



**HAL**  
open science

## Pre-calibrated Visuo-Haptic Co-location Improves Execution in Virtual Environments

Justine Saint-Aubert, Loïc P Heurley, Nicolas Morgado, Stéphane Régnier,  
Sinan Haliyo

► **To cite this version:**

Justine Saint-Aubert, Loïc P Heurley, Nicolas Morgado, Stéphane Régnier, Sinan Haliyo. Pre-calibrated Visuo-Haptic Co-location Improves Execution in Virtual Environments. *IEEE Transactions on Haptics (ToH)*, 2019, p. 588-599. hal-02947074

**HAL Id: hal-02947074**

**<https://hal.sorbonne-universite.fr/hal-02947074>**

Submitted on 23 Sep 2020

**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.

# Pre-calibrated Visuo-Haptic Co-location Improves Execution in Virtual Environments

Justine Saint-Aubert, Loïc P. Heurley, Nicolas Morgado, Stéphane Régnier and Sinan Haliyo

**Abstract**—A common approach to visio-haptic human-machine interfaces adopts a simpler design by shifting grounded force feedback away from the virtual scene. The alternative design favors intuitiveness by displaying visual and grounded force feedback at the same location (i.e. visuo-haptic co-location) but requires a sensibly more complex implementation and a tedious kinematic conception. The benefits of one approach over the other had not been fully investigated. Notably, (i) while users seem to better operate under co-located condition, it's not always the case. (ii) In the case of a desktop interface, the cost of a complex implementation to achieve co-location is challenged. We aim here to resolve (i) by conducting a user-centered experiment in which participants performed two generic tasks in co-located and delocated configurations, and comparing their performances. Additionally, we intend to fill the gap (ii) by testing a design without continuous head tracking, i.e with static co-location. Participants' performances are assessed in terms of execution time, accuracy and force variation, while their subjective experiences are collected via a survey. Findings indicate that co-located configurations lead to shorter execution times, more accurate motions, better management of forces and are largely preferred by users, even when the co-location is pre-calibrated statically.

**Index Terms**—Visual and Haptic Co-location; Virtual Reality; User Experiment ; Cable Device.

## 1 INTRODUCTION

VISUO-HAPTIC CO-LOCATION refers to spatial congruence between visual and haptic signals and is inherent to interactions with objects in everyday life. However, it is much more difficult to achieve in virtual reality (VR), where distinct devices convey visual and haptic feedback separately. This study focuses on desktop use-case with an interface composed of a monitor coupled with a grounded force feedback device. Head mounted display [1] are not considered here because they are more dedicated to short simulations or may be too immersive. Target applications are dedicated to a single user performing long and fine operations, such as robotic teleoperation, remote surgery, or computer aided design. An intuitive human-machine interaction is then a considerable benefit. When users perform a manual task in such setup, they are shown an avatar (virtual replica) representing their actions, while feeling the force feedback in their hand through the handle of a mechanical motion input device. The common straightforward approach avoids superimposing spatially the avatar and the handle. This visuo-haptic delocation results in a simpler implementation in contrast with visuo-haptic co-location which supposedly brings a more intuitive interaction but requires a more complex design. The main difficulty is that the display has to deliver visual feedback at the same apparent location at which the haptic device delivers force feedback. In particular, it is required to continuously track user's head or eyes such that the avatar

is superimposed with the end-effector of the haptic device.

Questions for human-machine interface designers then pertain to knowing first, in what extent the co-location helps users to operate and second, if a trade off can be found by implementing co-location without head tracking. Our purpose was to provide answers through a user-centered behavioral experiment.

### 1.1 Integration issues

How the nervous system combines signals from various sensory modalities to reach a unified percept is one of the main questions in cognitive psychology and neurosciences. Several theories explain this multimodal perceptual integration [2]. A popular one suggests a weak fusion, where inputs are individual boxes estimated separately [3] then combined linearly, each weighted according to its reliability [4], [5] [6]. The sensory binding depends on the temporal, structural, and, more importantly, spatial congruence between the various sensory signals [7]. Visual and haptic signals provided at different locations result in integration breakdown [8]. Evidence that signals refer to a single object can help to establish a correlation, for instance if stimuli are linked by a kinematic relation [9]. However, the implication of haptic stimuli is lowered in delocated configurations [10]. It could then be argued that executions may be impaired, especially that involved in actions requiring fine grained processing of force feedback.

Several studies have used prismatic distortions inducing either a slight or a massive delocation of visual signals compared to haptic signals. Accumulated results support the idea that discrepancies alter perception and action, and induce more or less important sensorimotor adaptations [11], [12], [13], [14]. Even if such adaptations allow to reach good

J. Saint-Aubert, S. Régnier and S. Haliyo are with Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, ISIR, F-75005 Paris, France e-mail: (saint-aubert@isir.upmc.fr).

L. P. Heurley and N. Morgado are with Université Paris Nanterre, Centre de Recherche sur le Sport et le Mouvement, CeRSM, F-92000 Nanterre, France.

\*This work was performed within the Labex SMART supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11 IDEX-0004-02.

Manuscript received ?; revised ?.

motor performances, training sessions of varying duration are necessary depending on the magnitude of adaptations. Moreover, one has to plan and control his actions in order to act. When we want to reach an object with the hand, it has to be located relative to the location of the body. There are 3 possible reference frames: objects are located either relative to the hand, the eye, or the entire body. Recent data suggests that visual location and haptic location do not necessarily rely on the same reference frame [15]. Thus, providing delocated visuo-haptic signals could induce a competition between various reference frames during hand-object interactions. Together, these data support the idea that actions could depend on the spatial integration of visuo-haptic signals and that delocating them could induce a decrease of performances during executions, an increase in sensorimotor adaptation phases and/or difficulties in locating objects in space.

In order to act, an additional critical aspect is the need for users to have a feeling of effector ownership. Such a feeling occurs naturally in everyday life, and various studies using the rubber-hand-illusion have shown that integration of visuo-haptic signals is involved. Indeed, when visual and haptic feedback is temporally synchronized, participants can feel a rubber hand as if it were their own hand [16], [17]. It is noteworthy that the visual signal (of the rubber hand) and the haptic one (of the actual hand) are slightly delocated (the rubber hand is placed near the actual hand, which itself is hidden). Even if such data support that some brain mechanisms can overcome a slight delocation of visuo-haptic signals, only a portion of participants report the rubber-hand-illusion while some never experience it. Furthermore, when the illusion occurs, it is not immediate, with visuo-haptic signals having to be temporally synchronized during several minutes (around 5 minutes). Also, the actual hand has to remain static. When participants move it, the illusion vanishes. And finally, even if the illusion occurs, it does not seem strong enough to influence actions performed with the actual hand [18]. These data support that delocation of visuo-haptic signals could deeply diminish the feeling of ownership. Such a disturbance can decrease the sense of agency (i.e. the feeling of controlling our actions; e.g. [19]). All in all, these converging lines of research support the idea that visuo-haptic delocation could affect users' execution.

## 1.2 Impact on Execution

### 1.2.1 Previous findings

Past studies tend to show that co-located configurations lead to better performance in terms of time, accuracy and efforts necessary to achieve a task. An immediate benefit is shorter execution times in co-located configurations when compared with delocated ones, in the case of a task in which participants had to align a virtual cube with other intangible cubes. With [20] or without haptic feedback [21] (i.e. hand was displayed at the same location than the virtual environment but did not get any force feedback), participants were faster in the co-located configuration. A speed gain in co-located configurations was also found during a pointing task [22] as well as during needle insertion into surgical body phantoms [23]. However, both these tasks were performed on real objects which provide much more

cues than current haptic devices [24], so care must be taken when interpreting these results. Shorter execution times in co-located configurations were also observed during a surgical task where participants had to cut a virtual gut without touching its surrounding wall [25].

A second improvement seems to be greater accuracy in the co-located configuration. In the gut cutting task, this was indirectly demonstrated by the higher number of walls touched in the delocated configuration. More direct proofs were obtained from a 3D Fitts tapping experiment [26] where a significant decrease in end point error was obtained in the co-located configuration. A similar result was found in a task in which participants had to juggle a ball with racket in a virtual environment [20]. A perfect positioning between the ball and the racket was required to successfully accomplish the task.

A third possible enhancement concerns a better management of forces, i.e. better motor-control. In a task where participants explored the softness of an object [27], authors concluded that in co-located configurations, the same perception can be achieved with less energy consumption than in delocated configurations.

### 1.2.2 Limitations

Although several studies argue for an enhancement of participants' performances during executions in co-located configurations, the issue is not thoroughly settled.

