A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?

Santiago Villarreal-Narvaez[1], **Jean Vanderdonckt**[1], **Radu-Daniel Vatavu**[2], **Jacob O. Wobbrock**[3]
[1]Université catholique de Louvain (UCLouvain), LouRIM, Louvain-la-Neuve, Belgium
[2]University Stefan cel Mare of Suceava, MintViz Lab | MANSiD Research Center,Suceava, Romania
[3]The Information School, DUB Group | University of Washington, Seattle, WA 98195, USA
santiago.villarreal, jean.vanderdonckt@uclouvain.be, radu.vatavu@usm.ro, wobbrock@uw.edu

ABSTRACT
Gesture elicitation studies represent a popular and resourceful method in HCI to inform the design of intuitive gesture commands, reflective of end-users' behavior, for controlling all kinds of interactive devices, applications, and systems. In the last ten years, an impressive body of work has been published on this topic, disseminating useful design knowledge regarding users' preferences for finger, hand, wrist, arm, head, leg, foot, and whole-body gestures. In this paper, we deliver a *systematic literature review* of this large body of work by summarizing the characteristics and findings of N=216 gesture elicitation studies subsuming 5,458 participants, 3,625 referents, and 148,340 elicited gestures. We highlight the *descriptive*, *comparative*, and *generative virtues* of our examination to provide practitioners with an effective method to (i) *understand* how new gesture elicitation studies position in the literature; (ii) *compare* studies from different authors; and (iii) *identify* opportunities for new research. We make our large corpus of papers accessible online as a Zotero group library at https://www.zotero.org/groups/2132650/gesture_elicitation_studies.

Author Keywords
Gesture elicitation; Survey; Systematic Literature Review.

CCS Concepts
•**Human-centered computing** → **Gestural input;** *User interface design; Participatory design; Empirical studies in interaction design;*

INTRODUCTION
Gesture elicitation studies (GES) were introduced in 2005 by Wobbrock *et al.* [122] in the form of a "guessability method," designed to enable the collection of end users' preferences for symbolic input, such as for the letter-like stroke gestures of the EdgeWrite alphabet [124]. The first application of the guessability method to hand-gesture input was conducted four years later by Wobbrock *et al.* [123] in the context of surface computing. Since then, GES have become a popular tool to inform the gesture user interface (UI) design, reflective of end users' behavior and preferences, for a variety of devices [7, 39,40,90,116], applications [42,78,81,105], environments [32, 54,58,66,92], and contexts of use [25,35,65,104,110]. To date, we were able to identify N=216 such studies published in peer-reviewed journals and conferences, reporting users' preferences for finger, hand, wrist, arm, head, leg, foot, and whole-body gestures for interactive devices, applications, and systems of all kinds; see Figure 1 for heatmap representations.

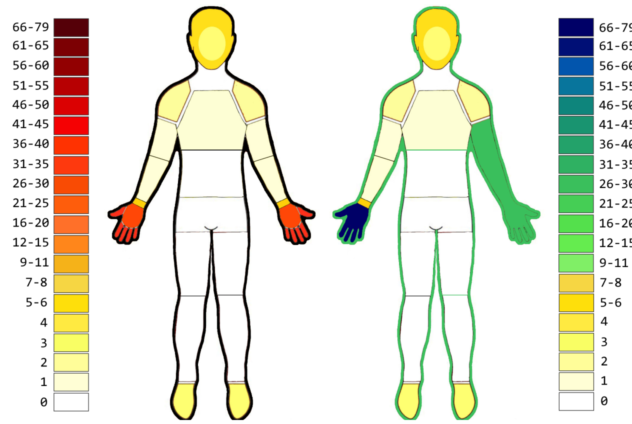

Figure 1. Heatmap representations of the number of GES that elicited gestures performed with specific body parts. ***Left:*** **gestures involving one body part only,** ***e.g.,*** **face or feet gestures.** ***Right:*** **gestures produced as the combination of at least two body parts,** ***e.g.,*** **hand and head gestures or wrist and forearm gestures.**

Consequently, an impressive amount of knowledge has resulted from the application of the GES method regarding: (i) the magnitude of users' agreement over gesture commands they would use for interactive applications, (ii) consensus gesture sets, (iii) design guidelines and recommendations [4,59], (iv) software tools [2,3,61,76,102], and (v) methodological variations and improvements of the original GES method [71,109,113–115]. However, while the community has been busy accumulating design knowledge regarding new types of gestures [21,28,110] or high-consensus gesture commands for new prototype devices [104,116,121] and new environments [17,24,30,32,54,95], the result was an ever-growing list of GES scattered through the various dissemination venues of the HCI scientific literature. In this context, *it is high time to systematize all this knowledge in a rigorous way for the community to be able to identify best practices and*

DIS '20, July 6–10, 2020, Eindhoven, Netherlands.

ACM ISBN 978-1-4503-6974-9/20/07 ...$15.00.
http://dx.doi.org/10.1145/3357236.3395511

use them effectively to advance scientific discovery in end-user gesture elicitation as well as the capacity to transfer results into actual user interfaces and applications.

In this work, we analyze the body of work on gesture elicitation in the form of a *systematic literature review* (SLR). The SLR method [12,49,56] is radically different from conducting a survey or from free-form discussion of related literature in that it follows a *rigorous, well-defined procedure guaranteed to produce reliable, reproducible results*. Moreover, we adopt the generic principles of "designing interactions, not interfaces" [8], which we distill into three virtues of our SLR:

V_1. The *descriptive* virtue: our SLR provides practitioners with a method to describe the underlying characteristics of any GES, including new studies to be published in the future and, thus, position them precisely in the literature. For instance, a practitioner may be interested in forming a clear understanding of the design and results of some new GES addressing gestures for smart rings [39], for which clear dimensions of analysis are needed. The corresponding research questions addressed by the descriptive virtue of our SLR with respect to some GES are: *What are the main meta-data of such GES?*, *What is the scope of the study?*, and *How were the results reported and analyzed?*

V_2. The *comparative* virtue: our SLR enables a method to compare GES from different authors. In our previous example, the practitioner may come across different GES for smart rings. By clearly describing those studies along various dimensions, the practitioner will be able to assess their overlapping degree, similarities, and key differences. The corresponding research questions for this virtue are: *In what ways are two or more GES similar?*, *How can one identify similar GES studies?*, and *Which studies relate to a specific topic?*

V_3. The *generative* virtue: our SLR makes it possible to identify areas of investigation in end-user gesture elicitation that were not covered or have been little explored by previous work. For example, after comparing previous work on gesture elicitation for smart rings, the practitioner should be able to identify new investigative opportunities. The corresponding research questions are: *Which areas require new or more GES?* and *Which areas need verification, validation, or consolidation?*

The contributions of this paper are as follows:

C_1. We conduct *the first systematic literature review on gesture elicitation*, for which we examine, characterize, and compare N=216 published, peer-reviewed studies. To this end, we introduce a detailed research method to guide our systematic literature review. We report findings from 5,458 participants, who were elicited for 3,625 referents, and proposed 148,340 gestures in response.

C_2. We focus on the *descriptive, comparative,* and *generative* virtues of our examination to provide the community with an effective method to (i) *understand* how current and new GES position in the literature; (ii) *compare* GES published by different authors; and (iii) *identify* opportunities for unexplored areas in gesture elicitation or areas in need of knowledge validation.

SURVEYS OF GESTURE ELICITATION STUDIES

A few papers discussed the literature on GES, but are all limited in their scope. For example, Vuletic *et al.* [107] conducted a SLR on hand gestures for user interfaces and mentioned the GES method, but GES were not the main goal of their investigation. Software tools for gesture elicitation were also surveyed [61], but with a focus on features and software qualities from the perspective of engineering interactive computing systems. The work closest to ours is Vogiatzidakis and Koutsabasis [118], who discussed GES for mid-air interaction, and analyzed a corpus of N=47 papers. While an impressive effort, their scope was limited to mid-air gestures only; in contrast, our examination addresses the entire GES literature, representing a scope five times larger (N=216).

CONCEPTS AND TERMINOLOGY

We present an example of a GES study to illustrate the principles of the method and summarize the main concepts.

Illustrative Example of a Gesture Elicitation Study

Imagine a designer that wishes to implement a gesture UI for a new smart ring prototype to enable users to control the lights of a smart home environment with commands such as turn lights on and off, dim lights, and make lights brighter. The designer assembles a group of potential users P, perhaps 20 in number, *i.e.,* $|P|$=20, to participate in a study. Each participant is presented with the effect of each function, *e.g.,* the intensity of the lights is increased for the "make lights brighter" function, and the participant is asked to propose a gesture command using the smart ring that would generate the effect just witnessed. At the end of the study, the designer has collected a set of 80 gestures = 4 (functions) $\times$ 20 (proposals). The designer now looks at the set of gestures elicited for each function to understand whether there are any gestures in agreement. If agreement turns out substantial and the sample of participants is representative of the population of end-users targeted by the smart ring application, then the designer can be confident that the elicited gestures are intuitive, and new users, not part of the original study, will likely guess, easily learn, and perhaps prefer the same types of gestures.

Glossary of Concepts and Terminology

We provide below an alphabetically ordered list of definitions for the concepts and terminology employed in GES.

Agreement: The situation in which the gestures of two or more participants are evaluated as identical or similar according to a set of rules, criteria, or similarity function and, thus, equivalent from the perspective of the target application. ***Example:*** The designer may consider that gesture direction and speed are more important than the amplitude of the finger movement, since the smart ring embeds an accelerometer, but not a sensor for measuring absolute position in space.

