



## Evaluation of AR Inconsistencies on AR Placement Tasks: A VR Simulation Study

Romain Terrier, Ferran Argelaguet Sanz, Jean-Marie Normand, Maud Marchal

### ► To cite this version:

Romain Terrier, Ferran Argelaguet Sanz, Jean-Marie Normand, Maud Marchal. Evaluation of AR Inconsistencies on AR Placement Tasks: A VR Simulation Study. EuroVR 2018 - International Conference on Virtual Reality and Augmented Reality, Oct 2018, London, United Kingdom. pp.190-210, 10.1007/978-3-030-01790-3\_12 . hal-01947356

**HAL Id: hal-01947356**

**<https://inria.hal.science/hal-01947356>**

Submitted on 6 Dec 2018

**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.

# Evaluation of AR Inconsistencies on AR Placement Tasks: A VR Simulation Study

Romain Terrier<sup>1,4</sup>, Ferran Argelaguet<sup>1</sup>, Jean-Marie Normand<sup>1,3</sup>, and Maud Marchal<sup>1,2</sup>

<sup>1</sup> Univ. Rennes, Inria, CNRS, IRISA, Rennes, France  
`firstname.lastname@irisa.fr`

<sup>2</sup> INSA, Rennes, France

<sup>3</sup> Ecole Central de Nantes, AAU UMR CNRS 1563, Nantes, France  
`jean-marie.normand@ec-nantes.fr`

<sup>4</sup> IRT b<>com, Rennes, France  
`romain.terrier@b-com.com`

**Abstract.** One of the major challenges of Augmented Reality (AR) is the registration of virtual and real contents. When errors occur during the registration process, inconsistencies between real and virtual contents arise and can alter user interaction. In this paper, we assess the impact of registration errors on the user performance and behaviour during an AR pick-and-place task in a Virtual Reality (VR) simulation. The VR simulation ensured the repeatability and control over experimental conditions. The paper describes the VR simulation framework used and three experiments studying how registration errors (e.g., rotational errors, positional errors, shaking) and visualization modalities (e.g., transparency, occlusion) modify the user behaviour while performing a pick-and-place task. Our results show that users kept a constant behavior during the task, i.e., the interaction was driven either by the VR or the AR content, except if the registration errors did not enable to efficiently perform the task. Furthermore, users showed preference towards an half-transparent AR in which correct depth sorting is provided between AR and VR contents. Taken together, our results open perspectives for the design and evaluation of AR applications through VR simulation frameworks.

**Keywords:** Registration Errors · Augmented Reality · VR Simulation · Interaction

## 1 Introduction

Current advances in Augmented Reality (AR) technology (e.g., Microsoft HoloLens or Meta 2 glasses) as well as in tracking capabilities [5] (with e.g., the release of Apple’s ARKit and Google’s ARCode SDKs) are showing the potential of AR applications in consumer grade applications (e.g., entertainment, education or maintenance). However, despite these recent advances, AR is still confronted to a number of challenges such as occlusion management between virtual and real objects, limitations in the field-of-view of AR devices or, as it is the main

focus in this paper, registration errors. Registration is generally referred to as the process of finding in real-time the position and orientation of virtual objects so that they can be integrated in a plausible way in the real world. When errors occur during this process, inconsistencies between real and virtual objects can arise. Such inconsistencies can be constant (e.g., fixed errors on position and orientation) or irregular (e.g., shakiness of the virtual content), and can potentially hinder user interaction or alter users' behavior. Furthermore, as mentioned by Azuma and Bishop [2] *"The human visual system is very good at detecting even small misregistrations [...]. Errors of just a few pixels are noticeable"*. As of today, most AR applications or SDKs still face many of these inconsistencies, especially when the user modifies the scene through direct interaction with real or virtual objects.

Studying how registration issues affect users when interacting with AR is of great importance but remains difficult to achieve. Indeed, since AR is typically presented on hand-held devices (tablets, phones) or Head-Mounted Displays (HMDs), it generally remains complex to propose repeatable and controlled user studies in AR environments. To that purpose, Virtual Reality (VR) provides a promising tool to evaluate, not only AR interfaces, but also to explore how current limitations of AR systems influence users' behaviors and interaction capabilities [12]. Carrying out controlled VR experiments enables to explore a particular subset of limitations while perfectly simulating other AR features (e.g., ensuring a perfect tracking).

In this paper, we explore, through a VR simulation of AR, how user behavior and performance is altered when registration errors occur. Indeed, while it has been shown that the perception of the co-existence of virtual and real objects can be altered by registration accuracy (e.g., misregistration distorts spatial relationships [15]), little is known of their impact on the interaction process. Through three different experiments participants were confronted with different degrees of registration errors and AR visualizations in a VR simulation environment while performing a pick-and-place task (a common manipulation task). The main goal of the experiments were (1) to explore how the intensity of registration errors alters the behavior of users, in particular, whether real or augmented content drive their interactions, and (2) to measure how irregular registration errors impact users' accuracy.

The remainder of the paper is structured as follows: Section 2 provides an overview on registration errors in AR as well as on how VR simulation is used to study AR systems. Then, Section 3 presents the VR simulation platform we used in our experiments. Sections 4, 5 and 6 detail three different experiments aiming to analyze the impact of registration errors on users' behaviors and performance. Finally, Section 7 presents a global discussion and Section 8 concludes the paper.

## 2 Related Work

### 2.1 Categorization of Registration Errors in AR

A critical issue in AR applications is when virtual information is misaligned with the real environment [3]. This misalignment is also called registration errors [10]. From the users' viewpoint, registration errors can be seen as if the AR information floats and can break the illusion that real and virtual objects co-exist [15].

Registration should be stable both spatially (virtual objects must be collocated in rotation and position with real objects) and temporally (motion of virtual and real objects should be synchronous). In the literature, several classifications of registration errors have been presented [9, 3]. Holloway [9] gives a precise definition of registration errors by decomposing them into four main metrics: linear, lateral, angular and depth registration errors. In contrast, the classification proposed by Azuma [3] focused on whether or not AR objects or the user's viewpoint were static (static) or in motion (dynamic). The remainder of this section is structured following the taxonomy of Azuma [3] where static errors represent a spatial incoherence between AR and real content while dynamic errors represent a temporal incoherence.

