Gesture Elicitation Studies for Mid-Air Interaction: A Review

Panagiotis Vogiatzidakis * and Panayiotis Koutsabasis [ID]

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

check for
updates

Abstract: Mid-air interaction involves touchless manipulations of digital content or remote devices, based on sensor tracking of body movements and gestures. There are no established, universal gesture vocabularies for mid-air interactions with digital content or remote devices based on sensor tracking of body movements and gestures. On the contrary, it is widely acknowledged that the identification of appropriate gestures depends on the context of use, thus the identification of mid-air gestures is an important design decision. The method of gesture elicitation is increasingly applied by designers to help them identify appropriate gesture sets for mid-air applications. This paper presents a review of elicitation studies in mid-air interaction based on a selected set of 47 papers published within 2011–2018. It reports on: (1) the application domains of mid-air interactions examined; (2) the level of technological maturity of systems at hand; (3) the gesture elicitation procedure and its variations; (4) the appropriateness criteria for a gesture; (5) participants number and profile; (6) user evaluation methods (of the gesture vocabulary); (7) data analysis and related metrics. This paper confirms that the elicitation method has been applied extensively but with variability and some ambiguity and discusses under-explored research questions and potential improvements of related research.

Keywords: elicitation method; mid-air interaction; gesture; review; survey

1. Introduction

Mid-air interaction is about touchless manipulations of digital content or remote devices, based on tracking of body movements, postures and gestures with non-intrusive sensors (or minimally-intrusive, mainly based on computer vision). Over the last few years, mid-air interaction has evolved into a distinguishable interaction style of human-computer interaction (HCI). The origins of mid-air interaction can be traced back in the late seventies in the MIT Media Room and the "Put That There demo" [1] and in the eighties at the live music performances of Vincent John Vincent and Francis MacDougall (https://www.youtube.com/watch?v=-zQ-2kb5nvs). In the last decade, we have witnessed popular gaming platforms that rest on mid-air interactions like the Wii and Xbox, impressive demos like the wearable SixthSense project [2] and several public installations in museums and technology-enhanced rooms [3]. Lately, mid-air interaction is explored as an alternative or complementary interaction style in several application domains that require touchless manipulation like mobile [4] and desktop micro gestures [5], gesture-based control of the TV [6] and other "smart" home appliances [7], remote interaction with distant displays in the wild [8] and in particular contexts (e.g., operating rooms) [9,10], interaction with smartwatches [11], in secondary driving tasks [12,13] and so forth.

There is not an established, universal gesture vocabulary for typical mid-air manipulations in any of these aforementioned application domains. On the contrary, it is widely acknowledged that "each input method is best at something and worse at something else" [14] and that "there is no

such thing as a universal gesture vocabulary for every application" [15]. Thus, the identification of appropriate gestures for mid-air user interactions in terms of criteria like discoverability, memorability, performance, reliability and comfort, is an important design decision. This must be made at the early stages of system development and it severely affects the development course of every mid-air application project as well as the user experience (UX) of intended users.

Emerging from the field of participatory design, gesture elicitation studies have been widely applied to help designers select the most appropriate gesture set for a given application. Although the method does not have a strict procedure, the main approach is to first define what operations (called "referents") have to be executed through gestures, then ask the end users to propose at least one gesture that they find preferable for each referent and finally to extract the gesture vocabulary after analysing the collected data [16]. The goal of a gesture elicitation study is to extract "good" gestures, which according to Morris et al. [17] are "gestures that meet certain design criteria such as discoverability, ease-of-performance, memorability, or reliability."

Over the last few years, several elicitation studies have been conducted to identify mid-air gesture vocabularies for many different application domains and contexts, devices, types of digital content and tracking technologies subsumed. A survey of elicitation studies of mid-air interaction is useful to HCI researchers and interaction designers for deepening the understanding of the many interrelated issues involved in their conduction, assessment and impact. This paper presents such a survey with the following research questions:

1.　What are the **application domains** of mid-air interaction elicitation studies? Given that mid-air interaction is researched in a wide application scope, it is useful to provide an overview of the diverse domains and possibly identify application trends.
2.　What is the level of **technological maturity of the systems at hand**? Since that the elicitation method is applied early in the requirements or design phases of the interactive systems development lifecycle, on which design ground do users make gesture proposals?
3.　What is the **basic process followed and its variations**? Given that the method does not have a strict procedure it is important to identify the main steps and their outcomes.
4.　What are **the dimensions of appropriateness** for gesture selection? Various dimensions are met in related work (like discoverability, memorability, performance, reliability, comfort, usability); it is useful to identify the degree to which these are considered in elicitation studies.
5.　What is the **profile of participants** in elicitation studies? As in every user-centred method, the number and the characteristics of participants are critical for the validity and generalizability of the results.
6.　How are **user proposals (about gestures) evaluated**? Various methods have been employed, it is useful to identify them as well as possible trends in this respect.
7.　What **data analysis** of user proposals is conducted and based on which **metrics**? Several metrics have been proposed and applied, with variation, in gesture elicitation studies that need to be reviewed and discussed.

The rest of the paper is structured as follows: Section 2 presents the paper selection process and criteria and the characteristics of the papers reviewed. Section 3 presents the findings of the survey in terms of the aforementioned research questions. Section 4 discusses the findings of this survey, identifying further challenges and trends. Section 5 presents the conclusions.

2. Method

A systematic method for paper selection and analysis was followed, which adopted the Quality of Reporting of Meta-Analyses (QUORUM) statement [18] and includes the following steps. The selection and analysis of papers took place in two periods: first in late May 2018 and late August 2018 (when the paper corpus was finalized).

Step 1. Potentially relevant publications identified and screened for retrieval

Source selection. We selected the following online digital libraries: ACM Digital Library, IEEEXplore, Taylor & Francis online, Springer (link), Elsevier (sciencedirect) and the Google scholar search engine. These online services provide direct access to the vast majority of high-quality journals and conferences of Human-Computer Interaction (HCI).

Search queries. At first, we explored search results with a few search terms that combined ("mid-air interaction" or "in-air interaction" or "touchless interaction" or "kinaesthetic interaction") and ("elicitation" or "user defined gestures" or "empirical study"). We observed that the most comprehensive lists of search results were obtained when we used two queries: ("mid-air interaction" and "elicitation study") and ("mid-air interaction" and "user defined gestures"). Therefore, we decided to retain these queries for all online sources.

Search constraints (refinement). We refined the results for the years 2011–2018 for all online sources. We set 2011 as the starting year because it was then that the first affordable sensors were released in the market. We shorted the results by "relevance". We approximately examined the first 100 search results of each online library (if available); since that we (gradually) observed that results after the first 100 were less relevant to our search. The online sources of ACM DL, Elsevier and Google scholar returned several hundred results on these queries, while the other three online sources returned only a few results. In addition, we used the feature "cited by", in cases of highly cited papers to identify more results that may have not been identified directly by the queries employed.

Criteria for screening. We screened more than 1000 search results (based on their title and summary) on the following criteria: (a) It potentially referred to an elicitation study of mid-air interaction (disambiguation was required since that some search results were clearly out of context), (b) it was a scientific paper (i.e., not a book, a thesis, an editorial, etc.), (b) it was accessible by the subscription from our academic institution (e.g., not a mere citation without a link to the source publication), (d) it was a relevant paper to this study, even if not an elicitation study per se (e.g., other review survey papers of mid-air interaction). A total number of 247 papers met those criteria.

Step 2. Publications retrieved for detailed evaluation

Further screening criteria. In this phase, the papers were reviewed by abstract and rapid examination of structure and content to: (a) exclude other relevant papers (identified in previous screening, phase (d)), (b) eliminate duplicates and (c) identify if they referred to an elicitation study, typically by containing a section on "user elicitation", or "gesture design" or "user definition of gestures". A total of 106 papers remained from this examination.

