Article

The Effects of Augmented Reality Companion on User Engagement in Energy Management Mobile App

Joo Chan Kim [†,*] , Saguna Saguna [†] and Christer Åhlund [†]

Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology,
931 87 Luleå, Sweden; saguna.saguna@ltu.se (S.S.); christer.ahlund@ltu.se (C.Å.)
* Correspondence: joo.chan.kim@ltu.se; Tel.: +46-7-2225-4785
† These authors contributed equally to this work.

Abstract: As the impact of global warming on climate change becomes noticeable, the importance of energy efficiency for reducing greenhouse gas emissions grows immense. To this end, a platform, solution, and mobile apps are developed as part of the European Union's Horizon 2020 research and innovation program to support energy optimization in residences. However, to ensure long-term energy optimization, it is crucial to keep users engaged with the apps. Since augmented reality (AR) and a virtual animal companion positively influenced user engagement, we designed an AR companion that represented the user's residence states; thereby making the user aware of indoor information. We conducted user evaluations to determine the effect of the AR companion on user engagement and perceived usability in the context of energy management. We identified that the user interface (UI) with AR (ARUI) barely affected user engagement and perceived usability compared to the traditional UI without AR (TUI); however, we found that the ARUI positively affected one of the user engagement aspects. Our results show AR companion integration's potential benefits and effects on energy management mobile apps. Furthermore, our findings provide insights into UI design elements for developers considering multiple interaction modalities with AR.

Keywords: user interface; user evaluation; user engagement; perceived usability; augmented reality; Internet of Things; energy management

1. Introduction

In 2023, climate change caused by global warming becomes a threat to humankind more than ever [1]. Therefore, attention to greenhouse gases' substantial impact on global warming has arisen, which leads to attempts to reduce gas emissions [2,3]. The European Union (EU) takes action for energy efficiency improvement to solve the greenhouse gas emission issue [4]. The "Adapt-&-Play holistic cost-effective and user-friendly innovations with high replicability to upgrade smartness of existing buildings with legacy equipment" (PHOENIX) project is one of the outcomes of the EU's Horizon 2020 research and innovation program to improve energy efficiency by optimizing energy consumption [5]. The PHOENIX project aims to increase the smartness of a building's energy management system, thereby increasing energy efficiency. As a solution, information about an apartment and advice on energy consumption optimization is served to occupants through a desktop browser-based dashboard. However, since a mobile app could provide additional benefits than a desktop browser in terms of ease of use and user engagement [6–8], a mobile app was discussed as an alternative platform to serve the PHOENIX solution.

In order to accomplish long-term energy optimization, a mobile app needs active users who continuously use the app and solution. As this user retention is affected by user engagement [9], improvement in user engagement is the requirement. Augmented reality (AR) has shown its effectiveness in engaging users with a mobile device, which is verified in various contexts, such as game [10], education [11], and tourism [12]. Regarding

user engagement, an animal companion presented in AR has the potential to benefit the user as well [13]. Meanwhile, AR is also used in energy management studies; however, fostering awareness of energy efficiency using intuitive data visualization [14–16] is the main purpose of AR rather than enhancing user engagement. To the best of our knowledge, the effect of an AR animal companion on user engagement in the energy management context has yet to be studied.

In this study, we aim to investigate the effect of an AR companion on user engagement within the context of an energy management mobile app. We also conduct a perceived usability test to understand how end-users perceive a mobile app with an AR companion. We develop two mobile apps on the PHOENIX platform to serve the energy optimization solution. Each mobile app features a distinct user interface (UI): one incorporates AR, while the other does not. We hypothesized that the UI with an AR companion increases user engagement (H_a^1) and perceived usability (H_a^2) compared to the UI without AR. We perform user evaluations on both UIs and compare the results to understand the effect of AR companion on user engagement and perceived usability. The following research questions are established for the research aim:

RQ1. How does an AR companion affect user engagement in the energy management mobile app?

RQ2. How does an AR companion affect perceived usability in the energy management mobile app?

In Section 2, we described related works about the effects of AR and studies of energy management mobile apps with AR. In Section 3, we present information about the participants of our user evaluation, the system architecture of the PHOENIX solution, the UIs, and the user evaluations. We describe results and discussion about user evaluation data in Section 4 and Section 5, respectively. Section 6 presents the conclusion.

2. Related Works

2.1. Effects of Augmented Reality

AR was studied for its effect on user engagement in various contexts. For example, Garay-Cortes and Uribe-Quevedo [17] designed an AR mobile game with a story of a character to make the players visit five landmarks in the real-world. The feedback from 50 players implied that 84% players were engaged with the play, and satisfied with the experience. For an indoor scenario, López-Faican and Jaen [10] developed a markerless-based AR mobile game, which supported multiplayer, for primary school students to improve their emotional intelligence and social communication. In total, 38 students experienced the multiplayer-supported AR mobile game and evaluated that the AR mobile game positively affected user engagement.

In education, Chang and Yu [18] developed an educational mobile app with image target-based AR to display educational 3D content to the students. The user evaluations with 93 students showed that the students experienced a positive influence on their learning attitude in biology, which was a hard class for them. When AR showed a positive effect on learning attitude, Chao and Chang [19] measured the technology acceptance scales of AR in terms of usefulness and ease of use from 63 students in a mathematics class. The developed mobile app for math education used image-based AR, which required a printed image as a marker for displaying AR content. A comparison test's results between the two student groups showed that the students with the AR mobile app were engaged and achieved better learning results than the students without the AR mobile app. In addition, Wen [11] showed that the educational AR mobile app designed for a Chinese language class not only engaged students but also enhanced their willingness to continue their participation in the learning process due to the interaction between the AR mobile app and students' activities in the real-world.

In tourism, Spadoni et al. [12] utilized AR in multiple ways to engage astronomy museum visitors. For example, the 3D recreation of an ancient astronomical artifact helped the visitors to understand how the artifact worked, and the AR animations about astronomy

were displayed along with written text on a panel to provide information effectively. In total, 23 participants evaluated the AR mobile app as 'engaging' rather than 'boring' because of the visual and audio stimuli. With an increased sample size, Dağ et al. [20] determined the effect of the AR mobile app developed for a museum's visitors. In total, 397 visitors' feedback was analyzed, and it was identified that the immersive experience produced by the AR mobile app had a positive effect on user engagement.

To engage users, researchers utilized virtual animals in AR [21] or non-AR [22] mobile devices and proved their effect on user engagement. In this context, an AR animal could be used as a companion to engage and support the user, similar to a virtual companion working with the user in a training system [23]. So far, AR animals have been utilized in several ways, such as for providing user engagement in a serious game within the air-quality monitoring context [24], for an educational purpose [25] and in the treatment of animal phobia, like a spider [26]. Norouzi et al. [13] showed the benefits of virtual animal companions regarding mental well-being using AR mobile environments. Accordingly, we expect an additional positive effect on user engagement by using an AR animal companion in our mobile app.

Regardless of the effect on user engagement, there were studies focused on the intuitiveness of AR for presenting visual assets. For example, Seitz et al. [27] used emojis to represent the state of Internet of Things (IoT) sensors. The emojis and additional icons, which represent the sensor value, were designed to deliver information about the IoT sensors to users as quickly as possible. Huo et al. [28] also developed an AR mobile app that worked with several IoT devices in a room. Huo et al. [28]'s app was different in that it could control IoT devices from the AR interface based on the proximity between the mobile device and the IoT devices. As Huo et al. [28] presented, AR was able to use for providing an intuitive interface for interacting with IoT devices, and this type of interface was extended to enable remote control as well [29]. On the other hand, AR was utilized to present a whole process of the user's work as a guideline by overlaying virtual content onto real-world objects. In these cases, a projector on a ceiling [30] or a head-mounted display [31] was used instead of a hand-held mobile device to allow users to use both hands on the work. The overlaying feature of AR, which enabled intuitive information presentation, provided a positive impact on the surgical environment as well [32].

2.2. Energy Management with Augmented Reality

In the energy management context, AR was mainly used for (1) simply visualizing IoT sensor data or (2) presenting a digital twin for data visualization and control interface. For example, Naret et al. [16] developed an AR mobile app that visualized electricity consumption data of appliances collected through IoT sensors in real-time. Data were displayed in AR text and a printed image was used as a marker for positioning the AR text. On the other hand, Purmaissur et al. [33] presented a mobile AR app combined with IoT sensors that displayed energy consumption information and air quality data on a mobile screen. The AR mobile app used image recognition to recognize IoT sensors for AR initialization, and information in AR was provided as text. Unlike Naret et al. [16]'s AR mobile app, Purmaissur et al. [33]'s AR mobile app enabled a remote control on one IoT sensor at a time, whereas Cho et al. [34] extended the remote control for up to five devices. Cho et al. [34] designed an AR mobile app that visualized a building's energy consumption states in an AR map, which was a digital twin of the place in the real-world, which was augmented on a miniature building. While checking the states of energy consumption through their mobile AR app, users could control the power of IoT devices in the place through the AR interface. However, both studies only developed and introduced the UIs without user evaluation regarding either user engagement or perceived usability aspects.

As AR technology matured, a need arose to evaluate how AR is perceived and its effect on user engagement. Bekaroo et al. [35] measured 40 university students' perceptions of AR in the self-learning app that was used to enhance their awareness of the appliances' energy efficiency. Bekaroo et al. [35] utilized object recognition and image recognition

to identify various-sized appliances, as large appliances were incompatible with their AR mobile app's object detection technique. The energy consumption information was calculated based on a timer, which implied the duration of appliance usage, and a manually assigned power usage value for the identified appliance to provide an energy efficiency performance. Once the calculation was completed, the energy efficiency information was displayed as text with an icon highlighted with different colors depending on the calculated energy efficiency score. Although the energy consumption information was an estimated value instead of sensor-based data, the students assessed that they could gain knowledge about the energy consumption information of each appliance. More than half of students positively perceived the AR mobile app as easy to use, enjoyable, useful, and wanted to use continuously [35].

Instead of estimating energy consumption, Mylonas et al. [36] used real-world data provided by IoT sensors in a school building for their AR mobile app to enhance the learning experience in a classroom. Mylonas et al. [36] developed a lab kit alongside an app that helped primary and secondary school students be aware of school energy consumption. The AR mobile app presented virtual texts with line graphs to show the energy consumption states without access to the web portal where all data were available. The user evaluation, which involved 106 primary and secondary school students, identified a positive influence on user engagement with their lab kit and the content delivered through the AR mobile app. However, their research was mainly focused on education and learning effects rather than the influence of AR on user engagement.

While the aforementioned AR mobile apps focused on educating users about energy consumption or energy efficiency, Alonso-Rosa et al. [37]'s AR mobile app centered on its functionality to empower users with real-time data. Alonso-Rosa et al. [37] developed a mobile app that provided energy information about kitchen appliances obtained from IoT sensors in AR. The energy information was presented by texts written on a half-transparent AR panel, which appeared by recognizing the pre-registered features of appliances in a camera view. Although IoT sensor values were expressed only in texts, various font colors were applied depending on IoT sensor values to intuitively represent the state of power usage level. Their research validated that 16 persons from a maintenance team perceived their AR mobile app as positive in usability and overall impression perspective [37].

