



HAL
open science

Measuring the Impact of Objects' Physicalization, Avatar Appearance, and their Consistency on Pick-and-Place Performance in Augmented Reality

Antonin Cheymol, Jacob Wallace, Juri Yoneyama, Rebecca Fribourg, Jean-Marie Normand, Ferran Argelaguet

► To cite this version:

Antonin Cheymol, Jacob Wallace, Juri Yoneyama, Rebecca Fribourg, Jean-Marie Normand, et al.. Measuring the Impact of Objects' Physicalization, Avatar Appearance, and their Consistency on Pick-and-Place Performance in Augmented Reality. IEEE Transactions on Visualization and Computer Graphics, 2025, 31 (5), pp.2268-2276. <10.1109/TVCG.2025.3549151>. <hal-05164567>

HAL Id: hal-05164567

<https://inria.hal.science/hal-05164567v1>

Submitted on 16 Jul 2025

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

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



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Measuring the Impact of Objects' Physicalization, Avatar Appearance, and their Consistency on Pick-and-Place Performance in Augmented Reality

Antonin Cheymol, Jacob Wallace, Juri Yoneyama, Rebecca Fribourg, Jean-Marie Normand, and Ferran Argelaguet

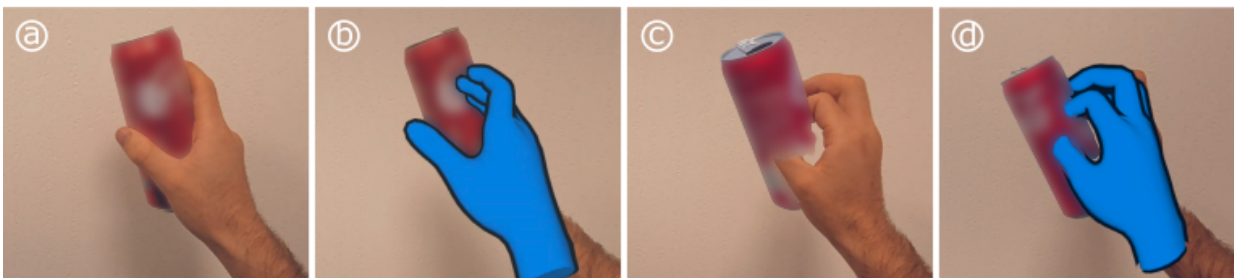


Fig. 1: Illustration of a user holding a can under different conditions of objects physicalization and hand appearance studied in this research. In (a) and (b) the user holds a *real* can, while in (c) and (d) they holds a *virtual* - non-physicalized - can. Moreover, in (a) and (c) the hand is displayed with its *real* appearance, while in (b) and (d) it is overlapped with a *virtual* avatar.

Abstract—Augmented Reality (AR) is a growing technology that enables interaction with both virtual and real objects. However, in order to support the future development of efficient and usable AR interactions, there is still a lack of systematic knowledge establishing basic interaction performance across different conditions. Therefore, in this paper, we report a user study measuring the impact of objects' physicalization (object's set composed of (i) virtual, (ii) real, or (iii) a composite mix of real and virtual objects) and hand appearance (hand's appearance displayed as (i) the real hand, (ii) an avatar, or (iii) dynamically adapting to the surrounding objects' physicalization) on the speed performance of a pick-and-place task. Overall, our results reveal that objects' physicalization plays a significant role in interaction performance, with the more real objects in a set the better the performance. Moreover, our results also suggest that pick-and-place interaction performances are mostly unaffected by the hand appearance. Interestingly, we also observed that interactions with real objects were less efficient as the object condition required the user to alternate between interactions with virtual and real objects (object condition (iii)), which provides novel insights into an important - mostly AR-specific - factor to consider for designing future AR interactions. Taken together, our results provide a rich characterization of different factors influencing different phases of a pick-and-place interaction, which could be employed to improve the design of future AR applications.

Index Terms—Augmented Reality, Avatar, Human-Computer Interaction (HCI)

1 INTRODUCTION

Augmented Reality (AR) technology has recently seen an increase in popularity, driven by technological improvements, along with greater access to cutting-edge industrial devices. Unlike Virtual Reality (VR), AR keeps the user's surroundings visible and, hence, interactable. This

enables ubiquitous utilization, by reducing the risk of users colliding with their surroundings, and permits flexible switches between interaction with real and virtual elements. For example, as AR utilization increases, users could regularly alternate from interacting with virtual objects to grabbing their phone on a table or opening their mailbox while walking back home. However, despite the growing interest in AR and the novel specific interaction opportunities that it opens, there is still a lack of systematic knowledge regarding users' basic interaction performance and the factors that influence it. Investigating this is crucial for establishing empirical foundations that would inform future design of efficient and usable interactions in AR. Note that in this paper, for virtual objects' manipulations, we focus on direct and near-field interactions (i.e., using grab/pinch gestures to select and move virtual objects [6, 21]). This is because they are the most basic, intuitive, and often preferred hand-free AR object manipulation technique [18, 24, 34], due to their equivalence to real-world interactions [22, 29].

Previous results have shown that interacting with physical objects leads to higher performance and usability compared to using gestures to manipulate virtual objects (e.g., objects which does not present any physical counterpart), whether it is in AR [4] or VR [20]. However, as stated above, AR users may also often end up alternating between physical and virtual interactions, hereafter referred to as "composite interaction". Such interaction would most likely require quick mental context-switching to adapt to the targeted interaction mode, which, in some cases, might impact the interaction patterns and performances.

• Antonin Cheymol is with Inria, Univ Rennes, CNRS, IRISA. E-mail: antonin.cheymol@inria.fr.

• Jacob Wallace is with Nantes Université, École Centrale Nantes LS2N-PACCE, UMR 6004, F-44000, Nantes, France. E-mail: jacob.wallace@inria.fr.

• Juri Yoneyama is with Inria, Univ Rennes, CNRS, IRISA. E-mail: juri.yoneyama@inria.fr.

• Rebecca Fribourg is with Nantes Université, École Centrale Nantes LS2N-PACCE, UMR 6004, F-44000, Nantes, France. E-mail: rebecca.fribourg@ec-nantes.fr.

• Jean-Marie Normand is with Nantes Université, École Centrale Nantes LS2N-PACCE, UMR 6004, F-44000, Nantes, France. E-mail: jean-marie.normand@ec-nantes.fr.

• Ferran Argelaguet is with Inria, Univ Rennes, CNRS, IRISA. E-mail: ferran.argelaguet@inria.fr.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxx

However, these impacts and their magnitudes have never been investigated and therefore, remain hypothetical. This highlights the need for further research and robust empirical evidence.