Step 3. Potentially appropriate publications for the review

Further screening criteria. In this phase, papers were cross-examined to (a) validate that each paper referred to a unique study (when more than one papers referred to the same study, the most comprehensive was kept), (b) exclude short papers and works in progress. Furthermore, papers were reviewed in further detail to ensure that they included reasonably sufficient information about the elicitation study. A total of 47 papers remained.

Step 4. Publications included for the review

General characteristics of the papers selected. The papers included in this survey span along the 5-year period like this (Table 1): 2011: 2 papers (4.3%); 2012: 5 papers (10.6%); 2013: 3 papers (6.4%); 2014: 6 papers (12.8%); 2015: 4 papers (8.5%); 2016: 8 papers (17%); 2017: 11 papers (23.4%); 2008: 8 papers (17%). Given that the paper selection was finalized in August 2018, papers to appear later in 2018 have not been included in this review.

Table 1. Number of papers examined within 2011–2018.

Year	Number of Papers	%
2011	2	4.3%
2012	5	10.6%
2013	3	6.4%
2014	6	12.8%
2015	4	8.5%
2016	8	17%
2017	11	23.4%
2018	8	17%
Total	**47**	**100%**

From the corpus of 47 publications (Table 2), there were 11 papers (23.4%) published in journals (in 10 journals) and 36 (76.6%) in conferences. A considerable number of papers selected (9 papers, 19%) were published in ACM CHI conference. Most papers were published in conferences rather than in journals, which may be attributed to the fact that the scope of elicitation studies is in user research and design. Papers on gesture elicitation do not often lead to development and evaluation of interactive systems. All publication venues are related to the field of HCI, most at its core, like the ACM CHI (Computer-Human Interaction), DIS (Designing Interactive Systems) and ITS (Interactive Tabletops and Surfaces) conferences and the International Journal of Human-Computer Studies and some in relation to applications of HCI like for example the Ubicomp conference.

Table 2. Citations of papers selected for this review and their publication venues.

Publication Type	Number	Citation
Journals (10 different journals)	11	[7,19–28]
Conferences	36	
CHI	9	[4,11,29–35]
DIS	5	[36–40]
ITS	4	[41–44]
Ubicomp	2	[45,46]
Other conferences (3DUIs, Academic Mindtrek, ACADIA, AVI, DAPI, Digital Home, DESRIST, Euro ITV, HCI Korea, ISS, IUI, MOBILEHCI, NordiCHI, Social Robotics, TEI, UAHCI)	16	[6,47–61]
Total	**47**	

3. Findings

In this section, we report on the findings based on the research questions of our review (posed in Section 1).

3.1. Application Domains

There is a considerable variety of application domains of mid-air interaction research, as well as variability in terms of the technological maturity of systems at hand (Table 3).

The control of various types of media, like music, videos, (types of) image collections and so forth, in the mid-air is a challenging design issue. A considerable number of elicitation studies (7/47, 14.9%) focus on mid-air media control, like the work of Ruiz et al. [34] who elicit a gesture set for controlling 360 degrees videos or the work of Siddhpuria et al. [40] who explore the use of discrete micro-gestures through a smartwatch for control of remote media.

Several studies employ mid-air interaction to improve the UX of (smart) TVs (6 out of 47, 12.8%), which have evolved beyond the passive TV-watching paradigm into interactive multimedia devices with features like web browsing, content manipulation, media playback and so forth. Notably,

in Reference [28] freehand gesture vocabularies for controlling the TV have been proposed, while in Reference [37] the gesture vocabulary is for blind users.

Mid-air control of public displays appeared in 12.8% of all studies examined (6/47). For example, Di Geronimo et al. [49] conduct an elicitation study to identify mid-air gestures to share data among mobile devices, Ref. [60] elicited user-defined gestures for defining virtual interaction spaces within a pervasive environment, while Rodriguez and Marquardt [44] present a gesture elicitation study on how to opt-in and opt-out from interactions with public displays to address the need for user registration and avoid "false-positive" content activation.

Generally, there are many other interesting domains of application for which elicitation studies have been conducted (Table 3) like mid-air manipulations in Virtual Reality [20,35,48,59], Smart Home environments [7,25,38,46] Mobile Devices [4,43,47,49] for Human-Robot/Drone manipulation [45,57,58], Augmented Reality [33,39], Desktop computer [30,50], In-Vehicle secondary driving task [52,56] Smartwatches [11,32], Gaming [19], CAD [53], Text readers [31], Operating rooms [23] and Digital exhibition [21].

Table 3. Domains and technological design maturity of applications employed in elicitation studies examined.

Domains	System	Working Prototype	Referents (Known Concept)	Total	%
Media	1	3	3	7	14.9%
TV	1	2	3	6	12.8%
Public Display	4	0	2	6	12.8%
Mobile Device	0	2	2	4	8.5%
Smart Home	0	1	3	4	8.5%
Virtual Reality	1	3	0	4	8.5%
Human-Robot/Drone Interaction	0	1	2	3	6.4%
Augmented Reality	0	1	1	2	4.3%
Desktop	0	0	2	2	4.3%
Secondary Driving Tasks	0	1	1	2	4.3%
Smart Watch	0	0	2	2	4.3%
Gaming	1	0	0	1	2.1%
CAD	0	0	1	1	2.1%
Text reading	0	1	0	1	2.1%
Operating room	0	1	0	1	2.1%
Digital exhibition	1	0	0	1	2.1%
Total	**9**	**16**	**22**	**47**	**100%**

3.2. Technological Maturity of Systems

Another issue that emerged from this survey is related to the technological development of systems at hand. Generally, elicitation studies are concerned with user research at the early stages of the design process, when it is not always possible or desirable to have a system or prototype. Therefore, the technological development of the system at hand may vary a lot, which affects the type of study.

We have identified three levels of technological design maturity:

Systems (fully-developed). These are employed in the gesture elicitation study (that may happen "in the wild") and they are possibly redesigned according to the results. For example, in the work of Lee et al. [46], an elicitation study was conducted with a "walk up and use" application about academic information installed on a public display in a university campus. Nine out of 47 elicitation studies (19.1%) were conducted with a fully developed system at hand.

Working prototypes. These are functional in the sense of providing interactive digital content and gesture tracking and demonstrating a working component of the system, however they are not fully-developed systems. For example, in the work of [54], the elicitation study takes place with reference to a working prototype of an image gallery. Sixteen out of 47 (34%) elicitation studies were conducted with working prototypes.

Referents. In this case, a set of referents (typically user commands) about a known system (e.g., the TV) is provided to users. For example, in the work of [19], three elicitation studies are conducted

to identify different gesture sets based on different body parts for intense gameplay. Almost half of the elicitation studies identified (22 out of 47, 46.8%) were conducted on the basis of referent sets about a known concept.

3.3. Gesture Elicitation Process and Variations

One of the most adopted user-elicitation methods is the "Guessability study" that was first introduced by Wobbrock et al. [62], as a unified approach for maximizing and evaluating the guessability of symbol input that was entered by users on a touchpad. This method was initially applied for interactive surfaces and later for mid-air gesture interactions. The main idea is that it is unrealistic to expect from novice, as well as expert users, to have the time or desire to undergo extensive training to learn new ways of interacting with a system.

In general, the guessability study starts by having participants presented with a referent (the effect of an action). Then, they are asked to propose a gesture that better matches (or is easy, intuitive etc.) the indented use. During the process, data collection takes place (video-audio recording, semi-structured interviews or think-aloud protocol, Likert scales questionnaires, etc.). At the end of the process, the gesture set is derived after data analysis with various quantitative and qualitative metrics and categorized into various taxonomies. Although the guessability study was initially applied for surface-based systems, it appears to be as the most adopted method (36 out of 47, 77%) for eliciting gestures for mid-air interaction systems (Table 4). Of the 47 papers reviewed, 9 papers (19.1%) followed the Wobbrock's methodology by the book (a. applying a Wizard-of-Oz, b. analysing proposed gestures with the metrics that Wobbrock et al. proposed (Level-of-agreement or Agreement-rate) and c. categorizing gestures into taxonomies), while 4 papers out of 47 (8.5%) enriched that methodology by applying additional metrics for analysing the gesture set.

