Trade-offs in Augmented Reality User Interfaces for Controlling a Smart Environment

Connor D. Flick
Western Kentucky University
Bowling Green, Kentucky, USA
connor.flick867@topper.wku.edu

Courtney J. Harris
University of New Orleans
New Orleans, Louisiana, USA
cjharri5@uno.edu

Nikolas T. Yonkers
University of South Florida
Tampa, Florida, USA
nyonkers@usf.edu

Nahal Norouzi
University of Central Florida
Orlando, Florida, USA
nahal.norouzi@knights.ucf.edu

Austin Erickson
University of Central Florida
Orlando, Florida, USA
ericksona@knights.ucf.edu

Zubin Choudhary
University of Central Florida
Orlando, Florida, USA
zubinchoudhary@knights.ucf.edu

Matt Gottsacker
University of Central Florida
Orlando, Florida, USA
gottsacker@knights.ucf.edu

Gerd Bruder
University of Central Florida
Orlando, Florida, USA
bruder@ucf.edu

Gregory F. Welch
University of Central Florida
Orlando, Florida, USA
welch@ucf.edu

ABSTRACT

Smart devices and Internet of Things (IoT) technologies are replacing or being incorporated into traditional devices at a growing pace. The use of digital interfaces to interact with these devices has become a common occurrence in homes, work spaces, and various industries around the world. The most common interfaces for these connected devices focus on mobile apps or voice control via intelligent virtual assistants. However, with augmented reality (AR) becoming more popular and accessible among consumers, there are new opportunities for spatial user interfaces to seamlessly bridge the gap between digital and physical affordances.

In this paper, we present a human-subject study evaluating and comparing four user interfaces for smart connected environments: gaze input, hand gestures, voice input, and a mobile app. We assessed participants' user experience, usability, task load, completion time, and preferences. Our results show multiple trade-offs between these interfaces across these measures. In particular, we found that gaze input shows great potential for future use cases, while both gaze input and hand gestures suffer from limited familiarity among users, compared to voice input and mobile apps.

CCS CONCEPTS

• Human-centered computing → Empirical studies in visualization; **• Computing methodologies → Mixed / augmented reality**.

SUI '21, November 9–10, 2021, Virtual

ACM ISBN 978-1-4503-9999-9/18/06??…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

KEYWORDS

Augmented Reality, 3D User Interfaces, Internet of Things, Connected Devices, Smart Home, Usability

ACM Reference Format:

Connor D. Flick, Courtney J. Harris, Nikolas T. Yonkers, Nahal Norouzi, Austin Erickson, Zubin Choudhary, Matt Gottsacker, Gerd Bruder, and Gregory F. Welch. 2021. Trade-offs in Augmented Reality User Interfaces for Controlling a Smart Environment. In *SUI '21: ACM Symposium on Spatial User Interaction, November 9–10, 2021, Virtual.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Over the past decade, interconnected devices existing as part of the Internet of Things (IoT) have been growing more prevalent among consumers in multiple settings, including home, office, healthcare, public, entertainment, and industry use, with an estimated 25 billion IoT devices that reached consumers by 2020 [30]. The Internet of Things is defined as a network of physical devices that are digitally connected together, using standard networking protocols to communicate information between digital and physical settings [15, 38]. As IoT devices continue to proliferate and replace more traditional non-connected devices at a growing rate, there exists a growing need to evaluate how users may interface with these devices physically and digitally.

IoT interfaces have largely built upon preexisting technologies that would be common and familiar to an average consumer. Mobile and web apps have particular dominance in this area, taking hold in home [49], industry [39], and healthcare environments [50], as a few examples. Additionally, intelligent virtual assistants (IVAs) using artificial intelligence to deliver and control Internet-based services have also gathered prevalence, with 20% of consumers using a smart speaker with an IVA (e.g., Amazon's Alexa) and 41% using a smartphone-based IVA [20]. These IVAs often include native support for IoT control, offering consumers the ability to avoid a myriad of apps in favor of a voice-operated IoT interface.

However, given the unique combination of physical and digital connections provided by IoT, augmented reality (AR) represents an opportunity to create novel and enhanced IoT interfaces that may offer greater accessibility over traditional interaction techniques [18, 40, 46]. AR, in one of its earliest characterizations, is defined as working to "merge electronic systems into the physical world instead of attempting to replace them" [47]. As IoT allows for physical objects to take on interconnected digital manifestations, AR interfaces are poised to complement this by introducing physical manifestations of interconnected digital objects. As AR continues to see strong growth among consumers [28], active consideration and continued research into AR's role in IoT systems is necessary.

Our hypothesis is that certain methods of controlling IoT devices via AR, particularly gestural or gaze-based approaches, will be perceived as superior to conventional IoT methods and interfaces. In this paper, we examine the potential synergies between AR and IoT and explore the development of two promising AR-IoT interfaces. We asked participants in a human-subject study to perform a predefined set of tasks using different AR and non-AR interfaces to elicit specific responses from common IoT devices and then complete questionnaires based on their experiences with each method of control. We assessed each method on a functional (pragmatic) and an aesthetic (hedonic) basis.

We specifically investigated the following research questions:

- RQ1 Are there user experience benefits of AR interfaces for IoT devices when compared to methods already in use?
- RQ2 Are there cognitive and performance benefits of AR interfaces for IoT devices compared to methods already in use?
- RQ3 How are user preferences for AR and non-AR interfaces influenced by factors such as privacy, security, and familiarity?

This paper is structured as follows. Section 2 presents an overview of related work. Section 3 describes the human-subject study. The results are presented in Section 4 and discussed in Section 5. Section 6 concludes the paper.

2 RELATED WORK

In this section, we provide an overview of related work on user interaction in smart connected environments. This section is split into two main subsections, the first of which covers common modern interaction methods for controlling IoT devices (e.g., voice input, mobile apps), and the second of which covers interaction methods enabled by AR technology (e.g., gaze input, hand gestures).

2.1 Traditional Interaction Methods

The most common method for controlling IoT smart devices in a modern day context is via the use of mobile and web apps that can be interfaced with other smart devices (e.g., smartphones, tablets) [9]. While these apps are well-suited for many use cases and environments in our society, such as smart home, smart city, and smart factory environments [31], multiple limitations currently exist, including interoperability challenges imposed by the prevalence of non-standardized apps [2]. Typically, each device has its own specific app for controlling the device's features, which limits the scalability and usability of this interaction method for IoT environments with large numbers of diverse types of devices [51].

Ledo et al. investigated an approach to overcome some of these limitations with mobile app interfaces called *proxemic-aware controls*, in which the location of the user's smartphone is tracked in relation to nearby IoT devices [24]. In this manner, the smartphone can show the locations of these IoT devices, and more detailed controls appear as the user approaches a particular device. While the proximity threshold is configurable in their approach and can be used to limit the number of IoT devices, for example from all devices in a house to all devices in the user's current room, the authors point out that if there is a high density of devices with similar distances to the user then this approach may need modification to further limit the amount of devices that appear on the smartphone. They mention that orientation tracking, or pointing toward the desired device may be a possibility for this.

Xiao et al. investigated a different approach to overcome these limitations, in which users can interact with IoT devices on their smartphone via tapping their smartphone to the device they want to use [48]. In this approach, after tapping the phone, the IoT device is identified through analysis of the electromagnetic emissions of the device, using a machine learning based classifier to determine the object type, and the appropriate interface appears on the user's phone. While this approach consolidates many apps into one interface, it is limited in that physical contact must be made with the device or an object associated with the device, potentially requiring the user to move to several locations to control multiple devices. This could be particularly troublesome for mobility-impaired users or for IoT devices placed in locations that are inconvenient to reach.

