When High Fidelity Matters: AR and VR Improve the Learning of a 3D Object

Thibault Louis
Univ. Grenoble Alpes
Grenoble, France
thibault.louis@univ-grenoble-alpes.fr

Jocelyne Troccaz
CNRS
Grenoble, France
jocelyne.troccaz@univ-grenoble-alpes.fr

Amélie Rochet-Capellan
CNRS
Grenoble, France
amelie.rochet-capellan@grenoble-inp.fr

Nady Hoyek
Universite Claude Bernard Lyon 1
Villeurbanne, France
nady.hoyek@univ-lyon1.fr

François Bérard
Univ. Grenoble Alpes
Grenoble, France
francois.berard@univ-grenoble-alpes.fr

ABSTRACT

Virtual and Augmented Reality Environments have long been seen as having strong potential for educational applications. However, research showing actual evidences of their benefits is sparse. Indeed, some recent studies point to unnoticeable benefits, or even a detrimental effect due to an increase of cognitive demand for the students when using these environments. In this work, we question if a clear benefit of AR and VR can be robustly measured for a specific education-related task: learning a 3D object.

We ran a controlled study in which we compared three interaction techniques. Two techniques are VR- and AR-based; they offer a High Fidelity (HF) virtual reproduction of observing and manipulating physical objects. The third technique is based on a multi-touch tablet and was used as a baseline. We selected a task of 3D object learning as one potentially benefitting from the HF reproduction of object manipulation. The experiment results indicate that VR and AR HF techniques can have a substantial benefit for education as the object was recognized more than 27% faster when learnt using the HF techniques than when using the tablet.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Mixed / augmented reality**; **Virtual reality**.

KEYWORDS

User study, Spatial augmented reality, Virtual reality, Head mounted display, Mental rotation, Learning task.

ACM Reference Format:
Thibault Louis, Jocelyne Troccaz, Amélie Rochet-Capellan, Nady Hoyek and François Bérard. 2020. When High Fidelity Matters: AR and VR Improve the Learning of a 3D Object. In *International Conference on Advanced Visual Interfaces (AVI '20), September 28-October 2, 2020, Salerno, Italy.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3399715.3399815

AVI '20, September 28-October 2, 2020, Salerno, Italy

ACM ISBN 978-1-4503-7535-1/20/09…$15.00
https://doi.org/10.1145/3399715.3399815

1 INTRODUCTION

Technological advances have made Virtual Reality and Augmented Reality environments easier to create and to deploy. As a consequence, VR and AR are becoming widespread. In the fields of learning and education, these virtual environments have long been seen as having great potential by offering a richer interaction with the learned notions than videos or interactive applications on desktop and tablet computers. This enhanced experience should translate into higher *learning outcome*.

However, while AR and VR environment seem to increase students' motivation and interest, empirical evidences of better learning outcome are difficult to find in the literature. Recent results point to unnoticeable benefits [26], or even a detrimental effect [24, 27]. This could be explained by an increase of *extraneous* cognitive load induced by the virtual environments. Theories of learning point to the intrinsic vs. extraneous cognitive load in knowledge acquisition [33]. Any learning task has an intrinsic cognitive load that does not depend on the medium used to learn. In contrast, the extraneous cognitive load induced by navigating VR or AR environments may be superior than that induced by using desktop applications.

Still, AR and VR environments both offer stereo and head coupling that have been shown to offer improved depth perception. This could translate to better learning when depth perception is critical to the learning task. In addition, active exploration from students was shown to be beneficial to visual recognition tasks [16] in spite of the higher extraneous cognitive load compared to a more guided observation (such as videos or static images). Furthermore, the extraneous cognitive load of AR and VR environments might be lowered if the virtual objects are manipulated with an *isomorphic* control: users' control on the input device is directly transferred on the virtual object; which allows for intuitive and efficient control of rotations [17].

In this study, we aimed at observing a measurable benefit on the *learning outcome* of AR and VR environments with isomorphic control. We qualify these environments has "high fidelity" (HF) 3D interaction in the sense that the visual perception and the control of the viewpoint and the objects is very similar to that in the physical world, in contrast to more traditional learning environments on desktop or tablet computers. The experimental task is the *learning of a 3D object*; which refers to the cognitive process of creating a mental representation of the *morphology* of an initially unknown object. This mental representation can then be used to recognize the object's morphology in a scene, regardless of its orientation. We chose this learning task as one that may benefit from the improved depth perception offered by HF 3D interaction, and as one that is related to actual learning tasks: students in anatomy need to learn 3D objects such as bones and organs. This is also the case for other students in science, technology, engineering, and mathematics (STEM).

We evaluate 3 interaction techniques. As the status quo, we use a tablet that allows active exploration of a 3D object with multi-touch interaction and a simple perspective rendering of the object. We compare this technique with two HF techniques that offer isomorphic control of the object and head-coupled stereo rendering. The techniques are a VR technique using an opaque HMD, and an innovative AR technique named Handheld Perspective Corrected Display (HPCD) introduced by Stavness et al. in 2010 [32]. The HPCD technique was recently shown as the most efficient and intuitive technique for 6D manipulation (rotations + translations) [23]. Both the VR and HPCD interaction techniques use a 30cm diameter sphere as a prop to manipulate the virtual objects. The study allows testing our main hypothesis: both HMD and HPCD should outperform the tablet in a 3D object-learning task. The more realistic rendering and more intuitive manipulation provided by the HF techniques should ease the construction of a mental representation of the object by reducing the information overload. This hypothesis was first expressed, but not tested, by Lee et al. [22].

Our main contribution is to bring strong empirical evidence of the benefit of HF 3D interaction on the *learning outcome* of a task of 3D object learning.

