AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Human Perception and Performance

ISSN: 0096-1523

2022, Vol. 48, No. 2, 174–189
https://doi.org/10.1037/xhp0000973

Interindividual Differences Influence Multisensory Processing During Spatial Navigation

Silvia Zanchi[1, 2, 3], Luigi F. Cuturi[1], Giulio Sandini[2], and Monica Gori[1]
[1] Unit for Visually Impaired People, Istituto Italiano di Tecnologia, Genova, Italy
[2] Robotics Brain and Cognitive Sciences, Istituto Italiano di Tecnologia, Genova, Italy
[3] DIBRIS Department, Università di Genova

When moving through space, we encode multiple sensory cues that guide our orientation through the environment. The integration between visual and self-motion cues is known to improve navigation. However, spatial navigation may also benefit from multisensory external signals. The present study aimed to investigate whether humans combine auditory and visual landmarks with improving their navigation abilities. Two experiments with different cue reliability were conducted. In both, participants' task was to return an object to its original location by using landmarks, which could be visual-only, auditory-only, or audiovisual. We took error and variability of object relocation distance as measures of accuracy and precision. To quantify interference between cues and assess their weights, we ran a conflict condition with a spatial discrepancy between visual and auditory landmarks. Results showed comparable accuracy and precision when navigating with visual-only and audiovisual landmarks but greater error and variability with auditory-only landmarks. Splitting participants into two groups based on given unimodal weights revealed that only subjects who associated similar weights to auditory and visual cues showed precision benefit in audiovisual conditions. These findings suggest that multisensory integration occurs depending on idiosyncratic cue weighting. Future multisensory procedures to aid mobility must consider individual differences in encoding landmarks.

Public Significance Statement

To navigate efficiently, we combined different sources of information available in the environment. Previous literature on spatial navigation focused on studying the integration between unisensory information external to the body such as visual cues and body-centered cues such as self-motion information based on vestibular and proprioceptive signals. Still, it remains to examine the mechanism underlying the integration of multisensory spatial cues exclusively external to the body, such as auditory and visual cues. The present study investigated the integration between auditory and visual points of reference during navigation. Our results revealed that only navigators who perceived auditory and visual cues as equally reliable benefit from the multisensory environment, achieving a more precise performance. Finding these interindividual differences in a homogeneous sample of adult participants emphasizes the role of idiosyncratic perceptual characteristics in spatial cognition and multisensory perception, likely explaining previous contrasting results on multisensory integration during spatial navigation. The outcome of this work has clinical relevance, highlighting the necessity to consider individual differences in perception to develop novel multisensory rehabilitation procedures for orientation and mobility in the case of sensory and motor disabilities.

Keywords: spatial navigation, multisensory integration, individual differences

Supplemental materials: https://doi.org/10.1037/xhp0000973.supp

---

Silvia Zanchi https://orcid.org/0000-0003-0572-0480
Luigi F. Cuturi https://orcid.org/0000-0001-8144-3740
Giulio Sandini https://orcid.org/0000-0003-3324-985X
Monica Gori https://orcid.org/0000-0002-5616-865X

The research presented here has been supported and funded by the Unit for Visually Impaired People, Istituto Italiano di Tecnologia (Genova, Italy). The research was partially supported by the MYSpace project (principal investigator Monica Gori), which has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant 948349). We thank all participants who took part in the experiments. All raw data have been made publicly available on the Zenodo repository (https://zenodo.org/record/5379613).

Silvia Zanchi, Luigi F. Cuturi, and Monica Gori conceived the studies and designed the experiments. Silvia Zanchi carried out the experiments, analyzed the data, and wrote the first draft of the article. All authors reviewed and approved the final version of the article for submission.

Correspondence concerning this article should be addressed to Monica Gori, Unit for Visually Impaired People, Istituto Italiano di Tecnologia, Via Enrico Melen 83, Genova 16152, Italy. Email: monica.gori@iit.it

Navigating and orienting through space are fundamental activities for our survival. To achieve efficient spatial navigation, we need to simultaneously combine multiple internal and external cues that are available in the environment. Many researchers acknowledge that humans can optimally integrate different sources of information, often modeled in Bayesian terms (Alais & Burr, 2004; Ernst & Banks, 2002; Fetsch et al., 2009; Gori et al., 2012). This ability occurs during spatial navigation as well (Bates & Wolbers, 2014; Butler et al., 2010; Chen et al., 2017; Nardini et al., 2008; Sjolund, 2016; Sjolund et al., 2018). Navigation with multisensory cues should reduce variability, meaning an increased precision and accuracy enhancement compared to navigation performance guided by a single cue (Bates & Wolbers, 2014; Nardini et al., 2008). According to the Bayesian maximum likelihood estimation (MLE) model, the amount of predicted improvement that occurs as a result of multisensory cues depends on the given weighting to each cue—that is, their reliability. The improvement over the best unimodal performance is at its maximum when the cues are equally reliable, while if one cue is more reliable than the other, the first dominates the second, and this strongly influences the behavior (Bates & Wolbers, 2014; Ernst & Banks, 2002; Gori et al., 2012).

Previous studies have investigated Bayesian optimal integration between information coming from visual landmarks, that is, visual fixed points of reference in space, and from self-motion cues, including vestibular and proprioceptive signals (Bates & Wolbers, 2014; Chen et al., 2017; Nardini et al., 2008; Sjolund et al., 2018). However, multisensory integration studies in navigation have shown contrasting results. For instance, some studies reported improvements both in accuracy and precision (Bates & Wolbers, 2014; Nardini et al., 2008). Other researchers have found clear improvements only for precision and single-cue dominance for accuracy (Zhao & Warren, 2015b). Finally, others have observed responses that have been consistent with a cue-competition model; in such cases, multiple cues compete with each other, which results in no increase in precision and constant error (response accuracy) between those found in unimodal conditions (Petrini et al., 2016). The heterogeneity in the investigation of multisensory integration during navigation raises the question of whether idiosyncratic factors may influence accuracy and precision when accomplishing the tasks. In the context of spatial information processing, a previous work introduced the notion of subjective discrepancy between multiple cues to explain the mixed prior findings (Cheng et al., 2007). According to this concept, the researchers suggested that regardless of the physical discrepancy between cues, if two cues are perceived as being largely discrepant, one cue will dominate over the other; if two cues are perceived as being similar, integration will occur (Cheng et al., 2007). Along these lines, previous research ascribed the absence of cue integration in a multisensory environment to perceived discrepancy between visual and self-motion cues used to navigate (Petrini et al., 2016). The differences in methodologies and subjective discrepancy across participants may be responsible for the divergent findings in multisensory navigation.

In a real-world environment, it is common to find visual landmarks associated with sounds—for example, a fountain that has the sound of water flowing or a bus at the bus station alongside the sound of its running engine. In addition, when vision is not reliable or absent such as in the case of visual impairments, spatialized auditory information can provide essential cues to orient and navigate. While the combination between self-motion and visual information is extensively investigated, it is still not clear whether spatialized auditory information in the surroundings can be combined with external visual cues. Recent research has investigated how auditory landmarks are used to reorient and navigate through space in a homing task (Jetzschke et al., 2017). In this research, the participants had to return to a previously learned "home" location by using landmarks that were placed at different locations. Results show that there were similarities between using visual and auditory landmarks to accomplish this task. Auditory cues alone can also help reorientation as they are used as geometric cues. Similar to what has been previously found regarding visual cues, an array of auditory landmarks can be encoded as a geometric configuration; in turn, this can provide people with information about the distance and direction among the auditory landmarks (Nardi et al., 2020). In addition, auditory sources in space provide people with enough spatial information to successfully orient themselves in an auditory equivalent of the Morris water maze (Viaud-Delmon & Warusfel, 2014), which is a classical paradigm used to assess spatial learning and memory in animals.

