

## Article

# Cross-Device Augmented Reality Annotations Method for Asynchronous Collaboration in Unprepared Environments

Inma García-Pereira <sup>1,\*</sup> , Pablo Casanova-Salas <sup>2</sup> , Jesús Gimeno <sup>1</sup> , Pedro Morillo <sup>1</sup>  and Dirk Reiners <sup>3</sup>

<sup>1</sup> Computer Science Department, University of Valencia, 46100 Burjassot, Valencia, Spain; [jesus.gimeno@uv.es](mailto:jesus.gimeno@uv.es) (J.G.); [pedro.morillo@uv.es](mailto:pedro.morillo@uv.es) (P.M.)

<sup>2</sup> Institute on Robotics and Information and Communication Technologies (IRTIC), University of Valencia, 46980 Paterna, Valencia, Spain; [pablo.casanova@uv.es](mailto:pablo.casanova@uv.es)

<sup>3</sup> Computer Science Department, University of Central Florida (UCF), Orlando, FL 32816, USA; [dirk.reiners@ucf.edu](mailto:dirk.reiners@ucf.edu)

\* Correspondence: [inmaculada.garcia-pereira@uv.es](mailto:inmaculada.garcia-pereira@uv.es)

**Abstract:** Augmented Reality (AR) annotations are a powerful way of communication when collaborators cannot be present at the same time in a given environment. However, this situation presents several challenges, for example: how to record the AR annotations for later consumption, how to align virtual and real world in unprepared environments or how to offer the annotations to users with different AR devices. In this paper we present a cross-device AR annotation method that allows users to create and display annotations asynchronously in environments without the need for prior preparation (AR markers, point cloud capture, etc.). This is achieved through an easy user-assisted calibration process and a data model that allows any type of annotation to be stored on any device. The experimental study carried out with 40 participants has verified our two hypotheses: we are able to visualize AR annotations in indoor environments without prior preparation regardless of the device used and the overall usability of the system is satisfactory.

**Keywords:** augmented reality; annotations; computer supported collaborative work; user-centered evaluation; human–computer interaction



**Citation:** García-Pereira, I.; Casanova-Salas, P.; Gimeno, J.; Morillo, P.; Reiners, D. Cross-Device Augmented Reality Annotations Method for Asynchronous Collaboration in Unprepared Environments. *Information* **2021**, *12*, 519. <https://doi.org/10.3390/info12120519>

Academic Editors: Ramon Fabregat, Jorge Bacca-Acosta and N.D. Duque-Mendez

Received: 5 October 2021

Accepted: 12 December 2021

Published: 14 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Collaboration is a promising area of research in the field of Augmented Reality (AR). One of the most common ways of classifying Computer Supported Collaborative Work (CSCW) is by using the space-time matrix developed in [1] and revised in [2]. It differentiates between four types of interaction, depending on whether or not the collaborators share the same physical space and whether the work is synchronous or asynchronous. Most CSCW systems using AR (hereafter AR-CSCW) focus on synchronous applications, both remote and face-to-face. In contrast, the problem of asynchronous collaboration is particularly underexplored [3,4]. Moreover, in recent years, distributed systems have received the most attention, due to the fact that the most studied scenario is that of remote experts assisting local users [5]. However, AR allows many other functionalities such as, for example, placing annotations to convey information about an item or location when producer and consumer cannot be present at the same time.

In this context, some studies raise challenges to overcome in the field of asynchronous collaboration through AR, among which we can highlight the in situ creation of georeferenced annotations [6] and the retention of annotations for use at a later time [3]. These considerations imply achieving a precise registration of virtual information in the physical world. To solve this problem, it is usually necessary to prepare the environment and the system in advance, as markers have to be placed or point clouds have to be recorded. If a collaborative AR system is to be used in unprepared environments, alternative anchoring modes have to be found. In addition, Irlitti et al. [3] point out the importance of considering

the technological asymmetry between users. Currently available AR systems differ widely from each other so that applications are built to run on only one type of device and migration to another is often very costly. Something as seemingly simple as communicating the position of an object between devices becomes a challenge [7].

All this makes it difficult to find asynchronous collaborative AR systems that can be used in unprepared environments, both indoors and outdoors, and with different devices. This is where we position our research. Our goal is to develop a collaborative tool for creating and reading AR annotations that is as universal as possible, which can be used anywhere and with any device. This has two fundamental implications: the inability to depend on prior preparation of the environment, so that no markers of any kind can be used, and not to depend on the implementation of a library that limits the type of device that can be used. This will require the collaboration of the user who, through a previous assisted step, will introduce the necessary anchor points for the correct alignment of the virtual information in the real world.

Although our method can potentially be applied in both outdoor and indoor environments, our experimental study has been developed in an indoor space. Thus, our primary hypothesis is that we are able to visualise AR annotations in indoor environments without prior preparation regardless of the device used. Our secondary hypothesis is that the overall usability of the system is satisfactory. In order to verify these hypotheses, an experimental study was run with the participation of 40 users. All of them used the application to find annotated objects in a room using AR. The experiment was performed with two devices (a high-specs smartphone and a low-specs smartphone), objective data were collected during the experiment and a subjective questionnaire was carried out at the end. The results obtained show that our hypotheses are correct.

To the best of our knowledge, this is the first work that performs an experimental study of an asynchronous, cross-device AR-CSCW in unprepared environments. This analysis will help researchers develop universal AR systems that can be used for collaboration between users in any environment and regardless of their device. This facilitates the development of novel cross-device AR applications in a wide variety of application fields.

The rest of the article is organized as follows. Section 2 reviews related work in the area of AR annotations, asynchronous and co-located AR-CSCWs and technological asymmetry. Section 3 describes the developed system. Next, Section 4 details the experimental study and in Section 5, the results of this experiment are presented and discussed. Finally, Section 6 summarizes the conclusions of the research and outlines the future work.

## 2. Related Work

One of the great advantages of AR technologies is that they have the ability to contextualise and locate virtual information in relation to the real world. In this sense, annotations are one of the most common uses of AR as they are a powerful way to provide users with more information about the world around them [8]. Wither et al. [8] define an AR annotation as virtual information that describes in some way an existing object and that is registered to it. This means that the virtual information of an AR annotation can take any format (text, image, sound, 3D models, etc.) and that the annotated element can be either a physical object, an area of the environment or a single point in space. This part of the physical world to which the virtual information is connected is called anchoring. Thanks to this geographical link, it is possible to communicate information about specific elements in the real world. This makes AR annotations a fundamental element of CSCWs and especially of asynchronous ones, when producer and consumer visit the same spaces at different points in time.

Although there is potentially a wide variety of application fields for asynchronous and co-located AR-CSCWs, to the best of our knowledge little work explores this context. We found some examples in the field of industry and construction [9,10], others that developed applications to geolocate points of interest [11], and specific systems for games

or sports [12]. The vast majority of the works found present ad hoc systems developed for a single type of device and in known environments.

As we mentioned above, one of the great challenges to overcome in the field of asynchronous AR collaboration is the in situ creation of geolocated annotations. Depending on the method used to position virtual information in the real world, a distinction is made between marker-based systems or markerless systems. However, both require some form of marker. Marker-based systems use artificial markers that must be placed in the real environment to be tracked in order to calculate the position and orientation of the virtual information. Markerless systems, on the other hand, use almost any part of the real environment as a marker that can be tracked to position the virtual information. Although they do not need artificial markers, they do require the creation of a point cloud that describes the environment, that is usually created in real time while the user is moving. In this case, to share the virtual information, the point cloud that links it to the real world must also be shared [13]. Other markerless systems take advantage of known elements that they know will be present in the environment, for example standardised symbols of electrical circuits [14] or geometric parameters of piano keyboards [15]. There are other works that, although they present systems that can be used in any environment, actually need some condition to work properly, such as pervasive point lights [16], street name signs [17] or table surfaces [18].