**Static Errors** Static errors are visible even if the user does not move his/her viewpoint (or the hand-held device) or when the real environment remains immobile. From the user's viewpoint, the AR content seems to be floating near its real position. Static registration errors are due to either optical distortions, tracking errors, mechanical misalignments or incorrect viewing parameters (i.e., field of view, tracker-to-eye position and orientation, inter-pupillary distance) [3]. Most of the time users perceive static errors as a constant gap between the desired and the actual position/orientation of the AR information even if the AR content has a perfect shape (see Fig. 3 for examples of static errors).

There are several ways to reduce static errors: improve calibration [1] or improve tracking techniques [11]. Nevertheless, the huge variability of environments (e.g., indoor, outdoor) and behaviors of tracked objects (e.g., static, in motion, slow, fast) makes it complex to provide error-prone solutions [19].

**Dynamic Errors** On the other hand, dynamic errors are only visible when the user's viewpoint or when objects are moving. Dynamic errors are mainly due to the latency of the system when there is a motion [3]. This delay (or latency) is the time between the moment when the tracking system computes the new position and rotation of the viewpoint and the moment when the virtual information is rendered at this position. This delay implies that virtual objects are not displayed with the right position and orientation at the right time. In order to reduce dynamic errors, four methods exist [3]: reduce system lag, reduce apparent lag, match temporal streams and predictive methods. Another characteristic of registration errors is that they are hard to predict. Such uncertainty

introduces a strong bias when evaluating AR systems as replicability is compromised. In order to overcome such limitations, VR has been proposed to simulate AR systems and thus evaluating them.

## 2.2 Simulating AR in VR

Virtual reality is a powerful tool to evaluate AR systems as it allows to simulate in a controlled way many AR features, enabling repeatable user studies. Several studies conducted in VR environments simulate AR systems in order to evaluate features such as latency [16, 15], field of view [17, 18] or visual realism [14].

Regarding latency, Lee et al. [13] replicated an AR study [6] in a VR simulation context. Their results showed that users' performance when moving a virtual ring along a virtual path was comparable between the simulated AR and the real AR studies. The VR simulator provided the feeling that the AR content was real thanks to a restricted transparent window in which the AR simulated content was displayed. Moreover, on top of the internal latency of the simulator due to tracker latency, computation time, render time and display time, Lee et al. [12] proposed to include "artificial latencies" between virtual objects and the real world. This enabled to add controlled latency to the simulated AR content and to analyze its effect. Other studies also linked latency with interaction performance showing a degradation of performance as latency increases [15]. Finally, Ragan et al. [16] studied the impact of jitter (visual shakiness of AR content) and showed that it is predominant over visual latency.

The Field of View (FOV) is a distinctive feature of HMDs that has also been evaluated in VR simulations. While comparing different AR HMDs with different FOV is prone to bias due to confounding factors (e.g., other HMDs characteristics) VR simulations enable to only alter the FOV parameter while minimizing confounding factors. For example, Ren et al. [17] showed that a wider FOV is better because it allows to display more information (2D annotations in their case) and to explore it more quickly. Moreover, a wider FOV also reduces users' head movements in the search task. Another important factor related to FOV corresponds to where guidance information regarding objects of interest is displayed (e.g., within the FOV, outside of the FOV, on the object, etc.). Users better focus during a manipulation task if AR guidance is drawn as a line connecting the hand of the user to the center of the searched object [18]. In another study, Baričević et al. [4] proposed to use VR to simulate user perspective rendering in AR (the view is rendered according to the user's point of view) in order to study its benefits over the classical device perspective (the view is rendered according to the point of view of the device's camera) for AR applications.

Furthermore, visual realism is also an important factor that simulated AR has to deal with. Lee conducted a study [14] about the impact of the realism of the VR environment (photo-realistic, etc.) on a user task. Results did not show relevant effects and the necessity to design a realistic VR environment has not been proved in this case. Finally, AR content presentation (transparency of the virtual content or occlusion management) has also been shown to impact

interaction with AR systems [7, 8]. In order to study its impact, we integrated those conditions into our experiments.

As a conclusion, although registration is a major challenge of AR, there is still a lack of studies on its impact in the interaction process. Such studies are complex to perform using existing AR systems since it is nearly impossible to ensure experimental conditions comparable across participants (e.g., registration errors will be hardly reproducible). To this end, inspired by existing works, we propose to study the impact of registration errors in a VR simulation of an AR interaction task.

### 3 Experimental Platform: VR Simulation of AR

As mentioned above, in order to be able to carry out repeatable and fully controlled user experiments, we chose to simulate our AR environments in VR. Our VR environment was designed to replicate a pick-and-place task of an AR application where the user has to position a cube (10x10x10cm) precisely onto a target. The cube and the target were lying on a table and both could be augmented (see Fig. 1, bottom).

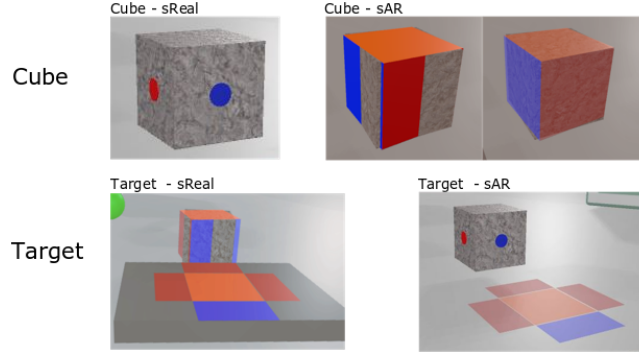
In our VR simulation, some virtual objects played the role of real objects in while others represented AR objects (*i.e.* virtual objects inserted into a real scene). As a consequence the virtual environment consisted of a set of “real objects”, or simulated Real (sReal) objects, and AR objects, or simulated AR (sAR) objects. The sReal objects could also be augmented by sAR objects.

Additionally, we could also manipulate the virtual AR FOV in the simulation. This allowed the simulation of both a device-based and of an HMD-based AR environment by adapting the size of a virtual window with a slightly different color. In our experiments, the virtual AR FOV covered  $\sim 90\%$  of the VR HMD’s FOV. We chose this value so that users would not be disturbed in their interaction with the sAR content. Fig. 1 (bottom) shows two first person views of the virtual environment where sAR and sReal objects co-exist.

The following subsections detail the different simulated conditions in terms of registration errors and AR content presentation. Only sAR objects were affected by registration errors (see Fig. 2) and sAR visualization conditions (see Fig. 3).

