



Article

Mid-Air Gesture Control of Multiple Home Devices in Spatial Augmented Reality Prototype

Panagiotis Vogiatzidakis and Panayiotis Koutsabasis *

Department of Product and System Design Engineering, University of the Aegean, 8410 Syros, Greece;
panosvog@syros.aegean.gr

* Correspondence: kgp@aegean.gr; Tel.: +30-693-219-9270

Received: 28 June 2020; Accepted: 28 August 2020; Published: 31 August 2020



Abstract: Touchless, mid-air gesture-based interactions with remote devices have been investigated as alternative or complementary to interactions based on remote controls and smartphones. Related studies focus on user elicitation of a gesture vocabulary for one or a few home devices and explore recommendations of respective gesture vocabularies without validating them by empirical testing with interactive prototypes. We have developed an interactive prototype based on spatial Augmented Reality (AR) of seven home devices. Each device responds to touchless gestures (identified from a previous elicitation study) via the MS Kinect sensor. Nineteen users participated in a two-phase test (with and without help provided by a virtual assistant) according to a scenario that required from each user to apply 41 gestural commands (19 unique). We report on main usability indicators: task success, task time, errors (false negative/positives), memorability, perceived usability, and user experience. The main conclusion is that mid-air interaction with multiple home devices is feasible, fairly easy to learn and apply, and enjoyable. The contributions of this paper are (a) validation of a previously elicited gesture set; (b) development of a spatial AR prototype for testing of mid-air gestures, and (c) extensive assessment of gestures and evidence in favor of mid-air interaction in smart environments.

Keywords: mid-air interaction; spatial augmented reality; home devices; smart environment; usability; user experience; elicitation study; MS Kinect

1. Introduction

Mid-air interaction is an established style of HCI (Human-Computer Interaction), in which users interact with distant displays and devices through body movements and gestures. The development of various body and gesture tracking technologies has led to the emergence of research and development of mid-air interaction in many domains such as remote manipulation of digital media on distant displays, interactive installations in public spaces, exergames, interactions with home devices, and others [1].

In principle, mid-air interaction presents many advantages: it is fast, accessory-free, ideal for “walk up and use” systems in public places or multiple surrounding systems or devices, it promotes hygiene since it does not require touching (a major requirement for increased research in operating rooms [2], and during the COVID-19 pandemic for any public space), it is also “magical” and engaging, etc. For these advantages to apply, mid-air interaction must be intuitive (easy to remember and apply, forgiving, etc.) and robust.

Empirical research in mid-air interaction often focuses on user elicitation of gestures, which results in a gesture set for the application at hand. The gesture elicitation methodology is well-documented [3,4], and has been applied to various domains of mid-air interaction [5]. Aspects of intuitiveness of mid-air gestures are predictively evaluated during the elicitation process, but this is tentative and must be validated in the practice of system use. Not much research on mid-air gesture elicitations has been

validated with interactive prototyping; despite this, a few studies have shown that user-defined gestures are not always the most usable [6]; interactive prototyping can also investigate the robustness of gestural interactions.

Most of the research in mid-air interaction deals with commands for a single application, display, or device. However, there are usage cases that would require simultaneous communication with multiple devices; for example, in the case of a smart home, car, or a technology-enhanced public space (e.g., industry or a classroom). In such scenarios, the user would have to address each one of the surrounding devices and instruct it with a gesture, while the gesture set should be consistent for similar operations among devices. Currently, there are very few studies of mid-air interaction with multiple devices [7–9] which have all progressed as far gesture elicitation and they have not been validated via interactive prototyping.

In previous work, we presented an in-depth elicitation study of upper-body mid-air gestures of a smart home device ecosystem [8]. In this paper, we present the continuation of that work, which follows a research-through-design approach [10]) and includes the (a) implementation of the mid-air gesture set (with Microsoft Kinect 2.0 SDK and Microsoft Visual Gesture Builder (<https://www.microsoft.com/en-us/download/details.aspx?id=44561>)) and a prototyping infrastructure of spatial augmented reality (S-AR) (built with Unity (<https://unity.com/>) and MadMapper (<https://madmapper.com/>)) and (b) the empirical evaluation of several aspects of the user experience (UX) of mid-air interactions in a two-phase laboratory test (with and without video-based help) with nineteen participants. The results of the empirical evaluation are very encouraging for the uptake of mid-air interaction with multiple devices in the home. The general approach followed, which included user elicitation, gesture refinement, and interactive prototyping, can be taken up in other investigations of mid-air interactions with multiple devices.

2. Related Work

Several studies have investigated gesture control for one home device, most usually the TV. In a review of 47 mid-air gesture elicitation studies [5], mid-air interaction with the smart home is investigated by 8.5% of papers. Older works in this respect present elicitation studies of mid-air gestures e.g., [11], as well as comparative studies to gestures on handhelds [12]. All these studies have not implemented the gestures, but they analyze several of their properties and make design recommendations. Some more recent studies combine elicitation with Wizard of Oz experiments, such as the work of Xuan et al., who compares mid-air gestures to remote control, collected data on perceived usability and experience, and concluded that “gesture control puts more mental stress and cognitive load on users, but it could improve the overall experience, compared to remote control” [13]. One major constraint of these types of studies is that they have not progressed to a more realistic context of the use of gestural interaction with interactive prototyping of gesture control accompanied by digital content, device indicators, etc. Furthermore, these studies are concerned with one home device only.

Mid-air interaction with a smart home device ecosystem (a set of devices) has been investigated in a few previous studies.

Choi et al. (2014) [7] have conducted a repeated elicitation study for 7 devices and 20 referents to investigate whether the top gestures proposed by participants are consistently repeated in a second study. They conclude that 65% of the top gestures selected in the first experiment were changed in the second, which indicates that there is variability in top gesture proposals even if the same users are involved in subsequent experiments.

Hoffman et al. [9] present an elicitation study to compare user preference between voice, touch, and mid-air gestures for smart home control. They define a set of five devices and eleven referents, considerably smaller than other studies. They conclude that voice commands or a touch display are clearly preferred compared to the use of mid-air gestures. They also do not proceed to interactive prototyping.

Vogiatzidakis and Koutsabasis [8] investigate a user-defined gesture vocabulary for basic control of a smart home device ecosystem consisting of 7 devices and a total of 55 referents (commands for device) that can be grouped to 14 commands (that refer to more than one device). The elicitation study was conducted in a frame (general scenario) of the use of all devices to support contextual relevance; also, the referents were presented with minimal affordances to minimize widget-specific proposals. The study explored mid-air gesture vocabulary for a smart-device ecosystem, which includes several gestures with very high, high, and medium agreement rates. This gesture set has been adopted (with a few modifications), implemented, and evaluated in the current paper.

Several other elicitation studies of gesture-based interactions assume a handheld device; for example, a smartphone [14], a Wiimote [15] or other custom handhelds such as the “smart pointer” [16] which emits visible light to select a particular device (similar to a small flashlight) and invisible infrared (IR) light to operate the device with gestures. Our work differs from these works since it is concerned with mid-air gestures with no handheld devices or other accessories.

A major shortcoming of aforementioned studies, which are otherwise fairly extensive and thorough, is that they have not validated the produced gesture sets through interactive prototyping. On the other hand, a few interactive prototypes of mid-air interaction with multiple home devices have been developed. However, these are somehow limited in scope and not informed by previous user research. For example, in [17] the design of a gesture-based prototype for context-sensitive interaction with smart homes is presented for 7 commands and 4 devices; their work contains a small number of commands and devices and does not involve gesture elicitation, but prototype development with designer-defined interaction techniques and usability testing. Ng et al. [18] also present a prototype for home automation for 5 devices and 2 commands only (switch on/off).

Our work contributes to current state of the art by validating a previously elicited gesture set for smart home control (7 devices, 41 referents) [8]; implementing a robust interactive prototype of spatial augmented reality which recognizes all gestures and translates them to system responses; assessing the usability, user experience, and learning of gestures in a scenario of mid-air gesture control among devices; and finally, by providing evidence that mid-air gesture control of multiple home devices is feasible, engaging, and fairly good in performance.

3. Study Design

3.1. Apparatus: A Spatial Augmented Reality Interactive Prototype

We develop the concepts and mid-air interactions in a spatial augmented reality (AR) prototype, which is based on projection mappings of digital content, device features, and feedback. Spatial AR has been successfully employed in scenarios of design evaluation of (interactive or not) systems in the past such as control panels and car dashboards [19], as well as in games and installations that support mid-air interactions (e.g., RoomAlive [20] and Room2Room [21]). In the case of our research, we adapt and develop this technology for interactive prototyping in a research through design approach [10], by which we develop a fairly robust interactive artifact and implement mid-air gestures to empirically evaluate user interaction. Spatial AR technology is suitable for prototyping because it affords directed projections and highlighting of digital controls, indicators and animated content onto (initially white-blank, foil) 3D surfaces while allowing the user to concentrate on required mid-air interactions (rather than individual features of devices).