Measure of agreement: A numerical measure quantifying the magnitude of agreement among the gestures elicited from participants. ***Example:*** From the 20 participants of the study, three subgroups of sizes 9, 7, and 5 emerge for the "dim lights" referent, so that all the participants from each subgroup are in agreement. The agreement score [122] computes $(9/20)^2+(7/20)^2+(5/20)^2 = .388$, and the agreement rate [114] returns $(9\cdot8+7\cdot6+5\cdot4)/(20\cdot19) = .353$.

Agreement score/rate: Various measures of agreement have been referred to as a "score" [122] or a "rate" [114].

Command: A signal that actuates the execution of a function in the user interface. Also referred to as a "gesture command," where the signal is represented by gesture input. ***Example:*** In response to the "dim lights" referent, a participant may propose a downwards movement of the finger.

Consensus rate: A measure of agreement [113] that employs dissimilarity functions and tolerances.

Consensus gesture set: The set of gestures that reached the largest agreement for the referents examined in the study. ***Example:*** Say that 11 participants believe that a downwards movement of the finger is best to effect "dim lights"; 9 participants consider than an upwards movement should effect "make lights brigher"; and 15 participants propose clockwise and counter-clockwise circle movements to turn the lights on an off. In that case, the consensus gesture set is composed of the downwards, upwards, clockwise, and counter-clockwise gestures of the finger wearing the smart ring.

Dissimilarity function: A numerical function that computes a real, positive value reflecting how dissimilar two gestures are [113]. ***Example:*** If the designer collects gestures numerically, *e.g.,* as a set of linear acceleration points, the Euclidean distance or the Dynamic Time Warping functions [112] could be used to automate the evaluation of agreement measures.

Elicitation: The process of making participants respond to referents and propose gestures to effect those referents.

End user: A potential user of the interactive device, application, or system for which gestures are designed; a sample of end users forms the "participants" of the GES.

Function: A feature of the user interface that can be controlled independently using a command; the same as "referent." ***Example:*** make lights brighter.

Gesture: A movement of a part of or the whole body performed in response to a referent.

Referent: A feature of the user interface that can be controlled independently using a command; the same as "function." ***Example:*** make lights brighter.

Symbol: Any artifact that evokes a referent in the form of a computer function, *e.g.,* stroke gestures, surface gestures, mid-air gestures, command keywords, voice commands, icons, button labels, menu items. Synonym: sign.

Participant: A subject, representative of the population of end users, that volunteered to participate in the GES.

RESEARCH METHOD

We outline our method for conducting the SLR, which we organized as a four-phase flow procedure consisting of Identification, Screening, Eligibility, and Inclusion phases, inspired by the approach of Liberati *et al.* [56]. We took specific care to implement each phase, so that our results could be readily reproduced. Fig. 2 shows an overview of our method represented in the form of a PRISMA[1] diagram [56].

[1] Preferred Reporting Items for Systematic reviews and Meta-Analyses.

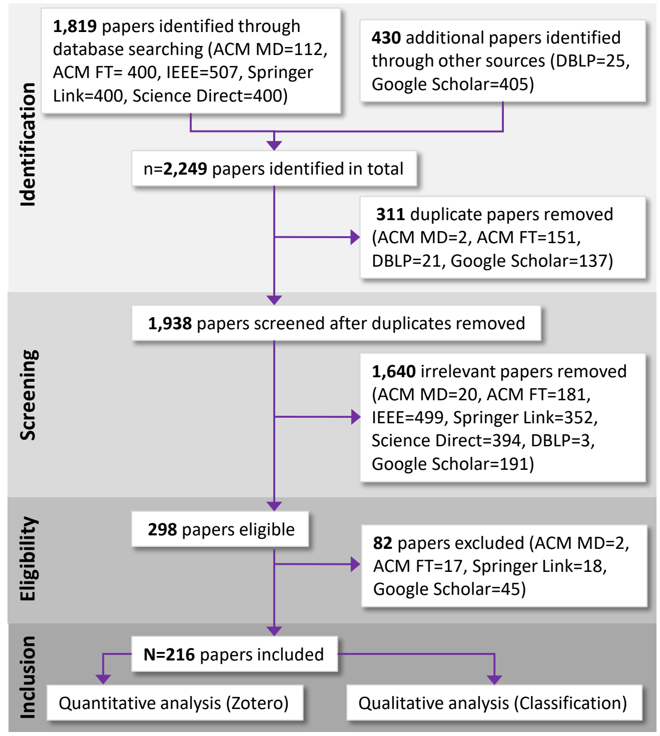

Figure 2. The four-phase flow diagram of our SLR examination.

Phase 1: Identification

We searched for papers potentially relevant to our topic of investigation using the following query:

```
Q = ("Gesture" AND "Elicitation" AND "Study")
```

which we ran on both *single-publisher libraries* (*e.g.,* IEEE Xplore) and *multi-publisher engines* (*e.g.,* Google Scholar).

We selected the following five major Computer Science digital libraries: (1) ACM DL[2]; (2) IEEE Xplore[3]; (3) Elsevier ScienceDirect[4]; (4) Elsevier Ei Compendex (also referred to as Engineering Village[5]); and (5) SpringerLink.[6] We also used two multi-publisher sources to ensure the *completeness* and *coherence* of the GES references, validate independent query results, and cover other publishers as well: (6) DBLP CompleteSearch[7] and (7) Google Scholar.[8] These two engines were employed more as a verification mechanism than for reference identification purposes.

To ensure the *stability* and *constancy* of our queries, we downloaded all the resources during one week. First, the list of references was compiled in one day (August 15, 2019), which we set as the last day for considering GES in our SLR. Individual papers were then progressively retrieved from the digital libraries over several subsequent days (until August 21, 2019) to avoid excessive requests that might forbid our authorized

[2] https://dlnext.acm.org
[3] http://ieeexplore.ieee.org/
[4] http://www.sciencedirect.com/
[5] https://www.elsevier.com/solutions/engineering-village/
[6] http://www.springerlinks.com
[7] http://www.dblp.org/search/
[8] http://scholar.google.com/

access to those libraries. The results of each query `Q` were retained according to the following rules:

1. *The "advanced search" rule.* Queries were run using the most advanced search feature of the digital libraries, and results were sorted in decreasing order of their relevance.
2. *The "minimum number of references" rule.* If the query returned a small number of references (less than 400), all those references were automatically retained.
3. *The "no missing references" rule.* If the query returned more than 400 references, the first 400 were automatically selected for further screening, and a manual check of the next 200 references was performed to make sure no relevant papers were discarded.

Table 1 shows the results of executing query `Q` for each source of references. For example, the ACM DL returned a number of 2,859 references in full-text search mode, of which 400 were retained; from those, 151 duplicates were identified when comparing against the results of other queries, and 198 references were excluded based on further screening criteria (see next), leaving 51 papers; see the second row of Table 1. Overall, our queries identified 1,938 references across all seven sources.

Phase 2: Screening

Each paper was evaluated with respect to its relevance to GES using criteria of *form* and *contents*, as follows:

Form. We retained only papers written in English that underwent a peer-review process, and for which the full text and the resulting gesture set was available. These included research papers published in peer-reviewed journals, conferences, symposiums, workshops, and excluded unreviewed references where the full gesture set was not published or accessible, such as patent descriptions, standards, extended abstracts, slideshows. PhD and Master theses were not considered, although they could in a future version. *Contents.* We retained only those papers that (1) explicitly introduced a GES for UI design; or (2) discussed GES, emphasizing at least one discriminating feature (*e.g.,* , scale [80], memorability [73], etc.); or (3) explicitly used a method to examine GES. A number of 1,640 irrelevant references were excluded by our screening, leaving 1,938 - 1,640 = 298 papers.

Phase 3: Eligibility

We considered ineligible and further excluded those papers that matched any of the following situations: (1) *independent research question*: the paper explicitly mentioned a GES, but did not report its results, or the goal of the GES was not gesture UI design, *e.g.,* Yin and Davis [127] conducted a GES to collect data for training a gesture recognizer; (2) *methodology-oriented*: the paper addressed methodological aspects of GES, but did not report an actual study, *e.g.,* Morris *et al.* [71] discussed legacy bias for gesture elicitation; or (3) *field mismatch*: the paper reported an actual elicitation study, but for another discipline, *e.g.,* a GES conducted to elicit user-defined gestures to train a robot. By using these rules, we further removed 82 papers, leaving a final corpus of N=216 studies for our examination.

Phase 4: Inclusion

This phase consisted in verifying *quantitative* and *qualitative* aspects of our corpus of papers, as follows.

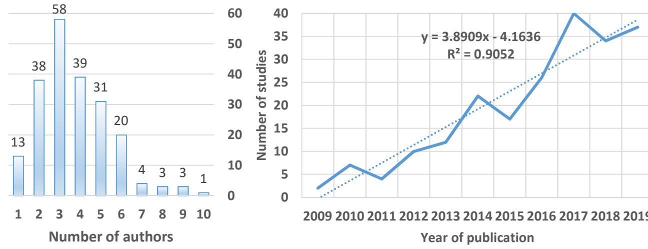

Figure 3. Number of authors per paper (a), over time (b).

Quantitative analysis. We employed the following tools to create a collection of papers and generate summary statistics:

1. Zotero,[9] a multi-platform bibliography management software tool use for managing our corpus publicly available at https://www.zotero.org/groups/2132650/gesture_elicitation_studies.
2. PaperMachines,[10] a Zotero extension for visualizations.
3. PDF2Text,[11] software for automatic extraction of text from PDF files. The extracted text was submitted to automatic language processing and analysis.