#### 3.1 Registration Errors

In order to study how registration errors would affect user interaction and performance in an AR pick-and-place task, we propose to simulate two different kinds of registration errors (see Fig. 2): constant and irregular. For ours experiments, we make the choice to focus the analysis on errors with only a rotation of the sAR content with respect of the sReal content around the Y-axis. Those errors are called rotational errors. The choice is motivated by the simplicity of the error that has only one degree of freedom. Translation errors are not in the scope of our study.



**Fig. 1.** Depiction of the different simulated objects (Top - cube; Bottom - target). The depiction is either sReal (Left) or sAR (Right). When the cube is sReal and the target is sAR, colored dots are painted on the sReal cube to indicate the orientation of the cube. When there is a sReal cube with its sAR cube and a sReal target, the sReal cube has no colored dot. The orientation is given by the sAR cube. The sAR objects have different conditions of visualization (transparency condition and opacity condition).

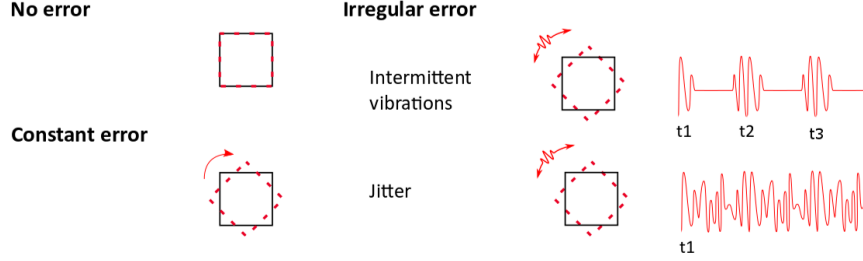
**Constant registration errors** introduce a constant misalignment between sAR and sReal objects (either in position or orientation). In this condition, no matter how the user manipulates the sReal cube, its sAR counterpart is always misaligned by the same amount either in position or rotation. We do not use the term “Static” registration errors proposed by Azuma [3] (that only happen when the user’s viewpoint does not move) since in VR we can simulate a constant error even with a dynamic environment (objects in motion or user’s changing his/her viewpoint). For simplicity, and due to the nature of the pick-and-place task, in order to avoid inter-penetrations we only considered rotation errors on the vertical axis (Y-axis). sAR content is rotated on its own Y-axis with respect to the sReal object.

**Irregular registration errors** are dynamic and variable over time. There are two kinds of such errors that we chose to name *jitter* and *intermittent vibrations*. Here again we chose not to use the term “Dynamic” [3] since the registration errors we simulate are not due to latency or delay but we rather wanted to study the effect on a non-constant registration error in a dynamic environment.

*Intermittent vibrations* are defined as discrete vibrations of the sAR object. More specifically, a pulse is generated on the sAR object’s Y-axis for a certain amount of time (e.g., for 0.1s every 1s). The pulse follows a noise function  $f$  (see Equation 1) with  $x_1, x_2$  being two real values  $\in [0, 1]$ , and  $k$  a real positive constant which determines the maximum absolute rotation error (in radians).

$$f(x_1, x_2) = \frac{\sin(2\pi x_2) * \sqrt{-2 \log(x_1)}}{3k} \quad (1)$$

On the other hand, “*jitter*” refers to a continuous vibration of the sAR object around its Y-axis. The “*jitter*” is generated continuously for a certain amount



**Fig. 2.** Simulation of AR rotational errors. A black solid square (resp. red dotted) represents a sReal (resp. sAR) object. Top-Left: no rotational error. sAR and sReal objects are perfectly aligned. Bottom-Left: constant error. The angle between sReal and sAR objects is constant over time. Right: two states, intermittent vibrations (rotational vibration that occurs from time to time, Top-Right) and jitter (rotational vibration around a position that constantly occurs, Bottom-Right).

of time and follows the same function  $f$  as *intermittent vibrations*. We refer the reader to the supplementary material which illustrates the intermittent vibrations and jitter conditions.

### 3.2 AR Content Presentation

The second aspect of the simulation we wanted to evaluate was the presentation of the AR content. Although numerous rendering styles could have been considered, we decided to focus on transparency of the AR content and occlusion management (whether or not AR objects were correctly occluded by real objects) of sAR objects, as illustrated in Fig. 3. In particular the two considered conditions were:

1. “Transparency” of the sAR content. A transparency of 100% makes the sAR object invisible while a transparency of 0% makes it completely opaque.
2. “Occlusion” management. This allows us to control whether the sAR content was properly (Occlusion ON) or not (Occlusion OFF) occluded by the sReal content.

## 4 Experiment 1: Analyzing the Effect of Constant Registration Errors in User Behaviour

The goal of the first experiment was to explore how registration errors influence the perception of co-existing sAR and sReal objects. In particular, the experiment aimed at determining which object is used as a *reference* (sAR or sReal) or in other words, which content drives the users’ actions. Furthermore, as different AR visualizations might play an important role on users’ behaviour, different sAR visualizations were evaluated.



### 4.1 Apparatus and Participants

Experiments were performed using an HMD (HTC Vive) and users were able to interact with the environment through a controller (an HTC Controller). During the experiment, the users were comfortably seated in an office chair. The experiment was conducted using the platform described in Sec. 3 which was implemented in Unity (5.6).

Twelve right-handed users (1 female, 11 males) participated in this experiment (age:  $M = 22.83$ ;  $SD = 2.66$ ). All participants had previous experience with AR and/or VR.

### 4.2 Experimental Protocol

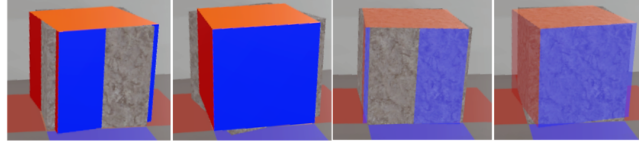
Upon arrival participants read and signed a consent form which briefly described the experiment and their rights. The consent form did not provide any information that could bias the users during the experiment. At the end of the experiment additional information was provided regarding the real purpose of the experiment. Users were told that they should consider the VR environment as their reality (*i.e.* with sReal objects) and that the AR content (*i.e.* sAR objects) which is displayed in the brown window is the AR. In addition, users had to fill out a pre-experiment questionnaire to gather background information (*e.g.* age, VR and AR experience, headset experience, laterality, visual impairment). Once users fully understood the experimental task, the experimenter equipped them with the HMD and they were immersed in the virtual experimentation room.