Another understudied aspect of AR interaction is the avatar, the virtual counterpart of the user's body. While this concept has been extensively explored in VR [13, 25, 37, 45], research on avatars in the context of AR remains relatively recent [19]. Beyond applications using avatars to overcome one's body limitations (e.g., by decoupling virtual and real positions to enable further and quicker reaching [46]), previous research has also found that a collocated avatar overlapping the user's hands might improve usability, as well as selection and pick-and-place performance [42, 43]. However, these results were obtained in a specific context, namely a precision task with virtual objects, and there are reasons to doubt its generalization to others. For example, authors suggest that these results are most likely due to the visibility the avatar offers of the interacting layer (i.e., where the system assumes the user's hand to be) [32, 42, 43], aligned with previous observations in VR [28]. Indeed, since the tracking system usually does not enable perfect alignment of the interactive layer with the real hand, the avatar provides more accurate information about its position, thereby helping to avoid movement errors or unintended collisions. However, we might expect this benefit to be less impactful for most AR interactions which typically do not require high movement precision (e.g., grabbing an object with a pinch gesture and placing it in a target area), calling for further investigations in other contexts. Furthermore, information about the position of the interactive layer is only relevant when interacting with virtual objects, whereas, conversely, users should rely on the position of their real hand when interacting with real objects. Therefore, similarly to how not seeing the interactive layer may decrease interaction performance with virtual objects, it should be investigated whether the avatar, by hiding the real hand, may decrease interaction performance with *real* objects. Finally, while the avatar is traditionally restricted to a fixed appearance, it has been proposed that it could also dynamically adapt to the context [11], which could be employed to provide additional information and facilitate interactions [10, 33]. Typically, it would be interesting to study the benefits of dynamically activating the avatar based on proximity to virtual or real objects. This could improve interaction performance by increasing the visibility of the relevant effector (whether it is the avatar or the real hand) and decreasing cognitive load by facilitating mental context-switching.

In order to explore these questions and provide empirical foundations for AR interactions, we report a user study measuring users' pick-and-place speed performance across different conditions of avatar and object physicalization. Specifically, there were three object conditions, which defined whether the manipulated objects were virtual, real, or composite (e.g., balanced between real and virtual objects). Moreover, the hand appearance variable compared the performance when displaying either the real hand, an overlapping avatar, or a hybrid condition (where the avatar activates and deactivates based on proximity with virtual or real objects, respectively).

2 RELATED WORK

We review hereafter previous works that guided this study, beginning with general research related to direct interaction in AR, followed by the specific impact of avatars on it.

2.1 Direct Interactions in AR

Among the diverse types of object interaction techniques in AR, direct interactions, which consist of interacting with virtual objects using gestures similar to those they would use in the real world (e.g., pinching or grabbing gestures), is one of the most common. Its popularity mostly lies in its intuitiveness, by leveraging user's existing real-world knowledge and skills [22]. As a result, it is usually faster to learn, more usable, and preferred by users [18, 24, 29, 34, 46]. As compared to far-field interaction, direct interaction also tends to optimize precision [26]. Finally, users tend to prefer and perform better when interacting with real objects compared to virtual ones (i.e., not physicalized), whether it is

in AR [4] or VR [20], resulting from the additional passive haptic feedback they can rely on. Physicalized interactions also lead to smoother and more natural movements in AR [20]. Moreover, previous studies revealing differences in the gestures employed to interact with real and virtual objects, respectively in the real world and in VR [2], suggest that similar differences likely exist in AR depending on object physicalization. Along those lines, previous works classifying systematically interaction with different types of objects, in real world [7, 15, 16] and in VR [3, 12], may inform the understanding of user's direct interaction in AR.

However, the specificity of the AR context should also be noticed, especially the possibility offered to users to easily alternate between interaction with real and virtual objects. To the extent of our knowledge, no study has investigated the user interaction performance in such composite interaction, which may hypothetically suffer from users having to adapt their interaction gestures and higher cognitive load.

2.2 The Impact of Avatar on Direct Interaction in AR

Recently, constructors of AR-enabled headsets, such as the HoloLens 2, Quest 3, and Vision Pro, have begun integrating hand-tracking cameras that allow for the display of an avatar - a virtual counterpart of the user's body - whose position and movements follow those of the user's real hands. While the avatar can be incorporated to elicit a subjective *sense of embodiment* [25] which can benefit the user experience [19], it also offers numerous benefits in terms of interaction. For example, by decoupling its position from the user's position, it can permit to interact with far-field virtual objects [46] or increase reaching speed [27]. However, by increasing the movement speed, the decoupling may also decrease precision and usability [18, 42]. Also, reducing the virtual hand's volume may make it easier to avoid collisions with virtual objects [41]. Furthermore, even when set to overlap the user's hand, the avatar can assist user's performance notably during precise interactions with virtual objects. Indeed, since the tracking system usually does not enable perfect alignment of the interactive layer with the real hand, the avatar provides more accurate information about its position, thereby helping to avoid movement errors or unintended collisions [32, 42, 43]. On another hand, it is also interesting to note that different avatars' realism does not necessarily impact the user speed and precision performance [36, 43]. However, despite displaying the user's hand position, it has been proposed that the avatar could also be utilized as a "signifier" of the interaction afforded by the avatar [37]. For example, in the context of VR, it was reported that in an object-picking task, directional cues displayed on the avatar would impact objects grabbing position [23]. Along those lines, it was even proposed that the avatar can dynamically alter its appearance to adapt to the context [11]. These propositions taken together, we can imagine that the avatar could be dynamically activated depending on whether the user needs to interact with a real or a virtual object. Such an approach could facilitate interactions by displaying the appropriate effector (whether it is the avatar or the real hand), and improving the user's understanding of affordances.

3 EXPERIMENT

We describe hereafter an experiment aiming at measuring the impact of object physicalization and hand appearance on interaction speed performance. To this end, participants performed several simple objects pick-and-place operations under different conditions. The main experimental task was inspired by Fitts' law-based tasks, in which a number of targets are laid within a circle [30]. In contrast, in our experiment, the pick-and-place operation required to move one object (real or virtual) to the opposite side of the circle (see Figure 4). We were not interested on the throughput but rather on exploring changes during ballistic and corrective motions. This experiment was approved by our institutional review board (OCELER).

3.1 Conditions

We investigated the impact of two independent variables, *object* and *hand* (Figure 1 provides illustrations of some of the experimental conditions). The **object** variable evaluated three conditions of objects' set's physicalization manipulated by participants during the pick-and-place

operation, either real, virtual (i.e., without physical counterpart), or composite (i.e., half the objects are real and half are virtual). Here, real and virtual objects' conditions allow for the comparison of interaction differences provoked by individual physicalization differences. The composite object condition further investigates the specific AR context, where users can easily alternate between different object physicalizations. The **hand** variable evaluated three conditions of the participants' active hand's visual aspect, either real (i.e., participants only see their real hand without any virtual element overlapping), virtual (an avatar, with matching movement and position, is constantly overlapping the user's hand), or hybrid (an avatar, with matching movement and position, appears as the hand approaches a virtual object, and disappears as it approaches a real object - illustrated in Figure 2). Here, real and virtual hand conditions allow for the comparison of interaction differences provoked by fixed appearance differences. The hybrid hand condition further investigates the potential benefit of an avatar dynamically adapting to the context. Note that therefore, the hybrid hand condition was only tested under the composite object condition. Thus a full factorial design was not possible. This led to a total of seven combinations of conditions, i.e., both real and virtual hand conditions tested with real, virtual, and composite object conditions, and the hybrid hand condition tested with the composite object condition.

