International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2024)
S. Hasegawa, N. Sakata and V. Sundstedt (Editors)

Do we study an archaeological artifact differently in VR and in reality?

Maxime Dumonteil[1], Valérie Gouranton[2], Marc Macé[3], Théophane Nicolas[4], and Ronan Gaugne[1]

[1] Univ Rennes, Inria, CNRS, IRISA, France
[2] Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France
[3] CNRS, Univ Rennes, Inria, IRISA, France
[4] Inrap, UMR Trajectoires, France

Abstract

The use of virtual reality (VR) in archaeological research is increasing year over year. Nevertheless, the influence of VR technologies on researchers' perception and interpretation is frequently overlooked. These device-induced biases require careful consideration and mitigation strategies to ensure the integrity and reliability of archaeological research results. Our aim is to identify potential interpretation biases introduced by the use of VR tools in this field, by analyzing both eye-tracking patterns and participant's behavior. We have designed an experimental protocol for a user study involving an analysis task on a corpus of archaeological artifacts across different modalities: a real environment and two virtual environments, one using a head-mounted device and the other an immersive room. The aim of this experiment is to compare participants' behavior (head movements, gaze patterns and task performance) between the three modalities. The main contribution of this work is to design a methodology to generate comparable and consistent results between the data recorded during the experiment in the three different contexts. The results highlight a number of points to watch out when using VR in archaeology for analysis and interpretive purposes.

Keywords: Virtual Reality, Eye tracking, Gaze behavior, Immersive room, HMD, Archaeology

1. Introduction

Archaeology is an inherently visual discipline, based on the observation of archaeological material [Opg21]. This core visualization activity has been extended to different techniques of representation of the original material, onto advanced 3D representations in virtual reality (VR). Many studies have examined the positive and negative aspects of 3D representations, especially when they incorporate reconstructed elements based on interpretations [FDC*19]. An explicit distinction is made between knowledge and 3D knowledge [Huv18], mainly because 3D models are produced with an intention, even in the case of digitization-based models. However, Molloy & Milić [MM18] advocate that 3D representations of archaeological material produced via faithful digitization techniques are helpful for the research activity as they " can be rotated and manipulated by a viewer, [...] compare to what was previously possible with only photographs and illustrations.".

In recent years, the integration of VR technology and 3D data within archaeological research has significantly advanced the field [GBL*22, PSA22, HBD*24]. This convergence has enabled archaeologists to employ innovative methodologies for interacting with artifacts in controlled virtual environments, thereby reducing the risk of physical damage. Consequently, conventional methodologies employed in archaeological investigation and interpretation are undergoing a significant transformation. However, this paradigm shift is not without challenges. One such challenge arises from the potential for bias introduced by the use of VR equipment such as head-mounted displays (HMD). While HMDs are useful for facilitating immersive experiences, they may inadvertently influence researchers' perceptions and interpretations, which could affect the integrity of archaeological analysis. Consequently, while VR has great potential for enhancing archaeological exploration, the implications of device-induced biases require careful consideration and the implementation of mitigation strategies to ensure the integrity and reliability of research outcomes.

Accordingly, we designed an experiment with the objective of discerning potential biases induced by the use of VR systems. As visual processing itself constitutes a significant portion of human cognitive processing, it is crucial to recognize its intricate nature, encompassing various types of ocular movements such as gaze fixation (focusing the gaze on a point with precision) or saccades (brief and fast movement of the eyes between two stable positions). While these movements serve as approximations of general ocular behavior, they also offer insights into attentional processes during visual perception [Car11]. In addition, eye-tracking is increasingly used

DOI: 10.2312/egve.20241371

in VR headsets, enabling many new applications to be developed in virtual contexts [Kow11].

The experimental protocol for the user study involves the analysis of a corpus of archaeological artifacts across three modalities: a real environment and two distinct virtual environments (a head-mounted VR device and an immersive room). The aim of this experiment is to compare participants' behavior (head movements, gaze patterns and task performance) between the three modalities.

2. Related Works

Some researches tend to develop archaeological interpretations based on the analysis of visual behavior. The work presented in [CBAPB19] seeks to demonstrate that our visual reactions and artifacts have evolved together to reflect changes in the social and cultural complexity of the corresponding societies, as our brains have evolved to develop selective attention. Another study [SGIF*22] about stone tools suggested that visual exploration induced by observed objects is influenced by their affordance and salience.

However, these two observations are based on photographic representations of the objects studied. Eye-tracking tools and software consider most of the time 2D material as a visualization support while more and more use cases need a 3D environment to support ecological observation and manipulation of the objects. Therefore some works such as [TSC*18] propose methods to map eye-tracking data on a 3D object and provide data models to visualize the gaze behavior in this context.

In addition, some studies tackle the question of behavioral differences between a real situation and a similar one in a virtual context, such as in [TIMC21], or in [MCM18] where eye-tracking is associated to EMG, to measure muscular activity between real and virtual environments. Accordingly, some bias could be identified in the use of VR tools compared to a similar experience but the use cases are generic and can hardly be generalized to an archaeological context.

Another experiment compared the eye movements of museum visitors in a real room and in its copy in virtual reality [GDZ21]. Their analysis revealed similarities in exploration patterns between the two conditions, focusing on eye fixations, and concluding that virtual environments are valid counterparts for this use case.

