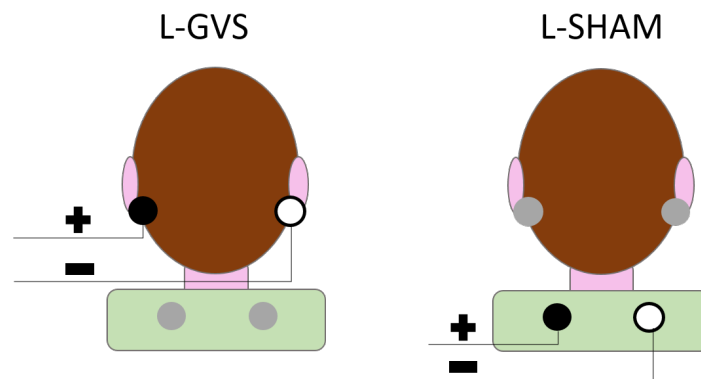
Figure 4.8 On the left, L-GVS configuration is shown, with the anode (+) on the left mastoid and the cathode (-) on the right mastoid. On the right, the L-SHAM configuration, with the two electrodes located at the basis of the neck.

4.2.3 Experimental design and procedure

Verbal instructions about the task were given to participants before the experimental session. The whole procedure was carried out in a dark room. To reduce the postural consequences of the GVS pulse, participants seated in a comfortable position rested their heads on a chinrest and placed their arms on their lap. Before the experiment, participants experienced a brief block of four L-SHAM and four L-GVS trials in a randomized order to familiarize themselves with the sensations generated by the stimulations (no feedback about the type of stimulation was provided). Our experimental design involved the presence of landmarks (visual or auditory) and vestibular signals. We asked participants to perform a Vestibular Detection Task (VDT), similar to the one used by Gallagher et al. (2020), and it was designed to follow a signal detection approach (Hautus et al., 2021; Macmillan and Creelman, 1991). We presented four blocks, two without landmarks, one with the visual landmark, and one with the acoustic landmark. In the visual landmark block, participants wore sunglasses to reduce the possibility of perceiving irrelevant visual cues available in the room; in the other conditions, participants had their eyes closed. We designed two blocks without landmarks to pair the two blocks with the landmarks, obtaining the same number of trials with and without the external stimuli. Each block contained 32 trials with GVS stimulation (vestibular stimuli present) and 32 trials with SHAM stimulation (vestibular stimuli absent), obtaining a total of 256 trials (32 trials x 2 types of stimulation x 4 blocks). GVS and SHAM stimulations were randomized within each block. The order of blocks was counterbalanced across participants so that the two blocks without landmarks were never consecutive (see an example in Figure 4.10). For landmarks blocks, the landmark was presented for 60 seconds prior to completing
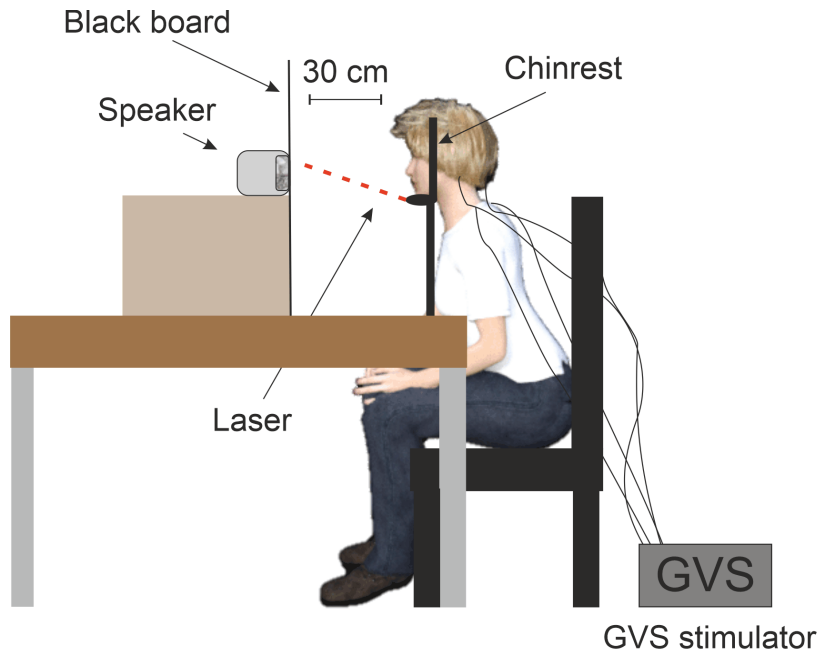
Figure 4.9 Experimental setup. The participant was comfortably seated on a chair, leaning on a chinrest with the arms on their lap. On the table, at 30 cm of distance, the experimenter placed the loudspeaker (that played the acoustic landmark) and a blackboard on which a red laser light was projected (i.e., the visual landmark). *The avatar representing the participants has been rendered in DAZ Studio 4.21 (DAZ productions, http://www.daz3d.com/).*

the detection task and continued throughout the entire block. Similarly, for blocks without landmarks, participants waited for 60 seconds before starting the experimental trials. On each trial, participants heard a beep to warn them to pay attention to any potential GVS-induced roll sensations but ignore any non-specific vestibular sensations, such as tingling under the electrodes' surface. A second beep 500 ms later indicated that participants should verbally respond 'yes' if they felt roll sensations or 'no' if they did not. GVS or SHAM stimulation was delivered between these two sounds. The experimenter used Matlab (R2019b, The MathWorks, United States) to trigger the stimuli and record participant responses.

4.2.4 Data analysis

A signal detection approach was used to analyze the responses in the VDT (Hautus et al., 2021; Macmillan and Creelman, 1991). We counted the number of hits (the number of trials in which L-GVS was present and the participant responded 'yes'), misses (the number of trials in which L-GVS was present and the participant responded 'no'), false alarms (the number of trials in which L-GVS stimulation was absent and the participant responded 'yes'),

Experimental Design

| 1 BLOCK | 1 BLOCK | 1 BLOCK | 1 BLOCK |
| Acoustic Landmark OFF | Acoustic Landmark ON | Visual Landmark OFF | Visual Landmark ON |
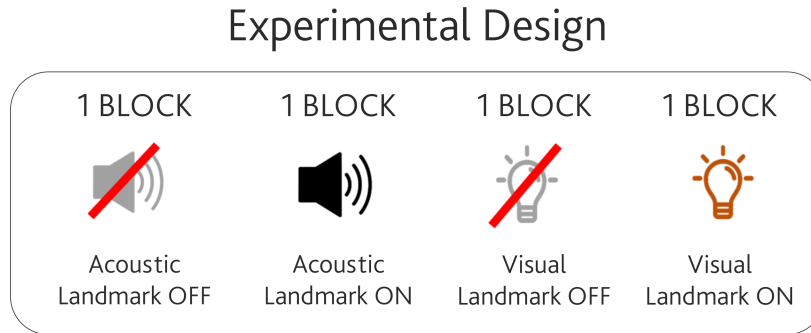
Figure 4.10 Experimental design. Here, it is depicted an example of the potential presentation of the blocks. As mentioned, we counterbalanced the order of ON vs. OFF blocks, and visual vs. acoustic blocks across participants, so that the two OFF block were never consecutive. In this example, we started with the acoustic blocks, specifically with the landamrk OFF condition.

and correct rejections (the number of trials in which L-GVS stimulation was absent and the participant responded 'no'). A correction was applied in case of zero correct answers or all correct answers for hits or false alarms (hits = 0 means never detecting the sensation induced by the GVS; false alarms = 0 means never mistaken a SHAM stimulation for GVS; all yes to GVS trials means always detect the GVS stimulation; all yes to SHAM trials means always mistaken the SHAM for GVS stimulation). Specifically, if we observed zero correct answers, we added 0.5 to 0; if we obtained all correct answers, we subtracted 0.5 from the number of given trials. Hit rates [P('yes'|GVS)] and false alarm rates [P('yes'|SHAM)] were used to calculate perceptual sensitivity (d prime; d'), namely the difference between z-transformed probabilities of hits and false alarms ($d' = z(Hit) - z(False\ Alarm)$). The response bias (C), i.e., the idiosyncratic criterion of participants to be more likely to answer yes or no regardless of the presented stimuli, was also calculated ($C = [z(Hit) + z(False\ Alarm)]/2$). Both d' and C were computed for each condition for each participant. Therefore, we divided our data according to the two sensory modalities of landmarks. Since we hypothesized an increased sensitivity to GVS elicited sensations in the presence of landmarks, we applied a one-tailed paired t-test on d' comparing blocks with and without landmarks (visual landmark ON vs OFF; acoustic landmark ON vs OFF). Conversely, we applied two-tailed paired t-tests to compare C, expecting no variation of individual response bias across conditions. In case of violation of the normality assumption (tested with Shapiro-Wilk's tests), we performed the corresponding permutation tests, using *perm.t.test* function from package *RVAideMemoire* in RStudio (version 3.6.2, 2019).

Table 4.1 Mean (SD) percentage hits, misses, false alarms, and correct rejections per each block.

Block/Responses	Hits (mean % ± SD)	Miss (mean % ± SD)	Correct Rejection (mean % ± SD)	False Alarms (mean % ± SD)
Acoustic landmark ON	60.2 ± 31.3	39.8 ± 31.3	91 ± 12.6	9.01 ± 12.6
Acoustic landmark OFF	65.7 ± 29.9	34.3 ± 29.9	87 ± 18.4	13 ± 18.4
Visual landmark ON	64.4 ± 33	35.6 ± 33	88.8 ± 16.5	11.2 ± 16.5
Visual landmark OFF	60.3 ± 26.4	39.7 ± 26.4	88.2 ± 16.1	11.8 ± 16.1

4.2.5   Results

The mean percentage and standard deviations (SD) of hits, misses, false alarms, and correct rejections by blocks are reported in Table 4.1. Regarding the comparison between visual landmark ON and OFF, the one-tailed paired t-test on d'revealed a significantly higher d'in the condition with the visual landmark than without it (t(25) = -1.85; p = .04; Cohen's d = -0.36, 95% CI [-0.8, 0.002]), suggesting greater sensitivity. The permutation paired t-test comparing the d'between acoustic landmarks ON and OFF revealed no such difference (t = 0.77; p = .78; Cohen's d = 0.15, [-0.25, 0.65]) (see Figure 4.11). Analyses on criterion C revealed neither differences for visual ON vs OFF conditions (paired t-test: t(25) = 0.6; p = .55; Cohen's d = 0.12, 95% CI [-0.26, 0.61]) nor for auditory ON vs OFF conditions (permutation paired t-test: t = 1.51; p = .12; Cohen's d = 0.3, [-0.63, 0.07]) (see Figure 4.12).

Overall, these results showed a slightly significant increased sensitivity to self-motion sensations elicited by GVS when in the presence of a visual landmark, while this condition did not have an impact on the individual response bias. Conversely, the presence of the acoustic landmark did not influence either participants' sensitivity or response bias.

4.2.6   Conclusion

In the present study, we investigated the interaction between external points of reference and vestibular inputs. We showed that an Earth-fixed landmark could increase the sensitivity to vestibular signals, but solely when the landmark consisted of a visually-encoded stimulus.