Qualitative analysis. The list of N=216 references was exported from Zotero into a spreadsheet, and qualitative classifications of each GES were manually performed with regards to various dimensions of analysis, *e.g.,* , number of participants, measures of agreement, etc.; see next section for details. An *inter-judge agreement procedure* was employed for this, as follows. Two HCI researchers, with 5 and 20 years of experience, respectively, and who were both authors of this paper, independently classified each paper based on the set of criteria. In case of disagreement, a third researcher with 9 years of experience, who was not an author of this paper, arbitrated a discussion between the first two researchers until consensus was reached.

RESULTS: DESCRIPTIVE VIRTUE

We present in this section an overview of the characteristics of published GES according to various dimensions of analysis, such as the number of participants involved, the types of gestures addressed, any complementary input modalities, and more. We seek to understand the audience of GES by looking *where* those studies were published and *which* studies have been most influential.

Authors, Venues, and Hot Topics in Gesture Elicitation

Figure 3b shows the number of GES conducted between 2009, the year when the first hand-gesture elicitation paper of Wobbrock *et al.* [123] was published, and 2019. The increasing number of studies (we found a linear regression with R^2=.90) suggests that the peak of GES has not been reached yet. Figure 3a highlights the number of authors involved in conducting and reporting GES. While a typical GES paper has three authors (60 studies, representing 27.8%), a number of studies were conducted by single authors [46,64,111], while others by up to eight [53,78], nine [4,74], or even ten authors [101], when multiple classification criteria or domains of expertise were needed [20,74].

[9] http://www.zotero.org/
[10] http://www.papermachines.org/
[11] http://www.pdf2text.com/

Source	Query	Rules	Duplicates	Excluded	Included	References
1. ACM DL (metadata)	112	112	2	22	88	[1–4,6,7,10,14,17,18,21,23–26,28,30,32–35,37–39,45–48,52–54,61,64–66,68,70, 73,75,79–81,85,86,89–91,93–98,100,101,103–105,111,113,115,116,121,123,129, 133,138,140,141,144,147,149–153,160,165,166,170,187,188,191,193,195–197, 199–201,203,204,208,209,211,214,217–222,224,227,229,233–235,238,242,249, 261–263,266,274,275,279,280,283,287–289,294,295,298–300,304,306,307,309, 313,318,322,329,341–343]
2. ACM DL (full-text)	2,859	400	151	198	51	
3. IEEExplore	507	507	0	499	8	[27,78,92,163,184,267,314,328]
4. ScienceDirect	>6,000	400	0	394	6	[51,99,145,146,243,320]
5. SpringerLink	1,662	400	0	370	30	[42,44,55,58,74,82,110,125,128,135,154,156,161,171,178,182,205,206,236,239, 252,256,268,273,282,319,333,334,339,340]
6. DBLP	25	25	21	3	1	[119]
7. Google Scholar	405	405	137	236	32	[5,20,59,63,72,84,126,134,143,158,168,176,189,192,210,215,216,226,231,250, 251,257,258,272,284,308,317,326,327,330,336,338]
Total	**>11,570**	**2,249**	**311**	**1,722**	**216**	

Table 1. Number of references returned by running queries on digital libraries and search engines.

Conference proceedings:	**171**	**79%**
CHI	38	17.6
ITS	10	4.6
MobileHCI	9	4.2
DIS	8	3.7
AVI	5	2.3
AutomotiveUI	5	2.3
NordiCHI	5	2.3
OzCHI	5	2.3
INTERACT	4	1.8
3 studies per venue: AISC, EICS, HCII, ICMI, IDC, PerDis, SUI, UbiComp, UIST	9 × 3 = 27	12.5
2 studies per venue: AHA, DAP, EuroITV, HFES, HMI, HRI, ISS, IUI, TEI, TVX	10 × 2 = 20	9.2
One study per venue: 3DUI, APCHI, ASSETS, CASA, CCHI, CHIuXID, CNS, DAA, DAD, DC, DIGI, EuroVis, FDG, GI, HCC, HCIK, ICTD, Interaccion, ISCC, ITIE, MC, Mindtrek, MM, MMUI, MobiSys, MoMM, MS, MUM, Perspective, Reco, ROMAN, SA, SAICSIT, SCIS, UIC	35×1 = 35	16.2
Journals:	**45**	**21%**
MTA	6	2.8
IJHCS	5	2.3
JAIS	3	1.4
IEEE Access, IJHCI, TOCHI, UAHCI	4 × 2 = 8	1.8
One study per venue: AS, BIT, CHB, CTW, EDAM, HF, IJMHCI, IMW, Informatics, Information, JCH, JCST, JIFS, JMUI, JUS, Machines, MTI, Pervasive, PLOS, Presence, PUC, TiiS, Visual	23 × 1 = 23	10.6

Table 2. Distribution of GES in the literature.

In terms of dissemination venues, Table 2 lists conferences and journals where GES have been published, with conferences taking a major lead (79%). Of these, CHI (38 papers), ITS (10), MobileHCI (9), and DIS (8 papers) had the highest number of GES. Journal articles account for 21% of all published GES, but no clear preferences can be identified. The variety of dissemination venues shows a body of gesture design knowledge that is scattered through the literature of HCI and, in some cases, of other disciplines, which is challenging for practitioners to locate specific information.

Table 3 lists the twelve most influential GES in decreasing order of their Google Scholar number of citations (from August 15, 2019). The hot topics that received most citations are directly proportional to the importance, the diffusion, or the novelty of the device studied: tabletops [38,72,73,122, 123], smartphones and tablets [47,90], smart TVs and smart homes [51,111], augmented reality [81], and shape-changing interfaces [53] or a drone [17], two emerging fields. The consideration of other categories of users than able-bodied adults, along with the analysis of its influence, appeared for blind participants [47], who analyzed the trade-off between the users' preference and the system's performance. Memorability [73] demonstrates the superiority of user-defined gestures over designer-defined or system-defined gestures with with respect to preference and recall rate. The body of GES knowledge counts for 5,907 citations with 856 citations per year on average.

Participants

Figure 4 presents a detailed overview of the number of GES in relation to the number of participants involved (vertical axis) and over time (horizontal axis)[12].

We found that the average number of participants was M=25 (SD=4), and the most frequent choice was 20 participants, a sample size that was adopted by 36% of the studies. This preference is likely explained by the fact that Wobbrock *et al.* [123] employed 20 participants in their original gesture elicitation study. Starting from this value, we highlighted three regions in Figure 4: (1) the "green" region, where the number of participants is at least 20 (51%); (2) the "yellow" region with less than 20 participants, but above 10 (38%); (3) and the "orange" region, with studies reporting results from a small number of participants, less than 10 (11%). Note, however, that a small sample size is justifiable in some cases, such as when the study examines end-users from populations that are more challenging to access, such as users with impairments, or populations of users under-represented, such as small children. Large sample sizes were employed in studies interested in cultural aspects [74], such as 340 participants to test culturally-aware touch gestures [65] or 227 participants involved in framed guessability for multiple devices [14]).

We also analyzed the ratio Male/Female[13] and found an average of 2.10 (SD=2.18, Mdn=1.13). Most of the studies (52%) involved more male than female participants; a parity of gender was found in 15% of the studies, and 17% studies had more female than males. A gender bias is clear from these figures, probably due to availability or volunteering aspects.

[12] Years were omitted when GES were absent. Four studies did not report their sample size and, thus, N=212 for this report.

[13] For the 183 GES studies that reported this information.

	Authors	Study	Application	Year	Cita-tions
1.	Wobbrock *et al.* [123]	User-defined Gestures for Surface Computing	interactive tabletops	2009	1070
2.	Ruiz *et al.* [90]	User-defined Motion Gestures for Mobile Interaction	smartphones	2011	373
3.	Kane *et al.* [47]	Usable Gestures for Blind People: Understanding Preference and Performance	tablets, smartphones	2011	256
4.	Wobbrock *et al.* [122]	Maximizing the Guessability of Symbolic Input	hand-stroke gestures	2005	253
5.	Morris *et al.* [72]	Understanding Users' Preferences for Surface Gestures	interactive tabletops	2010	228
6.	Vatavu [111]	User-defined Gestures for Free-hand TV Control	smart TVs	2012	169
7.	Piumsomboon *et al.* [81]	User-defined Gestures for Augmented Reality	Augmented Reality	2013	165
8.	Kühnel *et al.* [51]	I'm home: Defining and evaluating a gesture set for smart-home control	smart homes	2011	156
9.	Lee *et al.* [53]	How Users Manipulate Deformable Displays As Input Devices	unconventional displays	2010	144
10.	Nacenta *et al.* [73]	Memorability of Pre-designed and User-defined Gesture Sets	interactive tabletops	2013	143
11.	Cauchard *et al.* [17]	Drone & Me: An Exploration into Natural Human-Drone Interaction	human-drone interaction	2015	134
12.	Frisch *et al.* [38]	Investigating Multi-touch and Pen Gestures for Diagram Editing on Interactive Surfaces	surface computing	2009	125

Table 3. The twelve most influential GES, authors, and application domains, according to the number of Google Scholar citations.

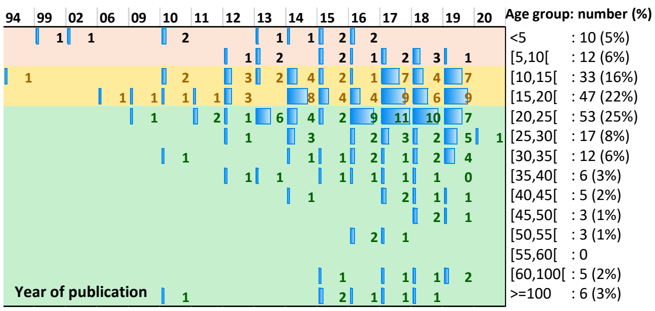

Figure 4. Number of participants by age groups involved over time.