3.1.1. Hardware and Setup

The home environment created is illustrated in Figure 1. The user is seated or standing in front of a Microsoft Kinect v2.0 (<https://developer.microsoft.com/en-us/windows/kinect/>), connected to a PC, which tracks the user's gestures which are applied with hands and the rest of the upper human body parts. These gestures trigger audio and visual feedback on the PC (Intel i7, 16 GB memory), and is displayed through a ceiling-mounted projector on the wall in 2D or is projection-mapped onto

dummy devices created by foil (Figure 2a). These devices are: (1) Air Conditioner, (2) Blinds, (3) Lights, (4) Amplifier with speakers, (5) Audio player, (6) Media-video player, and (7) TV. During the training session, an iPad is used by the researcher to select the appropriate video animation which is also projected on the wall (Figure 2b).

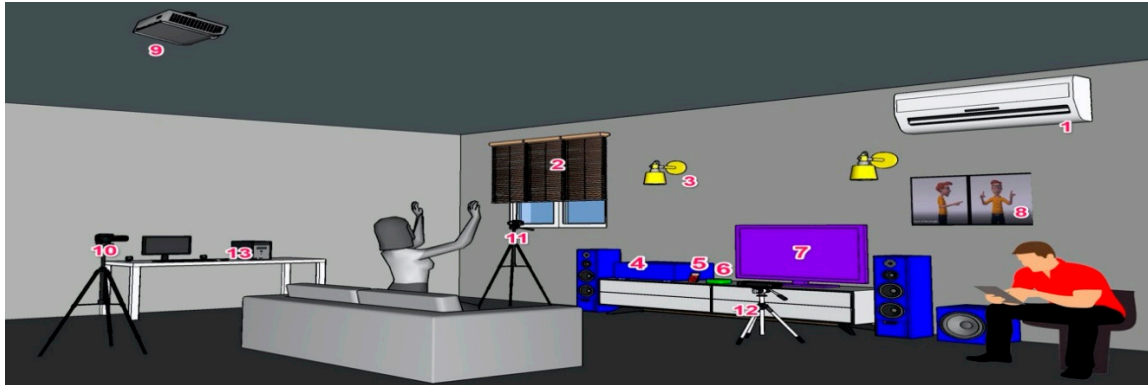


Figure 1. General setup of the experiment that includes: seven devices that the user can control (1–7), a frame displaying help videos on the wall (8), two cameras (10 and 11) for video capturing the experiment, a PC (13) connected to the Kinect sensor (12) and to a projector (9), the researcher that controls the help-videos through an iPad, and the user.

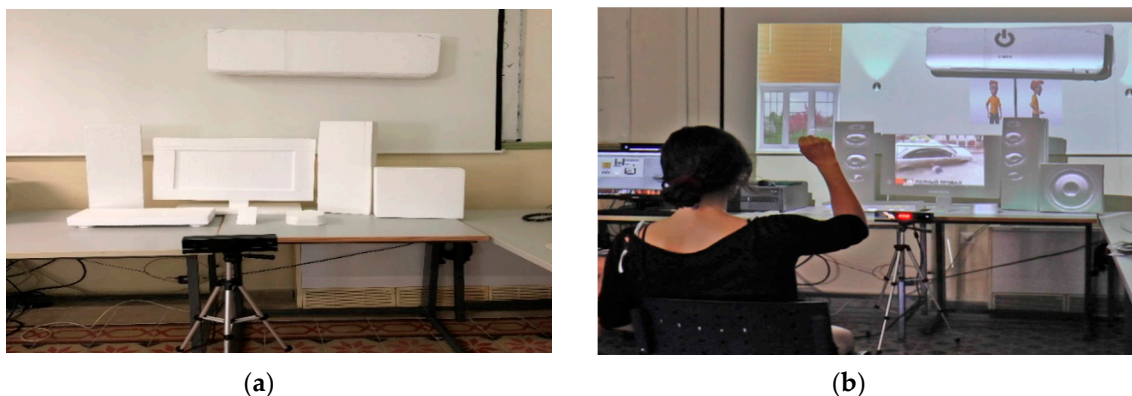


Figure 2. (a) Dummy devices created by foil; (b) Projection of 2D content on the wall (for blinds, lights and help-videos), and on the dummy devices with projection mapping.

3.1.2. Software

Several interconnected technologies and software were used to implement the prototype (Figure 3). To track the user's gestures, we have used Microsoft Kinect v2.0 sensor (<https://developer.microsoft.com/en-us/windows/kinect/>), connected to a PC (Intel i7, 16 GB memory) running Windows 10 and Microsoft Kinect SDK v2.0 (<https://www.microsoft.com/en-us/download/details.aspx?id=44561>). The prototype was developed in Unity (<https://unity.com/>) and C# including the seven aforementioned devices with the appropriate audio or visual feedback (animations) which is triggered by the user's commands. To connect Unity with the Kinect SDK we have used the plugin "Kinect v2 with MS-SDK (<https://assetstore.unity.com/packages/3d/characters/kinect-v2-examples-with-ms-sdk-and-nuitrack-sdk-18708>)".

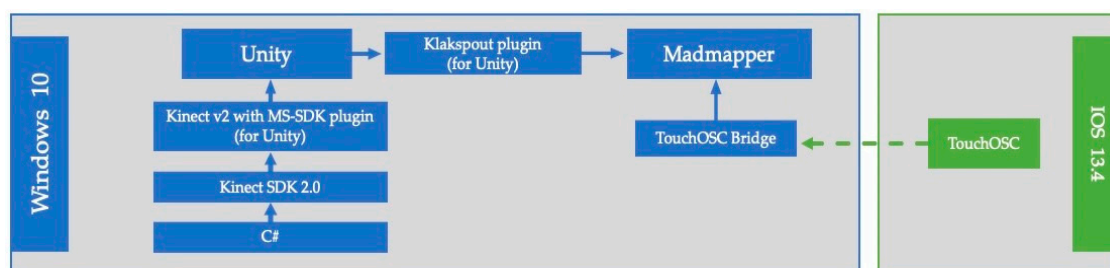


Figure 3. Software used in the study's prototype.

To assist the users, learn or recall the required gesture, we provided during the training session, a video animation of a 3D character performing the corresponding gesture from different viewing angles. These animations were created in Blender and merged together into a single video (for each gesture) with Final Cut Pro (<https://www.apple.com/final-cut-pro/>).

Madmapper (<https://madmapper.com/>) was used to project the animations from Unity, onto the wall and the dummy devices through the Klakspout (<https://github.com/keijiro/KlakSpout>) plug-in. Madmapper was also used to select the space and location on the wall, where the help-videos (with the 3d character performing the gesture) would be displayed. The researcher was able to maintain control wirelessly, which helps video to be displayed through use of an application called TouchOSC (<https://hexler.net/products/touchosc>) running on an iPad pro (with IOS 13.4). The TouchOSC application was communicating with Madmapper, through “TouchOSC-Bridge” software that was running on the PC.

The whole experiment was video recorded from 2 cameras; one capturing the user's gestures and the other capturing the devices. The video footage was then synced and merged side by side into one video clip for each user, before using it for analysis.

3.2. Gestures

The previous work of Vogiatzidakis and Koutsabasis [8] was used as a ground to form the gesture set that was implemented in the prototype. In that study, the proposed gesture set is based on the score of the Agreement Rate metric $AR(r)$ [22] after an elicitation study with 18 users (we refer to it as GS_{el} as it is the result of the user elicitation study (Table A1)).

3.2.1. Registration and Command Gestures

In GS_{el} , there are two types of gestures: *Registration-Gestures* and *Command-Gestures*.

A Registration-Gesture is unique for a device and is used to activate the tracking capabilities of this device (which becomes active). These gestures are iconic or deictic and denote a visual feature or an effect of the home device. For example, by drawing a rectangle in mid-air, the user registers the TV, and by holding out his/her arms he/she registers the air conditioner, and so on (Tables 1 and 2).

A Command Gesture is the same on many different devices when the operation is the same or similar. Consequently, for each device, there is only one registration gesture (which is unique), and many Command-Gestures (that can also be used in other devices). Most command gestures comprise typical gestures employed for manipulations like swipe up/down/right/left, hand (fist) open/close, etc. (Table 1 and Figure 4), although there are a few iconic or deictic gestures as well, e.g., the Shhh gesture.

Table 1. Consistent Gesture Set (GS_C) as it is formed through the steps of “Simplification for Consistency” routine.