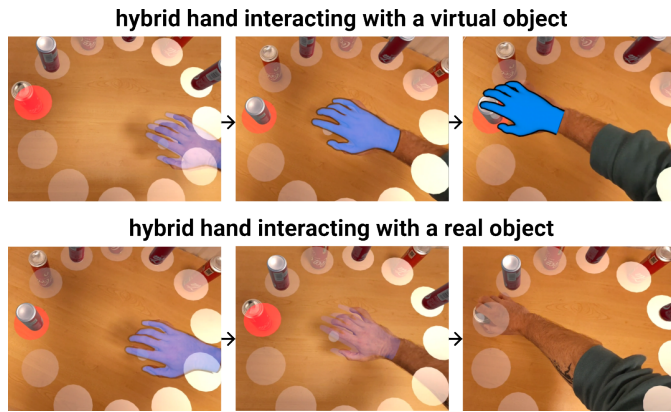


Fig. 2: Illustration of the hybrid hand interaction condition, interacting with either virtual (images at the top) or real (images at the bottom) objects. Vertically aligned images illustrate the user's hand approaching the target object over time. As it approaches, the hand appearance gradually adapts to the object, either by increasing or decreasing the overlapping avatar visibility (respectively in the virtual and real object conditions).

3.2 Participants

We recruited 21 participants (5 females, 16 males), aged between 22 and 36 years old ($M=27.2$, $SD=3.9$). 2 of them were left-handed, 19 right-handed. 4 of them declared having no previous experience with AR, 5 declared using it less than once a month, and 12 more than once a month. The recruitment phase was conducted in the laboratory for convenience, with careful consideration given to exclude any participant who had prior knowledge of the study's purpose.

3.3 Apparatus

Participants stood in front of a table, on which markers were arranged equidistantly in a semi-circle with a diameter of 64cm. The virtual content was presented from the Meta Quest 3 Head-Mounted Display (HMD), enabling visual see-through AR. The virtual and real manipulated objects consisted of identical slim cans of a diameter of 5.7cm and a height of 13.9cm (empty soda cans). We tried to match as best as possible the virtual appearance of the cans to resemble their real counterparts. The choice of using cans for the study was made as they were relatively simple to grasp, and because their circular shape allowed them to be grasped similarly regardless of their rotation. The (empty) real cans weighted 9.5g.

The avatar hand used in the virtual and hybrid hand conditions was blue, with a black outline, ensuring that it would be clearly noticeable, regardless of the participant's skin color. The choice of an abstract appearance was also made to avoid unbalanced dissimilarities between the avatar and the participants' hands (e.g., skin color or gender), which may impact the user experience [14,35]. Note that the avatar hand could not be occluded when overlapped by a real object (see Figure 1(b)) as it would also have been occluded by the participants' hand. In the hybrid hand condition, illustrated in Figure 2, the avatar was originally at 50% transparency, which linearly decreased or increased to 0% or 100% as the user respectively approached the target real or virtual object within a range of 60 to 35cm. To avoid any bias due to a faulty object recognition, we only considered the target-object to be grabbed (always known), to determine whether to display the virtual hand opaque or fully transparent. In the virtual hand condition, the virtual hand was always opaque. Participants could grab the virtual cans by performing a pinching gesture.

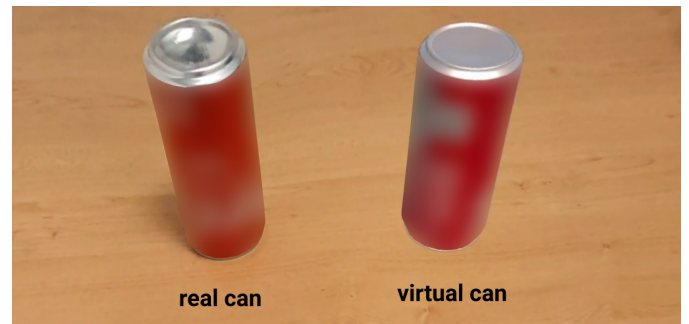


Fig. 3: Real (left) and virtual (right) cans utilized in our experiment

3.4 Protocol

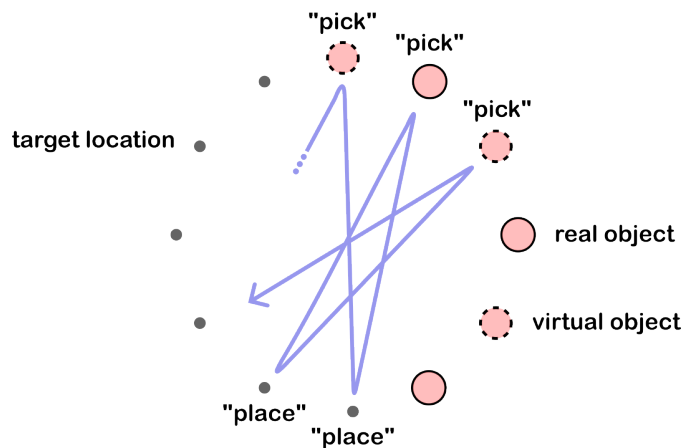


Fig. 4: Illustration of the pick-and-place task in *composite* object condition, where the hand trajectory is represented by the blue path, the real objects by the red circles with plain outline, the virtual objects by red circles with hashed outline, and the target locations by the grey dots.

After signing an informed consent form, participants filled in a demographic questionnaire and were briefed about the experiment process. They were not informed about the purpose of the experiment. Thereafter, participants were equipped with the HMD.

The experiment followed a within-subject design, and was divided into 14 blocks (two repetitions for each of the seven conditions). Each block was successively presented to the participants in a randomized order, and counterbalanced across participants.

The task was inspired from the Fitts' task and was designed to measure pick-and-place manipulations in any direction on a 2D

plane [38,40]. At the beginning of each task, six cans were arranged on top of the markers forming a semi-circle, facing 6 virtual empty can locations completing the circle (12 target locations in total). The task consisted in picking the first can clockwise in the semicircle, and placing it after the last can clockwise in the semicircle (i.e. the place at the opposite position on the circle - see Figure 4). At any time, the targeted object to pick or position to place was indicated by a virtual circle on the table. After 12 pick-and-place operations, the placement of the objects was the same as the initial configuration. For each block, participants performed in total 24 pick-and-place operations (two loops).

Virtual cans were placed automatically, while the real cans were placed manually by the experimenter. The set's composition in terms of virtual and real cans depended on the object condition. The first object to pick at the beginning of the task was always at the furthest position from the user on the circle. If any problem occurred that prevented the participant from completing the task, such as a can falling off the table, they could ask the experimenter to resume the task. This only happened a few times for each participant.

After completing each block, participants were asked to remove the HMD and were given the opportunity to discuss any aspect of the experiment. In total, the experiment lasted about 30min, explanations included.

3.5 Measures

First of all, given the nature of the task, pick-and-place, two different sub-tasks can be considered: (1) the reaching motion to pick the object, and (2) the translation of the object to place it at its final destination. For each sub-task, we were mainly interested in the task completion time (from the release of the previous object to the pick of the current object, and then to the placement of the object to the next location). In addition, to further analyze the potential impact on the task completion time, we were also interested in the maximum hand linear speed (ballistic phase) and the time taken within a 10cm range of the cans' picked and placed positions (corrective phase) for each sub-task [31].

During the experiment, participants' hand position and velocity were recorded at 90Hz, including additional information regarding the next target object and whether the hand was within a close vicinity of the target. Due to the impossibility of precisely determining when the real object was grabbed, the hand motion speed was used to infer the grab and release events for both virtual and real objects (see Section 4).