A related study [RFSK19] proposes a task where 6 artifacts are analyzed over 3D-printed copies of these objects, 3D models presented in optical see-through augmented reality, and 3D models presented on a 2D screen. Eye-tracking did not reveal any significant difference between the three conditions. This study illustrates the need for a thorough examination of the technological factors that can influence the interpretation of historical and cultural narratives in archaeological research through various modalities.

To conclude, the precise characterization of the difference in behavior induced by the VR environment compared to natural stimuli may help to understand the impact of immersive technology on perceptual processes in an archaeological context. In particular, this can identify the potential interpretation biases introduced by the use of virtual reality tools in this field.

3. Methods

The aim of the experiment is to study a person's behavior, and in particular their oculomotor behavior, during a specific archaeological task, carried out using different modalities. The task involves examining the shapes and ornamentation of potteries, a process that relies primarily on visual engagement and is usually carried out in laboratory environments. This task involves meticulous examination of artifacts, often requiring manual manipulation to facilitate access to fine or complex details.

A crucial aspect to consider is the design of a methodology to generate comparable and consistent results between the data collected during the experiment in the three distinct contexts.

3.1. Modalities

We consider three different modalities, with the aim of comparing how the same archaeological activity is carried out in each of them:

- Real modality: the task is performed on reproductions of the real artifacts;
- HMD modality: the task is performed in VR, on a digital 3D copy of the artifact, wearing a HMD;
- Immersive modality: the task is performed in VR, on a digital 3D copy of the artifact, in a large CAVE-like immersive room.

To gain a deeper understanding of user behavior in virtual environments, we considered both HMD and immersive modalities, as user behavior may differ slightly between the two. In the case of an immersive room, users are present with their own physical bodies and can therefore interact in a more natural way, which is particularly interesting for users who are unfamiliar with VR, as is still often the case for archaeologists. Nevertheless, given that HMDs are now commonly used for VR, thanks to their low price and the accessibility of software tools for developing 3D interactive environments, we also included this modality in our study.

3.2. Corpus

The corpus under consideration is composed of three distinct styles of pottery (Fig. 1). The 3 artifacts were selected to illustrate a variability of style and diversity in size, shape and texture (Tab. 1). To prevent damaging the original material, facsimiles of the potteries were used during the experiment. They were created by potters using traditional techniques. This selection of artifacts can induce a wide set of gaze patterns for each pottery.

Id	origin	dimensions	shape and texture
1	Neolithic France	11 cm high 12,5 cm diam.	Spherical shape with 2 tabs and motifs engraved on the outside
2	Antiquity North Africa	15.5 cm high 7.5 cm diam.	Jug with one handle and painted motifs
3	Bronze Age France	14 cm high 19 cm diam.	Bell shape without ornaments and one perforated knob

Table 1: *Main characteristics of the corpus*

Figure 1: *Corpus of copies of archaeological potteries (top) and their virtual counterparts (bottom)*

Figure 2: *Top, left: Real modality; Top, right: HMD modality; Bottom: Immersive modality (CAVE-like immersive room)*

Figure 3: *Eye-tracking glasses and stereoscopic glasses attached together with a 3D-printed clip*

3.3. Participant

A total of 18 participants were recruited on a voluntary basis, aged between 19 and 42 years ($M = 26.22, SD = 7.1$) with 8 females and 10 males. No requirements regarding the level of knowledge in VR and in archaeology were specified, nevertheless half of the participants often use virtual reality, three, just sometimes, five, only one time, and one never tried VR. Two participants were archaeologists.

3.4. Apparatus

The experiment was conducted using three distinct setups, each corresponding to one of the three modalities under investigation. For the Real modality, the participant stood in front of the artifact, which was placed on a turntable attached to a tripod (Fig. 2, top, left). They were able to turn the table and rotate it. The participant wore Tobii Pro Glasses 2, a portable eye-tracking system that can also record a video of the observed scene. A computer was placed near the participant to answer questions related to the experiment. The location of the screen was chosen so as to clearly distinguish the gaze on the artifact from that on the monitor.

The environment presented to the user in the other two modalities has been made as similar as possible to the real environment. For the HMD modality, participants stood and wore an HTC Vive Pro Eye VR headset (Fig. 2, top, right). As in the Real modality, they were allowed to move around the table and rotate the turntable using a controller. This HMD incorporates a Tobii eye-tracking system. The subject was required to use a single button to click on the virtual interface (by pointing with a beam) or to grasp and rotate the turntable, by pressing the button.

For the Immersive modality, participants stood in Immersia (www.irisa.fr/immersia), a large CAVE-like installation. This consisted of a 4-sided viewing screen measuring 10 m x 3 m x 3 m (width, depth and height) equipped with 14 WQXGA projectors (Fig. 2, bottom). Subjects' position and head movements were tracked using passive markers detected by an Optitrack optical tracking system with 16 cameras at a 120 Hz update rate. They wore Vulfoni stereoscopic glasses attached to Tobii eye-tracking goggles with a 3D-printed clip (Fig. 3). The subject interacted with the virtual environment using a wireless controller, also using a single button, located in the same place as in the HMD modality on the controller, in order to maintain interactions as similar as possible.

