Flavien Lebrun, Sinan Haliyo, Malika Auvray, David Gueorguiev, Gilles Bailly

To cite this version:

HAL Id: hal-04451467
https://hal.science/hal-04451467
Submitted on 12 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Flavién Lebrun, Université d’Evry Paris Saclay, IBISC, France
Sinan Haliyo, Sorbonne University, CNRS, ISIR, France
Malika Auvray, Sorbonne University, CNRS, ISIR, France
David Gueorguiev, Sorbonne University, CNRS, ISIR, France
Gilles Bailly, Sorbonne University, CNRS, ISIR, France

Visuo-haptic illusions enhance the user experience in virtual reality when physical resources are limited. In order to tailor these illusions to individual user characteristics, we argue in this paper that quantifying users’ visual and proprioceptive sensitivities is essential. We thus present a method - comprising an experimental protocol and Bayesian analysis - for capturing human visual and proprioceptive sensitivity in virtual reality. Our findings demonstrate that 1) in accordance with existing literature, the visual modality is more precise in the azimuthal direction, and proprioceptive precision is higher in depth; 2) the visual and proprioceptive precisions achieved in VR are of a similar magnitude, in contrast to real-world literature results; 3) it is possible to categorize participants based on their proprioceptive and visual precisions. These results will serve as a foundation for studying the impact of human sensitivities on the perception of visuo-haptic illusions.

CCS Concepts: • Human-centered computing → HCI theory, concepts and models; Pointing; Visualization techniques.

Additional Key Words and Phrases: Interaction technique, Models

Mots Clés et Phrases Supplémentaires: Techniques d’interaction, Modèles

Reference:

1 INTRODUCTION

Visuo-haptic illusions rely on a mismatch between the visual and haptic information coming from the same physical event. To solve this mismatch, the brain often favors perception coming from the visual modality [7, 46]. Thus, the visual information can “manipulate” the multi-modal perception of physical events and compensate for the lack of
realism of the haptic information. Visuo-haptic illusions are used in virtual reality (VR) to reduce the limitations of mechanical haptic interfaces by expanding their perceived workspace or resolution [1, 22]. They allow redirecting a user’s movements towards selected real objects, enabling a real object to serve as physical counterpart for several virtual objects [3]. They also provide physicality to virtual objects without or with limited haptic stimulations [38, 63].

A key parameter of visuo-haptic illusion is the amplitude, which corresponds to the gap between the visual and haptic information. When this amplitude is too important for the brain to solve the sensory mismatch, the users "detect" the illusion, which leads to a decrease in presence in VR. In practice, VR designers need to know when users are likely to detect illusions to maintain a high level of immersion. Importantly, the ability to detect illusions vary among individuals [37]. Our long term goal is to develop adaptive visuo-haptic illusion techniques, i.e., techniques that adapt the illusion to the perceptual abilities of each individual. Such adaptive techniques raise to challenges: (1) estimating the perceptual abilities of each individual and (2) elaborating a computational model.

In this article, we address the first challenge: how to measure the sensitivity of a given modality in virtual reality? More precisely, we aim to address the three following research sub-questions:

- **RQ1:** Is there an appropriate methodology to measure visual and proprioceptive sensitivities?
- **RQ2:** How do the sensory sensitivities of participants in Virtual Reality compare to the ones in the real world?
- **RQ3:** Is it possible to differentiate users based on their visual and proprioceptive sensitivities?

To address these research sub-questions we rely on experimental protocols and analysis methods derived from cognitive psychology [59]. These studies were conducted in a real environment, and we hypothesize that the precision of certain modalities may differ in VR (RQ2).

We first provide a brief overview of visuo-haptic illusion techniques and discuss how they correlate with the precision of the user’s sensory modalities. We then introduce a methodology to experimentally estimate these precisions for both a population and an individual. This methodology is applied in an evaluation campaign, which is analyzed using a conventional approach from the literature, as well as a Bayesian analysis technique. This analysis demonstrates, on one hand, that the precisions of certain modalities differ between VR and real environments. It also shows that Bayesian analysis allows for the calculation of individual parameters, which can be leveraged for personalizing an illusion.

## 2 RELATED WORK

In this section, we describe visuo-haptic illusions in VR and their applications.

### 2.1 Visuo-Haptic Illusions in Virtual Reality

Several interaction techniques based on visual illusions have been proposed in virtual reality. The common feature among these illusions is the manipulation of the visual rendering of the virtual environment. Although some of these illusions are based on a shift between visual and vestibular [48] or auditory information [50], we primarily focus on visuo-haptic illusions \(^1\). Indeed, this family of illusions enhances haptic interactions in VR. Almost all these illusions involve the manipulation of the visual rendering of the virtual hand of the user or of a tool manipulated by the user.

The manipulation of the visual rendering can involve the color or shape of objects or the user’s avatar’s skin to influence tactile perceptions of object characteristics, such as their stiffness [2, 4, 47]. However most of visuo-haptic illusions deceive proprioception by introducing a difference between the movements of the real body and those of the

---

\(^1\)These illusions are often collectively referred to as “pseudo-haptic” [38, 63], although a family of illusions called “redirection” [3, 35] is often dissociated from this classification.
corresponding virtual avatar. This can alter the perception of an object’s weight [13, 28, 53], the number of physical objects in the scene [3, 8, 34], or the stiffness of a spring [39]. For instance, the technique of “Hand Redirection” [3, 8, 33] introduces a gradual offset between a user’s real hand and their virtual avatar during a movement. This offset is continuously corrected, resulting in a change in the real hand’s trajectory. In extreme cases, the user’s hand can be redirected toward a single physical object when moving towards different virtual objects [3, 8, 34]. In this article we focus on the study of visuo-haptic illusion that involve a shift between the real hand (proprioceptive perception) and the virtual hand (visual perception).

