# The Effect of Room Complexity on Physical Object Selection Performance in 3-D Mobile User Interfaces

Maryam Rezaie, Morteza Malekmakan, Ali Asghar Nazari Shirehjini, and Shervin Shirmohammadi, *Fellow, IEEE*

***Abstract*—An important challenge in smart environments is how to manipulate the smart objects. Although mobile applications are typically used for controlling a smart environment, no previous study has evaluated the users performance in manipulating smart objects under different environmental complexities. This article presents an experimental comparison between three different selection techniques 3-D, 2-D, and physical user interfaces (UIs). We evaluate these techniques across two levels of environment complexity measuring 51 participants timing data and errors. Our results indicate that the 3-D UI is superior for task completion time and error, and the 2-D UI is not a better solution than the physical UI when the environment is not complex. The results also show the importance of considering the environment complexity in choosing the proper UI.**

***Index Terms*—Environment complexity, human-computer interaction (HCI), physical object selection, smart office, 3-D user interface (UI).**

## I. Introduction

USERS interact with their homes and offices through switches, buttons, and remote controllers. Recently, mobile devices, such as smartphones, are widely used for controlling smart environments. In general, interaction with physical surroundings includes tasks, such as exploring the target space, mental orientation within the (complex) environment, conducting object selection, and the manipulation of the selected physical objects [1]. In addition, when using 3-D user interfaces (UIs) to virtually represent and interact with physical environments, users need to conduct 3-D translation tasks [1] or virtual navigation [2]. In this context, the term navigation refers to the translation of users, such as walking around in a 3-D virtual space. In general, the task of navigation is about reaching the position of the target device to touch or select it. Navigation also contains the act of orientation (spatial awareness), which is the user's implicit knowledge of his/her position and orientation within the environment during and after travel [3]. Furthermore, to make changes to the environment, such as turning ON lights or changing TV channels, users need to select and manipulate the right device. Thus, interacting with physical surroundings includes the important tasks of finding, selecting, and manipulating target objects, which is referred to as physical object selection [4].

Although numerous researchers have proposed physical object selection techniques to meet specific application requirements and contexts, we are exclusively concerned with evaluating 3-D-based mobile interaction techniques, particularly the effect of environment complexity on object selection performance within the context of interacting with smart environments. Our objective is neither to develop or improve 3-D selection techniques, such as ray-casting [5] or snap-to-it [6], nor to compare various 3-D object selection techniques.

According to [7] and [8], 3-D UIs help users to better orient themselves within complex rooms. However, in certain situations, other selection techniques, such as physical pointing devices [9] or mid-air gestures [10], might be preferable and more usable. For example, some users might find it easier to switch ON a ceiling light using natural gestures instead of using mobile games or virtual reality-based remote controllers. Furthermore, while device-free natural mid-air gestures [10] might be the preferred technique for selecting and manipulating nearby physical objects in smaller rooms, they might prove difficult when attempting to interact with devices at a distance and in large complex rooms. Our previous study [11] has shown that in large and complex environments, users prefer 3-D-based mobile UIs over other techniques; however, the effect was reduced or absent in smaller and simpler rooms.

Determining when best to use mobile 3-D-based object selection techniques requires a scientific study of the effect of environment complexity and selection techniques on selection performance, which has not yet been empirically studied. Thus, to the best of our knowledge, our study is the first to conduct a two-factor analysis of variance (ANOVA) experiment to examine the effects of selection technique and environment complexity on physical object selection performance in a real-world smart meeting room. More specifically, we conducted a $2 \times 3$ between-subject design ANOVA experiment with 51 human subjects. The independent variables were selection techniques (2-D, 3-D, physical) and room complexity level (low,

Manuscript received July 10, 2019; revised December 26, 2019; accepted February 23, 2020. Date of publication April 30, 2020; date of current version July 14, 2020. This article was recommended by Associate Editor G. Serra. *(Corresponding author: Ali Asghar Nazari Shirehjini.)*

Maryam Rezaie is with the Department of Computer Engineering, Sharif University of Technology, Tehran 1458889694, Iran (e-mail: mrezaiee@ce.sharif.edu).

Morteza Malekmakan is with the School of Interactive Arts and Technology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: morteza_malekmakan@sfu.ca).

Ali Asghar Nazari Shirehjini and Shervin Shirmohammadi are with the University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: anazaris@uottawa.ca; shervin@eecs.uottawa.ca).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2020.2984750

high). Selection performance was measured using the selection time and task selection error metrics. We found both main and interaction effects of the selection techniques and room complexity on object selection performance. A post-hoc analysis was conducted using the Tukey's honestly significant difference (HSD) test method, and showed that the average selection time for participants in the 3-D UI group was significantly lower than the other two groups (physical and 2-D UI) in both low and high complexity rooms. Overall, our results showed that the use of 3-D UI increases user performance in object selection. In a less complex environment, no significant differences were evident between the selection performance using a 2-D-based mobile control interface and performing tasks manually.

The main contributions of this article are: 1) the first study that explores the role of environment complexity on user performance in physical object selection; 2) comparison of three types of UIs based on task completion time and error rates to select objects in the context of smart environments; and 3) proposing environment characteristics to differentiate between simple and complex environments.

The remainder of this article is organized as follows. Section II presents a brief review of existing studies that have examined UIs, selection techniques, and environment complexity. Section III describes research methods and hypotheses that were tested within this study. Section IV reports the results of this study; Section V discusses the results. Section VI concludes this article.

## II. Related Works

Selection of an object is one of the basic interactions in smart environments. Because of its importance, numerous studies have examined interaction techniques with smart environments using a mobile device or a smartphone. In this section, we first provide a review of techniques for selecting physical objects. We then discuss related studies that address evaluating physical selection techniques and discuss the difference between these studies and ours.

### A. Physical Object Selection Techniques

The universal remote controller (URC) [12], [13] was one of the early works that focused on interacting with smart appliances. The idea of URC is integrating all the device controllers in one controller. Object selection was done by pressing buttons or labels that represented devices [14]–[16].