3.5. Protocol

The experimental protocol was approved by the local ethical committee. It was designed to perform the same task successively in

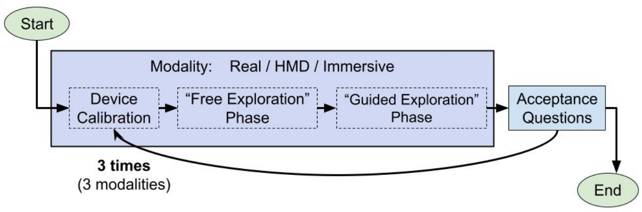

Figure 4: *Global structure of the experimental protocol*

Id	Question	Answers
0	Object color?	Black ; Wheat; Orange
1	Object condition?	Complete; Incomplete
2	Object quality?	Thin; Half rough ; Rough
3	Object size?	Wider than high; Higher than wide ; As wide as high ; Low and wide
4	Object shape?	Spherical; Conical ; Cylindrical ; Ovoid
5	Object decoration?	Paint; Incision ; Adding matter
6	Object profile analysis?	Not segmented (straight line); Mono segmented (curved line); Poly segmented (inflexion points)

Table 2: *Analysis grid: participants answered all questions of the grid for each condition*

Id	Question
PU1	Using the system improves my performances in my job.
PU2	Using the system in my job increases my productivity.
PU3	Using the system enhances my effectiveness in my job.
PU4	I find the system to be useful in my job.
PEOU1	My interaction with the system is clear and understandable.
PEOU2	Interacting with the system does not require a lot of my mental effort.
PEOU3	I find the system to be easy to use.
PEOU4	I find it easy to get the system to do what I want it to do.
OUT1	The quality of the output I get from the system is high.
OUT2	I have no problem with the quality of the system's output.
OUT3	I rate the results from the system to be excellent.
GAZE	The system does not obstruct my gaze.

Table 3: *Adapted Acceptance Questionnaire*

the three modalities, on three different artifacts, in order to compare participants' behavior under these different experimental conditions. The experiment was therefore divided into three iterations of the same task. The order of modalities and artifacts was counterbalanced to avoid introducing a learning bias.

The complete sequence of the experiment, for one participant, is shown in Fig. 4. First, each participant read and signed the informed consent form and completed a general questionnaire about previous VR experience, height, gender and age. Before each task iteration, a preparation phase involved selecting and positioning the artifact and starting the program. After placing the device associated with the current modality on the participant (eye-tracking glasses, HMD or stereoscopic glasses clipped to eye-tracking glasses), we calibrated the eye-tracking system following the standard procedure recommended by the provider.

Each iteration of the task followed the same scenario: after calibration, it began with a "free exploration" phase, followed by a "guided exploration" phase. The start of each phase was triggered by the participant clicking on a button (on a computer for the Real modality, or on a virtual interface for the other two modalities). During the "free exploration" phase, the participant had to respect a few rules specified orally:

- You cannot touch directly the artifact
- You can look at the artifact under any angle
- You can freely move around the artifact
- You can grab the turntable to rotate it horizontally

In the next phase, "guided exploration", the participant was asked to answer seven questions about artifact characteristics (Tab. 2), inspired by actual archaeological analysis grids [CDD02]. The goal of this form is to guide the user's observation of artifacts and encourage them to look at the potteries as a whole. The participant had to choose between several possible answers. Depending on the modality, the form was displayed on a virtual or real screen next to the artifact, leaving it visible. While completing the analysis grid, the participant was free to return to the artifact and view it again. Validation of the answer to the final question completed the task.

Once the task had been completed, the participant is required to fill an acceptance questionnaire (Tab. 3) based on the Technology Acceptance Model 3 (TAM3) [VB08]. The questions are organized into three distinct blocks: "Perceived Usefulness" (PU), "Perceived Ease of Use" (PEOU), and "Output Quality" (OUT). PU refers to the extent to which users believe that utilizing the technology will

Figure 5: *Augmented Reality markers used to track the participants' position relative to the object*

enhance their job performance. PEOU, on the other hand, represents the degree to which users perceive the technology as effortless to use. Finally, OUT encapsulates the users' belief that the technology will assist them in performing their tasks effectively [VB08]. All questions are presented as statements and answers are evaluated on a 7-point Likert scale (1 meaning strongly disagree and 7 strongly agree). We also added a question on the level of discomfort induced by the eye-tracking system on the field of view, using the same Likert scale (identified as GAZE).

3.6. Collected data

During the experimental part, in addition to responses to the analysis grid and questionnaires, we recorded the time of each phase, the position of the user's head and eye-tracking data: origin of the gaze (the central position of the eyeball) and gaze direction.

To record the user's head position and orientation, we used information directly available in the HMD and Immersive modalities, obtained from head tracking. For the Real modality, we placed ArUco augmented reality markers (Fig. 5) [GJMSMCMJ14b] all around the turntable and recorded video from the Tobii glasses camera to perform a pose estimation calculation during the post-processing step. The ArUco markers are also represented in the virtual scene of the HMD and Immersive modalities to maintain an identical environment for the user.

3.7. Hypotheses

Based on our method, three main hypotheses are considered:

- **H1**: The level of information that the user obtains about the artifact's characteristics does not differ according to modality.
- **H2**: The user's interactions with the environment are similar for each modality.
- **H3**: The user's gaze behavior is similar regardless of modality.

In order to evaluate the validity of hypothesis H1, a scoring system was defined based on the data obtained from the archaeological analysis grid completed during the experiment. The evaluation of hypotheses H2 and H3, respectively, is based on the analysis of user activity and oculomotor behavior. Regarding user activity, we consider the user's head movements in relation to the artifacts and the time needed to perform the task. Regarding gaze behavior, we consider eye movement characteristics, with a special focus on fixations, which can be used to infer attention or cognitive processes, and in relation to artifact areas of interest.

