*Article*

# Frame-Based Elicitation of Mid-Air Gestures for a Smart Home Device Ecosystem

**Panagiotis Vogiatzidakis *** and **Panayiotis Koutsabasis**

Department of Product and Systems Design Engineering, University of the Aegean, 84100 Mytilene, Greece; kgp@aegean.gr
* Correspondence: panosvog@aegean.gr; Tel.: +30-697-382-4408

**Abstract:** If mid-air interaction is to be implemented in smart home environments, then the user would have to exercise in-air gestures to address and manipulate multiple devices. This paper investigates 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 use of all devices to support contextual relevance; also, the referents were presented with minimal affordances to minimize widget-specific proposals. In addition to computing agreement rates for all referents, we also computed the internal consistency of user proposals (single-user agreement for multiple commands). In all, 1047 gestures from 18 participants were recorded, analyzed, and paired with think-aloud data. The study reached to a mid-air gesture vocabulary for a smart-device ecosystem, which includes several gestures with very high, high and medium agreement rates. Furthermore, there was high consistency within most of the single-user gesture proposals, which reveals that each user developed and applied her/his own mental model about the whole set of interactions with the device ecosystem. Thus, we suggest that mid-air interaction support for smart homes should not only offer a built-in gesture set but also provide for functions of identification and definition of personalized gesture assignments to basic user commands.

**Keywords:** frame-based; elicitation study; mid-air interaction; smart home; gesture proposal consistency; personalized gestures

## 1. Introduction

During the last decade, we have experienced an explosion in affordable motion tracking technologies, small enough to be integrated within home appliances or smart environments such as homes, industrial buildings, classrooms, labs, etc. These devices employ various technologies such as depth cameras, wearable assistive devices, and smartwatches to capture user's motion and translate it to a mid-air command to a specific device. Thus, mid-air interaction is becoming possible in various scenarios of pervasive and ubiquitous computing, and it is rapidly evolving into a distinguishable style of human–computer interaction (HCI) characterized by touchless, gesture-based interactions with remote displays or devices [1].

In mid-air interaction, there are neither universal gesture vocabularies nor established design patterns. The identification of appropriate gestures depends on the context of use and, thus, it is an important design issue. A user-centered approach to deal with gesture design is the gesture elicitation (or guessability) study, which is applied by designers to extract appropriate gestures that "meet certain design criteria such as discoverability, ease-of-performance, memorability, or reliability" [2].

The need for mid-air control of home devices can be argued for many use cases including user convenience (i.e., immediate (fast) control of a device, over remote control, which requires another control device), as well as for people that have temporary injuries or permanent mobility impairments

which affect their mobility inside the home. Mid-air interactions have been investigated in the smart home mainly to identify and craft gestures for single device interactions, like the TV [3–9] or various media (e.g., [6,10]). However, a smart home environment consists of several interconnected devices with variable sensing capabilities, which may be controlled and manipulated remotely. Thus, if mid-air interaction is to be implemented in smart home environments, then the user would have to exercise in-air gestures to address and manipulate multiple devices. A few studies have investigated mid-air interactions for a smart home device ecosystem; most of them employ gestures created by system designers [6,11–13] in contrast to a single study with user elicitation [14].

This paper presents an elicitation study of upper-body gestures for accessory-free, mid-air 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). We have investigated a potentially uniform gesture set for any device so that users would not have to remember different gestures for multiple devices in the home. The first result of this work is a user-defined gesture vocabulary for basic control of a smart home device ecosystem, which includes several gestures with very high, high, and medium agreement rates. Another important finding is that most users developed and applied their own mental model about the whole set of interactions with the device ecosystem, which resulted in high consistency within their gesture proposals across commands that apply to multiple devices. Thus, we suggest that mid-air interaction support for smart homes should not only provide a built-in gesture set, but also functions of identification and definition of personalized gesture assignments to basic user commands.

## 2. Related Work

### 2.1. On the Elicitation Study Method

A major issue in mid-air interaction design is the identification and crafting of gestures that are usable, comfortable, memorable, and discoverable. A user-centered approach to deal with gesture design is the gesture elicitation (or guessability) study, which aims at "*maximizing and evaluating the guessability of symbolic user input*" [15]. In brief, an elicitation study includes the following steps: (i) the researcher presents a referent (the effect of an action) to the participant; (ii) the participant proposes one or more gestures that better match the referent and possibly fills in a short post-task questionnaire; (iii) after all referents are presented, a wrap-up discussion or a questionnaire is filled in by the participant; (iv) the gesture set is derived from data analysis which mainly includes the calculation of agreement rates [16], among gesture proposals for each referent.

User elicitation is a genuinely user-centered method of requirements and design, however, it is not without some inherent shortcomings that need to be carefully addressed (as with any empirical research method). A primary concern is legacy bias [2], which leads users to propose gestures known from prior interactions because they cannot understand the fundamental capabilities of novel technologies; however, these proposals are often contextually irrelevant. Legacy bias can be reduced with the strategies of *production*, *partners*, and *priming*. 'Production' means to require users to propose many gestures for each referent, and 'Partners' suggests recruiting users in groups in order to leverage their ideas. Production is often encouraged in elicitation studies, in contrast to employing partners, since this increases the number of required participants. 'Priming' requires from users "*to think about the capabilities of a new form factor or sensing technology in order to produce ecologically valid gestures*". To increase the added value of priming, Cafaro et al. [17] propose the variation of 'Framed Guessability', which puts users within a 'frame', i.e., a scenario, to propose contextually relevant gestures (i.e., gestures and body movements that would make sense in the priming frame (scenario)).

### 2.2. Elicitation Studies for Control of Smart Home Devices

Several elicitation studies of mid-air gestures for control of home devices have been conducted. In a recent review of 47 mid-air gesture elicitation studies [18], mid-air interaction with the smart home

is investigated by 8.5% of papers examined while there are additional studies about control of digital content and devices that can be used in the home, like the TV (12.8%) and various media (14.9%).

Notably, elicitation studies of mid-air gestures for the smart home typically investigate user-defined gestures for interactions with one particular device. The most common device is the TV [3–9], which has evolved to an interactive multimedia device allowing web browsing, content manipulation, and media playback. There are many differences among these elicitation studies, including: differences among participants (sample sizes ranging from 18 to 81 users, apparent individual differences among participant groups, cultural differences); differences in the application of the gesture elicitation method; and different assumptions about sensor technology (e.g., some studies assume whole body tracking, others hand tracking only). Most importantly, a comparison of the results of the studies reveals that the proposed gestures are quite different, therefore a "*universal gesture set*" for mid-air interaction with smart TVs has not been obtained. Other home devices or activities investigated include music playback control [10] and "*entertainment environment*" [6].