The experimental task was a pick-and-place. Users had to pick a sReal cube augmented with an sAR cube (see Fig. 4) and place it at the center of a colored target (also a sReal object). Users had to orient the cube to match the color code, (*e.g.* the blue face of the sAR cube should face the blue marker on the target, etc., see Fig. 4). The field of view of the AR window was enough to ensure that all cubes fit the AR display. Participants were asked to be as precise as possible and no indication was given regarding which cube (sAR or sReal) had to be aligned on the target. Users performed several trials grouped in four blocks (see Sec. 4.3).

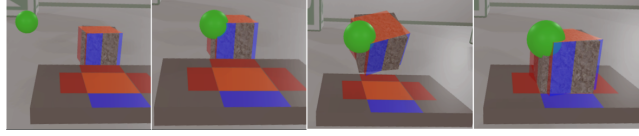
The users' virtual hand was represented as a green sphere to let them focus on the manipulation and not on their appearance. The green sphere was controlled with the HTC Vive controller (the trigger enabled to grab the sReal cube). Once users were pleased with the location of the cube, they had to signal the end of the trial by pressing the touchpad of the Vive controller. Basic physical simulation capabilities were enabled (gravity and collision detection with the virtual table). At the end of each block, participants had to fill out a questionnaire to gather their subjective impressions.

### 4.3 Experimental Design

In order to assess the effect of registration errors, we artificially considered a different range of constant registration errors. In particular, we considered seven



**Fig. 3.** AR presentation conditions. Each row corresponds to a *transparency* condition (50% and 0%). Each column corresponds to an *occlusion* condition (precise depth sorting vs. sAR always on top).

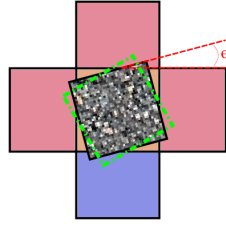


**Fig. 4.** Depiction of the first experiment's task. The user, using a virtual hand metaphor (see Sec. 4.2) had to pick and place the sAR/sReal cube on a target. Only the sAR cube was displayed in the AR overlay.

rotational mismatches  $\{15, 10, 5, 0, -5, -10, -15\}^\circ$ . In addition, we also considered four different visualizations for sAR objects. The visualization was by two independent variables (*transparency* and *occlusion*) with two levels each (see Figure 3) in order to analyze whether they biased the relationship between sAR and sReal content. The *transparency* of the sAR content was set to either 0% or 50%. *Occlusion* of the sAR content also had two options: ON (the sAR content is always visible) or OFF (the sAR content is correctly occluded by sReal objects depending on which object is closer to the user's viewpoint). To sum up, the experiment had a  $7 \times 2 \times 2$  factorial design with 4 repetitions for each condition, resulting in a total of 112 trials. The four combinations of transparency and occlusion were split into four blocks and counterbalanced using a Latin-Square design. For each block (28 trials), the order of the registration mismatch was randomized.

The dependent variables were the task-completion time (s), the position accuracy (cm) and the rotational accuracy (degrees). The task-completion time was measured from the moment the user grabs the cube until he/she validates the placement. The position accuracy was computed as the distance between the center of the sReal cube (sReal and sAR cubes shared the same center position) with respect to the center of the perfect placement position. The rotation accuracy was computed as the minimum angle between the sReal and sAR cubes with respect to the perfect placement rotation (see Figure 5). Participants had no explicit instruction whether they have to align the sReal or the sAR cube, considering the minimum provides a more insightful result. Positional and rotational accuracies are measured for each trial upon user validation of the placement.

Regarding subjective information, participants had to fill out a questionnaire after each block and at the end of the experiment. The aim of the after-block



**Fig. 5.** Depiction of how rotational error  $\theta_e$  is calculated. The blue, orange and red cross is the target. The textured square is the sReal cube. The green dotted line square is the sAR cube. In this example, the error is the angle between the sReal cube and the sReal target and not between the sAR cube and the sReal target because the angle of the error between the two sReal objects is the lowest.

questionnaires was to observe the progression of the state of the participants (i.e., tiredness, visual discomfort and task difficulty). A 7-point Likert scale was used where 1 meant “not at all” and 7 “meant completely”. The final questionnaire was designed to understand the behaviour of the participants and their impression on their performance. Therefore, the participants answered which reference they chose and if they placed the cube in relation to the sAR content or in relation to the sReal content. Then participants quantified the variability of the choice about the picked reference. Additionally, the participants evaluated on a scale of one (never) to seven (always) how often their choice changed. Finally, participants ranked the 4 conditions from their most preferred to their least preferred one. In the final questionnaire, we use the term “AR cube” to refer to the sAR cube and the term “real cube” to refer to the sReal one. Given this experimental design, we hypothesized that:

- H1.1** Smaller registration errors will result in lower rotation accuracy. The bigger is the mismatch the lower will be the ambiguity.
- H1.2** When the sAR content is semitransparent (*i.e.* Transparency 50%) the task completion time will be higher. Seeing sAR and sReal contents at the same time can be distracting.
- H1.3** When the sAR content is displayed on top (*i.e.* Occlusion ON) it will result in lower task completion times due to the same reasoning as H1.2.
- H1.4** As only rotational accuracy is considered and the sReal and sAR cubes share the same center position, no significant difference in positional accuracy is expected.

#### 4.4 Results

Regarding the objective measures, the analysis of the data (ANOVA analysis with post-hoc Tukey tests ( $\alpha > 0.05$ )) showed that the different independent variables had no impact on the dependent variables. All comparisons were not-significant. For the sake of simplicity we only report the two-way ANOVA anal-

ysis for the factors *Transparency* and *Occlusion*, pooling the data from the different levels of rotation mismatch. Anderson Darling tests were performed to ensure the normal distribution of the data.

**Rotation Accuracy.** No significant main effect was found for *Transparency* ( $F_{1,11} = 0.77, p = 0.399, \eta_p^2 = 0.06$ ) nor for *Occlusion* ( $F_{1,11} = 0.83, p = 0.382, \eta_p^2 = 0.07$ ). In overall, participants were extremely precise  $M = -0.12$  deg;  $SD = 0.93$  deg no matter the *reference* and the condition. In addition, as previously stated, the amount of registration mismatch did not influenced rotation accuracy. Thus, results do not support **H1.1**.