202 GES (93.5%) reported the age of the participants. Mean age was 27.57 years (*SD*=14.62, *Mdn*=23.4). Seven studies (3%) addressed children, 4 studies (2%) teenagers, and the vast majority of GES studies focused on adults (93%), of which 4 (2%) addressed older adults. Often, the background or the level of experience with technology is missing.

Body Parts Addressed by Gesture Elicitation

We found that some GES focused on specific body parts (*e.g.,* legs [99]) or combinations of body parts [78], while other studies examined gestures performed by the whole body [21]. From this perspective, we can identify a range of investigations that start from a specific body part and progress to the whole body, which we refer to as the "human gesture continuum." We consider this continuum as a structured mechanism, whereby gestures performed by the human body can be uniquely referenced. For instance, gestures performed with upper-body limbs can be decomposed into the various body parts that were involved in the production of the gesture, such as the face [89,93,99], head [28,99,120], head and shoulders [110], eyes [62], nose [87], mouth [27], shoulders [87], torso [99], and even belly gestures on the abdomen [117]. The arms are themselves subject to a gesture continuum: fingers [18], wrist [91], wrist and hand gestures [48], hands [5,37,82,116], forearm [52,91], arms [57], and skin-based gestures [44]. In terms of gestures produced by lower-body parts, we identified foot gestures [1,33,34], legs [98,99], and whole-body gestures [21,23]. To understand the coverage of body parts in the GES literature, we computed two measures:

1. The *Individual Frequency of Body Parts* (IFBP), showing often a part of the body, such as the hand [81] or head [28], was employed in a GES study.
2. The *Combination Frequency of Body Parts* (CFBP) of body parts, reporting how often specific body parts were combined, such as head and hand gestures [20].

The first measure reports on the importance of individual body parts, while the second on their combinations, including gestures performed at the level of the entire human body.

Figure 1 illustrates individual frequency values in the form of a heatmap directly on the human body, while the combination frequency is shown as colored outline of various body parts, *e.g.,* feet are shown in yellow while the outline of the human body is shown in dark orange. We found a number of 380 occurrences of individual body parts, of which 241 were combined. The most frequent gestures were distributed as follows: finger (144/380=38%; 39/241=16%), hand (139/380=37%; 28/241=12%), arm (35/380=9%; 1/241<1%), body (30/380=8%; 30/241=13%), wrist (8/380=2%; 8/241=3%), voice (7/380=2%; 7/241=3%), head (5/380=1%; 5/241=2%). Face, feet, shoulders, torso, or skin represent each 1% or less. However, hand and fingers are covered by 77/241 = 32% of combinations, followed by arm + hand + fingers (28/241=12%), and arm + hand (6/241=2%).

Referents

The average number of referents was *M*=20.10 (*Mdn*=22.03) with a wide variance (*SD*=5.23) indicating that the number of referents varies from a few (*Min*=1 [85]) to many (*Max*=70 [125]). Referents have been presented to participants graphically using images (most of the time), animations, videos, UI prototypes, either mock-ups or functional. This finding concurs with a study identifying the most frequent and efficient representations of referents [67]. Referents are often decided according to standard sets of categories [116,128]: *basic functions* (*e.g.,* TV on/off, Next/previous channel, Volume Up/down/mute, Open/Hide menu, and Open/Close item), *generic functions* (or *context-independent*: *e.g.,* Select simple choice, Select multiple choice, Select a date, Specify a number, an angle), and *specific functions* (or *context-dependent*: *e.g.,* TV Guide, Play/pause). Functions are described as being either *analogue* (when they induce a physical effect of real world, such as move, select, rotate, shrink, enlarge, pan, zoom/out, previous, next, maximize, and minimize) or *abstract* (when no physical effect of real world is induced, such as insert, delete, modify, cut, duplicate, paste, undo, menu access, and open.

Gesture Datasets

The average number of elicited gestures across all studies was M=716 (SD=1438, Mdn=375) with a minimum of 4 gestures [129] and a maximum of 12,240 [96,97]. After grouping gestures into clusters of identical or similar types, the average number of reported gestures was M=91 (SD=231, Mdn=24). Of these, an average of C=21 were selected to form consensus gestures sets. Often, C is smaller than the number of referents, since one gesture may be assigned to more than one referent[14].

Other Input Modalities

Besides gestures, other modalities have been elicited, either independently or in combination with gesture input, such as menu shortcuts [13], function keys [7], voice commands [55, 70], multi-device commands [6], and multimodal input [75]. In particular, the “Web on the Wall” study of Morris [70] elicited both gesture and voice commands, and was reproduced with multimodal commands [75].

Analysis

The most frequently reported measures of agreement have been: the agreement score A [122] (34%), the agreement rate AR [114] (51%), the maximum consensus [70] (5%), the consensus rate [70] (3%), the frequency of proposed gestures [89] (2%), and others (less than 5%). Most studies did not specify whether the measures were computed manually or automatically from a gesture set (*e.g.,* using tools such as AGaTE [114] or via spreadsheet calculations [110]).

Other measures have been reported besides agreement, such as: *Goodness-of-Fit*, a subjective assessment rating the participants’ confidence about how well the elicited gesture fit the referents [51,123]; *Memorability*, a recall rate of a gesture some time after it has been elicited [73], *Ease-of-Execution*, a subjective assessment expressing the participant’s perception of the ease of producing physically a gesture elicited [123,128]; *Ease-of-Conception*, a rating of how it is easy to come up with this gesture elicited for a referent [77]; *Enjoyability*, same for the subjective perception of playfulness [28], and validation through an *identification study*, which reverses an elicitation study by showing a new set of participants elicited gestures and asks them for the referent they think those gestures effect [3].

RESULTS: COMPARATIVE VIRTUE

All the words contained in the metadata of each paper, such as in their title, abstract, internal and external keywords, and all the terms contained in their corresponding PDF papers were extracted[11]) to feed a database of indexed keywords and tags. A visualization tool[15] exploits this database to produce several graphics and to browse the corpus along several structural dimensions, such as year, venue (*e.g.,* CHI, DIS, ITS), keyword (*e.g.,* agreement), type of gesture (*e.g.,* hand, motion gestures).

To determine important terms in the corpus, we built a Word Cloud based on Term Frequency-Inverse Document Frequency (TF-IDF), the most frequently applied weighting scheme in recommender systems [9]. Besides common terms found in this cloud (*e.g.,* gesture, participant, elicitation), other terms suggest that GES are indeed intended to design gestures for interaction, then to train a gesture recognizer afterwards (though not much covered), but not to consolidate gestures into a database. The most frequent terms cover either a device type (*e.g.,* table, TV) or a gesture class (*e.g.,* mobile, motion, mid-air gestures), or some limb (*e.g.,* hand, body, finger). However, they do not denote the labelling or the definition of individual gestures (*e.g.,* Swipe is the only one standing out) neither do they refer to a precise context of use. No common definition of referents stands out.

To better understand how these terms evolve over time, we computed a chronological tag cloud for every year from the first GES identified [84] until the last one [63] based on Dunning’s method [29]. In the beginning, GES were mostly focused an applying image processing to handle gestures, but they progressively switched to other capture (like raw data, video, manual recordings). Head and deictic gestures were the first considered around 2005, then surface gestures from 2009. From that time, full-body gestures appeared as well as the first decyphering schemes (*e.g.,* [69,77,81]) to systematically classify gestures. Motion gestures appeared in 2010, as well as the first considerations of various categories of user (*e.g.,* sighted vs blind users), but these categories remained basic.

Physical aspects related to multi-touch gestures in a smart environment were significantly considered in 2011 with different devices, but not really in a real context. The focus has been primarily put on the evolution of devices (*e.g.,* steering wheel in 2013, deformable displays in 2014, tangible interaction in 2015, wristwatch in 2018), then on participant categories (*e.g.,* different ages in 2016, visually impaired people in 2017).

The most emerging topics in 2019 concern the disagreement problem, the more profound consideration of different age groups (not just adults), public and social aspects, also raising the need to better analyze data sources and to use them in the rest of the development life cycle. This evolution is also revealed in a topic modelling based on a Sankey diagram [88] automatically generated from the GES contents: topics are sorted by most common, most coherent, and most variable, and can be presented as a stream, a stacked bar, a line graph, or a categorical histogram.

Any stream can then be explored to list GES satisfying the criteria selected and papers can be further explored. In particular, the flow representing the new gesture sets in new contexts of use, necessarily promoted by the most recent devices, is maintained over time. The interest for other topics, like finger-based gestures, classification schemes, and semantic aspects is decreasing to its minimum.

TAKEAWAYS: GENERATIVE VIRTUE

If the constraints imposed on the initial GES characterization are relaxed, then this characterization can be generalized to raise new ways of conducting and interpreting a GES, some of which have been partially explored, some others are left to be investigated. As takeaways, this section suggests several new lines for conducting a GES.

[14] In the original method [122], if the same gesture is being used for multiple referents, then a conflict resolution process assigns the gesture with the most agreement to its referent and the next referent receives its second-most agreement gesture. Many studies did not invoke this step, although it was part of the original procedure.

[15] See supplemental material for these outputs and for more details.