2 RELATED WORK

2.1 VR does not generally improve the learning outcome

In a recent study on the effect of immersive Virtual Reality on a biochemistry science lab simulation, Makranskya et al. mention that "Virtual reality is predicted to create a paradigm shift in education and training, but there is little empirical evidence of its educational value" [24]. Indeed, in their study comparing a desktop PC simulation vs. a VR simulation using a head-mounted display, they found that students reported being more present in the VR condition; but they learned less, and had significantly higher cognitive load based on the EEG measure. These results are in line with an earlier study from Parong et al.; which compared the instructional effectiveness of immersive virtual reality (VR) versus a desktop slideshow as media for teaching a biology lesson about how the human body works [27]. The results showed that students who viewed the slideshow performed significantly better on the posttest than the VR group. In both studies, the researchers used VR to immerse the students in a virtual environment related to the learned notions: biochemistry, circulatory system and parts of cells. The learned notions were not specifically related to the improved 3D perception offered by the VR interaction. The benefit of VR and AR HF 3D interaction may be found on notions that are more directly related to 3D perception.

2.2 Learning 3D objects

2.2.1 Constructing a mental representation of a 3D object. One of the key requirements for medical and STEM students is to be able to identify organs or mechanical parts; which are inherently 3D; from static 2D viewpoints (x-ray scans or line art drawings, for example). Learning to identify these objects usually relies heavily on the Mental Rotation (MR) abilities of the students. MR abilities have been shown to improve learning of various knowledge, including functional anatomy, spatial geometry, chemistry, surgery, architecture and engineering design [18, 36]. A common requirement of these domains is to be able to represent and manipulate 3D objects mentally. Guillot et al. [15] suggested that the score at MR tests such as the Vandenberg Mental Rotation Test (VMRT) [35] predict success in anatomy learning. It was observed that some students fail their exams because of low MR and more generally of low spatial abilities [29, 36]. As MR vary considerably among people, students with low MR abilities suffer from inequalities in STEM courses.

Guillot et al. [15] identified that MR plays a major role in the construction and memorization of the mental representation of a 3D object, especially when only static projective views are used (like pictures in a technical book). They suggest that using more faithful representations of the object like animated videos, physical models or 3D virtual data, may lower the impact of MR abilities when learning a 3D object.

2.2.2 Exploiting the benefits of VR interaction. High fidelity 3D interactive systems offer the possibility for the learner to interact with a more faithful representation of the learned object. Schnabel et al. [30] explored how immersive environments affect the way people create a mental model of a spatial volume. They tested an immersive stereoscopic virtual environment using an HMD alongside a non-immersive desktop environment displaying 2D representations. They found that participants using VR devices had a better understanding of the 3D volume and of its components than participants using 2D representations. Lee and Wong [21] investigated the impact of a VR learning environment to teach frog anatomy, in comparison with a classic PowerPoint lesson. Students with poor MR abilities greatly learned from the instinctive and implicative VR environment. The authors state that mentally recreating a 3D object from 2D representations produce extraneous cognitive load that may overload students' working memory, especially for those who struggled to perform mental spatial transformation tasks. Ye et al. observed similar behaviours when comparing VR environments vs. static representations in an assembly planning task [37]. Jan et al. compared active vs. passive exploration in a VR setup [19] and found similar results as Berney et al. [5]: students with low MR abilities greatly benefited from the active exploration; which was not the case with student having high MR abilities.

2.2.3 AR studies so far. VR devices shut users from their surrounding; which can be detrimental in a teaching context. This motivates the use of Augmented Reality setups that can offer high fidelity representations while staying connected to the physical surroundings. Moreover, hand-eye coordination is better in AR than in VR; which is is likely to improve object manipulation performances [20, 23]. Few studies to date have managed to show a benefit of AR in a teaching environment. Implementing a robust and convincing AR system (e.g. accurate, high resolution, high frame rate, and low latency) is still a challenge compared to VR setups; which could explain this lack of material. In 2011, Chen et al. tested the use of tangible models and AR models in an engineering graphics course [8]. While students enthusiastically welcomed the use of AR models, Chen et al. only observed a thin improvement in students' ability to map 3D objects to 2D images. The use of a tangible model, on the other hand, increased significantly their performances. However, the AR setup did not allow all rotations of the model, and the authors reported that it was sometimes used inappropriately. The use of a more *natural* AR device, i.e. offering a control over the virtual model closer to how we manipulate physical objects [4], may have resulted in performances more in line with those of the tangible model. Shin et al. studied how an observer perceives a complex 3D object in an AR scene when changing viewpoints either from observer movement or from object rotation [31]. They showed that moving around the object was more efficient to perceive it than rotating the object itself, highlighting the benefits of head coupling to interact with 3D scenes.

2.3 Use of VR and AR to learn to recognize 3D objects

If the benefits of 3D representations and active exploration have been demonstrated in a 3D object learning context, there is still no formal evidence of the impact of the level of realism of a 3D interaction on object recognition. In addition, we are not aware of any study that included a comparison of an AR and a VR condition in this learning context.

Studying high fidelity VR and AR technologies is an interesting lead to learn 3D objects efficiently by exploring them as *naturally* as possible. In addition, practicing manipulations with 3D objects could be an intuitive and easy method to train MR, as it requires the user to actively explore and rotate the objects. Indeed, the framework of *grounded cognition* [1] suggests future reactivations of the sensorimotor experience of the object, helping to create embodied MR strategies that could generalize to similar objects. This way, exploiting recent high fidelity VR and AR systems in courses could enhance both students' mental rotation and applied learning.

3 USER STUDY DESIGN

3.1 Overview

We designed a user study to observe how three interaction techniques affected the learning of a generic 3D object: the techniques include two high fidelity VR and AR interactions, and a low fidelity multi-touch tablet. We used a single object so as to limit participants' experiment duration (around 45 minutes) and cognitive demand. We call the learned object "the `sphinx`" (c.f. Figure 1, left).

Each participant was *trained* using one of the three interaction techniques. We chose this between-subject design to prevent any transfer of knowledge between sessions. Participants were repeatedly presented with objects viewed from various *initial viewpoints*. Objects were either the `sphinx` or an object with a slightly different shape. Participant could rotate the objects with their assigned interaction technique and observe them from many viewpoints. They had to identify whether the objects were the `sphinx` or not.

We measured the recognition of the `sphinx` in *recognition* sessions: participants were presented with a series of the same objects used in the *training* and viewed from various *fixed* viewpoints on a simple perspective tablet. For every presented object, participants had to decide whether the presented object was the `sphinx` or not.