2.2 Detection of Illusion

The brain rejects an illusion if it notices that something is wrong. Several studies aimed to estimate the maximum amplitude beyond which participants (population level) notice it or judge it uncomfortable [1, 7, 8, 21, 23, 30, 44, 51, 69]. However, these detection thresholds are low and can be increased by customizing the implementation of illusions to each participant’s characteristics (individual level). Lebrun et al. [37] show that it is possible, after a short calibration phase, to adjust the amplitude of hand redirection for each participant and thus increase the detection threshold in comparison with the one obtained at the population level. Groth et al. [24] also find that there exists a difference in gender for the detection of hand redirection illusions. They demonstrate that for a similar amplitude the detection probabilities of women are significantly lower. Yet, to our knowledge, no study has attempted to explain the differences between participants using computational models. We justify the importance of users’ sensitivities for the implementation of such models in the following section.

3 HYPOTHESIS AND THEORETICAL FOUNDATION

We make the hypothesis that the ability of the users to estimate the positions of the physical and virtual hand has an impact on the likelihood of detecting the Hand Redirection illusion. We describe here two mechanisms that can lead to the detection of illusion. The first is linked to a decrease in users’ sense of agency i.e. a decrease in the experience of controlling one’s own actions and, through them, changes in the external environment [25]. The second is linked to a decrease in the sense of body ownership (SoO). Body Ownership refers to one’s self-attribution of a body [20, 61]. In this section, we demonstrate the importance of estimating individual sensitivities – particularly visual and proprioceptive – to predict the break of SoA and SoO and thus predict the detection of an illusion. To achieve this, we describe some computational models linked to the decrease in SoA and SoO.

3.0.1 Models about the Sense of Agency. The SoA is a component of the sense of presence [31]. A decrease of SoA can lead to users noticing that something is off. The comparator model [18] can explain a decrease of SoA (see 10 for a detailed explanation of the comparator model). According to this model, the brain creates an efferent copy whenever a new motor command is generated. If the efferent copy matches the actual sensory perception, the movement is perceived as self-caused, and SoA arises [10]. On the opposite, if the sensory perception and the efferent copy are too different, it could lead to a semantic break [45] and a decrease in SoA. Thus, the actual sensory perception plays an important role. In our context, the brain perception of the hand comes from an integration of discrepant information from the visual and proprioceptive modalities. To evaluate the value of this perceived position, one needs to understand how these two information are weighted to come up with a unified perception.

Computational models related to the optimal multi-sensory integration can help us predict the estimation of the hand position from the integration of visual and proprioceptive information. For instance, according to the Gaussian
Estimated Position

Fig. 1. Illustration of the model of optimal multi-sensory integration. On the left, the brain combines the visual information ($x_v$) and the proprioceptive information ($x_p$) in an optimal manner. The estimated position $\hat{x}_{pv}$ is closer to $x_v$ than $x_p$ because the confidence in the visual information (small variance $\sigma_v$) is better than the proprioceptive one (large variance $\sigma_p$). The variance $\sigma_{pv}$ is smaller than either $\sigma_v$ or $\sigma_p$. On the right we illustrate a similar optimal integration for 2D positioning instead of 1D. Here the confidence in a positional information is represented by confidence ellipse. These ellipses have 3 parameters, the $x$ and $y$ variance and the covariance, that describe the relative elongation and orientation of the ellipse.

The estimation of a quantity by a combination of different modalities (e.g. vision-proprioception) depends on the confidence that the brain has in each modality. This confidence is characterized by the sensitivity of the corresponding modality. Consider the estimation of the $x$ position of an object (1D estimation) by the visual and proprioceptive modalities (Figure 1-Left). The sensitivity of these modalities can be mathematically represented by their corresponding variances: $\sigma_v^2$ for the vision and $\sigma_p^2$ for the proprioception. The final position estimation of the object by the brain $\hat{x}_{pv}$ is thus a weighted combination of the estimation made by the visual $x_v$ and proprioceptive $x_p$ modalities:

$$\hat{x}_{pv} = \frac{\sigma_p^2}{\sigma_v^2 + \sigma_p^2} x_v + \frac{\sigma_v^2}{\sigma_v^2 + \sigma_p^2} x_p$$

The final estimation $\hat{x}_{pv}$ is more precise than the most precise modality estimation (see Figure 1). Moreover, the less precise the estimation of a modality is (i.e. large variance), the less influence it has on the final estimation of the quantity. The integration is “optimal” in that sense that it minimizes the error of the estimation of the physical quantity.

The brain seems to follow this integration pattern for physical quantities such as the size of an object through vision and haptics [14], the orientation of the body in space through vision and the vestibular system [16, 70], the shape of an object by vision and touch [26] and the localization of our limbs during a movement [6, 57, 58, 65–67].