3.6 Hypotheses

Based on our design and the literature review, we formulated the following hypotheses related to the influence of the object condition (H1.x) and the influence of the hand condition (H2.x):

- **H1.1:** The *real* object condition will lead to a faster completion time than *virtual* and than the *composite* object conditions. Consistent with previous results reporting faster manipulation with physicalized objects [4].
- **H1.2:** The *composite* object condition will obtain a faster completion time than the virtual object condition, resulting from half objects being physicalized, and hence, faster to manipulate according to previous results [4].
- **H1.3:** Interactions with either virtual or real objects will show faster completion time in uniform object conditions (i.e., virtual and real object conditions) than in composite conditions, due to an expected higher cognitive load resulting from alternating between different types of interactions.
- **H2.1:** Consistent hand appearances corresponding to the object condition will lead to faster completion time, i.e., the *virtual* hand condition will be faster in the *virtual* object condition, and the *real* hand condition will be better in the *real* object condition. This would align with previous results reporting better performance when displaying an effector consistent with the manipulated object [32,42,43].
- **H2.2:** The *hybrid* hand condition will lead to higher performance than *virtual* and *real* hand conditions, as it permits to dynamically

align the effector to the context, consistent with the aforementioned results [32,42,43].

4 DATA ANALYSIS

Data from two participants were excluded from the analysis due to issues with data acquisition.

In order to compute the precise grab and release time, hand motion data was resampled at 60Hz and a low-pass filter (Butterworth, cut-off frequency of 2Hz) was used to remove the high frequency component of the signal (see Figure 5). Initially, raw data was flagged when the hand reached the new target (see Trigger Reached label in the Figure 5), a rolling window algorithm was used to compute local minima after the trigger enter event. The algorithm also accounted for miss-placements trials in which multiple corrective movements were performed. This method allowed to compute the estimates for the beginning and end time for each task. The initial pick action and the last place task for each round were discarded. This resulted in a total of 12236 observations (6118 pick and 6118 place operations). A visual analysis of the distribution was used to remove extreme values (21 observations exceeding 5 seconds).

The normality of the data was assessed using the Shapiro-Wilk test. When data did not meet the parametric assumptions, Friedman's test was utilized. The Nemenyi Test was used for the post-hoc pairwise analysis ($p < 0.05$). The data processing and the statistical analysis were performed using R Studio¹.

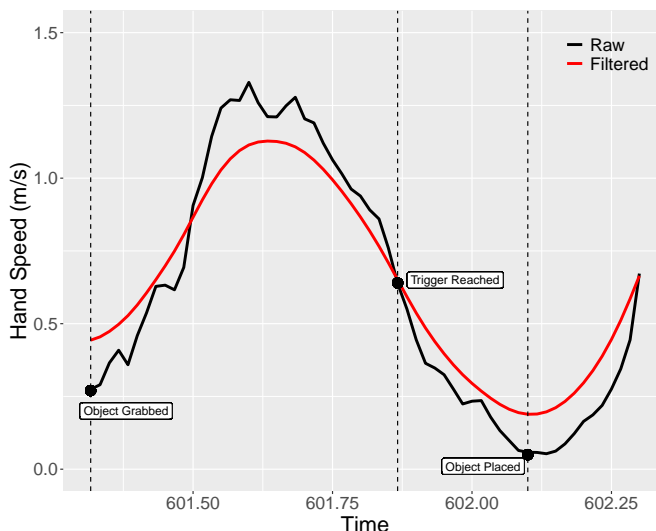


Fig. 5: Example of the evolution of the hand speed during a place task (black line). A low-pass filter was applied (red line) to remove high frequencies and enable the estimation of the time in which the the grab and place actions took place. The Trigger Reached event was logged during the experiment, and the Object Grabbed and Object Placed events were inferred.

5 RESULTS

In the result section, we refer to each combination of hand and object's variable with a notation of the form Hand/Object. With Hand referring to the hand variable, either R, V or H, respectively for real, virtual, or hybrid conditions; Object referring to the object variable, either R, V, C, respectively for real, virtual, or composite conditions. For example, a trial where participants interact with the virtual hand appearance with the composite object's set is referred to as V/C.

Hereafter, we provide a structured and concise summary of our study results, with a particular focus on the nature of the objects, as the impact on the different dependent variables was larger and consistent, while the impact of the hand representation was anecdotal. Moreover, in the

¹<https://www.r-project.org/>

supplementary material, we provide full access to the results of all the statistical analyses, including complete post-hoc test results.

5.1 Task Completion Time

Global Analysis The Friedman analysis total task time showed significant differences both for the pick ($\chi^2(6)=39.93, p<0.001$) and the place tasks ($\chi^2(6)=48.61, p<0.001$) (see Figure 6 left). Post-hoc tests for the pick task, showed that the R/R condition significantly achieved lower task completion times compared to all other conditions (except for V/R). The V/R condition also achieved lower task completion times (all, except for R/R and R/H). Same post-hoc results were obtained for the place task, except that the V/R condition also required significantly less time compared to the R/H condition.

While extending our analysis to the ballistic and correction phases, the Friedman tests showed significant differences for the ballistic time for the pick ($\chi^2(6)=73.22, p<0.001$) and place ($\chi^2(6)=53.752=6, p<0.001$) tasks. Post-hoc tests, showed the same pairwise differences as for the task completion time for both tasks (see Figure 6, right). Regarding the correction time, although the Friedman analysis was significant (pick: $\chi^2(6)=17.74, p=0.047$, place: $\chi^2(6)=15.50, p=0.016$), the post-hoc analysis did not show any significant pairwise difference. This results suggest that the conditions in which all the objects were real (R/R and R/V) achieved the highest performance, and the decrease of performance for the other tasks happened during the ballistic phase and not the corrective phase.

Interaction with real or virtual objects. In order to explore the impact of interacting with real and virtual objects within the same task, for each Hand/Object condition, we isolated the tasks depending on the nature of the virtual object (real or virtual). Although the number of repetitions is not balanced (e.g., more interactions with real objects in the R/R condition compared to the R/C condition), we report them as they provide insightful information. Regarding virtual objects, the different analysis were inconclusive, and only anecdotal differences were found. Suggesting that the interaction with virtual objects was not affected when real objects were also present.

In contrast, the analysis with real objects showed numerous significant differences. Regarding the *pick task*, differences were found for total time ($\chi^2(4)=32.21, p<0.001$), ballistic time ($\chi^2(4)=34.316, p<0.001$), and corrective time ($\chi^2(4)=17.347, p=0.002$) (see Figure 8). Post-hoc tests for total time and ballistic time, showed that participants took significantly more time to interact with real objects in the composite conditions (R/R and V/R conditions enabling faster interactions) for both pick and place tasks. For corrective time, participants only required significantly more time for the V/C condition compared to the others.

Regarding the *place task*, differences were found for the total ($\chi^2(4)=27.07, p<0.001$) and ballistic time ($\chi^2(4)=34.02, p<0.001$), but not for the corrective time ($\chi^2(4)=7.37, p=0.12$). Again, post-hoc tests for the time and ballistic time were consistent, both showing that participants required significantly more time to perform the task for the composite conditions, but this time with the exception of the V/R and the R/C comparison.