Finally, a major design choice was to spread participants' experimental sessions over two successive days in order to favor some stabilization of the learning process and especially, the integration of the sensorimotor experience to visual representations.

3.2 The sphinx and the distractors

Many objects from everyday life can easily be recognized from many viewpoints thanks to strong visual cues like color, size or texture patterns. Hence, memories of all possible viewpoints or mental rotations are *not* necessary to recognize them. As we focused on the influence of the interaction on the learning process, we used an unknown object that *does not* have such cues, such as a bone viewed on x-ray images. We created the `sphinx` as an abstract object that does not look like any real-life object to avoid any potential bias induced by participants' previous experiences. The `sphinx` is asymmetrical so that it has a different appearance when seen from different viewpoints. We created 3 similar objects to be used as distractors from symmetries and partial rotations of the `sphinx`. The distractors are illustrated alongside the `sphinx` on Figure 1 (right).

The `sphinx` has to be distinguished from the distractors either after rotations from an *initial* viewpoint during the training sessions or from a *fixed* viewpoint during the recognition sessions. For the initial and fixed viewpoints, we used 32 rotations from a *reference orientation* defined from 8 rotation axes: (1,0,0), (0,1,0), (0,0,1), (1,1,1) and their opposites, and from 4 rotation angles (40°, 80°, 120° and 160°).

3.3 Protocol details and rationales

Figure 2 presents the all experiment timeline. Every step of the protocol is detailed here after:

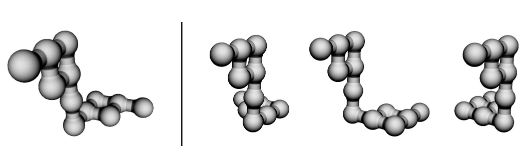

Figure 1: The `sphinx` (left) and the three distractors (right).

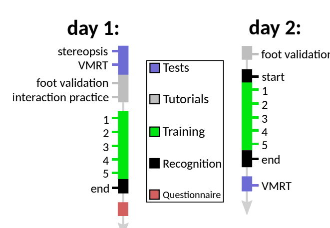

Figure 2: The experiment timeline per day.

Figure 3: Experimental setup showing the tablet (A) and the spherical device (B) for all training groups and during recognition (reco).

3.3.1 Pre- and Post- Tests. As two of the tested interaction techniques used stereo rendering, we screened participants to insure that they had good stereoscopic vision using the Stereo Optical RANDDOT stereopsis test.

With a between subjects design, the performance difference between groups could be partly explained by a difference in the initial performance in "3D object recognition" of the groups. As there is no standard test for this ability, we chose to balance the groups according to their MR abilities using the VMRT [35]. Previous work suggests that MR abilities should play an essential part in both learning and recognizing 3D objects [15, 25]. In addition, by asking participant to perform a second VMRT at the end of the experiment, we could assess how the 3 interaction techniques affected participant's MR abilities.

3.3.2 Interaction tutorials. Participants must tell if a shown object is the `sphinx` or not. To provide this answer, we asked participants to tap their right foot on the floor for the `sphinx`, or to tap their left foot otherwise. We chose this validation technique in order to avoid any disturbance in the working area that demanded the constant use of the two hands and a constant visual focus. In pilot tests, we also found it faster and more robust than a vocal validation. At the beginning of each day, participants went through a quick tutorial of about 1 minute to get used to this foot validation technique.

In addition, at the beginning of the first day, they familiarized with the interaction technique that they were assigned to in another 1-minute tutorial: they practiced the interaction with 3D letters (i.e. objects unrelated to the `sphinx`).

3.3.3 Training to learn the `sphinx`. On each day, participants executed 5 *training sessions* of 3 minutes. In each, they performed as many *training trials* as possible during the 3 minutes. A training trial began by a *target* object being displayed in one of the 32 initial orientations. The target object was either the `sphinx` or a distractor. Participants rotated the target object freely until they could answer with the foot validation technique. They immediately got a feedback that let them know if they answered correctly, and the next trial began. This repeated until the end of the session. The number of trials performed in a training session depended on the pace of the participants: the faster they answered, the greatest the number of answers in the session.

In the first training session, only one distractor was used: the target object was either the `sphinx` or this distractor. The tablet displayed help in the form of the distractor on the left and the `sphinx` on the right, both in their reference orientation. This help is illustrated in various conditions on Figure 3. Participants *had to* rotate the target object in its reference orientation and this was enforced by the system: they could not provide an answer before reaching this orientation. The first training session was designed to let participants discover the `sphinx` and to offer a first simple identification strategy, i.e. orienting the target object to get a familiar viewpoint where the `sphinx` was well known and could be easily recognized.

During the second training session, the help only showed the `sphinx` in its reference orientation. All 3 distractors were used, and participants could answer at any time. This session was included because we found in pilot studies that hiding the `sphinx` while introducing new distractors was too confusing.

In the last 3 training sessions, no help was provided: participants had to rely only on their memory to tell if the test object was the `sphinx`. The total number of training sessions was limited to 5 in order to maintain the duration of the experiment below 45 min. per day.

3.3.4 Recognition. Participants' recognition of the `sphinx` was measured during 3 minutes long *recognition sessions*. All participants regardless of their group executed recognition sessions in the same low fidelity modality. In a recognition trial, an object was displayed in one of the 32 orientations on the *tablet*. As soon as participants decided if the object was the `sphinx` or not, they answered with the foot validation technique, there was no feedback on the correctness of the answer, and a new recognition trial was started.

In addition to the recognition sessions at the end of each day, participants performed a recognition session at the beginning of day 2. This session allowed us to isolate the effect of a night of sleep on the knowledge they acquired the day before.

3.3.5 Questionnaire. We collected participants' subjective ratings with a brief questionnaire designed to provide first insights on users' acceptance of the various techniques. They rated 5 sentences on a scale from 1 ("I totally disagree") to 5 ("I totally agree"). The sentences were: "The experiment gave me pain in the eyes, headaches, or nausea." (mental pain), "I felt some muscular fatigue" (motor fatigue), "This experiment was fun" (fun), "I could manipulate the objects instinctively" (ease of learning) and "I could manipulate the objects precisely" (precision).