Some previous work provided a voice-based device selection approach [17]–[19]. In their methods, users say words or sentences for selecting and controlling physical devices. For example, Kang *et al.* [19] investigated the combination of voice-based interaction with free-hand gestures to select Internet-of-Things (IoT) devices. One disadvantage of voice-based interaction is that it is not proper for all environments; for instance, in a smart classroom, where a lecturer talks, unintended device selection may happen due to false signals.

Other tools such as laser pointers have also been used to help users select physical objects [9], [20]; for example, in PI control system [9], the light sensor on physical objects receives the light emitted from handheld projectors.

Some techniques that do not need users to hold intermediary devices are known as mid-air interaction. For instance, in [21] and [22] users can select devices by gaze. In AmbiGaze, a user approaches a device and sees the animated options which can be selected if followed by eyes. Though faster, gaze-based interaction has some drawbacks such as inaccuracy and unintentional selection [22].

Augmented reality (AR) is also considered for selecting physical objects. Oda and Feiner [23] created a system that uses AR to present physical objects to the user. In their AR-based system, users see the representation of the real-world objects. This virtual representation can be seen through head-worn displays, and users can select objects with the combination of gestures (with a barehand) and a Wii remote. Although this technique provides more selection accuracy in shared environments, it is more time-consuming since its a two-step action; the user has to select an interacting sphere first, then he/she can select the presented object. Raycasting is another selection technique that uses virtual rays to assist users in their selection tasks [24]. Users can cast rays using their hands [1] or their eyes and head [25] in the virtual environment. Once the user touches a point on a touch screen, he/she casts a ray from the touch point. Then, the ray selects the 3-D model that it reaches.

### B. Comparison Studies on Selection Techniques

Since there are numerous selection techniques to point or select objects, studies have been performed to compare the performance of users while interacting with the environment using these techniques. Such comparison studies help designers choose the best selection techniques based on the specific environment they are designing. There are numerous works that compares the selection techniques for virtual objects in virtual environments (e.g. [26], [27]); however, we are interested in works that compare physical object selection techniques, since physical object selection concerns working with real devices that make the selection process more complex. These specific type of comparison studies, which are related to ours, are described here.

In a 22-subject user study, Oda and Feiner [23] compared four physical object selection techniques—"laser pointer," "video share," "virtual hand," and "sphere select." A single-factor experiment was done in a laboratory, including quantitative measurements of selection accuracy and task completion time. Subjects shared an AR environment in which a subject (called indicator) selects physical objects using one of the said techniques, and other participating subjects are called recipients. The study found that when users share a similar view, laser pointer is better than the other techniques since it provides both accuracy and less task completion time; however, when indicator and recipients have different perspectives, the "sphere select" technique is the most accurate even though it is not as fast as the other techniques.

In a recent user study, Wei *et al.* [28] evaluated three UIs—2-D, 3-D, and speech-based UI in an immersive virtual environment. They studied the effect of UI type on user preferences and task completion time with 30 subjects. In

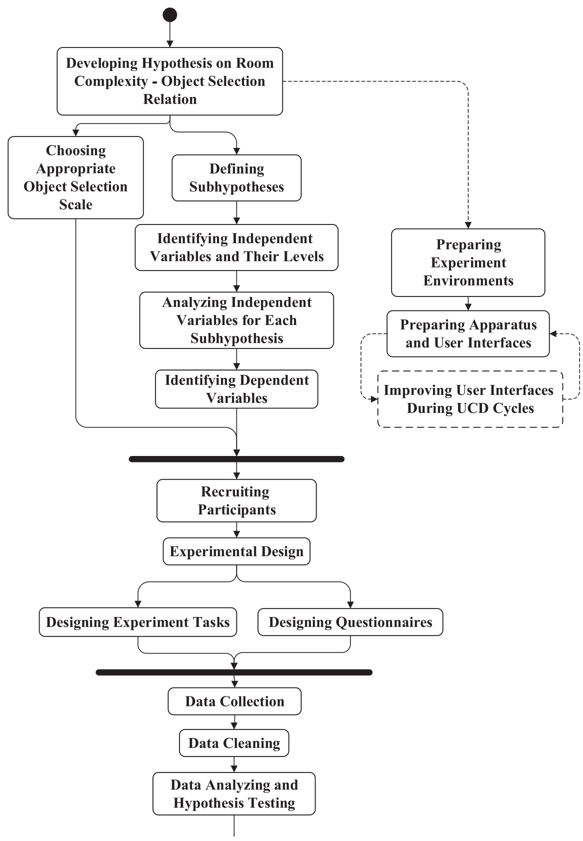


Fig. 1. Steps of the experimental procedure.

order to evaluate system usability, ease of learning, and overall satisfaction, users filled out questionnaires. As a result of their single-factor experiments, they found that even though the speech interface is faster, it has a higher rate of malfunction.

To sum up, in the above studies, the environment did not change throughout the experiment. Our work is different because we evaluate selection techniques in two environments with different levels of complexity, in order to know which selection technique is more suitable given the complexity of the environment. To the best of our knowledge, no other work has studied this environment complexity factor.

## III. Method

We designed an experiment to investigate the effects of UI-type and room complexity level on user performance when selecting objects by interacting with devices in smart environments. The experiment procedure is shown in Fig. 1.

### A. Goals and Hypothesis

The goals of the experiment were threefold: 1) determine how the type of UI affects the object selection performance; 2) investigate whether the characteristics of the rooms and devices affect the object selection performance; and 3) show that room features such as size, number of devices, number of controllers, and other features might help to differentiate a complex room from a simple one.

TABLE I
Characteristics of the Two Different Rooms

| Room characteristic | Conference room | AmI Lab |
|---|---|---|
| Room size | 83 $m^2$ | 23 $m^2$ |
| Number of interactable devices | 30 | 12 |
| Number of controllers (physical) | 11 | 11 |
| Density of devices | higher | lower |
| Density of controllers | higher | lower |
| Number of similar devices | 28 (5 groups) | 4 (1 group) |
| Controller complexity | higher | lower |
| Number of 3-D models (3-D UI) | 96 | 27 |

Before conducting the experiment, we hypothesized that if the environment complexity increases, the object selection performance will significantly decrease. We also hypothesized that using a 3-D UI will increase the object selection performance. Finally, we were also interested in examining the interaction effect of these two factors.