It is reasonable to believe that the multisensory integration between visual and auditory cues may improve navigation tasks. Investigating whether external audiovisual information is integrated would aid our understanding of navigation strategies in a realistic environment, in which multiple sensory cues are available. Moreover, landmarks provide the navigators with two pieces of information: First, they inform them about the position of the person that is moving relative to landmarks' locations, thus providing an egocentric reference; second, they give metric information about the relations among multiple landmarks in space, which allows for allocentric processing of space. However, to this point, few studies have investigated how multimodal landmarks can be processed and exploited during navigation tasks. Compared to using visual-only and auditory-only landmarks, the combination of audiovisual landmarks in virtual way-finding tasks leads participants to select shorter routes and to travel faster (Werkhoven et al., 2014); moreover, it improves their recognition of landmarks' positions in space (Karimpur & Hamburger, 2016). Likewise, compared with single-cue performances, audiovisual cues in a virtual environment lead to the most efficient navigation strategy during game-like experiences (Gröhn et al., 2005). 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 present study aimed to investigate (a) how combined audiovisual landmarks influence navigation performance and (b) whether audiovisual landmarks are optimally integrated to reduce variability. We hypothesized different outcomes according to the idiosyncratic perception of the available sensory cues. Consistent with the prediction of the MLE model (Ernst & Banks, 2002), if participants weighted both auditory and visual signals similarly, we expected to find the optimal multisensory integration of the two sources of information, which would result in reduced variability for the audiovisual trials. Conversely, if participants associated different weights to the two cues, we expected to observe nonoptimal integration; in turn, this would likely result in the predominance of the most reliable sensory information for spatial processing (visual cues). To test our hypotheses, we performed two experiments. In both, the participants had to navigate in a dark room to perform an object relocation task (Nardini et al., 2008). The participants' goal was to relocate a target object to its original position, orienting themselves using only landmark information that could be visual-only, auditory-only, or audiovisual. We also ran a conflict condition that had a spatial discrepancy between the visual and auditory landmarks to assess how participants weighted each cue. Since cue integration occurs when different sources of information are equally reliable (Ernst & Banks, 2002), we manipulated the reliability of the visual cues across the two experiments to account for the differences in spatial reliability between the two sensory modalities (Alais & Burr, 2004), which occur due to the high spatial reliability of the visual system relative to the auditory one. Thus, in Experiment 2, we decreased the reliability of the visual landmarks.

The results from both experiments revealed that there were two distinct groups of individuals, "integrators" and "nonintegrators." Our findings suggested that investigations of multisensory integration during navigation must consider individual differences in the ability to encode external landmarks. Notably, this factor might explain prior divergent results.

Experiment 1

Method

Participants

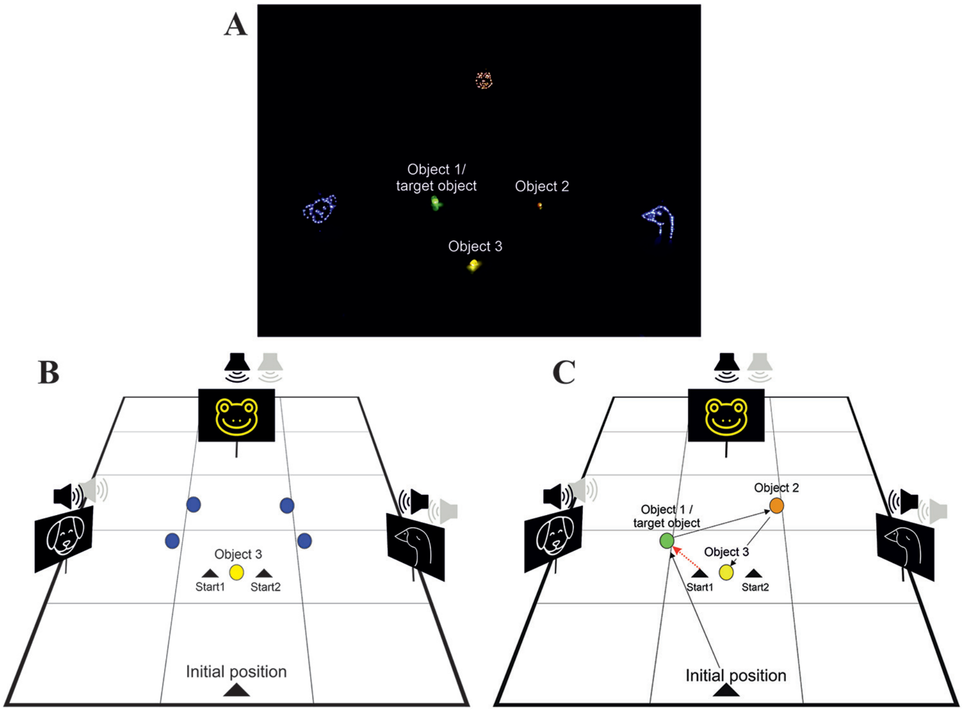
The sample size was determined by performing a priori power analysis, using pooled effect sizes observed in three previous experiments within the same study about cue combination during spatial navigation (Sjolund, 2016) for the main effect of condition on the measure of variability (standard deviation; SD). Specifically, we took into account Experiments 1a, 1b, and 2 because a main effect of the condition was found. We considered this particular effect because variation in SD according to the presence of one or multiple sensory cues is crucial evidence of multisensory integration. Sjolund's previous experiments (Sjolund, 2016) used generalized η^2 (η^2_G) as a measure of effect size for repeated-measures analysis of variance (ANOVA) because it provides comparability across different designs (Bakeman, 2005; Olejnik & Algina, 2003). For this reason, we adopted the same effect size in our analysis (see below in "Data Analysis" sections). The mean of reported effect sizes in Sjolund's (2016) experiments was $\eta^2_G = .20$. To compute the sample size, we performed a power analysis via the function ss.power.wa in the package BUCSS (Anderson & Kelley, 2020) in RStudio (Version 1.2.5033; 2019), the approach of which uses the observed F value and sample size from a previous study to predict the needed sample size based on the desired power and alpha. In Sjolund's experiments, the mean of the observed values of F was 19.17, and the mean sample size was 55. To achieve a power of .9 and an alpha of .05, the power analysis yielded a sample size of 20 participants (alpha prior = .05, assurance = .8). Thus, 21 participants (11 women, M age = 29.2 ± 3.2 years old) took part in the present experiment. All participants reported that they had normal or corrected-to-normal vision and an absence of hearing impairments. The study was approved by the ethics committee of the local health service (Ethical Committee, ASL 3, Genova, Italy), and it was conducted following the World Medical Association (2013). All the subjects provided informed written consent.

Apparatus and Stimuli

The experiment was carried out in a 300 cm × 500 cm dark real-world environment. Participants were required to explore the area and pick up three clearly detectable bright objects, which lay on the floor (Figure 1A). Then, they had to put the first collected object to its original position by orienting themselves with landmarks. The visual landmarks consisted of three hand-made LED arrays, each of which was fixed on a black sheet of paper that measured 21 cm × 29.7 cm. We arranged the LED arrays to resemble the shapes of animal faces, specifically a frog, a dog, and a duck (Figure 1A). The paper sheets with the LED arrays were mounted on adjustable height stands. From the initial position (Figure 1B), the central landmark (the frog) was straight ahead at a distance of 400 cm and at a height of 150 cm. The two lateral landmarks (the dog and the duck) were at 45° to the left and to the right of the initial position at a distance of 208 cm and at a height of 50 cm. The central landmark was made of yellow LED lights, while the lateral ones were made of white LED lights; thus, due to landmarks' colors, the straight-ahead direction was always well detectable from the initial position.

The auditory landmarks consisted of three loudspeakers (Stilgut YB202STGD, 6.6 × 6.6 × 6 cm) that were positioned in the room to correspond with the three visual landmarks. Moreover, they were mounted on the same stands, and they were 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. Each auditory landmark played a sound that semantically corresponded to the image of visual landmarks; specifically, the central one played a croaking frog, the left one played a barking dog, and the right one played a quacking duck. All three sounds were downloaded from a royalty-free sound web archive (https://freesound.org/). We chose animal sounds to make the landmarks distinguishable from one another and semantically congruent to the accompanying visual cues. Similar to previous studies (Jetzschke et al., 2017; Nardi et al., 2020; Viaud-Delmon & Warusfel, 2014), we provided steady auditory landmarks to increase auditory information spatial reliability and make it more comparable with visual cues to spatial navigation. It is plausible to assume that the continuous presence of the sounds would indeed increase the saliency of spatial information, while intermittent sounds are more likely to produce disorientation in the moments of silence, considering the overall poor spatial acuity of auditory information. Thus, during the conditions in which the auditory cues were available, the three sounds were constantly active. To avoid an excessively noisy setting, we regularly presented animal sounds, but never overlapping, with an interval of 800 ms maximum among stimuli. To provide steady sound from the landmark locations when the animal sounds were momentarily silent, we added a pink-noise background different for each animal sound (generated using the Audacity Digital Audio Editor Software) that played continuously (for more details, see Figure S1 in the online supplemental materials). 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 spatial discrepancy

Figure 1
Experimental Setup and Procedure



Note. 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 were visible (the green apple—Object 1/target object, the orange tangerine—Object 2, and the yellow pear—Object 3). Panel B: Experimental setup: Big central triangle indicates the initial position. The circles on the right and on the left represent 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 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 arrow). See the online article for the color version of this figure. "Dog" and "Duck" icons by Iconic, and "Frog" icon by Norbert Kucsera, from thenounproject.com. All icons adapted by Silvia Zanchi.

between visual and auditory landmarks, an additional triplet of loudspeakers was used; at these times, each 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 a pilot experiment showed that this discrepancy was unnoticeable to participants. Eleven participants experienced the left direction of conflict; the other 10 experienced the right direction of conflict.