While the individual response criterion did not vary across blocks, the sensitivity to the self-motion sensation elicited by the GVS was enhanced when staring at a visual landmark. This finding suggests that vestibular signals are processed by taking into account the surrounding environment. A great amount of research already showed that, while we move through space, we rely on multiple sources of spatial cues. Specifically, vestibular and visual

information are shown to be optimally integrated to efficiently perceive our self-motion and navigate (Butler et al., 2010; Chen et al., 2017; Deangelis and Angelaki, 2012; Fetsch et al., 2009; Kaliuzhna et al., 2018; Nardini et al., 2008; Petrini et al., 2016; Zhao and Warren, 2015b). The present finding shows that visual signals coming from the environment provide
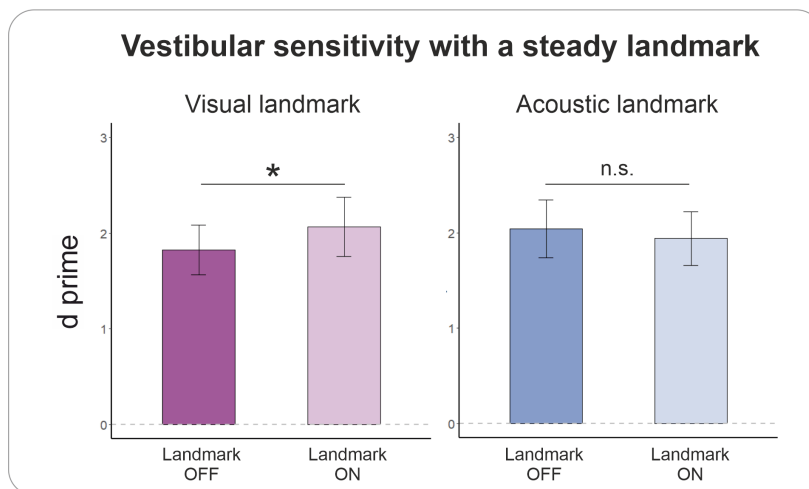
Figure 4.11 Results on d prime comparing landmark ON and OFF conditions for each landmark's modality. One-tail t-tests showed a higher d prime in landmark ON than landmark OFF condition in visual landmark modality. No difference was found between ON and OFF in acoustic landmark modality. Error bars are standard errors. * = p < .05; n.s. = not significant.
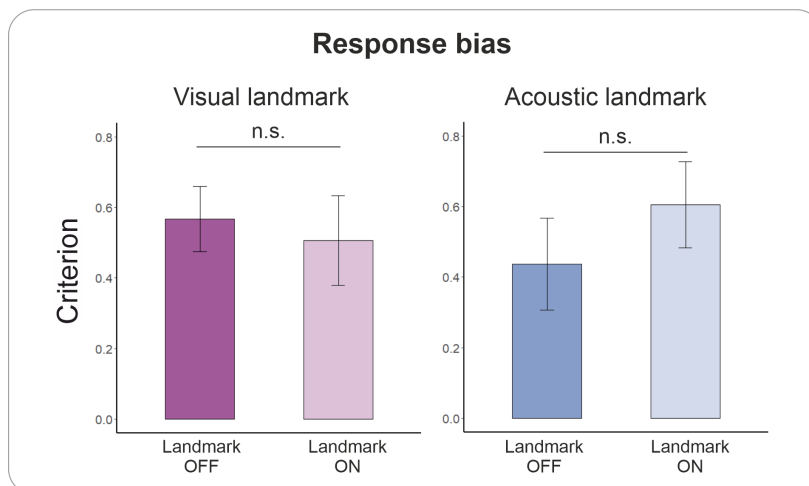
Figure 4.12 Results on Criterion comparing landmark ON and OFF conditions for each landmark's modality. Two-tail t-tests showed no difference between landmark ON and OFF conditions in both sensory modalities of landmarks. Error bars are standard errors. n.s. = not significant.

relevant information to enhance our ability to perceive inertial self-motion cues. This result suggests a further interaction between visual and vestibular systems in self-motion perception. Interestingly, we provide a single and small light as a landmark, placed in front of participants. It is plausible that varying the salience, the closeness, and the position of the landmark itself may lead to a different influence on vestibular processing: e.g., a more salient landmark may induce a greater increase of inertial self-motion cues sensitivity, providing an even more stable and trustworthy reference in the external world.

When in the presence of an acoustic landmark, we observed no changes in vestibular cues detection sensitivity. In the acoustic landmark block, we provide a steady pink noise sound located in front of the participant as an external point of reference. The sound was chosen to be as informative as possible: located in front of the participant to foster the detection of any sensation of deviation from the Earth-fixed vertical and easily localizable. Despite this consideration, the acoustic landmark has not been taken into account by the vestibular processing. This indicates once again the specific and robust functional connection between the visual and vestibular systems.

Chapter 5

General discussion and conclusion

The main aim of the present doctoral thesis is to investigate how multiple sensory cues, such as the visual, auditory, and vestibular, impact spatial processing and representation for effective spatial orientation and navigation. To this end, my Ph.D. project involved the study of various spatial tasks in multisensory environments. After a general introduction to the interaction between our sensory systems for spatial processing while exploring the external world (Chapter 1), I investigated audio-visual integration during spatial navigation (Chapter 2). Subsequently, I explored the interaction between the vestibular input and the available environmental cues by using different methodologies, such as passive motion stimuli (Chapter 3) and artificial stimulations (Chapter 4). The present last chapter will discuss and interpret the research work presented so far in this thesis, taking into consideration its theoretical and potential clinical implications.

5.1 Summary

Following the theoretical background displayed in Chapter 1, Chapter 2 focused on the ability to use Earth-fixed landmarks to efficiently navigate and on the interaction between hearing and vision. Specifically, it aimed at investigating the occurrence of multisensory integration between auditory and visual landmarks in an inferential spatial navigation task with healthy adults. The results specifically emphasized how, even in an apparently homogeneous group of participants, interindividual differences in perceptual reliance on external cues may lead to notably dissimilar spatial performances. The extensive analyses of the navigational abilities of each individual allowed us to discover novel aspects of multisensory integration, suggesting that idiosyncrasies should be taken into account in future investigations.

Chapter 3 presented two studies centered on audio-vestibular interaction in self-motion perception. In particular, the first study aimed at innovatively combining two newly developed simulation devices: the Rotational-Translational Chair to deliver passive motion stimuli and the 3D Tune-In Toolkit to build an acoustic virtual reality. Sighted and visually impaired participants were compared in a self-motion discrimination task that provided vestibular-only or audio-vestibular cues. Although the results are considered preliminary because of the small and variegate sample of visually impaired participants, this study allowed us to claim the valid combination of the two devices to be used for this clinical population as well. This finding provides the basis to develop multisensory and easily manageable methodologies to dig into spatial abilities in the presence of sensory disabilities. On the other hand, the second study highlighted how the vestibular inputs could help us understand the functional mechanisms underlying the processing of self-motion cues. In a self-motion discrimination task, we observed an inherent vestibular bias that led to systematically overestimating passive motion stimuli. This bias arguably represents how the brain attempts to build an inaccurate but optimal representation of the uncommon movements that deviate from the expected (i.e., most common) movements in daily life navigational activities.

Chapter 4 presented two studies focusing on exploring vestibular interaction with environmental spatial cues, using Galvanic Vestibular Stimulation as a simple but effective technique to elicit or interfere with vestibular signals during spatial perception tasks. In the first study, we used a stochastic waveform of stimulation that disrupted the vestibular inputs themselves. Temporarily impairing vestibular signals, we found that the vestibular system provides essential information about the distance of visually-encoded environmental stimuli. In the second study, we observed that the presence of an Earth-fixed visual landmark significantly improves the sensitivity to detecting artificially induced vestibular signals.

5.2   Theoretical and clinical implications

The research work done during my Ph.D. and displayed in this thesis provides new knowledge about how humans exploit multiple sensory cues for spatial processing, orientation, and navigation. To achieve this, different spatial tasks were employed: an inferential landmark-based navigational task (Chapter 2), a self-motion discrimination task (Chapter 3), a distance estimation task (Study 1, Section 4.1), and a self-motion detection task (Study 2, Section 4.2). Altogether, these methodologies allowed me to: i) investigate the multisensory integration mechanisms underlying spatial navigation; ii) establish how vestibular signals influence spatial encoding and spatial processing and how they interact with environmental landmarks for

orientation and navigation performances. Overall, the observed results stress the importance of considering all the spatially-relevant cues coming from every sensory source available.

5.2.1 Precision and accuracy in multisensory navigation

In the study described in Chapter 2, we showed that landmark-based spatial navigation is enhanced in precision by combining audio-visual cues, which are optimally integrated when they are equally reliable. Our data suggest that the integration of multisensory landmarks during navigation may depend on individual differences in cue weighting. Interestingly, our results are in line with Bentvelzen and colleagues' study (Bentvelzen et al., 2009), which found that only the group with comparable weights for auditory and visual information achieved the optimal audio-visual cue integration in a speed discrimination task. Our finding highlights how essential is to consider the natural variability across subjects, especially in terms of the heterogeneity of unimodal precision (Rohde et al., 2016). Age-related differences in multisensory perception have been previously found in the context of development (Cuturi and Gori, 2019; Gori et al., 2008; Nardini et al., 2008) and in the elderly population (Bates and Wolbers, 2014; de Dieuleveult et al., 2017), while idiosyncratic differences that occur regardless of age received little interest. One previous study that investigated the use of self-motion and visual information to navigate showed that the level of multisensory integration depended on the individual perceived discrepancy of the available sensory cues (the higher the perceived discrepancy, the lower the integration; Petrini et al. 2016). Another study investigating visuo-vestibular conflict observed individual differences in weighting the gravity cue, which was discarded when perceived as extremely discrepant relative to the visual one (Mcmanus and Harris, 2023). Finally, differences in sensory weights across individuals were previously found in the context of self-motion perception accomplished using visual and vestibular cues (Fetsch et al., 2009).

Which is the origin of these differences among individuals? Spatial navigation is a complex activity of human behavior. For this reason, it is puzzling to understand the underlying mechanisms, strategies, and preferences that lead to different navigational performances. People's navigation ability is influenced by their heterogeneous perceptual and cognitive processing, and how they explore the environment changes according to how they combine self-motion and external cues or use landmarks or other geometrical information (Wolbers and Hegarty, 2010). Many elements, such as distorted feedback and prior experiences, can affect spatial cues interaction (Zhao and Warren, 2018). For instance, previous research found that, when exposed to an unstable visual environment, people's reliance on visual cues may

be reduced (Chen et al., 2017; Zhao and Warren, 2015a), indicating sensory cue weighting flexibility. In the same way, different strategies or experiences might underlie the differences in participants' abilities to take advantage of the multisensory landmarks. Considering our experiments, the participants performed an inferential navigation task in which they were asked to use only landmarks to orient themselves (i.e., experiencing disorientation before each relocation response). As such, self-motion cues were never available to them except for the online execution of the response movement. One can speculate that participants that usually orient through space using visual landmarks could have been facilitated to be more precise using our visual cues compared to the others that mostly exploit self-motion or other information to navigate. Indeed, some navigators use orientation-based wayfinding strategies, and others use route-based strategies (Ishikawa, 2023): the first kind of navigator is focused on updating orientation according to external points of reference; the second one is focused on self-motion cues centered on the body. It is therefore plausible that participants who usually rely on route-based strategies may have been disadvantaged in performing the chosen landmark-based navigation task.

The presented findings revealed that the non-integrators group observed in our experiments ("dissimilar weight" group or DW group) shows numerically greater precision in exploiting visual cues and lower precision in exploiting auditory cues than integrators. Their performance can then be explained by the combination of these two aspects. It can be speculated that "favoring" the processing of one dominating sensory cue over another can originate from individual experiences. It is known that the auditory system performs better when processing temporal rather than spatial information (Burr et al., 2009; Gori et al., 2012). According to the cross-calibration theory (Burr and Gori, 2012; Gori, 2015), the more accurate sense within a specific task calibrates the less accurate sense and this process likely occurs during child development. It is then plausible the individuals in the non-integrators group experienced spatial navigation by relying mostly on visual information; this tendency could lead to greater expertise in this typology of environment exploration strategy, but, at the same time, it could make them deficient in perceiving cues from other sensory modalities. In this sense, it is plausible that unisensory-driven behavior could prevent the sharpening of the remaining senses.