Lastly, An et al. [38] proposed an AR mobile app tested on a tablet PC that allowed the user to monitor home appliances' energy consumption states in the form of text and line graphs. An et al. [38]'s AR mobile app utilized image recognition to detect the target appliance and presented the recognized appliance's data, which were collected through IoT sensors, on the tablet PC screen. The user evaluation with 21 participants showed that the AR mobile app had a positive effect on reducing energy consumption while avoiding a negative impact on user comfort. However, An et al. [38]'s AR mobile app had no AR companion, as other studies mentioned in Section 2.2.

To sum up, we identify that an AR animal companion is yet to be studied for its effect on user engagement in the energy consumption management context. Table 1 summarizes previous studies regarding an energy management mobile app with AR and compares the characteristics with our study.

Table 1. Comparisons of previous studies in energy management with augmented reality (AR). The column 'User evaluation' was marked with 'O' if the study had a user evaluation with results, regardless of whether it was about user engagement evaluation. The first row is the summary of this study.

Author	Platform	Marker	IoT Sensor Control	User Evaluation	AR Companion
-	Hand-held Mobile device	Human hand (palm)	O	O	O (cat)
Naret et al. [16]	Hand-held Mobile device	Printed image	X	X	X

Table 1. *Cont.*

Author	Platform	Marker	IoT Sensor Control	User Evaluation	AR Companion
Purmaissur et al. [33]	Hand-held Mobile device	Printed image	O	X	X
Cho et al. [34]	Hand-held Mobile device	Building model	O	X	X
Bekaroo et al. [35]	Hand-held Mobile device	Printed image and Physical object	X	O	X
Mylonas et al. [36]	Hand-held Mobile device	Printed image	X	O	X
Alonso-Rosa et al. [37]	Hand-held Mobile device	Physical object	X	O	X
An et al. [38]	Hand-held Mobile device	Physical object	O	O	X

3. Materials and Methods

3.1. Participants

From northern Sweden, we gathered 29 participants who showed interest in optimizing energy usage to reduce energy costs as they were paying a bill based on the energy use in their residences. The average age of all 29 participants was 44.38, and 22 out of 29 participants (75.86%) had prior knowledge about AR. We received feedback from 24 participants (P1–P24) who experienced both UIs in a day with a controlled environment that imitated the use of the PHOENIX solution in a real-world scenario. The average age of 24 participants (P1–P24) was 38.33, and we grouped them (i.e., Group A) in this study.

Five occupants (P25–P29) from four residences had IoT sensors installed for running the PHOENIX solution. Two married participants shared one residence, while the remaining three participants lived alone (three males and two females, average age of 73.40). These occupants were grouped as Group B in this study. Table 2 illustrates the demographic characteristics of all participants of the user evaluation.

Table 2. Demographic characteristics of 29 participants.

Characteristics		Number of Participants ($n = 29$)	Percentage (%)
Gender	Male	20	68.97
	Female	8	27.59
	N/A	1	3.45
Age	20–29	3	10.34
	30–39	12	41.38
	40–49	4	13.79
	50–59	5	17.24
	60–69	1	3.45
	70–79	4	13.79
Prior knowledge on AR	Yes	22	75.86
	No	7	24.14
Background knowledge	Energy science	11	37.93
	Computer science	8	27.59
	Wood science	2	6.90
	Administration	3	10.34
	N/A	5	17.24

3.2. System Architecture

We had the following three entities for serving notification messages to our mobile apps: (1) the participants' residences with IoT sensors installed for measuring temperature and CO_2, (2) the PHOENIX server that recorded indoor condition data and sent notification messages, and (3) Firebase server worked for our mobile apps. Figure 1 depicts the overall system architecture.
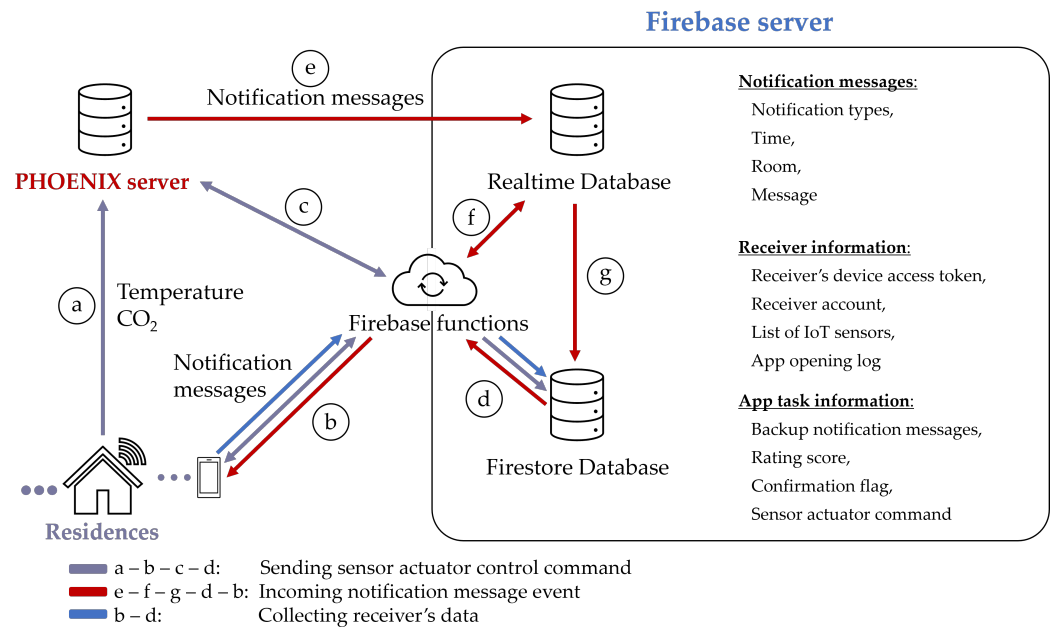

Figure 1. The system architecture of the PHOENIX solution consists of the PHOENIX server, Firebase server, and our mobile apps.

First, each residence installed an IoT sensor on a radiator in each room and one CO_2 sensor to check the residence's CO_2 level. Every IoT sensor on a radiator had an actuator; thus, the participant could control room temperature by changing the value of the radiator through the actuator. On the other hand, the CO_2 sensor could only measure the indoor CO_2 level without controlling the ventilation system. On average, 11 IoT sensors were installed in each apartment. All IoT sensors were connected remotely to the PHOENIX server for data recording and data analysis for producing notification messages (a in Figure 1).

Second, the PHOENIX server received IoT sensor data from the participants' apartments every five minutes. When the room temperature needed to be updated, the participant sent a control command for the sensor actuator in the target room through the mobile app. The sequence order (a–b–c–d) is highlighted by the purple lines in Figure 1. The PHOENIX server stored the IoT sensor data and the requests for changing a radiator temperature (c in Figure 1) created by a participant through a mobile device (b in Figure 1). The control command of the sensor actuator was also archived in the Firestore database (d in Figure 1).

The PHOENIX server had several algorithms to calculate when to send a notification message to inform advice that could be useful to a receiver for managing energy consumption while maintaining a proper indoor condition for their living. The information in the notification message differed depending on the time and temperature, which was labeled with different notification types. For example, if the room temperature was maintained at a higher degree than the room's average, the type of notification message was a heating alert. In the opposite case, a cooling alert was the type of notification message. In addition, when the period that energy prices become expensive in the near future or when the room temperature changes frequently during the high energy price period, a flexibility recommendation message is sent. Accordingly, each notification message contained a type of

notification, a time the notification message was issued, a room that triggered the notification, and a message on how to behave to achieve an ideal temperature while reducing energy consumption and maintaining the quality of residents' comfort level. Once the notification message was issued, it was stored in the Realtime database within the Firebase server (e in Figure 1).

Third, the Realtime database only stored incoming notification messages from the PHOENIX server. When the PHOENIX server issued a notification message, the Firebase function processed it to deliver to a rightful participant's device. The sequence order (e–f–g–d–b) is drawn by the red lines in Figure 1. The Firebase function monitored the Realtime database to check when it received a notification message (f in Figure 1). When the Realtime database received the notification message, the Firebase function also stored the notification message in the Firestore database (g in Figure 1). The notification message was organized based on the receiver. In order to identify the receiver, the receiver's information was collected when the participants opened our mobile apps with their accounts. On the first connection with their account on our mobile apps, the device access token used to identify the destination of notification messages from the Realtime database was stored (b and d represented with the blue lines in Figure 1). With the information about installed IoT sensors for each participant's apartment and the device access token for each participant's mobile device (d with the red line in Figure 1), the Firebase function identified and delivered every notification message to the rightful receiver (b with the red line in Figure 1). In addition, the Firebase server collected additional information for understanding how the participants perceived the mobile apps and notification messages (b and d represented with the blue lines in Figure 1). For example, the mobile app access history, which was recorded whenever the participants opened the apps, the rating score of a notification message when the participants rated a value of the message's information by clicking a star icon up to five stars, the boolean value of whether the participant read the notification message, and the IoT sensor actuator control command from the mobile apps.

3.3. Interface Designs

We designed two mobile apps with the following different UIs to measure the differences in user engagement scale and perceived usability from participants: (1) traditional UI (TUI) without AR and (2) AR-based UI (ARUI). Due to the differences in how the apps attempted to engage users, each UI supported different input and output interaction modalities—for example, ARUI provided interactions between the AR companion and the participants through additional modalities, such as hand tracking and device location. Table 3 summarizes the input and output interaction modalities each UI supported.

Table 3. Summary of input and output interaction modalities for traditional user interface (TUI) and augmented reality user interface (ARUI).

Modality	TUI	ARUI
Input	Touch	Touch Hand position and palm-side Location (distance) Voice (speech-to-text)
Output	Visual graphics Haptic (vibration) Audio (alarm)	Visual graphics Haptic (vibration) Audio (alarm) Audio (text-to-speech)

As default, finger touch interaction for input modality, visual graphics (i.e., images and texts) for output modality, and mobile device's vibration with an alarm sound to notify incoming messages were implemented in both UIs. Additionally, we added customizable font sizes and switchable languages to adjust with user preferences. Finally, both UIs pro-

vided a city temperature where the participants lived to enable them aware of the outdoor temperature; thus, they could adjust the indoor temperature based on the information.

3.3.1. Traditional User Interface

The TUI showed most information on 2D UI with visual graphics. Finger touch interaction was used as an input modality for ease of use since it was familiar to people with smartphones.