**Task-Completion Time.** No significant main effect was found for *Transparency* ( $F_{1,11} = 1.27, p = 0.284, \eta_p^2 = 0.10$ ) nor for *Occlusion* ( $F_{1,11} = 2.83, p = 0.121, \eta_p^2 = 0.20$ ). In overall, participants required  $M = 10.086s$ ;  $SD = 5.927s$  to complete the task. Although the mean task completion time was considerably high considering the nature of the task, we have to note that participants were requested to perform the task as accurately as possible. Thus, we cannot accept neither **H1.2** nor **H1.3**.

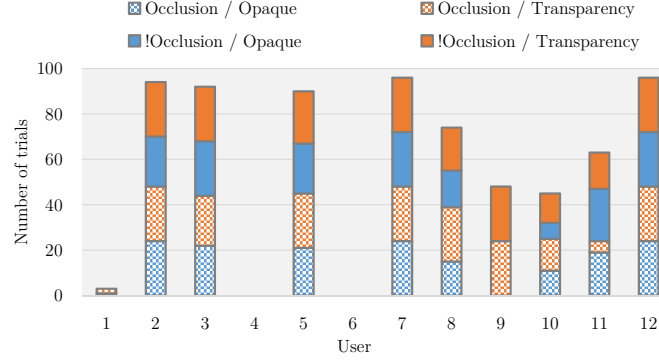
**Position Accuracy.** No significant main effect was found for *Transparency* ( $F_{1,11} = 2.31, p = 0.156, \eta_p^2 = 0.17$ ) nor for *Occlusion* ( $F_{1,11} = 0.36, p = 0.560, \eta_p^2 = 0.03$ ). In overall, participants were extremely accurate  $M = 0.15cm$ ;  $SD = 0.07cm$  no matter the visual condition. These results support **H1.4**. Yet, if any difference exists, considering the level of accuracy, it would be non-relevant.

#### 4.5 Reference Cube

After analyzing users' behaviors, we observed that the majority of participants had a different yet consistent behaviour during the entire experiment. In particular, we computed the object (sAR or sReal cube) that was considered as the *reference* by the user. The *reference* cube is the cube (sAR or sReal) that minimizes the error to the target orientation at the moment the user validates the trial. Fig. 6 shows the amount of time the sReal cube was chosen as *reference* for each user and for each condition. Cases in which the rotation mismatch was 0 were not considered as they do not provide any information. Six out of twelve users considered the sReal cube as the main *reference* (users 2, 3, 5, 7, 8, and 12), three users considered the sAR cube as the main *reference* (users 1, 4 and 6), one user (user 10) did the task considering that both cubes were the *reference* (minimizing the error between both cubes) and two users (9 and 11) changed *reference* in the middle of the experiment.

#### 4.6 Subjective Questionnaires

**After-Block Questionnaires** Concerning the task difficulty, users considered the task relatively easy  $M = 2.54$ ;  $SD = 1.18$  no matter the visual condition. Interestingly, small misalignments (e.g.,  $\pm 5$  deg) were reported to increase the difficulty of the task. Three users reported that "With small misalignments it was harder". Moreover, some users felt confused about the task: "I had the impression of making wrong choices" or "I was confused and a little bit disturbed



**Fig. 6.** Bar plot showing the number of trials in which for each participant the sReal cube was calculated as the *reference* object.

because of misalignments”. Concerning visual discomfort, it remained constant during the entire experiment  $M = 2.90$ ;  $SD = 1.43$ , which shows that the task and the set-up do not cause strong visual discomforts. Finally regarding tiredness, the mean value is almost constant and below 3  $M = 2.85$ ;  $SD = 0.92$ .

**Post-Experiment Questionnaire** Nine users (75%) stated that they picked a *reference* and kept it during all the experiment, whereas three users (25%) reported to change their *reference*. Six users out of twelve (50%) chose the sReal cube, three chose the sAR cube and the others changed during the manipulation. Only one user picked all the time an imaginary cube placed at the mid-position between the sAR cube and the sReal cube. These results confirm results from the previous section. Globally users pick a reference and keep it during all the task. According to users, the choice of the *reference* was made because the task seems easier for them with their choice of *reference*: “With AR it was more practical”. Finally, regarding user preferences on sAR visualization, the condition in which the sAR content was always visible and opaque was the less preferred one. Seven out of twelve participants (58.33%) answered that it was their less preferred choice. Yet, results do not show a clear user preference.

#### 4.7 Discussion

The main outcome of this experiment is that constant registration errors did not alter the way users performed the task nor their performance. The analysis of users’ behaviors as well as questionnaires showed that participants have a strong preference when choosing the main *reference* (either sReal or sAR) no matter the registration mismatch nor the visualization. Interestingly, while we hypothesized that the occlusion would have an influence on users’ behavior, it was not the case, nor any other visualization condition. One explanation could be that users did not need to change their *reference* (there was no significant effect on accuracy) since in all cases they were able to accurately place the chosen cube.

Another interesting finding is that there was no significant difference in task-completion time for the different sAR visualization conditions. As users were requested to be as accurate as possible, there was a moderate user variability. This could have decreased the power of the analysis. Nevertheless, we were expecting strong differences and it was not the case. Finally, the condition that we hypothesized to be the less optimal (no transparency and no occlusion) was the condition users preferred less.

## 5 Experiment 2: Analyzing the Effect of Irregular Registration Errors in User Behaviour

The results of the previous experiment shown that when constant registration errors occur, they do not have a strong influence on users' behavior, as the *reference* chosen by users rarely changed during the experiment. In order to explore in depth users' behavior, we performed a follow up experiment in which we explored the effects of irregular registration errors. The goal of this second experiment was to analyze the tolerance to irregular registration errors and measure if there was a threshold that triggers a shift in the selected *reference*. The experimental protocol and the apparatus of the experiment were the same as the first experiment. Upon their arrival, participants read and signed a consent form which briefly described the experiment and their rights. After users fully understand the experimental task, the experimenter equipped them with the HMD.

### 5.1 Participants

Twelve right-handed users (2 females, 10 males) participated in this follow up experiment ( $M = 23.25$ ;  $SD = 2.22$ ). Seven never had any experience with VR, five never had any experience with AR and six never used a HMD before.