It is plausible that our perceptual idiosyncrasies may arise from precocious experiences in childhood while interacting with others and the environment. It is indeed during childhood that children develop their ability to encode various sources of spatial information (inertial information, geometric cues, beacons, landmarks) and to combine cues (Newcombe, 2019). The progress of spatial abilities is tuned with motor development, interaction with the external

world, and feedback (Newcombe, 2019). In addition, literature concerning the absence of one sense during early life shows how this sensory disability substantially determines how the remaining senses process available cues. In particular, early visual disability influences the development of spatial representations based on the intact senses (Gori, 2015). Accordingly, severely visually impaired infants have been shown to rely more heavily on tactile cues than sighted infants (Gori et al., 2021). In addition, early blind individuals, differently from late blind and sighted, showed no biases in perceiving verticality judged haptically, demonstrating that visual sensory experience impacts the readout of vestibular information about the perception of gravity direction (Cuturi, 2022). In our study on self-motion discrimination ability in the visually impaired population (described in Section 3.4 of Chapter 3), we observed a slight improvement of spatial orientation precision in visually impaired participants when in the presence of auditory cues adding to vestibular ones, while, for the sighted population, the performance was similar in both conditions. This means that for sighted participants, vestibular cues are sufficient to reach their best performance, but for visually impaired people, the presence of external auditory landmarks shows to be useful to perform optimally, even though their processing of vestibular cues is not impaired per se (Moser et al., 2015). This suggests that the lack of one sense determines how the remaining senses process information and, consequently, it may lead to perceptual preferences for specific sensory cues. Besides, development and sensory disability are not the only backgrounds through which the flexibility of our perceptual systems can be observed. Extensive and repetitive experiences in adult life may lead healthy navigators to a substantial improvement in spatial abilities. Training can indeed enhance spatial working memory (Setti et al., 2018), and spatial orientation (Ishikawa and Zhou, 2020), likely involving changes in behavior and cortical structures (Maguire et al., 2003). Altogether, the present findings emphasize the importance of considering idiosyncratic perceptual characteristics and experiences when investigating multisensory integration.

If, on the one hand, we observed that precision had been modulated by individuals' perceptual preferences, on the other, the performance of accuracy has been firmly dominated by visual cues. Similarly, previous studies report that navigation accuracy can be determined by one cue at a time (in our case, the visual landmarks) (Zhao and Warren, 2015b). Relative to the visual-only condition, the presence of additional auditory cues did not either improve or worsen participants' relocation accuracy, suggesting that, in attempting to achieve high accuracy, all participants neglected the auditory information when both the visual and auditory landmarks were provided. Likewise, Zhao and Warren (2015b) found that, in a virtual homing task, visual cues dominated self-motion cues when determining the homing

direction, but at the same time, visual and self-motion cues were integrated to increase the precision. Thus, to explain their accuracy and precision results, the authors argued that cue competition and cue integration models coexist: this seems the case in our study as well. Crucially, our observed finding is consistent with the hypothesis that landmarks play a role in resetting the orientation of the path integration system (Etienne and Jeffery, 2004). The path integration system operating in the background (Cheng et al., 2007) could accumulate orientation errors during the relocation of the target object. Landmarks correct the orientation error that participants accumulate while basing on self-motion information to move. However, if the presence of landmarks ruled accuracy in a navigation task, one would expect comparable results for landmarks regardless of the encoding sensory modality, whether this is visual or auditory. Still, we observed a visual dominance in participants' spatial navigation accuracy. This finding then extends Zhao and Warren's (Zhao and Warren, 2015b) work by specifically suggesting that only *visual* landmarks might reset the orientation error of the path integration system, while auditory landmarks might not have the same role. It is plausible that highly reliable visual points of reference in the environment may correct the accumulating orientation error of the path integration system, while auditory cues may not be sufficiently reliable to play the same role. Along the same lines, in the study investigating self-motion perception ability in healthy participants (Section 3.5), we observed that accuracy in self-motion estimation was strongly dominated by vestibular cues, whilst spatial information conveyed by auditory stimuli was disregarded. Globally, these results suggest that in certain spatial contexts, according to the sensory cues available, accuracy may be determined by the processing of the most reliable cue for spatial orientation and navigation tasks.

The sound properties of auditory landmarks may play a role in determining whether they can facilitate orientation through space. In a real-world environment, auditory landmarks can be either continuous (e.g., a river flowing) or irregular (e.g., the playing bell of a church), likely conveying different levels of spatial information. In our experiments, we selected steady auditory landmarks, similarly to previous studies (Jetzschke et al., 2017; Nardi et al., 2020; Viaud-Delmon and Warusfel, 2014), because we meant to provide constantly active sound stimuli that may be more likely to be detected and used to spatially navigate. Considering the navigational performance obtained only with auditory landmarks (Chapter 2), our study extended previous findings, confirming that, even with less efficiency, it is possible to use auditory sources of information to orient themselves through space (Jetzschke et al., 2017; Nardi et al., 2020; Viaud-Delmon and Warusfel, 2014). This last piece of evidence is particularly relevant to studying navigation and orientation in more ecological

and real-life environments and exploring landmark-based navigation in the visually impaired population. Indeed, in the absence of vision, auditory landmarks may aid more complex spatial reasoning in blindness. With regard to this possibility, we actually found that visually impaired people may benefit from auditory cues when estimating self-motion sensations (see the study described in Section 3.4), even though the small size of our sample prevents us from robustly claiming the observed trend.

Overall, we can claim that audio-visual points of reference in the environment lead to optimal navigational precision for navigators who do not extremely favor vision over hearing. Nonetheless, accuracy is still dominated by vision. Despite the effort to make auditory cues as spatially informative as possible, auditory landmarks seem to be still not as reliable as visual ones. These findings must be taken into account when further exploring multisensory processing in spatial tasks with healthy and clinical populations, benefiting from the investigation of individual perceptual preferences to build tailored assessment and rehabilitation procedures.

5.2.2   Optimizing self-motion perception

The studies described in Chapter 3 presented the successful and novel combination of two simulation devices (the RT-Chair and the 3DTI Toolkit) to investigate audio-vestibular perception. In Study 1 (Section 3.4), we tested this tools' combination with sighted and visually impaired populations for the first time, performing a self-motion discrimination task. Results supported the possibility of taking advantage of the employment of simulation technologies to investigate spatial perception in healthy and clinical populations with the same methodology.

In addition, specifically for visually impaired participants, we also observed a feeble improvement in self-rotation estimations in the condition with audio-vestibular cues than in the only-vestibular condition, which brought their performance to the sighted participants' level. The potential conclusion to draw is twofold: on the one hand, even though prior research reported well-functioning vestibular perception in blind people (Moser et al., 2015), their use of vestibular inputs to estimate the traveled distance seems to be poorer without the aid of any other cues if compared with sighted; on the other hand, visually impaired people seem to process spatial auditory cues more profitably than sighted, benefiting from the additional spatial information that they convey. Crucially, navigators with visual disabilities rely on auditory signals on a daily basis. However, the small size and the heterogeneity of our visually impaired participants sample prevent us from generalizing the results we observed.

Future studies are needed to compare different groups of visually impaired people in the presented audio-vestibular interaction task.

Critically, factors such as duration of blindness and onset of the visual disability (congenital, acquired in the early or late stage of life) usually strongly influence performance in spatial tasks (Amadeo et al., 2019, 2020; Cappagli et al., 2017; Renier et al., 2014). For simpler tasks, such as auditory discrimination and localization (Battal et al., 2020; Doucet et al., 2005; Lessard et al., 1998), congenital or early blindness is associated with comparable or even superior performance to the sighted population. Along with behavioral abilities, early blind individuals show a functional reorganization of the occipital areas (originally designated to process visual cues) to compute the spatial components of auditory stimuli (Amadeo et al., 2019; Collignon et al., 2009). Nonetheless, according to cross-sensory calibration theory (Burr and Gori, 2012; Gori, 2015), vision is the sense most suitable to process spatial information, and it is a necessary requirement to calibrate the other senses for spatial processing in the first stages of life. Thus, congenital or early acquired blindness often negatively impacts complex auditory spatial skills. Conversely, late-acquired visual impairment allows for exploiting the visual calibration that occurred in childhood. Late blind subjects are indeed more similar to sighted individuals in locational judgments after a perspective change in small-scale space (Juurmaa and Lehtinen-Railo, 1994), spatial reasoning capabilities (Bertonati et al., 2020), and navigation tasks (Gori et al., 2017).

However, literature that concerns spatial abilities in congenital, early, and late blind people is greatly heterogeneous because of the extensive use of different experimental methodologies and individual differences (Thinus-blanc and Gaunet, 1997). For this reason, it is challenging to precisely predict the performances different groups of visually impaired individuals might have in the presented self-rotation estimation task. In detail, our task requires: i) efficiently localizing the two spatialized auditory landmarks relative to the body, one at a time, not necessarily identifying the metric relationship between the two (egocentric representation); ii) using the collected vestibular cues to estimate the traveled rotational distance; iii) combining the audio-vestibular information and generating a precise global estimation. Considering the possibility of processing one sound at a time relative to the body, I would expect a good ability to effectively localize the two auditory landmarks for all groups of blind individuals, regardless of the time of onset of the impairment or years of blindness. Regarding the ability to grasp spatial information from vestibular signals, it is known that, generally, the vestibular thresholds of visually impaired people in yaw rotations are comparable to the sighted ones (Moser et al., 2015); however, the ability to infer more complex spatial information from vestibular cues alone might be worse for congenitally blind

individuals than in sighted (Seemungal et al., 2007). In addition, previous studies showed that blind individuals who lacked visual inputs in the early stages of life (congenitally or early blind) show impaired performance in spatial tasks that require the integration of multiple and simultaneous sensory inputs if compared with both late blind and sighted individuals (see the review by Pasqualotto and Proulx 2012). Although we did not deliver simultaneous audio-vestibular cues, our stimuli required participants to extract spatial information from two different sensory modalities. Altogether, these considerations suggest the possibility of expecting improved performance in the audio-vestibular condition for late blind individuals but not for early or congenital blind ones. Furhter studies are needed to deeply explore these hypotheses.

In Study 2 of Chapter 3 (Section 3.5), we performed the same self-rotation perception task, focusing on estimation accuracy. The results revealed the presence of a robust overestimation bias in yaw rotations, showing resistance to the influence of the presented static auditory information, both automatically and actively explored.