We chose the position and orientation of visual and auditory landmarks so that from the exploration area, all three visual landmarks were recognizable and all three sounds were audible. The experimenter controlled when both auditory and visual landmarks would start and end by pressing a wireless mouse key, which remotely controlled the sound and light sources via Matlab (R2019b, The MathWorks, United States).

The three objects to be collected consisted of fruit toys made of semitransparent plastic (a green apple—Object 1, an orange tangerine—Object 2, a yellow pear—Object 3). The objects had small LED lights inserted inside to make them glow in the dark (Figure 1A). As occurred in the experiment by Nardini et al. (2008), the objects could be placed in different positions. In particular, the positions of Objects 1 and 2 varied across trials; however, Object 3 remained in the same position for the whole experiment, at 173 cm straight ahead of the initial position. Relative to the position of Object 3, Object 1 and Object 2 could have been placed in four different locations: 40° or 30° to the left or the right, at a distance of 140 or 103 cm. To ensure participants were required to go to both sides of the area, Object 2 was always placed on the opposite side of the room relative to Object 1. The participants experienced each of the four possible positions of Object 1 (henceforth, the target object) once in all conditions. To minimize the effects of cognitive fatigue due to trial long-lasting duration (i.e., 5 minutes), we followed the procedure of Nardini et al. (2008) and participants performed four trials in each condition.

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 (see Figure 1B). We selected the two starts to force participants to use only the landmark information to orient themselves and to make the reference of the Object 3 position unreliable to relocate the objects successfully.

In addition to being in a dark space, the participants were equipped with covering sunglasses that had increased opacity due to the addition of nylon filters over the lenses. Participants who had corrected-to-normal vision wore the sunglasses over their own glasses. The participants wore these specially designed 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; in turn, they were not able to see the stands, the loudspeakers, the floor, the walls, or the ceiling. Ad hoc procedures were applied to sanitize the environment and the setup at the beginning of each session to ensure that the participants and the experimenter were protected from COVID-19.

Design and Procedure

The experiment involved four conditions, depending on which landmark modality was provided: (a) auditory-only, (b) visual-only, (c) both auditory and visual (combined), and (d) both auditory and visual but with a spatial conflict between the auditory and visual information (conflict). After two unimodal practice trials (one auditory, one visual), the participants performed four experimental trials for each condition (for a total of 16 trials). The experiment took place over two separate sessions, which lasted about 45 min each; moreover, they were scheduled 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. In particular, 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 presentation of conditions was pseudorandomized across trials. We implemented this design so we could rule out any potential learning effect or influence among the conditions.

Each trial consisted of two different phases: the exploration phase, in which the subjects collected the objects and had the goal of remembering the target object location relative to the landmarks' configuration, and the response phase, in which the participants tried to relocate the objects to the original location of the target object. A disorientation procedure occurred between the two phases to disrupt the use of gathered self-motion information that could have been attained during the response phase. The experimental procedure was similar to the one used by Nardini and colleagues (2008) that consisted of an inferential spatial navigation task, in which participants had to infer new routes based on previously experienced spatial relationships (see Figure 1C). Indeed, this task allowed us to systematically investigate participants' ability to travel novel paths exploiting exclusively the spatial information conveyed by the configuration of landmarks. The procedure occurred as follows. Before entering the experimental room, the participants read the instructions of the experiment. Subsequently, they were guided through the dark environment, wearing the sunglasses, until they reached a swivel chair that was located at the initial position (see Figure 1B). Once they were seated on the chair, they were given passive noise-canceling headphones, and the experimenter rotated the chair to turn participants' back to the central landmark. Once the experimenter positioned the objects on each trial, the participants were explicitly told on which side of the room the target object was located (right or left) to ensure that they would recognize it. Then, the experimenter switched on the landmarks and asked the participants to take off the headphones and begin the exploration phase. Their primary goal was to pick up a transparent plastic bottle with the target object inside and to remember its original location relative to the landmarks. 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 disoriented for ∼20 s by being spun around in the chair. During the disorientation time, the participants wore the noise-canceling headphones again, and they closed their eyes to prevent them from detecting any external cues. To be certain that the participants lost their orientation, they were asked to point with their finger to the central landmark with their eyes still closed (the "frog"). If they answered correctly, they were rotated for a further 10 s until they were fully disorientated (maximum further disorientation moments = 3). When the disorientation procedure was finished, 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 to open their eyes. They were slowly rotated while seated to face the central landmark. When the landmarks were switched on, the experimenter encouraged the subjects to use all the 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 first session, the participants were guided out of the room with the sunglasses still on so that they could not see the experimental setup. Once the participants concluded the second session, they were allowed to remove the sunglasses and look at the setup. At this time, the experimenter explained to them the aim of the study. Moreover, the experimenter asked all the participants whether they noticed anything unusual during the experiment; no participant was explicitly aware of the spatial conflict.

Data Analysis

Behavioral Analysis

For each trial, the distance between the participant's response and the correct location of the objects was measured in centimeters. For each participant and for each condition, we quantified the performance by computing the mean response error, namely the constant error (CE). Moreover, the variability of errors was taken as an index of precision (the lower the variability, the higher the precision), which was measured as the SD of the responses. Before calculating the mean CE, 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 the 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 auditory condition, two from visual condition, one from combined condition, and three from conflict condition). For the variability measures, no responses met the outlier definition. In addition, one participant was removed from the analyses due to giving multiple outlying responses (n = 8 out of 16 trials) that prevented the calculation of the variables. As such, 20 participants' data were included in the final analyses.

The normality of the variables in each condition was verified using Shapiro-Wilk tests (results from these tests can be seen in the online supplemental materials). In case of normal distributions, we performed a repeated-measures ANOVA on each dependent variable, considering condition as the within factor. Generalized eta squared (η^2_G) was calculated as effect size. We applied a Greenhouse-Geisser correction whenever the Mauchly test suggested that there was a violation of the assumption of sphericity. When appropriate, we performed post hoc comparisons with pairwise t tests, calculating the Cohen's d as the effect size for paired comparisons with the cohen.d function in the effsize package (Torchiano, 2016) in RStudio. As reported in the effsize package documentation, for paired comparisons, the effect size is computed using the approach suggested in Gibbons et al. (1993). In particular, the function applies a correction taking into account the correlation of the two samples (see Borenstein et al., 2009). In case of violation of the assumption of normality, we performed nonparametric 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 evaluated to be significant when they were lower than .05 after applying the Bonferroni correction.

Modeling Analysis and Bayesian Predictions

It was necessary to use four experimental conditions so that we could determine 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 audiovisual 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 & Wolbers, 2014; Nardini et al., 2008). In the conflict condition, the relative proximity to the location given by the auditory cues ($rprox_A$) was calculated as follows:

$$rprox_A = (1/d_A)/(1/d_V + 1/d_A) = d_V/(d_A + d_V) \tag{1}$$

in which d_A and d_V are the predicted response distances based on (shifted) auditory-only and visual-only cues, respectively (Bates & 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 combined audiovisual information, the predicted variance σ^2_{V+A} was calculated as:

$$\sigma^2_{(V+A)} = w_V^2\sigma_V^2 + w_A^2\sigma_A^2 \tag{2}$$

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 1 would correspond to the empirical weights given to the cues.