3.8. Data post-processing

Some of the data collected during the experiment require one or more post-processing in order to obtain standardized data and comparable measurements between the three modalities.

3.8.1. User activity

As previously stated in Section 3.7, in addition to the duration of the task, the movements of users' heads in relation to artifacts are considered as a reference for comparison between modalities. A prerequisite is to get the position of the user in the environment, and for this step, each modality has a specific processing. For HMD and Immersive modalities, the user's head position is directly provided in Unity by the corresponding tracking system (section 3.4). As mentioned in Section 3.6, the Real modality employs image processing to compute these values from the video (recorded by the camera of Tobii glasses) using ArUco markers [GJMSMCMJ14a].

To characterize users' movements, two metrics are considered: (i) the distance between the user's head and the center of the artifact (UD), and (ii) a vertical angle (VA) to measure how much the user is above the artifact while looking at it. This angle is defined as the angle between two vectors originating from the center of the artifact, one directed towards the user's head, and one vertical vector pointing upwards. As the vertical angle value approaches zero, the head is positioned above the artifact. Furthermore, as the value approaches ninety degrees, the head is situated closer to the side of the object. It should be noted that for each modality, the same process is employed, as the data are already uniform at this processing step.

3.8.2. Oculomotor behavior

Oculometric measurements are derived from the collected data through two principal stages. Initially, 3D eye-tracking data is computed, linking the gaze to the 3D environment. This entails to calculate the projection of the gaze onto the elements observed by the user. For the Real modality, to track the elements in the environment and compute the 3D intersection point of the gaze on the object, we implemented the method developed by Takahashi et al. [TSC*18]. Subsequently, data on the spatio-temporal behavior of the gaze is computed. This may include gaze fixation patterns or rapid eye movements. The aforementioned calculations are performed in an identical manner across all three modalities, using Tobii's oculometric data.

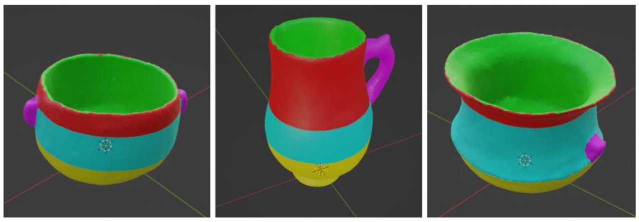

Figure 6: *Areas of Interest (AOI) for each artifact (red: top, blue: body, yellow: foot, green: inside, pink: raised parts*

Once the 3D data has been generated, it is possible to compute statistics about eye behavior. The primary focus is on examining the different types of eye movements, such as fixations and saccades and the associated metrics. Additionally, for each gaze event, the position of the gaze point and the direction of the gaze are analyzed. To detect such gaze events, we implemented an I-VDT algorithm (Velocity and Dispersion Threshold Identification) [LJMMGA20, KK13, DKV*22]. Our algorithm detects fixations across saccades using a spread threshold of 1.3° and a time threshold of 0.1s, as recommended by Llanes-Jurado [LJMMGA20].

The three gaze metrics measured in our study are : (i) duration of fixations (FD), (ii) duration of the first fixation (FFD), which provides additional information about the user's attentional focus [CL20] and (iii) frequency of fixations (FF). The last metric (FF) corresponds to the number of fixations on the artifacts divided by the time of the observation phase of the task. This metric is more representative than just the number of fixations during the task, as it does not depend on the duration of the task.

Furthermore, in order to perform a more comprehensive examination of gaze behavior, ocular events are considered in relation to areas of interest (AOIs) that correspond to a semantic breakdown of artifacts employed by ceramologists and archaeologists. To this end, five areas are defined on each artifact of the corpus and applied on their virtual copies (Fig. 6). The first area encompasses the lip and neck (i.e., the upper part of the artifact, indicated in red in the Fig. 6), while the second one encompasses the body (i.e. the central part of the object is defined as blue in Fig. 6). The lower part is the yellow area, the interior is the green area, and the pink area encompasses all raised parts, such as handles, knobs, or tabs. To delineate the upper, central, and lower portions of the artifact, the demarcations of the different areas are based on the inflection points on the object's profile. This abstraction of the structure of the artifacts allows the calculation of different metrics, such as the time spent on each part of the object, the number of visits to each AOI, and the transitions between different AOIs. To define a metric capable of identifying a classification from the least to the most visited AOI, we calculate the percentage of fixations time for each AOI. As for gaze events, the aforementioned calculations are performed in an identical manner across all three modalities, on the projection of gaze data in the 3D environment.

Proceedings published by Eurographics - The European Association for Computer Graphics.

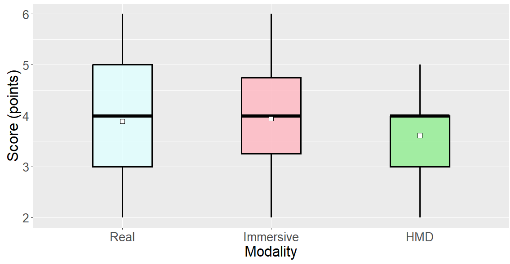

Figure 7: *Scores of the analysis grid per modality*

4. Results

4.1. Analysis grid score

We assigned a score to the analysis grid completed by each participant, counting 1 point for a correct answer, and 0 for a wrong one, according to a reference framework of answers validated by a ceramologist. The scores for each modality (Fig. 7) are respectively for Real, $m = 3.89, SD = 1.18$, Immersive, $m = 3.94, SD = 1.05$, and HMD, $m = 3.61, SD = 0.91$

A Friedman test indicates no statistically significant differences ($\chi^2 = 0.7, p = 0.71$) between the results obtained with the different modalities. Results were neither significantly different with respect to artifacts ($\chi^2 = 3.6, p = 0.17$) and iteration ($\chi^2 = 0.5, p = 0.79$).