In addition, a few studies have investigated particular interaction techniques for addressing [19] alternate devices or opt-in/opt-out from interactions [20], which are necessary in a multi-device interaction context. According to Freeman et al. [19] to address a device in mid-air interaction "*involves finding where to gesture so that their actions can be sensed, and how to direct their input towards that system so that they do not also affect others or cause unwanted effects*". In their approach, they propose an interaction technique where devices provide consistent light, tactile, and audio feedback to indicate to users how and where to gesture; off course, this presupposes that home devices are equipped with additional actuators. In a similar context, Walter et al. [21] investigated different ways to reveal the registration gesture which defines the beginning of mid-air interaction on a public display. This is an important affordance of any public device that can detect mid-air interactions, which can also contribute to avoiding the phenomenon of 'Midas' touch', i.e., activating a device accidentally. Also, in [20], a couple of interaction techniques for opt-in (wave with one or two hands) and opt-out (arms crossed and turn around with raised hand) from a distant display are empirically evaluated. Although these interaction techniques have not emerged from user elicitations, they have proven to be usable.

Recently a number of elicitation studies attempt to constrain the types of gestures investigated into particular types like on-skin [22] and single-hand microgestures [3]. Bostan et al. [22] investigate hand specific on-skin gestures for the control of a desktop interface and their main conclusion is that users tend to use their hands as a self-interface, where one hand plays the role of a surface and the other of a 'controller' to apply gestures similar to those of a multi-touch surface. Chan et al. [3], have conducted an elicitation study for single-hand microgestures, which can be tracked with wearable sensors like the Myo Armband. Another interesting approach of microgestures is proposed by Siddhpuria et al. [23], which assumes a smartwatch as a wearable sensor and actuator. These studies employ simple commands as referents which, to some extent, are similar to those required in a smart home environment. Furthermore, they explore interesting constraints on the type of features they investigate, which provides a more specific context of application; this is important, since that, most usually, the participants to an elicitation study are "*free, however, to suggest completely unrelated gestures for different effects*" [17]. Additionally, microgestures are discrete, which is an important requirement for ubiquitous environments, however, for the home this is perhaps not a high-priority requirement. However, all aforementioned studies investigated interactions with distant displays and, therefore, more research is required to validate their suitability to the context of a smart home device ecosystem.

Several studies have investigated mid-air interaction with a smart home device ecosystem however, one only was an elicitation study. Choi et al. [14] 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. In [12], the design of a gesture-based prototype for context-sensitive interaction with smart homes is presented for 7 commands and 4 devices; their

work did not involve gesture elicitation but prototype development with designer-defined interaction techniques and usability testing. Ng et al. [13] also present a prototype for home automation for 5 devices and 2 commands only (switch on/off). Other studies about pervasive smart home interactions either assume a handheld device (e.g., a Wiimote in [6]) or a smartphone used as a gesture surface only [11].

## 3. Study Design

### 3.1. Aim and Scope

The aim of the study is to identify a user-defined gesture vocabulary for mid-air control of a smart home device ecosystem consisting of 7 common devices: TV, speakers, video player, audio player, air-conditioner, lights, blinds. We are interested in identifying across-device interactions in which users would apply a single gesture for a command that applies for multiple devices in the home (e.g., the command 'switch on/off' applies for many devices). Therefore, we have also investigated user consistency of gesture proposals across devices, which has not been examined in previous studies.

### 3.2. Methodology and Procedure

In terms of methodology, we have followed a similar approach to framed guessability proposed by Cafaro et al. [17] in order to assist users in providing more intuitive and contextually relevant gestures. More specifically, we had prepared (a) a summary of a scenario of use all devices and (b) a more detailed a script of use of all home devices that was narrated to users at the time of the experiment. As a consequence, participants initially had to reflect on the scenario summary at the beginning of the process and then (during the process, at particular instances of the script) had to stop and think about how to apply a mid-air gesture to control a device. The process followed was identical for all users as well as the order by which the referents were presented. We did not record timings (thinking time, time-to-propose a gesture, etc.) because we followed a (concurrent) think-aloud approach which inevitably included some discussion with the researcher.

To minimize the legacy bias [2] or any other affordances that may stem out from particular devices, we had prepared referents (in a slideshow) with minimal illustrations of devices (without handlers or controls).

We focused on upper-body gestures assuming the user will be standing or seated, which is typical in the home. Furthermore, we were interested in 'walk up and use', 'bare hand and body' interactions that do not require any accessories or wearables; this implies that tracking technology would be ambient but oriented similarly to the location of the home devices.

We also instructed participants that they should think about interacting directly with the devices (illustrated in slides, like those of Figure 1) and not make assumptions about particular digital content or specific manipulations or user interfaces with menus, controls, buttons, etc.

During the unfolding of the scenario, when it was time for participants to apply a gesture, they were asked to produce more than one gesture for a referent, while for the referents of registration on/off, they were instructed to produce exclusive gestures for each device. During this procedure, they were encouraged to think aloud and verbally state the beginning and the end of each proposed gesture. The experiment took about one hour, on average, for each participant, and extensive notes were taken.

### 3.3. Participants

Eighteen participants took part in this study, 6 females and 12 males whose age ranged from 24 to 47 ($M_{age}$ = 37.8, $SD_{age}$ = 7.4). Most of them were recruited from our research lab environment. Their background was in computer science, engineering, and design. Ten participants had some experience with mid-air gesture interaction because they had participated in other empirical studies (concerning interactions with Leap motion and Kinect), however, none were highly experienced with the technology (e.g., a researcher in the field or a designer or developer).
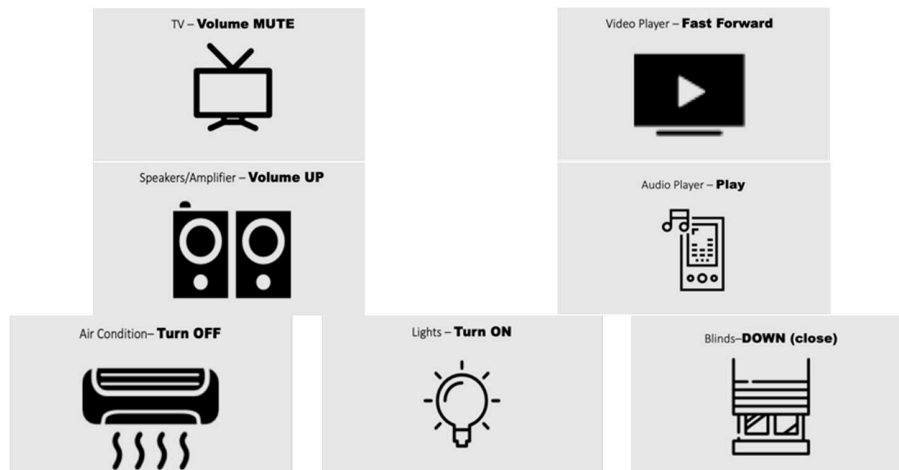
**Figure 1.** Sample slides for the list of referents.

*3.4. Apparatus*

We conducted the study in a home office room. A MacBook pro was connected to an external 26-inch monitor and the participants were seated in front of it in approximately 2 m. We used MS PowerPoint to display slides of each referent per device. A webcam was used to video capture the whole process with Camtasia Studio software (https://www.techsmith.com/video-editor.html).