Results are disparate across studies. While sometimes a greater accuracy is observed in one task, it is not the case in another one. Some studies did not even conform to these results at all. For instance, in an user experiment where participants had to use a stylus haptic device to navigate between cubes placed in a 3D maze game [20], no significant difference could be found between co-located and delocated configurations. The same conclusion was obtained during a 3D Fitts pointing exercise [28]. In some others studies performances in delocated configurations was even found to be better than in co-located configurations: In an experiment where participants were requested to point to the corner of a cube using a haptic device, accuracy was found to be better in delocated configurations [29]. However, when the exact same group of participants had to follow a spiral with the same set up, they were more accurate in co-located configurations. The nature of the task therefore seems to influence results. Unfortunately, all tasks tested in the literature are very loosely pertain to real executions in virtual environments, or lack external relevance as they are too specific.

## 1.3 Implementation Trade-Off

Co-location of visual and haptic feedbacks is matter of discussion because it complexifies the implementation. Indeed, the avatar in the virtual environment and the handle have to continuously occupy the same location from the user's point of view. It requires tracking his eyes or head in respect to the screen. It raises a first difficulty since the robustness of the tracking is often an issue. Once the different poses are inferred, they are used to find the correct location and orientation of a virtual camera in the VR environment. Consecutive frame changes are necessary in order to get

the correct perspective, which sum the relative error in each step. Consequently, the implementation of co-located feedback is more arduous than for delocated configuration for which none of these operations are critical.

In the case of a "desktop" interfaces which refer to a small working space, the head tracking may be less critical. In such devices, users' motions are quite limited and it seems that co-location can be achieved with an initial calibration and without continuous tracking. This idea is supported by results in a reaching task for targets where no differences in terms of performances were raised with and without head tracking after a pre-calibration [30]. All prior experiments presented required real-time tracking or HMD technologies that directly include trackers. The issue remains unanswered in case of desktop interface. We would like to explore here if a simpler implementation, with solely pre-calibrated co-location, is sufficient. In the affirmative, this could be an interesting trade-off for designers.

## 1.4 Aims of the Study and Methodology

This study investigates the effects of the relative location of visual and haptic feedbacks on user's manual execution on generic tasks and explores if a pre-calibrated static co-location is sufficient to improve performances.

For this purpose, two tasks are devised : a manipulation task consisting of moving an object along a line and an exploration task consisting of haptically circumventing a fixed obstacle. A total of 36 participants performed these tasks with a dedicated human-machine interface in a static (i.e without head tracking) co-located configuration (noted [C]) and two delocated configurations (noted [V'] and [H']). In the [V'], the visual feedback is shifted compared to the [C], while in the [H'] the haptic feedback is shifted compared to the [C]. Testing both [V'] and [H'] allows accounting for the posture difference between configurations and to differentiate the effects of co-location and the user posture. Position, time and force variability are recorded in all three configurations in order to evaluate performances. Furthermore, user's subjective experiences are collected with a post-study questionnaire.

## 2 METHODS

### 2.1 System Design

#### 2.1.1 Visual Display

The visual display is a computer monitor. Head-mounted displays (HMD) are not considered here for the desktop use-case. Hand access to the displayed scene and objects can be provided by using a 3D perspective, via stereoscopic cues. This requires the virtual scene to appear solely in front of the display as the hand can not go through the monitor and results in a very limited volume. The surface of the screen, also called the null parallax is a physical boundary and only the negative parallax volume is accessible.

Screen/mirror apparatus [24], [31] are the only technologies providing hand access to both positive and negative parallax spaces. The users see the screen reflected on a one-way mirror and put their hands below. They then can see their

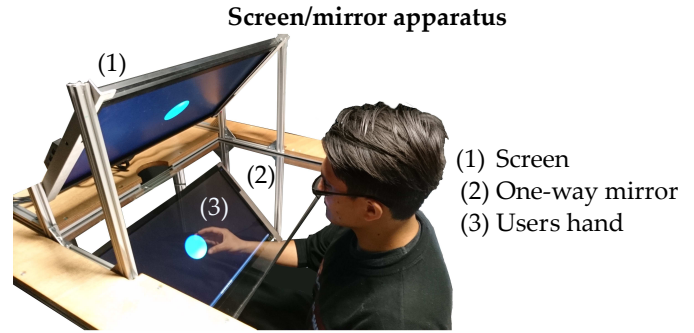


Fig. 1: Screen/mirror apparatus allowing co-location of the virtual environment and user's hand.

own hands inside the virtual environment reflected on the mirror as depicted in Fig. 1.

Note that if a HMD was used, a complex virtual representation of the hand would have been necessary.

#### 2.1.2 Co-Located Cable Interface

Current haptic technologies are not designed and are ill-adapted to achieve the co-located configuration [C]. The shared location with the virtual scene implies a non invasive device, yet most existing solutions are made of grounded rigid links. An interesting alternative consists in switching from rigid links to cables [32]. Such devices have already been implemented for the purpose of co-location with a large screen [33] but not for screen/mirror display. We have designed such a co-located desktop interface shown on Fig. 2, explained in detail in [34].

A screen (1) is tilted by  $\frac{\pi}{4}$  rad around the  $x_c$  axis from the vertical plane and placed above a one-way mirror (2). The haptic device is composed of five direct drive actuators (3) connected to a 5cm diameter spherical handle (4) via *Spectra* cables (5) through a system of pulleys (6). The actuators are arranged in an inverted pyramid with the base located 10 cm below the mirror in order to generate forces isotropically up to 5N applied to the handle within a cubic workspace about  $(37 \times 25 \times 20) \text{cm}^3$ .

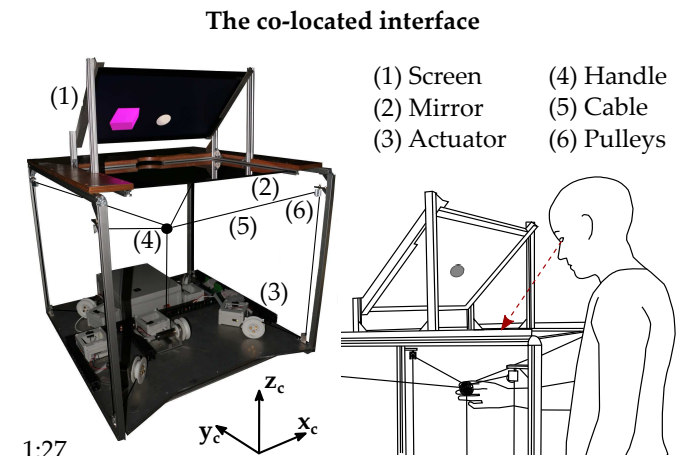


Fig. 2: Visual and haptic feedbacks are co-located by superimposing the reflection of the screen with a cable device.

Virtual items are displayed in the virtual scene using the open-source simulation software *Blender*<sup>1</sup>, in which the avatar is a faithful white copy of the real 5 cm diameter handle. The user's hand is not reproduced since it is visible across the mirror. When an interaction occurs, forces are generated by the haptic device in impedance mode. The spatial location is known via the length of wires and sent to the physics engine to control the avatar. God-object method performs the impedance/admittance conversion [35] to satisfy physical constraints and the coupling. When forces  $\mathbf{F}$  are detected by the physics engine in the software, each cable is loaded so as to build the equivalent mechanical action via the pseudo-inverse method specific to redundant mechanisms [36]. In no-contact situations, cables are simply taut and compensate each other's action.

The bandwidth is limited by the frame rate of the physics engine, including the communication with the haptic device, to around 500 Hz for both tasks tested in the experiment. Since the bandwidth of human sensing is about 300 Hz [37], the response time is adapted.

The monitor is capable of active 3D stereo with shutter-glass goggles. A pilot experiment tested the display with or without 3D and the choice is made to use only perspective cues. Users reported visual fatigue probably because the display did not succeed to keep up with the accommodation process of human eyes [38]. The conflict between accommodation and convergence on users' comfort has been established previously [39], [40].

Users operate the interface holding the handle of the device which they can freely move in the workspace while visualizing the avatar in the virtual scene and feeling the force feedback when in contact with other virtual objects.

### 2.1.3 Pre-calibrated Static Co-Location

For the purposes of the study, a static calibration, without head tracking, is employed. Fixed frame changes corresponding to the screen tilt and mirror image are implemented. Visual scaling and perspective between virtual and real items is guaranteed by setting the focal length and pose of the virtual scene as equal to those of the human eye. Finally the pose is simply adjusted to the users' height at the beginning of each session.