### 5.2 Experimental Design

Participants had to perform the same pick-and-place as in the first experiment task while varying the intensity of the registration error. We considered five different levels of intensity, defined by the amplitude of the vibration (the  $k$  parameter of Eq.1) and the time (discrete vs. continuous). The registration errors were presented by increasing intensity:

- L1** No Error: Perfect registration between sAR and sReal objects.
- L2** Low intensity intermittent vibration: a pulse of 0.11s is generated every 1s with  $k = 0.3rad$ .
- L3** High intensity intermittent vibration: a pulse of 0.11s is generated every 0.5s with  $k = 0.3rad$ .
- L4** Low intensity jitter: a continuous vibration is generated with  $k = 0.05rad$ .
- L5** High intensity jitter: a continuous vibration is generated with  $k = 0.3rad$ .

Only one sAR visualization was considered which corresponded with the most preferred condition of the first experiment (50% transparency and Occlusion OFF). For each intensity level, participants performed 5 repetitions of the pick-and-place task. To better account for users' choice of *reference*, constant registration errors were also introduced for each intensity level (four repetitions randomly chosen between  $\{-10, 10\}^\circ$  and one repetition had no constant registration error). The target is a sReal object once again. We were only interested in measuring the *reference* chosen for each trial, which we determined following the same approach as in the first experiment, choosing the object (sReal or sAR) which minimized the placement accuracy. We hypothesized that:

**H2.1** As the intensity of the irregular registration error increases the users will have the tendency to shift their *reference* towards the sReal object.

### 5.3 Results

Similar to the first experiment, the ANOVA analysis did not show any significant differences among, participants were extremely precise in the placement task (position accuracy:  $M = 1.65mm$ ;  $SD = 0.9mm$  and rotation accuracy:  $M = 0.86\text{ deg}$ ;  $SD = 0.95\text{ deg}$ ). Also, participants followed a similar behavior: once participants chose the *reference* (typically in the first few trials), they tend to keep it during the entire experiment (see Fig. 7 right). Six users had a clear tendency towards choosing the sAR cube as *reference* while the remaining six users chose the sReal cube as *reference*.

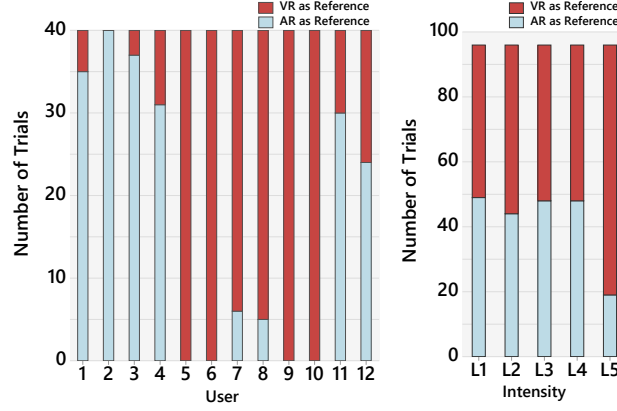
Regarding the strength of registration errors (see Fig. 7), we observed a change in behaviour for some users who chose the sAR reference in the L5 condition. Seven out of twelve users (58.33%) explained that they picked a *reference* and kept it during all the experiment. Four users (33.33%) said they changed their *reference*, explicitly mentioning that the change occurred at the beginning of the last condition (L5, high intensity jitter). Thus, in the presence of strong registration errors, and due to the inability to perform the task, the *reference* object was changed to the more reliable *reference*.

### 5.4 Discussion

Results of the second experiment tend to reinforce the fact that users make a strong reference choice at the beginning of the experience. Even with the addition of irregular registration errors, participants are reluctant to change their *reference*. However, if the error can discourage the completion of the task, they have the tendency to shift towards a more stable *reference*. In conclusion, **H2.1** is validated only in the presence of high intensity of jitter.

## 6 Experiment 3: Analyzing the Effects of Irregular Registration Errors on User Performance

One common element of the first two experiments is that sAR content augmented sReal objects (*i.e.* a sAR cube was displayed on top of a sReal cube). However,



**Fig. 7.** Choice of *reference* during the second experiment. (Left) Bar plot showing the choice of *reference* for each participant. (Right) Bar plot showing the choice for each condition.

in AR applications pure virtual objects (*i.e.* without any relation to real objects) are commonly used. In such situations, the pure virtual object will always be considered as the *reference* as there is no ambiguity. This third experiment explores the impact of irregular registration errors on user performance when the task is driven just by AR content. Can users perform effectively and efficiently in presence of irregular registration errors in such context? The participants in this experiment were the same group which participated in the second experiment.

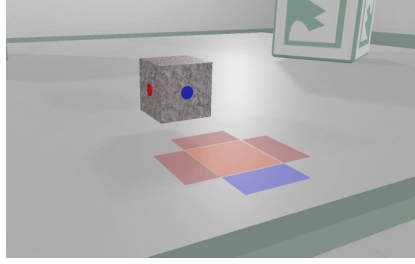
### 6.1 Experimental Protocol

Upon their arrival, participants read and signed a consent form which briefly described the experiment and their rights. After users fully understand the experimental task, the experimenter equipped them with the HMD. The task was a variation of the pick-and-place task of the first experiment. In this case, the sReal cube was not augmented, and the target location was a sAR object. Users had to pick the sReal cube and place it at the location indicated by the sAR target. sReal colored stickers were placed on the sReal cube to display the correspondence with the sAR target (see Fig. 8). At the end of the experiment participants had to fill out a questionnaire to gather their subjective impressions.

### 6.2 Experimental Design

The experiment had two independent variables which defined the *intensity* of the irregular registration error and the sAR *visualization*. Only the sAR target is affected by vibrations. Regarding registration errors, we considered (and simulated) the following five different levels of intensity (sorted by increasing intensity):





**Fig. 8.** Task description of the third experiment. The user had to place the sReal cube onto the sAR target by aligning the colored stickers placed on the sReal cube with the sAR target's colors.

- L1** No Error is applied to the target.
- L2** Low intensity intermittent vibration: a pulse of 0.11s is generated every 2s with  $k = 0.33rad$ .
- L3** Medium intensity intermittent vibration: a pulse of 0.11s is generated every 1s with  $k = 0.33rad$ .
- L4** High intensity intermittent vibration: a pulse of 0.11s is generated every 0.5s with  $k = 0.33rad$ .
- L5** High intensity jitter: a continuous vibration is generated with  $k = 0.33rad$ .