3.4 Interaction techniques

3.4.1 HMD condition. Participants wore an opaque HMD and got immersed in a virtual scene imitating the experimental room, as illustrated on Figure 3, left. The table, the tablet and the sphere were tracked and reproduced at their actual position in the virtual scene. This way, participants could grab and orient the virtual sphere instinctively by manipulating the real sphere. To improve contrast, we used a black background for the tablet and the sphere in the virtual scene. The sphinx or a distractor was displayed at the center of the virtual sphere, as if rigidly attached to it. Rotation of the object was thus isomorphic to the rotation of the physical sphere. This high fidelity condition offered Head-Coupled Stereoscopy and Isomorphic Control (HCSI).

3.4.2 HPCD condition. Introduced by Stavness et al. [32], Hand-held Perspective Corrected Displays (HPCDs) are volumetric displays that can be held and manipulated. The images displayed at the surface of the device are computed to create the illusion that the 3D scene is inside the device. Recent improvements include wireless operation, a lightweight device, and stereoscopic rendering [2]. Participants were holding the same sphere as in the HMD condition, but the sphere was used as an HPCD: it provided both input and graphical output as illustrated on Figure 3, 2nd from left. As with the HMD, this is a high fidelity condition that offers HCSI. The inclusion of HPCD was motivated by recent results that found HPCD superior than HMD for control [23].

3.4.3 TABLET condition. In this condition, all graphical feedbacks were displayed on the tablet. In pilot studies, we let participant use the tablet as they wanted, but a very large majority grabbed the tablet with their non-dominant hand and interacted with the dominant hand, i.e. they didn't leave the tablet on the table. This should not come as a surprise as it is coherent with Guiard's kinematic chain model [14]. In the experiment, we chose to enforce this dual handed use for uniformity concerns. The controllable object was rendered in the middle of the screen as illustrated on Figure 3, 3rd from left. Single finger interaction controlled the rotation of the object around the x- and y-axis of the screen, while two-finger interaction controlled the rotation in the tablet's plan. To be coherent with current tablets, stereoscopy and head coupling were not used. This low fidelity (simple perspective) condition is representative of systems that are widely available in learning institutions.

3.4.4 Secondary information. In all the conditions, the tablet was used to display secondary information during the trials (remaining time, success indicator, and the help when available), as illustrated on Figure 3. The 3D objects of the help where rendered using stereo and head coupling for the two HF conditions.

3.4.5 Recognition interaction. The objects were displayed on the center of the tablet without stereo or head coupling, as illustrated on Figure 3, right. Rotation interaction was disabled: HMD and HPCD participants left the sphere on the side, and removed the HMD or the shutter glasses; TABLET participants put the tablet back on the table. The remaining time was displayed above the objects.

3.5 Technical Setup

As system performances can be an important factor in human performances, we equalized as much as possible the performance of the system for the 3 interaction techniques. All participants' actions were sensed by the same optical tracking system (Optitrack with 10 Prime 13 camera at 240 Hz). The optical system offered high tracking performances in term of precision and stability. In particular, we measured the tracking jitters at 0.022mm of standard error; which was imperceptible. All the conditions were implemented in the same custom-developed C++ software running on an Intel Xeon 3.7Ghz computer with an NVidia GeForce GTX 1080 graphic card.

Images were created by the same projector for the HPCD and TABLET conditions (Barco F-50, 2560x1600@120 Hz pixels). Lights were turned off to maximize the contrast of the projection. The tablet was simulated by a 60 cm $\times$ 40 cm wooden board painted in white. Markers were attached to the board for tracking purpose. In the HMD condition, images were created by an HTC Vive (2160x1200@90 Hz pixels with a 110° field of view) and tracking was insured by the optical tracking system, i.e. we did not use the HMD's tracking.

We used the same tracking in all conditions, and the same projector in the HPCD and TABLET conditions. This contributed to the internal validity of the experiment. In addition, it avoided a predictable drop of performances in the TABLET condition induced by device latency and jitter [34]. Our tablet had the same 27ms latency as the HPCD; which was notably lower than current commercial tablets (around 80ms). The 0.022mm jitter from the tracker was clearly low enough as users' precision with a finger on a touch surface has been measured at 0.17 mm [3]. We estimated the pixel density at around 90 dpi in HPCD and TABLET, which is significantly lower than the one offered on current commercial tablets. We considered that this would not be a factor in our experiment due to the use of large objects.

Foot tapping was measured by attaching optical markers on both shoes of the participants. Touch interaction in the TABLET condition was implemented by attaching markers on participants' index and middle finger nails.

In the HMD and HPCD conditions, we used an expanded polystyrene sphere of 30 cm of diameter. Markers were attached for tracking purpose. In the HPCD condition, participant wore shutter glasses with attached markers. This provided for both active stereoscopy (at 60 Hz per eye) and head coupling.

4 METHODOLOGY

We welcomed 30 participants (12 women, 8 left handed, mean age 27.1 [18, 40]). After passing the VRMT, they were assigned to one of the 3 interaction technique groups. On the course of the 2 days they trained for 10 sessions of 3 minutes each (2(days) * 5(sessions) * 3(minutes) = 30 minutes of training) and executed 3 recognition sessions of 3 minutes. Our main focus was on the recognition trials that were designed to evaluate the effect of the interaction technique on the learning of the `sphinx`.

4.1 Measurements

Participant performance in each training and recognition session was characterized with different measurements that allowed the analysis of the speed-accuracy trade-off at different levels:

(1) The **answer time**, computed as the duration between the appearance of a target object and the tap of one of the feet, averaged over the session.
(2) The **success rate**, computed as the number of good answers divided by the total number of trials in the session.
(3) The **score** computed as follow: one point was attributed for every correct answer and one point was lost for every wrong answer in a session. This original score was used to highlight the speed/accuracy trade-off as participants who answered

slowly but with a good error rate had similar scores to those who answered more quickly but with a poorer error rate.

(4) The **amount of rotation** (for training sessions only), computed as the amount of rotation applied by participants to the target object during a trial.

Finally, we analyzed the difference in VMRT scores between the pre-test and the post-test to study the effect of the object learning on the MR ability.