### 2.1.4 The Delocated Configuration [V']

The same interface is used to set up delocated configurations in order to avoid bias that could be introduced by eventual mechanical differences. In the [V'] configuration, the visual feedback is delocated while keeping the haptic feedback (i.e the cable device) in the same location as in the [C] configuration (Fig. 3). The virtual scene is shown in front of the participants with the screen simply set upwards, barring the mirror. The frame change is adapted in order to limit the shift between handle and avatar to a fixed translation of  $\mathbf{O}_c\mathbf{O}_v=[0,+0.2,+0.4]\text{m}$  from  $\mathcal{R}_c\{x_c, y_c, z_c\}$ . The haptic sensation remains the same than in the [C] configuration.

### 2.1.5 The Delocated Configuration [H']

The [H'] configuration is set up by shifting the haptic feedback while keeping the visual feedback at the same

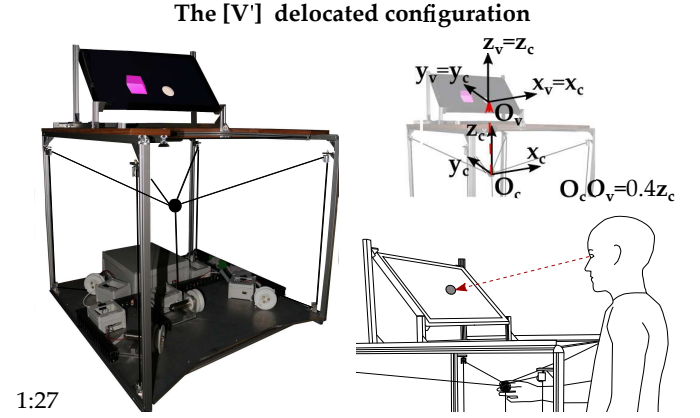


Fig. 3: In the [V'] delocated configuration, participants directly look at the screen without the mirror and operate in the same space as in the [C] configuration.

location as in the [C] configuration (Fig. 4). The comparison between [V'] and [H'] allows to evaluate the impact of user posture on execution and to differentiate it from the effect of the delocation. To set up [H'], the screen and mirror are shifted to the left side of the device. The participants are placed in front of the screen in the same way as in the [C] configuration and use the handle now located on their right side. For technical reasons the surface of the mirror is totally reflective, masking every object located below. As the participants' hands are outside of this zone the transparency is not an issue. The frame change is adapted to limit the shift between handle and avatar to a fixed translation of  $\mathbf{O}_c\mathbf{O}_h=[-0.2,-0.5,+0.2]\text{m}$  along  $[x_c, y_c, z_c]$ . The haptic sensation remains the same than in the [C] configuration.

## 2.2 Methodology

### 2.2.1 Participants

A random sample of 36 naive volunteers (half of them male) from the *Institut des Systèmes Intelligents et de Robotique*, Sorbonne Université, takes part in the experiment. The mean age is 25.4 years ( $s_{age} = 3.8$  years). They are selected

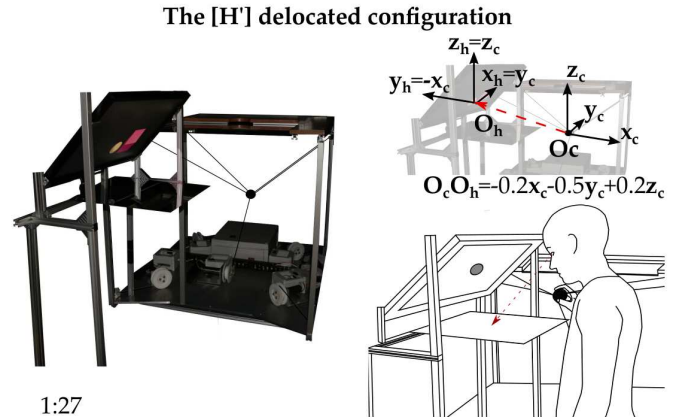


Fig. 4: In the [H'] delocated configuration, participants operate on the right side while observing an image at the same location as in the [C] configuration.

1. <https://www.blender.org/>

because they do not report any visual impairment (including color blindness), physical issues and are not regular consumers of psychoactive substances which could affect performance [41]. Only right-handed subjects according to Coren's handedness test [42] are recruited in order to perform the task in the [H'] configuration without bias.

### 2.2.2 Overview of Trials

Participants perform two tasks. In the *manipulation* task they guide a sphere from one point to another following a line. In the *exploration* task they probe and avoid an obstacle<sup>2</sup>. The virtual environment reproduces Newtonian physics without gravity and some amount of damping is applied to avoid excessive speeds and motions.

At the beginning of each task, the experimenter explains the overall procedure based on a pre-established script. She also informs participants that they can rest for as long as necessary between each trial. Next, participants perform one or two practice trials to ensure that they understand the procedure, after which pink noise is played to them and they perform the complete task.

### 2.2.3 Manipulation Task

At the beginning of a trial, a ghost grey line (0.25-m-long) with green and red extremities appears in the center of the screen (Fig. 5 (a)). Participants grasp the handle of the haptic device to control the avatar. They are requested to touch the green extremity, causing a tangible purple sphere to appear at a distant location (Fig. 5 (b)). They are asked to bring back the purple sphere in contact with the green extremity of the grey line by pushing it solely with the handle avatar. The line turns green to confirm contact (Fig. 5 (c)). They then push the purple object along the line (Fig. 5 (d)) to the red extremity (Fig. 5 (e)) where it should stay in contact for at least 2 seconds, to avoid ballistic strategies. After this, another trial starts (Fig. 5 (f)).

They perform a total of 20 trials with some variations. First of all, the orientation of the line changes from 0 to  $2\pi$  in the null parallax plane. Furthermore, different initial locations of the purple sphere are displayed, and its weight changes between 0.25 kg, 0.5 kg and 1 kg. The other features of the object (color, size, texture, friction=0.5) remain constant. The objectives of these variations is mainly to prevent participants from losing focus and to perform on-line adaptation. The order of variations is randomly predefined so that the same variations always appear in the same order between configurations and participants.

Other features are added in the virtual scene. Intangible but visible walls are shown around the scene to prevent participants from losing the purple sphere from view. A 0.15m-long box is thus constructed. If the purple sphere touches the boundaries the trial is considered a failure and the next trial automatically begins. Moreover, participants from a pilot study showed difficulties in moving the purple sphere efficiently, probably because of the lack of a stereoscopic-3D display. To help them, the purple sphere has the same dimensions as the handle of the haptic device and the background turns from black to dark red when the purple sphere moves off the line.

2. A third task in which they have to resist a perturbation was additionally conducted but not included in this study.

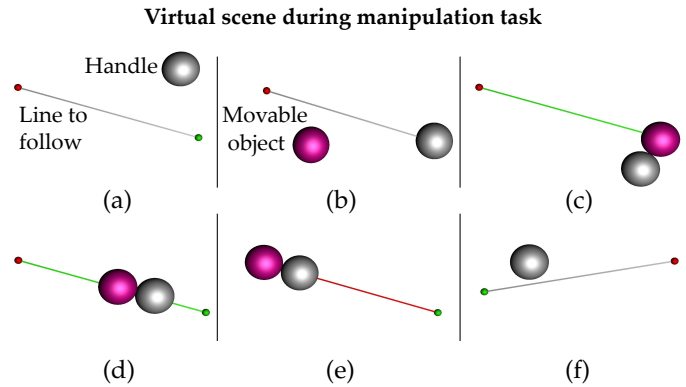


Fig. 5: Manipulation scene as seen by participants. They push the magenta sphere using the handle's avatar.

### 2.2.4 Exploration Task

At the beginning of a trial, a ghost grey line (0.25m-long) with green and red extremities appears in the center of the screen (Fig. 6 (a)). Participants first touch the green extremity with the avatar which coupled to the handle of the haptic device. The line turns green in order to confirm contact (Fig. 6 (b)). They are then requested to travel along the line towards the red extremity (Fig. 6 (c)) until they experience a contact indicating the presence of an invisible curved obstacle. They are asked to circumvent it by keeping contact (Fig. 6 (d)). These instructions make participants to explore the object quite naturally while their movements are sufficiently constrained to allow subsequent analyses. Once participants finish probing the obstacle, they proceed as before until they reach the red extremity (Fig. 6 (e)). After three seconds, another trial starts.

Participants perform a total of 30 trials with some variations. Again, the orientation of the line changes from 0 to  $2\pi$  in the null parallax plane. Furthermore, the location of the curved object as well as its shape vary along the line. The shape is created from one or several smooth spheres partially merged together. Thus, the distance traveled remains the same regardless of the direction of exploration chosen to circumvent the object. These variations are randomly predefined so that the same variations always appear in the same order between configurations and participants.

In the same way as in the manipulation task, participants are helped (except during contact with the object) by the background turning from black to dark red when the handle moves off the line.