### B. Environment

The level of environment complexity is an attribute of the environment that can be described by dimensions such as quantity (number of items) and variety (different kinds of the same item). The increase in the value of these dimensions will lead to the increase in perceived complexity, which is the users interpretation of the complexity [29]. Studies have shown that environment complexity or object density can affect user performance in some cognitive interaction tasks such as navigation [30]–[32], orientation [33], and object selection [34] in 3-D virtual environments.

As recommended by [29], we chose rooms that were quite different in terms of quantitative complexity metrics. Table I lists the characteristics of each room including the size of the room, the number of devices, the number of controllers (which makes choosing the right controller harder), the density of the devices, and the number of similar devices (which leads to similar controllers that make mapping harder). The higher the values of these attributes, the more complex the environment.

We conducted our experiment in two real environments with devices that were equipped to be controlled by two types of UI. One of the rooms, The Ambient Intelligence Laboratory (AmI Lab), resembled a smart workplace (see Fig. 2), and the other room was a smart conference room (see Fig. 3). Both rooms were located at Sharif University of Technology. The devices circled in Figs. 2 and 3 are interactable and could be manipulated using the UIs. To study the effects of UI-type on object selection performance, we considered three kinds of UI—2-D UI, 3-D UI, and physical UI.

2-D UI: As shown in Fig. 4, for each device in the room, the 2-D UI has a corresponding button with the same name as the device. To manipulate any device, the user has to find and tap on this corresponding button.

3-D UI: The 3-D UI is a collection of 3-D models of the real room. For each device in the room, there is a 3-D model in the scene, as shown in Fig. 5. When the user taps on a 3-D model

Fig. 2. AmI Lab was used in the experiment as the less complex room. The marked objects were interactive: 1) standing light; 2) lampshade; 3) printer; 4) ceiling lights; 5) air conditioner; 6) tea maker; 7) standing ventilator; and 8) monitor.

Fig. 3. Conference room was used as the more complex room. The interactable devices in the conference room are circled: 1) video projector; 2) electric projector screen; 3) different types of lamps; 4) air conditioners; and 5) electric roller blinds.

on the mobile application, the corresponding device will react (i.e., turn-ON/OFF, go up/down) according to its initial state.

Physical UI: We also compared the users' performance when manipulating the devices directly without any application; we call this type of UI "physical."

## C. Apparatus

We conducted the experiment in two rooms described above: 1) the AmI Lab which consisted of a tea maker, a fan, a chiller, a printer, table and floor lamp shades, a floor lamp, a monitor and four ceiling lamps; and 2) the conference room which consisted of a video projector, an electric projector screen, different kinds of lamps, two air conditioners, and three electric roller blinds.

Each participant was given a Sony Xperia Z phone with five-inch display to control the devices in the rooms. Two mobile applications which were made with the famous Unity3D engine were used as the UIs. Several z-wave wall plugs and smart relay switches were used in both rooms to make connections between the real devices and the applications. To minimize user distraction, we removed every object, tool, and device that was unrelated to the context of the smart environment.

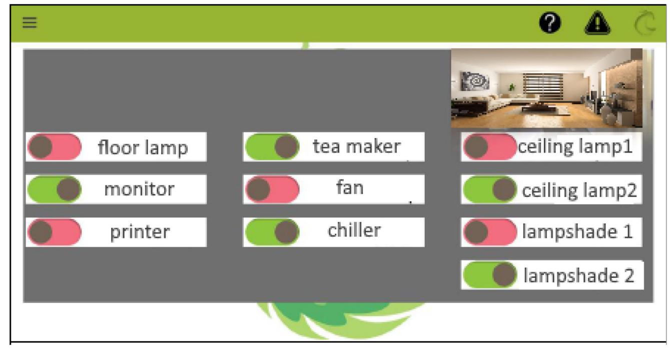


Fig. 4. 2-D UI for the less complex room.

(a) (b)

Fig. 5. (a) Conference room (higher complexity) and (b) 3-D UI used for physical object selection and manipulation in the conference room.

## D. Experimental Tasks

Each user was asked to perform a set of 11 structured tasks in a predefined order that was the same for all participants in each room. Each task involved manipulating a device by either turning it ON or OFF, or moving it up or down.

Each subject was randomly assigned to a UI-type (3-D, 2-D, physical). Participants assigned to the 3-D or 2-D UIs had to perform the tasks using the application, and the remaining participants used manual controllers. To record the exact object selection time, we eliminated the navigation and orientation times. Thus, for each task, before recording the time, we took the participant near the device or the controller of the device (eliminating the navigation time in the UI) and turned their body or feet to point toward the device (eliminating orientation time). Furthermore, prior to the experiment, the experimenter explained the scenario to each participant. Then, all the devices in the room and their positions were shown to the participant on a picture of the room to eliminate environment familiarization time (the time that the user would spend figuring out the environment and the available devices). Figs. 2 and 3 were used

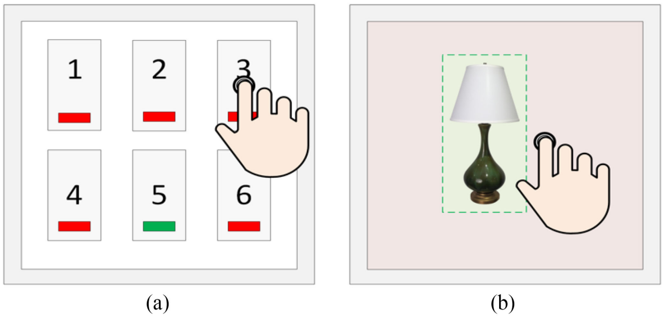


Fig. 6. Two kinds of selection errors. (a) Example for mis-map error in the 2-D UI. The user has wrongly selected button number 3 instead of button number 5. (b) Example for mis-tap error in 3-D UI. The user has clicked outside the lamp image.

to show to the participants where the interactable devices were located. Each participant looked at the picture for about 30 s.