4.2. Subjective questionnaires

To evaluate the results of each block of the TAM questionnaire, a PLS-SEM analysis was performed (using the SEMinR package with RStudio [RDV*24]) to obtain a score (from the composite score in the model) and also the average Likert scale score for each block. Results are presented in Tab. 4.

A Friedman test (made from Likert means results) indicates a statistically significant difference for each block of questions across the three modalities: PU ($\chi^2 = 13.4, p = 0.001$), PEOU ($\chi^2 = 26.9, p < 0.001$), and OUT ($\chi^2 = 21.3, p < 0.001$). Post-hoc Wilcoxon tests indicate the presence of statistically significant differences between the Immersive modality and the other two modalities, with the exception of the OUT block, for which the difference is significant between all modalities (Tab. 5). Composite scores and Likert results generate the same significant differences. The result for GAZE is also significantly different across modalities ($\chi^2 = 12.9, p = 0.001$), more specifically between Immersive modality and the others (Tab. 5).

	Real		Immersive		HMD	
PU	$m = 4.94$	$SD = 1.72$	$m = 3.69$	$SD = 1.31$	$m = 4.61$	$SD = 1.28$
PEOU	$m = 6.64$	$SD = 0.41$	$m = 5.05$	$SD = 1.04$	$m = 6.29$	$SD = 0.58$
OUT	$m = 6.09$	$SD = 1.05$	$m = 4.55$	$SD = 1.14$	$m = 5.33$	$SD = 0.93$
GAZE	$m = 2.0$	$SD = 1.33$	$m = 4.28$	$SD = 1.60$	$m = 2.89$	$SD = 1.71$

Table 4: *Likert scale results for the blocks of the TAM questionnaire*

Tests	PU	PEOU	OUT	GAZE
Real / Immersive	<0.01	<0.001	<0.001	0.003
HMD / Immersive	<0.01	<0.001	0.02	0.04
Real / HMD	0.86	0.13	0.03	0.25

Table 5: *Wilcoxon p-values with Bonferroni correction on TAM Likert blocks average scale*

	Real		Immersive		HMD	
"free" phase	$m = 49.3$	$SD = 25.1$	$m = 66.4$	$SD = 42.2$	$m = 58.9$	$SD = 38.9$
"guided" phase	$m = 83.5$	$SD = 42.5$	$m = 106.9$	$SD = 29.6$	$m = 105.4$	$SD = 46.5$
total	$m = 132.9$	$SD = 54.3$	$m = 173.2$	$SD = 45.0$	$m = 164.3$	$SD = 69.9$

Table 6: *Duration of the experiments' phases per modality*

Tests	"Free exploration" Phase Time	"Guided exploration" Phase Time	Total Time
Real / Immersive	0.14	0.08	0.08
HMD / Immersive	1.0	1.0	1.0
Real / HMD	0.39	0.22	0.09

Table 7: *Wilcoxon p-values with Bonferroni correction on "Time" metrics*

4.3. User activity

4.3.1. Task duration

The descriptive statistics regarding the duration of the experiment are presented in Tab. 6. A Friedman test indicates that the time allotted for the "free exploration" phase is not correlated with the modality ($\chi^2 = 4.11, p = 0.13$). In contrast, the "guided exploration" phase, during which participants responded to questions about the artifact's characteristics, resulted in a significantly different time between modalities ($\chi^2 = 7.44, p = 0.02$). However, Wilcoxon tests on these time metrics did not reveal any significant differences between pairs of modalities (Tab. 7).

A Friedman test indicates a significant difference ($\chi^2 = 12, p = 0.002$) in the Total Time metric in relation to the iteration order of the modalities used to perform the task. This result is also observable ($\chi^2 = 12.4, p = 0.002$) for the "guided exploration" time, but not for the "free exploration" time. A Wilcoxon post-hoc test with Bonferroni correction indicates that the difference is significant between the first iteration and the two others (resp. $p_{1,2} = 0.004$, $p_{1,3} = 0.002$) but not between the second and third iterations ($p_{2,3} = 1.0$).

4.3.2. Users' movements

The movement metrics collected during the experiment are presented per modality in Tab. 8. The results for UD and VA, calculated for each modality, are presented respectively in Fig. 8(a) and Fig. 8(b). A Friedman test indicates that both UD ($\chi^2 = 6.77, p = 0.03$) and VA ($\chi^2 = 14.33, p < 0.001$) are significantly different depending on the modality. Wilcoxon post-hoc tests with Bonferroni correction are presented in Tab. 9. They show significant differences for VA, between Real and the two VR modalities, but none for UD.