*3.5. Referents*

After the review of related work on commands and references for smart home devices (especially [4,5,9–11,14,24–29]), we identified a list of commands that were frequently used and applicable in multiple devices. We listed all commands and then removed those that referred to particular controls of a user interface. We ended up with twelve (12) commands, each one applying to at least 2/7 home devices of our scenario (Table 1). We have added two more commands to these: registration on and registration off, which apply to all devices. User registration is required to avoid the phenomenon of 'Midas Touch', which refers to that "every active hand action, even unintentional, could be recognized as a command by vision-based technology" [29].

We were concerned with the form and function of referents and the affordances these may convey to users. To minimize interaction affordances [30], the device icons did not include details of the controls (i.e., buttons, knobs, sliders etc.) since users might associate them with specific gestures (Figure 1).

**Table 1.** List of 7 devices, 14 commands, and 55 referents.

| Devices<br>Commands | TV | Speakers | Video<br>Player | Audio<br>Player | Air-<br>Conditioner | Lights | Blinds |
|---|---|---|---|---|---|---|---|
| Registration On | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Registration Off | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Turn On | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Turn Off | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Up | ✓ | ✓ | | | ✓ | ✓ | ✓ |
| Down | ✓ | ✓ | | | ✓ | ✓ | ✓ |
| Next | ✓ | | ✓ | ✓ | | | |
| Previous | ✓ | | ✓ | ✓ | | | |
| Mute | ✓ | ✓ | | | | | |
| Fast Forward | | | ✓ | ✓ | | | |
| Fast Rewind | | | ✓ | ✓ | | | |
| Play | | | ✓ | ✓ | | | |
| Pause | | | ✓ | ✓ | | | |
| Stop | | | ✓ | ✓ | | | ✓ |
| Referents | 9 | 7 | 11 | 11 | 6 | 6 | 5 |

### 3.6. Data Analysis

Data analysis included retrospective video analysis, gesture classification, and calculation of agreement rates. Retrospective video analysis included the review of the video recordings from each user and extracting the 'stroke' phase [31] for every gesture proposed. Then a list was made with a small description of the gesture and an index number (i.e., G1, G2, G3, ... , etc.) was assigned to each one of them. Gestures with similar motion or posture were considered as the same. MS Excel was used to list all the gestures index-number by user and by referent and to calculate the most popular gesture for each referent.

#### 3.6.1. Gesture Classification—Taxonomy

Numerous studies have categorized gesture preferences into taxonomies according to their characteristics, to gain some insights and get a better understanding of the users' design space. Most of them are based on the taxonomy proposed by Wobbrock et al. In 2009 [32], which involved gestures for interactive tabletops. Since then, similar taxonomies has been adapted for studies on mid-air interaction [11,29] with mobile phones [33], wearables [23,34,35], on-skin microgestures [22], augmented reality [36], etc. In our study, gestures are classified along five main dimensions: Nature, Form, Body Part, Flow, and Spatial (Table 2).

**Table 2.** Taxonomy of motion gestures based on collected gestures.

| Dimension | Category | Description |
|---|---|---|
| Nature | Symbolic<br>Metaphorical<br>Abstract | Gesture depicts a symbol.<br>Gesture indicates a metaphor.<br>Gesture-command mapping is arbitrary. |
| Form | Static<br>Dynamic | Gesture is a pose.<br>Gesture involves body movement. |
| Body Part | One hand<br>Both Hands<br>Full body | Gesture is performed with one hand.<br>Gesture is performed with both hands.<br>Gesture involves movement or pose of upper body. |
| Flow | Discrete<br>Continuous | Command is performed on gesture completion.<br>Command is performed during the gesture. |
| Spatial | 2-Dimensional<br>3-Dimensional | Gesture is performed on single axis.<br>Gesture is performed in 3D space. |

The Nature dimension refers to the mental model behind the gesture-meaning relationship and involves subcategories of symbolic, metaphorical, and abstract gesture types. Symbolic gestures depict a symbol, such as an "X" formed with both arms, or a rectangle sketched on the air with both hands to denote a cancelation of a command. Metaphorical gestures mimic real-world interaction, such as rotating a knob or moving a slider control up/down to adjust the volume. Any gesture that has an arbitrary mapping with the referent is characterized as abstract, such as opening or closing a fist.

Form dimension distinguished moving gestures and postures. Static gestures trigger the command after the body part(s) are still, in a specific location for a certain amount of time, such as forming a "+" with two pointers. In opposite, dynamic gestures, such as a hand-swipe to the left or right, involve a movement of body parts during the "stroke" phase and between the "preparation" and "retraction" phases [31] of the gesture.

The Body Part dimension indicates whether a gesture is performed with one hand or both hands. In cases where a gesture acts in combination with another part of the body, such as "pointing on eye", or "covering both ears", we characterize it as full body.

In the Flow dimension, gestures are categorized as either discrete or continuous. A discrete gesture triggers the command only when completed. For example, the gesture of pushing a button with a palm to turn on/off a device is considered as discrete since the command will be triggered only

when the user stretches her arm to a certain level. In Continuous gestures, the command acts as long as the gesture is performed.

The Spatial dimension is used to describe whether the gesture acts in a 2D plane or in 3D space [8]. Two-dimensional gestures involve gestures in which users move their hands in either the vertical or horizontal axis in a 2D plane, such as swipe with one hand left/right or up/down. Others, for example, hand rotational motion, or metaphorical gestures, such as holding arms with their hands, or covering their ears, require movement in three axes, and are categorized as 3-dimensional.

3.6.2. Gesture Consensus Level among Users (Agreement Rate)

To evaluate the level of gesture consensus between users, the gesture agreement rate (*AR*) was calculated using the Equation (1) [16] with the accompanying AGATe (AGreement Analysis Gesture Toolkit) software (http://depts.washington.edu/madlab/proj/dollar/agate.html), where *P* is the number of all proposals for a given referent *r*, and $P_i$ is a subset of the identical gesture proposals. *AR* can take values from 0 (no agreement) to 1 (total agreement).

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \in P} \left(\frac{|P_i|}{|P|}\right)^2 - \frac{1}{|P|-1} \tag{1}$$

For example, for the "Blinds-Stop" referent, we obtained 4 different gestures with popularity 13, 2, 2, and 1, from 18 users. The agreement rate is computed as follows:

$$AR(Blinds - Stop) = \frac{18}{18-1}\left[\left(\frac{13}{18}\right)^2 + \left(\frac{2}{18}\right)^2 + \left(\frac{2}{18}\right)^2 + \left(\frac{1}{18}\right)^2\right] - \frac{1}{18-1} = 0.52.$$

Agreement rates were calculated, and the referents were classified as "Very high", "High", "Medium" or "Low" based on the agreement intervals of >0.5, 0.3–0.5, 0.1–03 and ≤0.1 respectively, according to the classification proposed in [16].