Depending on the UI-type, participants were moved to the interaction zone of the related device and were asked to turn the device ON or OFF. The experimenter recorded the time taken to perform the task. For each task, the exact place and direction were marked to make the situation exactly the same for each participant. In each room, four of the 11 tasks related to the same device, in order to show how repeating the task and learning the act of mapping affected the object selection performance. In each room, we selected the operation of the ceiling lamps because they were similar in shape and had similar controllers which made the act of mapping complex in all three types of UI; in the physical UI, the switches were located near each other and looked the same; in the 2-D UI, the buttons were named with a number; for example, in the more complex room, the buttons were named lamps 1–4; and in 3-D UI, even though the 3-D models resembled the real devices, the similarity between the shapes was potentially misleading.

## E. Experimental Design

The experiment had a $2 \times 3$ between-subject design, meaning that each subject was assigned to only one of the six conditions. The independent variables were UI-type (2-D, 3-D, physical) and room complexity level (low, high).

The dependent variables for search performance included the average completion time of the tasks, and the mis-tap and mis-map errors. A mis-tap error is when the user knows the interaction button of the device to select an object but cannot use it correctly, and a mis-map error is when the user tries another button instead of the correct one. Fig. 6 shows examples of mis-taps for the 3-D UI and mis-maps for the 2-D UI. Additionally, if the user taps on the wrong 3-D model in the 3-D UI or chooses the wrong button in the physical UI, it is a mis-map error, and if the user chooses the right button to tap but has to repeat it due to misplacement of his/her finger, it is considered a mis-tap error.

We separated the mis-tap errors from the mis-map errors because mapping is an important part of the object selection; also, we wanted to obtain a more accurate understanding of whether using the UI will help users to perform the cognitive act of mapping.

TABLE II
APPROXIMATE DURATION OF EACH STEP OF THE PROCEDURE

| Procedure step | Duration (minutes) |
|---|---|
| Technology affinity and experience questions | 1 |
| Primary explanation about test | 0.7 |
| Subject familiarization with the devices | 0.7 |
| Instructions (app and tasks) | 1 |
| Pre-test necessary points | 1 |
| Test | 7 |
| Questionnaires (room complexity score, SUS, ASQ, net promoter, SEQ, SSQ, shape of hands) | 2 |

After the experiment, the participants were asked to complete the following questionnaires: room complexity score, which was a single question measuring the complexity level in subjects opinion; Simulator sickness questionnaire (SSQ) which measures the motion sickness that subjects may experience during exploring a 3-D environment (more popular in virtual reality environments); after scenario questionnaire (ASQ) and single ease question (SEQ) both of which determine how hard subjects feel the experiments tasks were; net promoter question [35] to determine user satisfaction; and system usability scale (SUS) questionnaire to determine whether user performance was affected by usability problems. We also drew the shape of the participants fingers to eliminate the fat finger problem if necessary [36].

## F. Participants

Of the 51 participants, who were aged 18–40 years old, 27 were male and the rest female. The participants included 43 students and 8 staff members. Their educational fields included computer engineering, economics, material, and chemical engineering. None of the participants had visited the rooms or used the applications before the experiment. Eight participants were assigned to each of the six conditions. Three were excluded due to their negative SSQ results, and we considered their data as invalid to eliminate the effect of sickness on user performance.

## G. Procedure

Each test lasted approximately 12–15 min (12 in the less complex room). Table II shows the steps of the experiment for each user and the duration of each step.

At the first step, the participants were welcomed and asked the following questions to determine their experience with touch cell phones and their experience in synthetic 3-D environments.

1) How often do you use touch cell phones or other touch devices?
2) How often do you play 3-D games or work with other 3-D applications?

Each participant rated the above on a five-point Likert scale, ranging from 1 (no experience) to 5 (frequent/daily use). The average scores for experience with touch devices was $M = 4$ ($SD = 0.73$) and with 3-D games was $M = 2$ ($SD = 0.76$).

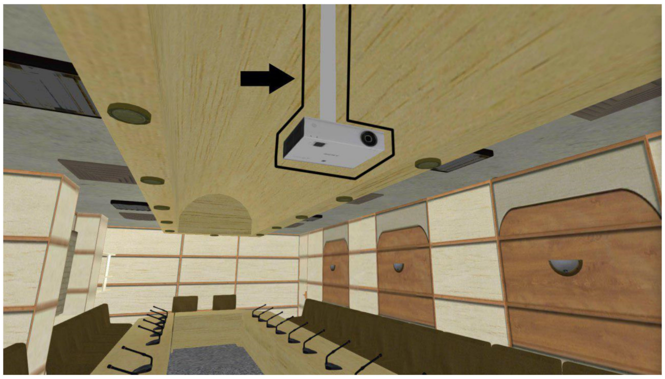

Fig. 7. Sample that the experimenter used to teach the participants how to use the 3-D UI application to select a projector.

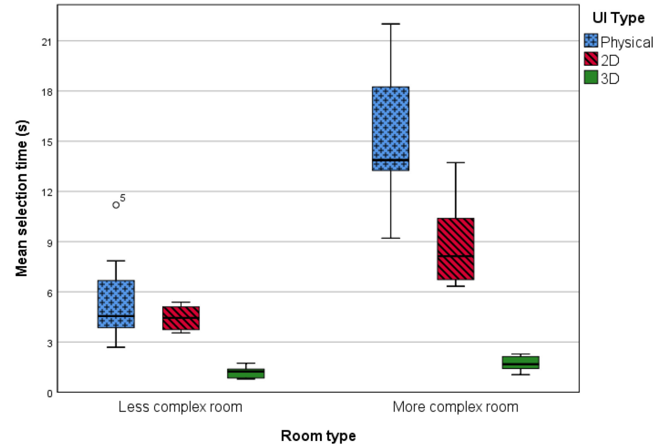


Fig. 8. Average object selection times for room complexity×UI type.

Next, without revealing the hypotheses, the participants were given a brief explanation about the test based on the UI with which they would work. Then, as explained before, to eliminate the orientation time, the experimenter showed the room and the devices to the participants using the pictures shown in Figs. 2 and 3 for the "less complex room" and "more complex room," respectively.