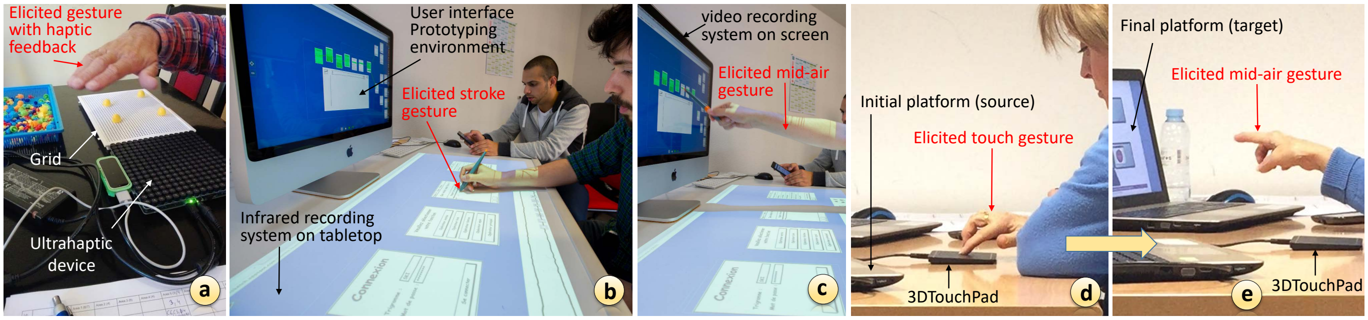

Figure 5. Suggestions for imaginary setups not covered in the SLR: with haptic modality (a), within a UI prototyping environment (b:by stroke gesture, c: by mid-air gesture), for cross-device interaction (d:touch gesture on the source platform followed by, e: mid-air gesture on the target platform.

Eliciting other symbols than gestures, other modalities

Most GES reviewed produce gestures only. Other symbols could be elicited as well, such as a command with parameters, menu items along with their mnemonic, shortcut, or function key, a function performed with one or many modalities (*e.g.,* graphical, vocal, tactile, haptic). For example, Fig. 5a relies on a physical grid with plots to elicit haptic feedback from a ultra-haptic device. Among the five primary human senses, vision and audition are covered much more than tactition, gustation, and olfaction, probably because our human brain filters signals so that the visual, auditory, and tactile channels respectively occupy 80%, 10%, and 5% of the total bandwidth, thus leaving less space for others. Beyond these senses, secondary senses, such as thermoception (heat), nociception (pain), or equilibrioception (balance) could also serve for eliciting new ranges of symbols. In these cases, the symbol should be captured easily in a cost-effective way, such as by a device, a video or any other recording system. Symbols are also ideally captured numerically, that is, they can be compared computationally. For a monomodal symbol, the modality conditions the recording system. Gestures and vocal commands are straightforward to capture, but nocipetion requires thermal devices. For a multimodal symbol, a minimal UI, *e.g.,* a prototype, should provide the participant with the right context of use to foster elicitation. Fig. 5b-c depicts a running GUI prototyping environment where the participant could elicit different symbols for the same GUI command: a stroke gesture (b) and/or a mid-air gesture (c). The participant elicits gestures according to the CARE properties [22]: one gesture (assignment), alternate gestures (equivalence), two combined gestures (redundancy) or two gestures making a single function (complementarity).

Eliciting more than one symbol

Most GES reviewed elicit one gesture per participant for each referent at a time, some repeat this procedure until a certain condition is met, such as the participant is happy with the final gesture, the score exceeds some threshold, a combination of measures reaches a certain value. GES should benefit more from methods proposed to increase variations of symbols elicited, such as the “Production principle” [71] where participants should elicit a certain number of symbols before they can move onto the next referent. The trial and error may also inform the elicitation process, namely by computing measures taking into account different votes with different scores ranking the alternatives. In this way, the agreement could be assessed not only on one symbol, but also on all candidates suggested by participants, not just the best one.

In the original GES, one symbol is finally assigned to each function. When there is a small number of functions, the consensus set remains reasonably small, therefore simple and easy to use. When there is a larger number of functions, the consensus set may confuse the end user because of too many symbols to reproduce, thus exceeding people’s cognitive limits. So, instead of mapping a single gesture or a single symbol to a single function, multiple symbols could be combined to trigger a function by relying on hierarchical structure and congruence [16]. For example, Fig. 5d-e depicts a GES for cross-device gesture interaction in which the participant elicits a touch gesture on a 3DTouchPad to copy an object on the source computer (d) and a mid-air gesture to paste it on the target computer (e), thus eliciting a touch+air gesture [19]. McNeill identifies two paradigms of interaction [69]: the “function-object” paradigm in which a function is first designated and then an object on which this function will be performed is identified, and the “object-function” paradigm in which an object is first selected and then submitted to a function that is subsequently specified. *Multi-symbol* elicitation could therefore support both paradigms by eliciting an symbol for the function and another for the object, whatever their ordering is. This approach would be particularly suitable for gestural interaction with a wide range of devices, like in rich smart environments, where too many gestures could be reduced to a smaller subset of gestures designating functions (on/off, move) and another gesture for the object (TV, radio).

Eliciting for zero to many referents

Most GES reviewed target one referent at a time. Zero to many gestures could be elicited as well for more than one referent, especially over time to explore retention, but also to investigate multimodality. The amount of elicited gestures that can be remembered over time by a participant is a function of frequency of usage, along with is relationship with physical and form factors. It is not because a participant elicits a gesture that it will be accurately recognized (recording the gesture raw data is useful to train a future recognizer and to test its recognition rate [113]), effectively remembered (the participant may have forgotten the gesture after some period of non-usage), and efficiently produced (fatigue, confusion, conflict may occur that prevent the participant to reproduce the adequate gesture in context, especially in demanding conditions). An empirical study could investigate the factors that influence the users’ preferences for gestures (*e.g.,* effectiveness, efficiency, but also naturalness, social acceptance) against the system’s performance (*e.g.,* how the recognizer will properly

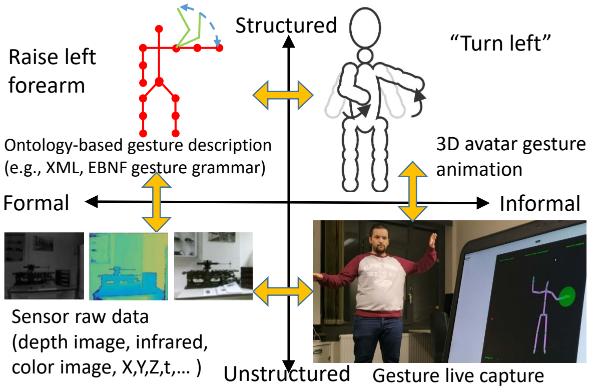

Figure 6. (In)formal, (Un)structured gesture representations.

work in context). Structured Equation Modelling (SEM) [106] may be an appropriate research method for approaching this question as it performs causal modelling between qualitative and quantitative variables, such as between user's preference and system's performance. A *referent* materializes a function depending on its context [15] by providing a scene before and after carrying out a function, with its description. A referent presumes not making any reference to how the function could be executed. A referent provides some guidance, but it can bias the participant. Therefore, we suggest a *function-less* or *referent-less* method, where participants are proposed to elicit any symbol without referring to any function or referent: "Please give me ten gestures that you would like to use with this new device". We have to investigate whether any significant difference exists between a referent-oriented and a referent-free method [60].

Multiple representations of referents and symbols

Referents used in reviewed GES exploit a wide range of representations [67], ranging from verbalizing the name of its associated function to an almost real application of the function without showing the symbol. The influence of the referent representation on the elicitation process should be further investigated depending on the context of use, such as whether using alternate representations for the same referent (*e.g.,* presenting a vocal referent, then a graphical equivalent) would affect the quality of elicitation. In most cases, a graphical representation is sufficient. Representations that are specific to a context could be used instead. Gestures elicited in GES reviewed, apart when they also serve afterwards for recognition [102], are typically captured in an informal, unstructured way, *e.g.,* by a video to be further interpreted by an observer. When the elicited gesture is really captured by the sensor, thus producing raw data, this representation could be transformed into an informal, structured representation (Fig. 6), and then submitted to interpretation to derive a formal and structured representation. Creating a gesture ontology, *e.g.,* based on an Extended Backus-Naur Form (EBNF) grammar leading to a XML-compliant representation [79], would enable expressing any gesture in a common format for computational analysis. This representation could automatically generate a formal, structured representation, such as a 3D animation of a gesture. The connections among these representations could be explored via different mechanisms, such as computer vision, gesture recognition, ontology querying, automatic avatar generation, image processing, and deep learning of the videos.

Closed vs. Open Elicitation

Most GES reviewed foster an *open elicitation*: participants propose whatever symbol they want in an unconstrained manner. A *closed elicitation* invites participants to select symbols from a predefined set, *e.g.,* via a mapping procedure, thus constraining their choice. Open elicitation keeps the process as natural and creative as possible, assuming that the elicitation captures all variables at once. The symbol set is unlimited, sometimes allowing too many symbols to be elicited, from which little or no consensus emerges. Closed elicitation tends to increase the agreement among participants due to their limited choice, but prevents them from being creative. An open elicitation should precede a closed one: after participants elicited their initial gestures and after computing the initial agreement, they enter into a closed elicitation to converge towards a better agreement or better other measures, thus limiting the size of the consensus set. This approach is proposed in *framed guessability* [14].