3.0.2 Models about the Sense of Ownership. The SoO is also a component of the sense of presence. Like the SoA, a decrease in SoO can lead to the detection of an illusion. Bayesian Causal inference models [36] explain the sense of ownership (SoO) in VR [32]. These models propose that the SoO emerges as the result of attributing all sensory information available about the body to a single common cause: the self-body [52, 54]. In our case the brain decides to attribute the visual ($x_v$) and proprioceptive ($x_p$) position of the hand to a common cause ($C=1$) or to a different cause ($C=2$). If the virtual hand is seen as the cause of both $x_v$ and $x_p$ ($C=1$) there is a SoO toward the virtual hand, else if the virtual hand is only the cause of $x_v$ ($C=2$), no SoO occurs (see Figure 2).
To select the most likely cause the brain infers each cause probability considering the available information: $P(C=1|X_v, X_p)$ and $P(C=2|X_v, X_p)$. The Bayes’s theorem is used to calculate these probabilities:

$$P(C = i|X_v, X_p) = \frac{P(X_v, X_p|C = i) \times P(C = i)}{P(X_v, X_p)}$$  \hspace{1cm} (2)$$

where $i$ equals 1 or 2. In this equation, $P(C = 1)$ is the prior probability of a common cause and $P(C = 2)$ the prior probability of a different cause. These two prior probabilities are fixed by the brain and do not depend on the sensory sensitivity. The term $P(X_v, X_p)$ is constant and can be disregarded when comparing $P(C = 1|X_v, X_p)$ and $P(C = 2|X_v, X_p)$. Finally $P(x_v, x_p|C = 1)$ and $P(x_v, x_p|C = 2)$ are the likelihoods of having the positions of the hand at $x_v$ and $x_p$ given $C=1$ or $C=2$. These likelihoods are calculated based on users’ visual and proprioceptive sensory sensitivities.

Therefore to predict both the breaks in SoA and SoO with computational models, it is necessary to first capture and measure users’ visual and proprioceptive sensory sensitivities.

4 MEASURES OF SENSORY SENSITIVITIES

Problem formulation. The task we consider in this work is a pointing task in VR with the right hand. The specificities of this task are that 1) two modalities are involved (vision and proprioception) and 2) a pointing movement involves multiple joints (wrist, elbow, shoulder and potentially finger) and 3) we consider two quantities (the final $x$ and $y$ positions). Our objective is to estimate the visual sensitivity $S_V$ of the users as well as the proprioceptive sensitivity of their right hand $S_{PrRH}$ for this task. While it is easy to estimate an overall sensitivity $GS_1$ by measuring, for instance, the dispersion of the final hand points, it is difficult to estimate the respective contribution of each modality. From a mathematical point of view, the problem consists of resolving a single equation with two unknowns:

$$GS_1 = f(S_V, S_{PrRH})$$  \hspace{1cm} (3)$$

Both sensitivities $S_V$ and $S_{PRH}$ can be decomposed into precision *i.e. the variance of the distribution*, and accuracy *i.e. the systematic bias*.

**Common experimental protocols.** 2 main experimental protocols have been proposed in the literature to measure the sensory sensitivity of modalities.

One protocol consists of simply asking participants to perform several times the same pointing task (estimating quantities). With several repetitions, the participants form a distribution whose variance and mean can be extracted. The mean inform on the accuracy and the variance on the precision. The global sensitivity of this task is thus the measure of this variance and mean. A variant of this protocol has been extensively used in health and sports to measure proprioceptive sensitivity for multiple joints (e.g. the ankle [12] or the knee [62]). It was possible because the task did not involve vision. In our context, the global proprioceptive sensitivity of the right hand $S_{PRH}$ also depends on the proprioceptive sensitivity of multiple involved joints. While it could be informative to measure the sensitivity of each joint, this approach is time-consuming, tedious and not necessary to study visuo-haptic illusion.

Another protocol [14, 41] is based on a 2AFC experiment. The participants compare two quantities (e.g., the size of an object [14]) and should estimate which one of the two quantities is smaller/larger (forced choice). The more finely a participant can discriminate the two quantities with one modality, the greater the accuracy of this modality is. This 2AFC protocol has been used in different contexts [49], [42], [27].

**Limitations.** These two protocols do not allow to distinguish the visual and proprioceptive sensitivities of each modality for a pointing task. One can consider suppressing one modality when performing the pointing task to measure the sensitivity of the other modality. However, it is extremely challenging. For instance, if the visual modality is suppressed, the stimulus, the target, is not visible. It would require to carry the hand of the participant to defined positions. Beyond the difficulty (time and setup) to implement such study, it remains that the participants would perform passive movements instead of active movements. Similarly, it would also be possible to suppress the proprioceptive modality, but it would require invasive operations (anesthesia). To solve this problem, Van Beers et al. [64] proposed an original method on which we build on to measure users’ sensitivities.