Additionally to these mobile and web app based interfaces, more recently, we have seen the advent of IVA-based interaction methods such as used by Amazon's Alexa or Google's Assistant, which utilize smart device technology to interact with IoT smart devices through their line of smart speakers [25]. These smart speakers integrate natural language processing and voice dialog systems to facilitate intuitive voice-based interaction between users and a virtual assistant. These methods have gained increasing popularity in our society, in particular since 2015 when Amazon widely released their first Echo smart speaker with native smart home device integration [34]. However, these methods have shown their own limitations and challenges from limited privacy controls [1] over diversity issues in recognizing voices [37] to scalability issues when it comes to the need for verbally referencing specific IoT devices [21], as IoT devices have to be named and uniquely identified, which can be difficult for people who are not aware of these names.

2.2 Augmented Reality Interaction Methods

AR-based interaction methods have been proposed as potential replacements or supplements for the aforementioned means for controlling IoT smart devices [17]. AR user interfaces can not only visualize relevant information regarding the state of IoT devices in the user's environment, similar to mobile apps, but they can also provide additional means of embodied input for users to interact with these devices, such as via gaze [10, 11, 35] or hand tracking. This coincides with the larger concept of the convergence of separate research fields such as AR and IoT to create a more seamless environment and enhance the user experience [19].

In 2002, pointing techniques for interacting with nearby devices were investigated by Swindell et al. [45]. In their work, they compared pointing-based interaction to a menu-based system and found that users exhibited significantly less cognitive load when using the pointing-based system, providing early support for smart interaction techniques compared to traditional menu style techniques. In 2007, Merril and Maes investigated gaze-based and pointing-based interaction techniques for obtaining information about nearby objects [29]. Their work employs infrared emission based detection of the user's intended device that can be achieved via looking toward the object (head gaze) or pointing at the object. Information about the targeted object is primarily delivered aurally through a Bluetooth earpiece as opposed to employing a visual display. Their results showed that participants had significantly quicker completion times for several search tasks when using the pointing-based interface compared to no interface at all (control group), whereas no significant effects were found for the gaze-based method. This work suggests that gesture-based interaction techniques may offer advantages over gaze-based techniques. In 2013, Chen et al. compared gazed-based interaction techniques to menu-based selection, where they found that users performed significantly better in terms of interaction speed and subjectively preferred the gaze-based technique over the menu-based one [7]. These papers together indicate that gaze-based and pointing-based interaction techniques have significant advantages over traditional menu-based techniques, although additional comparison between gaze and pointing techniques is needed to further establish their respective benefits and drawbacks.

Bittner et al. compared between using physical indicators and smartphone-based AR to gather information about the state of objects in the user's environment [3]. They found that the majority of users preferred the physical indicators, and users commented that it was inconvenient to interact with the environment since one hand was constantly required to hold the smartphone in order to view the AR information. These results indicate that AR HMDs may have an advantage over smartphone-based AR in this regard.

AR gesture-based interactions for controlling IoT devices were investigated by Sun et al. in 2019, where they compared an AR gesture-based interaction method with a more traditional smart phone based interface [44]. They found that interaction times were faster with the gesture-based system to interact with IoT devices.

3 EXPERIMENT

In this section we describe the experiment that we conducted to investigate the three research questions stated in Section 1. Participants were asked to interact with smart home IoT devices in the experiment, using four different user interfaces.

3.1 Participants

We recruited 23 participants for our experiment; 14 male and 9 female (ages 18 to 31, M=21.8, SD=3.5). The participants were members of the local university community: 21 in STEM fields and 2 in non-STEM fields. All of the participants had normal or corrected-to-normal vision; seven participants wore glasses during the experiment, and two wore contact lenses. One participant reported being color-blind, but as our experiment was designed with sufficient luminance differences among the user interfaces, we did not consider this a reason for exclusion. All participants were right-handed. When asked to rate their experience with *VR*, 2 reported no experience, 6 some, and 15 strong experience. For *AR*, 7 reported no experience, 7 some, and 9 strong experience. For *voice agents*, 1 reported some experience and 22 strong experience.

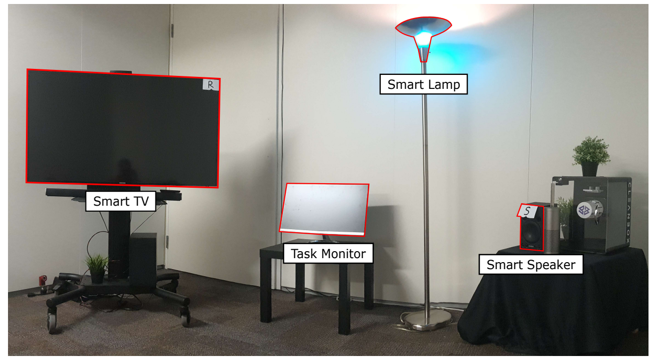


Figure 1: Annotated photo showing the room-scale smart home experimental setup with the three IoT devices and task monitor.

3.2 Materials

To investigate the research questions in an ecologically valuable setting, our materials included a room-scale smart home experimental setup with consumer-level devices and four user interfaces.

3.2.1 Physical Environment and Devices. We used a 2.1 m × 2.1 m isolated room to simulate a smart home environment for this experiment. We added three common IoT devices to the room for participants to interact with, selecting devices for their differences in modality:

- **Smart Lamp**: A Phillips Hue E26 RGB light bulb[1] was placed into a standing lamp.
- **Smart TV**: A Roku online streaming stick[2] was connected to a Samsung TV and used as the main function of the TV. Participants did not interact with any other TV controls except those explicitly related to the streaming stick.
- **Smart Speaker**: A Yamaha MSP3a Monitor speaker[3], receiving output from a hidden Amazon Echo, functioned as an interactive home speaker device.

The physical setup is shown in Figure 1. The arrangement of the room consisted of objects commonly found in a home environment, including a couch, chair, small coffee table, and various plants as well as a small table with a computer monitor atop it. The monitor was used to display instructions to the participants on what tasks were to be completed with the IoT devices. We used a UV CleanBox[4] to sanitize the equipment between use.

We used a Microsoft HoloLens 2 for the AR visual stimulus presentation (see Figure 3). The HoloLens 2 is an optical see-through AR HMD with a field of view of circa 54 degrees diagonally, with a

[1] https://www.philips-hue.com/
[2] https://www.roku.com/products/streaming-stick-plus
[3] https://usa.yamaha.com/products/proaudio/speakers/msp3a/
[4] https://cleanboxtech.com/

resolution of 47 pixels per degree of sight, and a refresh rate of 120 Hz. The HoloLens 2 leverages SLAM-based tracking [6] to localize itself with respect to the physical environment. For the rendering of the visual stimuli, we used the Unity 2019.4.26 game engine and its integration with the HoloLens 2.

3.2.2 Stimuli and User Interfaces. Informed by prior AR interaction methods (Section 2.2) we utilized four different methods of control over the IoT devices. Of these methods, two were custom AR implementations developed with the Unity game engine currently available to developers through usage of base HoloLens 2 features, and the other two served as our baseline conditions that are currently in use among everyday consumers:

Gaze Input. This AR-based technique allowed participants to interact with IoT devices in their environment by looking at them through head movements. For this method, we used spatial mapping to duplicate the physical study environment into the Unity game engine. Invisible planar objects were mapped synchronously over all IoT devices to recognize their locations. Using HoloLens 2 features to track participants' movements, they were instructed to hold their gaze for about one second on target objects to open each respective holographic menu (see Figure 2a). A red circular cursor was constantly projected to give users visual indication of where their gaze is, based upon their head position. Once the menus were opened, the same method was also used to interact with the buttons of each menu (see Figure 3).