Figure 2a shows the side menu where the participant could check and control IoT sensors (i.e., a sensor actuator). The drop-down menu for each room contained links to the installed IoT sensor. When a notification message came from the Firebase server, the notification message was presented with a colored outline representing the type of notification message (Figure 2b). Red and blue implied a heating and cooling alert, respectively.
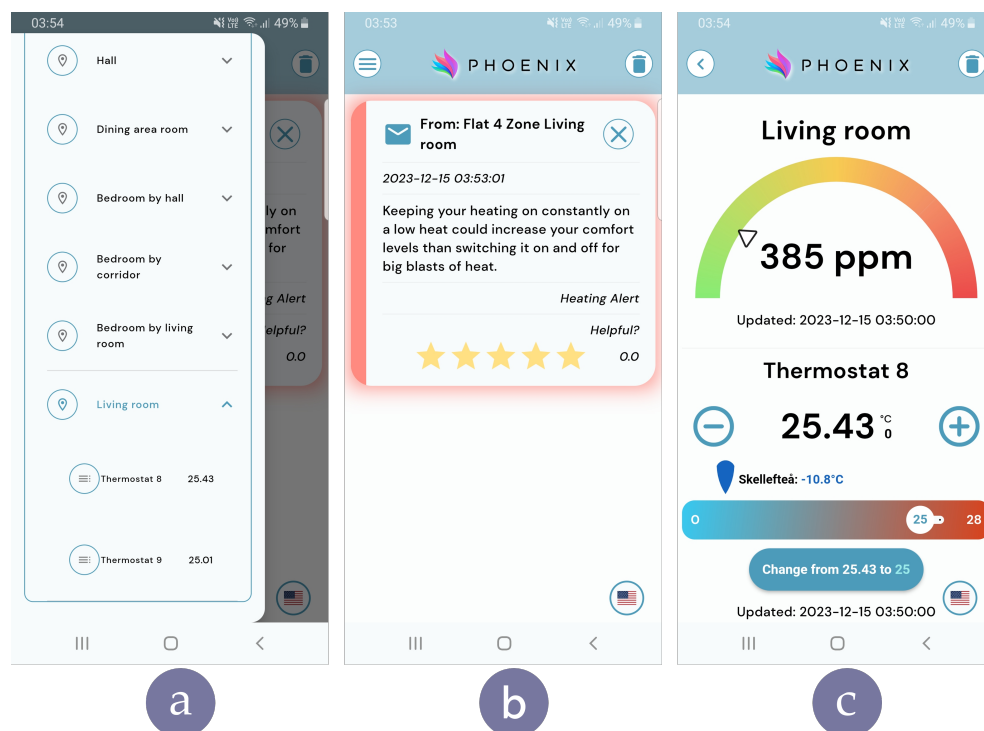

Figure 2. (**a**) Side menu for checking and controlling the IoT sensor values. (**b**) The notification message about a room. (**c**) After clicking the notification message, the page for detailed information on the apartment's room temperature and CO_2 level.

The participants could obtain more detailed information about the room in which the notification message was issued by clicking the message (Figure 2c). The CO_2 level of the apartment and room temperature were available, and the participants could also update the room temperature on this detailed page. The only audio output was a notification alarm when receiving a message. Haptic feedback was triggered along with the alarm sound to enhance the notification for participants.

3.3.2. Augmented Reality-Based User Interface

Unlike TUI, the ARUI required positional data to place an AR content. To reduce the burden of preparation for using the AR, we decided to use a human hand to provide the positional data for AR content rather than using additional materials, such as a printed marker or tangible object. Figure 3a illustrates the scene where the participant's hand became an anchor for placing the AR content. After three seconds of initialization, the AR companion would appear.
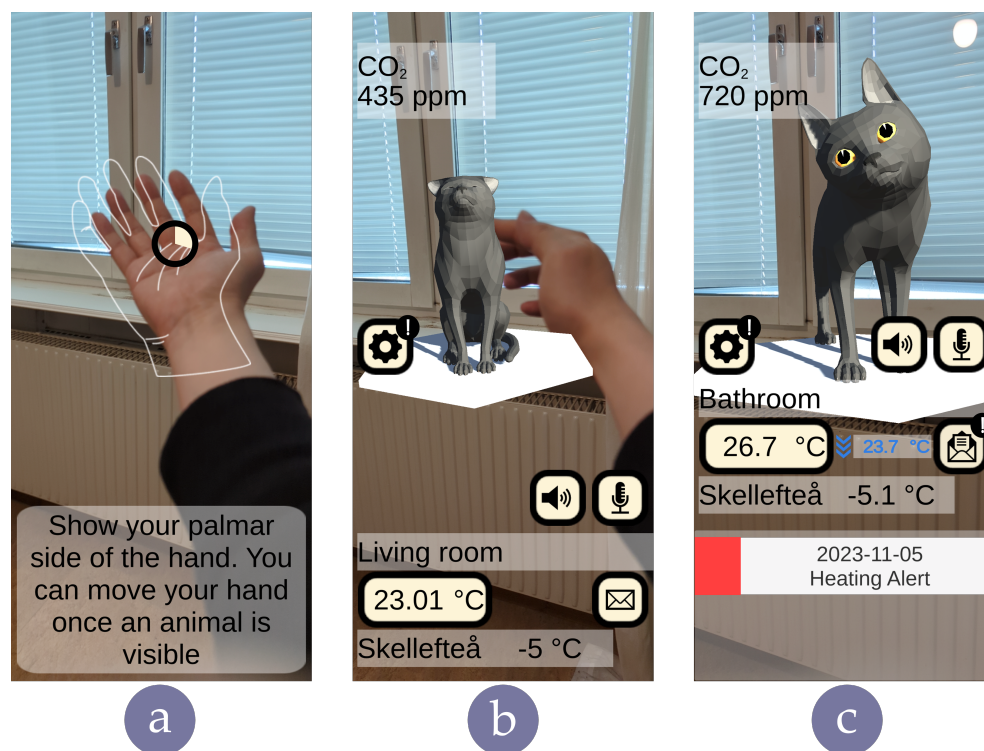
Figure 3. (**a**) Initialization of AR companion based on the hand position. (**b**) Interactive reaction based on the room temperature and hand tracking data. (**c**) A heating alert notification message for the room.

The AR companion played different animations based on the notification message type and the current room condition, which reflected a real-world ambient temperature. We designed each animation to represent the situation that the participant should be aware of and take action to resolve. Figure 3b shows the AR companion's sitting pose, representing that the room temperature was at the level that the participant could feel cozy. The AR companion was designed to play a unique animation for each room condition (e.g., ideal, too hot, or too cold). If the participant felt differently about what the AR companion expressed, we expected the participants to warm up or cool down the room temperature. In addition, the AR companion had several animations to react to the participants' behaviors, such as exhibiting a serene response when the participants pet its head with their hand (Figure 3b), staring at the participants if the participants got close (Figure 3c), and swinging his foot when the participants got too close to the AR companion.

The participants could change the room through the side menu that popped up after clicking the gear icon on the left side of the screen. The button with the microphone icon enabled a voice command for changing the room and its temperature, and the button with the speaker icon initiated an audio output that read the latest notification message delivered to that room.

The notification message could be checked by opening the bottom panel after clicking the button with the message icon on the right side of the screen (Figure 3c). From the ARUI, the participants could control the room temperature by pressing the button with a temperature value on the left side of the screen. The updated room temperature would be displayed next to the temperature control button in a colored text representing whether the changed room temperature is warmer (i.e., red) or cooler (i.e., blue) (Figure 3c).

3.4. Mobile App Development

To run the AR at a high frame rate, we deployed the mobile app with ARUI on Google's Pixel 7 (Android 13). We created the mobile app with TUI using Flutter (version 3.13.2),

a UI software development kit (SDK) published by Google [39], for any device with Android 9 or higher version. We utilized Firebase SDK to enable communication between the mobile apps and the Firebase server. Before distributing our mobile app, we tested every participant's device model on the Android emulator to ensure the app had no critical issues with resolution and performance.

The ARUI was implemented via Unity 2022 [40] instead of Flutter for the following three reasons: First, Unity is a versatile tool for designing AR scenes with intuitive UI and extendable packages such as the AR Foundation [41], which enabled distance-based interaction and allowed the utilization of Google's AR SDK, ARCore [42]. Second, Mano motion SDK [43] enabled vision-based hand tracking. Although the SDK supported various hand gestures as input commands, the recognition performance could become unstable due to the requirement of strict hand movement within a limited area and proper speed in performing gestures. Since we aimed to avoid invoking frustration while the participants interacted with our apps, we wanted to retain simplicity in the interaction modality, which could reduce the cognitive load for the participants. Therefore, we only used hand tracking to recognize hand position and palm side. Lastly, we used Android's native speech-to-text and text-to-speech features to implement a voice command as input and an audio speaker as output, respectively. The mobile apps accepted some keyword combinations as a command for changing a room to display. The audio speaker read aloud the text on the screen. Thereby, the participants could obtain information without reading the texts.

3.5. User Evaluation

3.5.1. Questionnaires

With the user engagement scale (UES) [44] and system usability scale (SUS) [45], we collected feedback about the use of the PHOENIX solution through the mobile apps with the TUI and ARUI from 24 participants (P1–P24, Group A). We utilized UES and SUS to understand whether the ARUI improved user engagement and perceived usability more than the TUI. For five occupants (P25–P29, Group B), we asked the additional 16 questions regarding the overall experience and impression of the PHOENIX solution that the participants experienced through the TUI for about three months (i.e., March–April, June–August). The 16 questions are listed in Appendix A.

The UES was utilized to measure the participants' experience regarding the level of engagement on each UI. We used a long version of the questionnaire with 30 questions, proposed by O'Brien et al. [44], to be answered by the participants with five scales from 'strongly disagree (1)' to 'strongly agree (5)'. We modified the words in the questions to fit into our context. Through the questions, user engagement was measured in the following four dimensions: "aesthetic appeal" (AE), "focused attention" (FA), "perceived usability" (PU), and "reward factor" (RW). Each dimension had from 5 to 10 questions. We randomized the order of questions and hid the dimension label to avoid potential biases that might influence participants' responses [44]. The 30 questions before randomization are presented in Appendix B.

We also asked about the participants' perceived usability of each UI. In total, 10 questions from SUS, proposed by Brooke [45], were used (Appendix C), and the participants were asked to select one option that reflected their feelings between strongly disagree (1) and strongly agree (5).

Both questionnaires were prepared in English and Swedish to reduce confusion among participants. Due to the sample size ($N \leq 30$), the collected questionnaire answers were analyzed with a paired sample *t*-test to determine whether a statistically significant difference exists and descriptive statistics to check the central tendency and dispersion of the data. Apart from the questionnaire answers, we coded user feedback that was spoken by participants during the user evaluations to gain insight into the user perspective.