4.2 Factors and hypothesis

The data were analyzed in relation with our main hypothesis and by testing the effects of the following factors on the measurements:

- The interaction **technique**, with 3 levels: TABLET, HMD, HPCD (between subjects factor);
- The recognition **session ID**, with 3 levels: day 1 end, day 2 start, and day 2 end (within-subjects factor).

We were expecting the recognition performances to follow the order: HPCD > HMD > TABLET; which would be observed on the scores, answer time and/or success rate. We were also expecting that the learning of the 3D object would have positive effect on the VMRT scores (the VMRT score improves between the pre and post tests). According to Cherdieu et al. [9], the effect of the technique may be clearer on day 2 than day 1, in particular for the recognition.

4.3 Statistics

We performed the analysis on the scores and success rates using 2 or 3 way mixed ANOVAs and pairwise t-tests with Holm correction.

For the analysis of answer time, we used a linear mixed model because of the wide differences in variability between the interaction techniques; which made the ANOVA unsuitable. With the linear mixed model (lme method from R package nlme) we could take into account variances disparity. The model takes the interaction technique and the session ID as fixed factor and the participants as random factor. We computed pairwise analysis from this model with multiple comparisons of means using Tukey contrasts (with glht method from R package multcomp, that includes correction for multiple comparisons). Finally, we produced correlation scores using Spearman's rank correlation coefficients to analyze the correlation between the answer time and the quantity of rotation performed during the training tasks. We also used Spearman's rank correlation coefficients to evaluate the correlation between the pre-VMRT scores and the scores obtained during the recognition sessions, in order to evaluate the relevance of VMRT as our balancing task.

5 RESULTS

5.1 Recognition

The Figure 4 reports the participants' success rate in recognition sessions, as well as the mean time they spent to answer to recognition trials. Results are grouped with respect to the interaction technique used during the training.

5.1.1 Success rate. We observed an effect of the session ID on the success rate ($F(2, 54) = 25.2$, $p = 1.8e-08$). The success rate improved during the experiment. However, there was no significant effect of the group ($F(2, 27) = 0.04$, $p = 0.96$), nor any interaction between the group and the session ID ($F(4, 54) = 1.51$, $p = 0.21$).

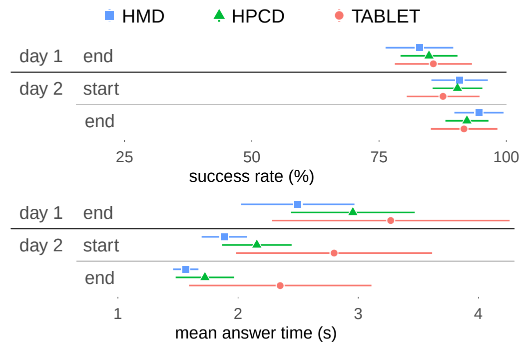

Figure 4: Success rate (top) and mean answer time (bottom) during the three recognition sessions with 95% confidence intervals, per groups.

Hence, all groups improved their success rate during the experiment, at a similar pace and reaching similar rates. The success rate was greater than 80% after the first day, and greater than 90% at the end of the experiment. We thus focus the recognition performance analysis on the answer time.

5.1.2 Answer time. Figure 4, bottom, reveals notable differences in answering time between the groups, especially during the second day of the experiment. It also reveals significant improvements of answering time during the sessions in all groups. Differences during the first day were not statistically significant. We ran a linear mixed model on day 2 data only. It confirmed the global effect of the technique on the answer time ($p = 0.013$). The interaction technique*session ID was removed from the model as its contribution was not significant ($p = 0.64$). Pairwise tests confirmed a significant difference between HMD and TABLET ($p = 0.007$). The difference between HPCD and TABLET was found to be close to significant ($p = 0.064$). The difference between HMD and HPCD was not found significant ($p = 0.23$). At the end of the second day, the average answering time with TABLET was 34% and 27% slower than with HMD and HPCD, respectively. We noticed on Figure 4, bottom, the very wide confidence intervals of the average mean answer times for TABLET when compared with the other groups. We present a possible explanation in the Discussion section.

5.1.3 Correlation between training and recognition. As expected in training experiments, participants' scores increased over the 8 training sessions with free manipulation (global effect of session: $F(7, 189) = 146.5$, $p < 2e-16$). This progression followed a similar scheme for all the techniques, with no clear technique x session ID interaction: $F(14, 189) = 1.19$, $p = 0.283$.

We computed the correlation between the average score during the training and the score in the post-training recognition session, for each participant and each day. We found a Spearman correlation coefficient of 0.78. This indicates a positive effect of the training to learn the `sphinx`.

5.2 Mental rotation abilities

Figure 5 shows the mean VMRT scores for each condition in pre- and post- tests. As VRMT was our group-balancing test, it is no

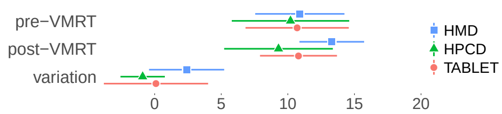

Figure 5: Mean scores of pre-VMRT and post-VMRT, and the difference between them, per group, with 95% confidence intervals.

surprise that the pre- test averages are all very similar. However, we did not find any significant effect of the group on the variation between pre- and post- tests ($F(2, 27) = 1.267$, $p = 0.298$).

Table 1 shows the Spearman correlation coefficient between participants' VMRT pre-test and participants' average recognition scores on the second day, grouped by interaction technique and all techniques combined. The general correlation coefficient (0.63, p=2e-04) indicates a positive correlation between the VMRT test and the recognition scores. We also observe that the correlation is stronger for TABLET than for the two other techniques.

5.3 Amount of rotation

During the training sessions with free manipulation (sessions 2 to 5 of each day), we measured a drop in correlation between the amount of rotation performed during a trial and the answering time from day 1 to day 2: we computed Spearman correlation coefficients of 0.71 for day 1 and 0.56 for day 2. These results indicate that the more participants proceeded through the experiment, the less they needed to rotate the target object to identify it. Toward the end of the experiment, more time may have been invested in the mental rotation of the object than in physical manipulations, suggesting an internalization of the task [10].