Hand Gestures. This AR-based technique allowed participants to interact with IoT devices by drawing their unique symbol identifier in mid-air to bring up their menu, followed by "Airtap" interaction [5]. For this method, we utilized legacy hand-tracking features from the HoloLens. A green sphere was placed over the participant's hand to give visual indication of where their gestural input location is. The participants were instructed to perform and hold a HoloLens Airtap by pressing their forefinger and thumb together. This action changed the green hand-tracking sphere to a red color, which indicated they were now in "drawing mode." The gesture's drawing was comprised of any movement from the tracked hand while the sphere was red. To give the user a visual indication of their drawing path, holographic grey capsules were placed anywhere the hand was tracked, until the user released the Airtap hold (see Figure 2b). Releasing the Airtap changed the hand's sphere color back to green. Then, the drawing was sent into a convolutional neural network machine learning algorithm, called PDollar Point-Cloud Gesture Recognizer[6], where it was compared to a set of pre-established gestures created before the study. Once the new gesture was compared to the pre-established gestures it returned a floating point value to represent its level of accuracy. We created a threshold value of 0.7 or higher to set a level of acceptability for the gestures and prevent gestures from being mistaken as another gesture, which was a common issue at higher thresholds.

Different drawing patterns were assigned to each of the IoT devices within the experimental space. Labels were placed near the devices to indicate their activation gestures to participants. For example, participants could interact with the speaker by drawing an 'S' and with the lamp by drawing an 'L'. However, the machine learning algorithm is not limited to recognizing such basic drawings and can recognize any pre-established drawing. For purposes of this study, a capital letter that could be drawn in a single stroke was arbitrarily assigned to each device to be the activation gesture. We decided on these symbols after pilot testing indicated that they appeared to be mistaken for each other by the Gesture Recognizer less often than other types of symbols we tested, such as numbers or abstract sketches of the shape/outline of the IoT devices. By drawing a specific gesture, a specific holographic menu appeared relative to the device they were trying to connect to. Once the menus were opened, Airtaps were used to interact with the buttons of each menu (see Figure 3).

Voice Input. This IVA-based technique allows participants to verbally control IoT devices by verbalizing the device's name to the voice agent, followed by verbal commands to change its device status. Therefore, we utilized an Amazon Echo device, which we connected to the aforementioned Yamaha MSP3a Monitor speaker. Because there are other popular IVAs on the market (e.g., Google Assistant), we changed the wake command to be "Hey Computer" to avoid too brand-specific interaction. Participants used the IoT smart devices' voice dialogs defined in the corresponding Amazon Alexa Skills to interact with the devices (see Figure 2c).

Mobile App. With this technique, participants used a consumer smartphone and mobile apps to interact with the IoT devices by touch. For the smartphone, we chose a basic commercially available phone, a Google Pixel 2 XL, running Android 11. We connected the devices to the smartphone via each device's respective app. All apps needed to control the IoT devices were placed on the homepage of the smartphone and were accompanied by various other apps to simulate the common look of most smartphone homepages (see Figure 2d).

3.3 Methods

We used a full-factorial within-subjects design in this experiment. As described in Section 3.2, the four conditions were as follows:

- **Gaze Input**
- **Hand Gestures**
- **Voice Input**
- **Mobile App**

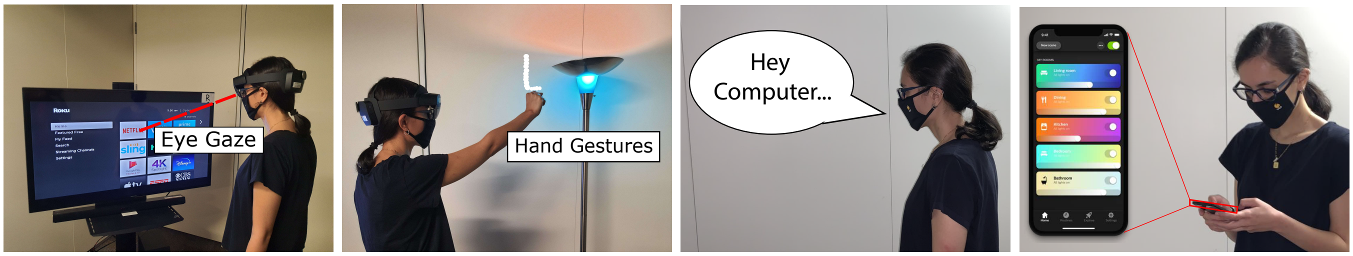
Each participant completed all four conditions. The testing order of the conditions was pre-generated and randomized.

3.3.1 Procedure. Prior to the experiment trials, participants first were asked to give their informed consent. Afterwards, they received a brief description of what AR and IoT devices are, as well as what they were being asked to do.

On the computer monitor (see Figure 1), participants were presented with a list of four tasks that involving all three smart devices in the room. Participants were instructed to complete the tasks in the order they were given, as well as given the choice to stand, sit, or move around as they completed the tasks.

Participants repeated this process four times in separate blocks; each time the method of control alternated between gaze input,

[5]https://docs.microsoft.com/en-us/dynamics365/mixed-reality/guides/authoring-gestures#air-tap
[6]https://assetstore.unity.com/packages/tools/input-management/pdollar-point-cloud-gesture-recognizer-21660



(a) Gaze Input (b) Hand Gestures (c) Voice Input (d) Mobile App

Figure 2: Annotated images of participants using the four methods to interact with the IoT devices in the experimental space.

hand gestures, voice input, and mobile app. Each participant used each method once, immediately after which they were asked to complete subjective questionnaires to provide feedback on that condition. The HoloLens headset was only worn for the gaze input and hand gestures methods, and was set aside for the voice input method, mobile app method, and survey completion. Prior to starting the next condition, the experimenter gave the participant a brief explanation of how to use the interaction technique, and the participants were given the opportunity to ask any relevant questions.

To prevent participants from repeating the same list of four tasks for all conditions, we developed four different variations of the tasks for each method, each of equal difficulty and consisting of subtle variations. For example, one variation's first task may have been to change the lamp's color to green, while for another variation the second task was to change the color of the lamp to blue. There were a total of 16 variations, 4 for each method. A Latin Square was used to randomize the variations per participant, where no two participants were completing the same tasks in the same order back to back. A Latin Square was also used to randomize the order in which each participant used each method of control, allowing for additional variety between participants.

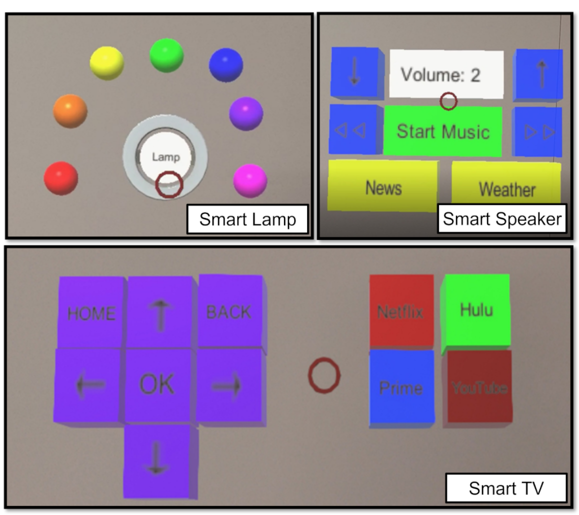


Figure 3: AR screenshots (via HoloLens Remoting in front of a plain wall) showing the basic smart home AR interfaces used in the gaze input and hand gesture conditions.

An experimenter remained in the physical space with the participants as they completed the tasks and discreetly timed each participant for the duration they used each method. Participants were not made aware they were being timed. Participants were able to freely interact with the space without disturbing the IoT devices, including sitting and standing as they wished.

Once participants experienced each user interface and completed the associated questionnaires, they were asked to complete an additional post survey questionnaire.

After completing the post survey, participants were given the opportunity to express any likes, dislikes, or general comments about the technology they interacted with. All participants received monetary compensation upon completion of the post-survey.

3.3.2 Measures. We collected both objective and subjective measures to understand the benefits or drawbacks of the different user interfaces.