In contrast, the alternation model predicts that the participants do not integrate cues; rather, they alternate between them. In this case, since the cues would be used as separated sources of information, the variance would increase (Nardini et al., 2008). In this model, the predicted variance σ^2_{V+A} was calculated as follows:

$$\sigma^2_{(V+A)} = p_V(\mu_V^2 + \sigma_V^2) + p_A(\mu_A^2 + \sigma_A^2) - (p_V\mu_V + p_A\mu_A)^2 \tag{3}$$

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, μ_V= 0 and μ_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 & 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 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 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 variance (Ernst & Banks, 2002). We computed the predicted optimal weighting for the auditory landmarks, w_A, and the visual landmarks, w_V, as follows:

$$\begin{aligned} w_A &= (1/\sigma_A^2)/(1/\sigma_V^2 + 1/\sigma_A^2) = \sigma_V^2/(\sigma_A^2 + \sigma_V^2); \\ w_V &= (1/\sigma_V^2)/(1/\sigma_A^2 + 1/\sigma_V^2) = \sigma_A^2/(\sigma_V^2 + \sigma_A^2) \end{aligned} \tag{4}$$

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 1) and predicted weights (Equation 4). 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:

$$\sigma_{(V+A)}^2 = \sigma_A^2 * \sigma_V^2/(\sigma_A^2 + \sigma_V^2) \tag{5}$$

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.

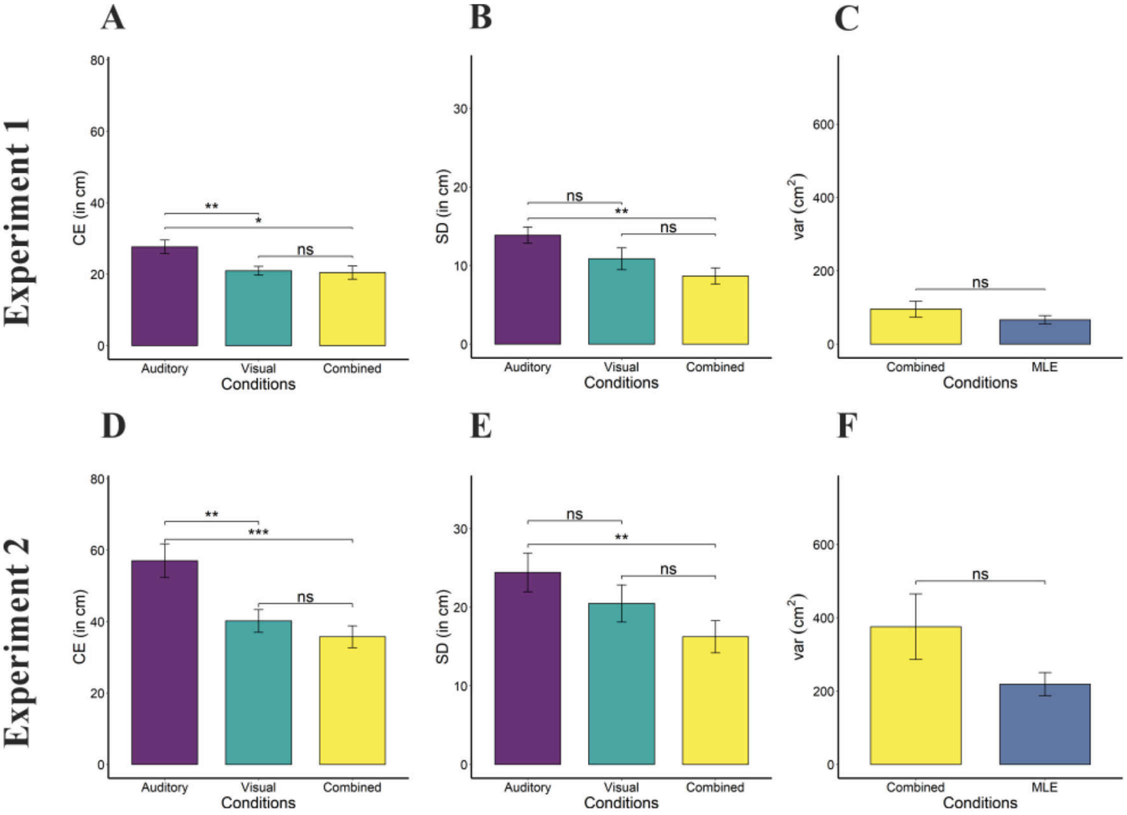
Results

The repeated-measures ANOVA on CE showed there was a main effect of condition, $F(2, 19) = 7.532$, $p = .005$, $\eta_G^2 = .17$, 95% CI [0, .36]. In particular, as shown in Figure 2A, this effect was greater in the auditory condition compared with the visual, $t(19) = 4.11$, *p* adjusted (adj) = .002, Cohen's $d = .89$, [.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 2B), the results suggested there was a main effect of condition on the SD as well, $F(2, 19) = 4.835$, $p = .014$, $\eta_G^2 = .15$, 95% CI [0, .34]. Specifically, the SD was higher, meaning that there was lower precision in the auditory

Figure 2

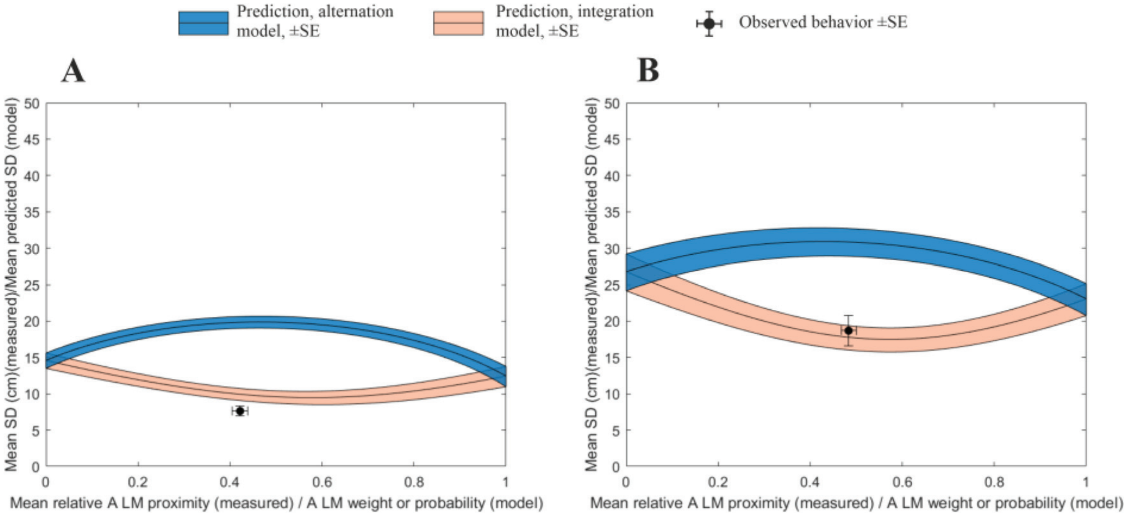
Constant Error (CE), Standard Deviation (SD), and Variance (var) in Each Condition in Experiment 1 (Upper Panels, A, B, and C) and Experiment 2 (Lower Panels, D, E, and F)



Note. In both experiments, CE in the auditory condition was significantly higher than in the other two conditions, which did not differ from each other (Panels A, D). In Expriment 1, SD was higher in the auditory condition if compared with the combined condition, while the other comparisons showed no differences (Panel B). The same pattern of results was found on SD in Experiment 2 (Panel E). Generally, Experiment 2 led to a worse performance, considering higher relocation error and greater variability. Variance in the combined condition did not significantly differ from maximum likelihood estimation (MLE) prediction (Panels C and F). Error bars are standard errors. ns = not significant. * $p < 0.5$. ** $p < 0.01$. *** $p < 0.001$. See the online article for the color version of this figure.

Figure 3

Model Predictions Versus Behavior in the Conflict Condition in Experiment 1 (Panel A) and Experiment 2 (Panel B)



Note. The curves represent the means of functions predicting mean standard deviations (SD) from different auditory landmark weights (integration model, lighter grey) or probabilities (alternation model, darker grey). 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). Note the poorer performance in Experiment 2. A LM = auditory landmarks; *SE* = standard error. See the online article for the color version of this figure.

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 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 the cues were available (Figure 2C). Moreover, we compared the observed behavior in the conflict condition with the predictions made using the integration and the alternation models (Figure 3A). 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 audiovisual 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 unimodal cue. We compared the empirical weights that were assigned to the auditory and visual cues (computed using Equation 1 above) and the weights predicted by the optimal integration (Equation 4). 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]; specifically, the auditory cues were weighted less than the visual ones.