5.4 Subjective results

Subjective ratings are illustrated on Figure 6. HMD was the only interaction technique reported to generate some mental pain. This reflects a well-known limitation of HMDs that makes some users feel nauseous, due to the distorted vision and motion sickness. The spherical input appears to be a little more tiring to use than the tablet (the overall HMD+HPCD average rating was 2.4 ($\pm$1.57) vs. 1.6 ($\pm$0.97) for TABLET). Participants of all groups found the experiment fun; which was encouraging for student acceptance in real-world contexts. They expressed that the device was accurate enough to execute the tasks; which is consistent with the overall high success rates (c.f. Figure 4, top). Both HMD and HPCD were found very instinctive to use, but some participants reported difficulties to rotate the objects exactly how they wanted using the tablet. To

	HMD	HPCD	TABLET	All
score - VMRT correlation	0.59	0.50	0.78	0.63
p-value	0.074	0.14	**0.008**	**2e-04**

Table 1: Spearman correlation coefficient (and corresponding p-value) between participants' pre-VMRT results and their recognition score of day 2; separated per groups or all united.

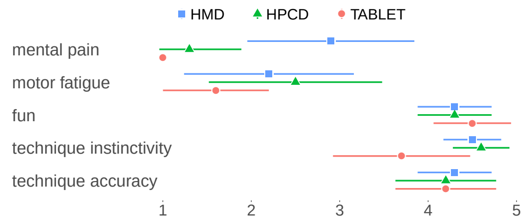

Figure 6: Subjective ratings per groups, with 95% confidence intervals.

investigate this difficulty, we processed the recorded logs of the training sessions with dictated manipulation (first training session of each day). The amount of non-optimal movements was 2.16$\times$ higher when using the tablet than when using the sphere (i.e. movements that do not turn the target object toward its reference orientation).

6 DISCUSSION

Our user study was designed to assess the effect of high fidelity 3D interaction in a 3D object learning task and on improving MR abilities. In the following, HF (High Fidelity) refers to both HMD and HPCD. Our results suggest that learning a 3D object with a HF interaction hastens its identification from different viewpoints when compared to learning the same object with a tablet interaction. This advantage was observed only when re-testing after a night of sleep.

MR abilities were more related to recognition performances in TABLET than in the HF conditions, suggesting that HF 3D interaction could reduce students' inequalities for learning.

We will now discuss the results in relation with our hypothesis, methodological aspects and new perspectives.

6.1 (HMD, HPCD) > TABLET

In the recognition sessions of day 2, the answering time was faster in the HF conditions than with TABLET. The measured performance improvement was quite large (more than 27%).

Two observations indicate that this result is robust. Firstly the average success rate with TABLET was lower than with HMD and HPCD on the last recognition session, i.e. when the improvement of answering time with the HF techniques was the largest. Secondly the display technique used to measure the recognition performances was arguably much similar to TABLET than to the HF techniques. In particular, HF participants learned the `sphinx` from head-coupled stereo renderings, but were asked to recognize it from simple perspective renderings such as the ones used in TABLET training. HF performances were thus superior to TABLET *in spite of* this experimental bias favoring the TABLET condition.

Furthermore, this indicates that any kind of mental representation of the `sphinx` that HF participants were able to build was robust to a change of appearance. This is an encouraging sign that HF interactions are worth being tested in ecological situation, e.g. training medical students on high fidelity virtual renderings of the organs. It is worth mentioning that the current availabilities of the two HF techniques that we tested are very dissimilar. HF VR is achievable today with a modest budget and plenty of readily available software such as Unity. Creating a similar level of fidelity in AR required a much larger budget and an important ad-hoc development effort. However, the current status of HF VR and AR will certainly evolve. For instance by reducing the mental pain and isolation from surroundings associated with VR's opaque HMDs, or by reducing the cost of HF HPCDs using OLED displays and inertial tracking. In any cases, our study focused on the fundamental usability of the various approaches that we tested. It indicates that it is worth pursuing both AR and VR approaches in a context of 3D object learning, as they bring a sizable benefit compared to the lower fidelity interaction.

6.2 Object Learning and MR abilities

6.2.1 Learning with low mental rotation abilities. Recognition scores of TABLET participants are highly correlated to their MR abilities as measured by the VRMT pre-test; resulting in a high variability of TABLET recognition scores. The correlation coefficients of the two other groups, however, suggest that HF training efficiency is less dependent on users' MR abilities. This could explain the differences in variability observed in Figure 4, bottom. It may also reveal different mental strategies in 3D object recognition according to the device used during learning, in agreement with embodied or grounded cognition theories [1, 11]. In particular, we hypothesized that the high fidelity reproduction of a physical object, offered by HF techniques, helps to create mental representation of the object, even for individuals with low MR abilities, by involving more "natural" sensorimotor support. Hence, using high fidelity 3D could facilitate object learning for people with poor MR abilities, and thus contribute to the reduction of an inequality in scientific courses.

6.2.2 VMRT as balancing task for between subjects designs in 3D object learning. As a between subjects design was required due to our test-retest protocol, we had to ensure that groups were balanced in regards to their ability to recognize 3D objects. We do not know of any test that predicts performances on such a complete cognitive task, hence using the VMRT was our best effort.

The correlations observed between the VMRT scores and recognition scores when all the training conditions are grouped together (c.f. Table 1, last column) indicate that mental rotation was indeed involved in recognition and that VMRT was a relevant measure to balance groups on 3D objects recognition. Other abilities might yet be involved, especially with HF, which could explain the lower correlation observed with HPCD and HMD.

6.2.3 Improvement of MR abilities. We did not find any evidence of an improvement of MR abilities, as measured using the VMRT. We chose the VMRT to measure participants' MR ability because it is widely used in the literature. However, the test does not appear to be sensitive enough to isolate an effect on such a short period. Indeed, we observed a large variability of participants' MR evolution regardless of the group. Some participants reported that the test was cognitively exhausting; which was even more problematic on the post-test that followed an already demanding experiment. In addition, our experiment confirmed that the variability of MR abilities across people is very large: we measured VMRT scores from 0 to 24 (the total range) in the pre-test. In the literature, studies using the VMRT to show an effect of different practices on MR abilities employ large cohorts (e.g. 200+ students of a class), and/or run over long training period (several weeks to one semester) [13, 18, 28].