Commands	Gestures from Step 1	Gestures from Step 2	Gestures from Step 3	Gestures from Step 4	Simplified Gesture Set (GS _C)
R ¹ . TV	Form a rectangle				Form a rectangle
R ¹ . Speakers	Form a circle				Form a circle
R ¹ . Video Player	Binoculars				Binoculars
R ¹ . Audio Player	Point on ear				Point on ear
R ¹ Air Conditioner	Hands holding arms				Hands holding arms
R ¹ . Lights	Clap				Clap
R ¹ . Blinds	Point (left hand)				Point (left hand)
Turn On		Closing fist		Opening fist	Opening fist
Turn Off		Swipe up			Closing fist
Up		Swipe down			Swipe up
Down				Swipe right	Swipe down
Next				Swipe left	Swipe right
Previous		Sssh			Swipe left
Mute					Sssh
Fast Forward				Move hand clockwise	Move hand clockwise
Fast Rewind				Move hand counterclockwise	Move hand counterclockwise
Play				Point (right hand)	Point (right hand)
Pause		Palm stop			Palm stop
Stop			Hands split		Hands split

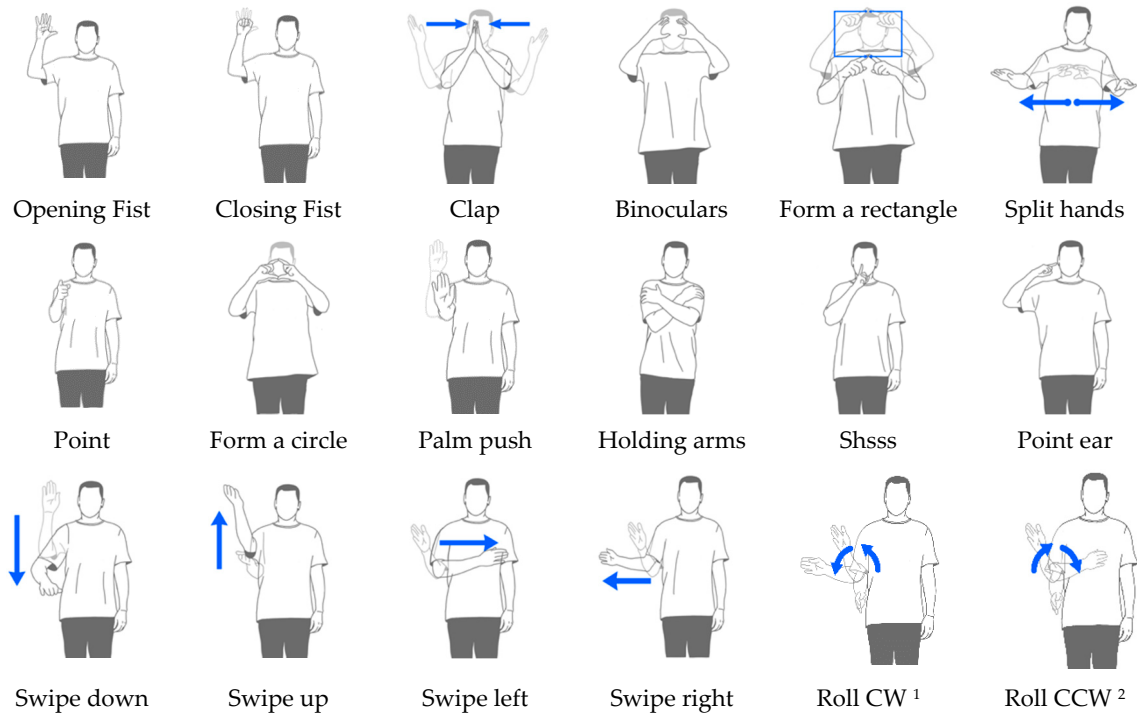
¹ R means registration gesture.**Figure 4.** Gesture Set used in the prototype, as it was formed after the simplification refinement. ¹ Clockwise, ² Counterclockwise.

Table 2. Scenario used for the evaluation of the prototype and the corresponding gestures needed for registering and commanding devices.

No.	Appliance	Task	Registration Gestures	Command Gestures
1	Lights	Turn on the lights	Clap	Opening Fist
2	Lights	Dim the lights down		Swipe down
3	TV	Turn on the TV	Form a rectangle	Opening Fist
4	TV	Switch to the next TV channel		Swipe right
5	TV	Mute the volume of the TV		Shsss
6	TV	Un-mute the volume of the TV		Shsss
7	Blinds	Open the window blinds	Point (left hand)	Swipe up
8	Blinds	Stop the blinds moving up		Split hands
9	TV	Switch to previous TV channel	Form a rectangle	Swipe left
10	Lights	Dim the light up	Clap	Swipe up
11	TV	Turn the TV volume down	Form a rectangle	Swipe down
12	TV	Turn the TV volume up		Swipe up
13	TV	Turn off the TV		Closing Fist
14	Movie Player	Turn on the movie player	Binoculars	Opening Fist
15	Movie Player	Play the movie		Point (right hand)
16	Movie Player	Fast forward the movie		Roll Clockwise
17	Movie Player	Pause the scene of the movie		Palm push
18	Movie Player	Play the movie		Point (left hand)
19	Movie Player	Fast rewind the movie		Roll Counterclockwise
20	Movie Player	Stop the movie		Split hands
21	Movie Player	Turn off the movie player		Closing Fist
22	Air Conditioner	Turn on the Air Conditioner	Holding arms	Opening Fist
23	Air Conditioner	Decrease the temperature		Swipe down
24	Speakers	Turn on the speakers	Form a circle	Opening Fist
25	Audio Player	Turn on the audio player	Point ear	Opening Fist
26	Audio Player	Play the song		Point (right hand)
27	Audio Player	Go to the next song		Swipe right
28	Audio Player	Fast forward the song		Roll Clockwise
29	Audio Player	Fast rewind the song		Roll Counterclockwise
30	Speakers	Mute the speakers	Form a circle	Shsss
31	Air Conditioner	Increase Air Conditioner Temperature	Holding arms	Swipe up
32	Air Conditioner	Turn off the air-condition		Closing Fist
33	Audio Player	Turn off the audio player	Point ear	Closing Fist
34	Speakers	Turn off the speakers	Form a circle	Closing Fist
35	Lights	Turn off the lights	Clap	Closing Fist
36	Blinds	Close the window blinds	Point (left hand)	Swipe down

The approach is that whenever the user wants to activate a device, he/she would have to perform the corresponding registration gesture first. Then, when the device is active, he/she can perform any of the available command-gestures on it. Even though some command-gestures are the same for many devices, only the active device can respond (i.e., each device stops tracking gestures after a while).

3.2.2. Simplification of Gesture Set for Consistency among Devices

Although the initial gesture set (GS_{el}) was a result of a user-centered design approach, we further analyzed it to reach further simplifications in order to improve consistency between devices and to minimize the number of different gestures that the user would have to remember. We simplified gestures following the steps below:

1. First of all, the gestures for the devices' registrations were left intact since each device must have a unique gesture that has to be performed before commanding it (when it is not activate).
2. For the commands that had only one gesture proposal, that gesture was selected and matched to the command. These commands were turn off, up (volume up, increase temperature, dim up the lights), down (volume down, decrease temperature, dim down the lights), mute, and pause.
3. The registration gestures (from step 1) and the matched gestures (from step 2), were then removed from the remaining commands. If a command is left with only one gesture, that gesture is then matched to the command.

4. For the remaining commands that had more than one proposed gesture, the gesture with the higher Level of Agreement was chosen as a match for the corresponding command and removed from the other commands.

Following the above routine, we ended up with a simplified and consistent gesture set as shown in Table 1. The new consistent gesture set, which we refer to as GS_C , is a simplified version of the one produced after the Elicitation Study (GS_{el}), since it contains in total 19 gestures, 5 less than GS_{el} .

3.2.3. Refinement Due to Technological Constraints

After the GS_C was formed, we implemented and tested these gestures in Microsoft Gesture Builder (VGB). There were few cases where we had to do further refinements, due to technological constraints:

1. In two cases there was an issue when two gestures intervened with each other. As a result, there was a conflict in gesture recognition since the Kinect sensor could not distinguish each of them with acceptable confidence. Therefore, more than one gesture was recognized at the same time. In our gesture set there was a conflict with “clap” and “form a rectangle” (both gestures had similar ending), as well as with “point” and “swipe left/right” (the beginning of the swipe left/right was similar with the pointing pose). In these cases, we had to decide which gestures to refine with minor changes, implement and test them again until we got distinct and acceptable recognitions. For example, we changed the “clap” gesture to be at the level of the shoulders, instead of the level of the spine base. Similarly, we changed the “point” gesture to have a thumb up.
2. We have implemented the gestures in Visual Gesture Builder (VGB), which supports discrete/static gestures that use a binary classifier that determines either if the gesture is performed or not and dynamic gestures that track the progress of the gesture over time. Continuous gestures (gestures that trigger a command as long as they are performed), such as “Roll CW or CCW” cannot be easily implemented in VGB since they do not have an end. Therefore, we decided to simplify them into dynamic gestures with the hand starting from the level of the spine base, performing a semicircle movement, and ending on either left or right side of the body as shown in Figure 4.