**Van Beers et al. protocol.** The Van Beers et al. protocol [64] aims to assess the precision of human hand proprioceptive localization. Their findings indicate that for proprioception, localization in the radial direction relative to the shoulder exhibits greater precision compared to localization in the azimuthal direction. Regarding visual localization, their observations reveal that localization in the azimuthal direction relative to the cyclopean eye demonstrates higher precision than localization in the radial direction. In their protocol the participants perform three different (pointing) tasks (instead of one). From each task, a general sensitivity is extracted providing from a unique combination of visual sensitivity $S_V$, proprioception sensitivity of the right hand $S_{PRH}$ and proprioception sensitivity of the left hand $S_{PLH}$ (see Figure 3). More precisely:

1. General sensitivity task 1 $GS_1$ is the combination of visual sensitivity $S_V$ and the proprioceptive sensitivity of the right hand $S_{PRH}$: $GS_1 = f(S_V, S_{PRH})$
2. General sensitivity task 2 $GS_2$ is the combination of visual sensitivity $S_V$ and the proprioceptive sensitivity of the left hand $S_{PLH}$: $GS_2 = g(S_V, S_{PLH})$
3. General sensitivity task 3 $GS_3$ is the combination of the proprioceptive sensitivity of the right hand $S_{PRH}$ and the proprioceptive sensitivity of the left hand $S_{PLH}$: $GS_3 = h(S_{PRH}, S_{PLH})$

6
In other words, one equation with two unknowns (which does not have a unique solution) is transformed into a system of three equations with three unknowns which has the advantage to have a unique solution.

Our general approach relies on this protocol. We now detail how we collected data in our context.

5 DATA COLLECTION

5.1 Participants and Apparatus

Twelve participants take part in the study. Their ages range from 25 to 32 years. There are 10 right-handed individuals and 2 left-handed individuals. Four participants wear glasses or contact lenses. None of the participants have any neuro-muscular disorders affecting their proprioception. All participants sign an informed consent form, and the study, in compliance with the Helsinki Declaration, is approved by the ethics committee of Sorbonne University (N° 2020 – CER- 2020- 61).

They wear an HTC Vive head mounted display tracked by an Optitrack system composed of 8 infrared cameras. The position and 3D orientation of the participants’ index finger is also recorded with the Optitrack system, using a glove with markers.

Participants are sitting in front of a transparent Plexiglas table. 7 haptic targets are placed under the table. The locations of these targets are the same than the virtual targets that the participants can see in the HMD (see Figure 4).
An additional haptic target indicates the starting position which is both felt and seen in VR. A pedal under the right foot of the participants is used to validate each trial. In the virtual environment, the participants do not see their hands.

5.2 Task and Conditions

The task consists of positioning the tip of their index finger as accurately as possible on one of the seven possible targets. Depending on the conditions, participants use either their right or left hand, virtual or physical targets.

- In the first condition, participants have to match as best as possible the position of the tip of the index finger of their **right hand** with the position of a visible **visual** target in the virtual environment : Right Hand - Visual Target \( R_H - V \). The visual target is represented with a green sphere of the size of a thumb (see Figure 4)
- In the second condition, participants have to match the position of the tip of the index finger of their **left hand** with the position of a **visual** target in the virtual environment (Left Hand - Visual Target \( L_H - V \))
- In the last condition, the participants have to match as well as possible the position of the tip of the index finger of their **right hand** on the table with the position of the index finger of their left hand placed under the table (Right Hand - Proprioceptive Target \( R_H - P \)). The left hand is positioned on a **physical** target. The physical targets have the same position and relative size as the visual targets, except that the thickness of the table separates them (~5mm).

Each participant performed the 3 conditions corresponding to the three tasks of the method (see Figure 3):
5.3 Procedure

We first explain the principle of the experiment to the participants and have them read and fill out a consent form. Then, they put on the HMD and glove, on their right hand for the $R_H - V$ and $R_H - P$ conditions and on their left hand for the $L_H - V$ condition. They then place their right foot on the pedal for all conditions. When they enter the VE, participants see a screen showing them their next task. They must press the foot pedal to begin the training phase. This phase is performed for each new condition.

In all conditions the participants, once the target has appeared, must place the hand wearing the glove (invisible in the VE) as best as possible on the position of this target. Once they are satisfied with the position of their hand, they must press the pedal. They must then return to the starting position to begin the next trial.

For the $R_H - V$ and $L_H - V$ conditions, the target is visual and appears as soon as the participants have reached the starting position. For the $R_H - P$ condition, the target is proprioceptive, it is the fingertip of the left hand. When participants start a new cycle (of $R_H - P$), a visual target appears. They must then place their left hand, under the table, on the haptic target corresponding to this position. They must then press the foot pedal. This causes the visual target to disappear and numerous visual targets to appear for 2 seconds. They act as a lure to make the participant forget the position of the visual target. The participant thus focuses only on the position of his left hand for the aiming task.

For the $R_H - V$ and $L_H - V$ conditions, participants perform 6 training cycles. For the $L_H - V$ condition, they perform 12 training cycles. Once the training phase is completed, the real experiment begins as soon as the participant is ready and presses the pedal. Participants aim 20 times at each of the 7 targets in the same way as they did in the training phase. The number of attempts already made is displayed on the virtual screen in front of them.

5.4 Design

This study follows a within-subjects design. The presentation order of the three conditions is counter-balanced between participants to avoid learning effects. For each condition, participants perform 20 blocks. In each block (7 trials), each target is presented one time in a random order. In summary the experimental design is: 12 participants x 3 conditions x 20 blocks x 7 targets = 5040 trials.

5.5 Measures

The 3 dependent variables are Precision and Accuracy of the index finger position as well as Time to reach the target.