3.5.2. Procedure

Figure 4 illustrates the procedure of the user evaluation for both UIs with 29 participants. Before the user evaluation, all participants signed a consent form for collecting their answers for research purposes and were informed that they could stop the user evaluation whenever they wanted. All participants voluntarily joined the user evaluation and received a gift as an incentive. We individually conducted the user evaluation for 24 participants (P1–P24, Group A) for up to 100 min by letting them experience our mobile apps firsthand and collecting feedback through questionnaires. Since we met the participants in our lab, we were required to replicate the situation that the user of the PHOENIX solution would face in a real-world scenario. Therefore, we manually issued notification messages as a sample, which were copies of actual notification messages sent by the PHOENIX server to one of five occupants (P25–P29, Group B), during the user evaluation of Group A. We also asked several simple tasks to help the participants fully experience all of the TUI's features and to check whether they understood how to use it. The participants experienced the TUI first ($\leq$30 min), and answered questions from UES and SUS regarding their experience with TUI ($\leq$20 min). The ARUI was evaluated after the TUI with the same procedure. Both UIs were installed on Google's Pixel 7, which could run both UIs with high frame rates.
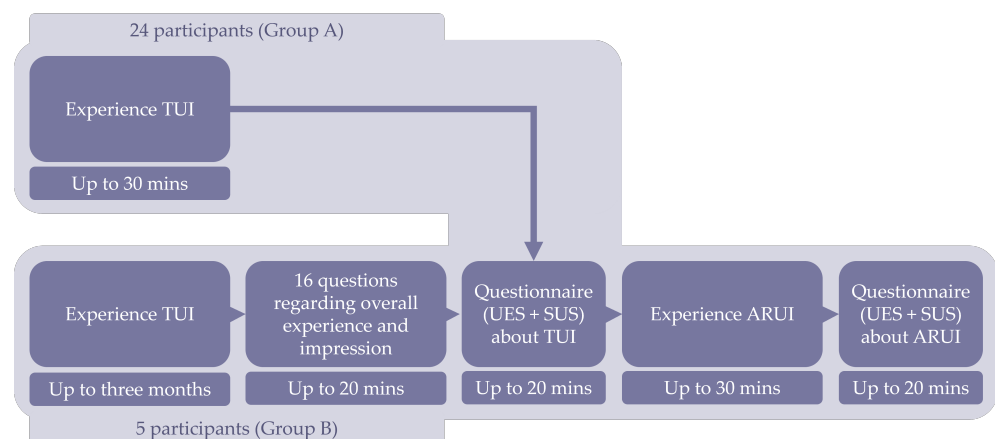

Figure 4. The procedure of the user evaluation.

We installed the mobile app with the TUI on Group B's participants' mobile devices to collect real-world experience-based feedback, as modern mobile devices were able to run our mobile app with TUI without a performance issue. On the contrary, another mobile app with the ARUI was prepared on Google's Pixel 7, which could run the AR with high frame rates, for user evaluation after the TUI. Group B's participants evaluated the mobile app with the TUI after three months of usage in their daily lives. By collecting feedback, we analyzed the overall experience and impression of our mobile app and PHOENIX solution. We then collected UES and SUS about the mobile app with the TUI before presenting the mobile app with the ARUI. The participants in Groups A and B used Google Pixel 7 to experience the AR features in the ARUI. We gave the participants up to 30 min to explore the ARUI and asked them to answer UES and SUS about the ARUI, which took 20 min. Since the mobile app with the ARUI was presented for a certain amount of time, we gave three primary tasks that the participants would have experienced if they used it in their daily lives, such as checking a notification message, finding out how to update a room temperature, and reading detailed information about a room and apartment. During the tasks, they navigated the ARUI freely; thereby, we expected them to experience the interaction with the AR companion.

4. Results

We aimed to determine whether user engagement was positively affected in the ARUI compared to TUI, and if so, which dimension of UES (Appendix B) in the ARUI

produced more positive experiences and what aspects affected user engagement. We then compared SUS (Appendix C) scores between the two UIs to determine whether the ARUI was perceived as more usable than the TUI.

In addition, we had feedback about the overall experience and impression of the PHOENIX solution and the mobile app with the TUI through 16 questions (Appendix A). To understand the end-user perspective in real life, we asked this questionnaire from the participants in Group B, who used the mobile app with the TUI in their daily lives for three months.

4.1. User Engagement Scale

The mean and standard deviation of Group A's UES for the UIs in the four dimensions ("aesthetic appeal: AE", "focused attention: FA", "perceived usability: PU", and "reward factor: RW") are displayed in Figure 5. AE, representing visual attractiveness, was rated higher in the TUI (M = 4.04, SD = 0.56) than ARUI (M = 4.01, SD = 0.81). PU asked about aspects of perceived usability, and the TUI (M = 4.36, SD = 0.42) was perceived as having better usability than the ARUI (M = 4.07, SD = 0.60). RW, which is positive experiential outcomes (e.g., willingness to recommend our app to others and having fun with the interaction), also scored a higher mean in the TUI (M = 3.96, SD = 0.55) than ARUI (M = 3.93, SD = 0.76). The overall engagement score is the sum of all the average scores of the UES dimensions (i.e., AE, FA, PU, and RW). The max score of overall engagement score is 20 and the TUI (M = 15.09, SD = 1.64) received a higher score than the ARUI (M = 15.01, SD = 2.51). Unlike the previous three dimensions and overall engagement score, FA, which represented a concentration level, was evaluated with a higher score in the ARUI (M = 3.01, SD = 1.00) than TUI (M = 2.73, SD = 0.80). To determine whether the data analysis result had a statistically significant difference, we conducted a paired sample *t*-test. Table 4 presents a paired sample *t*-test (right-tailed) results with a degree of freedom of 23.
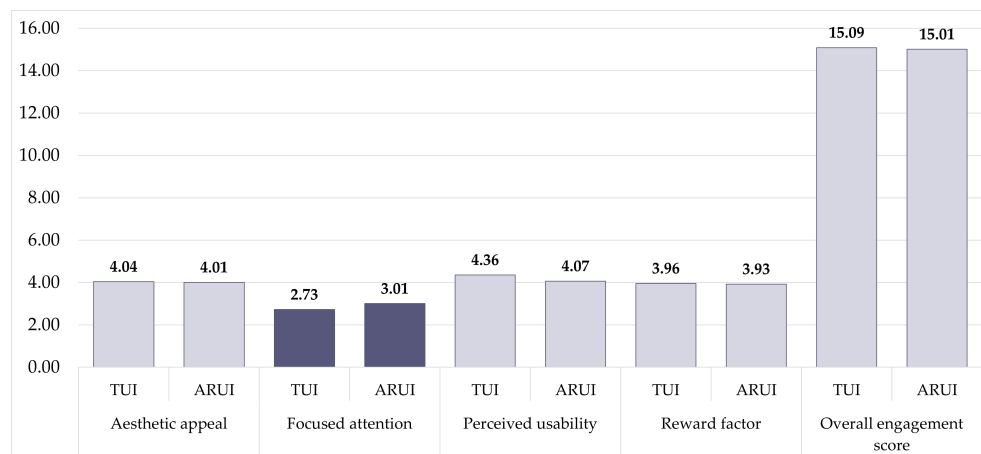

Figure 5. The user engagement scale (UES) result of Group A: means of the TUI and ARUI. The results are grouped into four dimensions, such as "aesthetic appeal (AE)", "focused attention (FA)", "perceived usability (PU)", and "reward factor (RW)", and into the overall engagement score, which is a sum of four scores of each dimension. The dark purple bars in the FA dimension mean the ARUI scored higher than the TUI.

Table 4. The result of the paired sample *t*-test (right-tailed) of Group A: mean difference (M_d), standard deviation for the differences (SD_d), standard error mean for the differences (SE), *p*-value (*p*), test statistic (*t*), and 95% of confidence interval of the difference (CI) in overall engagement score and each dimension ("aesthetic appeal (AE)", "focused attention (FA)", "perceived usability (PU)", and "reward factor (RW)") of UES.

Dimension	M_d	SD_d	SE	p	t	95% CI	
						Lower	Upper
AE	−0.03	0.78	0.16	0.42	−0.21	−0.31	0.24
FA	0.28	0.55	0.11	0.01	2.49	0.09	0.47
PU	−0.29	0.57	0.12	0.01	−2.52	−0.49	−0.09
RW	−0.03	0.40	0.08	0.34	−0.41	−0.17	0.11
Overall	−0.08	1.85	0.38	0.42	−0.21	−0.73	0.57

We used quantile–quantile plots (Q-Q plots) to confirm the normal distribution of the differences in the data. We then checked the *p*-value (*p*) and test statistic (*t*) to confirm whether we can accept the null hypothesis (H_0^1: There is no difference in user engagement between the ARUI and TUI). Since the *p* of AE and RW were higher than 0.05, we failed to reject the null hypotheses, which meant there were no significant differences between the ARUI and TUI in AE and RW dimensions for user engagement. Regarding the PU, *p* showed less than 0.05, which meant we might reject the null hypothesis. The 95% confidence interval also indicated the same evidence that we might reject the null hypothesis for the PU since it did not cross zero. However, the critical value of our case was 1.71 (degree of freedom = 23, α = 0.05, right-tailed test), and the *t* of PU was −2.52, which was far less than the critical value. Therefore, we failed to reject the null hypothesis for the PU since the *t* of PU was away from the rejection region. On the other hand, FA showed a *p* less than 0.05, 95% of a confidence interval that did not include zero, and a *t* of 2.49. As in our case (degree of freedom = 23, α = 0.05, right-tailed test), the critical value was 1.71, which was less than the FA's *t*. Therefore, we rejected the null hypothesis for FA and accepted the alternative hypothesis (H_a^1: The ARUI increases user engagement compared to the TUI) for FA dimension. In summary, there is no significant increase in the ARUI of three out of four dimensions of UES, such as AE (*t*(23) = −0.21, *p* = 0.42, 95% of confidence interval (CI) [−0.31, 0.24]), PU (*t*(23) = −2.52, *p* = 0.01, CI [−0.49, −0.09]), and RW (*t*(23) = −0.41, and *p* = 0.34, CI [−0.17, 0.11]). However, the paired sample *t*-test result showed a statistically significant increase in the ARUI at FA, *t*(23) = 2.49, *p* = 0.01, and CI [0.09, 0.47], meaning that the ARUI provided more engagement than the TUI in the aspect of FA. Overall, the paired sample *t*-test showed that there is no significant increase in the overall engagement score of the ARUI compared to the TUI, (*t*(23) = −0.21, *p* = 0.42, CI [−0.73, 0.57]), indicating that the ARUI did not increase user engagement; however, FA was significantly increased by the ARUI.

Table 5 shows the UES result of Group B, who experienced the PHOENIX solution with the mobile app of the TUI for three months. Since the sample size for this group was too small (*n* = 5) to conduct inferential statistical analysis, descriptive statistics (M and SD) were calculated to check the central tendency and dispersion of Group B's data. We found within Group B that the TUI's overall engagement score is higher than that of the ARUI due to the higher scores for the TUI in AE, PU, and RW dimensions. However, FA score of the ARUI is higher than the TUI, which is similar to Group A's result.

Figure 6 depicts the distribution of UI preference of the participants in Group B based on the mean of AE, FA, PU, and RW. The result showed that the ARUI was selected by four participants (80%), while the TUI was chosen more frequently (60%) in other dimensions, such as AE, PU, and RW. The 'Equal' refers to a case when both UIs received the same score from a participant.

Table 5. Group B's UES result: means (M) and standard deviations (SD) of the TUI and ARUI in the overall engagement score and the following four dimensions: AE, FA, PU, and RW.