3.3. Procedure

At the beginning of the evaluation process, introductory information was given to the users about the concept of a smart home environment and mid-air interaction control.

The users interacted with the smart home environment in a scenario, which was read to them by the researcher, and includes commands that they would have to perform on the 7 devices of our prototype (TV, Audio player, Video Player, Amplifier-Speaker, Air Conditioner, Blinds, and Lights). That scenario was read twice to all users. The first time (*training session*), users had to perform the desired gesture after this gesture was presented to them on the wall with a video animation of a 3D virtual character, while during the second time (evaluation session) the same scenario was used, but without help provided to them. Between the training and evaluation sessions, there was a *recap session*, which was used to check whether the users could remember the gestures. At the end of the evaluation session users were asked to fill in some questionnaires. The evaluation took 50 min on average for each user.

3.4. Scenario

During the evaluation process, a scenario that includes all the desired commands was used. The concept of the scenario is that the user returns to his/her home and while spending some time in the living room he/she operates some home appliances that are placed in front of him/her. The researcher reads the commands from the scenario and waits for the user to complete the task. If the task could not be completed, the researcher moves to the next command. The commands are listed in such an order so that the user uses the devices in an interchangeable way and not one-off. Table 2 lists the tasks of

the scenario, the Registration Gesture that is needed to activate the device (when it is inactive) and the Command Gesture that will trigger the command on the active device. Commands no. 5 and no. 15 are repeated twice in the scenario. However, in the data analysis we track results only from the first time they are performed.

3.5. Participants

Participants were recruited from academic research staff via an email invitation. Participation was voluntary and no reward was offered. All of them agreed to the goals of the study and the treatment of their personal data according to the General Data Protection Regulation (EU GDPR). Nineteen participants took part in the study, 5 females and 14 males whose age ranged from 26 to 49 ($M_{\text{age}} = 40$, $SD_{\text{age}} = 6.8$). They were personnel and postgraduate students of the university with a background in engineering and design. Almost half of them (9 out of 19) had some previous experience with mid-air interaction (mainly with Wii or Kinect). Four of the participants were left-handed.

3.6. Metrics and Data Collection

We collected performance-based metrics of *task-success*, *task completion-time*, and *errors* (*false positives* and *false negatives*) for both sessions (training, evaluation). These were supplemented with observational data, collected during the test as well as during post-test video analysis. In between the two sessions (training, evaluation) we conducted a gesture memorability test to all users.

Furthermore, we collected user-reported data with two standardized questionnaires: (1) SUS (System Usability Scale [23]), which is often used in usability evaluations, and consists of ten 5-point Likert statements about usability and computes a usability score within [0, 100] (The SUS usability score is very satisfactory when above 80, it is fairly satisfactory when between 60 and 80, and not satisfactory when below 60) and (2) UEQ (User Experience Questionnaire [24]), which consists of 26 pairs of terms with opposite meanings that the user can rate in a 7-point Likert scale; these terms reflect attractiveness, classical usability aspects (efficiency, perspicuity, dependability) and user experience (originality, stimulation).

During the test, at first, we collected demographic information from the users. Video footage was captured by two cameras, one facing the user and the other facing the projected devices. At the end of the experiment a post-study semi-structured interview was used to gain insight from the users, and the aforementioned questionnaires were provided to users to fill out.

4. Results

Nineteen users were asked to apply 36 gestures to the scenario, twice, which resulted in a total of 1368 gestures requested for all users. Measures of task success, task time, and errors were identified in approximately 40 h of videos, after the end of the experiments.

4.1. Task Success

We measured task success as a binary value (1 or 0) for all tasks. We considered the task unsuccessful after repeated failed attempts to perform a gesture (we set a high but reasonable time of 20 s per gesture to stop users from further attempts) or when users themselves gave up. The results for task success are shown in Table 3.

Device registration gestures were totally successful (100%) for all seven devices in both training and evaluation sessions. Of course, this is an excellent result for all users, who were able to address all devices successfully with mid-air gestures in order to start further interactions.

Table 3. Task success during training and evaluation session.

Tasks	Training Session	Evaluation Session
TV Registration	100%	100%
TV Turn on	100%	100%
TV Turn off	100%	100%
TV Next channel	95%	89%
TV Previous channel	100%	100%
TV Volume up	95%	79%
TV Volume down	95%	100%
TV Mute	100%	100%
Blinds Registration	100%	100%
Blinds up	100%	89%
Blinds Down	100%	100%
Blinds Stop	63%	68%
Lights Registration	100%	100%
Lights Turn on	100%	100%
Lights Turn off	95%	95%
Lights Dim Up	74%	84%
Lights Dim Down	100%	95%
Air Conditioner Registration	100%	100%
Air Conditioner Turn on	100%	100%
Air Conditioner Turn off	95%	100%
Air Conditioner increase temperature	79%	89%
Air Conditioner decrease temperature	100%	84%
Audio Player Registration	100%	100%
Audio Player Turn on	100%	100%
Audio Player Turn off	89%	89%
Audio Player Next	89%	84%
Audio Player Fast Forward	100%	100%
Audio Player Fast Rewind	100%	100%
Audio Player Play	100%	100%
Video Player Registration	100%	100%
Video Player Turn on	100%	100%
Video Player Turn off	100%	89%
Video Player Fast Forward	100%	100%
Video Player Fast Rewind	100%	100%
Video Player Play	100%	100%
Video Player Stop	84%	84%
Video Player Pause	89%	95%
Speakers Registration	100%	100%
Speakers Turn on	100%	100%
Speakers Turn off	95%	89%
Speakers Mute	89%	95%
Overall task success	96%	95%

With respect to the command gestures (and respective tasks), we saw very satisfactory results for most gestures (tasks). As shown in Table 3, many tasks were performed with absolute success (100%); such as the turn-on command (open hand gesture), which was applied five times in the scenario (for TV, lights, air conditioner, video player and speakers), twice (with/without help) by all nineteen users.

For some tasks, we saw some failures in task success. In most of those cases, users still performed very well overall, such as in next/previous (TV, Video player) and volume up/down (TV, audio, and video players) tasks. The less successful task was “Blinds Stop” with 63% (training) and 68% (evaluation). Interestingly, although the same gesture was used in both “Blinds Stop” and “Video Player Stop”, the latter task had a much higher success score of 84% in both sessions. A possible explanation is that “blind stop” was more frustrating for the user than “video player stop” since the blinds were moving up and therefore he/she had to be fast enough to command them before reaching the highest level.

In Table 4 we show the average gesture success score (i.e., aggregating the results of Table 3 for similar tasks among devices). We can see that in the training session was 96% and in the evaluation session was 95%. Although ideal task success must be 100%, and despite a few gestures for which there is certainly room for improvement, we consider this a fairly satisfactory and encouraging result for mid-air interaction with multiple devices.

Table 4. Gesture success during the training and the evaluation session. Between the two sessions the higher score is highlighted.

Gestures	Training Session (Help Enabled)	Evaluation Session (without Help)
Registration gestures ¹	100%	100%
Opening Fist (Turn On)	100%	100%
Closing Fist (Turn Off)	96%	94%
Swipe up (Up)	87%	85%
Swipe down (Down)	99%	95%
Swipe right (Next)	92%	87%
Swipe left (Previous)	100%	100%
Shsss (Mute)	95%	97%
Roll Clockwise (Fast Forward)	100%	100%
Roll Counterclockwise (Fast Rewind)	100%	100%
Point with right hand (Play)	100%	100%
Palm push (Pause)	89%	95%
Split hands (Stop)	74%	76%

¹ Registration gestures for all 7 devices of the prototype.

4.2. Task Time

Task (completion) time was measured in seconds for all successful tasks. Task time was calculated during both training and evaluation sessions from the users' video footage. It corresponds to the time duration needed by the user to start performing the gesture (including seeing help and/or thinking time) until the time the systems responded to it with digital (visual or audio) content or feedback. The results of task time are illustrated in Figure 5 (with 95% confidence intervals).

The pattern followed by users during task performance can be summarized as follows: (a) as soon as the user was instructed to perform the next task (by the researcher), (b) they viewed the gesture on video (training session, with video-enabled help (All videos lasted for approximately about one-half to one second, i.e., the time)), or they recalled the gesture (evaluation session, without help), (c) they applied the gesture, (d) they awaited system response, (e) in the case of no success they repeated steps (c) and (d) for a few attempts until the researcher requested for them to stop.

As shown in Figure 5, task completion time for mid-air gestures (i.e., all stages above) were performed within a few seconds on average for all gestures. For some tasks/gestures, average user performance is very satisfactory, like for example the tasks of video or audio player "play" (less than 3" in the evaluation session (without help)). For other tasks/gestures, average user performance is fairly satisfactory, such as the task TV registration (6.5" in the evaluation session (without help)).