### 2.2.5 Sequence of Configurations

Each participant complete the exploration task and then the manipulation task, respectively in all three configurations. A fully counterbalanced design is used to control potential effects of the session order, resulting in six groups in which participants are randomly assigned :

- Group 1 : [C] - [V'] - [H']
  - Group 2 : [C] - [H'] - [V']
  - Group 3 : [H'] - [V'] - [C]
  - Group 4 : [H'] - [C] - [V']
  - Group 5 : [V'] - [H'] - [C]
  - Group 6 : [V'] - [C] - [H']
- One session is about 45 minutes, subjects are then tested on different days in order to reduce fatigue and increase focus.

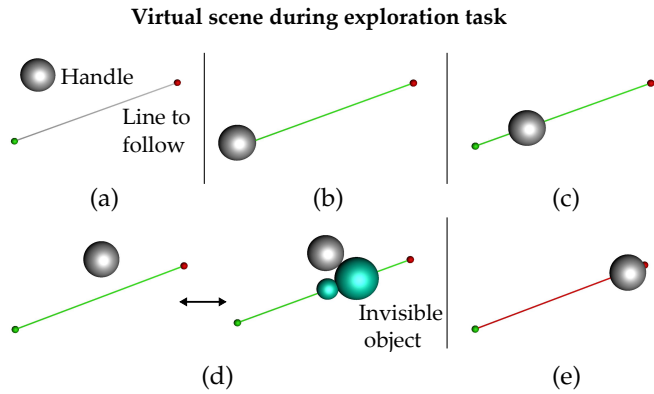


Fig. 6: Exploration scene as seen by participants. They follow the line with the handle and circumvent an invisible obstacle.

### 2.2.6 Collected Data

Information about time, position and force is recorded. The time is inferred using an *Intel Core i5-6600* cpu clock at 3.30 GHz but is limited to 2 ms precision by the physics engine frequency. The position is tracked via the cable system. With the hardware used, the system could theoretically achieve a precision below 1mm. However, an experimental set up is rarely perfect, so the position detection was pre-tested with an *Atracsys*<sup>3</sup> camera system commonly used for surgical applications. This system uses infrared signals and passive markers to detect the actual handle position with submillimetric accuracy at over 300 Hz. The locations detected with this system are compared to those recorded via the acquisition of cable lengths and a maximum absolute error of 5 mm at a distance up to 15 cm from the center of the device (i.e. the area of interest) was found.

Finally, contact forces are inferred in the physics engine with the god-object approach, with a spring-damper model. Since they relate to location of the proxy and the handle in the virtual world, the accuracy of the measurement depends on the cumulative error between the instant of contact and the penetration in the virtual world and is therefore complicated to estimate.

In the manipulation task, the time taken by participants to bring the sphere from its home position/the beginning to the end of the line is collected and defined as "*Travel time*". The recorded position is that of the purple sphere and defined as "*Travel distance*". The force corresponding to the contact force between the avatar and the purple sphere and is used to determine "*Contact force variability*". Finally, the percentage of wall hits is calculated and labeled "*Failed trials*". This last measurement influence all the other data since failed trials are left out in the estimation of the mean. In the exploration task, the time taken by participants to bring the handle from the green to the red extremity is recorded and once again defined as "*Travel Time*". The position information of the handle during this lapse of time is used to define a "*Distance to line*" measurement. It corresponds to the error surface between the line and the handle, computed as the sum of distances between 1000

points regularly spaced on the line and their corresponding nearest points of the handle center position. It is only recorded in the case where no contact with the probed object exists. Otherwise, when the handle touches the object, its position is used to determine the surface explored. It is then subtracted from the existing surface to obtain the "*Exploration error*"<sup>4</sup>, and is either negative when participants miss some surfaces or positive when deviations occur. Finally, the force is recorded and corresponds to the contact between the avatar and the invisible object. It is used to compute "*Contact force variability*".

### 2.2.7 Surveys

Participants' impressions are collected with a questionnaire. Points of particular interest concern the evaluation of their own performances, their difficulties during both tasks, their fatigue and relative preferences about the configurations. They fill out a questionnaire at the end of each of the three sessions relating to each individual configuration as well as a questionnaire at the end of the study which aim to compare all configurations. Subjects participate on three different days. Each one is dedicated to the test of one configuration and between 3 and 5 days elapse between the first and the last configuration. We deem this duration short enough to get reliable comparison. In any case, users' subjective judgment reported during the last session will be correlated with users' subjective judgment reported at the end of each session.

Participants answer questions on a visual analog scale (track bar) ranging from 0 to 100 and whose extremities are associated with verbal labels. They enter their observations by adjusting a cursor on the line using the computer mouse with a step precision of 1 (this information is not displayed to participants). The default location of the cursor is the middle of the line. Their answers are recorded only if participants click on the corresponding location so as to avoid missing answers. In the comparison questionnaire, three analog scales are presented for each question, one for each configuration. These three scales are all independent and displayed one above the other so as to simplify comparisons. The experimenter explain how to fill out these questionnaires before leaving the room.

The participants are asked several questions. Since some of them are intended to investigate similar concepts, we decide to gather questions are :

- **Self-assessed efficiency during manipulation/ line exploration/shape exploration** : The higher this combined score, the more participants felt efficient in performing the task.
- **Workload**. The higher this combined score, the more participants felt overloaded.
- **Intuitiveness** : The higher this score, the more participants felt that a given configuration is intuitive.
- **Spatial Spotting** : The higher this score, the more participants felt that they could spot their spatial location in a given configuration.
- **Preference** : The higher this score, the more participants preferred a given configuration.

3. Atracsys System fusionTrack 500. <https://www.atracsys-measurement.com/wp-content/documents/ftk500-datasheet.pdf>

4. More details on the "*Distance to line*" and "*Exploration error*" data can be found in Annex

The respective questions are listed in Tab. 1.

### 3 RESULTS

#### 3.1 Results

Results are shown as the means from every trial for all participants and for each configuration. Fig. 7 depicts results for the manipulation and Fig. 8 for the exploration. Results from the surveys are presented in Fig. 9.

##### 3.1.1 Orthogonal Contrast Analyses

The travel time, travel distance, mean force and force variability for each task (manipulation and exploration) are the dependent variables for the analysis of performances. For the analysis of the survey, the dependent variables are the self-assessed efficiency in manipulation, self-assessed efficiency in line exploration, self-assessed efficiency in shape exploration, workload, intuitiveness, spatial spotting and preference scores. In both analyses, the relative location of the visual and haptic feedback is the within-subject independent variable with three levels ([C], [H'] and [V']). Since the independent variable has more than two levels, conducting an omnibus repeated measure ANOVA to test its main effect would lead to a vague test of the hypotheses [43].

We therefore conduct separate analyses on each dependent variable for each task by testing two orthogonal contrasts. First, the delocated haptic [H'] configuration (coded as -1) is compared with the delocated vision [V'] configuration (coded as +1) without considering the co-located configuration [C] (coded as 0). Second, these two delocated configurations are grouped (both coded as -1) and compared to the co-located configuration (coded as +2). The results of the null hypothesis significance tests conducted with a significance threshold at 0.05 are provided for each analysis (i.e. test statistics and p-value) along with an unstandardized effect size estimate (i.e. mean difference) with its 95% confidence interval (CI) and a common standardized effect size  $\hat{\eta}_p^2$  expressing the percent of variance in the dependent variable explained by the independent variable.

For all the analyses, we checked whether our data were

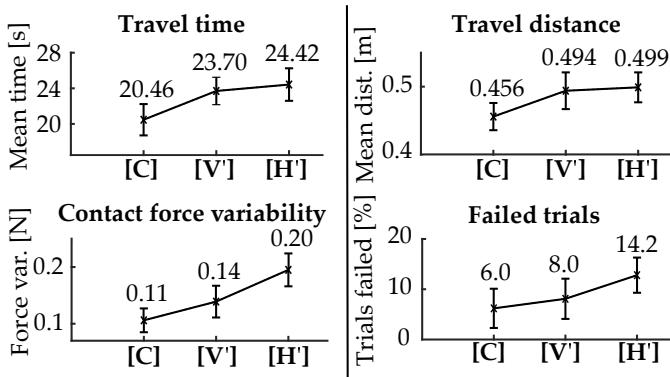


Fig. 7: Data collected during the manipulation task. Each graph presents the corrected mean of participant's mean values for all 20 trials respectively for the [C], [V'] and [H'] configurations. Error bars illustrate the 95% confidence interval for between-participant spread.