Regarding the sAR visualization, we only considered the occlusion condition with two levels, either the sAR content is occluded (Occlusion ON) or not (Occlusion OFF) by the real content. For both conditions the sAR target was displayed with a 50% transparency. The experiment had a 5x2 factorial design, each combination being repeated six times (60 trials per participant). To avoid ordering effects, the experiment was divided in four blocks in which the order of the sAR visualization was counterbalanced. For each block order the intensity level was randomized. The dependent variables were the task-completion time, the position accuracy and the rotation accuracy.

According to this experimental design, our hypotheses were that:

- H3.1** Occlusion OFF will have a negative impact on task-completion time.
- H3.2** Occlusion OFF will have a negative impact on position accuracy.
- H3.3** Occlusion OFF will have a negative impact on rotation accuracy.
- H3.4** as the intensity of the error increases task-completion time will increase.
- H3.5** as the intensity of the error increases position accuracy will decrease.
- H3.6** as the intensity of the error increases rotation accuracy will decrease.

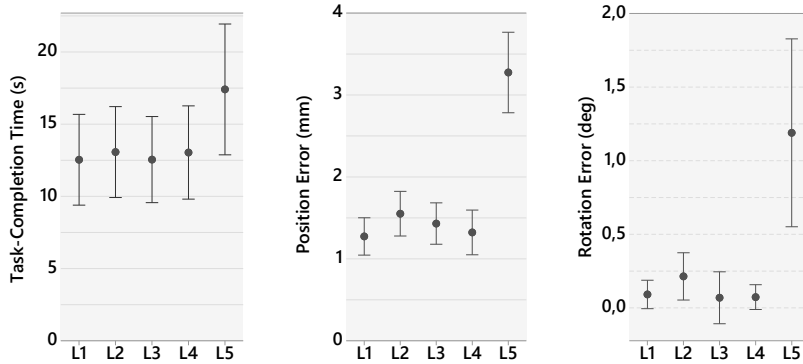
### 6.3 Results

Two-way ANOVA analysis were performed to determine the significance of registration errors and Occlusion conditions for each dependent variable. When needed, post-hoc Tukey tests were performed ( $\alpha > 0.05$ ). Anderson-Darling tests were performed to ensure normal distribution of the data.

**Task-Completion Time.** The ANOVA analysis showed a main effect on Intensity ( $F_{4,44} = 5.34, p < 0.001, \eta_p^2 = 0.33$ ), but there was no effect on Visualization ( $F_{1,11} = 0.49, p = 0.498, \eta_p^2 = 0.04$ ) nor on interaction effect ( $F_{4,44} = 0.76, p = 0.557, \eta_p^2 = 0.06$ ). Post-hoc tests showed that participants significantly required more time to perform the task for L5 than for the other conditions (see Fig. 9, left). These results do not support **H3.1** and only partially support **H3.4**.

**Position Accuracy.** The ANOVA analysis showed a main effect on Intensity ( $F_{4,44} = 47.66, p < 0.001, \eta_p^2 = 0.82$ ) and in Visualization ( $F_{1,11} = 15.16, p < 0.01, \eta_p^2 = 0.59$ ). No interaction effect was found. Post-hoc tests showed that the intensity level L5 resulted in significantly lower accuracy compared to the other four intensities (see Fig. 9, center), and that when the sAR content is not occluded by the sReal content ( $M = 1.97mm; SD = 1.05mm$ ) significantly (although slightly) decreased position accuracy compared to the opposite condition ( $M = 1.57mm; SD = 1.04mm$ ). These results support **H3.2** and only partially support **H3.5**.

**Rotational Accuracy.** The ANOVA analysis showed only a main effect for Intensity ( $F_{4,44} = 9.18, p < 0.001, \eta_p^2 = 0.46$ ). No main effect for Visualization ( $F_{1,11} = 0.34, p = 0.572, \eta_p^2 = 0.03$ ) and no interaction effects were found ( $F_{4,44} = 1.39, p = 0.300, \eta_p^2 = 0.10$ ). Post-hoc tests showed that for the Intensity variable, the L5 level induced significantly lower rotation accuracy than the other levels (see Fig. 9, right). These results do not support **H3.3** and only partially support **H3.6**.



**Fig. 9.** Confidence Intervals (95%) of the mean for each dependent variable. Left: task-completion time. Center: position accuracy. Right: rotation accuracy.

#### 6.4 Subjective Questionnaires

Eleven users out of twelve (91.66%) preferred when the sAR content was occluded by sReal objects (*i.e.* the Occlusion ON condition). Most of the users felt perturbed when the occlusion was not computed correctly: “it was easier to see

edges and borders of the cube when the AR target did not hide the real cube” and “when objects that should be behind everything appear in the foreground, it is not natural”. Regarding the amount of jitter, as in the second experiment, users only complained when the registration error did not allowed them to perform the task efficiently (i.e., the L5 level).

## 6.5 Discussion

Results of this experiment show that low to moderate levels of irregular (see Sec. 3.1) registration errors (L1 to L4) do not have a noticeable impact on user performance and accuracy. Surprisingly, no significant effects were found between levels L1 to L4. Indeed, this result shows a relatively strong tolerance of users towards registration errors. Furthermore, in terms of user preferences of the sAR visualization, although the objective results did not present any significant differences, the visualization minimizing the perceptual conflicts (i.e., correct occlusion management) was preferred.

## 7 General Discussion

In this paper, we have presented three experiments focusing on the impact on registration errors in an AR pick-and-place task. In order to ensure controlled and repeatable experimental conditions, the AR environment was simulated in VR. Data analysis has mainly focused on user behavior (how registration errors alter the way users interact) and user performance (how users can account for registration errors).

The first two experiments shown that participants chose a reference (the sAR content or the sReal content) from the very beginning and they were reluctant to change it afterwards. This choice of reference impacted their performance since it drove their actions (e.g., they either align the sReal or the sAR cube on the target). What was more interesting is that even in the presence of irregular registration errors, participants kept the chosen reference. Only when the task was compromised (e.g., when unable to perform the task) some users shifted their reference from the sAR content to the sReal content. Nevertheless, several users still kept the sAR content as reference when the task was compromised (L5 condition). These results show that an AR interface designer should unambiguously define which objects serve as reference to ensure that all users exhibit the same behavior.

In contrast, the third experiment focused on the impact of irregular registration errors on user performance, and more importantly on how users tolerate them. Results showed that, for the tested errors, the tolerance was high, and that errors of low to medium intensity did not significantly altered users’ performance. These results show that users had a strong adaptation to adverse situations.