5.2 Hand Speed

Global analysis The maximum hand speed for the picking-up motion was significant, $\chi^2(6)=93.74, p<0.001$, as was for the placing motion, $\chi^2(6)=83.91, p<0.001$ (see Figure 7 left). In overall, post-hoc test (see Figure 7 left) showed that the object condition, had a higher influence on the maximum hand speed. With higher hand speeds for real objects (R/R, V/R), followed by composite (R/C, C/C and V/C), and finally virtual objects (R/V, V/V).

The hand speed when reaching the target area also showed significant differences for pick ($\chi^2(6)=95.03, p<0.001$) and place ($\chi^2(6)=73.51, p<0.001$) (see Figure 7 right). Similar results were obtained as for the maximum hand speed, in which the nature of the objects played a more relevant role compared to the hand representation (see Figure 7 right).

Interaction with real or virtual objects Similar as the time analysis, the different tests considering only the interaction with virtual objects were not conclusive, and only anecdotal effects were found. Regarding real objects, we did find significant differences for the maximum hand speed for the place task ($\chi^2(4)=10.61, p=0.031$), and only a tendency for pick tasks ($\chi^2(4)=9.05, p=0.06$). We did post-hoc tests for both tasks, which showed a significant difference between the R/R and V/C conditions, R/R significantly higher hand speed (see Figure 9 left).

Friedman analysis also showed significant differences for the hand speed when reaching the trigger for pick ($\chi^2(4)=28.30, p<0.001$) and place ($\chi^2(4)=30.27, p<0.001$) tasks (see Figure 9 right). Post-hoc tests, for the pick task, showed that the hand speed was lower for the composite conditions compared to the R/R condition, while no significant differences were found between the R/R and V/R conditions. Significant differences were also found between the V/C and V/R conditions. Similar trends were also found for the place task, although differences were more mitigated.

6 DISCUSSION

We discuss hereafter the results of our experiment, which aimed at exploring the impact of objects' physicalization and avatar appearance on pick-and-place interactions in AR.

6.1 Impact of Objects Physicalization

To begin with, we observed that in the real object conditions, regardless of the hand appearance, participants obtained faster task completion, reached higher maximum speed, and attained the target at a faster speed, for both pick and place motions. These results support **HI.1**. Moreover, in the hybrid object condition, we also observed a higher maximum hand speed and hand speed when reaching the target area, in comparison to the virtual object condition. These results only partially support **HI.2**, as the difference was not observed for completion time. Taken together, these results are in line with previous results reporting faster interaction with physicalized objects [4]. This suggests that AR designers should prioritize physicalized objects if aiming at increasing interaction speed, even if it results in introducing a composite set of objects' physicalization instead of non-physicalized objects only. They also provide additional insights, by revealing that the physicalization seems to consistently impact composite object sets, i.e., interacting with a set evenly composed of both real and virtual objects tends to lead to faster interactions than a set of fully virtual objects, but lower than a set of fully real objects.

Interestingly, by separating ballistic and corrective movements for task completion times results, we also observed that the object physicalization has a higher impact on the reaching movement than on the manipulation (i.e., pick or place) itself. These results are in line with the significant differences we found in terms of maximum hand speed and hand speed when reaching the target area, which results mainly from the ballistic movement. While these results need to be further investigated to be well understood, we can already hypothesize that they could be due to the user focusing more on adjusting their hand in the case of virtual interactions, or reducing their movement speed to precisely control the movement endpoint, compensating for the absence of tactile feedback and opposed normal or friction forces (either between the hand and the can, or between the can and the table, respectively for pick and place motions)

6.2 Impact of Objects Consistency

Our results show that beyond the physicalization of the interacted object, the consistency of the object's set (i.e., whether it is composite or not) also influences the performance of interaction with real objects. Indeed, the results of interactions with real objects in the real object condition, in comparison to the composite object condition, showed faster task completion and hand speed when reaching the trigger, for both pick and place motions. They provide new insights for AR interaction designs, by suggesting that alternating between virtual and real objects' interaction in AR negatively impacts the capacity to efficiently interact with real objects. Interestingly, these results were obtained in both pick and place motions, indicating that not only the transition between

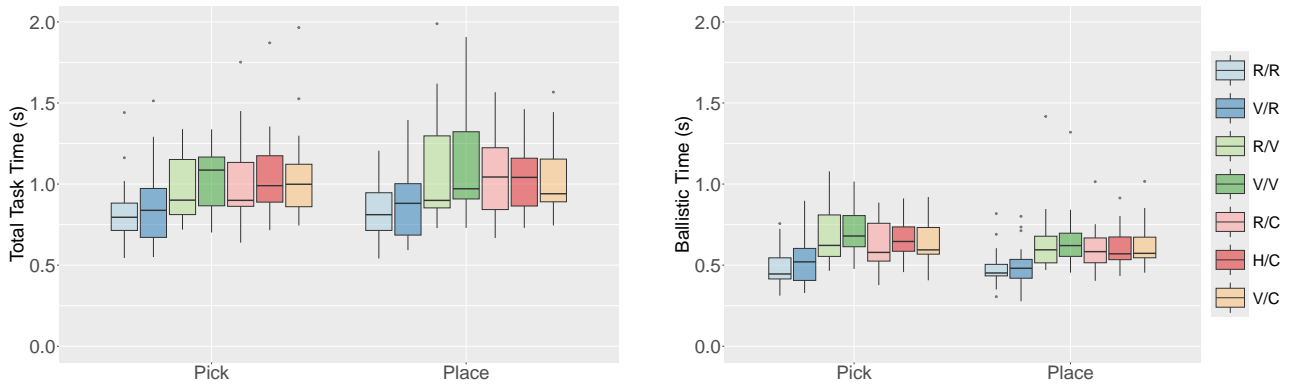


Fig. 6: Boxplots for the total (left) and ballistic (right) time for picking and placing tasks, grouped by conditions.

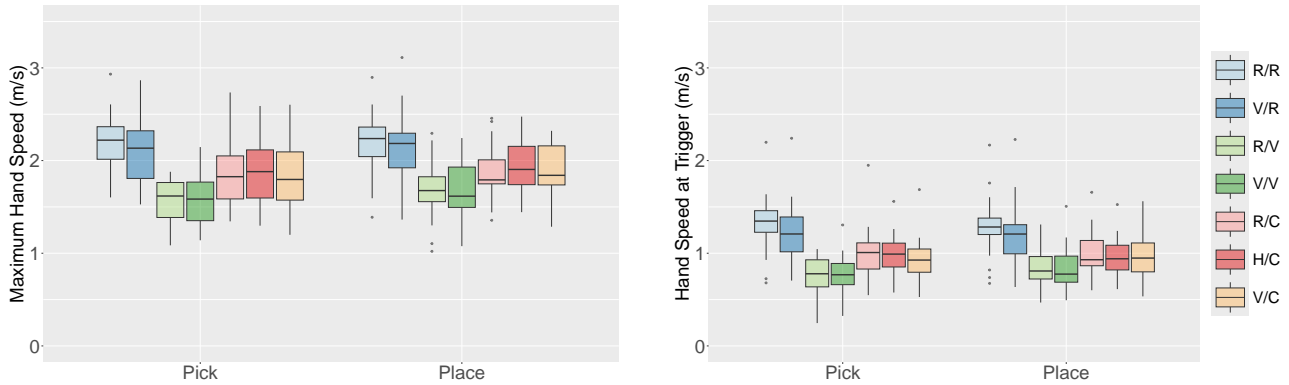


Fig. 7: Maximum hand speed (left) and hand speed at trigger (right) for the pick and place tasks, grouped by condition.