On the one hand, the modeling analysis revealed 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. Moreover, the empirical weights associated with each cue were the same as the predictions. On the other hand, the behavioral results suggest that multisensory integration did not fully occur. 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. Conversely, here, we observed no significant difference between the best unimodal condition (visual condition) and the combined one, even if the combined SD was lower than visual numerically. It is probable that the response variability in the visual condition already reached a plateau; thus, this would have prevented a multisensory improvement of any sort, which would have been in line with findings of previous literature (Chen et al., 2017). In addition, the visual cues revealed to be weighted significantly more than the auditory cues were. This prevented the predicted decreased variability in the combined condition that should have occurred when the unimodal cues had the same level of reliability (Fetsch et al., 2009). If there had been a great discrepancy between the weights associated with each cue, the more reliable cue would dominate the less reliable one, which would have resulted in a shift of the combined probability density toward the dominant sense (Ernst & Banks, 2002). Therefore, we performed a second experiment (Experiment 2), in which we reduced the reliability of the visual cues to make them comparable with the auditory ones; to achieve this parity, we moved all the landmarks farther from the target object locations. Indeed, previous studies showed that reliability decreased with visual landmarks that were farther away compared to closer ones (Chen et al., 2017; Zhao & Warren, 2015b).

Experiment 2

Method

Participants

Experiment 2 was planned to investigate the comparison between unimodal and multimodal conditions, controlling that auditory and visual cues were more comparable. The investigated variables and the expected effect size were the same as Experiment 1. Therefore, no new a priori power analysis was required; to choose the sample size of Experiment 2, we referred again to the previous a priori power analysis, which suggested a sample size of 20. Twenty-one participants took part in Experiment 2 (11 women, *M* age = 26.4 ± 4.9 years old), and none of them participated in Experiment 1.

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 494 cm × 597 cm in the environment to be explored. The position of Object 3 was the same as was used in Experiment 1, while 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 of 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 to 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. Similar to Experiment 1, ad hoc procedures were applied to sanitize the environment and setup at the beginning of each session to protect the participants and the experimenter from acquiring COVID-19.

Design and Procedure

The procedure was the same used in Experiment 1. Before starting the experimental session, participants of Experiment 2 performed three practice trials: two unimodal (one visual, one auditory) like Experiment 1 and one combined. We added this supplemental multisensory practice trial to allow participants to familiarize with the simultaneous presentation of visual and auditory cues. 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.

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 (.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^2 was 36 cm, according to the new selected conflict shift.

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 Figure 2D, CE was greater in the auditory condition when compared with the visual, $t(20) = 3.92$, p adj = .003, Cohen's $d = .88$, [.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 was found in Experiment 1.

In terms of variability (Figure 2E), 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 ± 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 that there was a near optimal reduction of variability when both cues were available (Figure 2F). To compare the model predictions and the behavior in the conflict condition (Figure 3B), 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 observed 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 1) and the weights predicted by the optimal integration (Equation 4). Here, in the conflict condition, the auditory and visual cue weights were .48 ± .02 and .52 ± .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 the auditory landmarks was as reliable as the information given by the visual landmarks.

It should be noted that moving the landmarks farther away, relative to the target object location, generally increased the difficulty of the spatial navigation task. In turn, when compared with the results from Experiment 1, this led to a global worsening of performance for all participants on all the dependent variables.

The absence of a significant improvement in precision when the audiovisual cues were present suggested that in Experiment 2, multisensory behavioral integration did not seem to occur. This conclusion is similar to what was observed in Experiment 1. Conversely, once again, the analysis on the model predictions and weights suggested a tendency to follow the optimal multisensory integration.

Relative Unimodal Weights

In summary, in both experiments, the behavioral results suggested that, in the presence of audiovisual cues, the participants followed the most accurate cue to relocate the objects. As such, the participants had similar performances in the visual and combined conditions, both of which differed from the auditory condition. Conversely, the comparisons between the observed behavior and the models revealed that the participants' performance was nearly consistent with the Bayes optimal prediction.

Despite the similarities between the combined condition and model predictions, the lack of substantial difference between the visual and combined variability prevented us from excluding the possibility that only the visual information determined the performance in the spatial navigation task. It is known that the MLE always predicts that there is greater bimodal precision than unimodal precision when the unimodal estimates are equally precise; the visual and auditory precisions did not differ in either of the two experiments. This suggests that, somehow, the two sensory modalities might be integrated. One possibility is that the individual differences in perception of each cue might have prevented our results from showing significant multisensory precision improvements. Thus, we decided to investigate whether the individual differences in the processing of the visual and auditory cues might induce differences among individual weighting.

Like in the research by Bentvelzen et al. (2009), we split the participants of both experiments into two groups, depending on how disparate 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 4, 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. In such cases, the multisensory variability would not differ from the MLE prediction. Conversely, the participants who associated the disparate weights with the visual and auditory cues would not show a reduction of multisensory variability. In the latter case, the variability in the audiovisual trials would be similar to the best unimodal one (visual). We predicted that the multisensory variance for them would be significantly greater than the MLE prediction. According to their predicted weighting, we divided the participants into two subgroups, following the procedure used in the research by Bentvelzen and colleagues (2009). We assigned participants whose data returned similar unimodal weights (the difference between weights was less than .5; $|w_V - w_A| < .5$) to a first group, which we called "similar weights" (SW; $n = 21$, 12 from Experiment 1, nine from Experiment 2, 11 women). Those that returned dissimilar unimodal weights (difference between weights higher than .5; $|w_V - w_A| > .5$) were assigned to a second group, called "dissimilar weights" (DW; $n = 20$, eight from Experiment 1, 12 from Experiment 2, 11 women).

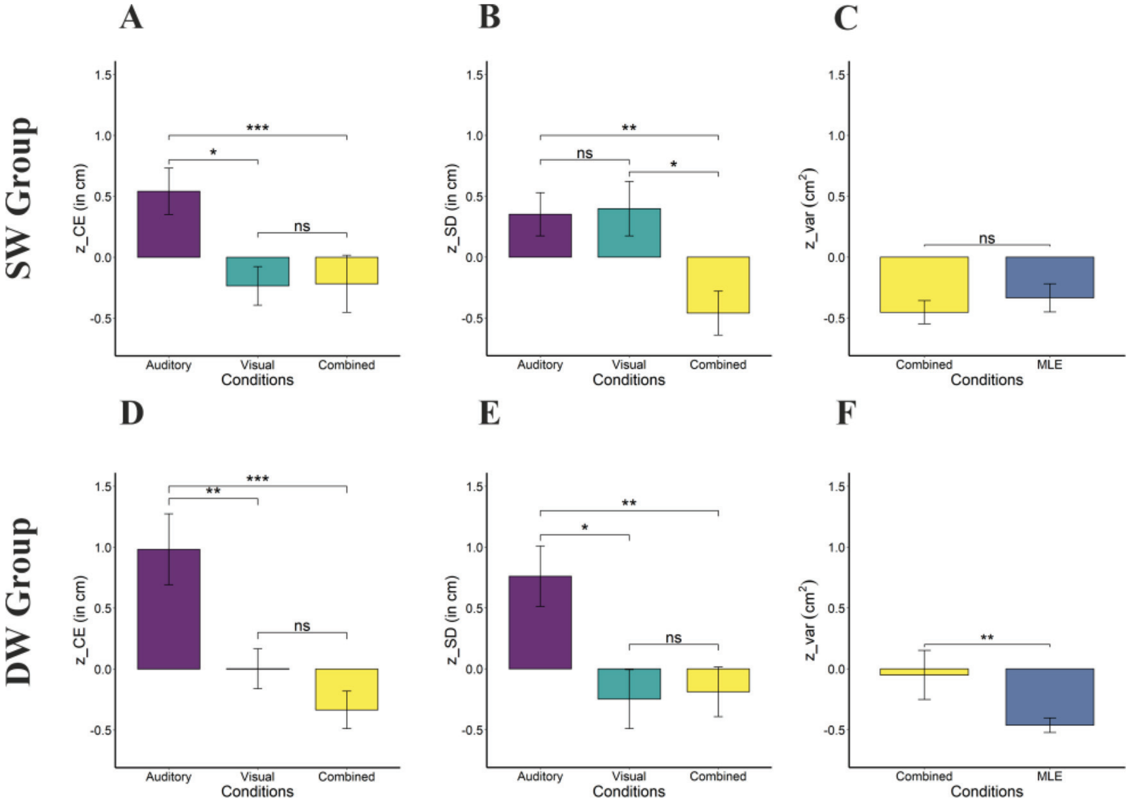
Since the setups of the two experiments were different, we observed different response magnitudes in them. To allow for comparison among responses of participants belonging to the two experiments' samples, we transformed the response data in the *z* scores. Then, we performed the same analyses that we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

For the SW group, the analysis on the standardized CE revealed a significant main effect of condition, $F(2, 20) = 8.68$, $p < .001$, $\eta^2_G = .14$, 95% CI [0, .33]. As shown in Figure 4A, the CE in the auditory condition was greater than those for the visual, $t(20) = 6.25$, p adj $< .0001$, Cohen's $d = .94$, [.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 on the accuracy measure again suggested a predominance of visual cues.