Along the different experiments, a secondary goal was to assess the impact of how AR content is presented. Simulating AR in VR enabled us to test a

wide range of visualizations which were difficult or almost impossible to achieve with current AR systems. The preference results are not ground-breaking: users prefer a perceptually-correct integration (e.g., correct occlusion management) as they provide a non-ambiguous layout. Yet, when occlusion management cannot be achieved (e.g., due to technical limitations) semi-transparent visualizations should be considered as they minimize perceptual conflicts. Regarding the impact on performance, interestingly, we did not find any significant difference between visualizations (i.e., transparency and occlusion). This result shows that for all conditions users were able to recover enough information from the sAR and sReal contents to successfully perform the task (e.g., a fully visible edge or corner of the cube).

## 8 Conclusion

Although registration algorithms are continuously being perfected, inconsistencies between sReal and sAR contents are still a major issue in AR applications. Results obtained our three user studies show that visual inconsistencies generated by registration errors can alter users' behavior as they require a subjective interpretation. According to their interpretation, the outcome of the interaction task might vary. This effect is clearly visible in the first and second experiments as the reference chosen (the sReal cube or sAR cube) determines the outcome of the task. In the same direction, visualization strategies should minimize inconsistencies between the AR and the real contents. However, if this is not possible, the interface designer must ensure that the visualization provides enough information to perform the task effectively.

The presented studies highlight the potential of evaluating registration errors in controlled VR simulations. Nevertheless, this paper has only focused on a particular subset of registration errors. Further studies should be conducted to evaluate other types of registration errors such as depth or lateral errors. This knowledge is paramount to better design AR applications and ensure they enable users to interact effectively and efficiently.

## References

1. Azimi, E., Qian, L., Kazanzides, P., Navab, N.: Robust optical see-through head-mounted display calibration: Taking anisotropic nature of user interaction errors into account. In: *IEEE Virtual Reality*. pp. 219–220 (2017). <https://doi.org/10.1109/VR.2017.7892255>
2. Azuma, R., Bishop, G.: Improving static and dynamic registration in an optical see-through hmd. In: *ACM Annual Conference on Computer Graphics and Interactive Techniques*. pp. 197–204 (1994). <https://doi.org/10.1145/192161.192199>
3. Azuma, R.T.: A survey of augmented reality. *Presence: Teleoperators and virtual environments* **6**(4), 355–385 (1997). <https://doi.org/10.1162/pres.1997.6.4.355>
4. Baričević, D., Lee, C., Turk, M., Höllerer, T., Bowman, D.A.: A Hand-held AR Magic Lens with User-perspective Rendering. In: *IEEE International Symposium on Mixed and Augmented Reality*. pp. 197–206 (2012). <https://doi.org/10.1109/ISMAR.2012.6402557>

5. Billinghurst, M., Clark, A., Lee, G.: A Survey of Augmented Reality. *Foundations and Trends® in Human-Computer Interaction* **8**(2-3), 73–272 (2015). <https://doi.org/10.1561/11000000049>
6. Ellis, S.R., Breant, F., Manges, B., Jacoby, R., Adelstein, B.D.: Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency. In: *IEEE Virtual Reality Annual International Symposium*. pp. 138–145 (1997). <https://doi.org/10.1109/VRAIS.1997.583063>
7. Ellis, S.R., Menges, B.M.: Localization of virtual objects in the near visual field. *Human factors* **40**(3), 415–431 (1998)
8. Furmanski, C., Azuma, R., Daily, M.: Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In: *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*. pp. 215–320. IEEE (2002)
9. Holloway, R.L.: Registration error analysis for augmented reality. *Presence: Teleoperators and Virtual Environments* **6**(4), 413–432 (1997)
10. Holloway, R.L.: Registration errors in augmented reality systems. Ph.D. thesis, University of North Carolina at Chapel Hill (1995)
11. Klein, G., Murray, D.: Parallel tracking and mapping for small ar workspaces. In: *ACM/IEEE International Symposium on Mixed and Augmented Reality*. pp. 225–234 (2007). <https://doi.org/10.1109/ISMAR.2007.4538852>
12. Lee, C., Bonebrake, S., Bowman, D.A., Hollerer, T.: The role of latency in the validity of AR simulation. In: *IEEE Virtual Reality*. pp. 11–18 (2010). <https://doi.org/10.1109/VR.2010.5444820>
13. Lee, C., Bonebrake, S., Höllerer, T., Bowman, D.A.: A replication study testing the validity of AR simulation in VR for controlled experiments. In: *IEEE International Symposium on Mixed and Augmented Reality*. pp. 203–204 (2009). <https://doi.org/10.1109/ISMAR.2009.5336464>
14. Lee, C., Rincon, G.A., Meyer, G., Hollerer, T., Bowman, D.A.: The Effects of Visual Realism on Search Tasks in Mixed Reality Simulation. *IEEE Transactions on Visualization and Computer Graphics* **19**(4), 547–556 (2013). <https://doi.org/10.1109/TVCG.2013.41>
15. Nabiyouni, M., Scerbo, S., Bowman, D.A., Höllerer, T.: Relative Effects of Real-world and Virtual-World Latency on an Augmented Reality Training Task: An AR Simulation Experiment. *Frontiers in ICT* **3**, 13 (2017). <https://doi.org/10.3389/fict.2016.00034>
16. Ragan, E., Wilkes, C., Bowman, D.A., Höllerer, T.: Simulation of augmented reality systems in purely virtual environments. In: *IEEE Virtual Reality*. pp. 287–288 (2009). <https://doi.org/10.1109/VR.2009.4811058>
17. Ren, D., Goldschwendt, T., Chang, Y., Höllerer, T.: Evaluating wide-field-of-view augmented reality with mixed reality simulation. In: *IEEE Virtual Reality*. pp. 93–102 (2016). <https://doi.org/10.1109/VR.2016.7504692>
18. Renner, P., Pfeiffer, T.: Evaluation of attention guiding techniques for augmented reality-based assistance in picking and assembly tasks. In: *ACM International Conference on Intelligent User Interfaces Companion*. pp. 89–92 (2017). <https://doi.org/10.1145/3030024.3040987>
19. Zhou, F., Duh, H.B.L., Billinghurst, M.: Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In: *IEEE/ACM International Symposium on Mixed and Augmented Reality*. pp. 193–202 (2008). <https://doi.org/10.1109/ISMAR.2008.4637362>