The participants were then instructed on how to use the application. To select each device, the 2-D UI group participants were advised to tap the button with a similar name as the device, while the 3-D UI group participants were advised to point at the similar 3-D model (cf. Fig. 7) to see the corresponding device reaction in the environment.

The experimenter explained some necessary points before beginning the test based on the outcomes of a small pilot test that we had conducted before the actual test. For instance, in the pilot test, we saw that users wanted to complete the task as quickly as possible, thus increasing their stress level. Therefore, before the actual test began, the experimenter told the participants that the test is designed to evaluate the overall system usability rather than selection performance, and that the tasks should be completed at the same speed as users would do in a real situation. In addition, to ensure we recorded accurate times, we asked the participants to start as soon as the experimenter finished the sentence that ordered the task. Once the main experiment started, participants were not allowed to request help about the UI mappings, i.e., which buttons or UI elements to use to control a specific device. This was because not only we were recording the task completion time, but also finding the right button was part of the task. Thus, we told the subjects that they could not interrupt the experiment to ask a question once the tests started.

At the beginning of the experiment, the first experimenter moved the participants to the determined positions and directed them toward the device. The first experimenter pointed to the desired device and asked the participant to turn the device ON or OFF according to the task. The second experimenter recorded the time using a stopwatch and wrote down the time, the mis-tap and the mis-map errors. This procedure was repeated for all 11 selection tasks.

## IV. Results

We present the results in the following four sections. Section IV-A presents the object selection performance, Section IV-B presents the effect of repeating a task on users performance, Section IV-C demonstrates the results related to the environment complexity, and Section IV-D presents other analyses of the data.

As mentioned before, for each selection task, we measured the selection time, the mis-map, and the mis-tap errors, while the participants performed the tasks. For hypothesis testing, we used two-factor ANOVA. The sample sizes in each group were equal $(n = 8)$.

### A. Object Selection Performance

A two-way ANOVA was performed to determine the influence of the two independent variables (room complexity and UI type) on the three dependent variables—average time to complete the tasks, average mis-tap errors, and average mis-map errors. Room complexity included two levels (less and more complex), and UI-type comprised three levels (physical, 2-D, and 3-D). Both effects were statically significant at the .05 significance level.

*1) Selection Time:* The F ratio of the main effect of room complexity ($F(1, 42) = 53.46$, $p \leq .001$, $\eta_p^2 = 0.56$) indicated a significant difference between the less complex room ($M =$ 3.70, SD $=$ 2.46) and the more complex room (M $=$ 8.59, SD $=$ 6.25). The main effect of UI-type yielded an F ratio of $F(2, 42) = 59.92, p < .001, \eta_p^2 = 0.74$, indicating a significant difference between the physical UI ($M = 10.37$, SD $=$ 6.05), the 2-D UI ($M = 6.62$, SD $=$ 2.94), and the 3-D UI ($M =$ $1.45$, SD $=$ 0.46). The interaction effect was also significant, $F(2, 42) = 15.93$, $p < .001$. Fig. 8 shows the mean selection times across the six conditions. Posthoc analysis using Tukeys HSD test showed that the average selection time for participants in the 3-D UI was significantly different from the two other groups in both environments. In the less complex room, the 2-D UI condition did not significantly differ from the physical UI condition. Taken together, the results show that the 3-D UI has a positive effect on user performance in object selection, but in

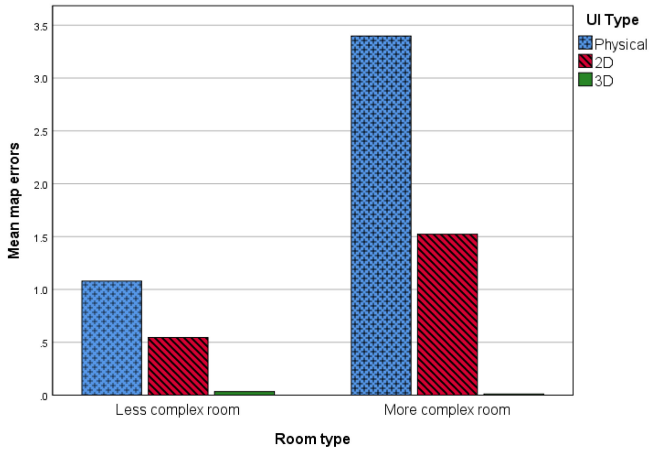


Fig. 9. Map error rate for room type (complexity) × UI type.

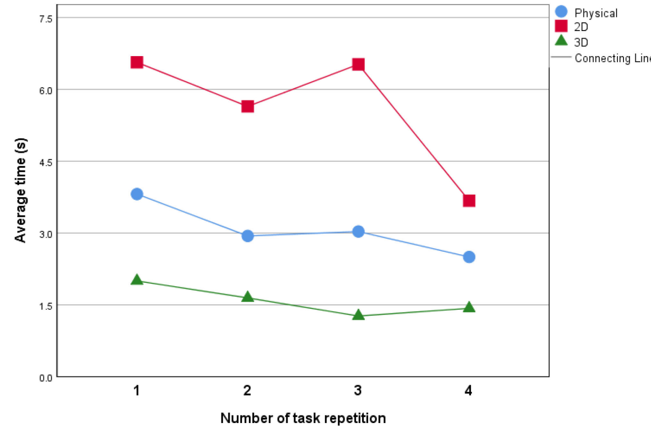


Fig. 10. Learning effect in the less complex room.

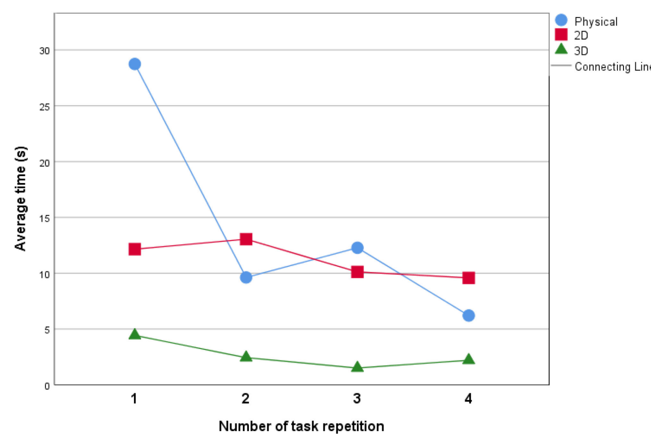


Fig. 11. Learning effect in the more complex room.

a less complex environment, using a 2-D UI compared to doing tasks manually might not increase user performance.