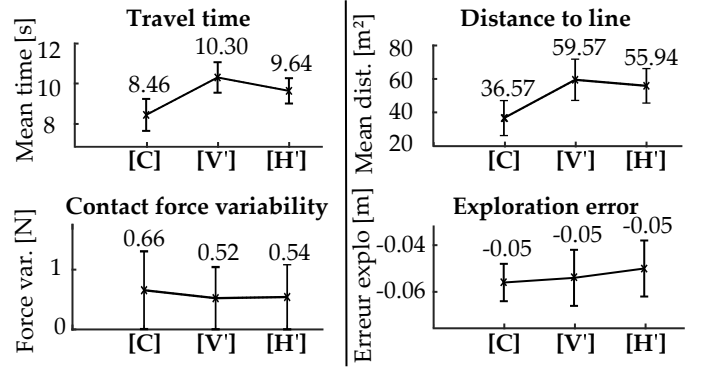


Fig. 8: Data collected in the exploration experiment. Each graph presents the corrected mean of participant's mean values for all 20 trials respectively for the [C], [V'] and [H'] configurations. Error bars illustrate the 95% confidence interval for between-participant spread.

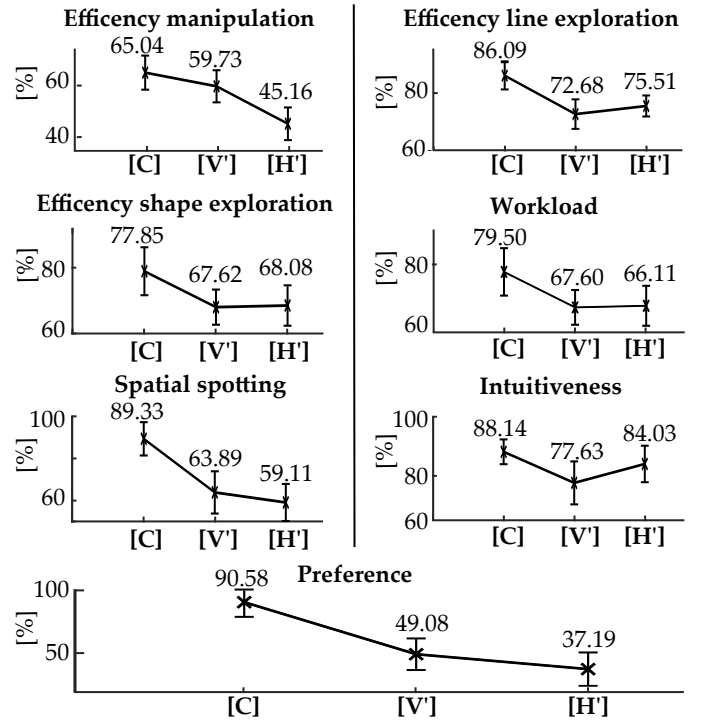


Fig. 9: Data collected in the survey. Each graph presents the corrected mean of participant's mean values for all 20 trials respectively for the [C], [V'] and [H'] configurations. Error bars illustrate the 95% confidence interval for between-participant spread.

reasonably compatible with the normality and homoscedasticity assumptions of the generalized linear model<sup>5</sup> by using quantile-quantile plots and residual-against-prediction plots (e.g. [45], [46]). Moderate violations of assumptions were observed for some analyses. However, we did not use data transformations to correct them because we judged that the analytic benefits did not compensate for interpretative costs. For all the analyses, we also checked the presence of out-

5. We handle the sphericity assumption common in within-subject designs by testing contrasts with only one degree of freedom, making this assumption unnecessary [44].



Questions	Verbal Labels
<b>Self-assessed efficiency manipulation</b>	Cronbachs $\alpha = 0.80$ , CI [0.64, 0.90]
Do you think that you succeeded in pushing the sphere along the line ?	[I didn't succeed at all - I perfectly succeeded]
Do you think that you succeeded in pushing the sphere along the line ?	[It was very difficult - It was very easy]
For each day, express the difficulty you had to follow the line.*	[It was very difficult - It was very easy]
<b>Self-assessed efficiency for line exploration</b>	Cronbachs $\alpha = 0.76$ , CI [0.58, 0.87]
Do you think that you succeeded in following the line ?	[I didn't succeed - I perfectly succeeded]
Do you think that you succeeded in following the line ?	[It was very difficult - It was very easy]
For each day, express the difficulty you had to follow the line.*	[It was very difficult - It was very easy]
<b>Self-assessed efficiency for shape exploration</b>	Cronbachs $\alpha = 0.81$ , CI [0.66, 0.90]
Do you think that you succeeded in circumventing the object while keeping in contact ?	[I didn't succeed - I perfectly succeeded]
Do you think that you succeeded in circumventing the object while keeping in contact ?	[It was very difficult - It was very easy]
For each day, express the difficulty you had to circumvent the object while keeping in contact.*	[It was very difficult - It was very easy]
<b>Workload</b>	Cronbachs $\alpha = 0.69$ , CI [0.49, 0.83]
Did you stay focussed during the entire task?	[Not at all - Totally]
Did you feel tired?	[Not at all - Completely]
For each day, express the difficulty you had to staying focussed.*	[It was very hard - It was very easy]
For each day, express your state of mental fatigue.*	[I was not tired at all - I was very tired]
For each day, express your state of physical fatigue.*	[I was not tired at all - I was very tired]
<b>Intuitiveness</b>	
Do you think that the interface is intuitive?	[Not at all - Totally]
<b>Spatial Spotting</b>	
For each day, express the difficulty you had to locate yourself in space.*	[It was very difficult - It was very easy]
<b>Preference</b>	
If you had the opportunity of using this device again, would you choose to ?*	[Not at all - Totally]

TABLE 1: Set of questions and verbal labels grouped for each item in the survey. The combination consistency is confirmed for every item by computing an average Cronbachs  $\alpha$  across configurations. Questions stamped with \* correspond to the cross-configuration evaluation performed at the tail end of the experiments.

lying trials for each participants and outlying participants in our sample. We defined an outlier as a data point for which the absolute studentized deleted residual (also called externalized studentized residual) is equal to or larger than 3 [44]. To limit the decrease of our sample size resulting from blindly discarding outliers, we conducted our analyses with and without outliers and discarded only those that dramatically changed our results. A non-trivial number of outliers is discarded. Some of them came from obvious bugs in the recording of the data while others came from the users struggling to operate in the virtual environment.

### 3.1.2 Complementary Analyses

As frequentist analyses have received a lot of criticism (e.g. [47]), the analyses are complemented with Bayes Factors (BF) which indicate the relative evidence for one hypothesis against another. We computed these with *JASP*<sup>6</sup> using the default multivariate Cauchy prior recommended by [48]. For more technical details, see [49]. In the analyses, the  $BF_{10}$ <sup>7</sup> is interpreted as the strength of evidence supporting the existence of an effect (H1) relative to the strength of evidence supporting the non-existence of that effect (H0). The classification proposed by [48] is used to qualify this strength of evidence as null (BF = 1), anecdotal (BF = [1, 3]), moderate (BF = [3, 10]), strong (BF = [10, 30]), very strong (BF = [30, 100]), and extremely strong (BF > 100). The complete statistical analysis is provided in Table 3.

6. <https://jasp-stats.org/>

7.  $BF_{10}$  expresses the strength of evidence supporting the non-existence of an effect (H0) relative to the strength of evidence supporting the existence of that effect (H1).

## 3.2 Co-Located [C] versus Delocated [H',V']

### 3.2.1 Manipulation Task

For the manipulation task, speed, accuracy and contact stability are found to be higher in co-located configurations. Indeed, we observe lower execution times. If we consider the mean total time necessary to perform the complete task to be about 6 min 50 s, a gain of 1 min and 12 s is obtained in the co-located configuration. In the same vein, the "Distance to Line" data show that accuracy is higher in the co-located configuration than in delocated ones, a result which is confirmed by the "Self-assessed efficiency manipulation" data from the survey. These findings support an enhancement of performances in term of speed and accuracy in the co-located configuration.

An increase of force variability "Contact force variability" is also observed in the delocated configurations. The co-located configuration somehow leads to more stable operation. This observation correlates with the higher percentage of "Failed Trials" in delocated configurations when compared to the co-located configuration.

### 3.2.2 Exploration Task

Again, speed and accuracy are found to be better in co-located configurations. An increase in execution time is observed in delocated configurations when compared to the co-located configuration. If we consider the mean total time necessary to perform the complete task to be about 4 min 14 s, a gain of 46 s is noticed in the co-located configuration. Moreover, the accuracy is higher in the co-located configuration during the no-contact phase where participants had to follow the ghost line. This result is confirmed by the "Distance to line" quantitative data and

the "Spatial spotting" as well as "Self-assessed efficiency line exploration" data from the survey. Nevertheless, these results do not hold during the contact phase. The "Contact force variability" and "Exploration error" data are not significantly different.