Concerning the variability (Figure 4B), the repeated-measures ANOVA on the standardized SD showed a significant main effect of the factor condition, $F(2, 20) = 7.543$, $p = .005$, $\eta^2_G = .17$, 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 $d = -.05$, [−.43, .33], while both the auditory, $t(20) = 3.49$, p adj $= .007$, Cohen's $d = .99$, [.29, 1.69], and the visual conditions, $t(20) = 2.73$, 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 normality tests failing; in turn, this revealed that no difference occurred ($V = 88$, $p = .355$, $r = .21$, [−.23, .60], $BF_{01} = 3.16$; Figure 4C). What should also be noted is the increase of the BF_{01} value if compared with the same comparisons done in both Experiment 1 (BF_{01} =

Figure 4

Standardized Constant Error (CE), Standard Deviation (SD), and Variance (var) Results on SW Group (Panels A, B, C) and DW Group (Panels D, E, F)



Note. 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 SD for the two unimodal conditions, which were significantly higher than SD in the combined condition (Panel B). Moreover, combined variance did not differ from maximum likelihood estimation (MLE) prediction (Panel C). Conversely, in the dissimilar weights (DW) group, the SD in the combined condition was not significantly lower than the best unimodal one (visual; Panel E), and the variance was significantly higher than the MLE prediction (Panel F). Error bars are standard errors. ns = not significant. * $p < .05$. ** $p < .01$. *** $p < .001$. See the online article for the color version of this figure.

2.45) and Experiment 2 ($BF_{01} = 1.78$), suggesting stronger evidence of integration in a Bayesian way. The analyses of the variability measures suggested that for the SW group, a 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, $F(2, 19) = 16$, $p < .001$, $\eta^2_G = .27$, 95% CI [.04, .46]; specifically, as shown in Figure 4D, the CE was greater in the auditory condition than in both the visual, $t(19) = 3.39$, p adj = .009, Cohen's $d = .90$, [.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 4E), 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^2_G = .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]. 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 revealed that there was a significant difference between the two ($V = 173$, p = .009, $r = -.568$, [−.82, −.20], $BF_{01} = .45$; Figure 4F). For the DW group, the overall results suggested that the participants followed only the visual cues even when the auditory cues were available. This prevented them from reaching the optimal variability reduction predicted by MLE.

General Discussion

In this study, we investigated multisensory integration through a landmark-based spatial navigation task. Performance was defined by the ability of participants to relocate an object based on visual-only, auditory-only, or audiovisual cues. The goals of the present two experiments were to test whether the presence of audiovisual landmarks could enhance performance during navigation and whether unisensory cues could be integrated in an optimal manner.

In Experiment 1, we observed that participants' accuracy in relocating the target object was similar across the visual-only and the audiovisual conditions and that the relocation precision using audiovisual cues was comparable to the visual-only performance. Nonetheless, the participants' precision with both visual and auditory cues was statistically similar to the precision predicted by the MLE model. In Experiment 2, we lowered the reliability of the visual cues to avoid a visual dominance over the auditory cues. At first, in this version of the experiment, we observed a pattern of results that was comparable to Experiment 1. Similar to a previous study about the integration of audiovisual cues (Bentvelzen et al., 2009), we aimed to unveil the potential influences on performance of the interindividual differences in single-cue weightings. Thus, we further explored the results by merging the results of both Experiments 1 and 2. In turn, we divided the participants into two subgroups, according to whether their predicted weights of the two unimodal cues were similar or highly discrepant (SW vs. DW). Consistent with our hypotheses, the participants in the SW group optimally integrated the auditory and visual cues, which resulted in them having increased precision in the audiovisual condition relative to both unisensory conditions. Conversely, the participants in the DW group did not benefit from the audiovisual cues; they accomplished their best precision with the visual-only cues. In the following text, the results are discussed with consideration of the across-individual heterogeneity in the perception of environmental cues.

Weak Evidence of Multisensory Integration in Experiments 1 and 2

Although we found partial evidence that supports multisensory cue integration in Experiments 1 and 2, single-cue dominance and idiosyncratic differences may be stronger factors that underlie people's ability to integrate multisensory landmarks while they are moving. For both Experiments 1 and 2, the precision in the audiovisual condition was statistically similar to the prediction of the MLE model, which is computed from the variance of the responses in unimodal conditions (Equation 5). In line with Bayesian integration principles, the mean of the measured weights associated with each sensory cue (computed from Equation 1) during conflicting trials was statistically equal to the mean of the weights predicted from the inverse of the unisensory variances (Equation 4). This result implies that the participants globally weighted the audiovisual cues according to their reliability, as has occurred in previous studies (Bates & Wolbers, 2014; Cheng et al., 2007; Nardini et al., 2008). Conversely, we observed no increment in precision in the audiovisual condition compared with the visual-only condition. These contrasting and ambiguous patterns of behavioral and modeling results keep us from firmly concluding about the presence of multisensory cue integration. These results usually occur in situations of unimodal dominance (Rohde et al., 2016); however, we found no significant difference in the measured precision between the two unimodal conditions, which reflects that the average levels of precision based on the auditory and visual cues were similar. Comparing the results of Experiment 1 and Experiment 2 shows that there were inconsistent findings for the unisensory weights. In Experiment 1, the average weights associated with the auditory cues were significantly lower than those for the visual cues. In turn, in Experiment 2, the average weights of the two cues were not statistically different. Such an inconsistency confirms that moving the landmarks farther away in Experiment 2 effectively allowed us to achieve our expected goal to lower the reliability of the visual landmarks, which made unisensory precision comparable across conditions. Nonetheless, in Experiment 2, where the two cues were weighted similarly on average, pooling the data of all participants and neglecting the interindividual differences prevented us from finding the predicted increase in precision of multisensory integration compared to the visual-only performance. Previous research by Bentvelzen and colleagues (2009) on the integration of audiovisual cues in speed perception has divided the participants between those who integrate multisensory information and those who do not. These authors suggest that the MLE integration mechanisms may need comparable component weights in order to take place. At the same time, the strong disparity between the unisensory weights prevented them from adopting multisensory integration strategies (Bentvelzen et al., 2009). Altogether, these reasons led us to thoroughly examine the patterns of the results in the data sets of both experiments, considering the interindividual differences in perception as being crucial for explaining such contrasting findings.

Individual Differences in Multisensory Precision

Spatial navigation ability benefits precision by combining audiovisual cues when they are weighted similarly. Although we tested a homogenous group of adult participants, we identified some subgroups that behaved differently based on their ability to integrate multisensory cues. Up to now, much of the research has focused on the development and the evolution of the multisensory integration across childhood (Cuturi & Gori, 2019; Gori et al., 2008, 2012; Nardini et al., 2008) and in the elderly population (Bates & Wolbers, 2014; de Dieuleveult et al., 2017). Idiosyncratic differences that occur regardless of age have not yet been fully explored. 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). Individual differences in sensory weights and precision were previously found, occurring in the context of visuo-vestibular self-motion perception (Fetsch et al., 2009), the perceptual basis to accomplish efficient spatial navigation. Indeed, many factors can influence the interactions of cues, such as distorted feedback and prior experiences (Zhao & Warren, 2018). Previous literature showed that exposure to an unstable visual environment could reduce people's reliance on visual cues (Chen et al., 2017; Zhao & Warren, 2015a). The data sets of Experiments 1 and 2 led us to identify two distinct subgroups of participants: the SW group that behaved as integrators and the DW group that behaved as nonintegrators. In the SW group with similar weights for each sensory cue, the results showed that precision in spatial navigation was better for the audiovisual 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 with highly disparate sensory weights, the multisensory precision was equal to the visual-only condition, and it did not show the substantial improvement that had been predicted by the MLE model. Specifically, in the DW group, when the audiovisual landmarks were available in the environment, the large difference between the auditory (lower) and the visual (higher) perceived reliability likely reduced the feasibility of being able to exploit the multimodal cues optimally; thus, the participants relied only on the best sensory information available to them. Our findings are in line with Bentvelzen and colleagues' (2009) results, which showed that only the group with comparable weights for auditory and visual information achieved the optimal audiovisual cue integration.