When viewing the average task time of the two test sessions (training/evaluation) comparatively, we can see that for most tasks (27/41) task time has improved; additionally, in twelve of those tasks, there is significant improvement according to 95% confidence intervals. Overall, users were slightly faster in the evaluation session since the average total time to complete the tasks of the scenario was 187.9 s (SD = 3.0) for the training session and 173 s (SD = 2.6) for the evaluation session.

These are very encouraging results for mid-air interaction with multiple devices: it is feasible and proves to be a very or fairly fast mode of interaction, in addition to its other advantages (accessory-free, etc.).

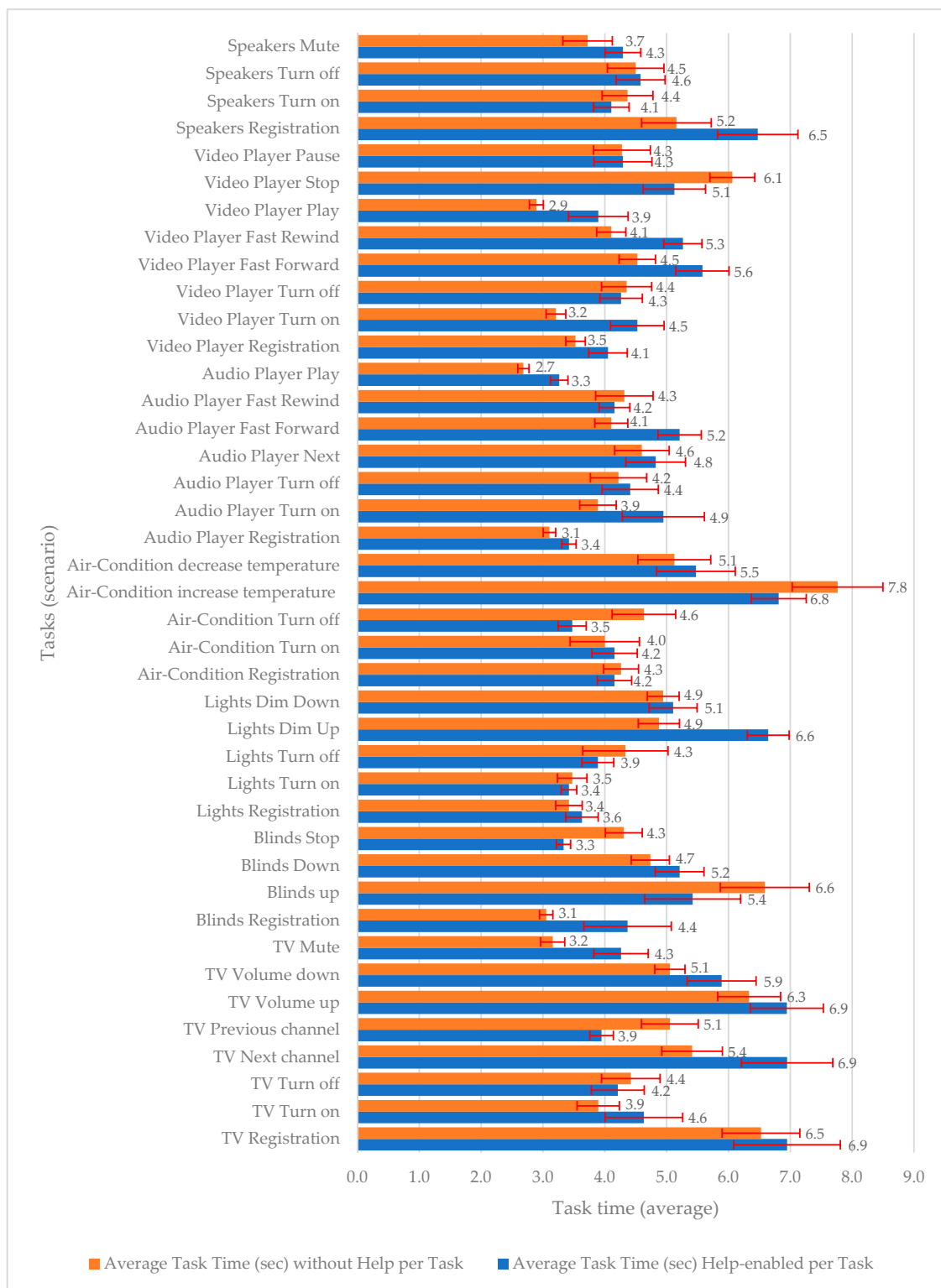


Figure 5. Task completion time.

4.3. Errors

We distinguish errors into false-negatives and false-positives. False negative errors occurred when the user applies the right gesture (even when gesture application was not rigorous) without system response. False positive errors occurred when the system responded accidentally or falsely.

4.3.1. False Negatives

Table 5 summarizes findings on false negative errors by showing average and median values per task for both test situations (with/without help). Median values are mostly zero, which means that most users did not present any false negative errors per task. Average false negative values were generally low, ranging in [0.2, 2.8] for the training session and in [0.2, 4.8] for the evaluation sessions. On the other hand, a few users had some difficulties with a few gestures (those with higher average false negatives).

Table 5. False negative errors (average, median) for learning and evaluation session of mid-air interaction with multiple devices.

False Negatives	Learning Session		Evaluation Session	
	Average	Median	Average	Median
TV Registration	1.7	0	3.7	0
TV Turn on	0.8	0	0.9	0
TV Turn off	0.4	0	1.1	0
TV Next channel	2.0	1	2.8	1
TV Previous channel	0.4	0	1.5	0
TV Volume up	1.4	1	3.4	2
TV Volume down	1.3	0	0.8	0
TV Mute	0.4	0	0.5	0
Blinds Registration	0.6	0	0.1	0
Blinds up	0.8	0	3.9	1
Blinds Down	1.3	0	1.2	0
Blinds Stop	1.8	0	2.6	0
Lights Registration	0.3	0	0.7	0
Lights Turn on	0.5	0	0.3	0
Lights Turn off	0.5	0	1.7	0
Lights Dim Up	2.6	2	3.0	1
Lights Dim Down	0.6	0	2.0	0
Air Conditioner Registration	0.3	0	0.6	0
Air Conditioner Turn on	0.6	0	0.8	0
Air Conditioner Turn off	0.3	0	1.3	0
Air Conditioner increase temperature	2.8	1	4.9	2
Air Conditioner decrease temperature	0.7	0	3.0	0
Audio Player Registration	0.2	0	0.1	0
Audio Player Turn on	0.8	0	1.5	0
Audio Player Turn off	1.0	0	1.7	0
Audio Player Next	1.7	0	3.8	0
Audio Player Fast Forward	0.8	0	0.8	0
Audio Player Fast Rewind	0.3	0	0.8	0
Audio Player Play	0.2	0	0.0	0
Video Player Registration	0.3	0	0.4	0
Video Player Turn on	0.6	0	0.3	0
Video Player Turn off	0.6	0	2.5	0
Video Player Fast Forward	0.8	0	0.8	0
Video Player Fast Rewind	0.6	0	0.8	0
Video Player Play	0.8	0	0.1	0
Video Player Stop	2.1	0	3.3	1
Video Player Pause	1.5	0	1.8	0

Table 5. Cont.

False Negatives	Learning Session		Evaluation Session	
	Average	Median	Average	Median
Speakers Registration	1.4	0	1.9	0
Speakers Turn on	0.4	0	1.3	0
Speakers Turn off	0.5	0	2.4	0
Speakers Mute	1.3	0	1.8	0

False negatives occurred typically when users did not apply the right gesture accurately. In those users we observed the following pattern: (a) the user views (in training) or remembers the gesture and begins to apply it; (b) the user applies the gesture loosely; for example (s)he didn't raise or extend their hand fully when required; (c) the system did not respond, (d) the user (immediately) understood that the application of the gesture was not correct, (e) the user re-applied the gesture rigorously, (d) the system responded. Apparently, this pattern occurred in a short time-lapse of a few seconds. In the evaluation session, the false negatives were increased. We noticed that this was mainly due to the fact that the virtual character helped the user to exercise the gesture more accurately, and not so much because it reminded them of the right gesture.

Overall, the results on false negatives reconfirm the very satisfactory performance of users in mid-air interaction with multiple home devices.

4.3.2. False Positives

Table 6 summarizes findings on false positive errors by showing average and median values per task for both test situations (with/without help). Median values are zero for all tasks, which means that most users did not present any false negative errors for any task. Average false positive values were generally low, and lower than false negatives, ranging in [0.0, 0.4] for the training session and in [0.0, 0.2] for the evaluation session.

Table 6. False positive errors (average, median) for learning and evaluation session of mid-air interaction with multiple devices.