Reversing elicitation

GES reviewed all share at least a *forward elicitation*, which was a condition for inclusion. *Identification studies* ([123], p. 1091) are the reverse of elicitation studies: the symbols are presented to participants, who are asked what they think those symbols would do or would cause to happen in the interactive system. Identification studies can serve as validation for elicited gesture sets, but this approach is rarely applied in reviewed GES. Ali *et al.* [3] formalize and discuss this process. With both elicitation and identification studies, more degrees of freedom: participants either propose the symbols they want with respect to a referent or not (*e.g.,* a gesture out-of-the-blue, a gesture associated to a known function) or vice versa (*e.g.,* propose a function and/or a referent for a gesture they already master). Legacy bias [71] prevents participants from proposing original, creative gestures in some circumstances, even modified [45], but could foster consistent gestures [50].

Multiple measures for elicited symbols

Although the agreement is the most frequent measure found in GES reviewed, each symbol could be assessed by one or many measures defined by a *measurement method*. A measure could be qualitative (*e.g.,* user satisfaction) or quantitative (*e.g.,* complexity of the symbol), objective or subjective. A qualitative objective measure could could describe what the gesture is, *e.g.,* based on a textual representation (Fig. 6): "The participant's hand moves with five fingers across the table." Any measurement method should be appropriate for the referent provided. While agreement is mostly computed by scores [114,123] and rates [115], computing agreement rigorously is not straightforward and alternate measurements exist [35,59,109,113]. Recently, agreement has been progressively superseded by disagreement computation [113,126]. Other measures could confirm or disconfirm the preferences elicited, such as methods for consolidating votes from participants on symbols: the *Condorcet method* [83], the *de Borda method* [31], and the *Dowdall method* [36]. Each method ranks symbols based on votes (*e.g.,* end users' preferences) and choose a winning symbol with their own rule: "select the symbol that beats any other symbol in exhaustive pairwise comparison" for Condorcet, "select the symbol that on average stands highest in the preferences" for de Borda. These methods sort symbols in different orders.

Other symbol assessment than with agreement

While (dis)agreement represents a key measure to be assessed among participants in reviewed GES, the overall evaluation of elicitation output could be assessed by other measures, which then strives for a more holistic computational model of acceptability of elicited symbols, not just based on agreement. For example, a GES could balance end users' agreement with the biomechanical capabilities of the gesture (*e.g.,* hand supination is more powerful than pronation). This approach reflects the mobility of some limbs and the tendency to rely on the most mobile limbs and gestures. Our SLR identified several other measures beyond agreement which could feed a more general weighted model. Such a model could express symbols as a vector of pairs (property, value) and rely on similarity metrics used in information retrieval and in artificial intelligence to relate items to each other. Madapena et al. [59] consider an overall assessment with other variables than merely agreement. The identification of these variables, their values, and their weightings remain to be determined in a cost function assessing the elicited symbol as a whole. This cost function, based on the Structured Equation Modelling, exhibits the capability of model-based optimization of gesture experience based on selected variables, such as agreement but not only (*e.g.,* agreement AND (fatigue OR memorability)).

Restarting from the context of use and reproducibility

Some GES are stripped from the context of use and the conditions in which the experiment took place, thus leaving the correct application of its results uncertain for the designer. To restore this importance, any GES should explicitly define this context (*e.g.,* user, device, environment) so that any repetition for the same context or any variation along any contextual dimension would facilitate the establishing of relationships among GES: by repetition (same context), by generalization (a dimension is more general than a previous one), by restriction (a dimension is more specific that a previous one), by dependency. Typical contextual variations include: conducting GES simultaneously for multiple *users* (*e.g.,* comparing sighted vs. blind users for the same functions [47], how culture influence gestures for internationalization [30,100]), multiple *platforms/devices/sensors* (*e.g.,* measure the consistency of gestures elicited by one participant for one referent on many devices [119], designing consistent gestures across several devices [26]), and in multiple environments (*e.g.,* eliciting finger gestures while standing vs. running [10]). Non-traditional devices, such as drones [94], intelligent textiles and clothes, knitted capacitive touch interfaces, shape-changing interfaces [68], particularly foldable, bendable, rollable displays, should be also investigated. For any very new device, a referent-free elicitation could be first conducted, especially when no gesture taxonomy already exists for this device, for example a squeezable interactive cushion. No GES exists for non-classical environments at home (*e.g.,* a kitchen, a bathroom, a sleeping room) and their furniture (*e.g.,* a table, a chair, a cabinet, a lamp), at work (*e.g.,* collaborative gestures in a corporate environment with analysis of spatial zones [86]), at leisure (*e.g.,* a sport room, a movie theater, outdoor sports). Thanks to computer vision, gestures can be elicited on any interaction surface beyond surface computing: floors (for foot-based interaction), walls, ceiling, windows [101], doors and their handles, etc. An explicit definition of the GES context of use is the cornerstone for reproducible research [41], defined by ACM via three badges: *repeatability* (same team, same experimental setup), *replicability* (different team with the same experimental setup), and *reproducibility* (different team and a different experimental setup). We found only one GES for replicability [103] and one for reproducibility [75], which seems very little in regard to the empirical validation found for other aspects. The same question arises for any modification of the GES procedure, such as in [45,71]. Software for conducting a GES distributed in time and space also appears (*e.g.,* [2,61] for conducting studies, [3] for analyzing the results via the crowd): participants propose symbols in their usual contexts of use, but no study exists that investigates the usage and the impact of such tools on a large scale.

Revisiting the elicitation procedure

Presentation of referents is independent of the experimenter, who is less likely to ask questions that may affect the participant. While examining the referent representation, the participant is concentrated on coming up with a proposal, and the experiment produces information that is grounded in the phenomenology of the subject [43]: "A photograph, a literal rendering of an element of the subject's world, calls forth associations, definitions, or ideas that would otherwise go unnoticed." Instead of providing the participant with predefined referents, participants could produce them and bring them in the experiment, thus making the experiment more natural to them. *Photo-elicitation* [11] encompasses a wide use of graphical material, such as pictures, but how and when that material is exploited varies among users and even studies: *photovoice*, *photonarrative*, *photoessay*, *photofeedback*, and *photointerview*. Photoessay [108] is typically a participatory design method where participants take their own photographs and arrange them with a scenario that make sense for them. This method involves participants more deeply as they are continually defining their context and expressing to the experimenter about how they behave in such a context. These methods empower the end users so that they can tell their stories and interpret their own user experience, culture or any other aspect to the experimenter instead of the other way around.

CONCLUSION AND FUTURE WORK

In this paper, we conducted a systematic literature review of the whole body of knowledge related to GES, resulting into a corpus of N=216 gesture elicitation studies conducted with 5,458 participants and 3,625 referents, eliciting a grand total of 148,340 individual gestures, from which 10,752 were agreed upon. We highlighted several results regarding the descriptive, comparative, and generative virtues to provide researchers and practitioners with concrete information on how to exploit and to further investigate this huge body of knowledge. We give access to our online collection along with its visualization tool to enable any interested party to perform any other searching, analysis, or other computations on the same corpus. We also refer to our web site for accessing the various source and result documents of this SLR, such as the base of initial documents, the spreadsheet containing the comparison of the 216 GES, and the full range of graphics. We are planning to continuously update this SLR to keep the corpus up-to-date.

REFERENCES

[1] Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. 2012. Putting Your Best Foot Forward: Investigating Real-world Mappings for Foot-based Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1229–1238. DOI:http://dx.doi.org/10.1145/2207676.2208575

[2] Abdullah X. Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. 2018. Crowdsourcing Similarity Judgments for Agreement Analysis in End-User Elicitation Studies. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. ACM, New York, NY, USA, 177–188. DOI:http://dx.doi.org/10.1145/3242587.3242621

[3] Abdullah X. Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. 2019. Crowdlicit: A System for Conducting Distributed End-User Elicitation and Identification Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 255:1–255:12. DOI: http://dx.doi.org/10.1145/3290605.3300485

[4] Leonardo Angelini, Francesco Carrino, Stefano Carrino, Maurizio Caon, Omar Abou Khaled, Jürgen Baumgartner, Andreas Sonderegger, Denis Lalanne, and Elena Mugellini. 2014. Gesturing on the Steering Wheel: A User-elicited Taxonomy. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '14)*. ACM, New York, NY, USA, 31:1–31:8. DOI: http://dx.doi.org/10.1145/2667317.2667414

[5] Leonardo Angelini, Denis Lalanne, Elise van den Hoven, Omar Abou Khaled, and Elena Mugellini. 2015. Move, Hold and Touch: A Framework for Tangible Gesture Interactive Systems. *Machines* 3, 3 (Aug. 2015), 173–207. DOI: http://dx.doi.org/10.3390/machines3030173

[6] Ilhan Aslan, Tabea Schmidt, Jens Woehrle, Lukas Vogel, and Elisabeth André. 2018. Pen + Mid-Air Gestures: Eliciting Contextual Gestures. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*. ACM, New York, NY, USA, 135–144. DOI: http://dx.doi.org/10.1145/3242969.3242979

[7] Gilles Bailly, Thomas Pietrzak, Jonathan Deber, and Daniel J. Wigdor. 2013. MéTamorphe: Augmenting Hotkey Usage with Actuated Keys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 563–572. DOI: http://dx.doi.org/10.1145/2470654.2470734

[8] Michel Beaudouin-Lafon. 2004. Designing Interaction, Not Interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)*. ACM, New York, NY, USA, 15–22. DOI: http://dx.doi.org/10.1145/989863.989865

[9] Joeran Beel, Bela Gipp, Stefan Langer, and Corinna Breitinger. 2016. Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries* 17, 4 (01 Nov 2016), 305–338. DOI: http://dx.doi.org/10.1007/s00799-015-0156-0