Rotational vestibular information plays a fundamental role in complex navigation paths of daily activities (Cuturi et al., 2021; Glasauer et al., 2002), but our perception of rotational signals proves nonetheless to be flawed. Previous works investigating self-motion perception found inaccuracies leading to overestimation (Israël et al., 1995; Ivanenko et al., 1997; Mackrous and Simoneau, 2011; Marlinsky, 1999). As mentioned earlier in this thesis, even though perceptual biases initially seem misleading, systematic distortions in perception can be functionally useful in daily life and help understand how the brain processes sensory information. Concerning self-motion perceptual bias, past studies reported an overestimation of the perception of heading directions, relying on either vestibular or visual cues (Crane, 2012a; Cuturi and MacNeilage, 2013). In this context, overestimating a specific heading direction can be interpreted as the result of having improved sensitivity to changes relative to the direction most relevant in daily life, i.e., the straight-ahead direction (Cuturi and MacNeilage, 2013). Over-representing lateral directions would then lead to detecting the slightest deviation from it, which would be swiftly corrected. Therefore, inherent perceptual biases are interpreted as the consequence of a system that attempts to optimize behavioral performance by taking advantage of prior experiences and knowledge (Cuturi, 2022; Knill and Pouget, 2004; MacNeilage et al., 2007). Likewise, the overestimation of yaw rotations we observed in the presented experiments has likely occurred due to a higher sensitivity to changes from the most relevant starting position at azimuth 0°. Similar to the literature on self-motion perception (Crane, 2012a; Cuturi and MacNeilage, 2013), we consider the observed overestimation bias to be an efficient adaptation of the brain that leads to a more functional

representation of the world. In this sense, our perception makes us highly sensitive to directions that are perceived as uncommon (deviated from the straight-ahead) to optimize the maintenance of the most common and useful ones (the straight-ahead). Both previously found heading biases and the rotational one presented here may represent a perceptual expansion of the actual traveled angular space that facilitate the detection of any (potentially even dangerous) deviations from the straight-ahead. Therefore, the fine-tuning of the vestibular processing to movements away from a preferential direction arguably reflects the tentative to achieve efficient spatial navigation and locomotion.

Interestingly, the neural network that encodes self-displacement information and the contribution of vestibular signals to traveled distance perception is not fully understood (Seemungal, 2015). As mentioned in the Introduction of this thesis (Section 1.1.2), vestibular inputs travel via vestibular thalamic pathways and then project to a highly distributed cortical network. Concerning high-level vestibular processing, previous research suggested that complex spatial processing based on vestibular inputs may be specifically supported by a temporo-parietal cortical network. For example, it has been shown that repetitive transcranial magnetic stimulation on the posterior parietal cortex (PPC) specifically disrupted healthy participants' ability to reproduce previously experienced angular displacements in a path integration task, but it did not affect their ability to perceive self-motion (Seemungal et al., 2008). This finding suggests that the neural substrates of spatial and motion perception based on vestibular cues do not overlap, as hypothesized by Seemungal (see Figure 5.1, Seemungal 2015). Another study with patients with acute hemisphere lesions (Kaski et al., 2016) has shown that only those who reported damage to the temporo-parietal junction (TPJ) had impaired performance in estimating traveled angular distance and motion duration, but presented a healthy-like vestibular motion perception. The self-rotation estimation task we used required participants not only to perceive vestibular cues but to process and interpret them in order to generate an accurate estimation of the traveled rotational displacement. Therefore, even if our study did not directly investigate the cortical substrate of the bias, the previous evidence described above leads us to speculate that the cortical temporo-parietal network may underlie the self-motion estimation mechanism that leads to the overestimation biases that we have observed.

Critically, in this study, we did not find any modulation of the overestimation bias by the present auditory cues. Previous literature has shown that when auditory information is provided at the same time as vestibular cues, they are optimally integrated to increase the precision in self-motion perception for low-frequency movements (i.e., below 0.5 Hz) (Shayman et al., 2020). According to the Maximum Likelihood Estimation (MLE) model, the
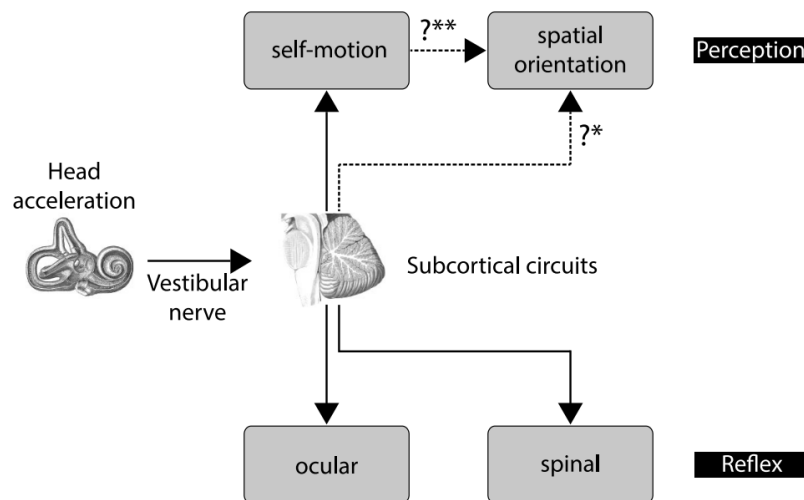
Figure 5.1 Hypothetical schema of vestibular signal flow. The figure depicts the pathway of vestibular signals into the brain and the branching off into reflex (lower level) and perceptual outputs (higher level). Vestibular motion perception is considered the primary vestibular perception. Vestibular spatial perception is derived from vestibular motion signals. The schema shows two alternatives for the derivation of vestibular spatial perception — a subcortical vs. a cortical derivation. *Reproduced with permissions from Seemungal, B. M. (2015). The components of vestibular cognition—motion versus spatial perception. Multisensory research, 28(5-6), 507-524.*

optimal estimation of combined multisensory cues corresponds to the weighted average of the two cues when they have equal reliability (Ernst and Banks, 2002). In contrast, if one cue is more reliable, the estimation shifts towards the former, resulting in its sensory dominance (Ernst and Banks, 2002). In our study, we observed no difference between vestibular-only and audiovestibular variability, i.e., same reliability. Besides, the absence of overlapping auditory and vestibular cues might be a limitation of our experimental design. The presence of acoustic information during motion may indeed increase the interaction between the external auditory cues and the vestibular ones. Thus, future studies may be addressed to provide auditory cues and vestibular cues simultaneously. Interestingly, previously in this thesis, auditory cues are already shown to be a weak source of spatial information than visual or vestibular information. In the previously mentioned studies about audio-visual landmarks integration and audio-vestibular interactions described in Chapters 2 and 4, auditory cues were not weighted as much as visual cues and did not properly interact with the other cues available for spatial processing. The only study in which it seems that the spatial components of auditory cues were taken into account was the study concerning the investigation of self-motion perception precision (Section 3.4), where visually impaired people were involved.

Thus, despite the consideration of making auditory signals as informative as possible, our results suggest that vision and inertial cues represent the predominant information for spatial processing in the sighted population.

Globally, the two studies presented in Chapter 3 highlight how humans process the sensory cues available to optimize their self-motion and navigational behaviors. The combination of two flexible existing simulation tools (the RT-Chair, Cuturi et al. 2020, and the 3DTI Toolkit, Cuevas-Rodríguez et al. 2019) allowed us to create a simple but effective self-orientation task to unveil the spatial perception mechanisms underlying multisensory spatial navigation abilities. Investigating how vestibular cues are processed and interpreted for self-motion estimation aid in uncovering the sturdy impact that the vestibular system has on neurocognitive functions in healthy and sensory-impaired populations (Ferrè and Haggard, 2020). Even though we have involved a small sample of participants with visual impairments in our study, we took advantage of two user-friendly technologies to systematically study self-rotation estimation performance based on audio-vestibular inputs. This technique reveals to be a promising methodology to help understand multisensory perception and implement easily new rehabilitation procedures in which the auditory-vestibular interaction may improve spatial representation abilities in visually impaired people. For instance, in the presence of sub-optimal self-motion estimation based on vestibular-only cues, the precise perception of spatialized auditory cues may serve as potential feedback to build a spatial link between traveled distance and the external auditory landmarks available, thus training vestibular processing for daily activities.

Therefore, these studies allowed us to shed new light on multisensory self-motion processing and provide neuroscience-based insights to build tailored assessment and rehabilitation procedures, which may enhance global spatial abilities for visually impaired people.

5.2.3   The vestibular-environment interaction

As mentioned earlier, we are extensively immersed in a multisensory environment. While navigating, we spatially encode both inertial and external cues. In Chapter 4, we emphasized the interaction between the vestibular inputs and environmental points of reference in a distance estimation task and a signal detection task. Indeed, I presented two studies employing Galvanic Vestibular Stimulation (GVS) as a technique to artificially stimulate the vestibular system during spatial tasks.

The first study (Section 4.1) explored the role played by vestibular signals in the spatial encoding of environmental targets. Vestibular cues strongly contribute to complex spatial

processing functions and heavily interact with the other sensory systems. Previous literature demonstrated that the vestibular system not only provides pivotal self-motion cues but also converts motion information to distance information to maintain spatial orientation (Kaski et al., 2016; Seemungal, 2015). Here we focused on the interaction between vestibular and external inputs for spatial encoding. In particular, we observed how, even in the absence of any self-motion cues coming from vestibular organs, artificial vestibular signals significantly contribute to the spatial estimation of visual targets located in depth. We indeed observed that interfering with vestibular signals while encoding visual-spatial cues affected the accuracy and precision in walking to a previously learned visual target location. These results demonstrate that the vestibular system participates in the encoding of the visual environment to interact with it efficiently, extending previous research showing the essential contribution of vestibular signaling in higher cognitive spatial functions.

Our finding revealed to be particularly essential to understand the multisensory interaction of spatial cues stemming from different sources, even in the preparation phase of efficient spatial navigation. To support our interpretation, the vestibular system notably never stops collecting information about the position of the head relative to the gravitational acceleration, transforming cues from a head-centered framework into cues in world-centered coordinates and allowing for efficient spatial orientation (Angelaki and Cullen, 2008). Gravity, like every linear acceleration, is perceived by the otoliths present in the inner ears. GVS is known to directly stimulate the entire vestibular nerve, eliciting responses from semicircular canals and otoliths (Stephan et al., 2005). Taking gravity as a constant point of reference, every head movement implies a new relationship between the organism and the surrounding space. Therefore, it is plausible that in our study, stochastic GVS might have interfered with graviception, thus disrupting a fundamental reference for the brain to process spatial information and thus modulating the perception of the relationship between the observer and environmental visually-encoded object. Along these lines, it is known that visual cues can significantly influence the perceptual read-out of head orientation relative to gravity (Dichgans et al., 1972; Gillingham and Previc, 1993; Howard and Hu, 2001).

Vestibular signaling interference seems to selectively influence the visual rather than auditory readout of spatial information. As already extensively mentioned, vision is the aptest sense for processing spatial information. Nevertheless, it is worth noticing that the human ability to determine the angular direction of a sound source in the horizontal plane is more accurate than distance estimation (Zahorik et al., 2005), which was the participants' task in the present study. Interestingly, previous research found a vestibular influence on auditory localization in the azimuthal plane (Lewald and Karnath, 2000). For this reason, we cannot

exclude that vestibular-auditory interaction may occur when acoustic targets are placed along the azimuth, becoming a greater informative spatial source. Future studies may be addressed to explore this interaction in the horizontal space. However, we previously observed in Section 3.5 that vestibular cues and horizontally spatialized auditory landmarks did not interact. Prior research that highlighted sensory integration between these two modalities (Kapralos et al., 2004; Shayman et al., 2020) involved simultaneous presentation and longer duration of the two stimuli if compared with our study with sGVS. It is plausible that, in order to take into account auditory information, there is a need to stress spatial components of auditory cues by making them more relevant. For example, it is already known that active self-motion can enhance the ability to discriminate sounds dislocated in depth (Genzel et al., 2018), demonstrating that the proprioceptive cues of body movements significantly contribute to vestibular-audio interactions. Moreover, spatial components conveyed by auditory cues can be made more informative by combining them with other congruent spatial cues, thoroughly considering multiple cues synchronization. Finally, previous findings showed that only when prompted to rely on a strategy based on auditory cues participants were able to use them to effectively reorient, whilst failing when required to use them spontaneously (Nardi et al., 2019). In our studies on self-motion discrimination (described in Chapter 3), participants were directly prompted also to use auditory cues. However, in the same trials where the auditory cue was present, more reliable sources of spatial information were also delivered, likely leading participants to rely on them and discard the available auditory signals. Thus, another possibility may focus on directly prompting navigators to use auditory-cue-based strategies, providing additional only-auditory cues trials. Critically, future studies may need to consider these aspects when focusing on multisensory integration in spatial navigation and probing the comparison between sighted and visually impaired populations.