False Positives	Learning Session		Evaluation Session	
	Average	Median	Average	Median
TV Registration	0.1	0	0.0	0
TV Turn on	0.4	0	0.0	0
TV Turn off	0.0	0	0.0	0
TV Next channel	0.1	0	0.0	0
TV Previous channel	0.2	0	0.1	0
TV Volume up	0.0	0	0.0	0
TV Volume down	0.0	0	0.0	0
TV Mute	0.0	0	0.1	0
Blinds Registration	0.0	0	0.0	0
Blinds up	0.0	0	0.2	0
Blinds Down	0.1	0	0.0	0
Blinds Stop	0.1	0	0.1	0
Lights Registration	0.1	0	0.0	0
Lights Turn on	0.1	0	0.1	0
Lights Turn off	0.1	0	0.0	0
Lights Dim Up	0.4	0	0.0	0
Lights Dim Down	0.0	0	0.0	0

Table 6. Cont.

False Positives	Learning Session		Evaluation Session	
	Average	Median	Average	Median
Air Conditioner Registration	0.1	0	0.1	0
Air Conditioner Turn on	0.1	0	0.0	0
Air Conditioner Turn off	0.0	0	0.0	0
Air Conditioner increase temperature	0.0	0	0.0	0
Air Conditioner decrease temperature	0.0	0	0.0	0
Audio Player Registration	0.0	0	0.0	0
Audio Player Turn on	0.1	0	0.1	0
Audio Player Turn off	0.2	0	0.0	0
Audio Player Next	0.1	0	0.1	0
Audio Player Fast Forward	0.3	0	0.1	0
Audio Player Fast Rewind	0.0	0	0.1	0
Audio Player Play	0.0	0	0.0	0
Video Player Registration	0.0	0	0.1	0
Video Player Turn on	0.0	0	0.0	0
Video Player Turn off	0.1	0	0.0	0
Video Player Fast Forward	0.2	0	0.2	0
Video Player Fast Rewind	0.3	0	0.0	0
Video Player Play	0.0	0	0.0	0
Video Player Stop	0.4	0	0.2	0
Video Player Pause	0.0	0	0.0	0
Speakers Registration	0.1	0	0.0	0
Speakers Turn on	0.1	0	0.0	0
Speakers Turn off	0.1	0	0.0	0
Speakers Mute	0.2	0	0.2	0

False positives occurred typically when users accidentally activated the system. This typically occurred in gestures that can be applied in two dimensions (i.e., have an opposite), such as swipe up/down, open/close hand. In those cases, the system recognized the opposite gesture and not the intended one, e.g., when the user was lowering his hand after an unsuccessful (false negative) swipe-up gesture.

False positives were few, which is also a very encouraging result. Overall, the results on false positives reconfirm the very satisfactory performance of users in mid-air interaction with multiple home devices.

4.4. Memorability

In between the training and evaluation sessions, users were asked to perform the 19 different gestures. This exercise was made in order to check whether users could remember the correct gestures and remind them the ones that could not recall. Memorability ratio was calculated based on the users' correct answers (Figure 6). Almost half of the gestures (10 out of 19, 52%) were 100% memorable, while the remaining gestures' memorability ration ranged from 79% to 95%. Users tended to confuse gestures that correspond to conceptually similar commands (such as "pause" and "stop") or devices (such as speakers—audio player), which can also be seen from their low memorability ratios. Generally, the memorability for all gestures was 94%, which is very high given that there were 19 gestures and users had only one session to learn and memorize them.

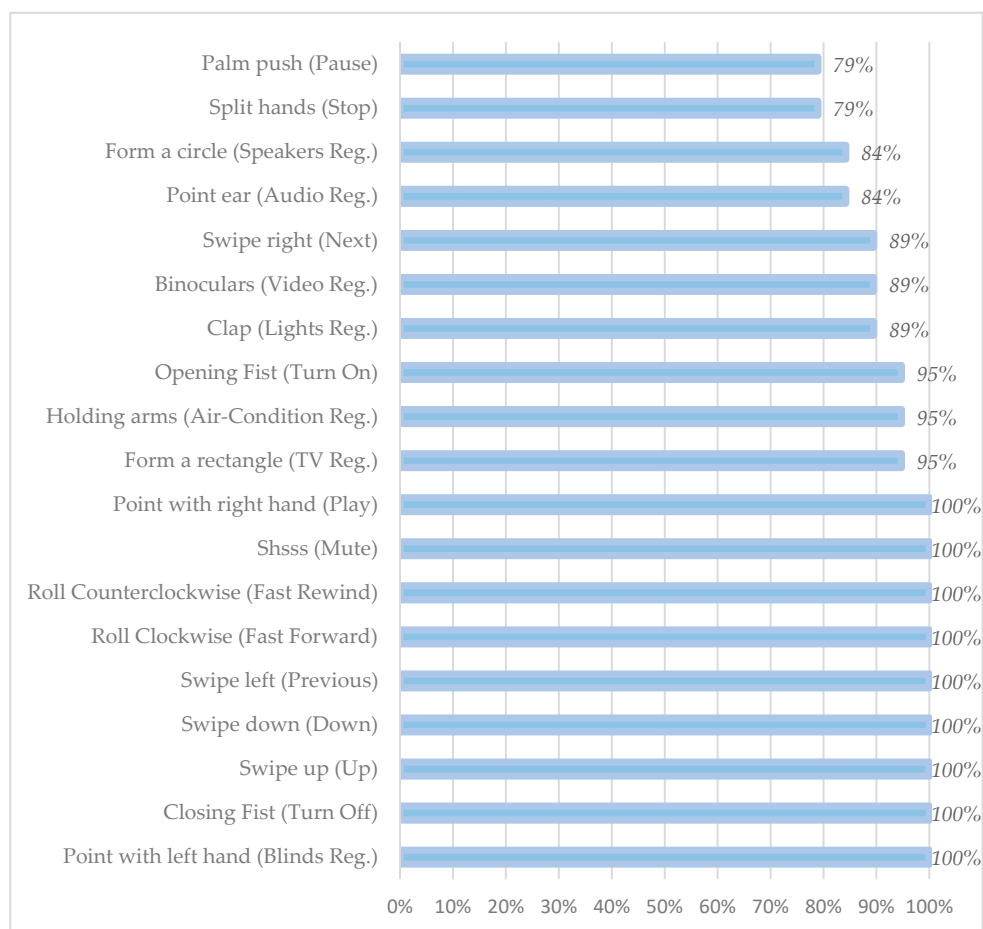


Figure 6. Memorability ratio.

4.5. Perceived Usability and User Experience

4.5.1. SUS

To assess users' perceived usability of the system we asked users to fill-in the System Usability Scale (SUS) questionnaire [23]. We calculated the SUS average score, which was 79.0. According to Tullis and Albert "an average SUS score under about 60 is relatively poor, while one over about 80 could be considered pretty good" [25]. Thus, this is another indication that the system usability was very satisfactory.

4.5.2. User Experience Questionnaire (UEQ)

To assess the user experience of the system we asked users to fill in the User experience Questionnaire (UEQ) [24]. The range of the scales is between -3 (horribly bad) and $+3$ (extremely good). According to Schrepp, "in real applications ... over a range of different persons with different opinions and answer tendencies it is extremely unlikely to observe values above $+2$ or below -2 ... the standard interpretation of the scale means is that values between -0.8 and 0.8 represent a neutral evaluation of the corresponding scale, values >0.8 represent a positive evaluation and values <-0.8 represent a negative evaluation" [26].

User responses to the UEQ were very positive and are depicted in Figure 7. All aspects of the system presented a very positive experience: attractiveness (pure valence dimension of the UX), aspects of pragmatic (goal-directed) quality (perspicuity, efficiency, and dependability) and aspects of hedonic (non-instrumental) quality (stimulation and novelty).

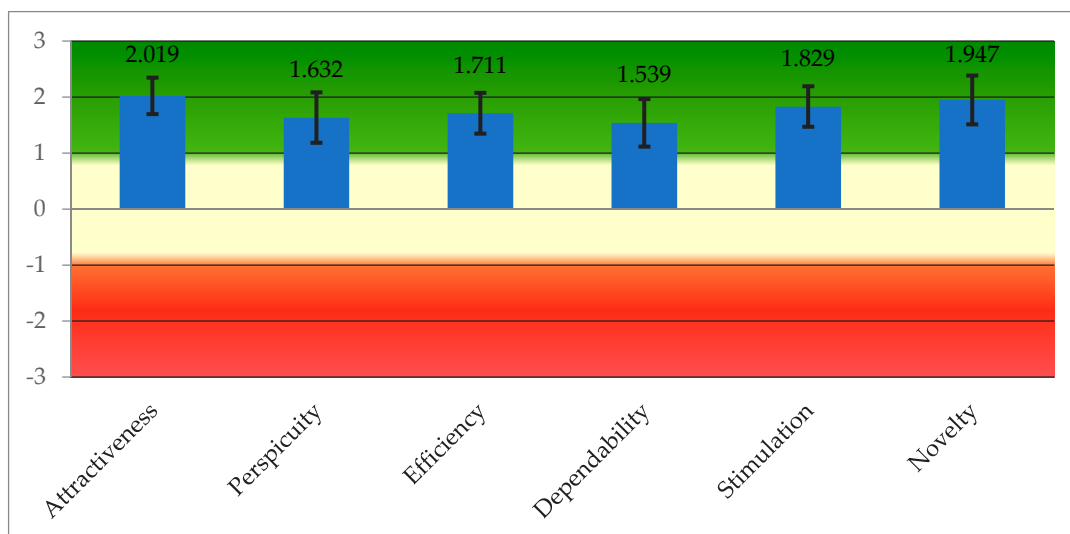


Figure 7. Average user responses to the UEQ (User Experience Questionnaire).