3.6.3. Gesture Consensus Level among Devices (Consistency Rate)

As noted above, our goal in this study is to also examine how consistent each user is when (s)he proposes gestures for similar referents on different devices (i.e., a single command). To quantify the user consistency, we employed the same formula as was used for calculating the agreement rate, since we detected many similarities. More specifically, agreement rate measures the agreement of the gestures proposed from many users, for a given referent on one device, while in our case, we examine the agreement of the gestures proposed from one user for a given command on many devices (Figure 2).
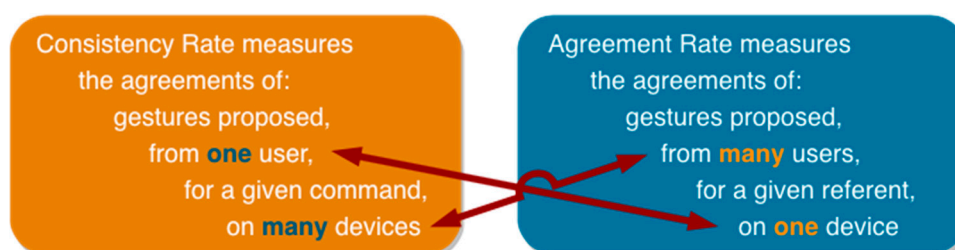


**Figure 2.** Comparison of agreement rate and consistency rate.

In short, we do not introduce a new metric, rather, we suggest a new way of using the agreement rate to evaluate user's gesture consistency among different devices. For simplicity reasons, we will

refer to it as "consistency rate—$CR(c)$" (Equation (2)), where $P$ is the number of all proposals for a given command $c$ which is similar in all devices, and $P_i$ is a subset of the identical gesture proposals.

$$CR(c) = \frac{|P|}{|P|-1} \sum_{P_i \in P} \left(\frac{|P_i|}{|P|}\right)^2 - \frac{1}{|P|-1} \qquad (2)$$

For example, for the command "Up" which is applied on five devices (i.e., referents: TV—volume up, speakers—volume up, air-conditioner—temperature up, lights—dim up, blinds—up) a user proposed 2 different gestures with popularity 4 and 1. We calculate the consensus rate as follows:

$$CR(Up) = \frac{5}{5-1}\left[\left(\frac{4}{5}\right)^2 + \left(\frac{1}{5}\right)^2\right] - \frac{1}{5-1} = 0.6$$

Users' $CR(r)$ for registration on and off was not evaluated since these referents must not be the same among different devices. We used the same classification intervals (Very high: >0.5, High: 0.3–0.5, Medium: 0.1–0.3, and Low ≤ 0.1) as in $AR(r)$.

## 4. Results

### 4.1. User-Defined Gesture Set

A total of 1047 gestures (108 unique) were proposed from 18 participants for 55 referents. Agreement rates were calculated for each referent. We selected the most popular gesture for each referent to construct the gesture vocabulary, which is shown in Table 3 and illustrated in Figure 3. From these gesture proposals, 3 (5%) were with very high agreement, 8 (15%) with high agreement, 13 (24%) with medium agreement and 31 (56%) with low agreement.
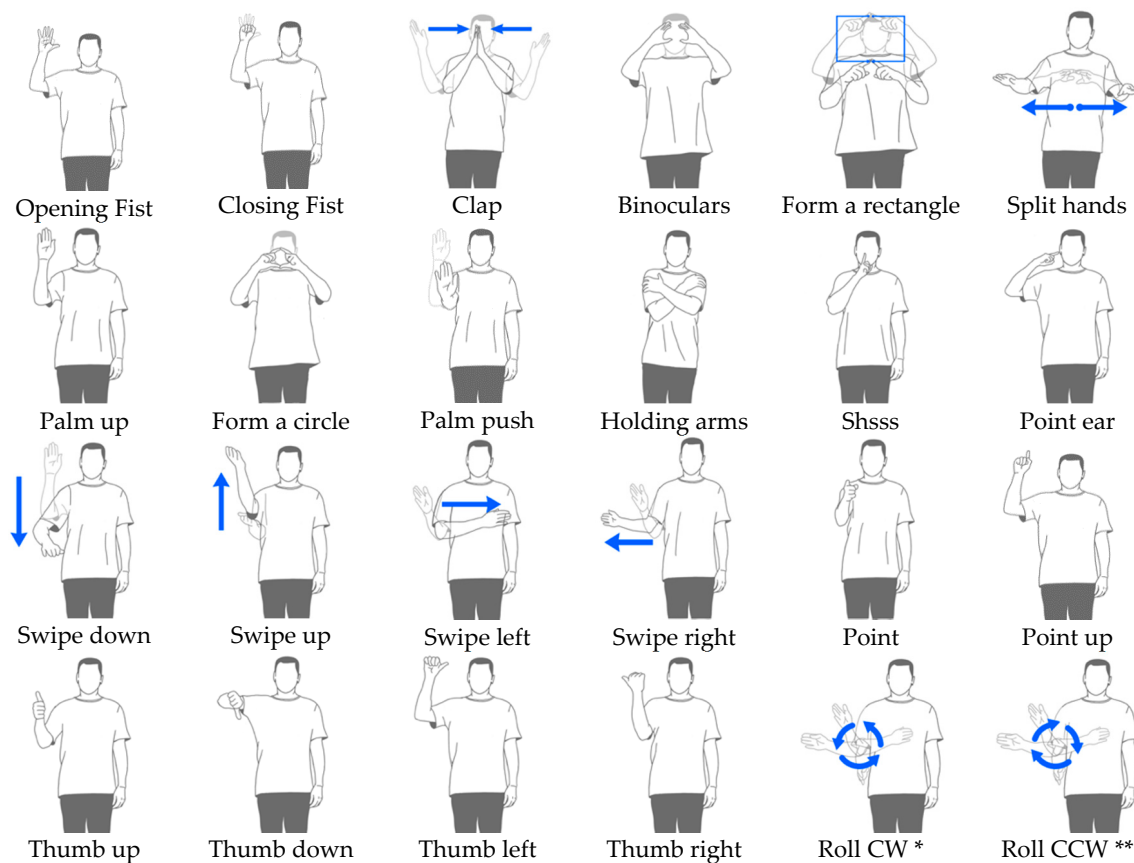


**Figure 3.** Proposed gestures with higher agreement rates (*: Clockwise, **: Counter-Clockwise).

**Table 3.** List of gestures for each command. * Gestures with both hands. Agreement rate (AR) for each command (c) is shown in parenthesis. Color of cells denote the classification of gesture agreement depending on the value of AR(c)—green color for AR ≥ 0.1 and white for AR < 0.1.