4.4. Oculomotor behavior

The three basic metrics collected during the experiment are presented per modality in Tab. 10. The results for the fixation du-

	Real		Immersive		HMD	
UD (m)	$m = 0.54$	$SD = 0.13$	$m = 0.58$	$SD = 0.16$	$m = 0.50$	$SD = 0.14$
VA (deg)	$m = 44.5$	$SD = 4.8$	$m = 51.6$	$SD = 4.8$	$m = 49.5$	$SD = 6.2$

Table 8: *Results for the users' movements metrics*

Tests	UD	VA
Real / Immersive	0.36	<0.001
HMD / Immersive	0.09	1.0
Real / HMD	1.0	0.04

Table 9: *Wilcoxon p-values with Bonferroni correction on user's movements*

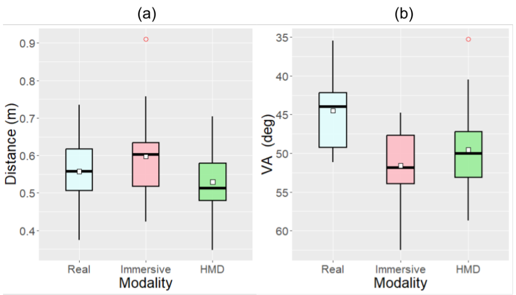

Figure 8: *User's movements metrics per modality: (a) average distance between the user's head and the artifact (UD) and (b) average vertical angle (VA) between the user and the artifact.*

ration per modality is presented in Fig. 9. A Friedman test indicates a significant difference between modalities for FD ($\chi^2 = 28.77, p < 0.001$) and for FF ($\chi^2 = 14.11, p < 0.001$), but no statistically significant difference for FFD ($\chi^2 = 3.11, p = 0.21$). The results of post-hoc Wilcoxon tests (Tab. 9) show significant differences between HMD and the other modalities for FF, and between all modalities for FD.

In order to visualize the oculomotor behavior with respect to the AOIs, we considered a straightforward abstract representation of all the artifacts where the five zones are easily identified (the fifth zone corresponding to the raised parts is represented by the little cube attached to the left side of the artifact). A classification of the different zones from least to most watched per modality is presented in Fig. 10.

A Friedman test has been applied to an encoding of the classification of the different parts of the artifacts with respect to the percentage of fixation time. The result indicates that there is no statistically significant difference ($\chi^2 = 0.2, p = 0.9$) between the watching order of the AOIs across the modalities. However it should be noted that a Friedman test indicates a significant difference ($\chi^2 = 13.7, p < 0.001$) between the watching orders across the artifacts. A post-hoc Nemenyi test indicates a significant difference between the artifacts 2 and 3 (cf Tab. 1)($p = 0.02$). The order of watching, per artifact and independently of the modality is presented in Fig. 11.

	Real		Immersive		HMD	
FD (s)	$m = 0.43$	$SD = 0.09$	$m = 0.30$	$SD = 0.07$	$m = 0.23$	$SD = 0.4$
FFD (s)	$m = 0.64$	$SD = 0.94$	$m = 0.21$	$SD = 0.16$	$m = 0.23$	$SD = 0.10$
FF	$m = 1.52$	$SD = 0.37$	$m = 1.43$	$SD = 0.38$	$m = 2.03$	$SD = 0.32$

Table 10: *Results for the eye-tracking metrics*

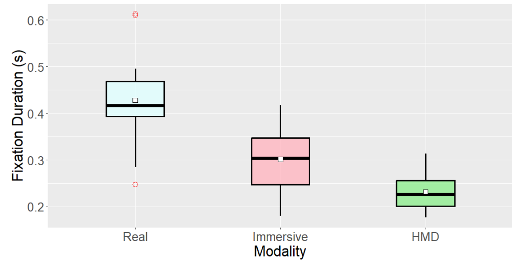

Figure 9: *Average fixation time, per modality*

Tests	FD	FF
Real / Immersive	<0.001	1.0
HMD / Immersive	0.001	<0.01
Real / HMD	<0.001	<0.001

Table 11: *Wilcoxon p-values with Bonferroni correction on Gaze metrics*

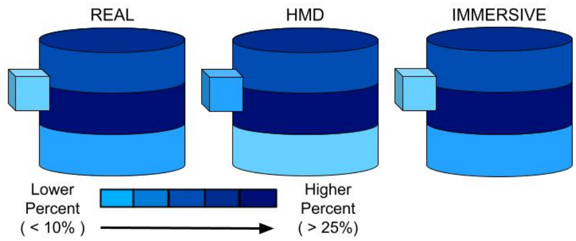

Figure 10: *Percentage of fixation time of the different AOI (per modality and considering all artifacts together)*

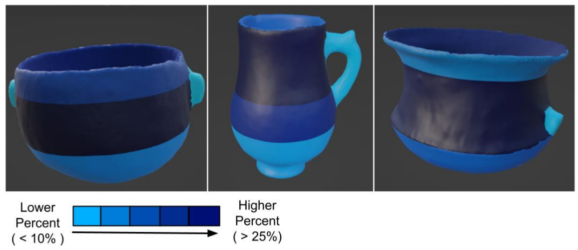

Figure 11: *Percentage of fixation time of the different AOI (per artifact, considering all modalities of exploration)*

5. Discussion

Our first hypothesis posits that the level of information obtained by the user about the artifact's characteristics does not differ across the modalities. The results on the analysis grid **validate H1** as there is no statistically significant difference between the three modalities. Additionally, this indicates that the corpus does not introduce a bias due to a particular attribute of a specific pottery type. The same observation with respect to the iteration of the task indicates an absence of learning bias.