We considered the following dependent variables:

- **User Experience**: We used the user experience questionnaire (UEQ) developed by Schrepp et al. [42] to assess participants' user experience with each condition. The questionnaire consists of 26 semantic differential items through which scores are calculated for six dimensions of: *attractiveness*, *perspicuity*, *efficiency*, *dependability*, *stimulation*, and *novelty*. Additionally, the dimensions of *perspicuity*, *efficiency*, and *dependability* can be grouped together to represent the *pragmatic* quality of the experience, and *stimulation* and *novelty* can be grouped together to infer the experience's *hedonic* quality. Answers were given on a 7-point scale.
- **Usability**: We used the system usability scale (SUS) developed by Brooke et al. [5] to assess the usability of each condition. Answers were given on a 1 to 5 scale to express agreement or disagreement with certain statements, where 1 is strongly disagree and 5 is strongly agree.
- **Task Load**: We used the raw version of the NASA Task-Load-Index (NASA TLX) questionnaire developed by Hart et al. [14] to assess the load introduced by each condition. The NASA TLX consists of six sub-scales of *mental demand*, *physical demand*, *temporal demand*, *effort*, *performance*, and *frustration*. Answers were given on a 1 to 20 scale to express

agreement or disagreement with certain statements, where 1 is strongly disagree and 20 is strongly agree.
- **Completion Time**: We stopped the time from the beginning of each condition to the completion of all four smart home tasks as described in Section 3.3.1.
- **Preferences**: We asked participants to indicate their subjective preferences and rank the four user interfaces from most preferred (rank of 4) to least preferred (rank of 1) for the aspects of *physically comfortable*, *familiarity*, *unnoticability*, *security*, and *privacy*. Additionally to these aspects, we asked the participants to rank the user interfaces depending on context to understand which ones they prefer for *smart home use*, *smart city use*, and *smart office use*.

We further debriefed the participants and asked them to verbalize additional qualitative observations and impressions.

4 RESULTS

We used parametric statistical tests to analyze the responses in line with the ongoing discussion in the field of psychology indicating that parametric statistics can be a valid and more informative method for the analysis of combined experimental questionnaire scales with individual ordinal data points measured by questionnaires or coded behaviors [32, 33]. We analyzed the responses with repeated-measures ANOVAs (one factor, four levels) and Tukey multiple comparisons with Bonferroni correction at the 5% significance level. We confirmed the normality with Shapiro-Wilk tests at the 5% level and QQ plots. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity was not met.

Additionally to the analysis of the combined questionnaire scales, for the analysis of the ranking scales we used Friedman tests and Wilcoxon Signed Rank tests with Bonferroni correction at the 5% significance level.

4.1 User Experience

The results for user experience (UEQ) are shown in Figure 4(a).

We found a significant main effect for *Attractiveness*, $F(3, 66) = 4.57$, $\boldsymbol{p = 0.006}$, $\eta_p^2 = 0.17$, for *Perspicuity*, $F(2.16, 47.58) = 9.63$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.30$, for *Efficiency*, $F(3, 66) = 5.28$, $\boldsymbol{p = 0.003}$, $\eta_p^2 = 0.19$, for *Dependability*, $F(3, 66) = 2.90$, $\boldsymbol{p = 0.041}$, $\eta_p^2 = 0.12$, for *Stimulation*, $F(3, 66) = 10.37$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.32$, for *Novelty*, $F(3, 66) = 25.78$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.54$, for *Pragmatic Quality*, $F(3, 66) = 6.74$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.23$, and for *Hedonic Quality*, $F(3, 66) = 21.23$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.49$. The significant post hoc-tests ($p < 0.05$) are shown in the figure.

4.2 Usability

The results for usability (SUS) are shown in Figure 4(b). We found a significant main effect for *SUS Score*, $F(2.1, 45.98) = 6.14$, $\boldsymbol{p = 0.004}$, $\eta_p^2 = 0.22$. The significant post hoc-tests ($p < 0.05$) are shown in the figure.

4.3 Task Load

The results for task load (NASA TLX) are shown in Figure 5(a).

We found a significant main effect for overall (average) *TLX Score*, $F(3, 66) = 5.94$, $\boldsymbol{p = 0.001}$, $\eta_p^2 = 0.21$, for *Mental Demand*, $F(3, 66) = 8.91$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.29$, for *Physical Demand*, $F(1.83, 40.32) = 18.59$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.46$, and for *Effort*, $F(3, 66) = 4.85$, $\boldsymbol{p = 0.004}$, $\eta_p^2 = 0.18$. The significant post hoc-tests ($p < 0.05$) are shown in the figure.

We found no significant main effect for *Temporal Demand*, $F(3, 66) = 0.29$, $p = 0.83$, $\eta_p^2 = 0.01$, for *Performance*, $F(3, 66) = 1.14$, $p = 0.34$, $\eta_p^2 = 0.05$, or for *Frustration*, $F(3, 66) = 2.62$, $p = 0.058$, $\eta_p^2 = 0.11$.

4.4 Completion Time

The results for completion time are shown in Figure 5(b). We found a significant main effect for *Completion Time*, $F(3, 66) = 12.10$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.36$. The significant post hoc-tests ($p < 0.05$) are shown in the figure.

4.5 Subjective Preferences

The subjective preferences of our participants are shown in Figure 6. For this analysis, we used Friedman tests and Wilcoxon Signed Rank tests, and we report Kendall's W to indicate effect sizes.

We found a significant main effect for the *Physically Comfortable* item, $\chi^2(3) = 36.60$, $\boldsymbol{p < 0.001}$, $W = 0.53$, for the *Familiarity* item, $\chi^2(3) = 46.25$, $\boldsymbol{p < 0.001}$, $W = 0.67$, for the *Unnoticeability* item, $\chi^2(3) = 30.60$, $\boldsymbol{p < 0.001}$, $W = 0.44$, for the *Security* item, $\chi^2(3) = 24.86$, $\boldsymbol{p < 0.001}$, $W = 0.36$, for the *Privacy* item, $\chi^2(3) = 44.79$, $\boldsymbol{p < 0.001}$, $W = 0.65$. The significant post hoc-tests ($p < 0.05$) are shown in the figure.

Also, we found a significant main effect for the *Smart Home Use* item, $\chi^2(3) = 19.07$, $\boldsymbol{p < 0.001}$, $W = 0.28$, for the *Smart City Use* item, $\chi^2(3) = 14.53$, $\boldsymbol{p = 0.002}$, $W = 0.21$, and for the *Smart Office Use* item, $\chi^2(3) = 14.64$, $\boldsymbol{p = 0.002}$, $W = 0.21$. The significant post hoc-tests ($p < 0.05$) are shown in the figure.