### 3.2.3 General Comparison

In both tasks, participants exhibit a lack of confidence in delocated configurations which is reflected by the "Self-assessed efficiency manipulation", "Self-assessed efficiency line exploration" and "Self-assessed efficiency shape exploration" data. It's also the case even when the task is well executed, based on the "Exploration error" data.

In general, participants prefer the co-located configuration over delocated configurations because they feel less overloaded and find the interface more intuitive, as depicted by the "Preference", "Workload" and "Intuitiveness" data.

## 3.3 Delocated [V'] versus Delocated [H']

### 3.3.1 Manipulation and Exploration Task

No significant differences in terms of time or accuracy error are observed in the manipulation task. The only significant finding concerned the "Contact Force Variability", which is higher in the [H'] configuration and correlated with more "Failed Trials". This is confirmed by the "Self-efficacy data" from the survey.

No significant differences are observed between delocated configuration during the exploration task.

### 3.3.2 General Comparison

The data relative to the general comparison between configurations do not show any significant difference. We had doubts about the fatigue caused by the [H'] configuration when compared to the [V'] configuration. Indeed, the distance between the elbow and the trunk in the former was larger than in the latter, so maintaining posture in the former was more tiring. However, no significant differences are found in terms of "Workload", leading us to conclude that participants were not biased by fatigue during operations.

## 4 DISCUSSIONS

### 4.1 Better Performances

The role of visio-haptic location on users' performances during a manipulation task is investigated. Past studies on the issue showed that operations were sometimes performed faster, sometimes more accurately in co-located configurations. The present study informs that time and accuracy are both better managed during the same application.

Additional result indicate a difference in contact stability, while previously only a change in the maximum forces was exhibited during a softness exploration tasks [27]. The stability issue here highlights a permanent effect that disturb operation all over the simulation. The difference between co-located and delocated configurations can seem quite low since of a total of 20 simulations, the mean increase of 4.66 % represents about 1 failed simulation, but we believe that this number is mainly low because of the dynamics implemented in the virtual environment. The inertia and dumping of the purple sphere are set in order to allow

its movement while giving participants some leeway before failing. Lower values for these parameters could lead to a higher number of fails. In summary, those findings strengthen that the co-location of visual and haptic feedback helps users to manipulate.

### 4.2 Task Dependency

The role of visio-haptic location on users' performances during an exploration task is also investigated. The operation mainly concerns the contact phase, however no difference are found between co-located and delocated configurations. The effect of the delocation then depends on the task performed. This evidence is consistent with the deviation of results between previous studies exhibited in the introduction section. Indeed, some did not conclude on an enhancement of performances in co-located configuration. The present study profits that the same participants used the same devices to perform the manipulation and the exploration task in the experiment. Hence, a reliable comparison can be performed.

A first factor that can explained the variation of performances between the tasks is the type of feedback. In the exploration task, the only visual cue that is available is the handle/avatar location, as the object is invisible. And since it doesn't concern direct information about the object but information derived from the force feedback, we can assume that haptics are somehow dominant in the process. If the visual cues are reduced, small conflicts should occur between visual and haptic cues regardless of the configuration. This is the case here. Unlike this phase, the no-contact phase of the exploration task and the manipulation task both involved visual and haptic feedback, that can lead to differences between configurations. A similar comment applies to a previous study that explored performances in a 3D Fitts task [28]. No haptic feedback was provided and no difference were found between co-located and delocated configurations. However, in the same experiment with haptic feedback [26], difference were identified.

Another factor that may influence results is the level of complexity. In the manipulation task, participants are required to handle the avatar and a dynamic object in the scene, while object to probe in the exploration task is static. Previous studies that involved complex operations (e.g needle insertion [23], surgery [23] or juggle [20]) found differences between co-located and delocated configurations.

A last, but not least, possible factor concerns the dimension of the space. The manipulation task is performed in 3D space while the exploration one is mainly executed in a plan. In past studies, no enhancement of performances was raised in co-located configuration during a pointing task [29]. However, the same participants followed a 3D spiral in co-located configuration with higher accuracy. Additionally, enhancement of performances was exhibited when an entire cube has to be aligned [20], [21], a task that required to handle correctly the position of each face in the 3D space.

### 4.3 Confidence Issue and Cognitive Load

A lower confidence coming from participants in delocated configuration is detected in the exploration task. A lower score for the [V'-H'] contrast than for the [C] in terms of

Data	Mean values	CI 95%	Frequentist analyses (Significativity)	Bayesian analyses (Bayes Factor)
<b>Manipulation task</b>				
Contrast	$\mu_{[H',V']-[C]}$			
Travel time	3.6s	[2.06, 5.12]	S, F(1,32) = 23.03, p < 0.001, $\hat{n}_p^2 = 0.42$	BF10 649.68 <b>Extremely Strong</b>
Travel distance	0.04m	[0.02, 0.06]	S, F(1,33) = 21.24, p = 0.0001, $\hat{n}_p^2 = 0.39$	BF01 413.62 <b>Extremely Strong</b>
Contact force variability	0.06N	[0.04, 0.08]	S, F(1,32) = 45.99, p < 0.001, $\hat{n}_p^2 = 0.59$	BF10 130729.24 <b>Extremely Strong</b>
Failed Trials	4.28%	[0.96, 7.59]	S, F(1,34) = 6.88, p = 0.01, $\hat{n}_p^2 = 0.17$	BF10 3.44 <b>Moderate</b>
Contrast	$\mu_{[H']-[V']}$			
Travel time	0.72s	[-0.96, 2.40]	N.S, F(1,32) = 0.78, p = 0.39, $\hat{n}_p^2 = 0.01$	BF01 0.33 <b>Moderate</b>
Travel distance	0.01m	[-0.02, 0.03]	N.S, F(1,33) = 0.15, p = 0.70, $\hat{n}_p^2 = 0.02$	BF01 5.08 <b>Moderate</b>
Contact force variability	0.06N	[0.03, 0.09]	S, F(1,32) = 14.58, p = 0.001, $\hat{n}_p^2 = 0.31$	BF10 52.30 <b>Very Strong</b>
Failed trials	4.66%	[0.97, 8.35]	S, F(1,34) = 6.60, p = 0.01, $\hat{n}_p^2 = 0.16$	BF10 3.07 <b>Moderate</b>
<b>Exploration task</b>				
Contrast	$\mu_{[H',V']-[C]}$			
Travel time	1.52s	[0.84, 2.29]	S, F(1,32) = 20.44, p < 0.001, $\hat{n}_p^2 = 0.39$	BF10 313.15 <b>Extremely Strong</b>
Distance to line	21.19m <sup>2</sup>	[12.09, 30.29]	S, F(1,34) = 22.40, p < 0.001, $\hat{n}_p^2 = 0.40$	BF10 134.35 <b>Extremely Strong</b>
Contact force variability	0.12N	[-0.26, 0.02]	N.S, F(1,17) = 3.44, p = 0.08, $\hat{n}_p^2 = 0.17$	BF01 3.87 <b>Null</b>
Exploration error	+0.004m	[-0.003, 0.011]	N.S, F(1,33) = 1.20, p = 0.28, $\hat{n}_p^2 = 0.04$	BF01 4.71 <b>Moderate</b>
Contrast	$\mu_{[H']-[V']}$			
Travel time	0.66s*	[-1.38, 0]	S, F(1,32) = 4.20, p = 0.049, $\hat{n}_p^2 = 0.12$	BF10 1.18 <b>Anecdotal</b>
Distance to line	3.62m <sup>2</sup> *	[-15.34, 8.10]	N.S, F(1,34) = 0.39, p = 0.53, $\hat{n}_p^2 = 0.01$	BF01 4.59 <b>Moderate</b>
Contact force variability	0.02N*	[-0.09, 0.13]	N.S, F(1,17) = 0.14, p = 0.72, $\hat{n}_p^2 = 0.1$	BF01 3.87 <b>Moderate</b>
Exploration error	0.004m	[-0.010, 0.017]	N.S, F(1,33) = 0.31, p = 0.58, $\hat{n}_p^2 = 0.01$	BF01 3.14 <b>Moderate</b>
<b>Survey</b>				
Contrast	$\mu_{[H',V']-[C]}$			
Efficiency manip.	12.59%*	[6.86,18.33]	S, F(1,32) = 20.02, p < 0.001, $\hat{n}_p^2 = 0.39$	BF10 277.57 <b>Extremely Strong</b>
Efficiency line explo.	11.99%*	[-16.17,-7.81]	S, F(1,34) = 34.12, p < 0.001, $\hat{n}_p^2 = 0.51$	BF10 11682 <b>Extremely Strong</b>
Efficiency shape explo.	10.00%*	[-15.91,-4.09]	S, F(1,34) = 11.84, p = 0.002, $\hat{n}_p^2 = 0.41$	BF10 21.62 <b>Strong</b>
Workload	12.65%	[9.24, 16.06]	S, F(1,34) = 56.82, p < 0.001, $\hat{n}_p^2 = 0.63$	BF10 1.37E+6 <b>Extremely Strong</b>
Intuitiveness	7.31%*	[3.68,10.95]	S, F(1,34) = 16.75, p < 0.001, $\hat{n}_p^2 = 0.33$	BF10 110.37 <b>Extremely Strong</b>
Spatial Spotting	27.83%*	[ 20.98,34.69]	S, F(1,35) = 67.98, p < 0.001, $\hat{n}_p^2 = 0.66$	BF10 1.115E+7 <b>Extremely Strong</b>
Preference	47.44%*	[37.29,57.60]	S, F(1,34) = 16.75, p < 0.001, $\hat{n}_p^2 = 0.33$	BF10 2.912E+8 <b>Extremely Strong</b>
Contrast	$\mu_{[H']-[V']}$			
Efficiency manip.	14.57%*	[-20.76, -8.39]	S, F(1,32) = 23.03, p < 0.001, $\hat{n}_p^2 = 0.42$	BF10 649.87 <b>Extremely Strong</b>
Efficiency line explo.	2.83%	[-1.5,7.16]	N.S, F(1,34) = 1.76, p = 0.192, $\hat{n}_p^2 = 0.51$	BF01 2.42 <b>Anecdotal</b>
Efficiency shape explo.	0.46%	[-4.36,5.28]	N.S, F(1,34) = 0.037, p = 0.847, $\hat{n}_p^2 = 0.26$	BF01 5.34 <b>Moderate</b>
Workload	1.57%	[-3.07, 6.04]	N.S, F(1,34) = 0.44, p = 0.51, $\hat{n}_p^2 = 0.01$	BF01 4.59 <b>Moderate</b>
Intuitiveness	6.40%	[-1.00, 13.80]	N.S, F(1,34) = 3.09, p = 0.09, $\hat{n}_p^2 = 0.08$	BF01 1.38 <b>Anecdotal</b>
Spatial Spotting	4.78%*	[-14.67, 5.11]	N.S, F(1,35) = 0.96, p = 0.33, $\hat{n}_p^2 = 0.03$	BF01 3.58 <b>Moderate</b>
Preference	11.89%	[1.46,25.24]	N.S, F(1,35) = 3.27, p = 0.08, $\hat{n}_p^2 = 0.09$	BF01 1.29 <b>Anecdotal</b>