Additionally, the results on gaze behavior appear to **validate H3**, since there is no difference in the classification from the least watched to the most watched of the AOIs, even if we have detected differences in some specific oculomotor behaviors based on the duration and quantity of fixations. However, it is very interesting to note that similarly to [GDZ21], we observed a decrease in average fixation time in HMD compared to the other modalities, as well as a higher number of fixations. In HMD modality, participants look more at the artifact (higher frequency of fixations) but for a shorter duration (lower average fixation time).

Another interesting observation regarding gaze behavior is the significant difference in the order of visiting the AOIs of the various artifacts. This strengthens the validity of the corpus, as the different characteristics of the vases induce different viewing behaviors.

The user behavior considered in our experiment differs significantly between the different modalities, thus **rejecting H2**. However, the magnitude of the differences for the different metrics allows to nuance this negative result. In fact, the difference of the average distance UD between the modalities is only a few centimeters (4 cm between Real and HMD and also between Real and Immersive), and the average difference of the vertical angle VA between the modalities is only a few degrees (5° between Real and HMD and 7° between Real and Immersive).

Regarding the time spent in the task, there are no significant differences between the modalities for the "free exploration" phase, which corresponds to the main observation phase. The significant difference in the total time of the task directly depends on the "guided exploration" phase, which was mainly used by the participants to fill in the analysis grid with few additional views of the artifacts. This difference can be explained by some difficulties in the virtual modalities, e.g. related to the use of interaction devices to fill in the analysis grid, especially in the immersive space, which caused a waste of time. We also measured an important learning effect between the first iteration of the task and the two following ones, since the analysis grid was always the same. This learning effect directly impacted the duration of the task (the total time decreased by 30.5 s in average between the first and the 2 following iterations), as users are already familiar with the questions presented in the analysis grid. However, it does not influence the analysis score, as discussed above.

The results of the TAM3 questionnaire indicate a markedly diminished level of appreciation for the immersive modality. Furthermore, informal feedback from participants at the conclusion of the experiment corroborates these findings. Similarly, visual comfort, as gauged by the GAZE score, exhibited a comparable trend. The discrepancy in results may be attributed to several factors. Primarily, the immersive room's size and level of technical sophistication may appear disproportionate to the task at hand, potentially eliciting negative appraisal regarding the use of this system in a business context. Additionally, the attachment method for the eye-tracking glasses to the stereoscopic glasses is somewhat weighty, and results in the stereoscopic glasses being displaced from the eyes, reducing the field of view. On the other hand, the level of acceptance, usability and visual comfort did not differ significantly between the Real and HMD modalities. This suggests that the use of 3D digital copies with virtual reality headsets could be an adequate tool for the study of artifacts by archaeologists. The archaeologists who participated gave positive feedback, particularly regarding the safe manipulation and interaction with the artifact permitted in VR.

6. Conclusions and future work

The aim was to investigate whether the use of VR tools introduces interpretation biases in the visual study of archaeological artifacts. To this end, an experimental protocol for a user study was designed. It involves an analysis task on a corpus of archaeological artifacts across different modalities: a real environment and two virtual environments, one using a head-mounted device and the other an immersive room. A methodology was designed to process comparable and consistent results between data recorded in the three contexts, with different devices. The comparative analysis is based on both eye-tracking patterns and participant's behavior.

The findings suggest a consistency in user behavior and performance across the modalities, despite the observed variability in the characteristics of the potteries within the corpus. In this way, we were able to observe both that participants looked at the vases in the corpus in significantly different ways, and that they observed them in similar ways across modalities, ultimately obtaining similar analysis scores. This suggests that VR can be employed in an archaeological context, without being a source of potential biases when analyzing artifacts inasmuch as the virtual elements are rendered with a quality similar or higher than used here. No significant differences were observed in the time spent looking at the objects or in the user's ability to retrieve information about the artifact's attributes in response to questions about the artifact's characteristics. The statistically significant discrepancies in user movements can be mitigated by their small amplitude. Upon examination of eye-tracking data, no distinction was discerned in the global gaze pattern, as determined by the AOIs' classification of exploration. However, discernible differences were observed in specific oculomotor behaviors, particularly in terms of fixation duration and quantity in HMD compared to reality. These findings regarding ocular behavior are aligned with those reported in existing literature [GDZ21].

Further research is required to encompass the activities of archaeologists in a more comprehensive manner, with a particular focus on other types of artifacts or archaeological materials. From a methodological perspective, it would be valuable to consider additional behavioral indicators, such as brain, body, and physiological activity. With regard to the measurement of oculomotor activity, the available data analysis tools are primarily designed for 2D images. The analysis of gaze in 3D environments could benefit from the development of more suitable measurement methods and tools.

Acknowledgments

This work was partially funded by Equipex+ Continuum ANR-21-ESRE-0030 and CNRS MITI AAP Digital Twin AJAX.