Table 4. User-based Elicitation methods employed in studies examined.

Method	Number of Papers	%
Wobbrock's "Guessability Study"	35	74.5%
Choice-based elicitation Study	5	10.6%
Combination of Wobbrock's and Nielsen's method	3	6.4%
Nielsen's "Intuitive and Ergonomic method"	2	4.3%
Other	2	4.3%
Total	**47**	**100%**

An interesting variation of the usual guessability method is the choice-based elicitation study which consists of two phases. The first one is a guessability study. The second phase, which can also be considered as a refinement technique, re-examines only the referents that scored low agreement during the previous phase. A survey is conducted and users select the most appropriate gestures from a predefined list of gestures, produced by the experts or the users. For instance, Dim et al. [37], while investigating mid-air gestures that would enable blind people to control the TV, conducted two elicitation studies. In the first one, gestures proposed by users were analysed on a basis of consensus for each referent. In the second study (also referred as "choice-based elicitation study" [19,37]), for those referents that scored a low consensus rate, a pre-defined list of representative gestures (proposed by experts) was presented to participants. Similarly, in a research investigating concurrent body gestures during intense gameplay, Silpasuwanchai and Ren [19], conducted a choice-based elicitation study, since the notion of simultaneous gestures was relatively uncommon for users. In their study, the list of pre-defined gestures was not proposed by the authors but it was populated by the two, previously conducted, elicitation studies. A similar approach of user-defined (instead of expert-defined) choice-based elicitation study was also utilized in Dong's et al. [22] research, in which the most frequently proposed gestures from a preliminary study, were presented to the users in a multiple-choice manner. Although the choice-based elicitation study is a time-consuming process,

it is claimed as a necessary complementary approach that improves creativity, especially for novel interfaces where users are not familiar with the design space.

Another user-elicitation method is the one proposed by [15] also referred as the "Intuitive and Ergonomic" Method. Drawing upon usability principles, they highlight the importance of designing gestures that are easy to perform and remember, intuitive, ergonomic and metaphorically logical. They imply that mid-air gesture interaction is not a panacea for every application and therefore it should be examined beforehand whether it is the most appropriate interaction technique for the system to be developed. Then, in order to produce intuitive gestures for a system, they propose a human-to-human non-verbal communication approach with the use of scenarios where users interact with the "operator" (i.e., the person who conducts the experiment) or another user by using gestures that are more appropriate for the specific function. In cases when the design of the interface or the feedback is to be tested, a human-to-computer approach can be applied by using a Wizard-of-Oz technique [63]. After collecting the proposed gestures and evaluating their ergonomic characteristics, the resulting gesture vocabulary is benchmarked in terms of memorability, stress and guessability. From our review, it appears that two studies (2 of 47, 4%) adopted the general approach of Nielsen's "Intuitive and Ergonomic" Method. In particular, [23] conducted an intuitive and ergonomic method to investigate and compare gestures proposed by experts and novices using a vision-based Anaesthesia-related system within an operating room, while [26], adopted Nielsen's approach to elicit the set of commands of a music player and the gestures that are more appropriate for each task.

Also, there are two studies (4%) that conducted an extensive user-based elicitation study, by combining both Wobbrock's and Nielsen's approach. More specifically, in Reference [7], a user-based guessability study was conducted to find the most appropriate gesture set for Mid-air interactivity within a smart environment, followed by the Intuitive and Ergonomic method to investigate memorability and performance aspects of the proposed gestures. In their study [27], investigate gesture-based TV-control, by adopting Wobbrock's methodology to collect the user-defined gestures for every task and Nielsen's techniques to initially identify the available commands of the intended system and then benchmark the gesture set in terms of memorability, comfort and gesture-command matching degree.

3.3.1. Controlling the Legacy Bias

A primary concern on user elicitation studies is legacy bias, which refers to "the prior experience with interfaces and technologies, that makes it hard to uncover new gestures for an emerging medium" [17]. According to Morris at al. [16] users tend to transfer prior knowledge to new technologies because biased interaction techniques minimize physical and mental effort and because sometimes users cannot understand the fundamental capabilities of novel technologies. Undoubtedly, legacy bias has a direct effect on the proposed gestures. Some researchers mention legacy bias as a pitfall of elicitation studies since it might limit the potential for producing gestures that take full advantage of emerging technologies and sensing capabilities [16,17].

Morris at al. [16] proposed three techniques (the 3-Ps) that might reduce the phenomenon of legacy bias; Priming, Production and Partners. Priming involves exposing users to a stimulus that sub-consciously influences the responses to future stimuli. 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. From our survey, 15 out of 47 (31.9%) studies adopted at least one of the aforementioned techniques to reduce legacy bias, while Production was the most frequent technique used (12 out of 15, 80%). For example, in Reference [51], users were kinaesthetically primed by lifting and moving boxes before the elicitation study, in References [54,55] users were prompted to produce more than 3 gestures for each referent and in Reference [41] users were recruited in groups to brainstorm several interactions and then to come up with the most preferable one.

Although legacy bias is considered as a factor that may not produce originality in gesture proposals, sometimes it is considered to have positive effects in elicitation studies. Köpsel and Bubalo [64] argue that biased gestures have, in most cases, the advantage of being simplistic, do not

require much time to be learned, or effort to be guessed, resulting in high agreement scores in elicitation studies. Such gestures are appropriate in cases when users do not have the time or desire to learn new interaction methods, or when the user's cognitive load should not be burdened. It becomes apparent that tackling or not the legacy bias is a matter of design decision and it mainly depends on whether the end product/system is meant to be a walk-up-and-use system or a system that would take full advantage of the novel interaction techniques. In our survey, 32 out of 47 studies (68.1%), did not utilize a technique to reduce legacy bias.

3.3.2. Referent Presentation

Following Wobbrock's terminology, the referent is the effect which is triggered by a gesture. Referents can be presented to the participants in various ways. Depending on the type and maturity of the prototype used, referents were either demonstrated through GUI animations [41,54], described as a text message on the screen, or verbally [4,28,37,61], presented as a video [51,61] or still images [19,24,34,55], or presented by manipulating the actual artefact [7,45].

3.3.3. Think-Aloud Protocol

Although the main method of elicitation studies are similar to Wobbrock's [62,65] or Nielsen's [15] approach, various complementary techniques were employed during the stage of user participation and gesture proposals. The most apparent one is the think-aloud protocol [66,67], which involves prompting participants to verbalize their thoughts, such as why they chose the proposed gesture or refer to other systems or previous experiences [54], as well as describing the gesture, especially the beginning and the end of it. The benefit here is two-fold, it helps the observer to understand the gesture delimiters [68] and it gives valuable insight on the mental models of users [69]. In our survey, 22 out of 47 (46.8%) studies implemented a think-aloud technique.

3.3.4. Wizard of Oz

The Wizard of Oz (WOz) method [70,71] has been claimed as a useful user-centred inquiry approach for novel interface researches when the design space is unknown or under investigation [69]. In the traditional WOz method, the participant has the impression that she is directly communicating with the system. In fact, this is done by an expert (the Wizard) who, in most cases, is hidden. The main process of the WOz method is, first to inform/present the participant for the task to be done, then allow her to start gesturing with the gestures that she prefers (with no expert intervention) and then present the effect of the gesture, giving the impression that the interaction is direct.

However, from the studies reviewed, although the term Wizard of Oz was highly referenced, the approach adopted was not the traditional one but rather a "reversed Wizard of Oz" process. In most studies, the effect (referent) was first presented to the participant, who then was prompted to suggest an appropriate gesture. Moreover, the expert was not hidden to the participant, suggesting that there was no significant interest in eluding the users for a direct human-to-system communication. In general, the Wizard of Oz technique that was employed in most studies (45 out of 47, 95.7%) was actually about the concept of having an expert to control the system and present the referent (to help participants better understand the task in question) and then ask the user to suggest a gesture.