| | TV | Speakers | Video Player | Audio Player | Air Condition | Lights | Blinds |
|---|---|---|---|---|---|---|---|
| **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) |
| **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) |
| **Turn On** | Point (0.04) | Opening fist (0.05) | Opening fist (0.01) | Palm push (0.05) | Opening fist (0.07) | Point up (0.05) | |
| **Turn Off** | Closing fist (0.05) | Closing fist (0.08) | Closing fist (0.04) | Closing fist (0.03) | Closing fist (0.07) | Closing fist (0.03) | |
| **Up** | Swipe up (0.52) | Swipe up (0.46) | | | Swipe up (0.28) | Swipe up (0.28) | Swipe up (0.33) |
| **Down** | Swipe down (0.52) | Swipe down (0.46) | | | Swipe down (0.33) | Swipe down (0.32) | Swipe down (0.33) |
| **Next** | Swipe left/right (0.34) | | Point right (0.14) | Swipe left/right (0.16) | | | |
| **Previous** | Swipe left/right (0.34) | | Point left (0.14) | Swipe left/right (0.16) | | | |
| **Mute** | Ssshhh (0.12) | Ssshhh (0.10) | | | | | |
| **Fast Forward** | | | Swipe left/right (0.17) | Move hand clockwise (0.09) | | | |
| **Fast Rewind** | | | Swipe left/right (0.17) | Move hand counterclockwise (0.09) | | | |
| **Play** | | | Palm up (0.03) | Point (0.08) | | | |
| **Pause** | | | Palm stop (0.09) | Palm stop (0.15) | | | |
| **Stop** | | | Palm stop * (0.07) | Hands split * (0.14) | | | Palm stop (0.52) |

During the study, almost one-third of the users (7 out of 18, 39%) showed difficulty in understanding the meaning of the registration on/off command, confusing it with the turn on/off command, especially at the beginning of the study. Many participants reported the concept of registration to a device as quite abstract and new for them, and in most cases, we had to provide further explanation about the similarity with the real world by comparing it with the act of grabbing the device's remote control before interacting with it. After understanding the idea of registration, many users asked to change their initial proposals, and two of them mentioned that it would be helpful to have a kind of feedback from the device (such as a green light) as long as the user is registered indicating that motion tracking has been activated. The majority of the participants came up with arbitrary gestures with low agreement rates, confirming the association of abstract referents and arbitrary gestures [4,11,37–39].

Swipe up/down gestures (Figure 3) showed great consensus across users for the referents of increasing and decreasing a device's function such as the volume (TV and speakers), temperature (air-conditioner), luminance (lights and blinds). Their agreement rates ranged above 0.3 for most of them (8 out of 10, 80%).

Many users attempted to reduce their cognitive load by using same gestures for either binary referents on the same device or for the same commands across devices. Binary referents have an on/off meaning, such as turn on/off and registration on/off. For those referents, most participants assumed that they could interchange these two binary states by using the same gesture, such as pushing an imaginary button with their palm to turn on and off the TV, or clapping to turn on and off the lights.

Surprisingly, even though participants used the same gestures for binary commands, these commands showed low agreement rates (Table 3).

Many users claimed that the referents are too many to remember in a real scenario. Most of them eventually figured out that while they are been registered on a device, any gesture they perform is ignored by the other devices, and they could therefore reuse the same gesture for the same commands on different devices. For instance, the gesture of "Shhsss" (Figure 3) was proposed for muting the volume on the TV and speakers.

Even though we repeatedly encouraged participants to propose more than one gestures for each referent, only 7 out of 18 (39%) did so. These 5 participants, although they were very productive at the beginning of the study (one of them was proposing about 3–4 gestures for each referent), as the study evolved, the rate of gesture production was diminishing and at the end, only single gestures were proposed for similar referents.

*4.2. Gesture Taxonomy*

In the *Nature* dimension of our taxonomy, the majority of gestures proposed were metaphorical (62%), followed by abstract (22%) and symbolic (16%) (Figure 4). In most cases, metaphorical gestures users were correlating their sense organs with a device, such pointing or covering their ear to interact with audio device, point on eyes to interact with TV or video player. Symbolic gestures involved either in-air sketching a symbol or a letter with their pointer finger (such as sketching a triangle, for the play command, binding it with the standard icon ▷ that can be found in most players), or forming a symbol with their hands (such as forming a rectangle with both hands as a symbolization of the TV's shape).
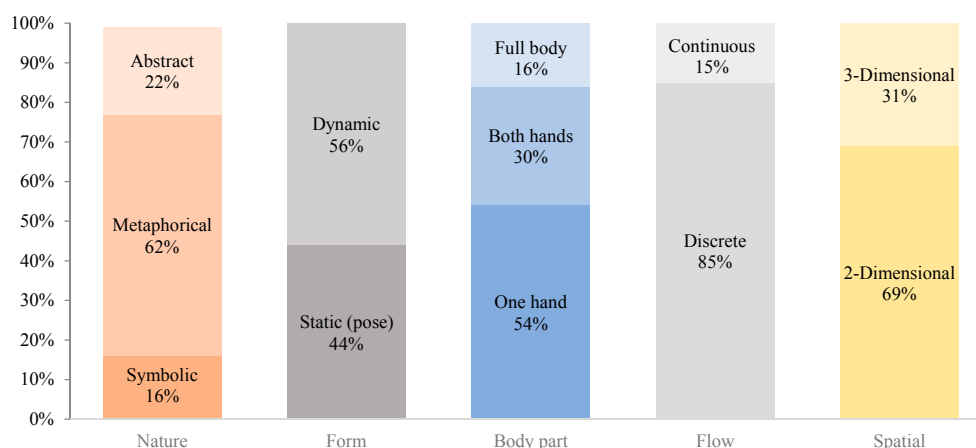


**Figure 4.** Gesture taxonomy for identified gestures.

The *Form* dimension has an almost equal distribution of gestures along its static (44%) and dynamic (56%) category. We observed that many of the static gestures proposed were used to indicate a direction (up/down for volume or luminance, left/right for next and previous) using their thumb or pointer finger.

In the *Body Part* dimension, more than half (54%) of the proposed gestures were performed with one hand, such as opening and closing a fist, swiping one arm left/right, in air sketching symbols, etc. The majority of the full body gesture proposed were also symbolic ones, such as pointing or covering eyes or ears, and touching their chest.

In the dimension of gesture *Flow*, continuous gestures were very few (15%) in comparison to discrete ones (85%). Commands like increase/decrease volume or fast-forward and fast-rewind were considered as continuous gestures by many of our participants and, in most cases, involved repeating movements. However, there were some users that proposed a hand posture, like a palm or thumb facing left/right, to keep the command on, as long as their hand was steady.

In regard to the *Spatial* dimension, almost two-thirds (69%) of the participants' gestures were 2-dimensional. Most of these gestures were swipe up/down/left/right, performed with one hand on a single axis of a two-dimensional plane. Other 3-dimensional gestures, such as clapping or splitting hands, were performed in 3D space.