*2) Mis-Map Error:* For each task in the physical UI, the device or the controller had a corresponding button to manipulate the device. Similarly, the 2-D UI had a correspondingly named button, and the 3-D UI had a 3-D model for interacting with the device. Each mistake when selecting the wrong button counted as a mis-map error. Two-way ANOVA was performed to measure the influence of room complexity and UI-type on mis-map errors. A significant main effect was evident for both room complexity ($F(2, 42) = 60.43$, $p < .001$, $\eta_p^2 = 0.59$) and UI-type ($F(2, 42) = 83.32$, $p < .001$, $\eta_p^2 = 0.79$). The interaction effect was also significant ($F(2, 42) = 23.35$, $p < .001$). Fig. 9 shows the average map error rate across the six conditions.

*3) Mis-Tap Error:* We recorded mis-tap errors when the participants did select the right button or 3-D model but failed to click it properly. Each user made a few such errors. ANOVA results indicate that no significant effect of room complexity exists for mis-tap errors ($F = 0.0$, $p = 1.000$). Similarly, UI type ($F = 2.28$, $p = .114$) and the interaction of these factors ($F = 3.02$, $p = .059$) had no effect on mis-tap errors.

## B. Learning Tasks

To examine the learning effect on completion time, 4 of the 11 tasks involved manipulating the same device. We ran one-way repeated ANOVAs on each group. The dependent variable was the average completion time, and the independent variable was the number of times the task was repeated. The results showed that for the physical UI in both environments, the learning task had a significant effect on completion time ($F(3, 5) = 4.00$, $p = .021$ and $F(3, 5) = 17.40$, $p < .001$). Figs. 10 and 11 show the results. The time needed to complete a learning task manually (without the application) includes the time that participants take to detect the corresponding switch. Taking this result and the completion time in the physical UI together, we can see that in the physical UI, even though it takes longer to complete the tasks at the first few trials, the participants learn the mapping between switches and the lamps after repeating the tasks several times.

For the 2-D UI in both environments, no significant effects of learning were found on completion time ($F(3, 5) = 1.89$, $p = 0.162$, $F = (3, 5) = 0.41$, $p = .743$), indicating that users forget or do not attempt to memorize the corresponding button for each lamp when repeating the tasks.

We did not expect to find a significant effect of learning in the 3-D UI, since it is easy to find and tap the corresponding lamp on the scene. The results showed no significant effect of learning when using the 3-D UI in a less complex environment ($F = 1.92, p = .157$). However, when using the 3-D UI in a more complex environment, a significant effect of learning was found ($F = 4.31$, $p = .01$). A possible explanation for this is user confusion in a complex environment. When the users observe the 3-D scene in the application with many similar lamps, they may become confused during their initial attempts; however, after several attempts, they get used to the locations of the 3-D objects, and thus, find the desired lamp faster.

## C. Environment Complexity

According to the room complexity criteria discussed in Section III-B, we assumed that the two rooms had different complexity levels. To validate this assumption, we asked each user to rate how complex they thought the room was from 1 (not at all complex) to 7 (extremely complex). These ratings showed that

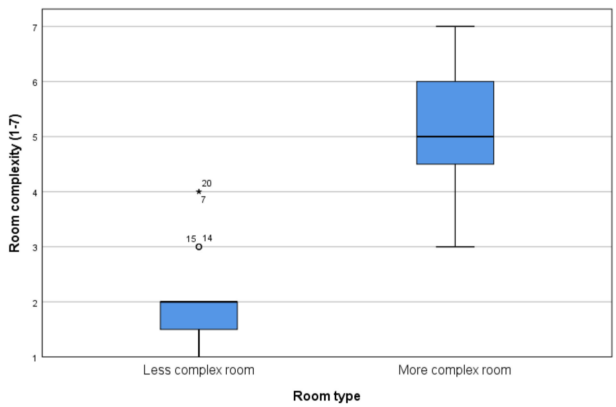


Fig. 12. Complexity scores for each of the environments.

users could recognize the complexity levels. An independent sample t-test on the complexity scores revealed that the level of complexity for the second room ($M = 5.13$, SD = 1.22) was significantly higher than that of the first room ($M = 2.04$, SD = .85); $t(46) = 10.086$, $p < .001$ (Fig. 12), which validates our assumption.

### D. Other Analysis

The SUS scores were 65.0–92.5 for the 2-D UI ($M = 82.50$, SD = 9.30) and 75–100 for the 3-D UI ($M = 89.37$, SD = 8.03), which shows that the usability of the applications did not affect user performance. The average score for the single question SEQ scores was 6.06 (SD = 1.1), which shows that the users had no problem understanding and completing the tasks.

For the 2-D UI, there were two promoters (score 9–10), ten passive users (score 7–8), and four detractors (score 1–6). Thus, the net promoter score (NPS) for the 2-D UI is -12. For the 3-D UI, there were 11 promoters and seven passive users which leads to an NPS of +68, considered as a high NPS score. As subjects were answering the net promoter question, we realized that recommending a physical UI to family and friends did not make sense for them as almost all of them asked for the meaning of this question. Since it was a between-subject experiment, they did not have other UIs in their mind to compare the physical UI with. Thus, we considered that their given score is not qualified enough to be reported.

## V. Discussion

We hypothesized that using a 3-D UI to select objects in smart environments would increase user performance. The results support our hypothesis. A feasible explanation is that the 3-D UI can simplify the complexity of working with devices, since users do not need to find a corresponding button on the device or learn how to set up the device. In addition, users can interact with the 3-D model of each device in 3-D scenes. Therefore, users do not need to map between the devices and controllers. Furthermore, our finding is in line with one of the interaction principles in Norman's book [37], which states that understandable mapping uses "the spatial correspondence between the layout of the controls and the devices being controlled." Simulated 3-D scenes hold to this principle since the devices are the controllers.