Dimension	M (SD)	
	TUI	**ARUI**
AE	3.64 (0.74)	3.36 (0.98)
FA	1.66 (0.63)	2.40 (0.47)
PU	3.68 (1.15)	3.18 (0.46)
RW	3.30 (1.29)	2.90 (0.65)
Overall	12.27 (3.42)	11.84 (1.39)
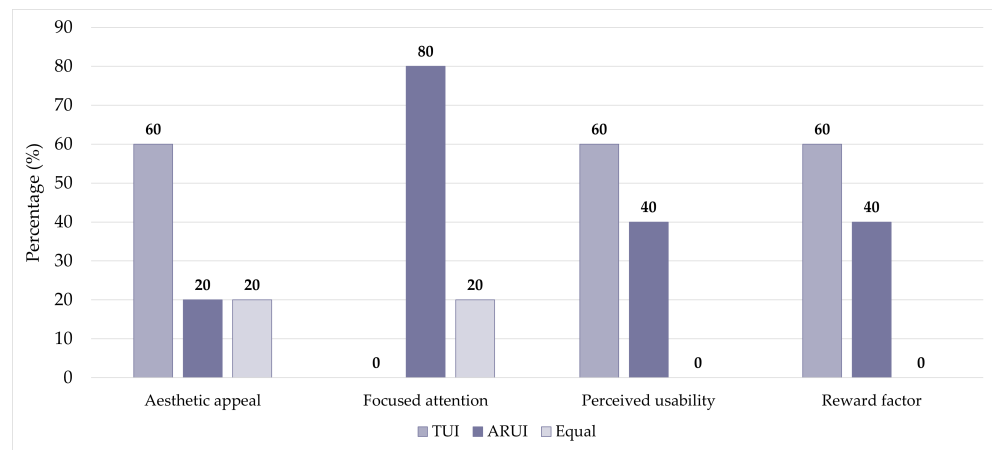
Figure 6. The distribution of preferred UI based on the mean of each UES dimension from the participants in Group B. 'Equal' indicates the means of both UIs are identical.

4.2. System Usability Scale

Figure 7 shows the comparison of each participant's SUS score between the TUI and ARUI. The mean of the overall SUS score of the TUI was 84.17 (SD = 8.71), and the ARUI was 78.13 (SD = 16.07). The bars with thick outlines indicate cases where the ARUI received a higher score than the TUI. Accordingly, six participants (P3, P5, P8, P11, P22, and P24) rated the ARUI higher than the TUI. To determine whether the result had a statistically significant difference, a paired sample *t*-test on SUS was conducted, and Table 6 describes its result.
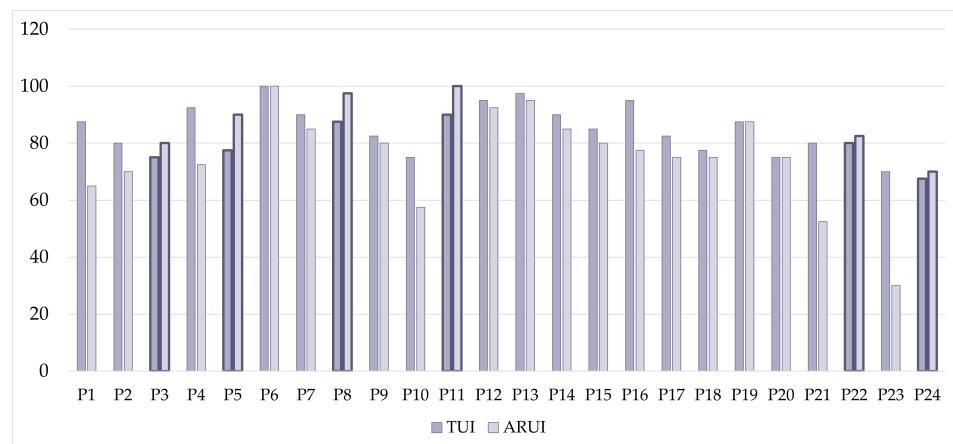

Figure 7. The result of Group A's SUS. The bars with thick outlines indicate the case that the ARUI (light purple bar) scored higher than the TUI (normal purple bar) by a participant.

Table 6. The result of the paired sample *t*-test (right-tailed) of Group A: mean difference (M_d), standard deviation for the differences (SD_d), standard error mean for the differences (SE), *p*-value (*p*), test statistic (*t*), and 95% of confidence interval of the difference (CI) of SUS.

Category	M_d	SD_d	SE	*p*	*t*	95% CI	
						Lower	Upper
SUS	−6.04	12.60	2.57	0.01	−2.35	−10.45	−1.64

The Q-Q plot was used to confirm the normality of the differences, and the paired sample *t*-test results indicated that there is no statistically significant difference in the right-tailed test of SUS scores, $t(23) = -2.35$, $p = 0.01$, CI [−10.45, −1.64].

Table 7 shows Group B's SUS scores for the TUI and ARUI. Three participants (P26, P27, and P28) perceived the mobile app with TUI had better usability than the app with ARUI. However, the average SUS score of the mobile app with the TUI was 69.00 (SD = 26.80), whereas the app with the ARUI was rated 72.00 (SD = 14.83) due to the large difference in SUS scores by P29. The data of P29 could be an outlier in a statistical analysis. However, since there was a difference in the period of time for experiencing each UI and the small sample size of Group B, conducting an inferential statistical analysis was inappropriate due to the low power of a statistical test with Group B's data. Therefore, we kept the P29's data to gain insight into what caused this difference between the two UIs.

Table 7. The system usability scale (SUS) score of the participants in Group B for the TUI and ARUI.

Participant	SUS Score	
	TUI	ARUI
P25	60.00	70.00
P26	77.50	65.00
P27	97.50	95.00
P28	82.50	75.00
P29	27.50	55.00

4.3. Overall Experience after Three Months of Use

Figure 8 represents the result of the overall experience and impression of the PHOENIX solution on the mobile app with TUI, answered by the participants in Group B. Each question's (Appendix A) mean with bars, and the mean range is depicted with different color intensities. For example, a bar with a mean over four is colored dark purple, whereas a mean between three and four is colored with a normal purple. A mean lower than three is filled with a light purple. We identified that the participants were highly interested in optimizing energy consumption to obtain energy cost reduction or other benefits (Q5–Q7 in Figure 8, M = 4.27). Instead of manual management, they preferred that the system support auto-management of indoor conditions based on personal configuration (Q12 and Q13, M = 3.90). The participants answered with an average score of 3.33 for the questions regarding the overall experience of the mobile app with the TUI (Q14 and Q16), the usefulness of the notification messages (Q4, Q10, and Q11), and the influence of energy management on their quality of comfort (Q1, Q3, and Q9). We found three questions received less than an average score of 3, such as a noticeable energy cost reduction (Q2, M = 1.80), an improvement in indoor air quality (Q8, M = 2.60), and frequent use of the mobile app (Q15, M = 2.00).
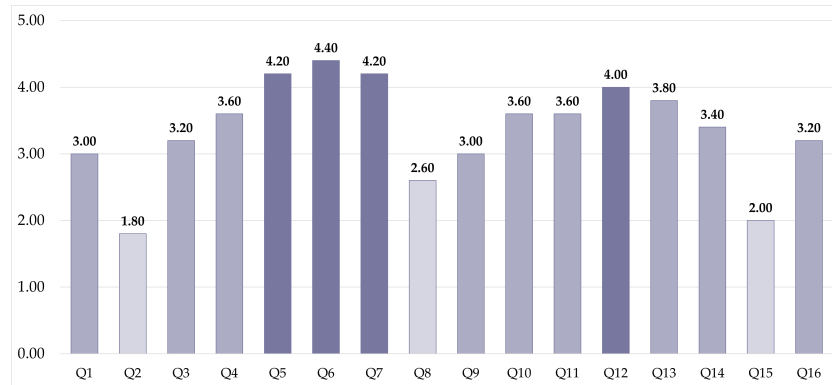
Figure 8. Means (M) of 16 questions (Appendix A) related to the overall experience and impression of the PHOENIX solution (i.e., notification message) on the mobile app with TUI after three months of use by the participants in Group B. The color of the bar implies a certain level of the mean (i.e., light purple: M < 3, normal purple: $3 \leq M < 4$, dark purple: $4 \geq M$).

4.4. Qualitative Findings

We analyzed feedback that the participants noted during the user evaluation. The feedback was analyzed and grouped into four categories, such as AR companion, ARUI, data and interface design, and personalization.

4.4.1. AR Companion

As Figure 9 illustrates, 62.50% of participants in Group A rated a higher score for the ARUI than the TUI in FA. Participants explicitly expressed their interest in the AR companion during the user evaluation. One participant (P9) noted that interacting with the AR companion was fun even after rating higher scores for the TUI in AE, PU, and RW. We then received various proposals to provide more engagement by adding content that users could enjoy. For example, the participants suggested adding more animations (P7, P17, and P29), audio dialogues or sound effects (P3), a feature that the AR companion can be changed to other animals (P2 and P10), and visual effects, such as icons, to emphasize the expression of the AR companion (P9, P17, and P20). While several participants focused on the entertainment aspect of the AR companion, two participants (P5 and P13) were interested in the AR's intuitiveness for effectively presenting data. P5 noted the following:

> "The pet (AR companion) represents the (room) condition explicitly, which means more easy to understand and more engaging". (P5)

Due to these aspects of entertainment and intuitiveness, 54.17% of participants in RW scored the ARUI higher than the TUI. Participants perceived the ARUI could be valuable for other people like kids [46] or older adults [47] who might prefer a UI with improved user engagement.
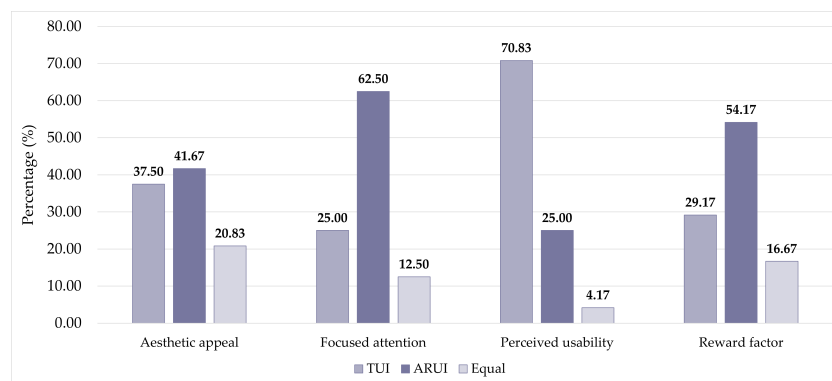

Figure 9. The distribution of user preference of UI for each dimension of UES in Group A based on mean values. 'Equal' refers to the cases if both UIs scored the same mean value.