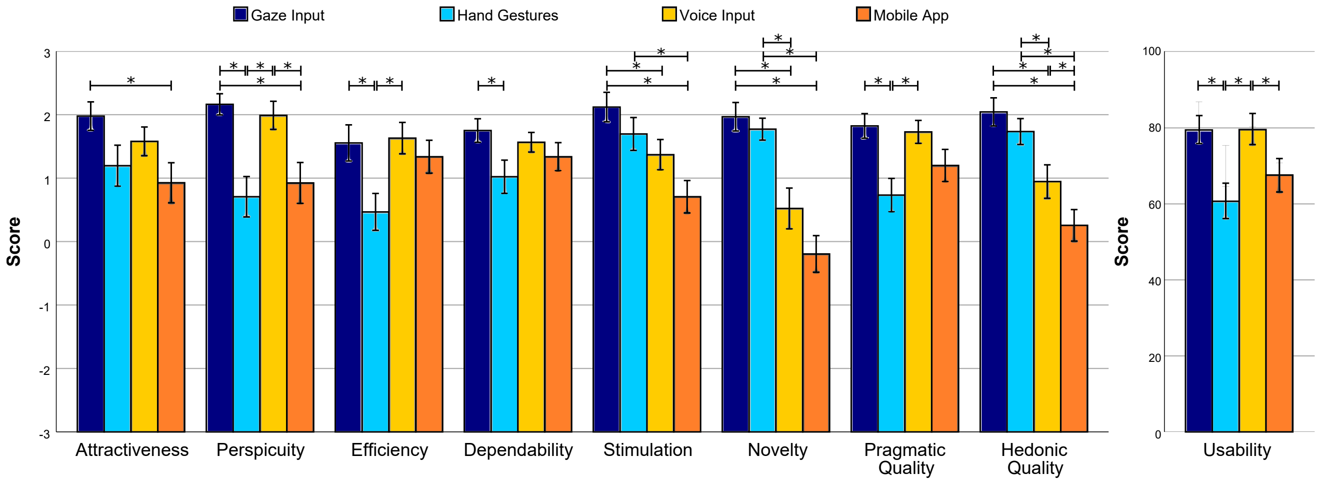
5 DISCUSSION

In this section, we summarize the main findings and discuss the implications for IoT device user interfaces. We discuss the different trade-offs between the user interfaces, and highlight how they are influenced by participants' familiarity with the technologies.

5.1 Influence of User Interfaces on User Experience (RQ1)

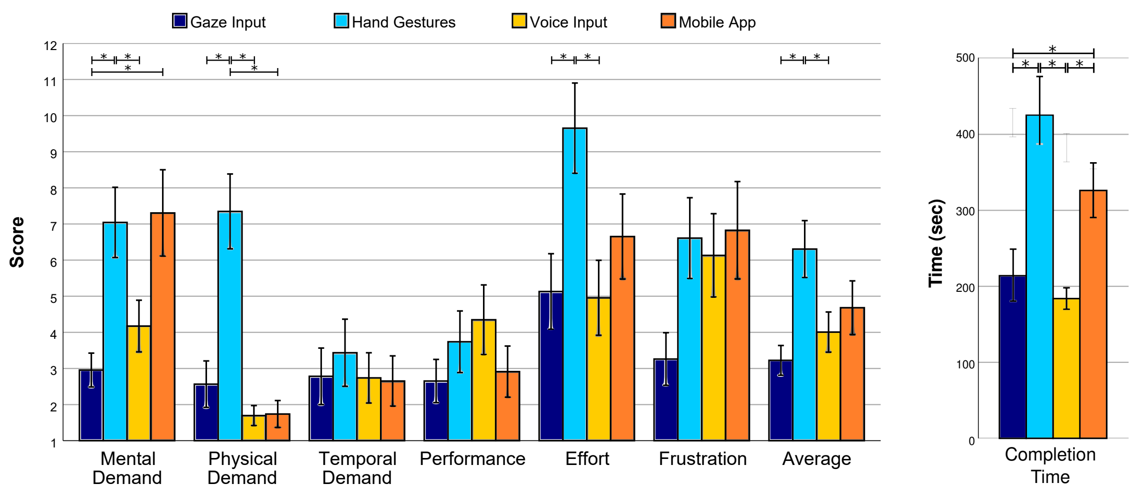
Here we discuss the influence of the different user interfaces on user experience factors by reviewing the significant effects in subjective ratings of user experience and usability by the participants (see Figure 4).

Attractiveness. Participants indicated that they saw the *gaze input* method as more attractive than the *mobile app*. The smartphone required users to navigate around a digital environment of several apps and menus unrelated to their goal, whereas the *gaze input* method required users to navigate their physical environment until they had line-of-sight with their target. The physical navigation with *gaze input* was likely more intuitive overall, coming closer



(a) UEQ (b) SUS

Figure 4: Subjective responses for the four experimental conditions: (a) User Experience Questionnaire (*higher is better*) and (b) System Usability Scale (*higher is better*). Error bars indicate the standard error, and the labeled whiskers indicate significant pairwise comparisons (* p < 0.05).



(a) NASA TLX (b) Time

Figure 5: Results for the (a) NASA Task-Load-Index questionnaire (*lower is better*) and (b) Completion Time (*lower is better*). Error bars indicate the standard error, and the labeled whiskers indicate significant pairwise comparisons (* p < 0.05).

to the familiar constant experience of the physical world and becoming more attractive when compared to the task of navigating through smartphone menus.

Perspicuity. The results for perspicuity suggest that *gaze input* and *voice input* were easy for participants to understand and learn how to use, whereas *hand gestures* and *mobile app* were more difficult. All participants were debriefed on how to use each method and were provided written use instructions to refer back to while completing the tasks, which was expected to eliminate learning differences. Voice and gaze interactions are often presented as natural user interaction mechanisms [13, 43], for instance, humans usually look at a given target object that they are interacting with, which may explain the higher perspicuity scores of these mechanisms. On the other hand, humans do not commonly use gestures to interact with objects and this lack of general familiarity may have influenced participants' perceptions of perspicuity for *hand gestures*. Interestingly, although use of mobile phones and applications is highly common in many societies, still, we observed lower perspicuity scores for the *mobile app* condition. We speculate that this may be due to the fact that participants were not presented with a single unified app interface when interacting with the different

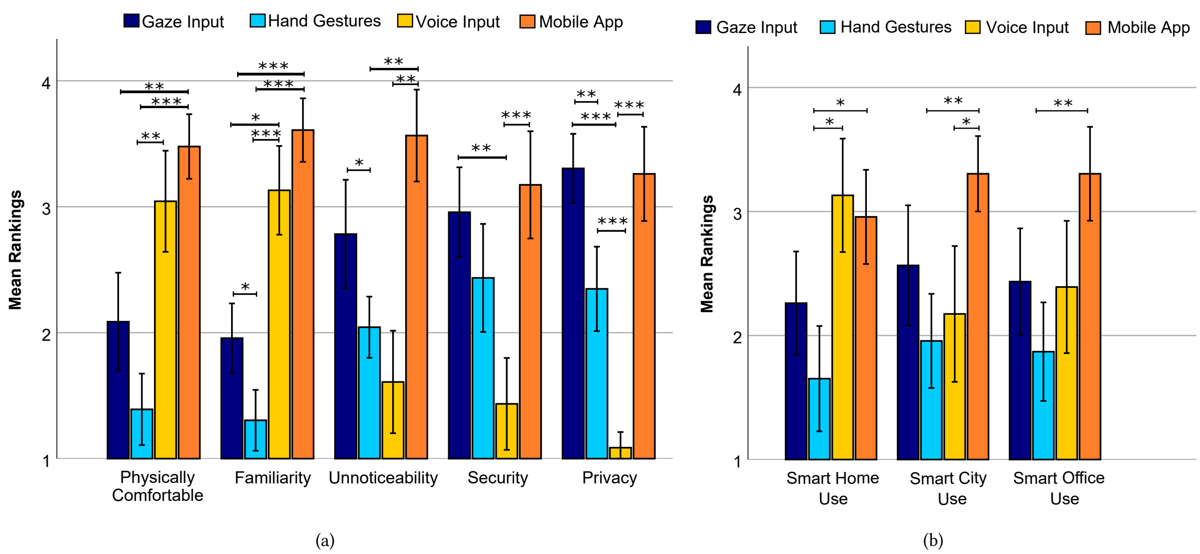


Figure 6: Mean preference rankings for the four conditions: (a) usability aspects and (b) smart environment contexts (*higher is better*; a rank of 4 indicates highest preference, a rank of 1 lowest preference). Error bars indicate 95% CI, and the labeled whiskers indicate significant pairwise comparisons (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

IoT devices and needed to utilize different apps for each device (see related challenges discussed in Section 2.1), which potentially resulted in lower perspicuity scores.