However, when an AR system has to be used in unprepared environments, none of these techniques can be used and alternatives have to be found. An unprepared environment is one in which the system has to be used without being able to place any kind of marker beforehand and without a previous point cloud capture. In the scientific literature, this term is used for both outdoor [19,20] and indoor [21,22] environments. This is, to the best of our knowledge, an unexploited field to date. An example is found in [23], where the user who makes the AR annotations creates a panoramic map with his mobile device of the environment where he wants to place the virtual information. This data is stored together with his GPS position. Later, another user can view these annotations by approaching the area using the GPS data and creating a new panorama.

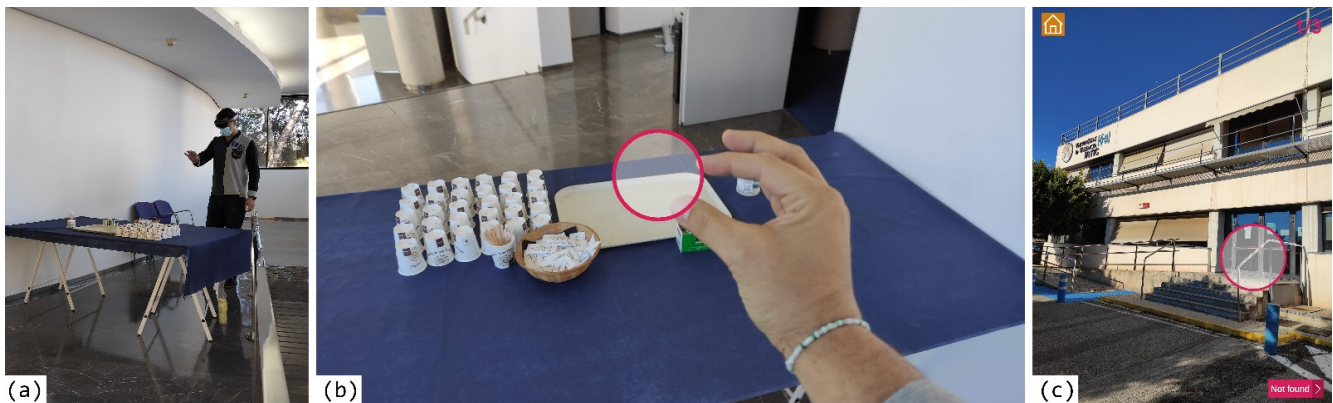
Another handicap to overcome in CSCW-AR is, as mentioned above, technological asymmetry. One of the fields where this idea has been most studied is Mixed Reality (MR), where AR and Virtual Reality (VR) devices are used in the same system [24–26]. In the field of AR alone, examples are scarcer. In [27,28], systems are presented in which different devices are used but each one of them has a different functionality and must be used synchronously. In [7,29], however, all the functionalities of the AR system can be used with the different devices considered, but again, users work simultaneously. One of the challenges posed by Speicher et al. [7] is to connect the coordinate systems and anchors of the different devices. To overcome this requirement, an initial manual calibration process by the users is required. Another example is the SDK provided by Microsoft in the Azure toolkit that supports the storage and sharing of AR annotations and other types of virtual content using the Azure Cloud services [30]. In this case, all the devices must use this SDK, not only to share information, but also to perform a parallel tracking process. In other words, all devices are using the same tracking to compute the location of the AR annotations, so there is no real technological asymmetry.

In this paper, we present a cross-device AR annotation technique that allows creating and visualising AR annotations with different devices asynchronously in unprepared indoor environment. The developed system uses a user-assisted calibration method that has proven to be fast, efficient, easy to learn and simple to use. In addition, all data related to AR annotations (virtual information and anchors) are stored using a data model independent of the device used to create or read them. Both users with high-specs devices and users with low-specs devices are able to visualise AR annotations present in the environment without any marker, without pre-loading point clouds into the system and without the need to use image recognition. The proposed method has been evaluated with 40 real users in an experimental study.

### 3. System Description

In order to validate our first hypothesis (we are able to visualise AR annotations in indoor environments without prior preparation regardless of the device used), we have developed an application that can be deployed on all types of devices with minimal variations. This implies that it is independent of the tracking system. The design of this application, as well as the calibration method and the visualisation of AR annotations, was created with our second hypothesis in mind: the overall usability of the system is satisfactory. This usability focuses on both the clarity of the graphical interface and the ease of use of the different functionalities of the system. To this end, the application has been designed to be very easy for users to use, allowing them to learn how to use it quickly.

Our AR annotation method has been tested in different environments (indoor, outdoor, wide spaces, confined spaces, etc.), with different devices (HMD and handheld devices) and for objects of different sizes. One of these scenarios was an area of about 500 m<sup>2</sup> in an outdoor parking where access gates to buildings, trees and traffic signs were annotated. HoloLens, Android smartphone and iPad were used. Figure 1c shows a screenshot of the application being used in this scenario with an iPad. Another scenario tested with the three devices mentioned above was an assembly hall and its access area. Figure 1a,b shows a user using HoloLens to select an annotation found in this scenario. In these preliminary tests, it was observed that lighting conditions affected the perception of the annotations, being more difficult to see outdoors (especially with HoloLens). This is due to the hardware's own characteristics, which are compromised in extreme lighting conditions. Otherwise, the handling of the application on the different devices and environments offered similar results. From these initial tests it was found that the designed annotation method could be suitable and to validate it we designed the system described below and the experimental study detailed in the following section.



**Figure 1.** Testing our AR annotation method in different scenarios with different devices: HoloLens (a,b) and iPad (c).

Among the different devices that allow the visualization of AR annotations (head-mounted displays, hand-held displays, projection-based displays, screen-based video see-through displays, etc.), we have selected the following two for this study: high-specs smartphone and low-specs smartphone. This selection is made based on the different location and tracking capabilities they have. In choosing these two devices, we ensure that there is an important difference between their tracking systems, as described below. In addition, we wanted to be consistent with our goal of achieving the most universal system possible, since Mobile AR Systems (MARS) are currently the most widespread AR systems on the market. This is mainly due to their low cost, ease of use and the fact that they are much less bulky than HMDs or projectors. Precisely for this reason, the experimental study carried out with these two devices is less prone to biases caused by users' lack of knowledge in the use of the technology (as could happen with Microsoft HoloLens, for example). To avoid biases related to the size of the device's screen or its processing capacity,



it has been decided to use the same hardware to simulate both devices: Xiaomi Mi MIX 2S octa-core with 6GB of RAM and a 5.99-inch screen. These two devices are described below:

- High-specs smartphone configuration. In this configuration the device relies on Google's ARCore platform in order to track its position and its surrounding environment. We have developed the app using Unity3D engine that provides an abstraction layer for AR development called ARFoundation. With this layer we can access the functionalities of both ARCore in android and ARKit in iOS devices.
- Low-specs smartphone configuration. To simulate this device all ARCore's functionalities are disabled, and tracking relies only on the gyroscope sensor present in the device. This type of tracking can only estimate the rotation of the device without 3D position information and limits the movement of the users as they only can rotate the device around them to find annotations. The accuracy of this tracking system may be highly dependent on the users' steadiness in handling the device.

### 3.1. Calibration Method

Since we want to develop a system that can be used in unprepared environments, we cannot use any kind of marker or prepare the space in any way. Prior to create any AR annotation, an initial setup is mandatory. In this step, three anchor points (which we will call "virtual anchors") must be geolocated in the real world to be used as a reference system to align the AR annotations. For each virtual anchor, its coordinates and an image of its virtual representation in the real environment are stored. This image will guide other users through the calibration process described below. The GPS coordinates of the device at the time of setup are also stored. In this study, this step is performed by a high-specs smartphone using Google's ARCore platform. This device is able to track flat surfaces and feature points to acquire 3D positional information of itself and its surrounding environment. When this process is over, the device is ready to create AR annotations, that is: place virtual information on real elements of the environment.