To summarize, the results of our study on vestibular encoding systematically demonstrate that, even in the absence of inertial information originating from self-motion, the vestibular system influences how we perceive the external world and, in particular, the spatial encoding of visual cues. This study provides critical insights not only into the understanding of healthy vestibular processing but also into the well-known difficulties in spatial processing observed in individuals with vestibular loss.

The second study of Chapter 4 (Section 4.2) explored the potential influence of Earth-fixed landmarks on detecting artificial vestibular self-motion sensations. In detail, we investigated how earth-fixed points of reference may modulate the sensitivity to detect self-motion sensations elicited by artificial vestibular stimulation. We observed that a simple and stable (thus, reliable) visual landmark placed in front of participants was sufficient to

enhance their sensitivity, providing useful information about the features of the (steady) environment and the stimulation. A great amount of previous research showed that vestibular and visual information is optimally integrated to efficiently perceive our self-motion and navigate (Butler et al., 2010; Chen et al., 2017; Deangelis and Angelaki, 2012; Fetsch et al., 2009; Kaliuzhna et al., 2018; Nardini et al., 2008; Petrini et al., 2016; Zhao and Warren, 2015b). The present finding extends prior research demonstrating that visual signals can provide relevant information to enhance our ability to perceive inertial self-motion cues. This means that, when processing spatial information during navigation and orientation, the presence of visual reference in space helps disentangle between self- vs object- motion, providing a reliable anchor to the environment. Prior research found evidence that additional vestibular inputs facilitate the discrimination of visual objects' directions of movement, particularly for the most eccentric directions relative to the straight-ahead (MacNeilage et al., 2012). Our study extended previous literature, demonstrating that also environmental objects can facilitate vestibular detection.

With an additional auditory landmark, we observed no improvement in vestibular cues detection sensitivity. This result is in line with findings observed in Section 3.5, where we observed the absence of audio-vestibular interaction. Previous studies showed an audio-vestibular interaction in balance tasks, generally reporting an advantage relative to a silent environment (Easton et al., 1998; Karim et al., 2018; Stevens et al., 2017; Zhong and Yost, 2013). Still, there is no agreement on the optimal features of the sounds that may influence the processing of the inertial cues. In the present study, we chose the sound position and typology to be as spatially informative as possible. Despite our considered choice, the acoustic landmark has not been taken into account by the vestibular processing, reinforcing the idea of how auditory cues are less informative than visual ones to provide accurate spatial information (see also the visual dominance on accuracy over hearing described in Chapter 2) and consequently may also have poor reliability if compared with the more relevant vestibular information. Possibly, future studies may focus on investigating more deeply the relationship between the complexity of the soundscape available and its interaction with self-motion cues. Indeed, richer environmental auditory information, i.e., multiple three-dimensional auditory sources, has been shown to improve self-motion perception, allowing for building an auditory spatial map between the sounds, used as landmarks, and the body (Gandemer et al., 2017).

Our study on vestibular-landmark interaction globally suggests that not only do multiple cues work together to be optimally integrated when available simultaneously but also that one cue can change the perceptual read-out of another cue coming from a different sensory source,

showing the robust and dynamic link between sensory systems conveying the information that helps build a coherent representation of space.

Taken together, results from these studies extended previous literature about vestibular cognition, demonstrating a fundamental role of vestibular inputs in the interactions with visual environmental cues. Our studies are in line with previous literature that shows how visual and vestibular cues are strongly connected from low-level (i.e., reflexes) to high-level cognitive functions. The presented findings shed new light on the pivotal role of the vestibular system in spatial processing and interaction with visual environmental signals. The two studies in Chapter 4 emphasized the importance of taking into account vestibular perception in high-level spatial contexts such as spatial orientation and navigation. This sensory system is the only one that constantly provides information about the relationship between the body and the environment (e.g., relative to gravity) and reaches a widespread network of cortical areas, thus precociously communicating with the other sensory systems. Crucially, the strict binding with the visual system in a healthy population makes vestibular functioning a rich resource for studying self-motion and navigational processes in blindness. Moreover, here we provided essential insights not only into the understanding of healthy vestibular processing but also into the difficulties in spatial processing observed in individuals with vestibular loss and into situations in which the vestibular system is altered, such as in spaceflight missions.

5.3   Concluding remarks

During my Ph.D., I worked on achieving a deeper understanding of how the brain processes mobility and locomotion skills. I designed different experimental procedures to investigate the brain mechanisms associated with mobility, orientation, and locomotion abilities in adults, studying how people exploit the multisensory sources of information available, including acoustic, visual, and vestibular cues. Specifically, I used methodologies that required participants, on the one hand, to actively explore the environment and, on the other, to passively perceive self-motion information. Regarding the former, I employed a landmark-based navigational task (Chapter 2) and reaching target while walking task (Section 4.1), both in real-world environments; regarding the latter, I employed a self-rotation perception task using a motion platform (Chapter 3) and a vestibular detection task using galvanic vestibular stimulation (Section 4.2). The heterogeneity of these methodologies and techniques allowed me to investigate self-motion perception, navigation, and orientation abilities in multisensory environments according to multiple perspectives, thus enabling me to generalize results in the broadest possible way.

The findings illustrated in this thesis allow for the following conclusions:

1. In spatial contexts, audio-visual integration depends on idiosyncratic perceptual traits. Navigators who rely especially on visual cues do not achieve optimal navigation precision, while navigators who equally base their performance on audio-visual cues are able to achieve optimal precision. This evidence highlights the importance of taking into account idiosyncratic perceptual differences, in order to extract their perceptual profile for any tailored assessment or rehabilitation procedures.

2. The vestibular system is a resourceful sensory basis to better understand the interaction between the body and the external environment, both with and without inertial cues available. The observed intrinsic overestimation bias in self-rotation perception allows for comprehending how a single sensory source works to optimize the effective achievement of spatial tasks (e.g., self-motion estimation). From our results, it was also firmly clear that spatialized auditory cues, even when providing important spatial information, do not continuously interact with vestibular inputs at high-level cognitive spatial processes.

3. Visual loss seems to favor the spatial use of auditory cues available in the environment over exploiting vestibular-only ones. The absence of visual-vestibular experience in blindness may indeed lead to weight spatial information originating from inertial cues to a lesser extent. For this reason, the brain arguably attempts to access different cues (e.g., auditory cues) to overcome the resulting impaired spatial processing.

Beyond theoretical relevance, the results of the current thesis may have repercussions for assessment and rehabilitation procedures following a sensory loss. Future research may dig deeper into the observed perceptual preferences, investigating whether it is the case even for interaction between other sensory systems, e.g., auditory and haptic, auditory and vestibular cues. From a technological perspective, the neuroscientific results obtained by this thesis provide the basis to develop novel science-driven rehabilitation devices for improving the impaired spatial navigation and orientation abilities of blind individuals. Moreover, the combination of simulation tools, explored specifically in Section 3.4 of Chapter 3, likely opens the door to easily implement technologies that allow for exploring preserved and impaired spatial abilities in the visually impaired population. As for rehabilitation, in case visually impaired individuals benefit from the presence of spatialized auditory cues for self-motion perception, there is potential to develop new technological devices by training their ability to associate the processing of external points of reference with the processing of

inertial cues originating from their bodies, enhancing their self-motion perception in daily life.

Scientific production

Papers published:

- **Zanchi, S.**, Cuturi, L.F., Sandini G., Gori M. (2022a), "Interindividual differences influence multisensory processing during spatial navigation", *Journal of Experimental Psychology: Human Perception and Performance, 48(2)*, 174.

- **Zanchi, S.**, Cuturi, L.F., Sandini G., Gori M. (2022b), "How much I moved: Robust biases in the self-rotation perception", *Attention, Perception and Psychophysics*, 1-14.

- **Zanchi, S.**, Cuturi, L.F., Sandini G., Gori M., "Evaluation of a motion platform combined with an acoustic virtual reality tool: a spatial orientation test in sighted and visually impaired people", *43rd Annual Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2021.

- Miehlbradt, J., Cuturi, L. F., **Zanchi, S.**, Gori, M., Micera, S. (2021). Immersive virtual reality interferes with default head–trunk coordination strategies in young children. *Scientific Reports, 11(1)*, 17959.

Papers under review and in preparation:

- **Zanchi, S.**, Cuturi, L.F., Sandini G., Gori M., Ferré E.R., "Vestibular contribution to visual target localisation"

- **Zanchi, S.**, Cuturi, L.F., Sandini G., Gori M., Ferré E.R., "Sensory landmarks for vestibular self-motion perception"

- **Zanchi, S.***, Cuturi, L.F.*, Cappagli, G., Sandini G., Gori M., "Development of allocentric representations using self-motion information"

*Zanchi S. and Cuturi L.F. are equally contributing to this paper.

Conferences

Talks

- XXVI Congresso dell'Associazione Italiana di Psicologia – Sezione sperimentale (Online, 2-4/09/2020). "L'influenza dei landmark sulle abilità di orientamento e navigazione spaziale". Invited speaker for a symposium.

- XXVII Congresso dell'Associazione Italiana di Psicologia – Sezione Sperimentale (Lecce (IT), 8- 10/09/2021). "Lo sviluppo della rappresentazione spaziale allocentrica".

- 8th International Conference on Spatial Cognition (Online, 13-17/09/2021). "Idiosyncratic benefits in multisensory spatial navigation".

- 43rd Annual Conference of the IEEE Engineering in Medicine and Biology Society (Online, 1-5/11/2021). "Evaluation of a motion platform combined with an acoustic virtual reality tool: a spatial orientation test in sighted and visually impaired people".

- 44th European Conference on Visual Perception (Nijmegen (NL), 28/08-1/09/2022). "Vestibular contribution to visual target localisation".

- XXX Congresso dell'Associazione Italiana di Psicologia (Padova (IT), 27-30/09/2022). "The role of vestibular system in spatial perception".

- Annual Experimental Psychology Society Meeting (London (UK), 4-6/01/2023). "Vestibular contribution to the encoding of target localization".

- Vision Sciences Society Meeting (St. Pete Beach (FL,USA), 19-24/05/2023). "Visual vs. Auditory Landmark for Vestibular Self-motion Perception".

- XXIX Congresso dell'Associazione Italiana di Psicologia – Sezione Sperimentale (Lucca (IT), 18-20/09/2023). "Environmental Landmarks for Vestibular Self-motion Perception". Planned.

Posters

- Vestibular-Oriented Research Meeting (Online, 16-17/02/2021). "Robust biases in the estimation of yaw rotation and the potential influence of auditory landmarks".

- 43rd European Conference on Visual Perception (Online, 23-27/08/2021). "Development of allocentric representations using self-motion information".

- International Multisensory Research Forum (Ulm (DE), 4-7/07/2022). "The relationship between vestibular thresholds and balance tasks in a multisensory environment".

- 27th European Low Gravity Research Association Biennial Symposium and General Assembly (Lisbon (PT), 6-9/09/2022). "Altered gravity disrupts spatial navigation".

References

Alais, D. and Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3):257–262.

Amadeo, M. B., Campus, C., and Gori, M. (2019). Impact of years of blindness on neural circuits underlying auditory spatial representation. *NeuroImage*, 191(July 2018):140–149.