3.4. On Gesture "Appropriateness"

The main goal of an elicitation study is to elicit appropriate gestures for mid-air interactions but how is appropriateness interpreted in elicitation studies of mid-air interaction?

A deeper look into the meaning of "appropriateness" in the papers examined (Table 5), reveals that a considerable number of studies (14 out of 47, 29.8%) investigate a gesture set that is a "better match" or "fit for purpose" for its intended use (without analysing this into a more specific meaning). For example, elicitation studies were conducted to elicit appropriate gesture sets for controlling a media player [54]

or for 3D travelling within a pseudo-universe [59], with the aim improve the user interaction and experience in general.

Table 5. Criteria of gesture selection

Criteria of Gesture Selection	Number of Studies	%
"Better match" (general)	14	29.8%
User's mental model		
Intuitive-Natural gestures	11	23.4%
Memorability	7	14.9%
Discoverability	3	6.4%
Learnability	3	6.4%
Other (Social acceptability, Guessability, Distinctness)	5	10.6%
Gesture's ergonomic characteristics		
Easy to perform	12	25.5%
Comfort	6	12.8%
Low fatigue	4	8.5%
Simple gestures	3	6.4%
Other (Number of hands, Concurrent gestures, Suitability of body parts)	5	10.6%
None	7	14.9%

Many elicitation studies focus on finding gestures that are easy to perform (12 out of 47, 25.5%). For example, in their study Ruiz et al. [4], were interested in gaining user insights about the ease of gesture application, after asking them to repeat each gesture application (five times).

An equivalent number of studies investigate whether gestures are intuitive or natural (11 out of 47, 23.4%) that is, gestures that are intuitive or natural in the sense that they "enable users to use the interface with little or no instructions" [72]. For instance, in their study Jahani et al. [52], highlighted the importance to find a set of mid-air gestures for in-vehicle secondary tasks, that are natural and intuitive to drivers, in order to avoid increasing their cognitive-load.

Another considerable number of studies (7/47, 14.9%) examined the memorability of proposed gestures which reflects on *"how easy users can recall the gesture set after some time of inactivity"* [72]. For example, Bostan et al. [36] investigated hand-specific on-skin gestures, showing that intuitive gestures were easily memorable, while Kühnel et al. [7], investigated the correlation of memorability with gesture suitability and perceived effort.

Many other dimensions of the appropriateness of a mid-air gesture were examined, such as comfort [11,23,27,36,44,60], perceived fatigue [24,36,40,50], discoverability [34,55,58] learnability (how easy gestures can be learned) [24,27,50], gesture simplicity [26,45,61], body parts suitability [19,73], concurrent gestures during intense gameplay [19] and unimanual or bimanual gestures [48,53]. Finally, a few studies (7 out of 47, 14.9%) did not mentioned any specific focus of the gestures proposed.

Most of the studies reviewed are driven by at least one dimension of appropriateness. These dimensions may be regarded as belonging into two more general categories: the *mental model of the users* (i.e., memorability, intuitiveness, discoverability, learnability and guessability) and the *ergonomic characteristics of the gestures* (ease of application, fatigue, simplicity, comfort, number of hand/body parts and concurrent gestures). Each dimension of gestural interaction is investigated with various empirical methods and techniques as shown in Section 3.6.

3.5. Participants: Number and Profile

In a user-centred approach, it is important to carefully select the participants, so they represent the user population adequately and appropriately. Participant selection (or recruitment) must address the questions of "how many participants are enough?" (sample size) and "what participant profiles are representative of the population?" The latter is based on qualitative criteria or possible previous analysis (i.e., user segmentation, personas, etc.).

The number of participants in the elicitation studies examined is shown in Table 6. Two studies employed only four participants [47,55] and another one employed 9 participants [26], while four studies employed from 35 to 89 users [22,29,53,57]. All other studies recruited between 10 to 30 participants. Generally, there was not much discussion about the required number of participants employed in gesture elicitation. However, the validity of the outcomes and recommendations of elicitation studies is significantly affected by the number of users employed.

Table 6. Number of participants in gesture elicitation studies.

No. of Participants	Number of Studies	%
≤10	3	6.4%
11–20	22	46.8%
21–30	18	38.3%
>30	4	8.5%
Total	**47**	**100.0%**

Regarding the participant profile (Table 7), we saw that about one-half of participant groups were drawn from the academic environment, that is, they were either students (23.4%) or a mix of students and researchers (academic staff 23.4%). These are often more accessible than other types of users and they often volunteer to participate in user-centred activities, to learn about the method or the technology, especially if they are rewarded. However, they may not always be representative of the user population; unfortunately, there was not much discussion about the representativeness, of participants employed in gesture elicitation studies.

Table 7. Participant profile in gesture elicitation studies.

Participant Profile	Number of Studies	%
Adults *	23	48.9%
Students	11	23.4%
Academic staff	11	23.4%
People with special needs	2	4.3%
Total	**47**	**100.0%**

(* One study also involved children).

The other half of participants (48.9%) were adult users of various characteristics. In general, participants were aged between 18 and 60 years old. Only in two works, the age range of the participants was wider, including elderly people [37] as well as a few children (an in-the-wild study) [41]. Regarding prior experience with mid-air gesture interaction, 17 of the 47 (36.2%) studies employed users with mixed experience to the technology examined. Another 9 studies (19.1%) exclusively employed experienced and 7 studies (14.9%) non-experienced participants, depending on whether they wished to control the legacy bias effect. Notably, 14 (29.8%) studies did not provide sufficient information to the aspect or prior user experience. Finally, 8 out of 47 studies (17%) applied gender balancing among the participant group.

3.6. User Evaluation of Proposed Gestures

An important aspect of an elicitation study is the user evaluation of proposed gestures according to a possible number of criteria, scales, as well as qualitative comments and remarks (Table 8).

Most elicitation studies (42 out of 47, 89.3%) include one or more user evaluation methods for gesture proposals (23 of the 42, include two or three methods). Generally, user evaluation methods may occur during the production of gestures (concurrently), and/or after the production of a single gesture ("post-task methods" of Table 8) and/or after the end of production of all gestures ("post-test methods").

More than one-third of the elicitation studies examined involve concurrent user evaluation. This was largely conducted with the think-aloud protocol (46.8%), in which participants are encouraged to

speak out their thoughts, feelings and opinions. A single study [44], placed users in pairs to discuss about their gesture productions and proposals, therefore adopting an approach co-discovery learning; this is an interesting approach since that researchers can concurrently observe the rationale of participant proposals, while participants themselves may be stimulated to argue for more during the production of a gesture.

More than half studies conduct user evaluations just after the production of a single gesture, repeatedly (post-task). In particular, 25/47 (53.2%) studies adopt a post-task rating scale for the gesture produced; this includes one or more questions about the appropriateness of the gesture from the user perspective (e.g., ease of use, fit for the task, etc.). An additional pair of studies (2/47) adopt post-task (short) interviews instead. One study adopted a post-test memorability test.

Table 8. User evaluation methods of proposed gestures in elicitation studies examined (possibly more than one method per study).

User Evaluation Method	Number of Methods	% in 47 Papers
Post-task rating scale	25	53.2%
Think-aloud (concurrent)	22	46.8%
Post-test interview	10	21.3%
Post-test questionnaire	7	14.9%
Post-task interview	2	4.3%
Post-test survey	2	4.3%
Usability test	2	4.3%
Post-test memorability test	1	2.1%
None	5	10.6%
Total	**62**	**100.0%**

Another significant number of studies (19/47, 40.4%) conduct user evaluation at the end of the elicitation procedure with various methods. Post-test interviews were employed in 21.3% of elicitation studies. A questionnaire was adopted in 14.9% of studies examined, which was often wider in scope than Likert scales, including questions about user opinions that extend gesture production issues. Post-test surveys were adopted in a couple of studies [52], which were online, confirmatory in nature and involved other participants than those of the elicitation study.

Finally, there are two studies [49,54] that proceeded to the technical development of gestures in a prototype and conducted usability evaluations of alternate gesture sets. These usability tests included typical usability metrics (time to task, errors) as well as measures of perceived usability (questionnaires) and fatigue.