The process of creating AR annotations has been carried out in a few simple steps: users select the point in space where they wish to place the virtual information, the system creates a semi-transparent sphere over these coordinates and users resize the sphere to highlight the object they wish to annotate. This process has been done by the evaluators and is not part of the experimental study conducted. This is because our aim is not to focus on the authoring part but to create a method to share the augmented space during asynchronous collaboration in unprepared environments regardless of the device used.

When a user is located in a certain space, GPS coordinates help to inform him if there are AR annotations nearby. To be able to display them correctly, an initial user-assisted calibration is necessary. First, the user must place three anchor points (which we will call "calibration points") in the appropriate position in the real environment assisted by the images stored in the initial setup described above. Afterwards, the system is able to calculate translation and rotation transforms that align the reference system of the user's device with the reference system of the AR annotations. This kind of user-assisted calibration is common in Spatial Augmented Reality applications as shown in [31].

The calculations made during the calibration process are slightly different depending on the type of device. In the case of the high-specs device, where it is possible to track flat surfaces, the calibration points selected by the user have a direct correspondence with the virtual anchors, so it is only necessary to calculate the translation and rotation matrices that minimize the distance between the set of three points and their corresponding control point. This matrix has been calculated with an equivalent implementation to the "estimateAffine3D" method included in the OpenCV library. For the calibration of the low-specs device, where only the gyroscope information is available, the same method cannot be used, so an algorithm based on the distance between straight lines has been implemented. When the user sets the location of a calibration point, three angles are obtained from the gyroscope (yaw, pitch and roll). These angles define a direction vector, which together with the 3D position of the virtual anchor defines a line (see Figure 2a).

Whenever the user has completed the three measurements from the same position with the minimum possible translation, we can calculate this position as the point of minimum distance between the three straight lines (see Figure 2b). In this calculation, it is necessary to consider that the rotations obtained by the device's gyroscope may be defined in Cartesian axes that are not aligned with the Cartesian axes of the virtual anchors, so in addition to the position it is also necessary to calculate the rotation transformation between the two reference systems.

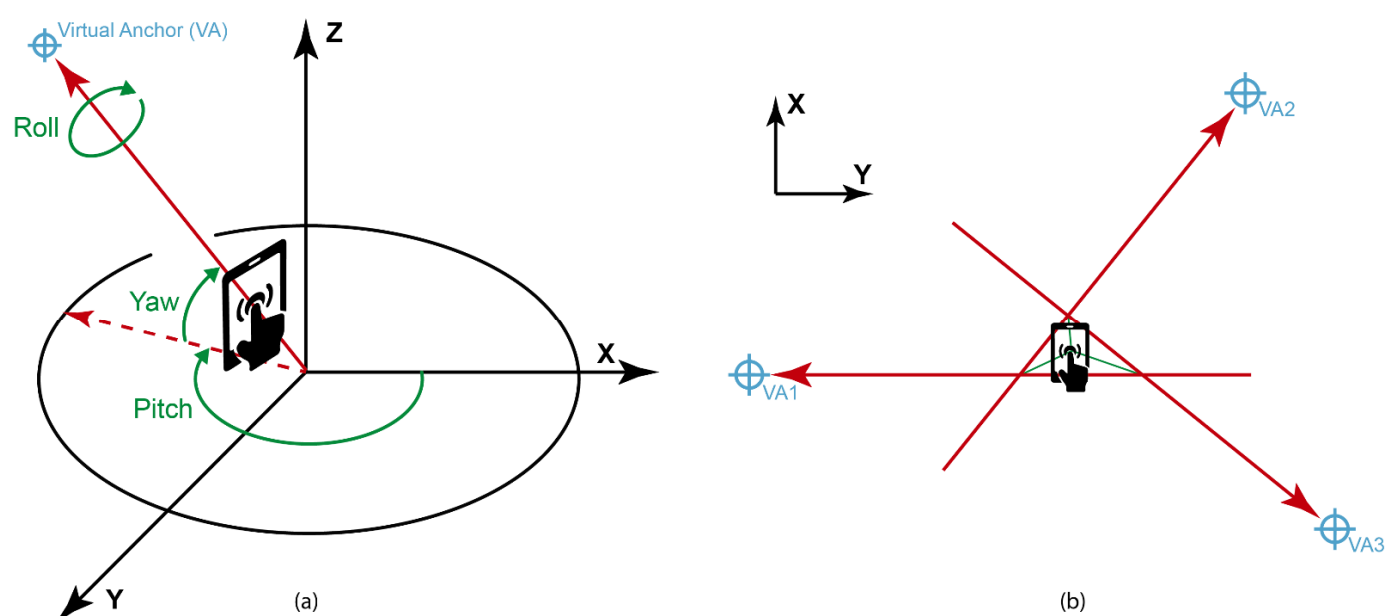


Figure 2. Isometric projection (a) and plant (b) of the calibration method for low-specs device.

### 3.2. Data Model

Based on the general taxonomy and the initial design of the data model presented in [32], we have developed an XML schema capable of describing any AR annotation regardless of the device used. An XML schema describes and constrains the structure and data of an XML document using the XSD language in a more powerful way than DTDs. Our schema makes it possible to store all the data necessary to describe an AR annotation: id, author, date, visibility, anchoring location, virtual information location, content and history of changes. Both the anchoring location and the virtual information location are defined by a reference system and coordinates with six degrees of freedom. The virtual information location also incorporates all data concerning the freedom of movement and its visual connection to the anchor point, if required. The content of the annotation is defined by key-value properties.

An example of the use of the developed XML schema with the AR annotations used in this study is shown in Scheme 1. It defines an AR annotation that is always visible, and the virtual content is always located in the same place as the anchoring. The anchoring has fixed coordinates and its reference system is the world. The “property” tags, composed of key-value pairs, easily define the spheres that we have used to annotate the objects in the environment. As can be seen, it is a simple but powerful structure that allows any type of annotation to be defined, both its anchoring to the real world and its virtual information. The developed schema ensures that this XML is valid and, in turn, has the potential to allow complex annotation definitions.

```

<?xml version="1.0" encoding="UTF-8"?>
<annotations>
<annotation id="H1">
<author>user2107201303</author>
<creationDate>2021-07-20</creationDate>
<visible>true</visible>
<visibilityType>fixed</visibilityType>
<anchoringLocation>
<positionRefSystem>world</positionRefSystem>
<coordinates>0.4384915,-1.52324259,-2.27827144,
0.369111031,163.683075,0.260549843</coordinates>
</anchoringLocation>
<virtualInformationLocation>
<positionRefSystem>world</positionRefSystem>
<orientationRefSystem>world</orientationRefSystem>
<maxDistanceAnchoring>0</maxDistanceAnchoring>
<minDistanceAnchoring>0</minDistanceAnchoring>
</virtualInformationLocation>
<content type="sphere">
<property key="diameterSize">
<value>100</value>
</property>
<property key="background color">
<value>255 255 255</value>
</property>
<property key="background alpha">
<value>25</value>
</property>
<property key="border color">
<value>216 30 91</value>
</property>
<property key="border size">
<value>5</value>
</property>
</content>
</annotation>
</annotations>

```

**Scheme 1.** XML that implements an AR annotation for any device.

This data model, together with the calibration process developed, makes our system a universal tool for the AR annotation process. In order to validate the proposed cross-device AR annotation method, we have completed an experimental study where forty real users were asked to conduct different tasks using our application.

## 4. Study

### 4.1. Protocol Design

The aim of our experimental study is to verify that we are able to visualize AR annotations in indoor environments without prior preparation regardless of the device used and that the usability of the developed application is high. That is why it is focused only on the user-assisted calibration process and on the visualization of the AR annotations present in the environment, not on their creation and the authoring process. For this purpose, we designed a set of tasks that were performed with a smartphone in which two different devices were simulated: one with low-specs and one with high-specs, as described in the previous section.