5.6 Differences with Van Beers et Al. Protocol?

In Van Beers et al. [64] experiment, participants were situated in a real-world setting and seated facing a table. They observed the reflection of visual targets affixed to a cardboard above the tabletop using a horizontally mounted mirror placed midway between the cardboard and the table. This setup created the illusion for the subjects that the targets were positioned on the table. Notably, they employed only three targets, a reduced number compared to our seven targets. These targets were positioned relatively close to the body: one approximately 30cm from the body, slightly on the left, and another approximately 20cm from the body, slightly on the right. The spatial coverage of these targets concerning the user closely resembled that of our targets. Consequently, we contend that, at least as an approximate comparison, it is viable to juxtapose the averages derived from our obtained values with those obtained in their study.
6 ANALYSIS

6.1 Mathematical Formalization of Sensory sensitivity

We assume that the sensory sensitivity can be represented by a confidence ellipse which contains the final positions of the index finger (see Figure 1). We work in this study with 95% confidence ellipses. The distribution of points to form this ellipse can be generated by a 2-D Gaussian function. A 2D Gaussian is defined by its mean $\mu = (\mu_x, \mu_y)$ representing the center of the ellipse and by its covariance matrix $\Sigma$ (see Figure 1) representing its shape:

$$\Sigma = \begin{pmatrix} \sigma^2_x & c_{xy} \\ c_{xy} & \sigma^2_y \end{pmatrix}$$

The covariance matrix has three parameters. The variances $\sigma^2_x$ and $\sigma^2_y$ characterize the elongation of the ellipse in a certain direction. The covariance $c_{xy}$ gives the orientation of the ellipse. For example in Figure 1 the ellipse illustrating the proprioceptive sensitivity (orange) is more elongated in the x direction and therefore has a higher $\sigma^2_x$ than $\sigma^2_y$. It is the opposite for the ellipse representing the visual sensitivity (blue). None of these ellipses are oriented perfectly horizontally, therefore their covariance is different from 0.

The covariance matrix $\Sigma$ represents the precision of the modality. Its accuracy corresponds to the offset between the target position and the center of the ellipse: $bi = (bi_x, bi_y)$. Note that in sensory integration models, the precision (i.e. covariance matrix $\Sigma$) is used to assign a weight to a modality.

6.2 Bimodal Data Analysis

The first step of the analysis is to put the observed data into an usable form. We have the final positions $(p_x, p_y)$ of the left and right hand indexes (FPI). We have 21 (7 targets * 3 conditions) distributions for each participant. For each distribution, we calculate the 5 parameters $(\mu_x, \mu_y, \sigma^2_x, \sigma^2_y, c_{xy})$ of the 95% confidence ellipse. However, these bimodal confidence ellipses correspond to the combination of two sensitivities. We then need to extract each unimodal sensitivity individually. We propose two methods of analysis.

6.2.1 Classical Analysis. The classical method of analysis is the one used by Van Beers et al. [64]. For each condition, the two distributions, expressing the precision of the two modalities involved, are independent. Indeed the two modalities are measuring different physical elements. Thus the bimodal covariance matrix obtained is simply the sum of the two unimodal covariance matrices. Similarly the bias (bi) of the mean of the bimodal distribution is the sum of the biases of the unimodal distributions. For the condition $R_H - V$ for example this is expressed as follows:

$$\Sigma_{R_H-V} = \Sigma_V + \Sigma_{P_{RH}}$$

$$bi_{R_H-V} = bi_V + bi_{P_{RH}}$$

where $\Sigma_{R_H-V}$ and $bi_{R_H-V}$ are the covariance matrix and bias parameters of the observed bimodal distributions with the experiment, and $bi_V, bi_{P_{RH}}, \Sigma_V, \Sigma_{P_{RH}}$, and $bi_{P_{RH}}$ are the bias and covariance matrix of the unimodal distributions representing the sensitivities of the visual ($\Sigma_V$) and of the right hand proprioceptive ($\Sigma_{P_{RH}}$) modalities. We can transpose the Equation 5 and the Equation 6.2.1 to the conditions $L_H - V$ and $R_H - P$. For each target, the 3 distributions obtained from the 3 conditions give a system of 3 equations with 3 unknowns (the unknowns being the 3 couples $(bi_V, \Sigma_V), (bi_{P_{RH}}, \Sigma_{P_{RH}}), (bi_{P_{LH}}, \Sigma_{P_{LH}})$). The Figure 3 illustrates the system of three equations with three unknowns. The resolution of this system provides the 3 couples of parameters describing each unimodal distribution. In summary:
• **Input (for each participant):**
  - 21 bimodal distributions \((bi, \Sigma)\) from 7 targets and 3 conditions

• **Algorithm for each target:**
  - Selection of the 3 bimodal distributions for each target (1 per condition)
  - Construction of the equation system with the 3 bimodal distributions
  - Solving the equation system

• **Output:**
  - 7 unimodal distributions \((bi, \Sigma)\) for each of the 3 modalities (visual, right and left hand proprioception)

This method directly relies on our observations and has the advantage of proposing an exact solution. However, the resolution of the equation system imposes the use of subtractions for the calculation of unimodal distributions. This can lead to parameters with impossible values (due to measurement noise) such as negative variances. This problem is corrected if the results are averaged over all participants. Indeed there are more positives than negatives result which when averaged give a positive result. This method therefore provides a tool for studying modality sensitivities at the population level, but is not appropriate at the individual level.