References

[Car11] CARRASCO M.: Visual attention: The past 25 years. *Vision research 51*, 13 (2011), 1484–1525. 1

[CBAPB19] CRIADO-BOADO F., ALONSO-PABLOS D., BLANCO, ET AL. M. J.: Coevolution of visual behaviour, the material world and social complexity, depicted by the eye-tracking of archaeological objects in humans. *Scientific reports 9*, 1 (2019), 3985. 2

[CDD02] CAULIEZ J., DELAUNAY G., DUPLAN V.: Nomenclature et méthode de description pour l'étude des céramiques de la fin du Néolithique en Provence. *Préhistoires Méditerranéennes*, 10-11 (Sept. 2002). 4

[CL20] CARTER B. T., LUKE S. G.: Best practices in eye tracking research. *International Journal of Psychophysiology 155* (2020), 49–62. 5

[DKV*22] DUCHOWSKI A. T., KREJTZ K., VOLONTE M., HUGHES C. J., BRESCIA-ZAPATA M., ORERO P.: 3d gaze in virtual reality: vergence, calibration, event detection. *Procedia Computer Science 207* (2022), 1641–1648. 5

[FDC*19] FERDANI D., DEMETRESCU E., CAVALIERI M., PACE G., LENZI S.: 3d modelling and visualization in field archaeology. from survey to interpretation of the past using digital technologies. *GROMA: Documenting Archaeology 4* (2019). 1

[GBL*22] GAUGNE R., BARREAU J.-B., LÉCUYER F., NICOLAS T., NORMAND J.-M., GOURANTON V.: extended reality for cultural heritage. In *Handbook of Cultural Heritage Analysis*. Springer, 2022, pp. 1405–1437. 1

[GDZ21] GULHAN D., DURANT S., ZANKER J. M.: Similarity of gaze patterns across physical and virtual versions of an installation artwork. *Scientific Reports 11*, 1 (2021), 18913. 2, 8

[GJMSMCMJ14a] GARRIDO-JURADO S., MUÑOZ-SALINAS R., MADRID-CUEVAS F. J., MARÍN-JIMÉNEZ M. J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition 47*, 6 (2014), 2280–2292. 5

[GJMSMCMJ14b] GARRIDO JURADO S., MUÑOZ SALINAS R., MADRID CUEVAS F. J., MARÍN JIMÉNEZ M. J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition 47*, 6 (June 2014), 2280–2292. 4

[HBD*24] HOSTETTLER M., BUHLKE A., DRUMMER C., EMMENEGGER L., REICH J., STÄHELI C.: Digital archaeology between hype and reality: The results of a survey on the use of 3d technologies in archaeology. In *The 3 Dimensions of Digitalised Archaeology: State-of-the-Art, Data Management and Current Challenges in Archaeological 3D-Documentation*. Springer International Publishing Cham, 2024, pp. 181–203. 1

[Huv18] HUVILA I.: The subtle difference between knowledge and 3d knowledge. *Hamburger Journal für Kulturanthropologie*, 7 (2018), 99–111. 1

[KK13] KOMOGORTSEV O. V., KARPOV A.: Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior Research Methods 45*, 1 (Mar. 2013), 203–215. 5

[Kow11] KOWLER E.: Eye movements: The past 25 years. *Vision research 51*, 13 (2011), 1457–1483. 2

[LJMMGA20] LLANES-JURADO J., MARÍN-MORALES J., GUIXERES J., ALCAÑIZ M.: Development and Calibration of an Eye-Tracking Fixation Identification Algorithm for Immersive Virtual Reality. *Sensors 20*, 17 (Jan. 2020), 4956. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute. 5

[MCM18] MIRZAEI M. A., CHARDONNET J.-R., MERIENNE F.: Comparing real and virtual object manipulation by physiological signals analysis: A first study. 2

[MM18] MOLLOY B., MILIĆ M.: Wonderful things? a consideration of 3d modelling of objects in material culture research. *Open Archaeology 4*, 1 (2018), 97–113. 1

[Opg21] OPGENHAFFEN L.: Visualizing archaeologists: a reflexive history of visualization practice in archaeology. *Open Archaeology 7*, 1 (2021), 353–377. 1

[PSA22] PSARROS D., STAMATOPOULOS M. I., ANAGNOSTOPOULOS C.-N.: Information technology and archaeological excavations: A brief overview. *Scientific Culture 8*, 2 (2022). 1

[RDV*24] RAY S., DANKS N. P., VALDEZ A. C., ESTRADA J. M. V., UANHORO J., NAKAYAMA J., KOYAN L., BURBACH L., BEJAR A. H. C., ADLER S.: seminr: Building and Estimating Structural Equation Models, June 2024. 6

[RFSK19] RAMKUMAR N., FEREYDOONI N., SHAER O., KUN A. L.: Visual behavior during engagement with tangible and virtual representations of archaeological artifacts. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays* (2019), pp. 1–7. 2

[SGIF*22] SILVA-GAGO M., IOANNIDOU F., FEDATO A., HODGSON T., BRUNER E.: Visual attention and cognitive archaeology: an eye-tracking study of palaeolithic stone tools. *Perception 51*, 1 (2022), 3–24. 2

[TIMC21] TAKAHASHI N., INAMURA T., MIZUCHI Y., CHOI Y.: Evaluation of the difference of human behavior between vr and real environments in searching and manipulating objects in a domestic environment. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)* (2021), IEEE, pp. 454–460. 2

[TSC*18] TAKAHASHI R., SUZUKI H., CHEW J. Y., OHTAKE Y., NAGAI Y., OHTOMI K.: A system for three-dimensional gaze fixation analysis using eye tracking glasses. *Journal of Computational Design and Engineering 5*, 4 (2018), 449–457. 2, 5

[VB08] VENKATESH V., BALA H.: Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences 39*, 2 (2008), 273–315. 4

Proceedings published by Eurographics - The European Association for Computer Graphics.