We conducted our experimental study in an indoor laboratory environment, shown in Figure 3. This environment is conducive to the vast majority of tracking systems, allowing us to focus on the evaluation of our calibration method without being influenced by external noise factors. In this room, three sets of AR annotations were prepared. The first one consists of a single AR annotation and is used for the training stage. The other two sets consist of three annotations each and are used for testing with each of the device configurations. The calibration points of the three sets are different but selected to be of equivalent difficulty. Each user performed a first test guided by an expert on the

Training Set and then performed the experiment twice: once on Set 1 with the low-specs configuration and once on Set 2 with the high-specs configuration. Users were divided into two groups: Group L started with Set 1 and Group H, with Set 2. This is to avoid biases related to tool learning.



**Figure 3.** Three hundred and sixty degree image of the environment in which the experimental study was carried out.

Since in this research we intend to make an application that is as universal as possible, people with different professions, ages and degree of previous experience in the use of AR technologies were recruited for the experiment. With the people who volunteered, a probability sampling was performed to randomly choose 40 participants. With each participant who came to the laboratory to perform the experiment, we carried out the following protocol:

*Presentation and description.* Before starting the experiment, it was explained to each user what the experiment would consist of (training, an experiment with each configuration of the device and a questionnaire). The differences between using one configuration or another and the tasks to be performed in the experiment were also explained. In addition, users were required to sign a compulsory informed consent where they declared to agree with the terms of the experiment.

*Training.* All users received a short briefing on how to use the application and the differences between using the mobile device in its high or low specs configuration. This was followed by a guided practice on how to place the three anchor points to calibrate the device and how to detect AR annotations present in the environment with both configurations.

*Experiment.* Each user performed the same set of tasks on the two device configurations. These tasks consisted of calibrating the device and finding three annotations in the room in the shortest possible time. During the experiment, information was saved on the total time for each test, the number of calibration attempts, the calibration time, the time to find each annotation, whether it was found or not, and whether the object identified as the annotated one was the correct one or another. In addition, photographs were recorded of the positioning of each anchor and of each AR annotation found by the users.

*Evaluation.* Once the users finished performing the tasks with both configurations, they were asked to complete a questionnaire. Table 1 shows the questions that make up this questionnaire. These questions were grouped in the six factors proposed by Witmer [33]: sensory factors (SF), control factors (CF), distraction factors (DF), ergonomic factors (EF), realism factors (RF), and other factors (OF). There are also four additional questions about user preference and recommendations regarding the application and their score for each of the configurations.



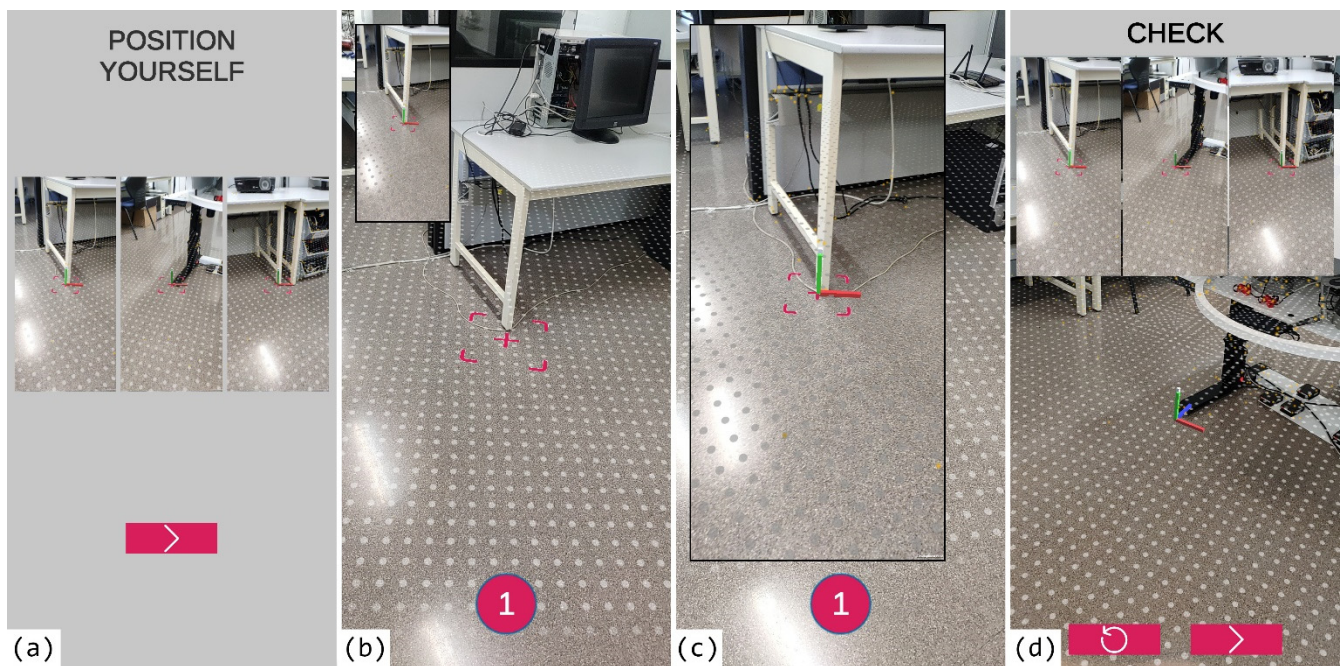
Table 1. Questionnaire.

Question	Factor
It was easy to calibrate the device (mark the three initial points).	RF
It was easy to find annotations with the high-specs configuration.	RF
It was easy to find annotations with the low-specs configuration.	RF
It was easy to find out which objects were annotated with the high-specs configuration.	RF
It was easy to find out which objects were annotated with the low-specs configuration.	RF
Not being able to move around with the low-specs configuration was NOT a problem.	CF
The use of the application did NOT require a great mental effort.	SF
The amount of information displayed on the screen was adequate	SF
The information displayed on the screen was easy to read	SF
The information displayed on the screen was easy to understand	SF
The use of the application did NOT require a great physical effort.	EF
The use of the smartphone during the experiment was comfortable (neck, shoulders, back, etc.)	EF
At no time did I feel that the smartphone was going to fall out of my hands.	CF
The handling of the application was simple and without complications.	CF
The handling of the application was natural	CF
The application responded to my actions adequately.	CF
I did NOT feel delays between my actions and the expected results.	CF
I quickly got used to the application	CF
I focused on the contents within the application and not on the mobile device.	DF
I think I have learned concepts and ideas about Augmented Reality annotations.	CF
I would like to use a similar application for other purposes.	OF
At the end of the experience I felt expert in the management of the application.	CF
I felt motivated during the experience.	OF
I liked the experience.	OF
What did you like most about the Augmented Reality annotation tool?	
What improvements or changes would you suggest?	
Rate the system of high-specs.	
Rate the system of low-specs.	

#### 4.2. Task Description

The experiment consists of two distinct parts. The first consists of calibrating the device so that the virtual information is properly positioned on the real objects that are being annotated. The second is to find the objects that have been previously annotated in the room. These tasks have been designed taking into account that, in the potential application contexts of this system, it will be an essential requirement that the search for AR annotations is a simple and clear process.

To perform the initial calibration, users are presented, on the device screen, with three photographs showing the location of the three virtual anchors that were used to take the AR annotations (see Figure 4). Users must position themselves at a point in the room from which they can see what these photographs show. Users are then shown the photographs one by one. By pressing with the finger on the picture, it becomes larger for a correct visualization (see Figure 4). Users have to find, with the device's camera and the help of a virtual sight, where each anchor point goes (see Figure 4). When they find the exact location, they click on a button on the graphical interface to register it. When the process is finished with the three virtual anchors, the three photographs are shown again and, in the real environment, the calibration points that the user has established (see Figure 4). In this way, they can verify that the calibration has been performed correctly. If so, users can move on to the next task; if not, users can perform the calibration again. When the task is finished, the application internally computes the time it took to perform the calibration process and the number of attempts.