3.7. Data Analysis and Metrics

The data analysis is conducted by the researchers and takes place after the user participation has ended. It involves gathering, organizing and coding the recorded data from the user-elicitation study, extracting the most appropriate gesture set, after being processed with various metrics and in many cases categorizing gestures into taxonomies.

3.7.1. Metrics

A significantly large number of studies (Table 9, 34 out of 47, 72.3%) utilized at least one metric to extract the gestures that better match to a specific referent, or to understand the conceptual complexity of a referent. The most apparent metric (30 out of 34, 88.2%), especially for those studies that follow Wobbrock's guessability study [65], is the "Level of agreement", which shows the level of consensus among participants for each referent. The Level of agreement was initially introduced by Wobbrock in 2005 [62] and was refined by Vatavu and Wobbrock in 2015 [74] which is called "Agreement Rate". Even though Agreement rate was claimed as an improved version of the Level of agreement, only half of the papers that were published after 2015 (8 out of 15, 53.3%) have utilized it [24,28,30,31,38–40,44].

Apart from Level-of-Agreement and Agreement-Rate (30 out of 47, 63.8%) the next most frequent metrics used were time-related (10 out of 47, 21.2%), which are the "Time-of-Thinking" and the "Time-of-Gesture-articulation". Time-of-thinking is the time the participant needs to think before defining a gesture and after the referent has been presented to her. In their research, Hoff et al. [51] utilized gesture thinking time to examine whether priming has a positive or negative effect in gesture proposal. Dim et al. [37] also used the thinking time as an indicator of how easily blind people can imagine their gestures. Kühnel et al. [7], examined the correlation of referents' conceptual complexity with the thinking time, showing that the longer the time needed, the higher the referent's complexity. Gesture articulation time is the time the user needs to perform a gesture. According to their findings Kühnel et al. [7], the gesture articulation time affects negatively the rate of easiness of performance.

Table 9. Metrics used in elicitation studies (possibly more than one per study).

Metric	Number	% (in 47 Studies)
Level of agreement A(r)	22	46.8%
Agreement Rate AR(r)	8	17%
Time of thinking (before proposing the gesture)	6	12.8%
Time of gesture articulation	4	8.5%
Consensus distinct	2	4.3%
Max-consensus	1	2.1%
Other	7	14.9%
None	13	27.7%

3.7.2. Gesture Taxonomy

Almost half of the papers reviewed (22 out of 47, 46.8%), analysed various characteristics of the proposed gestures by classifying them into different categories, called taxonomies. Taxonomies help researchers to gain some insights about the mental model of users [65] and guide designers to understand the type of gestures that are appropriate for various referents [40].

The most frequent taxonomy type is the Nature (Table 10, 13 out of 47, 27.7%) of the gesture which denotes the relationship between gesture and meaning/object [37] and has various dimensions such as symbolic, metaphorical, abstract, physical and deictic. Symbolic gestures are depictions of symbols, such as drawing a question mark in the air. Metaphorical gestures are linked to their meanings (not to their visual similarities), while abstract gestures map the interactive task arbitrarily. Physical gestures manipulate directly the content/object, such as scaling or rotating an object. Deictic gestures usually involve a stretched index finger, a palm or multiple fingers to indicate objects and directions.

Table 10. Types of taxonomies used in papers reviewed.

Taxonomy	Number of Papers	% in 47 Papers
Nature	13	27.7%
Symbolic	13	27.7%
Metaphorical	13	27.7%
Abstract	12	25.5%
Physical	12	25.5%
Deictic	1	2.1%
Body-part, number of hands	12	25.5%
Unimanual/Bimanual	12	25.5%
Full body	4	8.5%
Form	11	23.4%
Static/Dynamic	11	23.4%
Static pose with path	5	10.6%
Flow-Temporal	9	19.1%
Continuous–Discrete	9	19.1%
Other	19	40.4%

The number of hands (unimanual or bimanual), or body parts involved to a gesture was a frequent categorization type of the gestures proposed, in a significant amount of papers (12 out of 22, 25.5%). An equivalent number of studies (11 out of 22, 23.4%) also classified gestures in a taxonomy type called Form, which includes Static gestures (postures that does not vary over time), Dynamic gestures (gestures that involve body or hand movement) and Static gestures with path (gestures that involve hand movement while the hand pose remains the same). Gesture flow was another taxonomy frequently used (9 out of 22, 19.1%), which distinguish gestures into continuous and discrete describing whether the referent occurs during the gesture or after it respectively.

In a few studies, there has been an attempt to derive some design conclusions by contrasting the results from the metrics analysis with the taxonomy of the gesture vocabulary. For example, physical gestures were proved to have higher agreement rates, while abstract gestures require longer time to propose and articulate [7]. Although gesture taxonomization helps researchers to better understand the users' mental model, it was often the case that this categorization did not yield design guidance. Therefore, it appears that there is a need for more practical guidelines in this respect, for example regarding the number of body parts/joints employed for a gesture and respective estimations of stress or fatigue.

4. Discussion

In this section, we discuss our findings with respect to the current and future practice of conducting elicitation studies for mid-air interaction.

4.1. Variability of Application Domains and Systems' Technological Maturity

Our survey has identified that the gesture elicitation method is evolving into standard practice of a user-centred approach to mid-air interaction design. Its adoption is growing by researchers in various mid-air interaction scenarios involving interactive technologies that range from TV control to interaction with smartwatches or drones and it may also concern applications or services provided through these technologies.

It is interesting and fruitful that a user-centred design method is applied to novel interaction contexts and can reach to suggestions about design directions. However, it is equally important to capitalize on the results of elicitation studies in order to identify some prevailing gestures or gesture sets for basic mid-air interactions in any of these domains. This requires careful **reviews and assessments of the content of design suggestions for particular interactive technologies or contexts of use.** For example, in the work of [75], a systematic survey of mid-air hand gestures with interactive surfaces and displays is presented. More content-based surveys of this sort can help the community to summarize and reflect on previous results and provide feedthrough and inspiration for analogous design contexts.

This survey identifies that there is considerable variability of the technological design of the system under investigation. In many studies, there is not a prototype or system at hand but the referents are presented verbally or with cards (about user commands), which presupposes that the users have a good mental model about the referent. Other studies present prototypes to users or apply the WOz approach, which offers an orientation of the technological context but might 'contaminate' the gesture production process with legacy bias. Of course, the form of the referent mediates the response (production of gestures) and therefore the validity and quality of results. **Further research is required to identify the ways by which the form and media of the stimuli affect the responses in elicitation studies**.

4.2. On the Steps and Process of an Elicitation Study

Analysis of elicitation studies conducted in the papers reviewed, revealed some similarities in the process itself and the steps taken, as well as some variations. There are different approaches to

the methods proposed by Wobbrock, Nielsen and choice-based elicitation, which were evident in the studies reviewed.

In particular, in Nielsen's methodology, includes a step called "*Find the functions*" [15], from the pre-development stage in which users suggest the functions needed by the application with the help of scenarios. In the next step (*Collect the gestures*), a human-to-human non-verbal communication approach is adopted, where users, with the help of scenarios, are given the functions and are asked to find the matching gestures. A user-evaluation of the proposed gestures in terms of memorability, guessability and stress, is conducted during the final step.

Wobbrock's "*Guessability Study*" includes two stages. The purpose of the first stage is to collect gestures from users by showing them the referent (the effect of a gesture) while asking them to propose appropriate gestures. The second stage involves analysing the data collected using metrics to measure the consensus level among users for each referent and is conducted by the researchers.

A choice-based elicitation study can be considered as an enhanced variation of the guessability study. It consists of four stages and the first two are similar to the guessability study. The difference in this method is that for those referents that have scored low consensus level, a second round of investigation is conducted. So, in the next stage, experts create a list of gestures that are more appropriate for each referent (of those with low consensus level). That predefined list is then presented to the users in order to select the gesture that better match the corresponding referent. In a choice-based elicitation study, user participation is necessary during the first and the last step.