[10] Roger Boldu, Alexandru Dancu, Denys J.C. Matthies, Pablo Gallego Cascón, Shanaka Ransir, and Suranga Nanayakkara. 2018. Thumb-In-Motion: Evaluating Thumb-to-Ring Microgestures for Athletic Activity. In *Proceedings of the Symposium on Spatial User Interaction (SUI '18)*. ACM, New York, NY, USA, 150–157. DOI: http://dx.doi.org/10.1145/3267782.3267796

[11] Michael L. Boucher. 2018. *Participant Empowerment Through Photo-elicitation in Ethnographic Education. Research, New Perspectives and Approaches*. Springer International Publishing, Cham, Switzerland. DOI: http://dx.doi.org/10.1007/978-3-319-64413-4

[12] Pearl Brereton, Barbara A. Kitchenham, David Budgen, Mark Turner, and Mohamed Khalil. 2007. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software* 80, 4 (2007), 571 – 583. DOI: http://dx.doi.org/10.1016/j.jss.2006.07.009

[13] Daniel Buschek, Bianka Roppelt, and Florian Alt. 2018. Extending Keyboard Shortcuts with Arm and Wrist Rotation Gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 21, 12 pages. DOI: http://dx.doi.org/10.1145/3173574.3173595

[14] Francesco Cafaro, Leilah Lyons, and Alissa N. Antle. 2018. 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 (CHI '18)*. ACM, New York, NY, USA, 593:1–593:12. DOI:http://dx.doi.org/10.1145/3173574.3174167

[15] Gaëlle Calvary, Joëlle Coutaz, David Thevenin, Quentin Limbourg, Laurent Bouillon, and Jean Vanderdonckt. 2003. A Unifying Reference Framework for multi-target user interfaces. *Interacting with Computers* 15, 3 (2003), 289–308. DOI: http://dx.doi.org/10.1016/S0953-5438(03)00010-9

[16] John M. Carroll. 1982. Learning, using and designing filenames and command paradigms. *Behaviour & Information Technology* 1, 4 (1982), 327–346. DOI: http://dx.doi.org/10.1080/01449298208914457

[17] Jessica R. Cauchard, Jane L. E, Kevin Y. Zhai, and James A. Landay. 2015. Drone & Me: An Exploration into Natural Human-drone Interaction. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 361–365. DOI: http://dx.doi.org/10.1145/2750858.2805823

[18] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. 2016. User Elicitation on Single-hand Microgestures *(CHI ’16)*. ACM, New York, NY, USA, 3403–3414. DOI: http://dx.doi.org/10.1145/2858036.2858589

[19] Xiang ’Anthony’ Chen, Julia Schwarz, Chris Harrison, Jennifer Mankoff, and Scott E. Hudson. 2014. Air+Touch: Interweaving Touch + In-air Gestures. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST ’14)*. ACM, New York, NY, USA, 519–525. DOI: http://dx.doi.org/10.1145/2642918.2647392

[20] Zhen Chen, Xiaochi Ma, Zeya Peng, Ying Zhou, Mengge Yao, Zheng Ma, Ci Wang, Zaifeng Gao, and Mowei Shen. 2018. User-Defined Gestures for Gestural Interaction: Extending from Hands to Other Body Parts. *International Journal of Human–Computer Interaction* 34, 3 (March 2018), 238–250. DOI: http://dx.doi.org/10.1080/10447318.2017.1342943

[21] Sabrina Connell, Pei-Yi Kuo, Liu Liu, and Anne Marie Piper. 2013. A Wizard-of-Oz Elicitation Study Examining Child-defined Gestures with a Whole-body Interface. In *Proceedings of the 12th International Conference on Interaction Design and Children (IDC ’13)*. ACM, New York, NY, USA, 277–280. DOI: http://dx.doi.org/10.1145/2485760.2485823

[22] Joëlle Coutaz, Laurence Nigay, Daniel Salber, Ann Blandford, Jon May, and Richard M. Young. 1995. *Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE Properties*. Springer US, Boston, MA, 115–120. DOI: http://dx.doi.org/10.1007/978-1-5041-2896-4_19

[23] Suranjith De Silva, Michael Barlow, and Adam Easton. 2013. Harnessing Multi-user Design and Computation to Devise Archetypal Whole-of-body Gestures: A Novel Framework. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (OzCHI ’13)*. ACM, New York, NY, USA, 85–94. DOI: http://dx.doi.org/10.1145/2541016.2541020

[24] Linda Di Geronimo, Marica Bertarini, Julia Badertscher, Maria Husmann, and Moira C. Norrie. 2017. 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 (MobileHCI ’17)*. ACM, New York, NY, USA, 35:1–35:11. DOI: http://dx.doi.org/10.1145/3098279.3098530

[25] Nem Khan Dim, Chaklam Silpasuwanchai, Sayan Sarcar, and Xiangshi Ren. 2016. 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 (DIS ’16)*. ACM, New York, NY, USA, 204–214. DOI: http://dx.doi.org/10.1145/2901790.2901834

[26] Tilman Dingler, Rufat Rzayev, Alireza Sahami Shirazi, and Niels Henze. 2018. 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 (CHI ’18)*. ACM, New York, NY, USA, 419:1–419:12. DOI: http://dx.doi.org/10.1145/3173574.3173993

[27] Haiwei Dong, A. Danesh, Nadia Figueroa, and Abdulmotaleb. Saddik. 2015a. An Elicitation Study on Gesture Preferences and Memorability Toward a Practical Hand-Gesture Vocabulary for Smart Televisions. *IEEE Access* 3 (2015), 543–555. DOI: http://dx.doi.org/10.1109/ACCESS.2015.2432679

[28] Haiwei Dong, Nadia Figueroa, and Abdulmotaleb El Saddik. 2015b. An Elicitation Study on Gesture Attitudes and Preferences Towards an Interactive Hand-Gesture Vocabulary. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM ’15)*. ACM, New York, NY, USA, 999–1002. DOI: http://dx.doi.org/10.1145/2733373.2806385

[29] Ted Dunning. 1993. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics* 19, 1 (March 1993), 61–74. http://dl.acm.org/citation.cfm?id=972450.972454

[30] Jane L. E, Ilene L. E, James A. Landay, and Jessica R. Cauchard. 2017. Drone & Wo: Cultural Influences on Human-Drone Interaction Techniques. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ’17)*. ACM, New York, NY, USA, 6794–6799. DOI: http://dx.doi.org/10.1145/3025453.3025755

[31] Peter Emerson. 2013. The original Borda count and partial voting. *Social Choice and Welfare* 40, 2 (01 Feb 2013), 353–358. https://doi.org/10.1007/s00355-011-0603-9

[32] Hessam Jahani Fariman, Hasan J. Alyamani, Manolya Kavakli, and Len Hamey. 2016. Designing a User-defined Gesture Vocabulary for an In-vehicle Climate Control System. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction (OzCHI ’16)*. ACM, New York, NY, USA, 391–395. DOI: http://dx.doi.org/10.1145/3010915.3010955

[33] Yasmin Felberbaum and Joel Lanir. 2016. Step by Step: Investigating Foot Gesture Interaction. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI ’16)*. ACM, New York, NY, USA, 306–307. DOI: http://dx.doi.org/10.1145/2909132.2926057

[34] Yasmin Felberbaum and Joel Lanir. 2018. Better Understanding of Foot Gestures: An Elicitation Study. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI ’18)*. ACM, New York, NY, USA, 334:1–334:12. DOI: http://dx.doi.org/10.1145/3173574.3173908

[35] Leah Findlater, Ben Lee, and Jacob Wobbrock. 2012. Beyond QWERTY: Augmenting Touch Screen Keyboards with Multi-touch Gestures for Non-alphanumeric Input *(CHI ’12)*. ACM, New York, NY, USA, 2679–2682. DOI: http://dx.doi.org/10.1145/2207676.2208660

[36] Jon Fraenkel and Bernard Grofman. 2014. The Borda Count and its real-world alternatives: Comparing scoring rules in Nauru and Slovenia. *Australian Journal of Political Science* 49, 2 (2014), 186–205. DOI:http://dx.doi.org/10.1080/10361146.2014.900530

[37] Euan Freeman, Gareth Griffiths, and Stephen A. Brewster. 2017. Rhythmic Micro-gestures: Discreet Interaction On-the-go. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI 2017)*. ACM, New York, NY, USA, 115–119. DOI:http://dx.doi.org/10.1145/3136755.3136815

[38] Mathias Frisch, Jens Heydekorn, and Raimund Dachselt. 2009. Investigating Multi-touch and Pen Gestures for Diagram Editing on Interactive Surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS ’09)*. ACM, New York, NY, USA, 149–156. DOI: http://dx.doi.org/10.1145/1731903.1731933

[39] Bogdan-Florin Gheran, Jean Vanderdonckt, and Radu-Daniel Vatavu. 2018a. Gestures for Smart Rings: Empirical Results, Insights, and Design Implications. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS ’18)*. ACM, New York, NY, USA, 623–635. DOI: http://dx.doi.org/10.1145/3196709.3196741

[40] Bogdan-Florin Gheran, Radu-Daniel Vatavu, and Jean Vanderdonckt. 2018b. Ring x2: Designing Gestures for Smart Rings Using Temporal Calculus. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems (DIS ’18 Companion)*. ACM, New York, NY, USA, 117–122. DOI:http://dx.doi.org/10.1145/3197391.3205422

[41] Steven N. Goodman, Daniele Fanelli, and John P. A. Ioannidis. 2016. What does research reproducibility mean? *Science Translational Medicine* 8, 341 (2016), 341ps12–341ps12. DOI: http://dx.doi.org/10.1126/scitranslmed.aaf5027