4.4.2. ARUI

Participants in both groups noted the inconvenience of using their hands for the ARUI. They especially complained about the requirement of aligning their hands on the camera to spawn the AR companion (P11, P22, P24, P26, P27, P28, and P29). Some participants in Group A commented that they preferred the TUI over ARUI due to the familiarity of the UI design compared with apps designed for similar but different tasks, such as monitoring a robot cleaner, tracking parking prices, logging electricity usage, and managing a smart home (P1, P2, P6, P9, P10, and P17). Compared to the ARUI, the common characteristic of those apps was that the participants could see the data right after logging in to the app. The participants with ARUI had additional steps (i.e., holding a camera and showing a hand) to reach the data, and we speculate that it affected their preference for UI. P9 mentioned the following:

> "Depending on the users and context of use, the process to get information should be quicker and simpler or engaging to play". (P9)

In addition, the participants were familiar with reading numbers and graphs to gain information; therefore, understanding the condition of rooms through the AR companion could be fun but less satisfying with obtainable data due to missing details, such as numeric data. Hence, displaying AR on an object in the room (i.e., physical object) or in the air (i.e., markerless) simplifies the process of data acquisition (P9) rather than aligning a hand with the mobile camera, and a dialogue box on the AR companion for information display (P9) could be a potential solution for missing details.

4.4.3. Data and Interface Design

Participants in Group A complained about the quality of the notification message. Due to unfamiliar terminologies and ambiguous meanings of the sentences (P5, P18, P20, and P23), the participants in Group A had a hard time understanding the notification message. Moreover, this ambiguity of the notification message might cause a failure in optimizing energy usage due to unintentional effects. P23 noted the following:

> "My action can still affect a failure in energy optimization if the advice from the (PHOENIX) server is not clear and appropriate. For example, if CO_2 is too high, (the recommendation would ask to) open windows or use a ventilation system to refresh the indoor air. But that also leads to decrease the room temperature if I open a window in the winter season. This is not saving energy because if I feel cold after opening the windows, I would increase the temperature, so to use more energy". (P23)

In order to support understanding of rooms and apartment conditions, the necessity of various sensor data was mentioned as well since our apps provided temperature and CO_2 only. Data like electricity usage (P1 and P2), water usage (P11), temperature history (P14), sensor location (P17, P18, and P23), and saved costs by energy optimization (P23) were suggested as additional information. In this context, additional data visualization methods were requested to deliver data effectively rather than in a simple list, such as line graphs (P10 and P14), bar charts (P2), and 2D maps (P4, P11, P12, P17, P18, P20, and P23). When some participants argued about the missing sensor data, in the meantime, other participants (P4 and P7) pointed out the lack of supportive materials, like a tutorial, for understanding the UIs.

In relation to the lack of tutorials, an unexpected behavior of UI caused confusion among participants. For example, whenever a room was selected in the ARUI, the audio spoke the latest notification message for the selected room aloud. This behavior could only be prevented by clicking the microphone button to mute the audio and was never explained in the app (P7). Another example was that the voice command feature required the participant to speak specific words to activate the command. The app informed the available words while the system listened to user input (i.e., voice command) for 10 s. As a consequence, the participants gave a voice command after the system listening session ended because of the time spent reading the tutorial. Since the participants were unaware

of whether the system was still waiting for their input, the participants were annoyed and confused about the failure. Some participants (P15 and P20) liked the voice command feature, whereas many participants experienced the malfunction. If there were tutorials that explained how to control audio and use voice commands in detail, the participants could avoid discomfort when they tried these features. Supporting natural language for the voice command could resolve this inconvenience (P12) as well.

4.4.4. Personalization

Participants in Group A left feedback regarding a personalized system configuration for the ARUI. For example, the apps already provided four different font sizes and two interface theme colors; however, participants preferred to change to other sizes and colors and set them as default (P4, P10). In addition, participants wanted to adjust the threshold value of room temperature for receiving a notification message (P1, P19) that fitted their residence and daily energy usage pattern. Furthermore, participants who were less in favor of a cat wished to change the species of the AR companion to other animals, such as a dog (P2, P10). Auto energy management was one of the system features that the participants in Group A demanded, as it would enable one-time checks without periodical monitoring. For example, once a rule for the energy management system was set up based on time, date, current room temperatures, residents' presence, monthly energy cost, weather conditions, and outdoor temperatures (P1, P2, P8, and P10), residents could take advantage of the PHOENIX service even without being conscious of the notification message.

5. Discussion

5.1. Focused Attention and Reward Factor

Overall, participants in Group A preferred the TUI over ARUI (Figure 5); however, one dimension of UES that the participants in Group A preferred the ARUI over TUI with a statistically significant difference (Table 4) was FA ($t(23) = 2.49$, $p = 0.01$, CI [0.09, 0.47]). In addition, we found from the descriptive statistics of UES (Table 5) that the participants in Group B had similar experiences to those in Group A despite the fact that there were no inferential statistics for Group B. The participants in Group B rated a higher value to the ARUI than the TUI in FA, whereas the other dimensions of UES were evaluated with better scores in the TUI than in ARUI. We found feedback from the participants to explain the difference in the UES score at FA. When the participants played with the AR companion, they enjoyed interacting with it. Although we prepared only one animation for each scenario (e.g., when a room temperature is too hot, too cold, or ideal) in addition to a few motions (e.g., when a participant gets too close to the AR companion and when a participant pets the AR companion), the participants commented on how enjoyable interacting with the AR companion was. To engage participants further, the FA can be improved if the AR companion has more dynamism with additional features (e.g., switchable animal to cover various user preferences), details (e.g., sound effects and audio dialogue with visual effect, visual improvement to emphasize expression), and animations to react various situation while providing interactable elements [48,49]. Since AE, PU, and RW were proven to correlate with other dimensions, including FA [50,51], we speculate that the ARUI's score in AE and RW could potentially be positively affected if the ARUI's score in FA gets improved.

5.2. System Usability Scale and Perceived Usability

The SUS scores are expressed as percentiles rather than percentages. The average SUS scores of the TUI and ARUI by Group A (TUI: 84.17, ARUI: 78.13) and B (TUI: 69.00, ARUI: 72.00) participants were above 68. Therefore, both UIs of our mobile apps were considered to provide more than average experience since a commonly accepted average SUS score representing average experience is 68 [52]. We found that SUS of the ARUI by Group A had no statistically significant increase from the TUI (Table 6). In fact, there was a significant decrease of SUS in the ARUI compared to the TUI $t(23) = -2.35$, $p = 0.01$, CI [−10.45, −1.64], and Figure 7 showed that only six participants in Group A perceived better

usability in the ARUI over TUI. One dimension, PU, in the UES presented a similar result, which was a negative effect. We identified a significant decrease in the ARUI from TUI at PU of UES, $t(23) = -2.52$, $p = 0.01$, CI $[-0.49, -0.09]$.

In addition, we checked Group B's SUS and UES scores to find the characteristics of their results. Three out of five participants in Group B rated higher scores for the TUI than the ARUI in SUS, and Group B's UES data showed that the TUI in PU achieved a higher value than the ARUI. Although Group B's mean SUS score of ARUI (M = 72.00, SD = 14.83) was higher than the TUI (M = 69.00, SD = 26.80), we presume that P29's SUS score is an outlier that distorted the result of Group B. The potential reasons for P29's SUS score could be the difficulties that P29 faced during the TUI evaluation. For example, during the TUI evaluation, we lent P29 a tablet PC with a much larger display than a smartphone for better readability and easy control with relatively large-size buttons. However, experiencing the PHOENIX solution through the TUI was still a tough task for P29 due to physical constraints caused by aging (age $70 \leq$ P29). Multiple text readings and room temperature controls with finger touches were a burden to P29. Due to body movement difficulty, P29 was issued interacting with the AR companion using her hands. This issue was observed from other participants as well. We assume this issue negatively affected the perceived usability of the ARUI. However, despite an obstacle caused by a limited physical ability, we observed that P29 enjoyed interacting with the AR companion once P29 successfully summoned it. The user behavior-based reactions that the AR companion showed left a favorable impact on her. Moreover, the AR companion's room temperature-based animation appealed to her because she could quickly obtain high-level information about room conditions by watching the AR companion. Apart from the AR companion, P29 preferred the audio output in the ARUI since she could obtain the information by listening instead of reading. P29's case reminds us of the importance of multimodality in supporting various user contexts [53,54].

Regarding the low SUS scores of the ARUI compared to TUI from other participants in both groups, we assume the discomfort, which was caused by physical constraints of aging in hand alignment for using AR, affected the results since perceived usability was influenced by inconvenience. The participants needed to keep one of their hands in the air to spawn the AR companion while holding a mobile device in another hand, and the participants in Group B (age M = 73.40) were hardly able to do both in a stable condition. Even in a simple task, such as positioning the palm towards the mobile device's camera, the participants suffered from a shaky hand, resulting in instability of the AR companion initialization. Another case was the inconvenience of moving and stretching their arms to find a proper distance between the mobile device and their hand for good positioning of the AR companion. Due to the limited physical ability and unfamiliarity with such a posture, positioning the camera and the participant's hand required several attempts to find a proper position, which caused them to feel frustrated and inconvenienced in the first several tries. This discomfort could be resolved by using another device that provides more freedom in hand movement without aligning the view sight of a mobile device as a requirement, such as a glasses-type head-mount display.

Other reasons the participants in Groups A and B preferred the TUI over ARUI could be (1) the familiarity of UI in TUI with other commercial mobile apps for energy monitoring and smart home management, (2) the purpose of the app usage was to acquire information quickly, so the additional steps (i.e., showing a hand) for running ARUI felt like an obstacle to the participants, and (3) since the AR companion provided high-level information via its appearance only, missing numeric data caused uncertainty in understanding the situation. We could identify several solutions to resolve these issues, such as (1) increasing available data and adding different types of data visualization methods (e.g., a graph, chart, or map), (2) allowing different ways to visualize the AR companion (e.g., environmental object-based or markerless AR), and (3) displaying numeric data with the AR companion.

5.3. Data Clarification

Three issues regarding data that could potentially have negatively affected user perception of the ARUI were identified. First, Group B reported that the notification messages were perceived as informative advisors for improving energy efficiency (Q4, Q10, and Q11 in Figure 8). However, several Group A participants noted the notification message's obscurity due to unfamiliar terminology and a poorly written sentence. This issue caused uncertainty about the recommendation of the PHOENIX solution and invoked concerns about ambiguous purposes of actions. Participants proposed to (1) use simple terminology and (2) separate texts into two paragraphs. For example, a reason for the notification message issued and a recommendation for optimizing energy usage. The second issue was missing time series data (e.g., saved energy cost per hour, day, and month; history of energy usage) as other similar mobile apps provide. Since the participants wanted to see such time series data, we could resolve the issue by providing it in addition to others, such as water and electricity usage. Regarding additional data types, the participants wanted to have different data visualization methods rather than plain text in a list or panel, such as a line graph and bar chart. The participants also proposed utilizing a map of the user's residence to allow them to check an overview of their apartment's data. The last issue was the lack of instruction. The ARUI provided instructions on how to initiate an AR companion with a hand and how to use voice command control, whereas other features were unclear. This issue could be solved by providing support to users with text or video tutorials for each feature with enough time for practicing or providing personal assistance face-to-face [55].