**Figure 4.** Screenshots of the application during the calibration process.

For the search of AR annotations, a sequential process has been followed, so that users are presented with the annotations one by one. Therefore, until they find an annotation or reports that they are unable to find it, they cannot move on to the next one. Thus, once the system is calibrated, users are asked to find the first annotation. Users must move the device, which will have the camera active, to locate the AR annotation in the room. In this experiment, an AR annotation consists of a sphere surrounding the annotated object, as shown in Figure 5b. If the high-specs configuration is being used, users will be able to move around the room; otherwise, users must stand still at the location from which they performed the calibration and rotate around themselves. Once users find the annotation, they click on it and inform the evaluator of the actual object they think is the annotated one. Figure 5a shows a user clicking on a found AR annotation. The system saves a screenshot at the moment the user taps on the annotation and times how long it took to find it. The evaluator notes whether the correct object was found or not. If users cannot find the annotation, they can move to the next annotation by pressing a button on the screen. This annotation search process is repeated three times with different degrees of difficulty in terms of the annotated objects. The system also saves the total time it took the user to find (or not) the three AR annotations.

Once the search for the three AR annotations has been completed, the experiment is repeated with the other configuration of the device and with another set of annotations. These are of the same degree of difficulty as the previous ones but slightly different to avoid the biases that would be produced by the learning of their location during the experiment with the first configuration. Additionally, the calibration points are different for each configuration. With this in mind, the following sets of AR annotations have been prepared according to the device configuration:

- *Set 1 (low-specs configuration):* (L1) a computer monitor placed next to two other similar monitors, (L2) some filing cabinets placed between other objects of similar characteristics and (L3) an A4-size poster on the wall, placed next to others of the same size.
- *Set 2 (high-specs configuration):* (H1) an A4-size poster on the wall, placed next to others of the same size, (H2) a projector placed between other objects of similar characteristics and (H3) a computer monitor placed next to two other similar monitors.





**Figure 5.** User testing the application (a) and screenshot of the application during the AR annotation search (b).

In preparing these sets, we have introduced three degrees of difficulty in the annotations, based on their size and distance to similar nearby objects. Annotations L1 and H3 correspond to medium-large objects (between 38 and 44 cm) and with a relatively large distance between them (60 cm) (see Figure 6). Annotations L2 and H2 correspond to medium-sized objects (about 30 cm) and with a distance between them equal to their size (see Figure 6). L3 and H1 annotations correspond to objects of medium-small size (about 20 cm) and with a distance to other equal objects similar to their size (see Figure 6).



**Figure 6.** AR annotations that users had to find during the experimental study.

#### 4.3. Participants and Groups

We carried out the study involving 40 valid participants, that is: those who completed the entire experiment, from the training to the questionnaire, without interruptions. Out of these, 17 were women and 23 men. The participants' ages ranged between 19 and 60, distributed as follows: 11 participants under 30 years old, 9 between 30 and 39, 10 between 40 and 49 and 10 older than 50 years old. Care was taken to ensure that both the profession of the participants and their previous experience with AR technologies was as varied as possible. The mean knowledge in AR technologies and standard deviation was  $2.325 \pm 2.403$  (on a score of 0 to 6). The experiment was attended by administrative staff,

cleaning and security personnel, engineers, researchers and students from different areas, an economist, a journalist, etc.

We split the participants into two groups of 20 people (denoted as groups L and H), randomly assigning the participants to each group. Group L was composed of 11 men and 9 women with a mean age of 40 years and a previous AR experience of 2.65. Group H was composed of 12 men and 8 women with a mean age of 39.2 years and a previous AR experience of 2. The reason behind this separation is to check if the order in which the two devices configuration were used to complete the tasks of the experiment has a noticeable effect on how users perceive each one, as other similar works propose [34,35]. Participants in Group L first tested the low-specs configuration and participants in Group H first tested the high-specs configuration.

Several metrics were obtained during and after the experiment. The measurements came from the participants (through the questionnaire shown in Table 1 and the verbalization about which object they believes the annotated one is) and from the application (objective measures about user performance such as times or calibration attempts, and photographs of the positioning of each anchor and of each AR annotation found by the users). The questions listed in Table 1 are 7-scale Likert questions with 0 meaning strongly disagree, 1 disagree, 2 somewhat disagree, 3 neutral, 4 somewhat agree, 5 agree, and 6 strongly agree, except for the last two questions in which users rated the system from 1 to 10, with 1 being the worst score and 10 the best.

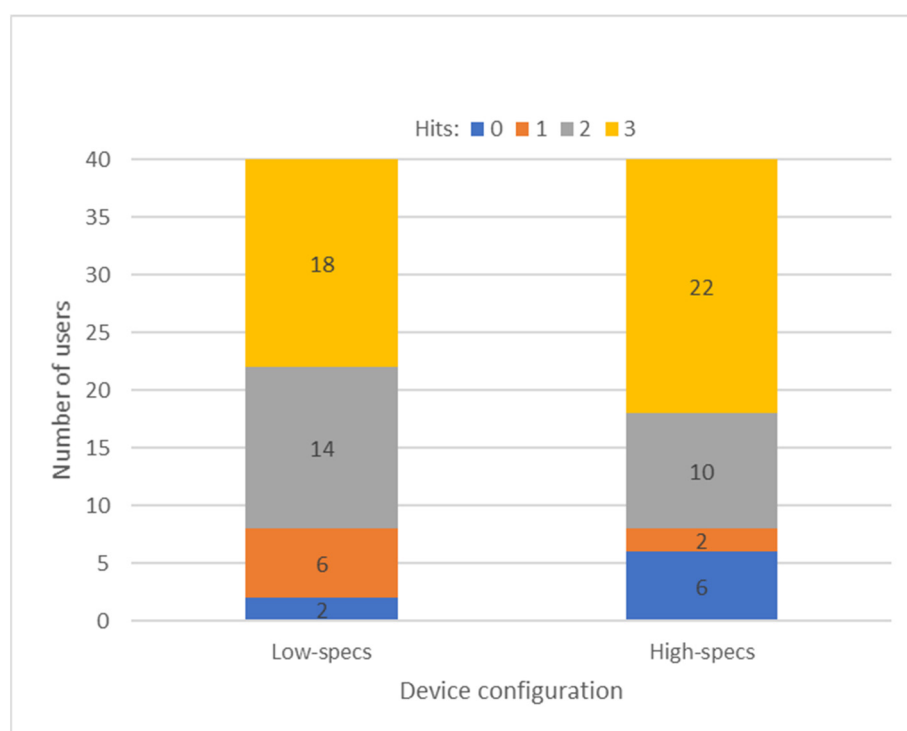
The questions shown in Table 1 are designed to test hypothesis 2 (the overall usability of the system is satisfactory), whereas the measures obtained during the experiment (especially the register of the annotated objects found) have been used to verify hypothesis 1 (we are able to visualize AR annotations in indoor environments without prior preparation regardless of the device used).

## 5. Results and Discussion

This section presents the statistical analysis of the sets of data obtained from the experiments conducted with real users. We will focus our analysis around three axes: annotated object hits (to validate our primary hypothesis), user satisfaction (to validate the secondary hypothesis) and execution times (to verify whether we are also doing it in an acceptable time).