TABLE 2: Statistical analyses of the  $\mu_{[H',V']-[C]}$  contrast and  $\mu_{[H']-[V']}$  contrast for the manipulation task, the exploration task and the survey. Mean values are provided for each data. The "\*" symbols stand for a negative value. For instance, in the exploration task  $\mu_{[H']-[V']}$  contrast, the mean travel time in the  $[H']$  configuration is 0.66 s less than in  $[V']$  configuration. The second last column shows the frequentist analyses, N.S stands for "not significant" and S for "significant". Finally, the last column corresponds to the bayesian analyses and shows the correlated classifications.

"Self-assessed efficiency" appears in the survey. Although their performances were similar in all 3 configurations, most participants indicated that they thought to have done better in the co-located case.

Our assumption is that this self-assessed efficiency score is correlated to subjective rating of mental effort, indicating a lower cognitive load [50], [51]. The task hence appears easier to the user because less demanding.

#### 4.4 Delocated $[V']$ Versus Delocated $[H']$

The direct comparison between  $[V']$  and  $[H']$  brings supplemental information. No differences in terms of time or accuracy are found. However, a difference in terms of contact stability and failed trials is apparent.

These findings are first relevant in order to judge the effect of the co-location of feedback. Indeed, they show that a part of the difference in term of stability and failed trials found

in the  $[H',V']-[C]$  contrast could be attributed to the change of posture. The other part is due to the delocation itself.

Additionally, the differences between delocated configurations are particularly interesting when co-location is not an option, for instance when visual and haptic have different working spaces. User's velocity, accuracy and management of forces are then affected by the fact that feedbacks are not co-located. But the management of forces is also impaired by the posture of the user. The type of delocation should then be considered.

#### 4.5 "Simple" Co-Location

All these conclusions are established with a pre-calibrated simple co-location, i.e without head tracking. This introduces a small error in superimposition of the handle and its virtual avatar when the user moves his head during the simulation. The improvement in execution in the co-located configuration, supported by two statistical analyzes

strengthen our conviction that the trade-off to drop the continuous head tracking is relevant. The approach would lend itself well for use in any desktop interfaces that do not directly dispose of head tracker or for which the precision of co-location is not critical.

#### 4.6 Stereoscopic Cues

One downside of our methodology is the absence of visual stereoscopic cues. Pilot experiments showed that users preferred to avoid 3D glasses. This shortcoming is probably stems from the light polarization introduced by the reflection resulting in a disagreeable experience. However, stereoscopic cues were absent in each configuration. This absence equally affects both collocated and delocated configurations in our design, hence comparative results remain valid. With stereoscopic cues, the reliability of the visual signal would have been higher in all cases, but how this impact is cognitively integrated remains to explore.

### 5 CONCLUSION AND PERSPECTIVES

The spatial congruence between visual and haptic feedbacks has been investigated in the case of a desktop display coupled with a grounded haptic device. In tested task users' execution is faster, more accurate and they better manage their motor-control in a co-located configuration. The also have a higher opinion of their performances indicating a lesser cognitive load. These findings are observed with a static pre-calibration. These argue that a simple co-location improves users' performances, without technical complexity of head tracking. Moreover, these indicate that an acceptable discrepancy range exists in the handle/avatar superposition. This range and its influence on users' performances can be explored further.