5. Discussion

5.1. Mid-Air Interaction with Multiple Home Devices Is Feasible and Fairly Satisfactory in Terms of Usability and User Experience

The results of the empirical evaluation provide evidence that mid-air interaction with multiple home devices is feasible, fairly easy to learn and apply, and enjoyable. All aspects of this empirical evaluation provide positive evidence: task success is high; task time is fairly satisfactory; errors are low; there is high memorability of gestures and fair learnability of the system (when performance is compared between learning and evaluation sessions), and last but not least, perceived usability scores are high (according to the SUS), as well as all factors of the user experience (UEQ questionnaire).

Regarding task time, this included (a) thinking time (or viewing video-enabled help), (b) gesture application and (c) system response. Presumably, in the case of remote control usage of multiple devices, users would presumably have to search and locate the appropriate remote controls, or reach to them and then locate buttons. Or, in the alternate case of mobile device control, the user would have to locate and reach the phone and then navigate through the UI. However, we have not tested these situations, which can be a dimension of further research.

Admittedly, positive evidence was not expected in all aforementioned dimensions since mid-air interactions often present usability issues such as the Midas' touch problem (accidental system activations), gesture distinctiveness, robustness, memorability, appropriateness, and others. These issues were not strongly present in this study, while it was evident that another prototyping cycle could further smoothen particular issues identified for a few users.

5.2. On a More Comprehensive Method that Moves from Defining Gestures to Testing Mid-Air Interactions

We have followed a research through design approach to investigate aspects of the UX of mid-air interaction with multiple devices, which included the following steps:

1. Elicitation study, which resulted in the first gesture set.
2. Designer refinement of gestures to:
 - a. *Maximize consistency among devices*, by viewing the data of elicitation and selecting a single gesture for each common operation (among devices) based on agreement scores.
 - b. *Prevent possible conflicts and unsatisfactory sensor tracking*, by considering simple rules of thumb for better sensor tracking. For the case of MS Kinect SDK, it is important to ensure

that each gesture includes a few distinctive points that correspond to skeletal joints or predefined hand gestures (open/closed hand, lasso). Furthermore, when using the Visual Gesture Builder, it is important to prepare a considerable number of videos of exercising gestures to model gestures accurately, with more than one user body types.

3. Implementation of the interactive prototype that is functional and reasonably realistic. The approach taken on spatial AR presents several advantages since it supports projections and highlighting of digital controls, indicators, and animated content onto 3D objects and surfaces (initially white-blank, made of foil).
4. Empirical evaluation of mid-air gestures in an extensive test which reported on several dimensions of performance and preference about usability and the user experience.

The aforementioned steps present a design research and prototyping cycle, which could be repeated—especially in steps 3 and 4—to further improve selected tasks and gestures. This approach could be adapted to other contexts of mid-air interactions with multiple devices, beyond the smart home, like for example mid-air interactions in vehicles, technology-enhanced public spaces, etc.

5.3. The “Device Registration Approach” as a Forcing Function to Avoid the Midas’ Touch Problem of Mid-Air Interaction with Multiple Devices

We have followed an approach of “device registration” as an option for controlling mid-air interactions with multiple devices. This approach requires from the user firstly (a) to address a device with a particular (unique) gesture for this device in order to make it active, and secondly (b) to perform mid-air gestures to interact with it. In HCI terminology, the requirement to address a device before the interaction is a forcing function, i.e., a designer-imposed, user behavior-shaping constraint that prevents undesirable user input made by mistake [27].

This approach actually minimized the Midas’ touch problem of accidental (device) activation and kept false positives at very low numbers. However, it may have presented a burden on some users who kept forgetting to address a device before starting to interact with it. We plan to further research this issue by introducing bimanual gestures that simultaneously provide registration and operation of a device.

5.4. Limitations of the Study

We refer to the limitations of the study in terms of its ecological validity [28], which is concerned with many factors like the context of use, the participants, the method, apparatus and prototype.

Given that we followed an experimental procedure, we set up the system in an academic laboratory and then invited participants to make use of it. This step, as part of an iterative design and prototyping process, can further include (in subsequent studies) a living lab experiment or a field study setup (i.e., in a home, with real devices). However, it is an advancement to the current state of the art, given that previous studies have not validated elicitation results with interactive prototyping.

Mid-air gesture control of remote devices concerns all people. For this study, we recruited participants from academic and research staff. Due to the voluntary character of the study, we observed that male participants were more numerous than women. Also, four of the participants were left-handed. Further studies should broaden the participant sampling.

The apparatus and prototype were built on spatial AR technology which afforded for interactivity, plausibility, and clarity. This is a major advantage of this research compared to other studies that remain at analyzing gesture elicitation results. Mid-air prototyping with multiple devices is difficult to prototype, and we have found that spatial AR can be a suitable technology for that.

6. Summary and Conclusions

This paper presented an assessment of mid-air interactions with multiple home devices on the basis of (a) a previous elicitation study which identified the gestures, (b) implementation of gestures

(MS Kinect SDK 2.0 and Visual Gesture Builder) and development of a spatial AR prototype (projection mapping interface via MadMapper and Unity) that allowed users to interact with digital content and device mock-ups, and (c) empirical evaluation of mid-air interactions with multiple devices on main usability indicators: task success, task time, errors (false negative/positives), memorability, perceived usability, and user experience.

The principal conclusion is that mid-air interaction with multiple home devices is feasible, fairly easy to learn, and apply, and enjoyable. The contributions of our work to the current state of the art are:

- The work presented in this paper validates a previously elicited gesture set for smart home control (7 devices, 41 referents) [8].
- We have implemented a robust interactive prototype of spatial augmented reality which recognizes all gestures and translates them to system responses.
- We assess the usability, user experience, and memorability of gestures in a scenario of mid-air gesture control among devices.
- We provide evidence that mid-air gesture control of multiple home devices is feasible, engaging, and fairly good in performance.

We followed a research-through-design approach as described in [10] which illustrates this as a model which emphasizes design and development iterations of “artifacts as vehicles for embodying what “ought to be” and that influence both the research and practice communities” [10]. The first iteration of our research (“artifact” in terms of [10]) is the elicitation study which resulted in the production and analysis of a set of gestures for mid-air interactions with multiple devices; here, the main research goal was to identify a consistent gesture set. The second iteration (or artifact) was the interactive software prototype of these gestures, which is implemented with spatial AR technology; here, the main research goal was to validate the usability and UX of the previously identified gesture set via interactive prototyping. Further work can develop other interactions (artifacts), such as the investigation of alternate gesture sets (i.e., a set without distinct gestures for registration with a device), the conduction of a living lab experiment [29], or a field study (that would require integration with real devices).

Author Contributions: Conceptualization, P.V.; Data curation, P.V.; Formal analysis, P.V. and P.K.; Investigation, P.V. and P.K.; Methodology, P.V. and P.K.; Project administration, P.K.; Resources, P.V.; Software, P.V.; Supervision, P.K.; Writing—original draft, P.V. and P.K.; Writing—review & editing, P.V. and P.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Appendix A

Table A1. List of gestures with the corresponding Agreement Rate (AR(r)) for each command which were proposed in the research of Vogiatzidakis and Koutsabasis [8]. We refer to this gesture set as GS_{el} . The last column shows the number of different gestures for each command. If for a command more than one gesture is proposed, the one with the higher AR(r) is kept unless it is the only option on another command. All registration gestures are kept since they are unique.