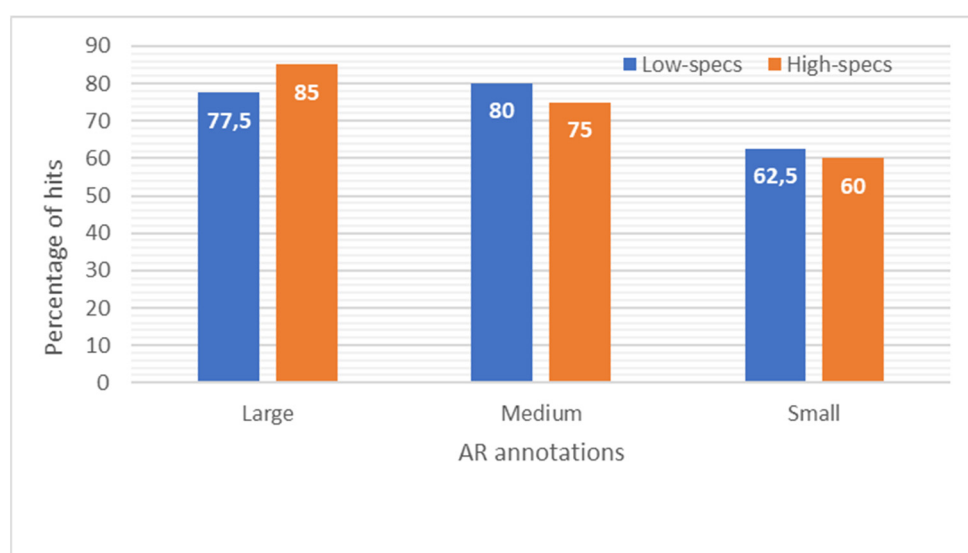
### 5.1. Annotations Found and Objects Correctly Identified

There is a difference between finding an AR annotation and correctly identifying the annotated object. Of the 40 participants in the study, only one failed to find two of the three AR annotations present in a set. This was because, during calibration of the device, he/she placed an anchor point on the wrong table. The remaining participants found all annotations with both device configurations. We will now analyze the hit rate for the annotated objects. Figure 7 shows the number of users who hit 0, 1, 2 or 3 of the annotated objects with each of the device configurations. With both the high-specs and low-specs devices, 80% of users were able to correctly identify 2 or 3 annotations out of the 3 presents in each set. On average, half of the users found all the annotated objects. Only 8 participants found one or none of the annotated objects.



**Figure 7.** Number of successful AR annotations for each user.

The number of hits differed according to the size of the annotated objects and their proximity to similar objects. Figure 8 shows how the hit rate is higher for medium and large objects than for small objects. Even so, the hit rate is equal to or higher than 60% in all cases. In addition, failures are normally distributed independently of age and calibration time, which will be discussed later. The results obtained do not allow us to state categorically whether either of the two devices was better than the other in hitting the annotated objects.



**Figure 8.** Percentage of hits for each of the AR annotations: large (L1 and H3), medium (L2 and H2) and small (L3 and H1).

It should be noted that the accuracy of the positioning of the annotations depends mainly on the prior calibration of the device by the user. Therefore, the care with which this is done directly influences the results obtained. We consider, therefore, that the results



obtained are satisfactory since, in the worst case (a device with low specifications to search for a small-medium sized AR annotation) we have obtained a 62.5% success rate. Even in cases where users misidentified the annotated objects, the AR annotations were displayed relatively close to the annotated objects. Figure 9 shows an overlay of all annotations found during the experiment. As can be seen, only in a few specific cases, the virtual information is displayed more than one meter away from the annotated object.



**Figure 9.** Overlay of the AR annotations found by the 40 study participants.

Therefore, based on these results, it can be said that our primary hypothesis is correct, as all users except one have visualized the AR annotations and a significant percentage of participants have associated them with the annotated object correctly regardless of the device used.

### 5.2. User Satisfaction

For all of the analyses detailed hereafter, significance tests were two-tailed and completed at the 0.05 significance level. First, we checked if the collected data follow a normal distribution. As a representative example, the Kolmogorov–Smirnov test ( $D = 0.344$  and  $p\text{-value} = 0.617$ ), the Anderson-Darling test ( $A = 0.567$  and  $p\text{-value} = 0.162$ ), and the Shapiro–Wilk test ( $W = 0.371$  and  $p\text{-value} = 0.471$ ) confirmed that the score dataset for high-specs smartphone configuration follows a normal distribution. Although for the sake of brevity we do not detail the rest of the normality tests, the same happened in the rest of datasets. Therefore, we can use parametric tests: the  $t$ -test and the Cohen’s test for paired and unpaired data, as well as a correlation study and a multifactorial ANOVA for analyzing relationships among the parameters in the experiment. Table 2 shows the average and the standard deviation of the results obtained from the questions shown in Table 1. The responses to the questionnaire are grouped around Witmer’s six factors. In addition, user ratings for each of the device configuration are included.

**Table 2.** Averages and standard deviations of the responses to the questionnaire for all participants.

Parameter	Mean $\pm$ SD	$t$	$p$	Cohen’s $d$
RF	5.145 $\pm$ 1.009	−1.037	0.306	−0.505
CF	5.251 $\pm$ 0.834	−2.046	0.048	−1.089
SF	5.35 $\pm$ 0.681	−3.072	0.004	−0.930
EF	5.675 $\pm$ 0.685	−1.16	0.253	−1.083
DF	5.410 $\pm$ 0.715	0.091	0.928	−1.520
OF	5.375 $\pm$ 0.632	−1.99	0.054	−4.349
SCORE H-S	9.218 $\pm$ 0.951	−1.594	0.119	−0.511
SCORE L-S	7.885 $\pm$ 1.855	−1.182	0.245	−0.379

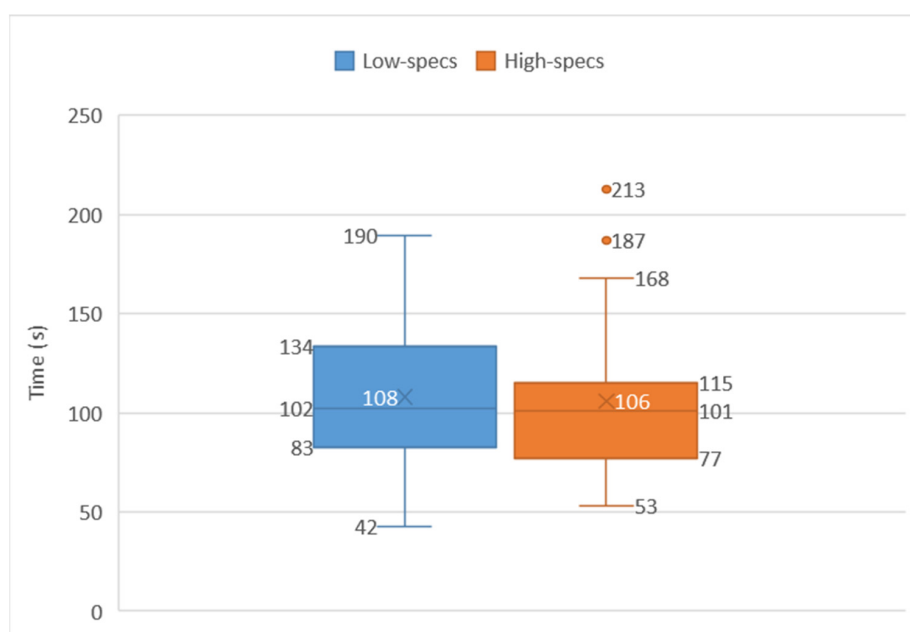
In all factors an average score above 5 was obtained. Recall that the maximum score was 6, so we can say that the user experience with the application has been very positive. If we analyze the differences between the two groups, we can see that only in CF and SF a  $p < 0.05$  was obtained in favor of Group L, which started the experiment with the low-specs configuration. These users scored better on the CF “Not being able to move around with the low-specs configuration was NOT a problem” and “At the end of the experience I felt expert in the management of the application” questions. On the other hand, the SF question “The amount of information displayed on the screen was adequate” obtained low scores (two twos and two threes) only from users of Group H. Moreover, if we analyze the final thoughts proposed by the users, seven of the participants from Group H referred to improvements related to the low-specs configuration compared to three users from Group L. Therefore, a possible explanation for these differences between groups could be that the users who tested the low-specs configuration second might have had a worse feeling and rated certain aspects of the application worse.