Efficiency, Dependability, and Pragmatic Quality. Participants rated the *gaze input* and *voice input* methods higher than *hand gestures* in terms of efficiency, suggesting that the perceived effort and time taken to complete tasks using the method was greater for *hand gestures* when compared to *gaze input* and *voice input*. The overall pragmatic qualities of each of the methods were rated similarly. In terms of dependability, participants rated *gaze input* higher than *hand gestures*. One factor in this could be the fact that we used a machine learning classifier in the *hand gesture* condition, compared to not relying on such classifiers in the *gaze input* condition. The machine learning classifiers used with the *hand gestures* method would occasionally fail to identify or misidentify user gestures. While we did not log the number of these occurrences in the study, we observed that in most cases participants were able to resolve them by drawing the symbol once again. However, these occurrences may have lowered the efficiency, dependability, and pragmatic quality scores as a result, especially when considering the greater physical and mental load required by the *hand gesture* condition in comparison to *voice input*, which also uses a machine learning model.

Stimulation, Novelty, Hedonic Quality. We found that participants rated *gaze input* and *hand gestures* significantly higher than the *mobile app* regarding stimulation, as well as rating *gaze input* higher than *voice input*. Our results also show that *gaze input* and *hand gestures* were rated higher than *voice input* or the *mobile app* in terms of novelty. This was also true for the hedonic quality, which also shows *voice input* being rated higher than the *mobile app*. Taken together, these results suggest a level of interest, engagement, enjoyment with the AR-based methods that is not reached in the other methods, indicating a larger preference to use AR as a method of control despite currently existing pragmatic drawbacks (such as the weight and discomfort of wearing the AR HMD).

System Usability Scale. Our results show *gaze input* being rated higher than *hand gestures*, and *voice input* being rated higher than the *mobile app* and *hand gestures* on the system usability scale. The scale and its results are largely a reflection of the pragmatic quality of the UEQ and its component parts, being subject to the same factors and reasoning for what may be underlying the results. The results generally indicate that the AR-based methods are able to perform at the same level, if not better than, the more traditional methods of control that exist, whether that be on the more usable or less usable end of the scale.

5.2 Influence of User Interfaces on Cognitive Aspects and Performance (RQ2)

Here we discuss the influence of the different user interfaces on cognitive and performance aspects by reviewing the significant effects in task load and completion times by the participants (see Figure 5).

Mental Demand. The *hand gestures* condition was rated as producing a significantly greater mental demand than *gaze input* or *voice input*, while the *mobile app* was rated as producing a greater mental demand than *gaze input* alone. These demands may be a result of the relative complexity of these menus and the greater number of actions required to navigate them, relative to *gaze input* or *voice input*. The *mobile app* condition requires users to navigate around several apps, in-app menus, and options for control and determine which of these, in the correct order, will complete the task. When compared to something like *voice input*, which is often designed to heavily reduce the voice dialogs the user must navigate and all of the associated information they carry, the *mobile app* method frequently presents as much information as reasonably possible to the user. As a result, the mental demands for the *mobile app* were likely higher overall. The steps needed to reach a menu in the *hand gestures* condition are also typically more involved and require multiple steps and specific hand movements to keep track of, creating a greater mental demand when compared to *gaze input*, which only has one immediately available step to access a menu for any IoT device. Due to the lack of accuracy in the gesture recognizer algorithm and its need for an accuracy threshold, we found most participants required multiple attempts to create a version of the gesture that would pass the threshold. This not only impacted mental demand but physical demand and completion time as well.

Physical Demand. As expected, the *hand gestures* method produced a greater physical demand than any other method. A participant using *hand gestures* would need to have their hand in front of their face to be detected by the HoloLens' hand tracking, sometimes for an extended period of time, which may produce physical strain on the back, bicep, and shoulder whereas the other methods may not. This observation is in line with previous research, sometimes denoted the "Gorilla Arm Syndrome" common to 3D mid-air interaction in VR/AR [16, 27].

Effort and Average. Our findings indicate that the *hand gestures* method was perceived as requiring significantly more effort as well as producing a greater subjective workload overall when compared to *gaze input* and *voice input.* Given the *hand gestures* condition required the highest combined mental and physical demand out of all of the methods, it is reasonable to assume that the reported increase in demand would likely result in a correlated increase in effort to meet these demands. The differences in dependability for the *hand gestures* condition likely also played a role, as errors in the machine learning classifier could cause participants to have to repeat a movement, which increases the effort needed despite maintaining the same mental and physical demands. These factors culminated in the average workload reported, as consistently elevated demand and matching effort from the *hand gestures* method led to a significantly increased burden on the end user. Notably, the differences between the *gaze input* and the *hand gestures* methods for average task load suggest that the factors previously described are centralized to the *hand gestures* method and are not indicative of AR-based control methods altogether.

Completion Time. Our participants typically took longer with the *hand gestures* and *mobile app* methods than the *gaze input* and *voice input* methods. Menu navigation presumably is a driving force behind this disparity, as the overall steps, redundant information, and complexity of reaching a target menu with the *hand gestures* and *mobile app* methods are relatively complex when compared to more immediate and command-based approaches with *voice input* and *gaze input.* Moreover, *gaze input* and *voice input* rely on very common, nearly intrinsic human behaviors whereas the *mobile app* and *hand gestures* rely on less common, more specialized, processes. Tapping into these common and instinctive behaviors likely simplified the *gaze input* and *voice input* methods for participants, reducing their completion times. The relatively high perspicuity of the *gaze input* and *voice input* methods likely also reduced their times, as participants likely required a shorter learning and exploration period to familiarize themselves with the method before attempting their tasks in earnest.

5.3 Factors Influencing Subjective Preferences of User Interfaces (RQ3)

Here we discuss the factors influencing user preferences of different user interfaces by reviewing the significant effects in subjective preference ranking by the participants (see Figure 6).

Familiarity. As expected, participants' ranking scores indicated that they were more familiar with the non-AR methods compared to the AR methods. Among the AR-based methods, *gaze input* was ranked as more familiar compared to *hand gestures.* Considering that gaze can be indicative of one's focus of attention [22], it is possible that our participants were more intuitively accustomed to this form of interaction.

Physical Comfort. Our findings indicated that participants ranked the *mobile app* method as more physically comfortable compared to both AR methods. We also found the *voice input* method to be ranked as more physically comfortable compared to the *hand gesture* method. It is possible that users rated the AR methods as less comfortable for reasons similar to what was found with the work by Bittner et al. [3]. In their work, users noted that it was inconvenient to use a smartphone-based approach compared to an approach in which the IoT information is visible in the environment without use of an additional display. In our work, it is possible that participants considered the act of switching contexts by donning the AR HMD when rating the interaction methods based on physical comfort, as it is certainly more convenient to pull a phone out of one's pocket than it is to don an AR HMD. As AR HMD technology advances and improves in form factor, such devices may eventually complement or take the place of the ubiquitous smartphones that we use today. Should this occur, we believe that this issue of physical comfort may at least be partially resolved.

Smart Home, City, and Office. Participants' preferences for the use of the methods in different environments changed when comparing the more private environments (i.e., smart home) with the less private ones (i.e., smart city and smart office). Both non-AR methods were ranked higher than the *hand gestures* method for the *Smart Home Use* item, while only the *mobile app* was ranked higher than *hand gestures* for the *Smart City Use* and *Smart Office Use* items. Additionally, the *mobile app* was also preferred over *voice input* for the *Smart City* item. This may be due to the lower perceived privacy afforded by a smart city environment and also the potentially higher ambient noise levels.