	TV	Speakers	Video Player	Audio Player	Air Conditioner	Lights	Blinds	Different Gestures
Registration On	Form a rectangle (0.11)	Form a circle (0.03)	Binoculars (0.09)	Point on ear (0.04)	Hands holding arms (0.07)	Clap (0.05)	Point (0.04)	7
Registration Off	Form a rectangle (0.05)	Form a Circle (0.02)	Binoculars (0.06)	Point on ear (0.02)	Hands holding arms (0.06)	Clap (0.03)	Point (0.01)	7
Turn On ²	Point (0.04)	Opening fist (0.05)	Opening fist (0.01)	Palm push (0.05)	Opening fist (0.07)	Point up (0.05)		4
Turn Off ¹	Closing fist (0.05)	Closing fist (0.08)	Closing fist (0.04)	Closing fist (0.03)	Closing fist (0.07)	Closing fist (0.03)		1
Up ¹	Swipe up (0.52)	Swipe up (0.46)			Swipe up (0.28)	Swipe up (0.28)	Swipe up (0.33)	1
Down ¹	Swipe down (0.52)	Swipe down (0.46)			Swipe down (0.33)	Swipe down (0.32)	Swipe down (0.33)	1
Next ²	Swipe right (0.34)		Point right (0.14)	Swipe right (0.16)				2
Previous ²	Swipe left (0.34)		Point left (0.14)	Swipe left (0.16)				2
Mute ¹	Ssshhh (0.12)	Ssshhh (0.10)						1
Fast Forward ³			Swipe right (0.17)	Move hand clockwise (0.09)				2
Fast Rewind ³			Swipe left (0.17)	Move hand counterclockwise (0.09)				2
Play ²			Palm up (0.03)	Point (0.08)				2
Pause ¹			Palm stop (0.09)	Palm stop (0.15)				1
Stop ³			Palm stop (0.07)	Hands split (0.14)			Palm stop (0.52)	2

¹ only one gesture is proposed, ² the gesture with the highest AR(r) is chosen, ³ the gesture with the higher AR(r) is not chosen because it was the only option or with the highest AR(r) in another command, Green, represents gestures form the Consistent Gesture Set (GS_C).

References

1. Koutsabasis, P.; Vogiatzidakis, P. Empirical research in mid-air interaction: A systematic review. *Int. J. Hum. Comput. Interact.* **2019**. [\[CrossRef\]](#)
2. Hötter, A.M.; Pitton, M.B.; Mildenerberger, P.; Düber, C. Speech and motion control for interventional radiology: Requirements and feasibility. *Int. J. Comput. Assist. Radiol. Surg.* **2013**, *8*, 997–1002. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Nielsen, M.; Störing, M.; Moeslund, T.B.; Granum, E. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 409–420. ISBN 978-3-540-21072-6.
4. Wobbrock, J.O.; Morris, M.R.; Wilson, A.D. User-defined gestures for surface computing. In Proceedings of the 27th International Conference on Human factors in Computing Systems—CHI 09, Boston, MA, USA, 4 April 2009; ACM Press: New York, NY, USA, 2009; p. 1083.
5. Vogiatzidakis, P.; Koutsabasis, P. Gesture elicitation studies for mid-air interaction: A review. *Multimodal Technol. Interact.* **2018**, *2*, 65. [\[CrossRef\]](#)
6. Koutsabasis, P.; Domouzis, C.K. *Mid-Air Browsing and Selection in Image Collections*; ACM Press: New York, NY, USA, 2016; pp. 21–27.
7. Choi, E.; Kwon, S.; Lee, D.; Lee, H.; Chung, M.K. Towards successful user interaction with systems: Focusing on user-derived gestures for smart home systems. *Appl. Ergon.* **2014**, *45*, 1196–1207. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Vogiatzidakis, P.; Koutsabasis, P. Frame-based elicitation of mid-air gestures for a smart home device ecosystem. *Informatics* **2019**, *6*, 23. [\[CrossRef\]](#)
9. Hoffmann, F.; Tyroller, M.-I.; Wende, F.; Henze, N. User-defined interaction for smart homes: Voice, touch, or mid-air gestures? In Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia—MUM '19, Pisa, Italy, 27–29 November 2019; pp. 1–7.
10. Zimmerman, J.; Forlizzi, J.; Evenson, S. *Research through Design as a Method for Interaction Design Research in HCI*; ACM Press: New York, NY, USA, 2007; p. 493.
11. Chen, M.; Mummert, L.; Pillai, P.; Hauptmann, A.; Sukthankar, R. Controlling your TV with gestures. In Proceedings of the International Conference on Multimedia Information Retrieval, Philadelphia, PA, USA, 29–31 March 2010; pp. 405–408.
12. Radu-Daniel, V. A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. *J. Ambient Intell. Smart Environ.* **2013**. [\[CrossRef\]](#)
13. Xuan, L.; Daisong, G.; Moli, Z.; Jingya, Z.; Xingtong, L.; Siqi, L. Comparison on user experience of mid-air gesture interaction and traditional remotes control. In Proceedings of the Seventh International Symposium of Chinese CHI, Chinese CHI '19, Xiamen, China, 27–30 June 2019; pp. 16–22.
14. Kühnel, C.; Westermann, T.; Hemmert, F.; Kratz, S.; Müller, A.; Möller, S. I'm home: Defining and evaluating a gesture set for smart-home control. *Int. J. Hum.-Comput. Stud.* **2011**, *69*, 693–704. [\[CrossRef\]](#)
15. Vataavu, R.-D. There's a world outside your TV: Exploring interactions beyond the physical TV screen. In Proceedings of the 11th European Conference on Interactive TV and Video, Como, Italy, 24–26 June 2013; pp. 143–152.
16. Vorwerg, S.; Eicher, C.; Ruser, H.; Piela, F.; Obée, F.; Kaltenbach, A.; Mechold, L. Requirements for gesture-controlled remote operation to facilitate human-technology interaction in the living environment of elderly people. In *Human Aspects of IT for the Aged Population. Design for the Elderly and Technology Acceptance*; Lecture Notes in Computer Science; Zhou, J., Salvendy, G., Eds.; Springer International Publishing: Cham, Switzerland, 2019; Volume 11592, pp. 551–569. ISBN 978-3-030-22011-2.
17. Neßelrath, R.; Lu, C.; Schulz, C.H.; Frey, J.; Alexandersson, J. A gesture based system for context—Sensitive interaction with smart homes. In *Ambient Assisted Living*; Wichert, R., Eberhardt, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 209–219. ISBN 978-3-642-18166-5.
18. Ng, W.L.; Ng, C.K.; Noordin, N.K.; Ali, B.M. Gesture based automating household appliances. In Proceedings of the 14th International Conference on Human-Computer Interaction, Orlando, FL, USA, 9 July 2011.
19. Port, S.R.; Marner, M.R.; Smith, R.T.; Zucco, J.E.; Thomas, B.H. Validating spatial augmented reality for interactive rapid prototyping. In Proceedings of the 2010 IEEE International Symposium on Mixed and Augmented Reality, Seoul, Korea, 13 October 2010; IEEE: Seoul, Korea, 2010; pp. 265–266.

20. Jones, B.; Shapira, L.; Sodhi, R.; Murdock, M.; Mehra, R.; Benko, H.; Wilson, A.; Ofek, E.; MacIntyre, B.; Raghuvanshi, N. RoomAlive: Magical experiences enabled by scalable, adaptive projector-camera units. In Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology—UIST '14, Honolulu, HI, USA, 5–8 October 2014; pp. 637–644.
21. Pejsa, T.; Kantor, J.; Benko, H.; Ofek, E.; Wilson, A.D. Room2Room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing—CSCW '16*; ACM Press: San Francisco, CA, USA, 2016; pp. 1714–1723.
22. Vatavu, R.-D.; Wobbrock, J.O. Formalizing agreement analysis for elicitation studies. In Proceedings of the 33rd Annual ACM Conference, Seoul, Korea, 18–23 April 2015; pp. 1325–1334.
23. Brooke, J. SUS—A Quick and Dirty Usability Scale. Available online: <https://hell.meiert.org/core/pdf/sus.pdf> (accessed on 19 June 2020).
24. Schrepp, M.; Hinderks, A.; Thomaschewski, J. Construction of a benchmark for the User Experience Questionnaire (UEQ). *Int. J. Interact. Multimed. Artif. Intell.* **2017**, *4*, 40. [CrossRef]
25. Tullis, T.; Albert, W. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*, 2nd ed.; Elsevier: Burlington, MA, USA, 2013; Interactive Technologies.
26. Schrepp, D.M. *User Experience Questionnaire Handbook*; 15. Available online: <https://www.ueq-online.org/Material/Handbook.pdf> (accessed on 21 June 2020).
27. Dix, A.; Finlay, J.; Abowd, G.; Beale, R. (Eds.) *Human-Computer Interaction*, 3rd ed.; Pearson/Prentice-Hall: Harlow, UK; New York, NY, USA, 2004; ISBN 978-0-13-046109-4.
28. Carter, S.; Mankoff, J.; Klemmer, S.R.; Matthews, T. Exiting the cleanroom: On ecological validity and ubiquitous computing. *Hum.-Comput. Interact.* **2008**, *23*, 47–99. [CrossRef]
29. Markopoulos, P.; Rauterberg, G.W.M. LivingLab: A white paper. *IPO Annu. Prog. Rep.* **2000**, *35*, 53–65.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).