6.3 On running multi-day controlled experiments

Studies on learning in the educational field typically span several days, for example when investigating the role of gesture in different memory tasks [7]. In a study from Cook et al. the significant improvement on learning could be observed only several weeks after the learning session [12]. Multi-day controlled (laboratory) experiments are less common in the HCI field: the focus is often on motor control and low level perception processes that in theory do not rely on users' higher level cognitive abilities. But even motor control tasks can induce high-level cognitive adaptation processes, as shown by Cattan et al. in a long-term experiment of target tracking [6]. In our experiment, recognition results in all groups where mingled at the end of the first day. Had our experiment lasted only one day, it would not have been possible to isolate any effect of the interaction technique. The progress of recognition performances between day 1 end and day 2 start is striking: even though participants did nothing related to the experiment in between, their performances increased sensibly in all conditions. This observation calls for HCI studies that span more than one day, especially when the mobilization of high-level cognitive processes is expected.

6.4 Limitations of the study

6.4.1 Generalization to ecological contexts. The `sphinx` was designed after the objects used in the VMRT; which has been shown to significantly correlates to general visuospatial abilities and in particular to success in anatomical exams [15]. For the experimental task, we took inspiration from a frequent task for medical and STEM students: identifying organs or mechanical parts from line art drawings (i.e. objects that have no color nor texture) and from a static viewpoint. However, we only tested a single, non-ecological, object due to the constraints on the experimental protocol and participants' time and cognitive efforts. Further studies are thus required to test the generality of the experiment's results.

6.4.2 Effect of a night of sleep. We attributed the score improvement between the 2 first recognition sessions to the consolidation of memory during the sleep, a well-studied phenomenon in cognitive science. However the effect could also be related to a test-retest effect. A control group passing the test and re-test the same day would be required to conclude.

7 CONCLUSION

We presented the first study that formally measured the effect of high fidelity vs. low fidelity 3D interaction on the *learning outcome* of a task of 3D object learning. Studying such a high-level cognitive process required the design and implementation of a complex experimental protocol. It revealed a sizeable improvement in recognition performances with head-coupled stereoscopy and isomorphic control compared to an interaction with a standard perspective and multi-touch control. This work contributes to a better understanding of HF 3D interaction by providing a novel proven benefit. In addition, this study should create a strong incentive to pursue the use of both VR and AR approaches in medical and STEM applied learning studies.

Acknowledgements: This work was supported by the An@tomy 2020 project ANR-16-CE38-0011.

REFERENCES

[1] Lawrence W. Barsalou. 2008. Grounded Cognition. *Annual Review of Psychology* 59, 1 (jan 2008), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

[2] Francois Berard and Thibault Louis. 2017. The Object Inside: Assessing 3D Examination with a Spherical Handheld Perspective-Corrected Display. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, Denver, CO, 4396–4404. https://doi.org/10.1145/3025453.3025806

[3] François Bérard and Amélie Rochet-Capellan. 2012. Measuring the linear and rotational user precision in touch pointing, In ACM International Conference on Interactive Tabletops and Surfaces (ITS) (Cambridge, Massachusetts, USA). *ITS*, 183–192. https://doi.org/10.1145/2396636.2396664

[4] François Bérard and Amélie Rochet-Capellan. 2015. The Transfer of Learning As HCI Similarity: Towards an Objective Assessment of the Sensory-Motor Basis of Naturalness, In ACM Conference on Human Factors in Computing Systems (CHI) (Seoul, Republic of Korea). *CHI*, 1315–1324. https://doi.org/10.1145/2702123.2702359

[5] Sandra Berney, Mireille Bétrancourt, Gaëlle Molinari, and Nady Hoyek. 2015. How spatial abilities and dynamic visualizations interplay when learning functional anatomy with 3D anatomical models. *Anatomical Sciences Education* 8, 5 (feb 2015), 452–462. https://doi.org/10.1002/ase.1524

[6] Elie Cattan, Amélie Rochet-Capellan, Pascal Perrier, and François Bérard. 2017. Does Practice Make Perfect?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, Denver, CO, 5619–5629. https://doi.org/10.1145/3025453.3025585

[7] Kuo-Jen Chao, Hui-Wen Huang, Wei-Chieh Fang, and Nian-Shing Chen. 2013. Embodied play to learn: Exploring Kinect-facilitated memory performance. *British Journal of Educational Technology* 44, 5 (aug 2013), E151–E155. https://doi.org/10.1111/bjet.12018

[8] Yi-Chen Chen, Hung-Lin Chi, Wei-Han Hung, and Shih-Chung Kang. 2011. Use of Tangible and Augmented Reality Models in Engineering Graphics Courses. *Journal of Professional Issues in Engineering Education and Practice* 137, 4 (oct 2011), 267–276. https://doi.org/10.1061/(asce)ei.1943-5541.0000078

[9] Mélaine Cherdieu, Olivier Palombi, Silvain Gerber, Jocelyne Troccaz, and Amélie Rochet-Capellan. 2017. Make Gestures to Learn: Reproducing Gestures Improves the Learning of Anatomical Knowledge More than Just Seeing Gestures. *Frontiers in Psychology* 8 (oct 2017), 1689. https://doi.org/10.3389/fpsyg.2017.01689

[10] Mingyuan Chu and Sotaro Kita. 2008. Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General* 137, 4 (2008), 706–723. https://doi.org/10.1037/a0013157

[11] Andy Clark. 1998. *Being there: Putting brain, body, and world together again*. MIT press.