Such interindividual differences may be the expression of previous idiosyncratic experiences of our participants in navigation. Along these same lines, different spatial navigation strategies might underlie the differences in participants' abilities to take advantage of the multisensory landmarks. People's navigation ability is influenced by their heterogeneous perceptual and cognitive processing; for instance, navigators differ according to how they can combine self-motion and environmental cues and how they use landmarks or other geometrical information in the environment (Wolbers & Hegarty, 2010). In the present experiments, the participants were asked to perform an inferential navigation task using only landmarks to orient themselves (experiencing disorientation before each relocation response). As such, self-motion cues were never available to them except for the online execution of the movement. One can speculate that the differences found among participants may be related to their natural navigation preferences; specifically, it is possible 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. Because it is already known that internal cues are optimally integrated with external visual points of reference in the environment (Bates & Wolbers, 2014; Chen et al., 2017; Nardini et al., 2008), future research may investigate how multisensory external cues may be combined with self-motion information, taking into account participants' individual strategies of navigation. Indeed, this research will extend our knowledge on the features of individual spatial abilities in a more ecologically valid environment. Our findings emphasize the importance of considering individual perceptual characteristics and experiences when investigating multisensory integration.

Another potential source of individual differences in our experiments might lie in the differences in perception and use of visual cues. Before entering the room, participants wore sunglasses covered by a nylon filter to increase opacity. Although each participant had normal or corrected-to-normal vision and experienced the same level of visual opacity, it is possible that they were affected in varying degrees in their use of visual information depending on their idiosyncratic sensory preferences. Indeed, it is plausible that participants that rely more on visual cues in daily life could have a greater disadvantage in orienting with opaque visual landmarks. Further studies may focus on these differences in the use of visual information during spatial navigation tasks.

Sensory Weighting in Conflict Condition

Testing a condition using a spatial conflict between the auditory and visual cues suggested that participants integrate—rather than alternate—between cues. Following the modeling analyses of Nardini and colleagues (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 substantially better than the prediction made using the alternation model. This suggests that participants did not switch between auditory and visual cues, which is in line with previous outcomes on adult navigation performances (Bates & Wolbers, 2014; Nardini et al., 2008). Intriguingly, in Experiment 1, the observed precision was significantly better than the integration model prediction. This reveals that participants were more precise in the conflict condition than predicted based on their empirical unimodal weights. In the integration model, weights are computed with consideration of the relative proximity (Equation 1); this consists of the predicted distance of responses relative to the visual and conflicting auditory landmarks, namely reflecting that this is 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 other words, in line with previous literature (Zhao & Warren, 2015b), accuracy and precision may have followed different principles of cue interaction, revealing to be independent measures of behavior. Supporting this view, in our experiments, the participants achieved high precision regardless of the accuracy of their responses, meaning that they were able to integrate the audiovisual cues without being influenced by the spatial conflict between the cues. Even if no difference was found between the empirical and predicted weights, as had occurred in previous studies (Bates & Wolbers, 2014; Chen et al., 2017; Nardini et al., 2008), 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 4).

Sensory Cue Domination in Accuracy

Overall, accuracy of the outcomes suggests that there is a visual dominance in spatial navigation rather than multisensory-related benefits when audiovisual landmarks are present. Indeed, accuracy was equal in conditions with and without auditory cues; this indicates that the processing of the visual-only information was ample for participants to achieve the best accuracy performance. This finding is consistent with previous literature that stipulated that navigation accuracy is determined by one cue at a time (in our case, the visual landmarks) and that combining multiple cues selectively tends to increase precision (Zhao & Warren, 2015b). Although cue combination can lead to integrating or alternating between cues instead of leading to reduced accuracy (Petrini et al., 2016), our results point toward the dominance of one sensory modality (visual) over the other in all participants belonging to both the SW and DW groups. Likely, to try to achieve high accuracy, all the participants disregarded the auditory information when both the visual and auditory landmarks were available to them. Thus, the presence of both cues did not either improve or worsen participants' relocation accuracy. During a homing task in a virtual environment, Zhao and Warren (2015b) found that visual cues dominated self-motion cues when determining the homing direction; however, at the same time, visual and self-motion cues were integrated to increase the precision. Thus, the authors argued that there was a coexistence of the cue competition and cue integration models to explain their accuracy and precision results. According to the interpretation of these researchers, this finding is consistent with the hypothesis that landmarks reset the orientation of the path integration system during navigation, but they do not lead to precision in the performance. In other words, the presence of landmarks corrects for the orientation error that participants accumulate while moving based on self-motion information. In the present study, we excluded the role of the path integration system by performing a disorientation procedure before the response phase in each trial. However, self-motion information was available in the response phase; thus, the path integration system operating in the background (Cheng et al., 2007) could have accumulated orientation errors while the participants relocated the target object. 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. Despite this prediction, we observed a visual dominance in participants' spatial navigation accuracy. This finding may extend Zhao and Warren's (2015b) work by specifically suggesting that *visual landmarks* might reset the orientation error of the path integration system that affects the accuracy; in contrast, *auditory landmarks* might not have the same role. If this is correct, it means that only 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 provide enough reliable information to do so. Future studies could directly test whether auditory landmarks compete with path integration for determining orientation by investigating the combination of self-motion information and progressively more conflicting auditory cues.

Visual Versus Auditory Relocation Accuracy

In line with the previous literature (Alais & Burr, 2004; Gori et al., 2012), our results show that participants are more accurate in performing spatial tasks based on visual rather than auditory sensory information. 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 & Gori, 2012; Gori, 2015), the more accurate sense within a specific task calibrates the less accurate sense during child development. In spatial tasks, vision is the most accurate sense for spatial perception because it specifies location more reliably than audition is able to (Gori, 2015). Studies investigating a congenitally blind population showed that the participants had impaired abilities in achieving auditory spatial tasks (Gori et al., 2014; Vercillo & Gori, 2016); thus, this indicates the crucial role of vision in processing spatial information, especially in the early years of life. In the present study, the DW group behaved as if vision was still the dominating sense; in turn, this prevented the participants in this group from benefiting from the redundancy of the multisensory environment. Our findings revealed that the more accurate sense dominates the other sense for the members of the group who showed a sensory-specific higher performance in the context of performing complex navigation tasks. In other words, the participants who were already highly precise in the visual-only condition did not improve their performance when the visual and auditory landmarks were presented together. To make the auditory sources of information comparable to the visual ones, participants with this advanced level of precision would probably require an increased degradation of visual stimuli (e.g., moving the landmarks even farther away). Alternatively, an increase of reliability of the auditory cues could accomplish the same thing. One potential explanation for the DW group's behavior is the individual differences mentioned above. It is likely that participants in this group experienced spatial navigation by relying mostly on visual information; this tendency could lead to an increase of their expertise in regard to this typology of environment exploration, but 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. Future investigation could explore whether this kind of population is identifiable even during spatial tasks that do not require navigation. Moreover, it could be beneficial to study the different developmental trajectories of the integrators and nonintegrators from childhood into old age.

Regarding the use of spatial auditory information, sound properties of the auditory landmarks may have a role in the ability to spatially orient through space. In our experiments, we provided steady auditory landmarks, as done in previous studies (Jetzschke et al., 2017; Nardi et al., 2020; Viaud-Delmon & Warusfel, 2014), because we intended to provide constantly active sound stimuli that may be more likely to be detected and used to spatially navigate (see "Apparatus and Stimuli" section for details). Still, in a real-world environment, auditory landmarks can be either continuous (e.g., a river flowing) or irregular (e.g., a ball bouncing), likely conveying different levels of spatial information. Future studies may directly compare continuous and intermittent sounds to specifically investigate how such properties of auditory cues influence spatial navigation and multisensory integration.