The overall average score obtained for the evaluated system was above 9 for the high-specs configuration and almost 8 for the low-specs configuration. The study of these datasets showed statistically significant differences between the obtained scores when participants rated the app for the low-specs configurations and the app for high-specs configurations with  $t[76] = 3.994$  and  $p = 0.000$ , Cohen’s  $d = 0.587$ . In addition to these high scores, to the question “What did you like most about the Augmented Reality annotation tool?” 15 participants answered that the application was easy and simple to use. All these results make us affirm that our second hypothesis has also been validated.

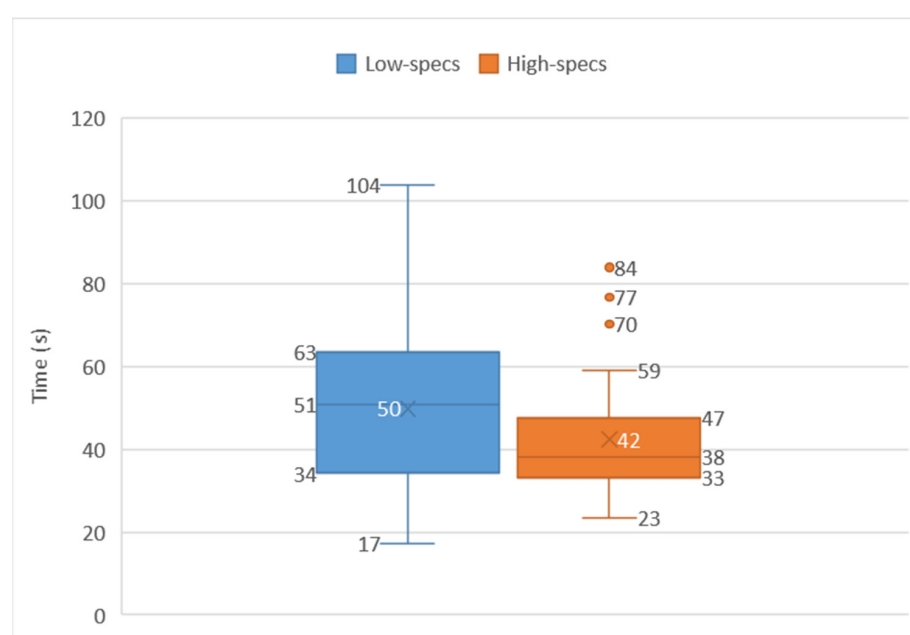
### 5.3. Execution Times

Although the above data have already verified our two hypotheses, we also wanted to analyze the time taken by users to carry out the tasks, as this will give us an idea of the efficiency of our application and, in particular, of our calibration method.

The mean total time to complete the tasks (calibration and annotation search) was very similar for both configurations and less than two minutes, as Figure 10 shows. In addition, all annotations were found, on average, in less than 30 s. There was also little difference in mean calibration time between configurations and in both cases, it took less than one minute to calibrate the device, as Figure 11 shows. Only 6 users with low-specs and 8 with high-specs made more than one calibration attempt. Moreover, both groups took less time to calibrate with the second configuration used. Therefore, although it may initially be tedious to burden the user with the task of calibrating the device, it has been found to be a quick and easy process with great benefits, such as using the device in unprepared environments.



**Figure 10.** Total time taken by participants to complete all tasks (calibration and annotation search).



**Figure 11.** Time spent by participants to calibrate the device.

All these data corroborate that the developed system meets the objectives of our research: to be able to find AR annotations in indoor environments without prior preparation, with different types of devices and, moreover, in a simple way.

## 6. Conclusions and Future Work

AR annotations have great potential in asynchronous collaborative contexts but present practical problems such as prior preparation of the environment and technological diversity. Therefore, it is important to find alternatives that allow their use in unprepared environments and with different types of devices when creator and consumer are not present at the same time. However, it is not easy to find publications in the academic literature that take all these factors into account. To address this shortcoming, this paper presents a method that allows creating and visualizing annotations with different AR devices without any kind of marker or image recognition. To verify that the developed system fulfils its function and that it is easy to use by any user, a study has been carried out with 40 participants who have used the application on a high-specs smartphone and on a low-specs smartphone.

From the results of this article, we can conclude that our primary hypothesis (“we are able to visualize AR annotations in indoor environments without prior preparation regardless of the device used”) is corroborated, since only one participant was not able to find some of the AR annotations and 80% of the users correctly identified 2 or 3 of the 3 annotated objects in the environment. Moreover, in a negligible number of cases the location error was greater than one meter. Our secondary hypothesis (“the overall usability of the system is satisfactory”) is also corroborated by the experimental data, since all factors analyzed were rated on average by users above 5 on a scale of 6. In addition, both device configurations scored high: 9.2 for high-specs and 7.9 for low-specs. Finally, the times taken to complete the tasks demonstrate that our system is efficient. In particular, the device calibration times corroborate that our method of anchoring virtual information in unprepared environments is practical and easy to use and learn for any user.

The results presented here can be helpful for future research on AR-CSCW that aim to be as universal as possible. This can be applicable in a wide variety of fields: industry, construction, heritage, tourism, games, etc. In fact, thanks to the data model used and the calibration method we have just validated, it is possible to create an AR annotation cloud that is completely independent of the device. Moreover, it would be relatively easy to present in the environment annotations taken by different users at different times even if they have different anchor points. With the GPS position stored, we would know that in a given space there are several groups of annotations and the necessary translations could be made to display all of them with the same reference system. Finally, we plan to test our data model and calibration method with other AR devices, such as head-mounted or projection-based displays, in different environments and with more complex annotations.

**Author Contributions:** Conceptualization, I.G.-P. and J.G.; methodology, I.G.-P. and P.M.; software, P.C.-S. and J.G.; validation, I.G.-P. and P.C.-S.; formal analysis, I.G.-P. and D.R.; investigation, I.G.-P. and P.C.-S.; data curation, P.M.; writing—original draft preparation, I.G.-P.; writing—review and editing, P.C.-S., J.G., P.M. and D.R.; supervision, P.M. and D.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is part of the I+D+i project RTI2018-098156-B-C55, supported by Spanish MCIN and ERDF A way to make Europe.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Acknowledgments:** I.G.-P. acknowledges the Spanish Ministry of Science, Innovation and Universities (program: “University teacher formation”) to carry out this study.