### 6.2.2 Bayesian Analysis

To study our results at the individual level, we propose to turn the problem around. Instead of starting from the bimodal distribution and try to extract the unimodal distributions, we search the optimal unimodal parameters that will generate the observed (FPI) bimodal distribution. Indeed, with a Bayesian inference approach \([5]\), the objective is not to find the exact value of one parameter (here the covariance matrix \(\Sigma\)) but rather a posterior probability law on this parameter \(\Sigma\). That is to say the probability distribution of the parameter value after taking into account the experimental data. A Bayesian inference approach requires a prior law of the parameters that we want to estimate. The prior law reflects the knowledge about the value of the parameter before observing the data.

To implement the Bayesian inference approach, we use the ABC (Approximate Bayesian Computation) algorithm already used in the HCI community by \([29]\). The principle of ABC is to repeatedly generate a set of unimodal parameter values drawn from prior law. We can then calculate the bimodal gaussian law parameters and generate the bimodal points distribution. Finally, we compare the simulated distribution to the observed distribution (FPI). If the distance between the two distributions is close enough, the parameter values are accepted. We repeat this process \(n\) times. For each parameter, the accumulation of their accepted values define the parameters’ probability law.

We have chosen to calculate the biases by the classical method presented in subsubsection 6.2.1 to limit the number of parameters. In summary:

• **Input for each participant:**
  - 21 bimodal distributions \((bi, \Sigma)\) from 7 targets and 3 conditions
  - 3 prior laws of \(\Sigma\) parameters, one per modality
  - The values of the unimodal biases \((bi)\) calculated with the classical method

• **Algorithm for each target:**

\[
\text{WHILE } n < 10000:\
\begin{align*}
&\text{- step 1: drawing of parameters } \Sigma_V, \Sigma_{PMR}, \text{ and } \Sigma_{PL}\text{ based on the chosen prior law (in total there are } 3*3 = 9 \text{ parameters).} \\
&\text{- step 2: generation of the bimodal distributions } (bi_{R_{H-V}}, \Sigma_{R_{H-V}}), (bi_{L_{H-V}}, \Sigma_{L_{H-V}}) \text{ and } (bi_{R_{P-V}}, \Sigma_{R_{P-V}}) \text{ (see Equation 5 and Equation 6.2.1)}
\end{align*}
\]

11
- step 3: creation of three $PFI_{sim}$ datasets generated following the 3 2D Gaussian laws $(bi_{MD-V}, \Sigma_{MD-V})$, $(bi_{MG-V}, \Sigma_{MG-V})$ and $(bi_{RH-P}, \Sigma_{RH-P})$
- step 4: evaluation of the distance $d(PFI_{obs}, PFI_{sim})$ for the 3 conditions $RH-V$, $LH-V$ and $RH-P$
- step 5: IF $d(PFI_{obs}, PFI) < \epsilon$ the parameters are accepted. ELSE the parameters are deleted

END WHILE

- final step: Draw the 9 parameters’ probability laws from the accumulation of accepted parameter values

- **Output:**
  - 9 probability laws on the covariance matrices $\Sigma$ for the 3 modalities (visual, proprioceptive right hand and proprioceptive left hand).

The average of these probability laws is used to compare this method with the previous classical method, which provides unique values. To apply the ABC algorithm we used the software package ELFI [40]. We use a variation of the classical ABC algorithm called Sequential Monte Carlo ABC and compare the euclidean distance between the variance and the mean of the real and simulated distribution. This implementation was made under python.

6.3 Frame of Reference
The two variances $\sigma_x^2$ and $\sigma_y^2$ are expressed with respect to x and y coordinates. However, it is more interesting to express them in the frame of reference of the participant: depth and azimuth. Each covariance matrix obtained is thus now expressed along these axes.

7 RESULTS
7.1 Bimodal Data
The Figure 5 illustrates the distribution (PFI) averaged over all participants for the three conditions.

7.1.1 Precision. We observe that the confidence ellipses of the $RH-V$ (right hand visual target; green) and $LH-V$ (left hand visual target; orange) conditions are close to a circle and slightly more elongated in the depth direction. In contrast, the confidence ellipses of the $RH-P$ condition (right hand proprioceptive target; purple) are more elongated in the azimuthal direction.

---

Fig. 5. Precision and accuracy for the seven targets and the three conditions. Left: Pointing with the right hand; Center: Pointing with the left hand; Right: Proprioceptive pointing with the right hand.
The Figure 6 compares precision in virtual reality (our data) with the precision in a real environment (data from [64]). The results suggest that the precision in virtual reality are aligned with those in a real environment (see section 5.6) [56, 64] : proprioception is more precise with respect to depth (Figure 6-left) and vision is more precise with respect to azimuthal direction (Figure 6-center).

The Figure 6-right also suggests that users are more precise in a real environment than a virtual one. We observed indeed on Figure 6-left a large depth variance of the conditions involving vision ($R_H - V$ and $L_H - V$) which is not compensated for by the small azimuthal variances (Figure 6-center).

7.1.2 Accuracy. We observe that participants tend to overreach targets in depth especially for the conditions involving vision. There is also a lateral overreach: for the condition visual pointing with the right hand ($R_H - V$), the pointing is shifted to the left, while for the condition pointing with the left hand ($L_H - V$), it is shifted to the right. Finally, for the condition with proprioceptive target ($R_H - P$), we observe a rightward overshoot.

The data on accuracy were analyzed only qualitatively in the original article [64]. Therefore, we cannot make quantitative comparisons. Qualitatively, our results on accuracy in virtual reality mirror those observed in a real environment, particularly for depth overreach [11, 17] and lateral overreach [9, 64] in a task with vision. Similarly, lateral overreach for a proprioceptive task corroborates the results obtained in a real-world environment[55, 64].