Conclusions

To conclude, in this research, we showed that landmark-based spatial navigation benefits in precision due to the combination of audiovisual cues, which are optimally integrated when they are weighted similarly. Our data suggest that individual differences in cue weighting are responsible for the integration of multisensory landmarks during navigation. This finding confirmed that it is necessary to consider the natural variability across subjects, especially in terms of the heterogeneity of unimodal precision (Rohde et al., 2016). In addition, this research extended previous findings, which stated that it was possible for people to use auditory sources of information to orient themselves efficiently through space (Jetzschke et al., 2017; Nardi et al., 2020; Viaud-Delmon & Warusfel, 2014). This last evidence is particularly relevant to studying navigation and orientation in more ecological and real-life environments and exploring landmark-based navigation in visually impaired people. Since research on inferential navigation in early blind people showed inconsistent results (Cuturi et al., 2016), investigating the integration of self-motion egocentric information and external auditory landmarks could shed new light on the representation of space in this population. Overall, these findings provide the basis for developing tailored novel rehabilitation tools and procedures to improve people's orientation and mobility based on multiple sensory modalities, taking into consideration people's idiosyncratic perceptual traits.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Anderson, S. F., & Kelley, K. (2020). BUCSS: Bias and uncertainty corrected sample size (R package version 1.2.1) [Computer software]. https://CRAN.R-project.org/package=BUCSS

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379–384. https://doi.org/10.3758/BF03192707

Bates, S. L., & Wolbers, T. (2014). How cognitive aging affects multisensory integration of navigational cues. *Neurobiology of Aging*, *35*(12), 2761–2769. https://doi.org/10.1016/j.neurobiolaging.2014.04.003

Bentvelzen, A., Leung, J., & Alais, D. (2009). Discriminating audiovisual speed: Optimal integration of speed defaults to probability summation when component reliabilities diverge. *Perception*, *38*(7), 966–987. https://doi.org/10.1068/p6261

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, *198*(1), 49–57. https://doi.org/10.1007/s00221-009-1933-z

Burr, D., & Gori, M. (2012). Multisensory integration develops late in humans. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 345–363). Taylor & Francis Group.

Butler, J. S., Smith, S. T., Campos, J. L., & Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision*, *10*(11), Article 23. https://doi.org/10.1167/10.11.23

Chen, X., McNamara, T. P., Kelly, J. W., & Wolbers, T. (2017). Cue combination in human spatial navigation. *Cognitive Psychology*, *95*, 105–144. https://doi.org/10.1016/j.cogpsych.2017.04.003

Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychological Bulletin*, *133*(4), 625–637. https://doi.org/10.1037/0033-2909.133.4.625

Cuturi, L. F., Aggius-Vella, E., Campus, C., Parmiggiani, A., & Gori, M. (2016). From science to technology: Orientation and mobility in blind children and adults. *Neuroscience and Biobehavioral Reviews*, *71*, 240–251. https://doi.org/10.1016/j.neubiorev.2016.08.019

Cuturi, L. F., & Gori, M. (2019). Biases in the visual and haptic subjective vertical reveal the role of proprioceptive/vestibular priors in child development. *Frontiers in Neurology*, *9*, Article 1151. https://doi.org/10.3389/fneur.2018.01151

de Dieuleveult, A. L., Siemonsma, P. C., van Erp, J. B. F., & Brouwer, A. M. (2017). Effects of aging in multisensory integration: A systematic review. *Frontiers in Aging Neuroscience*, *9*, Article 80. https://doi.org/10.3389/fnagi.2017.00080

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *The Journal of Neuroscience*, *29*(49), 15601–15612. https://doi.org/10.1523/JNEUROSCI.2574-09.2009

Gibbons, R. D., Hedeker, D. R., & Davis, J. M. (1993). Estimation of effect size from a series of experiments involving paired comparisons. *Journal of Educational Statistics*, *18*(3), 271–279. https://doi.org/10.3102/10769986018003271

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. https://doi.org/10.1163/22134808-00002478

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, *18*(9), 694–698. https://doi.org/10.1016/j.cub.2008.04.036

Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time. *Frontiers in Integrative Neuroscience*, *6*, Article 77. https://doi.org/10.3389/fnint.2012.00077

Gori, M., Sandini, G., Martinoli, C., & Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain*, *137*(1), 288–293. https://doi.org/10.1093/brain/awt311

Gröhn, M., Lokki, T., & Takala, T. (2005). Comparison of auditory, visual, and audiovisual navigation in a 3D space. *ACM Transactions on Applied Perception*, *2*(4), 564–570. https://doi.org/10.1145/1101530.1101558

Jetzschke, S., Ernst, M. O., Froehlich, J., & Boeddeker, N. (2017). Finding home: Landmark ambiguity in human navigation. *Frontiers in Behavioral Neuroscience*, *11*, Article 132. https://doi.org/10.3389/fnbeh.2017.00132

Karimpur, H., & Hamburger, K. (2016). Multimodal integration of spatial information: The influence of object-related factors and self-reported strategies. *Frontiers in Psychology*, *7*, Article 1443. https://doi.org/10.3389/fpsyg.2016.01443

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs (R package version 0.9.12-4.2) [Computer software]. https://CRAN.R-project.org/package=BayesFactor

Nardi, D., Carpenter, S. E., Johnson, S. R., Gilliland, G. A., Melo, V. L., Pugliese, R., Coppola, V. J., & Kelly, D. M. (2020). Spatial reorientation with a geometric array of auditory cues. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*. Advance online publication. https://doi.org/10.1177/1747021820913295

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, *18*(9), 689–693. https://doi.org/10.1016/j.cub.2008.04.021

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, *8*(4), 434–447. https://doi.org/10.1037/1082-989X.8.4.434

Petrini, K., Caradonna, A., Foster, C., Burgess, N., & Nardini, M. (2016). How vision and self-motion combine or compete during path reproduction changes with age. *Scientific Reports*, *6*, Article 29163. https://doi.org/10.1038/srep29163

Rohde, M., van Dam, L. C. J., & Ernst, M. (2016). Statistically optimal multisensory cue integration: A practical tutorial. *Multisensory Research*, *29*(4-5), 279–317. https://doi.org/10.1163/22134808-00002510

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Sjolund, L. A. (2014). *Cue integration and competition during navigation* [Master's thesis, Iowa State University]. Iowa State University Digital Repository. https://doi.org/10.31274/etd-180810-952

Sjolund, L. (2016). *Bayesian integration of spatial navigation cues* [Doctoral dissertation, Iowa State University]. Iowa State University Digital Repository.

Sjolund, L. A., Kelly, J. W., & McNamara, T. P. (2018). Optimal combination of environmental cues and path integration during navigation. *Memory & Cognition*, *46*(1), 89–99. https://doi.org/10.3758/s13421-017-0747-7

Torchiano, M. (2016). effsize: Efficient effect size computation (R package) [Computer software]. R Foundation for Statistical Computing.

Vercillo, T., & Gori, M. (2016). Blind individuals represent the auditory space in an egocentric rather than allocentric reference frame. *Electronic Imaging*, *2016*, 1–5(5). https://doi.org/10.2352/ISSN.2470-1173.2016.16.HVEI-096

Viaud-Delmon, I., & Warusfel, O. (2014). From ear to body: The auditory-motor loop in spatial cognition. *Frontiers in Neuroscience*, *8*, Article 283. https://doi.org/10.3389/fnins.2014.00283

Werkhoven, P., Van Erp, J. B. F., & Philippi, T. G. (2014). Navigating virtual mazes: The benefits of audiovisual landmarks. *Displays*, *35*(3), 110–117. https://doi.org/10.1016/j.displa.2014.04.001

Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences*, *14*(3), 138–146. https://doi.org/10.1016/j.tics.2010.01.001

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association*, *310*(20), 2191–2194.

Zhao, M., & Warren, W. H. (2015a). Environmental stability modulates the role of path integration in human navigation. *Cognition*, *142*, 96–109. https://doi.org/10.1016/j.cognition.2015.05.008

Zhao, M., & 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. https://doi.org/10.1177/0956797615574952

Zhao, M., & Warren, W. H. (2018). Non-optimal perceptual decision in human navigation. *Behavioral and Brain Sciences*, *41*, Article e250. https://doi.org/10.1017/S0140525X18001498

Received June 7, 2021
Revision received September 15, 2021
Accepted September 30, 2021 ■