Unnoticeability, Security, Privacy. Our participants' subjective preference rankings indicated the advantage of the *mobile app* and *gaze input* methods compared to the *voice input* and *hand gestures* methods in matters of noticeability, security, and privacy. Looking at the security item, we found that both the *mobile app* and *gaze input* were ranked as more secure compared to *voice input.*

Additionally, in comparison to *voice input*, all methods were ranked as affording more privacy, while *gaze input* was also ranked as more private than *hand gestures*. This may be a reflection of pre-existing perceptions and expectations about security and privacy for the technologies underlying each method. Smart speakers with voice-based virtual assistants are often seen as insecure, untrustworthy, and lacking adequate privacy safeguards, especially when compared to a smartphone that may allow for more transparent and granular privacy controls [1, 12, 23]. There are also existing concerns about AR data privacy (e.g., gaze-based interaction) [36], but the relative unfamiliarity with AR technologies when compared with smart speakers and voice-based virtual assistants may have caused participants to be less attuned to these risks. Last, we observed that the *mobile app* was ranked as less noticeable compared to the *voice input* and *hand gesture* methods, and the same pattern was observed between *gaze input* and *hand gestures*. These findings are understandable, considering that nowadays individuals carry and use their phones everywhere and utilizing gaze as an interaction mechanisms is inherently a less conspicuous approach to an external observer compared to verbalizing a command or drawing a gesture. However, considering the increase in research supporting gesture-based interaction (e.g., selection and manipulation of virtual content), it is possible to speculate that as people's familiarity with gesture-base interaction grows the more unnoticeable it may become [41].

5.4 Limitations

Participants were comparatively familiar with the non-AR methods in this study, while the AR-based methods were largely new to them. Since level of familiarity can affect interactions with the technology, it is possible that some of the measured benefits of the more traditional methods were due to familiarity and not necessarily due to the mechanisms of the approaches themselves [4]. As AR technology becomes more accessible to the general population and users become more familiar with interacting with AR user interfaces, it is possible that user preferences and performance in using AR-based interaction methods may shift over time. This is something that should be considered as future work continues to investigate the intersection of AR and IoT devices.

While some of the questions in our surveys asked about user preferences in different contexts, e.g., smart home, smart city, and smart office contexts, participants only experienced the interaction techniques within our single experimental environment. It is possible that these preferences may be different if the user is given the opportunity to perform the interactions in each of the mentioned contexts, as there may be factors associated with those contexts that were not considered by the users when making their selections. Such factors could include physical conditions such as ambient noise and lighting, as well as social conditions, such as the amount of people in the environment and the level of familiarity between the user and others around them.

6 CONCLUSION AND FUTURE WORK

In this paper, we presented a comparative evaluation of four user interfaces for controlling IoT smart devices: two AR-based and two non-AR user interfaces. Specifically, we tested *gaze input*, *hand gestures*, *voice input*, and *mobile app*. Our results revealed multiple trade-offs with respect to the pragmatic and hedonic qualities of these methods, with hedonic qualities generally favoring AR-based methods and pragmatic qualities favoring *gaze input* and *voice input*. Overall, our results suggest that *gaze input* has great potential for future IoT use cases, receiving high rankings and ratings compared to the other methods, even considering that participants reported being less familiar with the AR-based methods. Despite relatively high ratings for hedonic and aesthetic measures indicating a willingness to try AR-based methods by participants, both *gaze input* and *hand gestures* still suffer from limited familiarity among typical users, affecting perspicuity and overall perceptions of the methods. We discussed limitations of our study and avenues for future research studying AR interfaces for IoT devices.

Future Work. Based on our results, future work may focus on user experience improvements of the AR-based methods. First, we believe that basic improvements of the *hand gestures* method, specifically more comfortable 3D drawing (e.g., [26]) and improved machine learning for the drawing recognition, could significantly improve the user experience. It also may be valuable to explore more intuitive gestures. For instance, users may prefer specific devices and actions to have iconic gestures associated with them rather than the symbolic characters of this study [8]. In addition to improving these methods, it also may be valuable to explore more ways to use AR to interact with IoT devices, apart from *gaze input* and *hand gestures*. For example, it may be useful to design interaction methods that combine multiple AR and IoT interaction methods or leverage input from other parts of the body.

ACKNOWLEDGMENTS

This material includes work supported in part by the National Science Foundation under Award Number 1852002 (Dr. Jeffrey Forbes, CNS) and Collaborative Award Numbers 1800961, 1800947, and 1800922 (Dr. Ephraim P. Glinert, IIS) to the University of Central Florida, University of Florida, and Stanford University respectively; the Office of Naval Research under Award Numbers N00014-21-1-2578 and N00014-17-1-2927 (Dr. Peter Squire, Code 34); and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch).

REFERENCES

[1] N. Abdi, M. Ramokapane, and J. M. Such. 2019. More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, United States.

[2] G. Aloi, G. Caliciuri, G. Fortino, R. Gravina, P. Pace, W. Russo, and C. Savaglio. 2017. Enabling IoT interoperability through opportunistic smartphone-based mobile gateways. *Journal of Network and Computer Applications* 81 (2017), 74–84. https://doi.org/10.1016/j.jnca.2016.10.013

[3] B. Bittner, I. Aslan, C. T. Dang, and E. André. 2019. Of Smarthomes, IoT Plants, and Implicit Interaction Design. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Tempe, Arizona, USA) *(TEI '19)*. Association for Computing Machinery, New York, NY, USA, 145–154. https://doi.org/10.1145/3294109.3295618

[4] A. Blackler, V. Popovic, and D. Mahar. 2010. Investigating users' intuitive interaction with complex artefacts. *Applied Ergonomics* 41, 1 (2010), 72–92. https://doi.org/10.1016/j.apergo.2009.04.010

[5] J. Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[6] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard. 2016. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics* 32, 6 (2016), 1309–1332.