7.2 Unimodal Data

The Figure 7 illustrates the unimodal precision and accuracy for the classical analysis method. The Figure 8 does the same for the Bayesian Analysis. Regardless the method of analysis, we observe less precision in the depth direction for the visual modality, and in the azimuthal direction for the proprioceptive modalities. Note that the depth and azimuthal directions depend on the modality, the reference point being the middle of the eyes projected on the plane of interest for the visual modality and the positions of the shoulders projected on the plane of interest for the proprioceptive modalities (left and right). With this convention, the shapes of the ellipses are relatively stable whatever the target, especially for the right proprioceptive modality.

Regardless the method, the areas of the ellipses are close when moving from one modality to the other, this result is different from the literature [64], where the visual modality has a better precision than the proprioceptive modality at least in terms of the area of the confidence ellipses. This result is discussed in section 8. We observe that one of the
ellipses of the right proprioceptive modality is atrophied (see the rightmost ellipse of the central graph of the Figure 7). This is due to the analysis method involving subtraction and resulting in negative or almost zero variances (see section 6.2.1). This impossible result is due to the imprecision of the measurement of bimodal distributions. The Bayesian Analysis avoids these negative variances. The unimodal distributions obtained with the Bayesian Analysis have smaller areas than with the classical method. The shapes and direction of the ellipses are similar in both cases.

7.2.1 Detailed Study of the Precision Parameters. The averaged azimuthal and depth variances and covariances by modality are shown in Figure 9. We can see that the azimuthal variance is more than 2 times larger on average for right proprioception and almost 2 times larger for left proprioception than the depth variance. On the other hand, the visual variance is a little lower in the azimuthal direction than in the depth direction. Above all, we obtain azimuthal and depth variances for the visual modality that are much larger than the results of Van Beers et al. These observations are verified for both methods of analysis.

7.3 Individual Level
At the moment, we have only considered the averaged results across all participants, which has notably enabled us to compare them to previous studies that also presented results at the population level. Here we consider the results at the individual level. We have illustrated on the Figure 10 the mean of the confidence ellipses for each participant for the 3 modalities, and for the two analysis methods. For each participant it is reassuring to find a similar behavior for the two treatment methods. For example, if a participant has larger areas than his neighbors in the classical treatment, it is the same for the Bayesian Analysis. But the areas in the classical treatment are not a simple proportional increase of the areas in the Bayesian Analysis. Indeed, we do not observe the same ratios between the 3 modalities for the same
participant for both methods. For example, participant 2 has more precision with proprioception than with vision in Bayesian analysis, but it’s the opposite with classical analysis. Moreover we observe that the inter- and intra-participant variations are more marked for the classical method with a significant standard deviation for most of the participants. Even with less pronounced inter- and intra-participant differences with the Bayesian analysis, it is still possible to observe interesting patterns in certain participants. For example participant 7 has balanced precision across the three modalities. In contrast, participant 9 has a visual precision almost twice as low as his/her proprioceptive precisions. These results are discussed in section 8. Note that to streamline the article, the analysis of the evolution of precision with distance and orientation was placed in appendix B.

8 DISCUSSION
Our short-term goal was to assess the visual and proprioceptive sensory sensitivities of our participants, focusing on three research questions:
**RQ1: Is there an appropriate methodology to measure visual and proprioceptive sensitivities?** The protocol proposed by Van Beers [64] enables the extraction and differentiation of visual and proprioceptive sensitivities in hand movements. However, this methodology does entail certain limitations, such as the time required for conducting the data collection task, the great number of trials needed to obtain a representative distribution, and the increased complexity in data processing. Particularly, we showed that the traditional data processing method involving subtractions could yield impossible values due to the random nature of the targeting task. Instead, we decided to use a Bayesian analysis method (ABC). We demonstrated that it avoids outliers, such as negative variances or atrophied precisions. Moreover, it reduces uncertainty in our data, allowing for a better interpretation of trends.

**RQ2: How are the sensory sensitivities of participants in Virtual Reality compared to those in the real world?** Our results exhibit several consistencies with the existing literature. Similar to [11, 17], our participants tended to overshoot the targets in depth, and similar to [9, 64], they tended to overshoot laterally. Moreover, proprioception was more precise with respect to depth, and vision is more precise with respect to the azimuthal direction [56, 64]. However, contrary to the literature in the real world, in VR, the visual and proprioceptive accuracy measured are of the same order of magnitude. We see several possible explanations. One is that in condition RH − P, the tactile information coming from the deformation of the experimental table could have helped participants to increase their precision. Another is that participants in condition RH − P remember the position of the visual target. The addition of lures after the disappearance of the visual target may not be sufficient to forget the position of this target. In this case, the distribution (PFI) in the RH − P condition would be smaller than it should be, favoring the proprioceptive sensitivities against the visual sensitivity. Finally, visual sensitivities depend on the quality of visual feedback. Despite the good quality of current visual VR interfaces (HMD), the resulting VR vision is not yet similar to real-world vision. These poor visual feedback can thus explain the “bad” visual sensitivities in VR.