5.4. Interface Design and Aesthetic Appeal

The participants in Group A noted that unexpected interactions in UI confused them about using features like audio output and voice commands. We speculate that utilizing tutorials to explain each feature precisely could help users understand the UI, and the involvement of end-users during the app development might identify the issues causing user frustration in the early stage of development [56]. In the design of UI, the ARUI aimed to achieve high readability of data and UI elements from a real-world background; therefore, the visual appearance of ARUI was relatively less prioritized than TUI during the design process. As a consequence, every button in the ARUI used thick outlines to be easily distinguishable apart from the real world. All data were placed in one panel that could be hidden at the bottom of the screen. In addition, unlike the TUI, the ARUI did not adjust its UI size to fit the font size. Since AE is related to visual appearance and attractiveness, redesigning the ARUI by referring to AR [57] and non-AR [58] mobile app design principles could improve the AE. Furthermore, PU was proven to correlate with AE. Therefore, we speculate a possibility that the ARUI's AE could be positively affected if PU gets improved by updating UI with additional tutorials or user-centered design [56].

5.5. Personalization

There was a concern regarding understanding energy usage and cost while keeping comfort at an acceptable level [59]. The acceptable level is a subjective borderline. For example, the PHOENIX solution's policy to issue notification messages for every participant was the same. As a result, two participants (P25 and P29) reported the discomfort they experienced while following the interventions (Q9–Q11 in Figure 8). Therefore, personalized solutions were required to adjust the PHOENIX solution's policy to personal preferences. For example, the threshold for issuing a notification message should be adjustable for each user. In the context of personalization, other app features should also be customizable, such as the species of the AR companion, default font size, and theme color. In addition, the participants wanted a system that could manage the indoor conditions automatically (Q7 and Q8 in Figure 8). If the apps support auto-management based on personalized configuration, the chance of discomfort, which is caused by a human error (e.g., missing an update) and a mismatch between the PHOENIX solution's suggestion and a user's preference, could be reduced.

5.6. Limitations

We gained insights into the PHOENIX solution and the UI designs from 29 participants; however, the sample size in this study is too small to generalize for all age groups, especially for Group B. Moreover, the participants in Group B were end-users who experienced the apps for long-term use, whereas Group A experienced the apps for a shorter duration. We separated the participants into two groups based on the period of time they experienced the apps. This was performed to ensure that the duration of user evaluation and environment were identical for the participants in each group, hence securing the homogeneity of data. Future studies should consider having more number of participants who can have long-term experience with the apps.

6. Conclusions

In this study, we designed and conducted user evaluations for two mobile apps with unique UIs to study the effect of the AR companion on user engagement and perceived usability. The TUI consisted of 2D images and texts, finger-touch interaction for input, and a sound alarm with vibration for output modalities. On the other hand, the ARUI provided multiple modalities for input and output with the AR companion, representing the room condition, spawned on the ARUI using the user's hand. The user interacted with the AR companion with their hand in the real-world. The purpose of the AR companion was to engage the users, thereby increasing user retention on the apps and improving the energy efficiency of a residence as a result. We evaluated the PHOENIX energy management solution and the UIs in terms of user engagement and perceived usability using 29 participants. The UES and SUS results indicate that AR barely affects overall user engagement score and perceived usability; however, the ARUI was preferred over TUI in one dimension of UES; that is, FA. Based on the SUS results, the participants in both groups perceived that both UIs provided more than an average experience (SUS score ≥ 68). We identified design elements that affected user engagement and perceived usability, and we also identified potential solutions to improve user engagement and perceived usability. Future research should study the effect of user engagement and perceived usability on energy efficiency, which was out of the scope of this study. We expect our results to give insight into interaction modality and UI design for researchers and developers interested in understanding an AR companion's effect in an energy management mobile app.

Institutional Review Board Statement: This study did not include any sensitive personal data, as defined by the Act (2003:460) on ethical review of research involving humans (Lag (2003:460) om etikprövning av forskning som avser människor) from Sweden's legislature (Sveriges Riksdag).

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

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Abbreviations

The following abbreviations are used in this manuscript:

EU	The European Union
PHOENIX	Adapt-and-play holistic cost-effective and user-friendly innovations with high replicability to upgrade smartness of existing buildings with legacy equipment
AR	Augmented reality
IoT	Internet of Things
UI	User interface
TUI	Traditional UI
ARUI	AR-based UI
SDK	Software development kit
UES	User engagement scale
AE	Aesthetic appeal
FA	Focused attention
PU	Perceived usability
RW	Reward factor
SUS	System usability scale
M_d	Mean difference
SD_d	Standard deviation for the differences
SE	Standard error mean for the differences
p	P-value
t	Test statistic
Q-Q plot	Quantile–quantile plot

Appendix A

The 16 questions regarding the user experience and impression of the "Adapt-&-Play holistic cost-effective and user-friendly innovations with high applicability to upgrade smartness of existing buildings with legacy equipment" (PHOENIX) solution (i.e., notification messages) that the participants received through the traditional user interface (UI) without augmented reality (AR) (TUI).

1. You have noticed a reduction in your energy usage since using the service;
2. You have noticed a reduction in your energy costs since using the service;
3. You have not felt any thermal discomfort during the interventions;
4. You find the notifications of the service useful;
5. You know the meaning of demand-response (DR) events;
6. You would be interested in engaging with DR events to reduce your energy costs;
7. You would be interested in engaging with DR events to enhance electric grid's operations, in exchange for reduced energy costs or other benefits;
8. You have noticed an improvement in indoor air quality when using the service;
9. You have felt thermal comfort during the interventions;
10. You find service's comfort notifications useful;
11. You find service's health notifications useful;
12. You would like the air quality to be adjusted automatically;
13. You would like the indoor thermal conditions to be adjusted automatically;
14. By using and navigating the app you did not face any problems or challenges;
15. You use the app to monitor or manage your energy usage often;
16. The overall user experience of the app, including ease of use, reliability, and functionality is good.

Appendix B

The user engagement scale (UES)'s 30 questions in four dimensions.
Aesthetic appeal (AE)

1. This PHOENIX app was attractive;
2. This PHOENIX app was aesthetically appealing;

3. I liked the graphics and images of the PHOENIX app;
4. The PHOENIX app appealed to be visual senses;
5. The screen layout of the PHOENIX app was visually pleasing.

Focused attention (FA)

1. I lost myself in this experience;
2. I was so involved in this experience that I lost track of time;
3. I blocked out things around me when I was using the PHOENIX app;
4. When I was using the PHOENIX app, I lost track of the world around me;
5. The time I spent using the PHOENIX app just slipped away;
6. I was absorbed in this experience;
7. During this experience I let myself go.

Perceived usability (PU)

1. I felt frustrated while using this PHOENIX app;
2. I found this PHOENIX app confusing to use;
3. I felt annoyed while using this PHOENIX app;
4. I felt discouraged while using this PHOENIX app;
5. Using this PHOENIX app was taxing;
6. This experience was demanding;
7. I felt in control while using this PHOENIX app;
8. I could not do some of the things I needed to do while using the PHOENIX app.

Reward factor (RW)

1. Using the PHOENIX app was worthwhile;
2. I consider my experience a success;
3. This experience did not work out the way I had planned;
4. My experience was rewarding
5. I would recommend the PHOENIX app to my family and friends;
6. I continued to use the PHOENIX app out of curiosity;
7. The content of the PHOENIX app incited my curiosity;
8. I was really drawn into this experience;
9. I felt involved in this experience;
10. This experience was fun.

Appendix C

The system usability scale (SUS).

1. I think that I would like to use this system frequently;
2. I found the system unnecessarily complex;
3. I thought the system was easy to use;
4. I think that I would need the support of a technical person to be able to use this system;
5. I found the various functions in this system were well integrated;
6. I thought there was too much inconsistency in this system;
7. I would imagine that most people would learn to use this system very quickly;
8. I found the system very cumbersome to use;
9. I felt very confident using the system;
10. I needed to learn a lot of things before I could get going with this system.

References

1. Costello, A.; Romanello, M.; Hartinger, S.; Gordon-Strachan, G.; Huq, S.; Gong, P.; Kjellstrom, T.; Ekins, P.; Montgomery, H. Climate change threatens our health and survival within decades. *Lancet* **2023**, *401*, 85–87. [CrossRef] [PubMed]
2. Lashof, D.A.; Ahuja, D.R. Relative contributions of greenhouse gas emissions to global warming. *Nature* **1990**, *344*, 529–531. [CrossRef]
3. Yoro, K.O.; Daramola, M.O. CO_2 emission sources, greenhouse gases, and the global warming effect. In *Advances in Carbon Capture*; Woodhead Publishing: Cambridge, UK, 2020; pp. 3–28. [CrossRef]

4. European Commission; Statistical Office of the European Union. *Eurostat Regional Yearbook: 2023 Edition*; Publications Office: Luxembourg, 2023.

5. Phoenix H2020—Upgrading Smartness of Existing Buildings through Innovations for Legacy Equipment. Available online: https://eu-phoenix.eu/ (accessed on 29 December 2023).

6. Khan, I.; Hollebeek, L.D.; Fatma, M.; Islam, J.U.; Rather, R.A.; Shahid, S.; Sigurdsson, V. Mobile app vs. desktop browser platforms: The relationships among customer engagement, experience, relationship quality and loyalty intention. *J. Mark. Manag.* **2023**, *39*, 275–297. [CrossRef]

7. Jiang, T.; Yang, J.; Yu, C.; Sang, Y. A Clickstream Data Analysis of the Differences between Visiting Behaviors of Desktop and Mobile Users. *Data Inf. Manag.* **2018**, *2*, 130–140. [CrossRef]

8. Adepu, S.; Adler, R.F. A comparison of performance and preference on mobile devices vs. desktop computers. In Proceedings of the 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 20–22 October 2016; pp. 1–7. [CrossRef]

9. Hamari, J.; Koivisto, J.; Sarsa, H. Does Gamification Work?—A Literature Review of Empirical Studies on Gamification. In Proceedings of the 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 6–9 January 2014; pp. 3025–3034. [CrossRef]

10. López-Faican, L.; Jaen, J. EmoFindAR: Evaluation of a mobile multiplayer augmented reality game for primary school children. *Comput. Educ.* **2020**, *149*, 103814. [CrossRef]

11. Wen, Y. Augmented reality enhanced cognitive engagement: Designing classroom-based collaborative learning activities for young language learners. *Educ. Technol. Res. Dev.* **2021**, *69*, 843–860. [CrossRef]

12. Spadoni, E.; Porro, S.; Bordegoni, M.; Arosio, I.; Barbalini, L.; Carulli, M. Augmented Reality to Engage Visitors of Science Museums through Interactive Experiences. *Heritage* **2022**, *5*, 1370–1394. [CrossRef]

13. Norouzi, N.; Bruder, G.; Bailenson, J.; Welch, G. Investigating Augmented Reality Animals as Companions. In Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Beijing, China, 10–18 October 2019; pp. 400–403. [CrossRef]

14. Angrisani, L.; Bonavolontà, F.; Liccardo, A.; Moriello, R.; Serino, F. Smart Power Meters in Augmented Reality Environment for Electricity Consumption Awareness. *Energies* **2018**, *11*, 2303. [CrossRef]