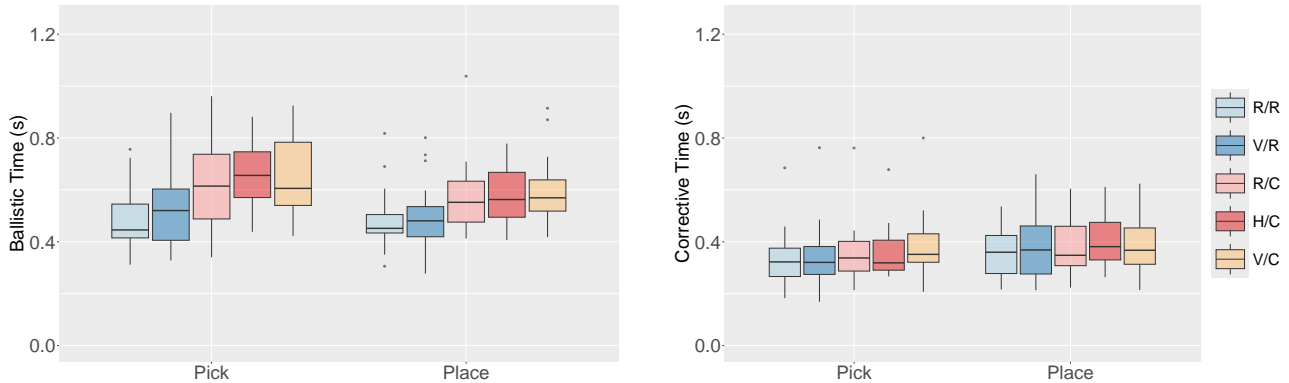


Fig. 8: Ballistic (left) and corrective (right) time for the pick and place tasks only considering interactions with real objects, grouped by condition.

virtual and real interaction was impacted (given that the user picked before placing the object, the transition between the previous and current interaction occurred as the pick interaction started). However, the difference was not observed for interactions with virtual objects. This might be due to a higher impact of altering real object interactions, considering that they require more precise movements. However, this is only a hypothetical explanation, and more studies are required to better understand this difference. Taken together, these results partially support **H1.3**.

6.3 Impact of Hand Appearance

We did not find any difference between hand conditions (i.e. when different hand conditions are presented under similar object conditions). These results do not support **H2.1** and **H2.2**, and do not align with previous results reporting and influence of the hand appearance on

different interaction tasks [32, 42, 43]. The difference most likely lies in the nature of the task. While previous results studied precision-based tasks, the task considered in our study did not require high precision, as participants were only instructed to place the object in the target destination, without enforcing a precise placing position. Moreover, it is interesting to note that even the virtual and hybrid hand conditions did not provide accurate occlusion cues when grabbing real objects, this did not impact performance measurements significantly. In this context, we can imagine that the subtle position differences between real or virtual hand positions and the affordances provided by the hand representation have less influence on performance. However, due to the repetitive nature of the task (same motions and same objects), we can also hypothesize that participants relied less on the hand visual information, and more on haptic and proprioceptive information. This supposition is supported by most participants reporting in post-experiment discus-

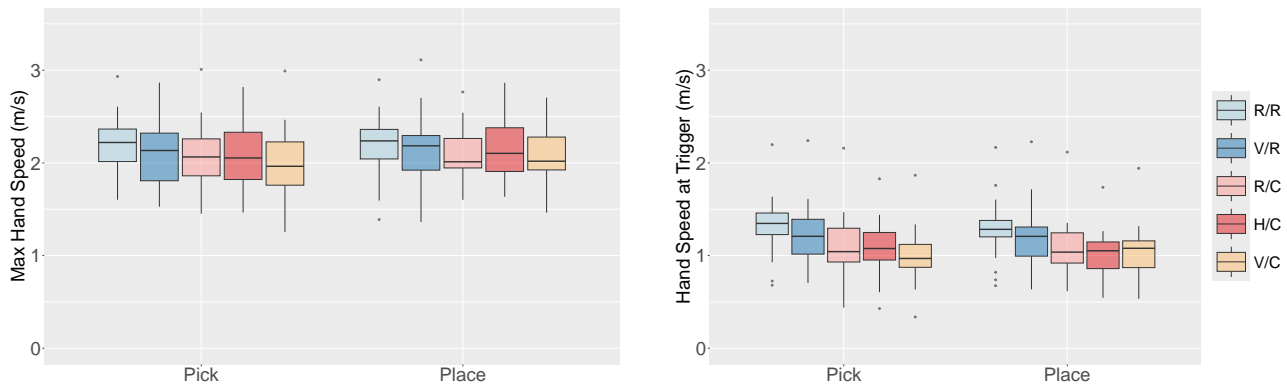


Fig. 9: Hand speed when reaching the trigger for the Real objects (left) and the virtual objects (right) for the pick and place tasks, grouped by condition.

sions that they did not notice or pay attention to the hand’s appearance. Nevertheless, tasks involving objects with different shapes might better highlight the differences in the hand representation. Taken together, these results also suggest that in this specific context, AR designers may choose the hands’ appearance based on their impacts on other aspects of the user experience, considering the little impact it plays on the interaction speed.

7 LIMITATIONS AND FUTURE WORKS

We develop hereafter key research perspectives resulting from our study’s results and its limitations. Before addressing perspectives involving different protocols, we should mention that our results were extracted from a relatively small sample size (19 participants), calling for replications including additional participants.

7.1 Expanding the Understanding of the Impact of Objects Physicalization and Avatar Appearance

Our results were obtained under a specific interaction context, which should be broadened in the future to expand our understanding of the impact of the tested conditions. To begin with, future works could explore the impact of different interaction tasks. Indeed, different tasks could mediate the negative impact caused by the lack of haptic feedback in interactions with virtual objects, but also the user’s needs to rely visually on the hand position, and hence, the impact of the hand appearance. Therefore, future research should explore the impact of different object’s *characteristics*, in terms of size, weight, shape, or even texture [8]. Moreover, different *types of interactions*, such as press, push, rotations, or more complex daily object manipulations, could also differently influence the user performance and the impacts of our conditions. Regarding the hand representation, we cannot rule out that our results could have been impacted by the avatar implementation, such as the precise occlusion from physical objects, or the hand appearance (even though previous results in VR seem to indicate that different avatars’ appearance does not impact pick-and-place performances [36]). Moreover, it would be interesting to observe how different virtual augmentations could reduce the negative impact caused by the absence of objects’ physicalization. For example, visual cues indicating contact zones between the hand and a virtual object - which obtained contrasting results in VR [9, 44] - could specifically help reduce the impact of the lack of haptic feedback in AR.