In general, the design and conduction of elicitation studies vary on the assumptions and research goals related to envisaged contexts of use. It seems that the strong points of the Wobbrock's approach are that it is a simple and practical procedure accompanied by a solid mathematical groundwork on how to analyse gestures before committing them to the gesture set. Nielsen's approach focusses on user-evaluation of the resulting gesture set considering several criteria beyond guessability in this method, user participation is essential in all the steps of the study [69]. Last but not least, the choice-based elicitation method can provide further validation of intermediate gesture proposals with the conduction of surveys in additional user groups.

All three variations of the elicitation method are to be conducted in a controlled environment like a classroom or a computer lab and with an instrumental (task-based) procedure. Of course, controlled studies have particular advantages for research but they inherently ignore contextual factors like other people's presence, environmental conditions variability (e.g., lighting, noise, etc.) and they may bypass important aspects of gesture appropriateness like gesture variability [76], which can be measured and captured automatically. There is recent work on gesture elicitation in more authentic contexts ("in-the-wild"), without a task-based procedure, like in the work of [77]. **Further work can compare the results of gesture elicitation between lab and field studies, as well as adopt elicitation method for in-the-wild settings**. This must also consider recent developments of agreement metrics in between-subjects elicitation studies [74].

4.3. On the Number and Profile of Participants in Gesture Elicitation Studies

Throughout our survey, we have found a scarce discussion on the criteria of participant selection and recruitment in elicitation studies. In many studies, we saw relatively homogeneous groups of users, like students or researchers, despite the wide appeal of the domains of applications examined. This may harm the validity of design suggestions, since it is plausible to assume that agreement scores of gesture proposals would differ (significantly) between diverse user groups of mid-air interaction applications. In a similar vein, there was not much discussion about the number of participants required in elicitation studies.

Lessons from other user-centred methods like usability and card-sorting studies indicate that there is not a "magic number" of minimum users for every study but a reasonably small range of carefully selected participants can yield wide-in-scope results and recommendations in particular contexts, possibly in a user-centred approach which can include repeated studies. More specifically,

there has been a long held discussion about the minimum number of users in usability tests with more recent opinions in agreement to that "*if you are interested in identifying major usability issues as part of an iterative design process, you can get useful feedback from three or four representative participants . . . as the design gets closer to completion, you need more participants*" [78]. In the context of card-sorting studies, according to [79], "*reasonable structures are obtained from 20–30 participants.*" Therefore, further research is required in this respect, which can provide **methodological guidance on how many users are required for elicitation study**, which and can help the designers and practitioners to better validate the results of their studies.

4.4. On the Dimensions of Gesture Appropriateness, User Evaluation and Data Analysis

Our survey identifies a number of dimensions considered by researchers in search of appropriate gestures for mid-air interaction, with most frequent those of "ease of application", "intuitive", "fit for purpose". Some studies are more exploratory, attempting to identify these dimensions during the elicitation process, for example, with the think-aloud protocol. In most studies, researchers are making use of self-developed Likert scales about those dimensions (either post-task or post-test). In a few studies, some standardized questionnaires have been employed, like the NASA-TLX (perceived mental and physical effort) [80]. Thus, the perceived appropriateness of a gesture is a source of differentiation in elicitation studies. Further work is needed towards the proposal and validation of **an instrument of post-task user assessment of the dimensions of gesture appropriateness**. Additionally, these user estimations about dimensions of gesture appropriateness might be taken into account into the calculation of agreement scores.

As expected, the level of agreement on gesture proposals is the metric most often used in elicitation studies. However, we have identified that some studies attempt to go further than agreement scores and assess gesture appropriateness on other grounds, like measured usability [54] and measured physiological risk [24]. In addition, other measures of fatigue that may be integrated to elicitation studies like consumed endurance [81] and the distance measured by hands [82], which can be automatically calculated provided an interactive system with gesture sensing capabilities is in-place. Thus, another area of further work is to **further validate the results of gesture elicitation with technical tests about measured usability and fatigue**. There are some works in this respect that need to be combined with elicitation studies like the work on consumed endurance [81].

4.5. On the Results of Elicitation Studies: Implications for Design

The typical outcome of an elicitation study is a (set of) gesture(s) for each referent (operation or user command), based on user agreement rates (or other metrics). Many elicitation studies have produced tables with listings of gestures for referents in the aforementioned domains of mid-air interaction. We need to ask ourselves is this is sufficient information for a designer or a developer to carry out detailed design and system implementation.

For example, a short description of a gesture (e.g., To swipe) does not specify important details of the gesture, such as: if it is performed with fingers or the hand, what human joints participate in the gesture and need to be monitored by the sensor, what is the time duration of the gesture and so on. A few studies have identified such factors, like the work of Riener et al. [13], who investigate the *interaction space*, that is, the physical 3D space that is available or preferable for a driver to apply mid-air gestures for secondary driving tasks. Therefore, an area of further work is to develop a **protocol for reporting results of mid-air elicitation studies that specifies detailed design information**.

Another related issue is whether user preferred gestures *at design-time* are indeed the most usable at the end of the process. There are some studies that indicate otherwise, such as in Koutsabasis and Domouzis [54], who proceed to an implementation of alternate mid-air gestures (produced from an elicitation study) and test their usability with many metrics: task time, errors, perceived usability, perceived effort and so forth, They conclude that the most usable and more preferred gesture for the manipulation of image collections (hand sideways extension) was different than the gesture originally

preferred in the user elicitation study (swipe). Notably, the same users participated in both studies. This is an interesting result that indicates that the context of user production of gesture proposals is very important and should be carefully prepared so that it is realistic. Further to that and despite that Morris et al. [83] have shown the benefits of elicited gestures to designed gestures, **additional factors (notably those related to performance, actual fatigue and usability) affect the user acceptance at the system implementation level.** To investigate these factors at design time is a big challenge for future elicitation studies.

4.6. Limitations of This Survey

The aim of this paper was to provide a review on the elicitation studies in mid-air interaction design. As with any survey, our approach has limitations. Our survey is a process-based analysis, focusing on the constituting elements and phases of the application of the method with a breadth of studies examined in terms of their domain of application. As a consequence, the discussion of the content and results (gesture proposals) of elicitation studies has been brief.

In addition, we have reviewed a sample of elicitation studies that was determined by the query method employed and the criteria of selection that have inevitably constrained the sample into ways that may not be easily assessed at the time of writing. For example, one of the criteria for selection of papers was to focus on full-papers, which might have left some high-quality short papers outside the corpus examined.

This review is limited to mid-air gesture elicitation studies alone. These are a large corpus of elicitation studies but there are also other gesture types, such as (multi-)touch, whole-body, user-defined gesture input for wearables, that have been investigated with this methodology. We did not broaden the scope of our review to these domains for reasons of motivation and also because this would lead to an inflation of surveys. Furthermore, most of these studies rest on referents rather than on working systems or prototypes.

5. Conclusions

This paper presented a survey of elicitation studies in mid-air interaction design. The survey is systematic in the sense that it followed an analytical approach to the selection and examination of related papers. It is critical to the extent that it discusses several issues and possible shortcomings of the elicitation studies identified, as well as it identifies a number of directions for further work. We envisage that this survey can contribute to a better understanding of elicitation studies in current and future mid-air interaction scenarios and applications and that researchers and practitioners in mid-air interaction design will be stimulated by the facts and ideas presented in this survey, reflect on the issues identified, enrich their knowledge of the state-of-the-art on conducting elicitation studies and possibly re-think and improve their own work and practice.

Author Contributions: Both P.V. and P.K. conceived the presented idea, contributed to the planning, design and writing of the paper. P.V. analyzed the data from the papers reviewed and focused on the Methods and Findings sections of the manuscript. P.K. focused on the Introduction, Method and Discussion sections while he supervised the project.