15. Ho, N.; Chui, C.K. Monitoring Energy Consumption of Individual Equipment in a Workcell Using Augmented Reality Technology. In *Technologies and Eco-Innovation towards Sustainability I*; Hu, A.H., Matsumoto, M., Kuo, T.C., Smith, S., Eds.; Springer: Singapore, 2019; pp. 65–74. [CrossRef]

16. Naret, S.; Annop, T.; Supharoek, S.; Thanadon, J. Visualizing electrical energy consumption with virtual augmented technology and the internet of things. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *1094*, 012010. [CrossRef]

17. Garay-Cortes, J.; Uribe-Quevedo, A. Location-based augmented reality game to engage students in discovering institutional landmarks. In Proceedings of the 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA), Chalkidiki, Greece, 13–15 July 2016; pp. 1–4. [CrossRef]

18. Chang, R.C.; Yu, Z.S. Using Augmented Reality Technologies to Enhance Students' Engagement and Achievement in Science Laboratories. *Int. J. Distance Educ. Technol.* **2018**, *16*, 54–72. [CrossRef]

19. Chao, W.H.; Chang, R.C. Using Augmented Reality to Enhance and Engage Students in Learning Mathematics. *Adv. Soc. Sci. Res. J.* **2018**, *5*. [CrossRef]

20. Dağ, K.; Çavuşoğlu, S.; Durmaz, Y. The effect of immersive experience, user engagement and perceived authenticity on place satisfaction in the context of augmented reality. *Library Hi Tech.* 2023, *ahead-of-print*. [CrossRef]

21. Thirumaran, K.; Chawla, S.; Dillon, R.; Sabharwal, J.K. Virtual pets want to travel: Engaging visitors, creating excitement. *Tour. Manag. Perspect.* **2021**, *39*, 100859. [CrossRef]

22. Chi, N.C.; Sparks, O.; Lin, S.Y.; Lazar, A.; Thompson, H.J.; Demiris, G. Pilot testing a digital pet avatar for older adults. *Geriatr. Nurs.* **2017**, *38*, 542–547. [CrossRef] [PubMed]

23. Mostajeran, F.; Steinicke, F.; Ariza Nunez, O.J.; Gatsios, D.; Fotiadis, D. Augmented Reality for Older Adults: Exploring Acceptability of Virtual Coaches for Home-based Balance Training in an Aging Population. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–12. [CrossRef]

24. Pokric, B.; Srdjan, K.; Dejan, D.; Maja, P.; Vladimir, R.; Zivorad, M.; Petar, K.; Dejan, J. Augmented Reality Enabled IoT Services for Environmental Monitoring Utilising Serious Gaming Concept. *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl.* **2015**, *6*, 37–55.

25. Savitri, N.; Aris, M.W.; Supianto, A.A. Augmented Reality Application for Science Education on Animal Classification. In Proceedings of the 2019 International Conference on Sustainable Information Engineering and Technology (SIET), Lombok, Indonesia, 28–30 September 2019; pp. 270–275. [CrossRef]

26. Suso-Ribera, C.; Fernández-Álvarez, J.; García-Palacios, A.; Hoffman, H.G.; Bretón-López, J.; Baños, R.M.; Quero, S.; Botella, C. Virtual Reality, Augmented Reality, and In Vivo Exposure Therapy: A Preliminary Comparison of Treatment Efficacy in Small Animal Phobia. *Cyberpsychol. Behav. Soc. Netw.* **2019**, *22*, 31–38. [CrossRef] [PubMed]

27. Seitz, A.; Henze, D.; Nickles, J.; Sauer, M.; Bruegge, B. Augmenting the industrial Internet of Things with Emojis. In Proceedings of the 2018 Third International Conference on Fog and Mobile Edge Computing (FMEC), Barcelona, Spain, 23–26 April 2018; pp. 240–245. [CrossRef]

28. Huo, K.; Cao, Y.; Yoon, S.H.; Xu, Z.; Chen, G.; Ramani, K. Scenariot: Spatially Mapping Smart Things within Augmented Reality Scenes. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–13. [CrossRef]

29. Hadj Sassi, M.S.; Chaari Fourati, L. Architecture for Visualizing Indoor Air Quality Data with Augmented Reality Based Cognitive Internet of Things. In *Advanced Information Networking and Applications*; Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; Volume 1151, pp. 405–418. [CrossRef]

30. Mengoni, M.; Ceccacci, S.; Generosi, A.; Leopardi, A. Spatial Augmented Reality: An application for human work in smart manufacturing environment. *Procedia Manuf.* **2018**, *17*, 476–483. [CrossRef]

31. Xue, Z.; Yang, J.; Chen, R.; He, Q.; Li, Q.; Mei, X. AR-Assisted Guidance for Assembly and Maintenance of Avionics Equipment. *Appl. Sci.* **2024**, *14*, 1137. [CrossRef]

32. Jud, L.; Fotouhi, J.; Andronic, O.; Aichmair, A.; Osgood, G.; Navab, N.; Farshad, M. Applicability of augmented reality in orthopedic surgery—A systematic review. *BMC Musculoskelet. Disord.* **2020**, *21*, 103. [CrossRef]

33. Purmaissur, J.A.; Towakel, P.; Guness, S.P.; Seeam, A.; Bellekens, X.A. Augmented-Reality Computer-Vision Assisted Disaggregated Energy Monitoring and IoT Control Platform. In Proceedings of the 2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC), Mon Tresor, Mauritius, 6–7 December 2018; pp. 1–6. [CrossRef]

34. Cho, K.; Jang, H.; Park, L.W.; Kim, S.; Park, S. Energy Management System Based on Augmented Reality for Human-Computer Interaction in a Smart City. In Proceedings of the 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 11–13 January 2019; pp. 1–3. [CrossRef]

35. Bekaroo, G.; Sungkur, R.; Ramsamy, P.; Okolo, A.; Moedeen, W. Enhancing awareness on green consumption of electronic devices: The application of Augmented Reality. *Sustain. Energy Technol. Assessments* **2018**, *30*, 279–291. [CrossRef]

36. Mylonas, G.; Amaxilatis, D.; Pocero, L.; Markelis, I.; Hofstaetter, J.; Koulouris, P. An educational IoT lab kit and tools for energy awareness in European schools. *Int. J. Child-Comput. Interact.* **2019**, *20*, 43–53. [CrossRef]

37. Alonso-Rosa, M.; Gil-de Castro, A.; Moreno-Munoz, A.; Garrido-Zafra, J.; Gutierrez-Ballesteros, E.; Cañete-Carmona, E. An IoT Based Mobile Augmented Reality Application for Energy Visualization in Buildings Environments. *Appl. Sci.* **2020**, *10*, 600. [CrossRef]

38. An, J.; Yeom, S.; Hong, T.; Jeong, K.; Lee, J.; Eardley, S.; Choi, J. Analysis of the impact of energy consumption data visualization using augmented reality on energy consumption and indoor environment quality. *Build. Environ.* **2024**, *250*, 111177. [CrossRef]

39. Flutter—Build Apps for Any Screen. Available online: https://flutter.dev/ (accessed on 29 December 2023).

40. Technologies, U. 2022 LTS Long Term Support Release Overview | Unity. Available online: https://unity.com/releases/lts (accessed on 29 December 2023).

41. Getting Started with AR Foundation | ARCore. Available online: https://developers.google.com/ar/develop/unity-arf/getting-started-ar-foundation (accessed on 29 December 2023).

42. Build New Augmented Reality Experiences that Seamlessly Blend the Digital and Physical Worlds | ARCore. Available online: https://developers.google.com/ar (accessed on 29 December 2023).

43. Manomotion—ManoMotion. Available online: https://www.manomotion.com/ (accessed on 29 December 2023).

44. O'Brien, H.L.; Cairns, P.; Hall, M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *Int. J. Hum.-Comput. Stud.* **2018**, *112*, 28–39. [CrossRef]

45. Brooke, J. Sus: A "quick and dirty'usability". In *Usability Evaluation in Industry*; CRC Press: London, UK, 1996; pp. 189–194.

46. Chen, Y.F.; Janicki, S. A Cognitive-Based Board Game with Augmented Reality for Older Adults: Development and Usability Study. *JMIR Serious Games* **2020**, *8*, e22007. [CrossRef]

47. Aladin, M.Y.F.; Ismail, A.W.; Salam, M.S.H.; Kumoi, R.; Ali, A.F. AR-TO-KID: A speech-enabled augmented reality to engage preschool children in pronunciation learning. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *979*, 012011. [CrossRef]

48. Johnson, W.L.; Rickel, J.W.; Lester, J.C. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *Int. J. Artif. Intell. Educ.* **2000**, *11*, 47–78.

49. Kang, S.H.; Feng, A.W.; Leuski, A.; Casas, D.; Shapiro, A. The Effect of an Animated Virtual Character on Mobile Chat Interactions. In Proceedings of the 3rd International Conference on Human-Agent Interaction, Daegu, Republic of Korea, 21–24 October 2015; pp. 105–112. [CrossRef]

50. O'Brien, H.L.; Toms, E.G. Examining the generalizability of the User Engagement Scale (UES) in exploratory search. *Inf. Process. Manag.* **2013**, *49*, 1092–1107. [CrossRef]

51. O'Brien, H.; Cairns, P. An empirical evaluation of the User Engagement Scale (UES) in online news environments. *Inf. Process. Manag.* **2015**, *51*, 413–427. [CrossRef]

52. Lewis, J.R.; Sauro, J. Item benchmarks for the system usability scale. *J. Usability Stud.* **2018**, *13*, 158–167.

53. Farage, M.A.; Miller, K.W.; Ajayi, F.; Hutchins, D. Design Principles to Accommodate Older Adults. *Glob. J. Health Sci.* **2012**, *4*, 2. [CrossRef] [PubMed]

54. Schiavo, G.; Mich, O.; Ferron, M.; Mana, N. Trade-offs in the design of multimodal interaction for older adults. *Behav. Inf. Technol.* **2022**, *41*, 1035–1051. [CrossRef]

55. Gomez-Hernandez, M.; Ferre, X.; Moral, C.; Villalba-Mora, E. Design Guidelines of Mobile Apps for Older Adults: Systematic Review and Thematic Analysis. *JMIR mHealth uHealth* **2023**, *11*, e43186. [CrossRef] [PubMed]

56. Abras, C.; Maloney-Krichmar, D.; Preece, J. User-Centered Design. In *Encyclopedia of Human-Computer Interaction*; Bainbridge, W., Ed.; Sage Publications: Thousand Oaks, CA, USA, 2004; Volume 1, pp. 445–456.

57. Liang, S. Establishing Design Principles for Augmented Reality for Older Adults. Ph.D. Thesis, Sheffield Hallam University, Sheffield, UK, 2018.

58. Liu, Y.; Zhang, Q. Interface Design Aesthetics of Interaction Design. In *Design, User Experience, and Usability. Design Philosophy and Theory*; Marcus, A., Wang, W., Eds.; Springer International Publishing: Cham, Switzerland, 2019; Volume 11583, pp. 279–290. [CrossRef]

59. Brown, C.J.; Markusson, N. The responses of older adults to smart energy monitors. *Energy Policy* **2019**, *130*, 218–226. [CrossRef]