7.2 Expanding the Understanding of the Impact of Composite Interactions

Our results also revealed that composite objects’ settings might negatively impact the interaction performance with real objects, due to the inconsistency of interaction type. Future research should expand these early results, especially by exploring factors that could impact this phenomenon. To begin with, the resemblance between virtual and real objects may play a role by highlighting the distinction in terms of required interaction and facilitating re-adaptations. Moreover, a larger

discrepancy between real and virtual object interactions (e.g., in terms of required hand aperture or strength [2]) due to different tasks or object characteristics, could negatively impact the user’s capacity to alternate between physical and virtual interactions. Additionally, settings where each object’s characteristics are inconsistent within virtual or real object sets (e.g., different shapes), and hence require different hand manipulations [2], could also further increase the user’s difficulty in alternating between virtual and real interaction. Finally, performance could be different in manipulation tasks involving higher levels of cognitive load, such as assembly tasks of puzzles, as they could decrease the attention users would allocate to the adaptation of interaction type. Investigating this latter factor could help in understanding the impact of such phenomenon in diverse ecological situations, where users often manipulate objects while focusing on higher-level cognitive tasks (e.g., talking on the phone, watching a video, or trying to remember something).

7.3 Expanding to Other Metrics

In this study, we measured the impact of objects’ physicalization and hand appearance on a pick-and-place interaction, from the perspective of a *specific metric of task performance*. Of course, different measurements could have led to partially different results or revealed other aspects of the user experience. Starting with our task design. It has to be mentioned that contrary to the classical Fitts’ law task [38], we did not measure the user performance under different distances and target sizes. Considering different index of difficulties would have allowed the computation of the throughput for each interaction modality enabling other analysis. Moreover, based on the Optimized Initial Impulse Model [31], we defined the ballistic and corrective phases by heuristically determining (i) when objects (real and virtual) were considered grabbed or placed, and (ii) the distance from the target that delineates ballistic and corrective phases. The heuristics employed were based on the hand speed motion, which consistently showed acceleration and deceleration motions for each task (see Figure 10). Nevertheless, we cannot exclude that a different analysis could have led to partially different results. Additional tracking methods could be used in order to precisely measure the motion of real objects, to ensure more accurate measurements.

Beyond task performance measurement, the understanding of the overall user experience would also benefit from considering additional metrics that were not included in this study (mostly to reduce the complexity of the interaction and protocol). For example, the perceived usability, usually measured through questionnaires [5, 17], could provide additional insights into the user’s subjective experience of the interaction. Moreover, it would have been interesting to measure the user’s cognitive load [39] across different conditions. Specifically, it could have allowed for a better understanding of its potential impact on composite interaction settings. Finally, it would have also been interesting to measure how much the participants felt that they were *embodying* the different avatars, namely their subjective “sense of embodiment” [19, 25], which is known to impact different aspects the

user's interaction in VR [1].

8 CONCLUSION

In this paper, we systematically measured the impact of three objects' physicalization conditions (virtual, real, and composite) and three hand appearance conditions (real, virtual, and hybrid), on a pick-and-place task in AR. Our results revealed three main tendencies. *First*, the objects' physicalization has a significant impact on the interaction performance, with the more real objects in a set the better the performance. *Secondly*, pick-and-place interaction performances are mostly unaffected by the hand appearance. *Thirdly*, interactions with real objects are less efficient as the object condition requires the user to alternate between interactions with virtual and real objects (composite object condition). This latter result provides novel insights into an important - mostly AR-specific - factor to consider for the design of future AR interactions. Taken together, our results provide a rich characterization of various factors influencing the different phases of a pick-and-place interaction, which could be employed to improve the design of future AR applications.

REFERENCES

- [1] F. Argelaguet, L. Hoyet, M. Trico, and A. Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE virtual reality (VR)*, pp. 3–10. IEEE, 2016. 8
- [2] A. D. Blaga, M. Frutos-Pascual, C. Creed, and I. Williams. A grasp on reality: understanding grasping patterns for object interaction in real and virtual environments. In *2021 IEEE international symposium on mixed and augmented reality adjunct (Ismar-Adjunct)*, pp. 391–396. IEEE, 2021. 2, 7
- [3] A. D. Blaga, M. Frutos-Pascual, C. Creed, and I. Williams. Vr-grasp: A human grasp taxonomy for virtual reality. *International Journal of Human-Computer Interaction*, pp. 1–17, 2024. 2
- [4] E. Bozgeyikli and L. L. Bozgeyikli. Evaluating object manipulation interaction techniques in mixed reality: Tangible user interfaces and gesture. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 778–787. IEEE, 2021. 1, 2, 4, 5
- [5] J. Brooke. Sus: A quick and dirty usability scale. *Usability Evaluation in Industry*, 1996. 7
- [6] V. Buchmann, S. Violich, M. Billingham, and A. Cockburn. Fingartips: gesture based direct manipulation in augmented reality. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pp. 212–221, 2004. 1
- [7] I. M. Bullock, T. Feix, and A. M. Dollar. Human precision manipulation workspace: Effects of object size and number of fingers used. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5768–5772. IEEE, 2015. 2
- [8] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar. Benchmarking in manipulation research: Using the yale-cmu-berkeley object and model set. *IEEE Robotics & Automation Magazine*, 22(3):36–52, 2015. 7
- [9] R. Canales, A. Normoyle, Y. Sun, Y. Ye, M. D. Luca, and S. Jörg. Virtual grasping feedback and virtual hand ownership. In *ACM Symposium on Applied Perception 2019*, pp. 1–9, 2019. 7
- [10] A. Cheymol, R. Fribourg, A. Lécuyer, J.-M. Normand, and F. Argelaguet. Avatar-centered feedback: Dynamic avatar alterations can induce avoidance behaviors to virtual dangers. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 91–100. IEEE, 2024. 2
- [11] A. Cheymol, A. Lécuyer, J.-M. Normand, and F. Argelaguet. Beyond my real body: Characterization, impacts, applications and perspectives of “dissimilar” avatars in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 2
- [12] A. Dalia Blaga, M. Frutos-Pascual, C. Creed, and I. Williams. Freehand grasping: An analysis of grasping for docking tasks in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 749–758, 2021. doi: 10.1109/VR50410.2021.00102 2
- [13] D. Dewez, L. Hoyet, A. Lécuyer, and F. Argelaguet Sanz. Towards “avatar-friendly” 3d manipulation techniques: Bridging the gap between sense of embodiment and interaction in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021. 2
- [14] T. D. Do, C. I. Protko, and R. P. McMahan. Stepping into the right shoes: The effects of user-matched avatar ethnicity and gender on sense of embodiment in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 3
- [15] T. Feix, I. M. Bullock, and A. M. Dollar. Analysis of human grasping behavior: Correlating tasks, objects and grasps. *IEEE transactions on haptics*, 7(4):430–441, 2014. 2
- [16] T. Feix, I. M. Bullock, and A. M. Dollar. Analysis of human grasping behavior: Object characteristics and grasp type. *IEEE transactions on haptics*, 7(3):311–323, 2014. 2
- [17] K. Finstad. The usability metric for user experience. *Interacting with computers*, 22(5):323–327, 2010. 7
- [18] M. Frutos-Pascual, C. Creed, and I. Williams. Head mounted display interaction evaluation: manipulating virtual objects in augmented reality. In *Human-Computer Interaction-INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part IV 17*, pp. 287–308. Springer, 2019. 1, 2
- [19] A. Genay, A. Lécuyer, and M. Hachet. Being an avatar “for real”: a survey on virtual embodiment in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):5071–5090, 2021. 2, 7
- [20] L. Gerini, F. Solari, and M. Chessa. Passive haptic feedback for more realistic and efficient grasping movements in virtual environments. In *International Conference on Extended Reality*, pp. 3–22. Springer, 2023. 1, 2
- [21] T. Ha, S. Feiner, and W. Woo. Wearhand: Head-worn, rgb-d camera-based, bare-hand user interface with visually enhanced depth perception. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 219–228. IEEE, 2014. 1
- [22] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-based interaction: a framework for post-WIMP interfaces. In *Proc. SIGCHI conference on Human factors in computing systems*, pp. 201–210, 2008. 1, 2
- [23] A. Kadri, A. Lécuyer, J.-M. Burkhardt, and S. Richir. The visual appearance of user's avatar can influence the manipulation of both real devices and virtual objects. In *IEEE symposium on 3D user interfaces*, pp. 291–292, 2007. 2
- [24] H. J. Kang, J.-h. Shin, and K. Ponto. A comparative analysis of 3d user interaction: How to move virtual objects in mixed reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 275–284. IEEE, 2020. 1, 2
- [25] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012. 2, 7
- [26] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, 68(10):603–615, 2010. 2
- [27] S. Kyian and R. Teather. Selection performance using a smartphone in vr with redirected input. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, pp. 1–12, 2021. 2
- [28] C. Lougiakis, A. Katifori, M. Roussou, and I.-P. Ioannidis. Effects of virtual hand representation on interaction and embodiment in hmd-based virtual environments using controllers. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 510–518. IEEE, 2020. 2
- [29] A. Macararas, A. N. Antle, and B. E. Riecke. What is intuitive interaction? balancing users' performance and satisfaction with natural user interfaces. *Interacting with Computers*, 27(3):357–370, 2015. 1, 2
- [30] I. S. MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction*, 7(1):91–139, 1992. 2
- [31] D. E. Meyer, R. A. Abrams, S. Kornblum, C. E. Wright, and J. Keith Smith. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological review*, 95(3):340, 1988. 4, 7
- [32] E. Normand, C. Pacchierotti, E. Marchand, and M. Marchal. Visuo-haptic rendering of the hand during 3d manipulation in augmented reality. *IEEE Transactions on Haptics*, 2024. 2, 4, 6
- [33] M. Prachyabrued and C. W. Borst. Visual feedback for virtual grasping. In *IEEE symposium on 3D user interfaces*, pp. 19–26, 2014. 2
- [34] M. Quandt, D. Hippert, T. Beinke, and M. Freitag. User-centered evaluation of the learning effects in the use of a 3d gesture control for a mobile location-based augmented reality solution for maintenance. *DELBA@ EC-TEL*, 3, 2020. 1, 2
- [35] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. "these are not my hands!" effect of gender on the perception of avatar hands in virtual reality. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 1577–1582, 2017. 3