References

1. Bolt, R.A. "Put-That-There": Voice and Gesture at the Graphics Interface. 1980. Available online: http://www.paulmckevitt.com/cre333/papers/putthatthere.pdf (accessed on 8 September 2018).
2. Mistry, P.; Maes, P. SixthSense: A wearable gestural interface. In Proceedings of the ACM SIGGRAPH ASIA 2009 Sketches, Yokohama, Japan, 16–19 December 2009.
3. Koutsabasis, P.; Vosinakis, S. Kinesthetic interactions in museums: Conveying cultural heritage by making use of ancient tools and (re-) constructing artworks. *Virtual Real.* **2017**, *22*, 103–118. [CrossRef]

4. 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.

5. Wacharamanotham, C.; Todi, K.; Pye, M.; Borchers, J. Understanding finger input above desktop devices. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014.

6. Wu, H.; Wang, J. User-Defined Body Gestures for TV-based Applications. In Proceedings of the 4th International Conference on Digital Home (ICDH), Guangzhou, China, 23–25 November 2012.

7. 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]

8. Gentile, V.; Malizia, A.; Sorce, S.; Gentile, A. Designing Touchless Gestural Interactions for Public Displays In-the-Wild. In *Interaction: Interaction Technologies, Proceedings of the 17th International Conference on Human-Computer Interaction, Los Angeles, CA, USA, 2–7 August 2015*; Springer: Cham, Switzerland, 2015.

9. O'Hara, K.; Gonzalez, G.; Sellen, A.; Penney, G.; Varnavas, A.; Mentis, H.; Criminisi, A.; Corish, R.; Rouncefield, M.; Dastur, N.; et al. Touchless interaction in surgery. *Commun. ACM* **2014**, *57*, 70–77. [CrossRef]

10. Tan, J.H.; Chao, C.; Zawaideh, M.; Roberts, A.C. Informatics in radiology: Developing a touchless user interface for intraoperative image control during interventional radiology procedures. *Radiographics* **2013**, *33*, 61–70. [CrossRef] [PubMed]

11. 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.

12. Ohn-Bar, E.; Trivedi, M.M. Hand Gesture Recognition in Real Time for Automotive Interfaces: A Multimodal Vision-Based Approach and Evaluations. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 2368–2377. [CrossRef]

13. 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.

14. Kinect, U. Kinect Human Interface Guidelines. 2014. Available online: http://download. microsoft.com/download/6/7/6/676611B4-1982-47A4-A42E-4CF84E1095A8/KinectHIG.2.0.pdf (accessed on 8 September 2018).

15. Nielsen, M.; Störring, 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, Proceedings of the 5th International Gesture Workshop, GW, Genova, Italy, 15–17 April 2003*; Springer: Berlin/Heidelberg, Germany, 2003.

16. Ruiz, J.; Vogel, D. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-Body Gestures with Low Arm Fatigue. 2015. Available online: https://hci.cs.colostate.edu/files/6614/3387/9908/paper2504.pdf (accessed on 8 September 2018).

17. Morris, M.R.; Danielescu, A.; Drucker, S.; Fisher, D.; Lee, B.; Schraefel, C.; Wobbrock, J.O. Reducing legacy bias in gesture elicitation studies. *Interactions* **2014**, *21*, 40–45. [CrossRef]

18. Moher, D.; Eastwood, S.; Olkin, Y.; Rennie, D.; Stroup, D.F. Stroup Improving the quality of reports of meta-analyses of randomized controlled trials: The QUOROM statement. *Oncol. Res. Treat.* **2000**, *23*, 597–602. [CrossRef] [PubMed]

19. Silpasuwanchai, C.; Ren, X. Designing concurrent full-body gestures for intense gameplay. *Int. J. Hum. Comput. Stud.* **2015**, *80*, 1–13. [CrossRef]

20. Jahani, H.; Kavakli, M. Exploring a user-defined gesture vocabulary for descriptive mid-air interactions. *Cogn. Technol. Work.* **2017**. [CrossRef]

21. Manghisi, V.M.; Uva, A.E.; Fiorentino, M.; Gattullo, M.; Boccaccio, A.; Monno, G. Enhancing user engagement through the user centric design of a mid-air gesture-based interface for the navigation of virtual-tours in cultural heritage expositions. *J. Cult. Herit.* **2018**, *32*, 186–197. [CrossRef]

22. 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]

23. Jurewicz, K.A.; Neyens, D.M.; Catchpole, K.; Reeves, S.T. Developing a 3D Gestural Interface for Anesthesia-Related Human-Computer Interaction Tasks Using Both Experts and Novices. *J. Hum. Factors* **2018**. [CrossRef] [PubMed]

24. 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.* **2017**. [CrossRef]

25. Radu-Daniel, V. A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. *J. Ambient. Intell. Smart. Environ.* **2013**, *5*, 187–211. [CrossRef]

26. 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. *Multimedia Syst.* **2011**, *18*, 15–31. [CrossRef]

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

28. Zaiṭi, 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]

29. 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.

30. 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.

31. Dingler, T.; Rzayev, R.; Shirazi, A.S.; Henze, N. Designing Consistent Gestures Across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018.

32. Malu, M.; Chundury, P.; Findlater, L. Exploring Accessible Smartwatch Interactions for People with Upper Body Motor Impairments. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18), Montreal, QC, Canada, 21–26 April 2018.

33. Piumsomboon, T.; Clark, A.J.; Billinghurst, M.; Cockburn, A. User-defined gestures for augmented reality. In *Human-Computer Interaction—INTERACT, Proceedings of 14th IFIP TC 13 International Conference, Cape Town, South Africa, 2–6 September 2013*; Springer: Berlin, Gremany, 2013; pp. 282–299.

34. 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.

35. Yan, Y.; Yu, C.; Ma, X.; Yi, X.; Sun, K.; Shi, Y. VirtualGrasp: Leveraging Experience of Interacting with Physical Objects to Facilitate Digital Object Retrieval. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–13.

36. 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.

37. 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.

38. Gheran, B.-F.; Vanderdonckt, J.; Vatavu, R.-D. Gestures for Smart Rings: Empirical Results, Insights, and Design Implications. In Proceedings of the 2018 on Designing Interactive Systems Conference, Hong Kong, China, 9–13 June 2018; pp. 623–635.

39. Pham, T.; Vermeulen, J.; Tang, A.; MacDonald Vermeulen, L. Scale Impacts Elicited Gestures for Manipulating Holograms: Implications for AR Gesture Design. In Proceedings of the 2018 on Designing Interactive Systems Conference 2018, Hong Kong, China, 9–13 June 2018; pp. 227–240.

40. 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.

41. Morris, M.R. Web on the wall: Insights from a multimodal interaction elicitation study. In Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, Cambridge, MA, USA, 11–14 November 2012.

42. Nebeling, M.; Ott, D.; Norrie, M.C. Kinect analysis: A system for recording, analysing and sharing multimodal interaction elicitation studies. In Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems, Duisburg, Germany, 23–26 June 2015; pp. 142–151.

43. Pyryeskin, D.; Hancock, M.; Hoey, J. Comparing elicited gestures to designer-created gestures for selection above a multitouch surface. In Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, Cambridge, MA, USA, 11–14 November 2012; pp. 1–10.

44. 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.

45. Cauchard, J.R.; E, J.L.; Zhai, K.Y.; Landay, J.A. Drone & me: An exploration into natural human-drone interaction. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 5–11 September 2015; pp. 361–365.

46. Lee, S.-S.; Chae, J.; Kim, H.; Lim, Y.; Lee, K. Towards more natural digital content manipulation via user freehand gestural interaction in a living room. In Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Zurich, Switzerland, 8–12 September 2013; pp. 617–625.

47. Aslan, I.; Buchwald, I.; Koytek, P.; André, E. Pen + Mid-Air: An Exploration of Mid-Air Gestures to Complement Pen Input on Tablets. In Proceedings of the 9th Nordic Conference on Human-Computer Interaction, Gothenburg, Sweden, 23–27 October 2016; pp. 1–10.