*4.3. Users Gesture Consistency among Devices*

For every user, we calculated the consistency rate for the 12 similar commands (excluding 2 commands about registration on/off, as explained earlier) among the 7 devices and their averages are shown in Figure 5. Based on the same classification intervals as [16], 3 users (17%) had medium consistency, 2 users (11%) had high consistency, and the majority of them (13 out of 18, 72%) showed very high consistency. Interestingly, there was no participant with low CR(c) (>0.1).
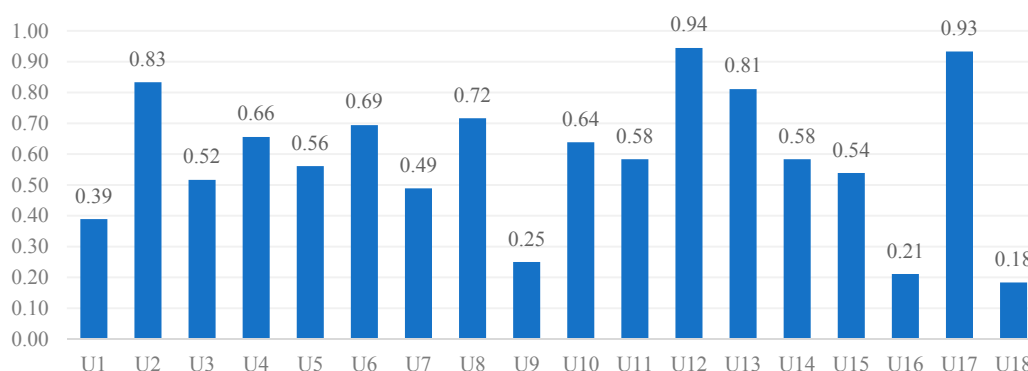


**Figure 5.** Average consistency rate per user.

At the beginning of elicitation sessions, after the participants were presented with the summary of the scenario (frame), we observed that most users (15/18) spent some time (ranging from about half a minute to about 3 minutes) to think about a potentially universal gesture set for commands. During this think-aloud process, most users asked for clarifications about what types of gestures might be preferable or what parts of the body to employ, or to remind them about the devices and commands. Clarifications were given in a neutral manner. Therefore, the scenario summary presented to them at the beginning of the elicitation session helped most participants to reflect and prepare themselves for proposing meaningful gestures.

As the study evolved, most users felt more confident about their proposals and needed less time to think. A possible reason is that participants created a mental model for interacting with the whole ecosystem, and after that, it became very natural to use same gestures for conceptually similar commands. However, three users begun proposing gestures quite fast, and it was soon apparent that they had obviously misunderstood the necessity of the registration gesture and the difference between registration and turn-on command. In such cases, we intervened and clarified the meaning of registration gesture/command. Most of these users (apart from one) were asked to reform their proposals, and in some cases, we had to roll-back the procedure and start from the beginning. Therefore, during the elicitation process, we attempted to keep users within the scenario given as well as eliminate misunderstandings, when detected.

We identified three users with very high consistency rate (Figure 5), who all spent some time at the beginning of the process to elaborate the concept of our scenario and think about their proposals. All of them admitted seeking for gestures that would fit in most referents in order to reduce their cognitive load, and when they started to produce their preferences, they seemed very confident about their choices. To illustrate with an example, user U12 was concerned from the beginning of the process, about technical details of implementation, such as how many cameras are tracking his movements, how these are set up in the room, their vision angle, etc. He implied that he was trying to understand how the motion tracking ecosystem works as he was seeking to find a way to use only one gesture

to register to all devices in order to minimize the total number of gestures needed. That user was 100% consistent (1 proposed gesture per command) for 11 out of 12 different commands. Similarly, U2 and U17 were 100% consistent in 9 and 10 respectively, out of 12 commands. All three users had in common the fact that they spend some time in the beginning to understand the concept and to find a way to minimize the different gestures needed for all the commands. Furthermore, they all had some experience with mid-air interactions.

On the other hand, three users with average consistency (bellow 0.3) reported that they could not find a similar pattern across the referents. These users had no prior experience with gestural interactions. We found that they were creative in their gesture proposals, but often these were contextually irrelevant. For example, U16 (CR = 0.21) could not anticipate at the beginning of the process, the total number of unique gestures that she would have to remember. Even when she realized that, just before the end of the process, she remained to her choices, implying that she wanted to be more kinesthetically expressive. Also, U9 and U18 seemed like they did not entirely understand the meaning of registration.

## 5. Discussion

### 5.1. Comparison to Other Gesture Sets

Choi et al. [14], have presented a repeated elicitation study for 7 devices (air-conditioner, TV, lights, blinds, curtains, door, window) and 20 referents, without investigating a registration gesture. If we compare our results with the (second) gesture set (which presented significantly higher agreement rate), we can see that there are only a few gestures that are alike: the swipe left/right for changing TV channels, and the swipe up/down for air-conditioning. Furthermore, several gestures are similar to those identified in our study, however for other operations. For example, they identify a pointing gesture for turn on/off the air-conditioner, while in our study, this is what was proposed for the blinds. A possible explanation for this mismatch could be that in their research, Choi et al. [14] assume "*that technical barriers on the gesture recognition are completely resolved*" while, in our study, we approach this scenario more holistically, and we are concerned about the communication isolation between the user and the device at a time. Since we believe that this is a major design challenge and implementation issue for an interconnected ecosystem to function robustly, we have introduced the registration command which will be used to switch communication with another device.

In [40], a prototype of gestural control of household appliances is presented, which includes ten (10) designer-defined, single-hand gestures for 2 devices: 6 gestures for the radio (on/off, increase/decrease volume, next/previous station) and 4 for lights (on/off, increase/decrease illuminance), while they address the devices with a pointing gesture. There is little agreement between their findings and ours. Our findings suggest the use of an iconic gesture for addressing a device, which is safer in terms of avoiding accidental activation. Many of their gesture definitions are similar to ours, for example, they also include swipe up/down, swipe left/right, but they match these gestures to other operations than those identified in our study.

Comparisons to other gesture sets [6,10,12,13] yield similar results: there is some agreement in produced gestures, however, the operations typically differ. There are many reasons for this, including apparent cultural differences among participants (between studies), sampling issues (demographics, previous experience, etc.), as well as method application issues.

Notwithstanding these known issues, it seems that many similar gestures have been identified, but for different operations; in other words, the gestures are here, but the exact interaction techniques remain to be designed. Definitely, in an iterative user-centered design process, several other steps follow up with gesture elicitation, including gesture modeling (to particular joints of the human body), user interface design, and prototyping (e.g., especially for operations with little gesture agreement); gesture recognition engine design and prototyping; empirical testing of the usability and UX of the end system. During this process, it has been shown that user-defined gestures are not always the most usable, and that measured usability significantly affects user preferred gestures [41]. Nevertheless, the results of the study presented in this paper are a good starting point for system designers and

developers because they refer to the wider range of devices and referents compared to other studies and they have been produced with attention to the context of use.

### 5.2. Increased Consistency Rate Raises the Requirement for Personalized Gesture Construction

We have examined the gesture consensus level among devices for every single participant (named: consistency rate) and we have found that consistency rate is very high for most, which is interesting for many reasons.

We observed that most participants (15 from 18), sooner or later, formed a mental model for the device ecosystem, as a whole, that allowed them to propose similar gestures for similar commands. We observed several strategies about forming the mental model: some participants attempted to 'mimic another ecosystem', such as the "in-vehicle ecosystem", which was claimed to have some similarities (the driver is also seated, and there are some similar devices like, speakers, mp3 player, air condition, windows). Others attempted to bind the name of the referent with an iconic gesture (e.g., mid-air drawing of an 'S' to stop). Others performed gestures drawn from human-to-human communication like the 'Sshhh' and 'Point ear' (Figure 3). Regardless of the strategy, the fact that most participants attempted to group interactions among devices reveals that future interactions with a smart home device ecosystem will have to consider this requirement and allow users to define gestures on-the-fly.