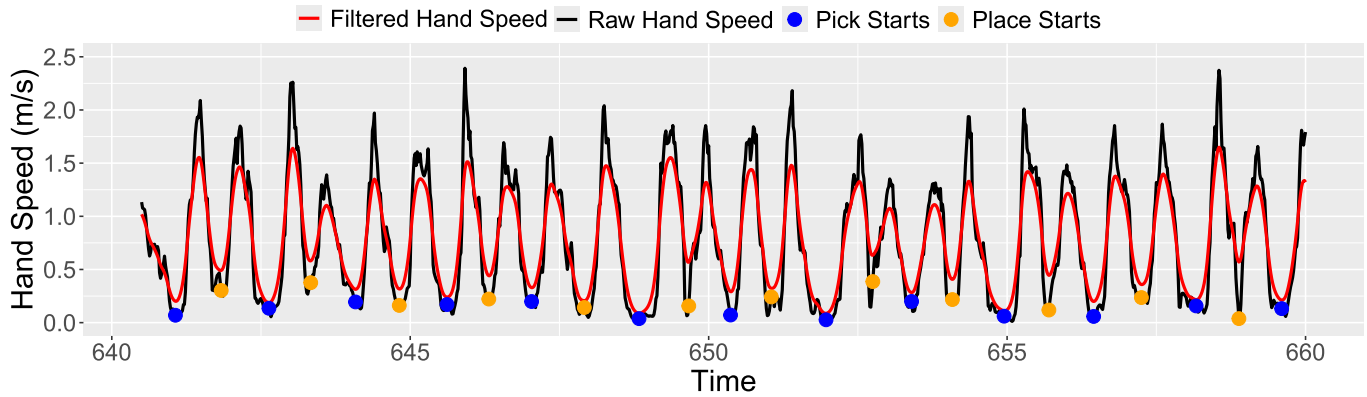


Fig. 10: Example of one participant's hand speed data over time for a complete block, with raw and filtered data (respectively black and red curves). Blue and Orange dots represent the deducted start of pick and place interactions, respectively, consisting of local minima after the hand entered the vicinity of the target position (within a 10cm radius around the target).

- [36] V. Schwind, J. Leusmann, and N. Henze. Understanding visual-haptic integration of avatar hands using a fits' law task in virtual reality. In *Proceedings of Mensch Und Computer 2019*, pp. 211–222. Association for Computing Machinery, 2019. 2, 7
- [37] S. Seinfeld, T. Feuchtnr, A. Maselli, and J. Müller. User representations in human-computer interaction. *Human-Computer Interaction*, 36(5-6):400–438, 2021. 2
- [38] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of fits' law research in hci. *International journal of human-computer studies*, 61(6):751–789, 2004. 4, 7
- [39] J. Sweller. Cognitive load theory, 2011. 7
- [40] R. J. Teather, W. Stuerzlinger, and A. Pavlovych. Fishtank fits: a desktop vr testbed for evaluating 3d pointing techniques. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pp. 519–522, 2014. 4
- [41] T. Q. Tran, H. Shin, W. Stuerzlinger, and J. Han. Effects of virtual arm representations on interaction in virtual environments. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–9, 2017. 2
- [42] R. Venkatakrishnan, R. Venkatakrishnan, R. Canales, B. Raveendranath, C. C. Pagano, A. C. Robb, W.-C. Lin, and S. V. Babu. Investigating the effects of avatarization and interaction techniques on near-field mixed reality interactions with physical components. *IEEE Transactions on Visualization and Computer Graphics*, 30, 2024. 2, 4, 6
- [43] R. Venkatakrishnan, R. Venkatakrishnan, B. Raveendranath, C. C. Pagano, A. C. Robb, W.-C. Lin, and S. V. Babu. Give me a hand: Improving the effectiveness of near-field augmented reality interactions by avatarizing users' end effectors. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2412–2422, 2023. 2, 4, 6
- [44] S. Vosinakis and P. Koutsabasis. Evaluation of visual feedback techniques for virtual grasping with bare hands using leap motion and oculus rift. *Virtual Reality*, 22(1):47–62, 2018. 7
- [45] F. Weidner, G. Boettcher, S. A. Arboleda, C. Diao, L. Sinani, C. Kunert, C. Gerhardt, W. Broll, and A. Raake. A systematic review on the visualization of avatars and agents in ar & vr displayed using head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2596–2606, 2023. 2
- [46] M. Whitlock, E. Harnner, J. R. Brubaker, S. Kane, and D. A. Szafrir. Interacting with distant objects in augmented reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 41–48. IEEE, 2018. 2