48. Chen, L.-C.; Cheng, Y.-M.; Chu, P.-Y.; Sandnes, F.E. The Common Characteristics of User-Defined and Mid-Air Gestures for Rotating 3D Digital Contents. In *Universal Access in Human-Computer Interaction. Interaction Techniques and Environments, Proceedings of the 10th International Conference, UAHCI 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, 17–22 July 2016*; Springer: Cham, Switzerland, 2016; pp. 15–22.

49. 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.

50. Havlucu, H.; Ergin, M.Y.; Bostan, İ.; Buruk, O.T.; Göksun, T.; Özcan, O. It Made More Sense: Comparison of User-Elicited On-skin Touch and Freehand Gesture Sets. In *Distributed, Ambient and Pervasive Interactions, Proceedings of the 5th International Conference, DAPI 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, 9–14 July 2017*; Springer: Cham, Switzerland, 2017; pp. 159–171.

51. Hoff, L.; Hornecker, E.; Bertel, S. Modifying Gesture Elicitation: Do Kinaesthetic Priming and Increased Production Reduce Legacy Bias? In Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, Eindhoven, The Netherlands, 14–17 February 2016; pp. 86–91.

52. 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, Proceedings of the 12th International Conference, DESRIST 2017, Karlsruhe, Germany, 30 May–1 June 2017*; Springer: Cham, Switzerland, 2017; pp. 104–121.

53. Khan, S.; Tunçer, B. Intuitive and Effective Gestures for Conceptual Architectural Design: An Analysis of User Elicited Hand Gestures for 3D CAD Modeling. 2017. Available online: http://papers.cumincad.org/cgi-bin/works/Show?acadia17_318 (accessed on 8 September 2018).

54. 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.

55. Lee, L.; Javed, Y.; Danilowicz, S.; Maher, M.L. Information at the wave of your hand. In Proceedings of the HCI Korea, Seoul, Korea, 10–12 December 2014; pp. 63–70.

56. May, K.R.; Gable, T.M.; Walker, B.N. Designing an In-Vehicle Air Gesture Set Using Elicitation Methods. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; pp. 74–83.

57. Obaid, M.; Häring, M.; Kistler, F.; Bühling, R.; André, E. User-Defined Body Gestures for Navigational Control of a Humanoid Robot. In *Social Robotics, Proceedings of the 4th International Conference, ICSR 2012, Chengdu, China, 29–31 October 2012*; Springer: Berlin, Gremany, 2012; pp. 367–377.

58. 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.

59.   Ortega, F.R.; Galvan, A.; Tarre, K.; Barreto, A.; Rishe, N.; Bernal, J.; Balcazar, R.; Thomas, J.-L. Gesture elicitation for 3D travel via multi-touch and mid-Air systems for procedurally generated pseudo-universe. In Proceedings of the 2017 IEEE Symposium on 3D User Interfaces (3DUI), Los Angeles, CA, USA, 18–19 March 2017; pp. 144–153.

60.   Rateau, H.; Grisoni, L.; De Araujo, B. Mimetic interaction spaces: Controlling distant displays in pervasive environments. In Proceedings of the 19th International Conference on Intelligent User Interfaces, Haifa, Israel, 24–27 February 2014; pp. 89–94.

61.   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.

62.   Wobbrock, J.O.; Aung, H.H.; Rothrock, B.; Myers, B.A. Maximizing the guessability of symbolic input. In Proceedings of the Human Factors in Computing Systems, Portland, OR, USA, 2–7 April 2005; pp. 1869–1872.

63.   Schiavo, G.; Ferron, M.; Mich, O.; Mana, N. Wizard of Oz Studies with Older Adults: A Methodological Note. 2016. Available online: https://www.iisi.de/wp-content/uploads/2018/07/schiavo_et._al._-_wizard_of_oz_studies_with_older_adults_a_methodological_note.pdf (accessed on 8 September 2018).

64.   Köpsel, A.; Bubalo, N. Benefiting from legacy bias. *Interactions* **2015**, *22*, 44–47. [CrossRef]

65.   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, Boston, MA, USA, 4–9 April 2009; pp. 1083–1092.

66.   Ericsson, K.A.; Simon, H.A. Verbal reports as data. *Psychol. Rev.* **1980**, *87*, 215. [CrossRef]

67.   Fonteyn, M.E.; Kuipers, B.; Grobe, S.J. A Description of Think Aloud Method and Protocol Analysis. *Qual. Health. Res.* **1993**, *3*, 430–441. [CrossRef]

68.   Ren, G.; O'Neill, E. 3D selection with freehand gesture. *Comput. Graph.* **2013**, *37*, 101–120. [CrossRef]

69.   Othman, N.Z.S.; Rahim, M.S.M.; Ghazali, M.; Anjomshoae, S.T. Creating 3D/Mid-air gestures. In Proceedings of the 2016 International Conference on Advanced Informatics: Concepts, Theory and Application (ICAICTA), Penang, Malaysia, 16–19 August 2016; pp. 1–6.

70.   Green, P. The Wizard of Oz: A Tool for Rapid Development of User Interfaces. *Proc. Hum. Factors Ergon. Soc. Annu. Meet* **1985**, *29*, 470–474. [CrossRef]

71.   Kelley, J.F. An empirical methodology for writing user-friendly natural language computer applications. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 12–15 December 1983; pp. 193–196.

72.   Nielsen, M.; Moeslund, T.B.; Störring, M.; Granum, E. Gesture Interfaces. In *HCI beyond the GUI: Design for Haptic, Speech, Olfactory and Other Nontraditional Interfaces*; Morgan Kaufmann: Burlington, MA, USA, 2008.

73.   Park, H.-J.; Park, J.; Kim, M.-H. 3D Gesture-based view manipulator for large scale entity model review. In *AsiaSim: Asian Simulation Conference, Proceedings of Asia Simulation Conference 2012*; Springer: Berlin, Gremany, 2012; pp. 524–533.

74.   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.

75.   Groenewald, C.; Anslow, C.; Islam, J.; Rooney, C.; Passmore, P.; Wong, W. Understanding 3D Mid-Air Hand Gestures with Interactive Surfaces and Displays: A Systematic Literature Review. In Proceedings of the 30th International BCS Human Computer Interaction Conference: Fusion! Poole, UK, 11–15 July 2016.

76.   Erazo, O.; Rekik, Y.; Grisoni, L.; Pino, J.A. Understanding Gesture Articulations Variability. In *Human-Computer Interaction—INTERACT 2017, Proceedings of the 16th IFIP TC 13 International Conference, Mumbai, India, 25–29 September 2017*; Springer: Cham, Switzerland, 2017; pp. 293–314.

77.   Oka, K.; Lu, W.; Özacar, K.; Takashima, K.; Kitamura, Y. Exploring in-the-Wild Game-Based Gesture Data Collection. In *Human-Computer Interaction—INTERACT 2017, Proceedings of the 16th IFIP TC 13 International Conference, Mumbai, India, 25–29 September 2017*; Springer: Cham, Switzerland, 2017; pp. 97–106.

78.   William, A.; Tullis, T. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*; Morgan Kaufmann: Burlington, MA, USA, 2013.

79.   Tullis, T.; Wood, L. How Many Users Are Enough for a Card-Sorting Study? In Proceedings of the Usability Professionals Association Conference, Atlanta, GA, USA, 21–24 June 2011.

80. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv. Psychol.* **1988**, *52*, 139–183. [CrossRef]

81. Hincapié-Ramos, J.D.; Guo, X.; Moghadasian, P. Consumed Endurance: A metric to quantify arm fatigue of mid-air interactions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014.

82. Bossavit, B.; Marzo, A.; Ardaiz, O. Hierarchical Menu Selection with a Body-Centered Remote Interface. *Interact. Comput.* **2014**, *26*, 389–402. [CrossRef]

83. Morris, M.R.; Wobbrock, J.O.; Wilson, A.D. Understanding users' preferences for surface gestures. In Proceedings of the Graphics Interface 2010, Ottawa, ON, Canada, 31 May–2 June 2010; pp. 261–268.