Therefore, personalized gesture production would allow users to revisit the given gesture set and alter it to match their own mental model and preferences. Generally, personalization of gesture interaction is recognized as a significant challenge which can address the diversity and cultural background belonging to different (communities of) users [42]. Several toolkits for gesture recording have been developed over the last few years [43–46], however, only a few studies have integrated user elicitation with gesture personalization in a prototype system like Wu et al. [29] for TV viewing and Albertini et al. [47], for manipulation of archaeological data. Apparently, a usable system for personalized construction of gestures in a multi-device smart home ecosystem presents several challenges which remain to be addressed by future work.

### 5.3. Attending the Application Context in User Elicitation

In this frame-based elicitation study, we have attempted to inculcate contextual conditions to participants by narrating a scenario of use of all devices and asking participants at particular times to interact with a device and propose a mid-air gesture. This approach was followed in order to maximize environmentally valid gesture proposals. We also found that it was more entertaining for the participants since they were encouraged to think about the scenario to produce intuitive gestures. This was achieved for most users, as revealed by the consistency rates. Therefore, in the same line to others [17,42], we suggest that an elicitation study needs to prime participants with a "context" in order to maximize environmental validity and intuitiveness of gesture proposals.

### 5.4. Attending Referents' Affordances

We have designed the referents of this study so that these did not include details of the controls (i.e., buttons, knobs, sliders etc.), since users might associate them with specific gestures. There are several approaches of referent presentation in elicitation studies [18]: GUI animations, text messages on screen, described verbally, still images, and illustrations of manipulations of the actual artifacts. Apparently, if a referent has been designed in detail, it would have included widgets and handles which present affordances to participants of a gesture elicitation study. However, user elicitation (as a pre-design method) is typically conducted before detailed user interface design; in this case, it is important that the design of referents does not include unnecessary user interface controls or widgets that present—unwanted—affordances about particular user manipulations.

## 6. Conclusions

This paper presents an elicitation study of upper-body gestures for accessory-free, mid-air control of a smart home device ecosystem. We have investigated a potentially uniform gesture set for any device so that users would not have to remember different gestures for multiple devices in the home.

The proposed user-defined gesture vocabulary includes several gestures with very high, high, and medium agreement rates. Additionally, we have seen that most users developed and applied their own mental model about the whole set of interactions with the device ecosystem, which resulted in high consistency within their gesture proposals across commands that apply to multiple devices; this raises the requirement for personalized gesture construction in the home.

We have followed an approach for user elicitation that inculcates contextual conditions to participants by narrating a scenario of use and asking participants at particular times to propose a mid-air gesture. We found that this approach was entertaining for the participants and that they achieved high consistency rates in their gesture proposals. More specifically, we first presented participants with a scenario summary, which we found that helped most participants to reflect and prepare themselves for proposing meaningful gestures. Additionally, during the elicitation process a scenario script (that had been prepared beforehand) was narrated by the researchers to until it was time for the participant to propose a gesture (the same scenario was narrated to all participants). We found that this process assisted most participants in producing contextually relevant gestures.

The future work of this research is to refine the design and develop the proposed gesture set in a virtual smart home prototype or a living lab and empirically test and evaluate various usability and UX aspects using qualitative metrics such as users' satisfaction, comfort, fatigue, appropriateness for tasks, as well as quantitative metrics such as task success, errors, time of completion, and others.

**Author Contributions:** Conceptualization: P.V. and P.K.; Data preparation: P.V.; Data analysis: P.V. and P.K.; Methodology, P.V. and P.K.; Project administration, P.K.; Writing—original draft: P.V.; Writing—review & editing: P.K.

## References

1. Koutsabasis, P.; Vogiatzidakis, P. Empirical Research in Mid-Air Interaction: A Systematic Review. *Int. J. Hum. Comput. Interact.* **2019**, 1–22. [CrossRef]
2. Morris, M.R.; Danielescu, A.; Drucker, S.; Fisher, D.; Lee, B.; Wobbrock, J.O. Reducing legacy bias in gesture elicitation studies. *Interactions* **2014**, *21*, 40–45. [CrossRef]
3. Chan, E.; Seyed, T.; Stuerzlinger, W.; Yang, X.-D.; Maurer, F. User Elicitation on Single-Hand Microgestures. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 3403–3414.
4. Dim, N.K.; Silpasuwanchai, C.; Sarcar, S.; Ren, X. Designing Mid-Air TV Gestures for Blind People Using User- and Choice-Based Elicitation Approaches. In Proceedings of the 2016 ACM Conference on Designing Interactive Systems, Brisbane, Australia, 4–8 June 2016; pp. 204–214.
5. Dong, H.; Danesh, A.; Figueroa, N.; Saddik, A.E. An Elicitation Study on Gesture Preferences and Memorability Toward a Practical Hand-Gesture Vocabulary for Smart Televisions. *IEEE Access* **2015**, *3*, 543–555. [CrossRef]
6. Vatavu, 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.
7. Vatavu, R.D.; Zaiti, I.A. Leap gestures for TV: Insights from an elicitation study. In Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, Newcastle Upon Tyne, UK, 25–27 June 2014.
8. Wu, H.; Wang, J. User-Defined Body Gestures for TV-based Applications. In Proceedings of the 2012 Fourth International Conference on Digital Home, Guangzhou, China, 23–25 November 2012; pp. 415–420.
9. Zaiți, I.-A.; Pentiuc, Ş.-G.; Vatavu, R.-D. On free-hand TV control: Experimental results on user-elicited gestures with Leap Motion. *Pers. Ubiquitous Comput.* **2015**, *19*, 821–838. [CrossRef]

10. Löcken, A.; Hesselmann, T.; Pielot, M.; Henze, N.; Boll, S. User-centred process for the definition of free-hand gestures applied to controlling music playback. *Multimed. Syst.* **2011**, *18*, 15–31. [CrossRef]

11. 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.

12. 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.

13. Ng, W.L.; Ng, C.K.; Noordin, N.K.; Ali, B.M. Gesture Based Automating Household Appliances. In Proceedings of the International Conference on Human-Computer Interaction, Orlando, FL, USA, 9–14 July 2011; Springer: Berlin/Heidelberg, Germany, 2011.

14. Choi, E.; Kwon, S.; Lee, D.; Lee, H.; Chung, M.K. Can User-Derived Gesture be Considered as the Best Gesture for a Command? Focusing on the Commands for Smart Home System. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Boston, MA, USA, 22–26 October 2012; Human Factors and Ergonomics Soc.: Santa Monica, CA, USA; Volume 56, no. 1. pp. 1253–1257.

15. Wobbrock, J.O.; Aung, H.H.; Rothrock, B.; Myers, B.A. Maximizing the guessability of symbolic input. In Proceedings of the CHI'05 Extended Abstracts on Human Factors in Computing Systems, Portland, OR, USA, 2–7 April 2005; ISBN 1-59593-002-7.

16. Vatavu, R.-D.; Wobbrock, J.O. Formalizing Agreement Analysis for Elicitation Studies. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 1325–1334.