[12] Susan Wagner Cook, Zachary Mitchell, and Susan Goldin-Meadow. 2008. Gesturing makes learning last. *Cognition* 106, 2 (feb 2008), 1047–1058. https://doi.org/10.1016/j.cognition.2007.04.010

[13] Andreas Dünser, Karin Steinbügl, Hannes Kaufmann, and Judith Glück. 2006. Virtual and augmented reality as spatial ability training tools. In *Proceedings of the 6th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction design centered HCI - CHINZ '06*. ACM Press, Christchurch, New Zealand, 125–132. https://doi.org/10.1145/1152760.1152776

[14] Yves Guiard. 1987. Asymmetric Division of Labor in Human Skilled Bimanual Action. *Journal of Motor Behavior* 19, 4 (dec 1987), 486–517. https://doi.org/10.1080/00222895.1987.10735426

[15] Aymeric Guillot, Stéphane Champely, Christophe Batier, Patrice Thiriet, and Christian Collet. 2006. Relationship Between Spatial Abilities, Mental Rotation and Functional Anatomy Learning. *Advances in Health Sciences Education* 12, 4 (jul 2006), 491–507. https://doi.org/10.1007/s10459-006-9021-7

[16] Karin L. Harman, G. Keith Humphrey, and Melvyn A. Goodale. 1999. Active manual control of object views facilitates visual recognition. *Current Biology* 9, 22 (2016/08/30 1999), 1315–1318. https://doi.org/10.1016/S0960-9822(00)80053-6

[17] Ken Hinckley, Joe Tullio, Randy Pausch, Dennis Proffitt, and Neal Kassell. 1997. Usability analysis of 3D rotation techniques, In ACM Symposium on User Interface Software and Technology (UIST) (Banff, Alberta, Canada). *UIST*, 1–10.

[18] Nady Hoyek, Christian Collet, Olivier Rastello, Patrick Fargier, Patrice Thiriet, and Aymeric Guillot. 2009. Enhancement of Mental Rotation Abilities and Its Effect on Anatomy Learning. *Teaching and Learning in Medicine* 21, 3 (jul 2009), 201–206. https://doi.org/10.1080/10401330903014178

[19] Susan Jang, Jonathan M. Vitale, Robert W. Jyung, and John B. Black. 2017. Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education* 106 (2017), 150 – 165. https://doi.org/10.1016/j.compedu.2016.12.009

[20] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. 2017. Augmented reality versus virtual reality for 3d object manipulation. *IEEE transactions on visualization and computer graphics* 24, 2 (2017), 1038–1048.

[21] Elinda Ai-Lim Lee and Kok Wai Wong. 2014. Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education* 79 (oct 2014), 49–58. https://doi.org/10.1016/j.compedu.2014.07.010

[22] Elinda Ai-Lim Lee, Kok Wai Wong, and Chun Che Fung. 2009. Educational Values of Virtual Reality: The Case of Spatial Ability. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* 3, 6 (2009), 934 – 938. http://waset.org/Publications?p=30

[23] Thibault Louis and Francois Berard. 2017. Superiority of a Handheld Perspective-Coupled Display in Isomorphic Docking Performances. In *Proceedings of the Interactive Surfaces and Spaces on ZZZ - ISS '17*. ACM Press, Brighton, United Kingdom, 72–81. https://doi.org/10.1145/3132272.3134124

[24] Guido Makransky, Thomas S. Terkildsen, and Richard E. Mayer. 2019. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction* 60 (2019), 225 – 236.

[25] Frank Meijer and Egon L. van den Broek. 2010. Representing 3D virtual objects: Interaction between visuo-spatial ability and type of exploration. *Vision Research* 50, 6 (mar 2010), 630–635. https://doi.org/10.1016/j.visres.2010.01.016

[26] Ghulam Muhammad, Muhammad Hussain, Muneer Al-Hammadi, Hatim Aboalsamh, Hassan Mathkour, and Amir Saeed Malik. 2016. Short-term and long-term memory analysis of learning using 2D and 3D educational contents. *Behaviour & Information Technology* 35, 11 (2016), 958–967. https://doi.org/10.1080/0144929X.2016.1212094 arXiv:https://doi.org/10.1080/0144929X.2016.1212094

[27] Jocelyn Parong and Richard E. Mayer. 2018. Learning science in immersive virtual reality. *Journal of Educational Psychology* 110, 6 (2018), 785–797.

[28] David Passig and Sigal Eden. 2001. Virtual Reality as a Tool for Improving Spatial Rotation among Deaf and Hard-of-Hearing Children. *CyberPsychology & Behavior* 4, 6 (dec 2001), 681–686. https://doi.org/10.1089/109493101753376623

[29] Kevin Rochford. 1985. Spatial learning disabilities and underachievement among university anatomy students. *Medical Education* 19, 1 (jan 1985), 13–26. https://doi.org/10.1111/j.1365-2923.1985.tb01134.x

[30] Marc A. Schnabel and Thomas Kvan. 2003. Spatial Understanding in Immersive Virtual Environments. *International Journal of Architectural Computing* 1, 4 (dec 2003), 435–448. https://doi.org/10.1260/147807703773633455

[31] Do Hyoung Shin, Phillip S. Dunston, and Xiangyu Wang. 2005. View changes in augmented reality computer-aided-drawing. *ACM Transactions on Applied Perception* 2, 1 (jan 2005), 1–14. https://doi.org/10.1145/1048687.1048688

[32] Ian Stavness, Billy Lam, and Sidney Fels. 2010. pCubee: a perspective-corrected handheld cubic display. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, Atlanta, GA, 1381–1390. https://doi.org/10.1145/1753326.1753535

[33] John Sweller. 2011. CHAPTER TWO - Cognitive Load Theory. Psychology of Learning and Motivation, Vol. 55. Academic Press, 37 – 76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

[34] Robert J Teather, Andriy Pavlovych, Wolfgang Stuerzlinger, and I Scott MacKenzie. 2009. Effects of tracking technology, latency, and spatial jitter on object movement. In *Proceedings of the IEEE Symposium on 3D User Interfaces - 3DUI*. 43–50. https://doi.org/10.1109/3DUI.2009.4811204

[35] Steven G. Vandenberg and Allan R. Kuse. 1978. Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills* 47, 2 (dec 1978), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

[36] Jonathan Wai, David Lubinski, and Camilla P. Benbow. 2009. Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology* 101, 4 (2009), 817. https://doi.org/10.1037/a0016127

[37] Nong Ye, Prashant Banerjee, Amarnath Banerjee, and Fred Dech. 1999. A comparative study of assembly planning in traditional and virtual environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 29, 4 (1999), 546–555.