**RQ3: How to differentiate users based on their sensory sensitivities?** We observed significant differences among participants for both analysis methods. We also observed that for Bayesian Analysis, the differences between participants are smaller. The small differences observed are certainly a consequence of the priors used in the analysis. The same priors are assigned to each participant, and the final values can be seen as a correction of the prior based on the experiment data. Therefore, it is not surprising to observe values that are relatively close and centered around the prior values for Bayesian analysis, and this can be seen as a limitation. However, we also observed a small standard deviation for most of the participants and the modalities for this analysis compared to classical analysis. Thus, any differences between participants become more significant and exploitable. In conclusion, Bayesian analysis appears to be more readily applicable for distinguishing participants based on their visual and proprioceptive sensory sensitivities. Classical analysis, while more direct, would require a larger number of trials to prevent inconsistencies in the final values.

In addressing these three research questions, we made a first step toward our long-term goal to develop adaptive visuo-haptic illusion techniques. Recently, Lebrun et al. [37] demonstrated the feasibility of adapting redirection amplitudes to each participant to increase the probability of not detecting the illusion. Their method requires a calibration task before appropriately adjusting the illusion. On the contrary, we aim to predict the optimal illusion amplitude thanks to users’ sensory sensitivities. They could serve as input for a future sensorimotor model predicting the trajectories of users’ hands and enabling the utilization of the comparator model [18]. We plan to use the stochastic optimal feedback control model [60] (SOFC). The interesting aspect of SOFC is the consideration of noises in motor commands and sensory perceptions. To implement the model, it is necessary to set a value for sensory noise. Usually, a significantly larger noise is assigned to proprioception compared to vision. The originality of our work is to be able to set relevant proprioceptive and visual noises for each participant. In particular, we would have more balanced visual and proprioceptive noises.
Additionally, it may be worthwhile to consider the causal inference model [36] and predict how far a virtual hand (or other limbs) can be offset from her real counterpart without disturbing the user. A user with good visual accuracy compared to his/her proprioceptive accuracy would be, for example, less likely based on the causal inference model to distinguish the virtual hand from the real hand.

9 CONCLUSION

In this paper, we motivate the relevance of studying users’ visual and proprioceptive sensitivities to study the detection of visio-haptic illusions. We aim at using these sensitivities as input for computational model predicting the consequences of illusion implementation. We had two contributions. Our first contribution is methodological. We adapted an experimental protocol from cognitive psychology to human-computer interaction (HCI) to capture the precision of each modality from multimodal tasks, and we used a Bayesian analysis method, Approximate Bayesian Computation (ABC), not employed in the study we are drawing inspiration from for our empirical data. Our second contribution is empirical. We collected and analyzed data from 12 participants to gain a better understanding of their behavior in virtual reality.

10 AKNOWLEDGEMENT

We thank all study participants for their time, and colleagues and reviewers of IHM’24 for their helpful comments. This work is partly funded by the ANR NeuroHCI ANR22-CE33-0006-01 and the ANR Show-me ANR-20-CE33-0010.

REFERENCES

According to this model, the SoA occurs when the prediction of the sensory consequence of an action matches the actual sensory feedback at the end of the action. When the brain creates a motor signal, it also generates a copy of this actual sensory feedback, which is referred to as the comparator model (see Figure 11).

The mechanism that is frequently cited as the best explanation of SoA is the so-called comparator model. According to this model, the SoA occurs when the prediction of the sensory consequence of an action matches the actual sensory feedback at the end of the action. When the brain creates a motor signal, it also generates a copy of this actual sensory feedback, which is referred to as the comparator model (see Figure 11).
motor command called "efferent copy". The sensory perception is then predicted based on this efferent copy [43, 68]. When the predicted and actual perceptions match, a strong SoA is experienced [19]. The comparator model can explain why we experience Agency over a virtual avatar. The pertinence of the comparator model is reinforced by the fact that the user representation does not have to be realistic. Even if the arm of the user is represented by a sphere in VR, the SoA over the sphere is possible if it mimics arm movement.

B EVOLUTION OF PRECISION WITH DISTANCE AND ORIENTATION

Interpreting the evolution of unimodal precision as a function of target position is difficult by looking only at the confidence ellipses in figures Figure 7 and Figure 8. We therefore display on Figure 12 the evolution of the mean of the areas of the confidence ellipses (and their standard deviation) as a function of the distance to the reference point of each modality (the depth r) for the two analysis methods. We note that for both analysis methods and for all modalities the areas increase with depth. Even if the classical treatment method presents important area variations, the large standard deviations prevent a clear reading of the results. This is why in the rest of this section we only present the curves obtained with the Bayesian Analysis.

On the Figure 13 we observe the azimuthal and depth variances as a function of r for the Bayesian analysis. We explain by these graphs the evolution of the areas of the Figure 12. We notice that the azimuthal variances of the proprioceptive modalities are increasing with r in contrast to their more stable depth variances. On the contrary, the visual modality is more stable in azimuth and varies strongly in depth. Thus it seems that the precision are stable in their most precise direction (depth for vision and azimuth for proprioception).

We study on the Figure 14 the evolution of the variances and the area of the confidence ellipses as a function of the orientation θ of the target with respect to the reference axis of each modality (perpendicular to the thorax and centered: at the middle of the body for vision, 200mm to the right or to the left for the respective proprioceptions). The azimuthal
variances are increasing with $\theta$ and the depth variances are decreasing. This explains the relative stability of the area with respect to $\theta$. Note that it is the visual modality that seems to be the most sensitive to $\theta$.

Exploiting these evolutions can help create from little data the precision of a modality and a prediction of how it will be impacted by a shift in space.