### REFERENCES

- [1] P. R. Desai, P. N. Desai, K. D. Ajmera, and K. Mehta, "A review paper on oculus rift-a virtual reality headset," *arXiv preprint arXiv:1408.1173*, 2014.
- [2] J. J. Clark and A. L. Yuille, *Data fusion for sensory information processing systems*. Springer Science & Business Media, 2013, vol. 105.
- [3] D. Marr, "Vision: A computational investigation into the human representation and processing of visual information. freeman and company," *San Francisco*, 1982.
- [4] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and modeling of depth cue combination: In defense of weak fusion," *Vision research*, vol. 35, no. 3, pp. 389–412, 1995.
- [5] M. S. Landy and H. Kojima, "Ideal cue combination for localizing texture-defined edges," *JOSA A*, vol. 18, no. 9, pp. 2307–2320, 2001.
- [6] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, 2002.
- [7] B. E. Stein and M. A. Meredith, *The merging of the senses*. The MIT Press, 1993.
- [8] S. Gepshtein, J. Burge, M. O. Ernst, and M. S. Banks, "The combination of vision and touch depends on spatial proximity," *Journal of Vision*, vol. 5, no. 11, pp. 7–7, 2005.
- [9] N. B. Debats, M. O. Ernst, and H. Heuer, "Kinematic cross-correlation induces sensory integration across separate objects," *European Journal of Neuroscience*, vol. 46, no. 12, pp. 2826–2834, 2017.
- [10] M. Congedo, A. Lécuyer, and E. Gentaz, "The influence of spatial delocation on perceptual integration of vision and touch," *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 3, pp. 353–357, 2006.
- [11] H. L. Pick and J. C. Hay, "Adaptation to prismatic distortion," *Psychonomic Science*, vol. 1, no. 1-12, pp. 199–200, 1964.
- [12] P. Sachse, U. Beermann, M. Martini, T. Maran, M. Domeier, and M. R. Furtner, "the world is upside down—the innsbruck goggle experiments of theodor erismann (1883–1961) and ivo kohler (1915–1985)," *Cortex*, vol. 92, pp. 222–232, 2017.
- [13] M. Yarossi, T. Manuweera, S. V. Adamovich, and E. Tunik, "The effects of mirror feedback during target directed movements on ipsilateral corticospinal excitability," *Frontiers in human neuroscience*, vol. 11, p. 242, 2017.
- [14] O. Ossmy and R. Mukamel, "Behavioral and neural effects of congruency of visual feedback during short-term motor learning," *NeuroImage*, vol. 172, pp. 864–873, 2018.
- [15] M. Avillac, S. Deneve, E. Olivier, A. Pouget, and J.-R. Duhamel, "Reference frames for representing visual and tactile locations in parietal cortex," *Nature neuroscience*, vol. 8, no. 7, p. 941, 2005.
- [16] M. Botvinick and J. Cohen, "Rubber hands feeltouch that eyes see," *Nature*, vol. 391, no. 6669, p. 756, 1998.
- [17] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, "First person experience of body transfer in virtual reality," *PloS one*, vol. 5, no. 5, p. e10564, 2010.
- [18] M. Kammers, F. de Vignemont, L. Verhagen, and H. C. Dijkerman, "The rubber hand illusion in action," *Neuropsychologia*, vol. 47, no. 1, pp. 204–211, 2009.
- [19] K. Ma and B. Hommel, "The role of agency for perceived ownership in the virtual hand illusion," *Consciousness and cognition*, vol. 36, pp. 277–288, 2015.
- [20] D. Swapp, V. Pawar, and C. Loscos, "Interaction with co-located haptic feedback in virtual reality," *Virtual Reality*, vol. 10, no. 1, pp. 24–30, 2006.
- [21] M. Mine, "Exploiting proprioception in virtual-environment interaction," *Unpublished Doctoral thesis, University of North Carolina, Chapel Hill*, 1997.
- [22] J. P. Rolland, F. A. Biocca, T. Barlow, and A. Kancherla, "Quantification of adaptation to virtual-eye location in see-thru head-mounted displays," in *Virtual Reality Annual International Symposium, 1995. Proceedings.* IEEE, 1995, pp. 56–66.
- [23] B. Wu, R. L. Klatzky, D. Shelton, and G. D. Stetten, "Psychophysical evaluation of in-situ ultrasound visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 6, pp. 684–693, 2005.
- [24] E. D. Graham and C. L. MacKenzie, "Physical versus virtual pointing," in *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM, 1996, pp. 292–299.
- [25] D. D. Lev, R. Rozengurt, T. Gelfeld, A. Tarchenshvili, and M. Reiner, "The effects of 3d collocated presentation of visuo-haptic information on performance in a complex realistic visuo-motor task," in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications.* Springer, 2010, pp. 432–437.
- [26] M. J. Fu, A. D. Hershberger, K. Sano, and M. C. Çavuşoğlu, "Effect of visuo-haptic co-location on 3d fitts' task performance," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on.* IEEE, 2011, pp. 3460–3467.
- [27] A. Widmer and Y. Hu, "Effects of the alignment between a haptic device and visual display on the perception of object softness," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 6, pp. 1146–1155, 2010.
- [28] R. J. Teather, R. S. Allison, and W. Stuerzlinger, "Evaluating visual/motor co-location in fish-tank virtual reality," in *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference.* IEEE, 2009, pp. 624–629.
- [29] P. Olsson, F. Nysjö, S. Seipel, and I. Carlbom, "Physically co-located haptic interaction with 3d displays," in *Haptics Symposium (HAPTICS), 2012 IEEE.* IEEE, 2012, pp. 267–272.
- [30] J. Boritz and K. S. Booth, "A study of interactive 3 d point location in a computer simulated virtual environment," in *VRST*, vol. 97. Citeseer, 1997, pp. 181–187.
- [31] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson, "Holodesk: direct 3d interactions with a situated see-through display," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2012, pp. 2421–2430.
- [32] M. Sato, "Development of string-based force display: Spidar," in *8th International Conference on Virtual Systems and Multimedia.* Citeseer, 2002.
- [33] L. Bouguila, M. Ishii, and M. Sato, "Effect of coupling haptics and stereopsis on depth perception in virtual environment," in *Proc. of*

the 1st Workshop on Haptic Human Computer Interaction, 2000, pp. 54–62.

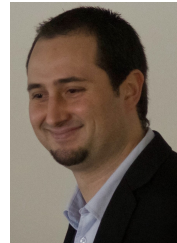
- [34] J. Saint-Aubert, S. Régnier, and S. Haliyo, "Cable driven haptic interface for co-localized desktop vr," in *Haptic Symposium, 2018 IEEE*. IEEE, 2018.
- [35] C. B. Zilles and J. K. Salisbury, "A constraint-based god-object method for haptic display," in *Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on*, vol. 3. IEEE, 1995, pp. 146–151.
- [36] R. Kurtz and V. Hayward, "Dexterity measures with unilateral actuation constraints: the n+ 1 case," *Advanced robotics*, vol. 9, no. 5, pp. 561–577, 1994.
- [37] E. Samur, *Performance metrics for haptic interfaces*. Springer Science & Business Media, 2012.
- [38] N. Padmanaban, R. Konrad, T. Stramer, E. A. Cooper, and G. Wetstein, "Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays," *Proceedings of the National Academy of Sciences*, p. 201617251, 2017.
- [39] N. Hiruma and T. Fukuda, "Accommodation response to binocular stereoscopic tv images and their viewing conditions," *SMPTE journal*, vol. 102, no. 12, pp. 1137–1140, 1993.
- [40] T. Inoue and H. Ohzu, "Accommodation and convergence when looking at binocular 3d images,[in] human factors in organizational design and management iii, k. noro and o. brown, jr," 1990.
- [41] J. Wittenborn, "Effects of benzodiazepines on psychomotor performance." *British journal of clinical pharmacology*, vol. 7, no. S1, 1979.
- [42] S. Coren, *The left-hander syndrome: The causes and consequences of left-handedness*. Simon and Schuster, 2012.
- [43] M. Brauer and G. McClelland, "L'utilisation des contrastes dans l'analyse des données: Comment tester les hypothèses spécifiques dans la recherche en psychologie?" *L'année Psychologique*, vol. 105, no. 2, pp. 273–305, 2005.
- [44] C. M. Judd, G. H. McClelland, and C. S. Ryan, *Data analysis: a model comparison approach to regression, ANOVA, and beyond*. Routledge, 2017.
- [45] W. D. Berry, *Understanding regression assumptions*. Sage Publications, 1993, vol. 92.
- [46] J. Fox, *Regression diagnostics: An introduction*. Sage, 1991, vol. 79.
- [47] G. Cumming, *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, 2013.
- [48] E.-J. Wagenmakers, M. Marsman, T. Jamil, A. Ly, J. Verhagen, J. Love, R. Selker, Q. F. Gronau, M. Šmíra, S. Epskamp *et al.*, "Bayesian inference for psychology. part i: Theoretical advantages and practical ramifications," *Psychonomic bulletin & review*, vol. 25, no. 1, pp. 35–57, 2018.
- [49] J. N. Rouder, R. D. Morey, P. L. Speckman, and J. M. Province, "Default bayes factors for anova designs," *Journal of Mathematical Psychology*, vol. 56, no. 5, pp. 356–374, 2012.
- [50] F. Paas, P. Ayres, and M. Pachman, *Assessment of Cognitive Load in Multimedia Learning Theory, Methods and Applications*, 11 2008, pp. 11–35.
- [51] F. A. Haji, D. Rojas, R. Childs, S. de Ribaupierre, and A. Dubrowski, "Measuring cognitive load: performance, mental effort and simulation task complexity," *Medical Education*, vol. 49, no. 8, pp. 815–827, 2015. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/medu.12773>



**Justine Saint-Aubert** obtained her Ph.D in Robotics at Sorbonne University in 2019, after a M.Sc. degree in Mechatronics Systems for Rehabilitation by Sorbonne University and University Degli Studi di Brescia. She's interested in human perception and interaction, especially around haptic. She investigates cognitive models for multisensory integration and their use for intuitive interfaces within virtual environments.



**Loïc P. Heurley** is an associate professor at Nanterre University in the Center for Research on Sport and Movement (CeRSM, EA 2931). He is internationally acknowledged for his work in Cognitive Psychology. His main goal is to better understand cognitive processes underlying how humans interact with their environment (perception-action coupling). He is particularly expert on the level of compatibility between various information sources during manual actions.



**Nicolas Morgado** is an associate professor at Paris Nanterre University and is affiliated to the Center for Research on Sport and Movement (CeRSM, EA 2931). He works on the influence of physical and social action constraints on the perception of spatial layout since 2009. More particularly, he is interested in how the effort associated with reaching an object influences people's visual perception of its distance. He has also worked on the way social action constraints and visual context influence physical performance.



**Stéphane Régnier** is a Professor at the Institute of Intelligent Systems and Robotics (ISIR), Sorbonne-Université. Currently, he is the dean of Science faculty at Sorbonne-Université. His research interests include micro and nanomanipulation, teleoperation and haptic feedback at the nanoscale, micromechatronics, and biological cell characterization. He has published more than 150 papers in journals and conferences. He is the associate editor for IEEE Transaction on Robotics and Journal of Micro-Bio Robotics.



**Sinan Haliyo** is an associate professor at Institute of Intelligent Systems and Robotics (ISIR), Sorbonne University and has been active in robotics since 1999. His main topics of interest include control and design issues, physical interactions and user interfaces for micro-scale applications in assembly, characterization and user training. He also takes a particular interest in human-computer interaction issues in remote handling and teleoperation, especially with haptics and multimodal interfaces.