**Conflicts of Interest:** The authors declare no conflict of interest.



## References

1. Bullen, C.V.; Johansen, R. *Groupware, A Key to Managing Business Teams*; Technical Report; MIT Sloan School of Management: Cambridge, MA, USA, 1988.
2. Ellis, C.A.; Gibbs, S.J.; Rein, G. Groupware: Some issues and experiences. *Commun. ACM* **1991**, *34*, 39–58. [\[CrossRef\]](#)
3. Irlitti, A.; Smith, R.T.; Itzstein, S.V.; Billinghamurst, M.; Thomas, B.H. Challenges for Asynchronous Collaboration in Augmented Reality. In Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Merida, Mexico, 19–23 September 2016; pp. 31–35. [\[CrossRef\]](#)
4. Pidel, C.; Ackermann, P. Collaboration in Virtual and Augmented Reality: A Systematic Overview. In *Augmented Reality, Virtual Reality, and Computer Graphics*; Springer: Cham, Switzerland, 2020; pp. 141–156. [\[CrossRef\]](#)
5. Ens, B.; Lanir, J.; Tang, A.; Bateman, S.; Lee, G.; Piumsomboon, T.; Billinghamurst, M. Revisiting collaboration through mixed reality: The evolution of groupware. *Int. J. Hum. Comput. Stud.* **2019**, *131*, 81–98. [\[CrossRef\]](#)
6. Sereno, M.; Wang, X.; Besancon, L.; McGuffin, M.J.; Isenberg, T. Collaborative Work in Augmented Reality: A Survey. *IEEE Trans. Vis. Comput. Graph.* **2020**. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Speicher, M.; Hall, S.D.; Yu, A.; Zhang, B.; Zhang, H.; Nebeling, J. XD-AR: Challenges and Opportunities in Cross-Device Augmented Reality Application Development. *Proc. ACM Hum. Comput. Interact.* **2018**, *2*, 7:1–7:24. [\[CrossRef\]](#)
8. Wither, J.; DiVerdi, S.; Höllerer, T. Annotation in outdoor augmented reality. *Comput. Graph.* **2009**, *33*, 679–689. [\[CrossRef\]](#)
9. Irizarry, J.; Gheisari, M.; Williams, G.; Walker, B.N. InfoSPOT: A mobile Augmented Reality method for accessing building information through a situation awareness approach. *Autom. Constr.* **2013**, *33*, 11–23. [\[CrossRef\]](#)
10. Jalo, H.; Pirkkalainen, H.; Torro, O.; Kärkkäinen, H.; Puhto, J.; Kankaanpää, T. How Can Collaborative Augmented Reality Support Operative Work in the Facility Management Industry? In Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Seville, Spain, 18–20 September 2018; pp. 41–51. [\[CrossRef\]](#)
11. Ioannidi, A.; Gavalas, D.; Kasapakis, V. Flaneur: Augmented exploration of the architectural urban space. In Proceedings of the 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, 3–6 July 2017; pp. 529–533. [\[CrossRef\]](#)
12. Daiber, F.; Kosmalla, F.; Krüger, A. BoulderAR: Using augmented reality to support collaborative boulder training. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2013; pp. 949–954. [\[CrossRef\]](#)
13. Kasahara, S.; Heun, V.; Lee, A.S.; Ishii, H. Second surface: Multi-user spatial collaboration system based on augmented reality. In *SIGGRAPH Asia 2012 Emerging Technologies*; ACM: New York, NY, USA, 2012; pp. 1–4. [\[CrossRef\]](#)
14. Martín-Gutiérrez, J.; Fabiani, P.; Benesova, W.; Meneses, M.D.; Mora, C.E. Augmented reality to promote collaborative and autonomous learning in higher education. *Comput. Hum. Behav.* **2015**, *51*, 752–761. [\[CrossRef\]](#)
15. Huang, F.; Zhou, Y.; Yu, Y.; Wang, Z.; Du, S. Piano AR: A Markerless Augmented Reality Based Piano Teaching System. In Proceedings of the 2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, 26–27 August 2011; Volume 2, pp. 47–52. [\[CrossRef\]](#)
16. Ahuja, K.; Paredy, S.; Xiao, R.; Goel, M.; Harrison, C. LightAnchors: Appropriating Point Lights for Spatially-Anchored Augmented Reality Interfaces. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, New York, NY, USA, 20–23 October 2019; pp. 189–196. [\[CrossRef\]](#)
17. Tregel, T.; Dutz, T.; Hock, P.; Müller, P.N.; Achenbach, P.; Göbel, S. StreetConqAR: Augmented Reality Anchoring in Pervasive Games. In *Serious Games*; Springer: Cham, Switzerland, 2020; pp. 3–16. [\[CrossRef\]](#)
18. Lee, T.; Hollerer, T. Hybrid Feature Tracking and User Interaction for Markerless Augmented Reality. In Proceedings of the 2008 IEEE Virtual Reality Conference, Reno, NV, USA, 8–12 March 2008; pp. 145–152. [\[CrossRef\]](#)
19. Azuma, R.; Weon Lee, J.; Jiang, B.; Park, J.; You, S.; Neumann, U. Tracking in unprepared environments for augmented reality systems. *Comput. Graph.* **1999**, *23*, 787–793. [\[CrossRef\]](#)
20. Höllerer, T.; Wither, J.; DiVerdi, S. “Anywhere Augmentation”: Towards Mobile Augmented Reality in Unprepared Environments. In *Location Based Services and TeleCartography*; Gartner, G., Cartwright, W., Peterson, M.P., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 393–416. [\[CrossRef\]](#)
21. Afif, F.N.; Basori, A.H. Orientation Control for Indoor Virtual Landmarks based on Hybrid-based Markerless Augmented Reality. *Procedia Soc. Behav. Sci.* **2013**, *97*, 648–655. [\[CrossRef\]](#)
22. Xu, K.; Prince, S.J.D.; Cheok, A.D.; Qiu, Y.; Kumar, K.G. Visual registration for unprepared augmented reality environments. *Pers Ubiquit Comput.* **2003**, *7*, 287–298. [\[CrossRef\]](#)
23. Langlotz, T.; Wagner, D.; Mulloni, A.; Schmalstieg, D. Online Creation of Panoramic Augmented Reality Annotations on Mobile Phones. *IEEE Pervasive Comput.* **2012**, *11*, 56–63. [\[CrossRef\]](#)
24. Casas, S.; Portalés, C.; García-Pereira, I.; Gimeno, J. Mixing Different Realities in a Single Shared Space: Analysis of Mixed-Platform Collaborative Shared Spaces. In *Harnessing the Internet of Everything (IoE) for Accelerated Innovation Opportunities*; IGI Global: Hershey, PA, USA, 2019; pp. 175–192. [\[CrossRef\]](#)
25. García-Pereira, I.; Gimeno, J.; Pérez, M.; Portalés, C.; Casas, S. MIME: A Mixed-Space Collaborative System with Three Immersion Levels and Multiple Users. In Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Munich, Germany, 16–20 October 2018; pp. 179–183. [\[CrossRef\]](#)



26. Hoppe, A.H.; Westerkamp, K.; Maier, S.; van de Camp, F.; Stiefelhagen, R. Multi-user Collaboration on Complex Data in Virtual and Augmented Reality. In Proceedings of the HCI International 2018—Posters' Extended Abstracts, Las Vegas, NV, USA, 15–20 July 2018; pp. 258–265.
27. Butz, A.; Hollerer, T.; Feiner, S.; MacIntyre, B.; Beshers, C. Enveloping users and computers in a collaborative 3D augmented reality. In Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99), Washington, DC, USA, 20–21 October 1999; pp. 35–44. [[CrossRef](#)]
28. MacWilliams, A.; Sandor, C.; Wagner, M.; Bauer, M.; Klinker, G.; Bruegge, B. Herding sheep: Live system for distributed augmented reality. In Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, Tokyo, Japan, 10 October 2003; pp. 123–132. [[CrossRef](#)]
29. Baillard, C.; Fradet, M.; Alleaume, V.; Jouet, P.; Laurent, A. Multi-device mixed reality TV: A collaborative experience with joint use of a tablet and a headset. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, New York, NY, USA, 8–10 November 2017; pp. 1–2. [[CrossRef](#)]
30. Azure Spatial Anchors | Microsoft Azure. Available online: <https://azure.microsoft.com/es-es/services/spatial-anchors/> (accessed on 5 August 2021).
31. Portalés, C.; Casanova-Salas, P.; Casas, S.; Gimeno, J.; Fernández, M. An interactive cameraless projector calibration method. *Virtual Real.* **2020**, *24*, 109–121. [[CrossRef](#)]
32. García-Pereira, I.; Gimeno, J.; Morillo, P.; Casanova-Salas, P. A Taxonomy of Augmented Reality Annotations', Valletta, Malta. 2020. pp. 412–419. Available online: <https://www.scitepress.org/Link.aspx?doi=10.5220/0009193404120419> (accessed on 13 April 2021).
33. Witmer, B.G.; Singer, M.J. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence Teleoperators Virtual Environ.* **1998**, *7*, 225–240. [[CrossRef](#)]
34. Juan, M.-C.; García-García, I.; Mollá, R.; López, R. Users' Perceptions Using Low-End and High-End Mobile-Rendered HMDs: A Comparative Study. *Computers* **2018**, *7*, 15. [[CrossRef](#)]
35. Polvi, J.; Taketomi, T.; Yamamoto, G.; Dey, A.; Sandor, C.; Kato, H. SlidAR: A 3D positioning method for SLAM-based handheld augmented reality. *Comput. Graph.* **2016**, *55*, 33–43. [[CrossRef](#)]