17. Cafaro, F.; Lyons, L.; Antle, A.N. Framed Guessability: Improving the Discoverability of Gestures and Body Movements for Full-Body Interaction. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–12.

18. Vogiatzidakis, P.; Koutsabasis, P. Gesture Elicitation Studies for Mid-Air Interaction: A Review. *Multimodal Technol. Interact.* **2018**, *2*, 65. [CrossRef]

19. Freeman, E.; Brewster, S.; Lantz, V. Do That, There: An Interaction Technique for Addressing In-Air Gesture Systems. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 2319–2331.

20. Rodriguez, I.B.; Marquardt, N. Gesture Elicitation Study on How to Opt-in & Opt-out from Interactions with Public Displays. In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, Brighton, UK, 17–20 October 2017; pp. 32–41.

21. Walter, R.; Bailly, G.; Müller, J. StrikeAPose: Revealing mid-air gestures on public displays. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 2 May 2013; pp. 841–850.

22. Bostan, I.; Buruk, O.T.; Canat, M.; Tezcan, M.O.; Yurdakul, C.; Göksun, T.; Özcan, O. Hands as a Controller: User Preferences for Hand Specific On-Skin Gestures. In Proceedings of the 2017 Conference on Designing Interactive Systems, Edinburgh, UK, 10–14 June 2017; pp. 1123–1134.

23. Siddhpuria, S.; Katsuragawa, K.; Wallace, J.R.; Lank, E. Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments. In Proceedings of the 2017 Conference on Designing Interactive Systems, Edinburgh, UK, 10–14 June 2017; pp. 1111–1122.

24. Fariman, H.J.; Alyamani, H.J.; Kavakli, M.; Hamey, L. Designing a user-defined gesture vocabulary for an in-vehicle climate control system. In Proceedings of the 28th Australian Conference on Computer-Human Interaction, Launceston, Australia, 29 November–2 December 2016; pp. 391–395.

25. Jahani, H.; Alyamani, H.J.; Kavakli, M.; Dey, A.; Billinghurst, M. User Evaluation of Hand Gestures for Designing an Intelligent In-Vehicle Interface. In *Designing the Digital Transformation*; Maedche, A., vom Brocke, J., Hevner, A., Eds.; Springer International Publishing: Cham, Switzerland, 2017; Volume 10243, pp. 104–121. ISBN 978-3-319-59143-8.

26. Riener, A.; Weger, F.; Ferscha, A.; Bachmair, F.; Hagmuller, P.; Lemme, A.; Muttenthaler, D.; Pühringer, D.; Rogner, H.; Tappe, A. Standardization of the in-car gesture interaction space. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Eindhoven, The Netherlands, 28–30 October 2013; pp. 14–21.

27. Rovelo Ruiz, G.A.; Vanacken, D.; Luyten, K.; Abad, F.; Camahort, E. Multi-viewer gesture-based interaction for omni-directional video. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 4077–4086.

28. Vatavu, R.-D. User-defined gestures for free-hand TV control. In Proceedings of the 10th European Conference on Interactive tv and Video, Berlin, Germany, 4–6 July 2012; pp. 45–48.

29. Wu, H.; Wang, J.; Zhang, X. User-centered gesture development in TV viewing environment. *Multimed. Tools Appl.* **2016**, *75*, 733–760. [CrossRef]

30. Dix, A.; Finlay, J.; Abowd, G.; Beale, R. *Human-Computer Interaction*, 3rd ed.; Pearson/Prentice-Hall: Harlow, UK; New York, NY, USA, 2004; ISBN 978-0-13-046109-4.

31. McNeill, D. *Hand and Mind: What Gestures Reveal about Thought*; University of Chicago Press: Chicago, IL, USA, 1992.

32. 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, 4–9 April 2009; pp. 1083–1092.

33. Ruiz, J.; Li, Y.; Lank, E. User-defined motion gestures for mobile interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; pp. 197–206.

34. Arefin Shimon, S.S.; Lutton, C.; Xu, Z.; Morrison-Smith, S.; Boucher, C.; Ruiz, J. Exploring Non-touchscreen Gestures for Smartwatches. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 3822–3833.

35. Di Geronimo, L.; Bertarini, M.; Badertscher, J.; Husmann, M.; Norrie, M.C. Exploiting mid-air gestures to share data among devices. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, Vienna, Austria, 4–7 September 2017; pp. 1–11.

36. Piumsomboon, T.; Clark, A.J.; Billinghurst, M.; Cockburn, A. User-defined gestures for augmented reality. In Proceedings of the CHI '13 Extended Abstracts on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; pp. 955–960.

37. Chen, Z.; Ma, X.; Peng, Z.; Zhou, Y.; Yao, M.; Ma, Z.; Wang, C.; Gao, Z.; Shen, M. User-Defined Gestures for Gestural Interaction: Extending from Hands to Other Body Parts. *Int. J. Hum. Comput. Interact.* **2018**, *34*, 238–250. [CrossRef]

38. Obaid, M.; Kistler, F.; Kasparavičiūtė, G.; Yantaç, A.E.; Fjeld, M. How would you gesture navigate a drone? A user-centered approach to control a drone. In Proceedings of the 20th International Academic Mindtrek Conference, Tampere, Finland, 17–18 October 2016; pp. 113–121.

39. Ruiz, J.; Vogel, D. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 3347–3350.

40. Guesgen, H.W.; Kessell, D. Gestural Control of Household Appliances for the Physically Impaired. In Proceedings of the Twenty-Fifth International FLAIRS Conference, Marco Island, FL, USA, 16 May 2012.

41. Koutsabasis, P.; Domouzis, C.K. Mid-Air Browsing and Selection in Image Collections. In Proceedings of the International Working Conference on Advanced Visual Interfaces, Bari, Italy, 7–10 June 2016; pp. 21–27.

42. Malizia, A.; Bellucci, A. The artificiality of natural user interfaces. *Commun. ACM* **2012**, *55*, 36–38. [CrossRef]

43. Fails, J.A.; Olsen, D.R. A Design Tool for Camera-based Interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Ft. Lauderdale, FL, USA, 5–10 April 2003.

44. Kim, J.-W.; Nam, T.-J. EventHurdle: Supporting designers' exploratory interaction prototyping with gesture-based sensors. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI '13, Paris, France, 27 April–2 May 2013; pp. 267–276.

45. Klemmer, S.; Landay, J. Toolkit Support for Integrating Physical and Digital Interactions. *Hum. Comput. Interact.* **2009**, *24*, 315–366. [CrossRef]

46. Maynes-Aminzade, D.; Winograd, T.; Igarashi, T. Eyepatch: Prototyping camera-based interaction through examples. In Proceedings of the 20th annual ACM symposium on User interface software and technology, Newport, RI, USA, 7–10 October 2007; pp. 33–42.

47. Albertini, N.; Brogni, A.; Olivito, R.; Taccola, E.; Caramiaux, B.; Gillies, M. Designing natural gesture interaction for archaeological data in immersive environments. *Virtual Archaeol. Rev.* **2017**, *8*, 12–21. [CrossRef]