To measure the effect of UI-types in each room, we applied a one-way ANOVA. We expected that an application that has buttons to control the environment is better than physically manipulating the devices. The one-way ANOVA and the post hoc comparison in the more complex environment indicate that indeed the 2-D UI reduces the selection time significantly. However, in the less complex environment, even though the average task completion time in the 2-D UI is less than the physical environment, they were not significantly different. A possible explanation is that, in a less complex environment, the 2-D UI users must look for the corresponding UI button among a small number of simple elements, while physical UI users must look for the button on the simple devices with a limited number of similar devices which makes it easy to understand the mapping. Thus, in this environment, using a 2-D application does not significantly increase the mapping performance. In other words, when the user intends to select one of the similar devices in the smart environment (such as turning ON a light from a set of lights), the mapping time will not decrease; it just transfers from physical buttons to the 2-D UI elements, leading to performance similarity in both UIs. On the contrary, in the complex environment, our results show that using the 2-D UI is significantly better than using no UI because there are many similar devices with similar corresponding buttons, and it is difficult to learn many devices. These features in the complex environment can confuse physical UI users and lead to many mis-map errors. However, having named buttons presented in one application scene could reduce the learning time of the controllers, the mapping time, and the task completion time. In addition, we eliminated navigation time in our study; thus, we cannot confirm that 2-D is not a proper solution for complex smart environments, because using a mobile application notably decreases at least the navigation time.

The net promoter results show that users were more satisfied with the 3-D UI than the 2-D UI. A possible explanation for this result is that the users were interacting with the simulated scene of the real world through an application. Thus, the 3-D UI seems more attractive than the 2-D UI in which the user can only see and interact with some buttons. Additionally, the results showed that our 3-D UI reduced the object selection time with fewer errors and this increase in user performance correlated with user satisfaction. The complexity scores showed that the users could distinguish between the complexity of our environments. Thus, we used appropriate environments with different levels of complexity.

One shortcoming of our study is that the context of experiments is limited to a smart office. Also, due to our limited resources, we managed to conduct the experiments in only two different rooms. In addition, we compared three kinds of UIs to select objects. Future studies can address these challenges.

## VI. Conclusion

In this article, we conducted controlled experiments to examine the effect of environment complexity and the UI type on users' physical object selection performance. Through the experiment, we studied a 2-D and a 3-D user interface in two environments with different levels of complexity. We found that both the UI-type and the environment complexity have significant effects on users' selection tasks completion times. In addition, we found that a 3-D UI simulating the devices in the environment provides the best user performance in object selection, and users prefer to use this UI to control the room. Furthermore, we found that, in a less complex environment, a 2-D UI is not better than selecting a physical device with barehands.

In future work, it would be interesting to evaluate the effect of environment complexity on user performance when navigating the environment with different techniques.

## REFERENCES


[1] F. Argelaguet and C. Andujar, "A survey of 3D object selection techniques for virtual environments," *Comput. Graph.*, vol. 37, no. 3, pp. 121–136, 2013, May 2013, doi: 10.1016/j.cag.2012.12.003.

[2] L. Vanacken, T. Grossman, and K. Coninx, "Multimodal selection techniques for dense and occluded 3D virtual environments," *Int. J. Human-Comput. Stud.*, vol. 67, no. 3, pp. 237–255, 2009, doi: 10.1016/j.ijhcs.2008.09.001.

[3] D. A. Bowman, D. Koller, and L. F. Hodges, "Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques," in *Proc. IEEE Annu. Int. Symp. Virtual Reality*, 1997, pp. 45–52, doi: 10.1109/VRAIS.1997.583043.

[4] W. Delamare, C. Coutrix, and L. Nigay, "Designing disambiguation techniques for pointing in the physical world," in *Proc. 5th ACM SIGCHI Symp. Eng. Interactive Comput. Syst.*, 2013, pp. 197–206, doi: 10.1145/2494603.2480309.

[5] R. J. Teather and W. Stuerzlinger, "Pointing at 3D target projections with one-eyed and stereo cursors," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2013, pp. 159–168, doi:10.1145/2470654.2470677.

[6] A. A. de Freitas, M. Nebeling, X. Chen, J. Yang, A. S. K. Karthikeyan Ranithangam, and A. K. Dey, "Snap-to-it: A user-inspired platform for opportunistic device interactions," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2016, pp. 5909–5920, doi:10.1145/2858036.2858177.

[7] W. Stuerzlinger and C. A. Wingrave, "The value of constraints for 3D user interfaces," in *Virtual Realities*. Berlin, Germany: Springer, 2011, pp. 203–223, doi: 10.1007/978-3-211-99178-7_11.

[8] T. Chandler *et al.*, "Immersive analytics," in *Proc. Big Data Visual Analytics*, 2015, pp. 1–8.

[9] D. Schmidt, D. Molyneaux, and X. Cao, "Picontrol: Using a handheld projector for direct control of physical devices through visible light," in *Proc. 25th Annu. ACM Symp. User Interface Softw. Technol.*, 2012, pp. 379–388, doi: 10.1145/2380116.2380166.

[10] V. Mäkelä, T. Heimonen, and M. Turunen, "Magnetic cursor: Improving target selection in freehand pointing interfaces," in *Proc. Int. Symp. Pervasive Displays*, 2014, pp. 112–117, doi: 10.1145/2611009.2611025.

[11] A. A. N. Shirehjini and A. Semsar, "Human interaction with IOT-based smart environments," *Multimedia Tools Appl.*, vol. 76, no. 11, pp. 13 343–13 365, 2017, doi: 10.1007/s11042-016-3697-3.

[12] J. Dvorak, O. Krejcar, and L. K. Cheng, "Application of universal remote control of non-smart home appliances for smart home concepts," *Int. J. Digit. Enterprise Technol.*, vol. 1, no. 3, pp. 276–291, 2019, doi: 10.1504/IJDET.2019.097847.