Amadeo, M. B., Campus, C., and Gori, M. (2020). Years of Blindness Lead to "Visualize" Space Through Time. *Frontiers in Neuroscience*, 14(August):1–14.

Angelaki, D. E. and Cullen, K. E. (2008). Vestibular System: The Many Facets of a Multimodal Sense. *Annual Review of Neuroscience*, 31(1):125–150.

Aubert, H. (1861). Eine scheinbare bedeutende Drehung von Objecten bei Neigung des Kopfes nach rechts oder links. *Archiv für Pathologische Anatomie und Physiologie und für Klinische Medicin*, 20(3-4):381–393.

Aytekin, M., Moss, C. F., and Simon, J. Z. (2008). A sensorimotor approach to sound localization. *Neural Computation*, 20(3):603–635.

Bates, S. L. and Wolbers, T. (2014). How cognitive aging affects multisensory integration of navigational cues. *Neurobiology of Aging*, 35(12):2761–2769.

Battal, C., Occelli, V., Bertonati, G., Falagiarda, F., and Collignon, O. (2020). General Enhancement of Spatial Hearing in Congenitally Blind People. *Psychological Science*, pages 1–11.

Bense, S., Stephan, T., Yousry, T. A., Brandt, T., Dieterich, M., Stephan, T., and Yousry, T. A. (2000). Multisensory Cortical Signal Increases and Decreases During Vestibular Galvanic Stimulation (fMRI). *Journal of neurophysiology*, 85(2):886–899.

Bentvelzen, A., Leung, J., and Alais, D. (2009). Discriminating audiovisual speed: Optimal integration of speed defaults to probability summation when component reliabilities diverge. *Perception*, 38(7):966–987.

Bertonati, G., Tonelli, A., Cuturi, L. F., Setti, W., and Gori, M. (2020). Assessment of spatial reasoning in blind individuals using a haptic version of the Kohs Block Design Test. *Current Research in Behavioral Sciences*, 1(November):100004.

Bigelow, R. T. and Agrawal, Y. (2015). Vestibular involvement in cognition: Visuospatial ability, attention, executive function, and memory. *Journal of Vestibular Research: Equilibrium and Orientation*, 25(2):73–89.

Blakemore, S. J., Wolpert, D., and Frith, C. (2000). Why can't you tickle yourself? *NeuroReport*, 11(11):R11–R16.

Blouin, J., Gauthier, G. M., van Donkelaar, P., and Vercher, J.-L. (1995). Encoding the position of a flashed visual target after passive body rotations. *Neuroreport*, 6(8):1165–1168.

Brandt, T. and Danek, A. (1994). Vestibular Cortex Lesions Affect the Perception of Verticality. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 35(4):403–412.

Brandt, T., Glasauer, S., Stephan, T., Bense, S., Yousry, T. A., Deutschlander, A., and Dieterich, M. (2002). Visual-vestibular and visuovisual cortical interaction: new insights from fMRI and PET. *Annals of the New York Academy of Sciences*, 956(1):230–241.

Bruggeman, H., Piuneu, V. S., Rieser, J. J., and Pick, H. L. (2009). Biomechanical Versus Inertial Information: Stable Individual Differences in Perception of Self-Rotation. *Journal of Experimental Psychology: Human Perception and Performance*, 35(5):1472–1480.

Burr, D., Banks, M. S., and Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, 198(1):49–57.

Burr, D. and Gori, M. (2012). Multisensory Integration Develops Late in Humans. In *The Neural Bases of Multisensory Processes*.

Butler, J. S., Campos, J. L., and Bülthoff, H. H. (2014). Optimal visual–vestibular integration under conditions of conflicting intersensory motion profiles. *Experimental Brain Research*, 233(2):587–597.

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

Cabolis, K., Steinberg, A., and Ferrè, E. R. (2018). Somatosensory modulation of perceptual vestibular detection. *Experimental Brain Research*, 236(3):859–865.

Cappagli, G., Finocchietti, S., Cocchi, E., and Gori, M. (2017). The impact of early visual deprivation on spatial hearing: A comparison between totally and partially visually deprived children. *Frontiers in Psychology*, 8(MAR):1–8.

Chan, E., Baumann, O., Bellgrove, M. A., and Mattingley, J. B. (2012). From objects to landmarks: The function of visual location information in spatial navigation. *Frontiers in Psychology*, 3(304):1–11.

Chen, X., McNamara, T. P., Kelly, J. W., and Wolbers, T. (2017). Cue combination in human spatial navigation. *Cognitive Psychology*, 95:105–144.

Cheng, K., Shettleworth, S. J., Huttenlocher, J., and Rieser, J. J. (2007). Bayesian Integration of Spatial Information. *Psychological Bulletin*, 133(4):625–637.

Cheng, Z. and Gu, Y. (2018). Vestibular system and self-motion. *Frontiers in Cellular Neuroscience*, 12(November):1–9.

Collignon, O., Voss, P., Lassonde, M., and Lepore, F. (2009). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Experimental Brain Research*, 192(3):343–358.

Crane, B. T. (2012a). Direction Specific Biases in Human Visual and Vestibular Heading Perception. *PLoS ONE*, 7(12):1–15.

Crane, B. T. (2012b). Roll aftereffects: Influence of tilt and inter-stimulus interval. *Experimental Brain Research*, 223(1):89–98.

Cuevas-Rodríguez, M., Picinali, L., González-Toledo, D., Garre, C., de la Rubia-Cuestas, E., Molina-Tanco, L., and Reyes-Lecuona, A. (2019). *3D tune-in toolkit: An open-source library for real-time binaural spatialisation*, volume 14.

Cuturi, L. F. (2022). Perceptual Biases as the Side Effect of a Multisensory Adaptive System: Insights from Verticality and Self-Motion Perception. *Vision*, 6(3):53.

Cuturi, L. F., Aggius-Vella, E., Campus, C., Parmiggiani, A., and Gori, M. (2016). From science to technology: Orientation and mobility in blind children and adults. *Neuroscience and Biobehavioral Reviews*, 71:240–251.

Cuturi, L. F., Alborno, P., Cappagli, G., Volta, E., Volpe, G., and Gori, M. (2021). The influence of yaw rotation on spatial navigation during development. *Neuropsychologia*, 154:107774.

Cuturi, L. F. and Gori, M. (2019). Biases in the visual and haptic subjective vertical reveal the role of proprioceptive/vestibular priors in child development. *Frontiers in Neurology*, 10(JAN):1–10.

Cuturi, L. F. and MacNeilage, P. R. (2013). Systematic Biases in Human Heading Estimation. *PLoS ONE*, 8(2).

Cuturi, L. F. and MacNeilage, P. R. (2014). Optic flow induces nonvisual self-motion aftereffects. *Current Biology*, 24(23):2817–2821.

Cuturi, L. F., Torazza, D., Campus, C., Merello, A., Lorini, C., Crepaldi, M., Sandini, G., and Gori, M. (2020). The RT-Chair : a Novel Motion Simulator to Measure Vestibular Perception. *In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE.*, pages 3318–3322.

Dallal, N. L., Yin, B., Nekovářová, T., Stuchlík, A., and Meck, W. H. (2015). Impact of vestibular lesions on allocentric navigation and interval timing: The role of self-initiated motion in spatial-temporal integration. *Timing and Time Perception*, 3(3-4):269–305.

Day, B. L., Séverac Cauquil, A., Bartolomei, L., Pastor, M. A., and Lyon, I. N. (1997). Human body-segment tilts induced by galvanic stimulation: A vestibularly driven balance protection mechanism. *Journal of Physiology*, 500(3):661–672.

de Dieuleveult, A. L., Siemonsma, P. C., van Erp, J. B., and Brouwer, A. M. (2017). Effects of aging in multisensory integration: A systematic review. *Frontiers in Aging Neuroscience*, 9(MAR):1–14.

De Vrijer, M., Medendorp, W. P., and Van Gisbergen, J. A. (2009). Accuracy-precision trade-off in visual orientation constancy. *Journal of Vision*, 9(2):1–15.

Deangelis, G. C. and Angelaki, D. E. (2012). Chapter 31Visual – Vestibular Integration for Self-Motion Perception Authors AND THE UTILITY OF VISUAL – VESTIBULAR. In *The Neural Bases of Multisensory Processes*.

Desantis, A., Roussel, C., and Waszak, F. (2014). The temporal dynamics of the perceptual consequences of action-effect prediction. *Cognition*, 132(3):243–250.

Deshpande, N. and Patla, A. E. (2005). Dynamic visual-vestibular integration during goal directed human locomotion. *Experimental Brain Research*, 166(2):237–247.

Dichgans, J., Held, R., Young, L. R., and Brandt, T. (1972). Moving Visual Scenes Influence the Apparent Direction of Gravity Published by : American Association for the Advancement of Science Stable URL : https://www.jstor.org/stable/1735275 REFERENCES Linked references are available on JSTOR for this article : reference # references _ tab _ contents You may need to log in to JSTOR to access the linked references . 178(4066):1217–1219.

Doucet, M. E., Guillemot, J. P., Lassonde, M., Gagné, J. P., Leclerc, C., and Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Experimental Brain Research*, 160(2):194–202.

Easton, R. D., Greene, A. J., DiZio, P., and Lackner, J. R. (1998). Auditory cues for orientation and postural control in sighted and congenitally blind people. *Experimental Brain Research*, 118(4):541–550.

Edwards, M., O'Mahonys, S., Ibbotson, M. R., and Kohlhagen, S. (2010). Vestibular stimulation affects optic-flow sensitivity. *Perception*, 39(10):1303–1310.

Eguchi, S., Hirose, G., and Miaki, M. (2019). Vestibular symptoms in acute hemispheric strokes. *Journal of Neurology*, 266(8):1852–1858.

Ehrenstein, W. H. and Reinhardt-Rutland, A. H. (1996). A cross-modal aftereffect: Auditory displacement following adaptation to visual motion. *Perceptual and Motor Skills*, 82(1):23–26.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433.

Etienne, A. S. and Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus*, 14(2):180–192.

Ferrè, E. R., Alsmith, A. J., Haggard, P., and Longo, M. R. (2021). The vestibular system modulates the contributions of head and torso to egocentric spatial judgements. *Experimental Brain Research*, 239(7):2295–2302.

Ferrè, E. R. and Haggard, P. (2020). Vestibular cognition: State-of-the-art and future directions. *Cognitive Neuropsychology*, 37(7-8):413–420.

Ferrè, E. R., Longo, M. R., Fiori, F., and Haggard, P. (2013). Vestibular modulation of spatial perception. *Frontiers in Human Neuroscience*, 7(OCT):1–7.

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., and Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *Journal of Neuroscience*, 29(49):15601–15612.

Fink, G. R., Marshall, J. C., Weiss, P. H., Stephan, T., Grefkes, C., Shah, N. J., Zilles, K., and Dieterich, M. (2003). Performing allocentric visuospatial judgments with induced distortion of the egocentric reference frame: an fMRI study with clinical implications. *NeuroImage*, 20(3):1505–1517.

Finocchietti, S., Cappagli, G., and Gori, M. (2015). Encoding audio motion: spatial impairment in early blind individuals. *Frontiers in psychology*, 6(1357):1–6.

Finocchietti, S., Cappagli, G., and Gori, M. (2017). Auditory spatial recalibration in congenital blind individuals. *Frontiers in Neuroscience*, 11(FEB):1–7.

Fitzpatrick, R. C. and Day, B. L. (2004). Probing the human vestibular system with galvanic stimulation. *Journal of Applied Physiology*, 96(6):2301–2316.