[7] Y. Chen, B. Zhang, C. Tuna, Y. Li, E. A Lee, and B. Hartmann. 2013. *A context menu for the real world: Controlling physical appliances through head-worn infrared targeting.* Technical Report. UC Berkeley Department of Electrical Engineering and Computer Sciences.
[8] A. El Ali, J. Kildal, and V. Lantz. 2012. Fishing or a Z? Investigating the Effects of Error on Mimetic and Alphabet Device-Based Gesture Interaction *(ICMI '12)*. Association for Computing Machinery, New York, NY, USA, 93–100. https://doi.org/10.1145/2388676.2388701
[9] M. A. El Khaddar and M. Boulmalf. 2017. Smartphone: the ultimate IoT and IoE device. *Smartphones from an applied research perspective* 137 (2017).
[10] A. Erickson, N. Norouzi, K. Kim, J. J. LaViola, G. Bruder, and G. F. Welch. 2020. Effects of depth information on visual target identification task performance in shared gaze environments. *IEEE transactions on visualization and computer graphics* 26, 5 (2020), 1934–1944.
[11] A. Erickson, N. Norouzi, K. Kim, R. Schubert, J. Jules, J. J. LaViola, G. Bruder, and G. F. Welch. 2020. Sharing gaze rays for visual target identification tasks in collaborative augmented reality. *Journal on Multimodal User Interfaces* 14, 4 (2020), 353–371.
[12] J. Foehr and C. C. Germelmann. 2020. Alexa, Can I Trust You? Exploring Consumer Paths to Trust in Smart Voice-Interaction Technologies. *Journal of the Association for Consumer Research* 5, 2 (2020), 181–205. https://doi.org/10.1086/707731 arXiv:https://doi.org/10.1086/707731
[13] L. P. Fu, J. Landay, M. Nebeling, Y. Xu, and C. Zhao. 2018. Redefining natural user interface. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–3.
[14] S. G. Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.
[15] International Telecommunication Union 2012. *Recommendation ITU-T Y.2060: Overview of the Internet of things.* International Telecommunication Union.
[16] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani. 2017. *Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion.* 3328–3339.
[17] J. C. Kim, T. H. Laine, and C. Åhlund. 2021. Multimodal Interaction Systems Based on Internet of Things and Augmented Reality: A Systematic Literature Review. *Applied Sciences* 11, 4 (2021). https://doi.org/10.3390/app11041738
[18] K. Kim, M. Billinghurst, G. Bruder, H. Duh, and G. F. Welch. 2018. Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017). *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 24, 11 (2018), 2947–2962.
[19] K. Kim, N. Norouzi, D. Jo, G. Bruder, and G. Welch. 2022. The Augmented Reality Internet of Things: Opportunities of Embodied Interactions in Transreality. In *Handbook of Augmented Reality*, A. Y. C. Nee and S. K. Ong (Eds.). Springer, Switzerland.
[20] B. Kinsella. 2018. *Over Half of Smartphone Owners Use Voice Assistants, Siri Leads the Pack.* https://voicebot.ai/2018/04/03/over-half-of-smartphone-owners-use-voice-assistants-siri-leads-the-pack/
[21] A. S. Lago, J. P. Dias, and H. S. Ferreira. 2020. Conversational Interface for Managing Non-trivial Internet-of-Things Systems. In *Computational Science – ICCS 2020*, V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. Sloot, S. Brissos, and J. Teixeira (Eds.). Springer International Publishing, Cham, 384–397.
[22] S. RH Langton. 2000. The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology: Section A* 53, 3 (2000), 825–845.
[23] J. Lau, B. Zimmerman, and F. Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (Nov. 2018), 31 pages. https://doi.org/10.1145/3274371
[24] D. Ledo, S. Greenberg, N. Marquardt, and S. Boring. 2015. Proxemic-Aware Controls: Designing Remote Controls for Ubiquitous Computing Ecologies. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 187–198. https://doi.org/10.1145/2785830.2785871
[25] G. López, L. Quesada, and L. A. Guerrero. 2017. Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 241–250.
[26] P. Lubos, G. Bruder, O. Ariza, and F. Steinicke. 2016. Touching the Sphere: Leveraging Joint-Centered Kinespheres for Spatial User Interaction. In *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. 13–22.
[27] P. Lubos, G. Bruder, and F. Steinicke. 2015. Influence of Comfort on 3D Selection Task Performance in Immersive Desktop Setups. *Journal of Virtual Reality and Broadcasting (JVRB)* 12, 2 (2015), 1–14.
[28] M. Mekni and A. Lemieux. 2014. Augmented reality: Applications, challenges and future trends. *Applied Computational Science* 20 (2014), 205–214.
[29] D. Merrill and P. Maes. 2007. Augmenting Looking, Pointing and Reaching Gestures to Enhance the Searching and Browsing of Physical Objects. In *Pervasive Computing*, A. LaMarca, M. Langheinrich, and K. N. Truong (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–18.
[30] P. Middleton, T. Tsai, M. Yamaji, A. Gupta, and A. Reub. 2017. Forecast: Internet of Things — Endpoints and Associated Services, Worldwide, 2017. (December 21 2017).
[31] J. Miranda, N. Mäkitalo, J. Garcia-Alonso, J. Berrocal, T. Mikkonen, C. Canal, and J. M. Murillo. 2015. From the Internet of Things to the Internet of People. *IEEE Internet Computing* 19, 2 (2015), 40–47. https://doi.org/10.1109/MIC.2015.24
[32] C. Mircioiu and J. Atkinson. 2017. A comparison of parametric and non-parametric methods applied to a Likert scale. *Pharmacy* 5, 2 (2017), 26.
[33] G. Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education* 15, 5 (2010), 625–632.
[34] N. Norouzi, G. Bruder, B. Belna, S. Mutter, D. Turgut, and G. Welch. 2019. *A Systematic Review of the Convergence of Augmented Reality, Intelligent Virtual Agents, and the Internet of Things.* Springer International Publishing, Cham, 1–24. https://doi.org/10.1007/978-3-030-04110-6_1
[35] N. Norouzi, A. Erickson, K. Kim, R. Schubert, J. J. LaViola, G. Bruder, and G. Welch. 2019. Effects of shared gaze parameters on visual target identification task performance in augmented reality. In *Symposium on Spatial User Interaction*. 1–11.
[36] J. Outlaw and S. Persky. 2019. Industry review boards are needed to protect VR user privacy. In *World Economic Forum, August*, Vol. 29.
[37] D. Pal, C. Arpnikanondt, S. Funilkul, and V. Varadarajan. 2019. User Experience with Smart Voice Assistants: The Accent Perspective. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. 1–6. https://doi.org/10.1109/ICCCNT45670.2019.8944754
[38] K. K. Patel and S. M. Patel. 2016. Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges. *International Journal of Engineering Science* 6122 (2016).
[39] S. R. Prathibha, A. Hongal, and M. P. Jyothi. 2017. IOT Based Monitoring System in Smart Agriculture. In *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*. 81–84. https://doi.org/10.1109/ICRAECT.2017.52
[40] Z. Rashid, J. Melià-Seguí, R. Pous, and E. Peig. 2017. Using Augmented Reality and Internet of Things to improve accessibility of people with motor disabilities in the context of Smart Cities. *Future Generation Computer Systems* 76 (2017), 248–261. https://doi.org/10.1016/j.future.2016.11.030
[41] S. S. Rautaray and A. Agrawal. 2015. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review* 43, 1 (Jan. 2015), 1–54. https://doi.org/10.1007/s10462-012-9356-9
[42] M. Schrepp, A. Hinderks, and J. Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (2017), 103.
[43] K. A Shatilov, D. Chatzopoulos, L. Lee, and P. Hui. 2019. Emerging natural user interfaces in mobile computing: A bottoms-up survey. *arXiv preprint arXiv:1911.04794* (2019).
[44] Y. Sun, A. Armengol-Urpi, S. N. Reddy Kantareddy, J. Siegel, and S. Sarma. 2019. MagicHand: Interact with IoT Devices in Augmented Reality Environment. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1738–1743. https://doi.org/10.1109/VR.2019.8798053
[45] C. Swindells, K. M. Inkpen, J. C. Dill, and M. Tory. 2002. That One There! Pointing to Establish Device Identity. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology* (Paris, France) *(UIST '02)*. Association for Computing Machinery, New York, NY, USA, 151–160. https://doi.org/10.1145/571985.572007
[46] G. Welch, G. Bruder, P. Squire, and R. Schubert. 2019. *Anticipating Widespread Augmented Reality: Insights from the 2018 AR Visioning Workshop.* Technical Report. University of Central Florida and Office of Naval Research.
[47] P. Wellner, W. Mackay, and R. Gold. 1993. Computer-augmented environments: back to the real world. 36, 7 (Jul 1993), 24+. https://link.gale.com/apps/doc/A14042355/CDB?u=orla57816&sid=bookmark-CDB&xid=c4452318 Article.
[48] R. Xiao, G. Laput, Y. Zhang, and C. Harrison. 2017. *Deus EM Machina: On-Touch Contextual Functionality for Smart IoT Appliances.* Association for Computing Machinery, New York, NY, USA, 4000–4008. https://doi.org/10.1145/3025453.3025828
[49] H. Yang, W. Lee, and H. Lee. 2018. IoT Smart Home Adoption: The Importance of Proper Level Automation. *Journal of Sensors* 2018 (22 May 2018), 11. https://doi.org/10.1155/2018/6464036
[50] Z. Yang, Q. Zhou, L. Lei, K. Zheng, and W. Xiang. 2016. An IoT-cloud Based Wearable ECG Monitoring System for Smart Healthcare. *Journal of Medical Systems* 40, 12 (29 Oct 2016), 286. https://doi.org/10.1007/s10916-016-0644-9
[51] I. Yaqoob, E. Ahmed, I. A. T. Hashem, A. I. A. Ahmed, A. Gani, M. Imran, and M. Guizani. 2017. Internet of Things Architecture: Recent Advances, Taxonomy, Requirements, and Open Challenges. *IEEE Wireless Communications* 24, 3 (2017), 10–16. https://doi.org/10.1109/MWC.2017.1600421