[13] T. Kim, H. Lee, and Y. Chung, "Advanced universal remote controller for home automation and security," *IEEE Trans. Consum. Electron.*, vol. 56, no. 4, pp. 2537–2542, Nov. 2010.

[14] J. Rekimoto and K. Nagao, "The world through the computer: Computer augmented interaction with real world environments," in *Proc. User Interface Softw. Technol.*, 1995, pp. 29–36.

[15] A. A. N. Shirehjini, "A novel interaction metaphor for personal environment control: Direct manipulation of physical environment based on 3D visualization," *Comput. Graph.*, vol. 28, no. 5, pp. 667–675, 2004.

[16] R. Piyare, "Internet of things: ubiquitous home control and monitoring system using android based smart phone," *Int. J. Internet Things*, vol. 2, no. 1, pp. 5–11, 2013, doi: 10.5923/j.ijit.20130201.02.

[17] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon, "Design and evaluation of a smart home voice interface for the elderly: Acceptability and objection aspects," *Pers. Ubiquitous Comput.*, vol. 17, no. 1, pp. 127–144, 2013, doi: 10.1007/s00779-011-0470-5.

[18] B. Parameshachari, S. K. Gopy, G. Hurry, and T. T. Gopaul, "A study on smart home control system through speech," *Int. J. Comput. Appl.*, vol. 69, no. 19, 2013, doi: 10.5120/12080-8244.

[19] R. Kang, A. Guo, G. Laput, Y. Li, and X. Chen, "Minuet: Multimodal interaction with an Internet of Things," in *Proc. Symp. Spatial User Interact.*, 2019, Art. no. 2, doi: 10.1145/3357251.3357581.

[20] S. N. Patel and G. D. Abowd, "A 2-way laser-assisted selection scheme for handhelds in a physical environment," in *Proc. Int. Conf. Ubiquitous Comput.*, 2003, pp. 200–207, doi: 10.1007/978-3-540-39653-6_16.

[21] E. Velloso, M. Wirth, C. Weichel, A. Esteves, and H. Gellersen, "Ambigaze: Direct control of ambient devices by gaze," in *Proc. ACM Conf. Designing Interactive Syst.*, 2016, pp. 812–817, doi: 10.1145/2901790.2901867.

[22] E. Velloso, J. Turner, J. Alexander, A. Bulling, and H. Gellersen, "An empirical investigation of gaze selection in mid-air gestural 3D manipulation," in *Human-Computer Interaction*. Cham, Switzerland: Springer, 2015, pp. 315–330, doi: 10.1007/978-3-319-22668-2_25.

[23] O. Oda and S. Feiner, "3D referencing techniques for physical objects in shared augmented reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2012, pp. 207–215, doi: 10.1109/ISMAR.2012.6402558.

[24] M. R. Mine, "Virtual environment interaction techniques," Univ. North California Chapel Hill, Chapel Hill, NC, USA, Tech. Rep. TR95-018, 1995.

[25] L. Sidenmark and H. Gellersen, "Eye & head: Synergetic eye and head movement for gaze pointing and selection," in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol.*, 2019, pp. 1161–1174, doi: 10.1145/3332165.3347921.

[26] Y. Jang, I. Jeon, T.-K. Kim, and W. Woo, "Metaphoric hand gestures for orientation-aware VR object manipulation with an egocentric viewpoint," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 113–127, Feb. 2017.

[27] H. Lee, S.-T. Noh, and W. Woo, "TunnelSlice: Freehand subspace acquisition using an egocentric tunnel for wearable augmented reality," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 128–139, Feb. 2017.

[28] Y. Weiß, D. Hepperle, A. Sieß, and M. Wölfel, "What user interface to use for virtual reality? 2D, 3D or speech–a user study," in *Proc. Int. Conf. Cyberworlds*, 2018, pp. 50–57, doi: 10.1109/CW.2018.00021.

[29] M. Cummings, F. Sasangohar, K. Thornburg, J. Xing, and A. DAgostino, "Human-system interface complexity and opacity—Part I: Literature review," , Massachusetts Institute of Technology, Cambridge, MA, USA, 2010. [Online]. Available: https://www.researchgate.net/profile/Farzan_Sasangohar/publication /263315585_Human-system_interface_complexity_and_opacity/links/0deec53accc4413 fb2000000.pdf.

[30] A. W. Juliani, A. J. Bies, C. R. Boydston, R. P. Taylor, and M. E. Sereno, "Navigation performance in virtual environments varies with fractal dimension of landscape," *J. Environmental Psychol.*, vol. 47, pp. 155–165, 2016, doi: 10.1016/j.jenvp.2016.05.011.

[31] D. A. Bowman, D. Koller, and L. F. Hodges, "A methodology for the evaluation of travel techniques for immersive virtual environments," *Virtual Reality*, vol. 3, no. 2, pp. 120–131, 1998, doi: 10.1007/BF01417673.

[32] R. P. Darken and B. Peterson, "Spatial orientation, wayfinding, and representation." in *Handbook of Virtual Environment Technology*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2001, pp. 493–518, doi: 10.1080/13506280444000058.

[33] C. Allison and E. S. Redhead, "Factors influencing orientation within a nested virtual environment: External cues, active exploration and familiarity," *J. Environmental Psychol.*, vol. 51, pp. 158–167, 2017, doi: 10.1016/j.jenvp.2017.03.016.

[34] L. Vanacken, T. Grossman, and K. Coninx, "Exploring the effects of environment density and target visibility on object selection in 3D virtual environments," in *Proc. IEEE Symp. 3D User Interfaces*, 2007, pp. 117–124, doi: 10.1109/3DUI.2007.340783.

[35] E. Goodman, M. Kuniavsky, and A. Moed, *Observing the User Experience: A Practitioner's Guide to User Research*. New York, NY, USA: Elsevier, 2012.

[36] R. L. Potter, L. J. Weldon, and B. Shneiderman, "Improving the accuracy of touch screens: An experimental evaluation of three strategies," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 1988, pp. 27–32, doi: 10.1145/57167.57171.

[37] D. Norman, *The Design of Everyday Things*. New York, NY, USA: Basic Books, 1988.