Francl, A. and McDermott, J. H. (2022). Deep neural network models of sound localization reveal how perception is adapted to real-world environments. *Nature Human Behaviour*, 6(January).

Fraser, L. E., Makooie, B., and Harris, L. R. (2015). The Subjective Visual Vertical and the Subjective Haptic Vertical Access Different Gravity Estimates. *PLoS ONE*, 10(12):1–20.

Frissen, I., Campos, J. L., Souman, J. L., and Ernst, M. O. (2011). Integration of vestibular and proprioceptive signals for spatial updating. *Experimental Brain Research*, 212(2):163–176.

Gallagher, M., Choi, R., and Ferrè, E. R. (2020). Multisensory Interactions in Virtual Reality: Optic Flow Reduces Vestibular Sensitivity, but only for Congruent Planes of Motion. *Multisensory Research*, 33(6):625–644.

Gandemer, L., Parseihian, G., Kronland-Martinet, R., and Bourdin, C. (2017). Spatial cues provided by sound improve postural stabilization: Evidence of a spatial auditory map? *Frontiers in Neuroscience*, 11(357):1–11.

Genzel, D., Schutte, M., Owen Brimijoin, W., MacNeilage, P. R., and Wiegrebe, L. (2018). Psychophysical evidence for auditory motion parallax. *Proceedings of the National Academy of Sciences of the United States of America*, 115(16):4264–4269.

Gillingham, K. K. and Previc, F. H. (1993). Spatial orientation in flight. *AL- TR-1993–0022. Brooks Air Force Base, Armstrong Laboratory, Texas.*

Glasauer, S., Amorim, M. A., Viaud-Delmon, I., and Berthoz, A. (2002). Differential effects of labyrinthine dysfunction on distance and direction during blindfolded walking of a triangular path. *Experimental Brain Research*, 145(4):489–497.

Gori, M. (2015). Multisensory integration and calibration in children and adults with and without sensory and motor disabilities. *Multisensory Research*, 28(1-2):71–99.

Gori, M., Campus, C., Signorini, S., Rivara, E., and Bremner, A. J. (2021). Multisensory spatial perception in visually impaired infants. *Current Biology*, 31(22):5093–5101.e5.

Gori, M., Cappagli, G., Baud-Bovy, G., and Finocchietti, S. (2017). Shape perception and navigation in blind adults. *Frontiers in Psychology*, 8(JAN):1–12.

Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, 18(9):694–698.

Gori, M., Sandini, G., and Burr, D. (2012). Development of visuo-auditory integration in space and time. *Frontiers in Integrative Neuroscience*, 6(SEPTEMBER):1–8.

Gozli, D. G., Aslam, H., and Pratt, J. (2016). Visuospatial cueing by self-caused features: Orienting of attention and action–outcome associative learning. *Psychonomic Bulletin and Review*, 23(2):459–467.

Grabherr, L., Nicoucar, K., Mast, F. W., and Merfeld, D. M. (2008). Vestibular thresholds for yaw rotation about an earth-vertical axis as a function of frequency. *Experimental Brain Research*, 186(4):677–681.

Greenlee, M. W., Frank, S. M., Kaliuzhna, M., Blanke, O., Bremmer, F., Churan, J., Cuturi, L. F., Macneilage, P. R., and Smith, A. T. (2016). Multisensory integration in self-motion perception. *Multisensory Research*, 29(6-7):525–556.

Gröhn, M., Lokki, T., and Takala, T. (2005). Comparison of Auditory, Visual, and Audiovisual Navigation in a 3D Space. *ACM Transactions on Applied Perception (TAP)*, 2(4):564–570.

Gu, Y. (2018). Vestibular signals in primate cortex for self-motion perception. *Current Opinion in Neurobiology*, 52:10–17.

Gu, Y., Angelaki, D. E., and DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11(10):1201–1210.

Gu, Y., DeAngelis, G. C., and Angelaki, D. E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nat Neurosci.*, 10(8):1038–1047.

Guldin, W. O. and Grüsser, O. J. (1998). Is there a vestibular cortex? *Trends in Neurosciences*, 21(6):254–259.

Harris, L. R. (2022). The Perception of Self-Orientation. *Oxford Research Encyclopedia of Psychology*.

Hautus, M. J., Macmillan, N. A., and Creelman, C. D. (2021). *Detection Theory. A User's Guide*.

Hilliard, D., Passow, S., Thurm, F., Schuck, N. W., Garthe, A., Kempermann, G., and Li, S. C. (2019). Noisy galvanic vestibular stimulation modulates spatial memory in young healthy adults. *Scientific Reports*, 9(1):1–11.

Honda, A., Maeda, K., Sakamoto, S., and Suzuki, Y. (2022). Effects of visually induced self-motion on sound localization accuracy. *Applied Sciences (Switzerland)*, 12(1).

Howard, I. P. and Hu, G. (2001). Visually induced reorientation illusions. 30(ii):583–600.

Ishikawa, T. (2023). Individual Differences and Skill Training in Cognitive Mapping: How and Why People Differ. *Topics in Cognitive Science*, 15(1):163–186.

Ishikawa, T. and Zhou, Y. (2020). Improving cognitive mapping by training for people with a poor sense of direction. *Cognitive Research: Principles and Implications*, 5(1):1–19.

Israël, I., Sievering, D., and Koenig, E. (1995). Self-rotation estimate about the vertical axis. *Acta Oto-Laryngologica*, 115(1):3–8.

Ivanenko, Y., Grasso, R., Israël, I., and Berthoz, A. (1997). Spatial orientation in humans: Perception of angular whole-body displacements in two-dimensional trajectories. *Experimental Brain Research*, 117(3):419–427.

Jabbari, Y., Kenney, D. M., von Mohrenschildt, M., and Shedden, J. M. (2021). Vestibular cues improve landmark-based route navigation: A simulated driving study. *Memory and Cognition*.

Jetzschke, S., Ernst, M. O., Froehlich, J., and Boeddeker, N. (2017). Finding home: Landmark ambiguity in human navigation. *Frontiers in Behavioral Neuroscience*, 11(132):1–15.

Juurmaa, J. and Lehtinen-Railo, S. (1994). Visual experience and access to spatial knowledge. *Journal of Visual Impairment and Blindness*, 88(2 I):157–170.

Kalia, A. A., Schrater, P. R., and Legge, G. E. (2013). Combining Path Integration and Remembered Landmarks When Navigating without Vision. *PLoS ONE*, 8(9).

Kaliuzhna, M., Prsa, M., Gale, S., Maire, R., and Blanke, O. (2018). Optimal visuo-vestibular integration for self-motion perception in patients with unilateral vestibular loss. *Neuropsychologia*, 111:112–116.

Kapralos, B., Zikovitz, D., Jenkin, M. R., and Harris, L. R. (2004). Auditory Cues in the Perception of Self Motion. *In Audio Engineering Society Convention. Audio Engineering Society.*, (116):1–14.

Karim, A. M., Rumalla, K., King, L. A., and Hullar, T. E. (2018). The effect of spatial auditory landmarks on ambulation. *Gait and Posture*, 60(August 2016):171–174.

Karimpur, H. and Hamburger, K. (2016). Multimodal integration of spatial information: The influence of object-related factors and self-reported strategies. *Frontiers in Psychology*, 7(SEP):1–12.

Karmali, F., Goodworth, A. D., Valko, Y., Leeder, T., Peterka, R. J., and Merfeld, D. M. (2021). The role of vestibular cues in postural sway. *Journal of Neurophysiology*, 125(2):672–686.

Karn, T. and Cinelli, M. E. (2019). The effect of galvanic vestibular stimulation on path trajectory during a path integration task. *Quarterly Journal of Experimental Psychology*, 72(6):1550–1560.

Kaski, D., Quadir, S., Nigmatullina, Y., Malhotra, P. A., Bronstein, A. M., and Seemungal, B. M. (2016). Temporoparietal encoding of space and time during vestibular-guided orientation. *Brain*, 139(2):392–403.

Kim, J., Barnett-Cowan, M., and Macpherson, E. A. (2013). Integration of auditory input with vestibular and neck proprioceptive information in the interpretation of dynamic sound localization cues. *Proceedings of Meetings on Acoustics*, 19.

Kitagawa, N. and Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, 416(6877):172–174.

Knill, D. C. and Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719.

Kolarik, A. J., Moore, B. C., Zahorik, P., Cirstea, S., and Pardhan, S. (2016). Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, and Psychophysics*, 78(2):373–395.

Konkle, T. and Moore, C. I. (2009). What can crossmodal aftereffects reveal about neural representation and dynamics? *Communicative and Integrative Biology*, 2(6):479–481.

Kontsevich, L. L. and Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39(16):2729–2737.

Lackner, J. R. and DiZio, P. (2005). Vestibular, proprioceptive, and haptic contributions to spatial orientation. *Annual Review of Psychology*, 56:115–147.

Lessard, N., Paré, M., Lepore, F., and Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature*, 395(6699):278–280.

Lewald, J. and Karnath, H. O. (2000). Vestibular influence on human auditory space perception. *Journal of Neurophysiology*, 84(2):1107–1111.

Li, Z. and Durgin, F. H. (2016). Perceived azimuth direction is exaggerated : Converging evidence from explicit and implicit measures. *Journal of Vision (2016)*, 16:1–19.

Lobel, E., Kleine, J. F., Bihan, D. L. E., Leroy-willig, A., Berthoz, A., Kleine, J. F., Bihan, D. L., and Leroy-willig, A. (1998). Functional MRI of Galvanic Vestibular Stimulation. *Journal of neurophysiology,*, 80(5):2699–2709.

Long, R. and Giudice, N. (2010). Establishing and maintaining orientation for mobility. *Foundations of orientation and mobility*, 1:45–62.

Loomis, J. M. (1999). *Human navigation by path integration*.

Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., and Fry, P. A. (1993). Nonvisual Navigation by Blind and Sighted: Assessment of Path Integration Ability. *Journal of Experimental Psychology: General*, 122(1):73–91.

Loomis, J. M., Silva, J. A. D., Philbeck, J. W., Fukusima, S. S., Loomis, J. M., Silva, J. A. D., and Philbeck, J. W. (1996). Visual Perception of Location and Distance. 5(3):72–77.

Lopez, C. and Blanke, O. (2011). The thalamocortical vestibular system in animals and humans. *Brain Research Reviews*, 67(1-2):119–146.

Lopez, C., Lenggenhager, B., and Blanke, O. (2010). How vestibular stimulation interacts with illusory hand ownership. *Consciousness and Cognition*, 19(1):33–47.

Lubetzky, A. V., Gospodarek, M., Arie, L., Kelly, J., Roginska, A., and Cosetti, M. (2020). Auditory Input and Postural Control in Adults: A Narrative Review. *JAMA Otolaryngology - Head and Neck Surgery*, 146(5):480–487.

MacDougall, H. G., Moore, S. T., Curthoys, I. S., and Black, F. O. (2006). Modeling postural instability with Galvanic vestibular stimulation. *Experimental Brain Research*, 172(2):208–220.

Mackrous, I. and Simoneau, M. (2011). Visuo-vestibular interaction: Predicting the position of a visual target during passive body rotation. *Neuroscience*, 195:45–53.

Macmillan, N. A. and Creelman, C. D. (1991). *Detection theory: A user's guide*. Psychology press.

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

MacNeilage, P. R., Zhang, Z., DeAngelis, G. C., and Angelaki, D. E. (2012). Vestibular facilitation of optic flow parsing. *PLoS ONE*, 7(7).

Maguire, E. A., Spiers, H. J., Good, C. D., Hartley, T., Frackowiak, R. S., and Burgess, N. (2003). Navigation expertise and the human hippocampus: A structural brain imaging analysis. *Hippocampus*, 13(2):250–259.

Marlinsky, V. V. (1999). Vestibular and vestibulo-proprioceptive perception of motion in the horizontal plane in blindfolded man - II. Estimations of rotations about the earth-vertical axis. *Neuroscience*, 90(2):395–401.

Mcmanus, M. and Harris, L. R. (2023). Enhancement of visual cues to self-motion during a visual/vestibular conflict. *PLoS ONE*, 18(3):1–14.

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

Moore, S. T., MacDougall, H. G., Peters, B. T., Bloomberg, J. J., Curthoys, I. S., and Cohen, H. S. (2006). Modeling locomotor dysfunction following spaceflight with Galvanic vestibular stimulation. *Experimental Brain Research*, 174(4):647–659.

Morey, R. D. and Rouder, J. N. (2018). BayesFactor: computation of bayes factors for common designs.

Moser, I., Grabherr, L., Hartmann, M., and Mast, F. W. (2015). Self-motion direction discrimination in the visually impaired. *Experimental Brain Research*, 233(11):3221–3230.

Myers, J. C., Mock, J. R., and Golob, E. J. (2020). Sensorimotor Integration Can Enhance Auditory Perception. *Scientific Reports*, 10(1):16–18.

Nardi, D., Anzures, B. J., Clark, J. M., and Griffith, B. V. (2019). Spatial reorientation with non-visual cues: Failure to spontaneously use auditory information. *Quarterly journal of experimental psychology (2006)*, 72(5):1141–1154.

Nardi, D., Carpenter, S., Johnson, S., Gilliland, G., Melo, V., Pugliese, R., Coppola, V., and Kelly, D. (2020). Spatial Reorientation with a Geometric Array of Auditory Cues. *Quarterly Journal of Experimental Psychology*, page 174702182091329.

Nardini & Cowie (2012). *The development of multisensory balance, locomotion, orientation and navigation*.

Nardini, M., Jones, P., Bedford, R., and Braddick, O. (2008). Development of Cue Integration in Human Navigation. *Current Biology*, 18(9):689–693.

Negen, J., Heywood-Everett, E., Roome, H. E., and Nardini, M. (2018). Development of allocentric spatial recall from new viewpoints in virtual reality. *Developmental Science*, 21(1).

Newcombe, N. S. (2019). Navigation and the developing brain. *Journal of Experimental Biology*, 222(Suppl_1), jeb186460.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1):97–113.

Oppenländer, K., Utz, K. S., Reinhart, S., Keller, I., Kerkhoff, G., and Schaadt, A. K. (2015). Subliminal galvanic-vestibular stimulation recalibrates the distorted visual and tactile subjective vertical in right-sided stroke. *Neuropsychologia*, 74:178–183.

Pasqualotto, A. and Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neuroscience and Biobehavioral Reviews*, 36(4):1179–1187.

Pasqualotto, A., Spiller, M. J., Jansari, A. S., and Proulx, M. J. (2013). Visual experience facilitates allocentric spatial representation. *Behavioural Brain Research*, 236(1):175–179.

Petrini, K., Caradonna, A., Foster, C., Burgess, N., and Nardini, M. (2016). How vision and self-motion combine or compete during path reproduction changes with age. *Scientific Reports*, 6:1–10.

Prins, N. (2016). *Psychophysics: a practical introduction*. 2nd edition.

Prins, N. and 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(JUL):1–14.

Prsa, M., Gale, S., and Blanke, O. (2012). Self-motion leads to mandatory cue fusion across sensory modalities. *Journal of Neurophysiology*, 108(8):2282–2291.

Raiser, T. M., Flanagin, V. L., Duering, M., van Ombergen, A., Ruehl, R. M., and zu Eulenburg, P. (2020). The human corticocortical vestibular network. *NeuroImage*, 223:117362.

Ramkhalawansingh, R., Butler, J. S., and Campos, J. L. (2018). Visual-vestibular integration during self-motion perception in younger and older adults. *Psychology and Aging*, 33(5):798–813.

Renier, L., De Volder, A. G., and Rauschecker, J. P. (2014). Cortical plasticity and preserved function in early blindness. *Neuroscience and Biobehavioral Reviews*, 41:53–63.

Ricciardi, E., Bottari, D., Ptito, M., Roder, B., and Pietrini, P. (2020). The sensory-deprived brain as a unique tool to understand brain development and function. *Neuroscience and Biobehavioral Reviews*, 108:78–82.

Riecke, B. E. (2016). Using spatialized sound to enhance self-motion perception in virtual environments and beyond: Auditory and multi-modal contributions. *Canadian Acoustics - Acoustique Canadienne*, 44(3):148–149.

Rieser, J. J., Ashmead, D. H., Talor, C. R., and Youngquist, G. A. (1990). Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception*, 19(5):675–689.

Rohde, M., Van Dam, L. C., and Ernst, M. O. (2016). Statistically optimal multisensory cue integration: A practical tutorial. *Multisensory Research*, 29(4-5):279–317.

Ronca, A. E., Fritzsch, B., Bruce, L. L., and Alberts, J. R. (2008). Orbital Spaceflight During Pregnancy Shapes Function of Mammalian Vestibular System. *Behavioral Neuroscience*, 122(1):224–232.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, 16(2):225–237.

Rumalla, K., Karim, A. M., and Hullar, T. E. (2015). The effect of hearing aids on postural stability. *Laryngoscope*, 125(3):720–723.

Salomon, R., Szpiro-Grinberg, S., and Lamy, D. (2011). Self-motion holds a special status in visual processing. *PLoS ONE*, 6(10):2–8.

Seemungal, B. M. (2015). The Components of Vestibular Cognition — Motion Versus Spatial Perception. 28:507–524.

Seemungal, B. M., Glasauer, S., Gresty, M. A., and Bronstein, A. M. (2007). Vestibular perception and navigation in the congenitally blind. *Journal of Neurophysiology*, 97(6):4341–4356.

Seemungal, B. M., Rizzo, V., Gresty, M. A., Rothwell, J. C., and Bronstein, A. M. (2008). Posterior parietal rTMS disrupts human Path Integration during a vestibular navigation task. *Neuroscience Letters*, 437(2):88–92.

Seiwerth, I., Jonen, J., Rahne, T., Schwesig, R., Lauenroth, A., Hullar, T. E., and Plontke, S. K. (2018). Influence of hearing on vestibulospinal control in healthy subjects. *Hno*, 66(July):49–55.

Seriès, P. and Stocker, A. A. (2009). Is the Homunculus "Aware" of Sensory Adaptation? *Neural computation*, 21(12):3271–3304.

Setti, W., Cuturi, L. F., Cocchi, E., and Gori, M. (2018). A novel paradigm to study spatial memory skills in blind individuals through the auditory modality. *Scientific Reports*, 8(1):1–10.

Setti, W., Engel, I., Cuturi, L. F., Gori, M., , and Picinali, L. (2021). The Audio-Corsi: an acoustic virtual reality-based technological solution for evaluating audio-spatial memory abilities. *Journal on Multimodal User Interfaces*, 1-12.

Shayman, C. S., Peterka, R. J., Gallun, F. J., Oh, Y., Chang, N. Y. N., and Hullar, T. E. (2020). Frequency-dependent integration of auditory and vestibular cues for self-motion perception. *Journal of neurophysiology*, 123(3):936–944.

Shirai, N. and Ichihara, S. (2012). Reduction in sensitivity to radial optic-flow congruent with ego-motion. *Vision Research*, 62:201–208.

Siegler, I., Viaud-Delmon, I., Israel, I., and Berthoz, A. (2000). Self-motion perception during a sequence of whole-body rotations in darkness. *Experimental Brain Research*, 134(1):66–73.

Sjolund, L. (2016). Bayesian integration of spatial navigation cues. *Graduate Theses and Dissertations*.

Sjolund, L. A. (2014). Cue integration and competition during navigation.

Sjolund, L. A., Kelly, J. W., and McNamara, T. P. (2018). Optimal combination of environmental cues and path integration during navigation. *Memory and Cognition*, 46(1):89–99.

Smith, P. F. (2012). Interactions between the vestibular nucleus and the dorsal cochlear nucleus: Implications for tinnitus. *Hearing Research*, 292(1-2):80–82.

Stephan, T., Deutschländer, A., Nolte, A., Schneider, E., Wiesmann, M., Brandt, T., and Dieterich, M. (2005). Functional MRI of galvanic vestibular stimulation with alternating currents at different frequencies. *NeuroImage*, 26(3):721–732.

Stevens, M. N., Barbour, D. L., Gronski, M. P., and Hullar, T. E. (2017). Auditory contributions to maintaining balance. *Journal of Vestibular Research: Equilibrium and Orientation*, 26(5-6):433–438.

Sun, H. J., Campos, J. L., Young, M., Chan, G. S., and Ellard, C. G. (2004). The contributions of static visual cues, nonvisual cues, and optic flow in distance estimation. *Perception*, 33(1):49–65.

Thinus-blanc, C. and Gaunet, F. (1997). Representation of Space in Blind Persons: Vision as a Spatial Sense? Psychological bulletin(1):20–42.

Torchiano, M. (2016). Effsize: efficient effect size computation (R package). *Vienna: R Foundation for Statistical Computing. software*.

Väljamäe, A. (2009). Auditorily-induced illusory self-motion: A review. *Brain Research Reviews*, 61(2):240–255.

Valko, Y., Lewis, R. F., Priesol, A. J., and Merfeld, D. M. (2012). Vestibular labyrinth contributions to human whole-body motion discrimination. *Journal of Neuroscience*, 32(39):13537–13542.

Viaud-Delmon, I. and Warusfel, O. (2014). From ear to body: The auditory-motor loop in spatial cognition. *Frontiers in Neuroscience*, 8(SEP):1–9.

Wei, X. X. and Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience*, 18(10):1509–1517.

Werkhoven, P., Van Erp, J. B., and Philippi, T. G. (2014). Navigating virtual mazes: The benefits of audiovisual landmarks. *Displays*, 35(3):110–117.

Wolbers, T. and Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences*, 14(3):138–146.

Wright, M. (2005). Open Sound Control: An enabling technology for musical networking. *Organised Sound*, 10(3):193–200.

Xie, Y., Bigelow, R. T., Frankenthaler, S. F., Studenski, S. A., Moffat, S. D., and Agrawal, Y. (2017). Vestibular loss in older adults is associated with impaired spatial navigation: Data from the triangle completion task. *Frontiers in Neurology*, 8(APR):1–9.

Yoder, R. M. and Taube, J. S. (2014). The vestibular contribution to the head direction signal and navigation. *Frontiers in Integrative Neuroscience*, 8(APR):1–13.

Zahorik, P., Brungart, D. S., and 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.

Zhao, M. and Warren, W. H. (2015a). Environmental stability modulates the role of path integration in human navigation. *Cognition*, 142:96–109.

Zhao, M. and Warren, W. H. (2015b). How You Get There From Here: Interaction of Visual Landmarks and Path Integration in Human Navigation. *Psychological Science*, 26(6):915–924.

Zhao, M. and Warren, W. H. (2018). Non-Optimal Perceptual Decision in Human Navigation. *Behavioral and Brain Sciences*, 469(3):319–323.

Zhong, X. and Yost, W. A. (2013). Relationship between postural stability and spatial hearing. *Journal of the American Academy of Audiology*, 24(09):782–788.

Zink, R., Bucher, S. F., Weiss, A., Brandt, T., and Dieterich, M. (1998). Effects of galvanic vestibular stimulation on otolithic and semicircular canal eye movements and perceived vertical. *Electroencephalography and Clinical Neurophysiology*, 107(3):200–205.