[42] Robin Guérit, Alessandro Cierro, Jean Vanderdonckt, and Jorge Luis Pérez-Medina. 2019. Gesture Elicitation and Usability Testing for an Armband Interacting with Netflix and Spotify. In *Information Technology and Systems (Advances in Intelligent Systems and Computing)*, Álvaro Rocha, Carlos Ferras, and Manolo Paredes (Eds.). Springer International Publishing, 625–637. DOI: http://dx.doi.org/10.1007/978-3-030-11890-7_60

[43] Douglas Harper. 1988. Visual sociology: Expanding sociological vision. *The American Sociologist* 19, 1 (01 Mar 1988), 54–70. DOI: http://dx.doi.org/10.1007/BF02692374

[44] Hayati Havlucu, Mehmet Yarkin Ergin, İdil Bostan, Oğuz Turan Buruk, Tilbe Göksun, and Oğuzhan Özcan. 2017. It Made More Sense: Comparison of User-Elicited On-skin Touch and Freehand Gesture Sets. In *Distributed, Ambient and Pervasive Interactions (Lecture Notes in Computer Science)*. Springer, Cham, 159–171. DOI: http://dx.doi.org/10.1007/978-3-319-58697-7_11

[45] Lynn Hoff, Eva Hornecker, and Sven Bertel. 2016. 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 (TEI ’16)*. ACM, New York, NY, USA, 86–91. DOI: http://dx.doi.org/10.1145/2839462.2839472

[46] Hessam Jahani-Fariman. 2017. Developing a User-defined Interface for In-vehicle Mid-air Gestural Interactions. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion (IUI ’17 Companion)*. ACM, New York, NY, USA, 165–168. DOI: http://dx.doi.org/10.1145/3030024.3038277

[47] Shaun K. Kane, Jacob O. Wobbrock, and Richard E. Ladner. 2011. Usable Gestures for Blind People: Understanding Preference and Performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’11)*. ACM, New York, NY, USA, 413–422. DOI: http://dx.doi.org/10.1145/1978942.1979001

[48] Frederic Kerber, Markus Löchtefeld, Antonio Krüger, Jess McIntosh, Charlie McNeill, and Mike Fraser. 2016. Understanding Same-Side Interactions with Wrist-Worn Devices. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI ’16)*. ACM, New York, NY, USA, 28:1–28:10. DOI: http://dx.doi.org/10.1145/2971485.2971519

[49] Barbara Kitchenham, Rialette Pretorius, David Budgen, O. Pearl Brereton, Mark Turner, Mahmood Niazi, and Stephen Linkman. 2010. Systematic literature reviews in software engineering – A tertiary study. *Information and Software Technology* 52, 8 (2010), 792–805. DOI: http://dx.doi.org/10.1016/j.infsof.2010.03.006

[50] Anne Köpsel and Nikola Bubalo. 2015. Benefiting from Legacy Bias. *interactions* 22, 5 (Aug. 2015), 44–47. DOI:http://dx.doi.org/10.1145/2803169

[51] Christine Kühnel, Tilo Westermann, Fabian Hemmert, Sven Kratz, Alexander Müller, and Sebastian Möller. 2011. I’m home: Defining and evaluating a gesture set for smart-home control. *International Journal of Human-Computer Studies* 69, 11 (Oct. 2011), 693–704. DOI:http://dx.doi.org/10.1016/j.ijhcs.2011.04.005

[52] DoYoung Lee, Youryang Lee, Yonghwan Shin, and Ian Oakley. 2018. Designing Socially Acceptable Hand-to-Face Input. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. ACM, New York, NY, USA, 711–723. DOI: http://dx.doi.org/10.1145/3242587.3242642

[53] Sang-Su Lee, Sohyun Kim, Bopil Jin, Eunji Choi, Boa Kim, Xu Jia, Daeeop Kim, and Kun-pyo Lee. 2010. How Users Manipulate Deformable Displays As Input Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1647–1656. DOI: http://dx.doi.org/10.1145/1753326.1753572

[54] Hoo Yong Leng, Noris Mohd Norowi, and Azrul Hazri Jantan. 2017. A User-Defined Gesture Set for Music Interaction in Immersive Virtual Environment. In *Proceedings of the 3rd International Conference on Human-Computer Interaction and User Experience in Indonesia (CHIuXiD '17)*. ACM, New York, NY, USA, 44–51. DOI: http://dx.doi.org/10.1145/3077343.3077348

[55] Xuan Li, Daisong Guan, Jingya Zhang, Xingtong Liu, Siqi Li, and Hui Tong. 2019. Exploration of Ideal Interaction Scheme on Smart TV: Based on User Experience Research of Far-Field Speech and Mid-air Gesture Interaction. In *Design, User Experience, and Usability. User Experience in Advanced Technological Environments (Lecture Notes in Computer Science)*, Aaron Marcus and Wentao Wang (Eds.). Springer International Publishing, 144–162. DOI: http://dx.doi.org/10.1007/978-3-030-23541-3_12

[56] Alessandro Liberati, Douglas G. Altman, Jennifer Tetzlaff, Cynthia Mulrow, Peter C. Gø tzsche, John P.A. Ioannidis, Mike Clarke, P. J. Devereaux, Jos Kleijnen, and David Moher. July 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine* 6, 7 (July 2009), 1–22. DOI: http://dx.doi.org/10.1371/journal.pmed.1000100

[57] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle Arms-down Mid-air Interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 63–71. DOI: http://dx.doi.org/10.1145/2807442.2807489

[58] Yihua Lou, Wenjun Wu, Radu-Daniel Vatavu, and Wei-Tek Tsai. 2017. Personalized gesture interactions for cyber-physical smart-home environments. *Science China Information Sciences* 60, 7 (July 2017), 072104. DOI:http://dx.doi.org/10.1007/s11432-015-1014-7

[59] Naveen Madapana, Glebys Gonzalez, Richard Rodgers, Lingsong Zhang, and Juan P. Wachs. 2018. Gestures for Picture Archiving and Communication Systems (PACS) operation in the operating room: Is there any standard? *PLOS ONE* 13, 6 (June 2018), e0198092. DOI:http://dx.doi.org/10.1371/journal.pone.0198092

[60] Nathan Magrofuoco, Jorge Luis Pérez-Medina, Paolo Roselli, Jean Vanderdonckt, and Santiago Villarreal. 2019. Eliciting Contact-Based and Contactless Gestures With Radar-Based Sensors. *IEEE Access* 7 (2019), 176982–176997. DOI: http://dx.doi.org/10.1109/ACCESS.2019.2951349

[61] Nathan Magrofuoco and Jean Vanderdonckt. 2019. Gelicit: A Cloud Platform for Distributed Gesture Elicitation Studies. *Proc. ACM Hum.-Comput. Interact.* 3, EICS (June 2019), 6:1–6:41. DOI: http://dx.doi.org/10.1145/3331148

[62] Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. 2012. Eye-based Head Gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 139–146. DOI: http://dx.doi.org/10.1145/2168556.2168578

[63] Francisco J. Martínez-Ruiz, Sebastian F. Rauh, and Gerrit Meixner. 2020. Understanding Peripheral Audiences: From Subtle to Full Body Gestures. In *Human Interaction and Emerging Technologies (Advances in Intelligent Systems and Computing)*, Tareq Ahram, Redha Taiar, Serge Colson, and Arnaud Choplin (Eds.). Springer International Publishing, 489–495. DOI: http://dx.doi.org/10.1007/978-3-030-25629-6_76

[64] Kohei Matsumura. 2015. Studying User-Defined Gestures Toward Off the Screen Interactions. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (ITS '15)*. ACM, New York, NY, USA, 295–300. DOI: http://dx.doi.org/10.1145/2817721.2823496

[65] Dan Mauney, Jonathan Howarth, Andrew Wirtanen, and Miranda Capra. 2010. Cultural Similarities and Differences in User-defined Gestures for Touchscreen User Interfaces. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 4015–4020. DOI: http://dx.doi.org/10.1145/1753846.1754095

[66] Keenan R. May, Thomas M. Gable, and Bruce N. Walker. 2017. 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 (AutomotiveUI '17)*. ACM, New York, NY, USA, 74–83. DOI: http://dx.doi.org/10.1145/3122986.3123015

[67] Erin McAweeney, Haihua Zhang, and Michael Nebeling. 2018. User-Driven Design Principles for Gesture Representations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 547, 13 pages. DOI: http://dx.doi.org/10.1145/3173574.3174121

[68] John C. McClelland, Robert J. Teather, and Audrey Girouard. 2017. Haptobend: Shape-changing Passive Haptic Feedback in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 82–90. DOI: http://dx.doi.org/10.1145/3131277.3132179

[69] David McNeill. August-December 1981. Action, Thought and Language. *Cognition* 10, 1-3 (August-December 1981), 201–208.

[70] Meredith Ringel Morris. 2012. Web on the Wall: Insights from a Multimodal Interaction Elicitation Study. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 95–104. DOI: http://dx.doi.org/10.1145/2396636.2396651

[71] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing Legacy Bias in Gesture Elicitation Studies. *interactions* 21, 3 (May 2014), 40–45. DOI:http://dx.doi.org/10.1145/2591689

[72] Meredith Ringel Morris, Jacob O. Wobbrock, and Andrew D. Wilson. 2010. Understanding Users' Preferences for Surface Gestures. In *Proceedings of Graphics Interface 2010 (GI '10)*. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 261–268. http://dl.acm.org/citation.cfm?id=1839214.1839260

[73] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-designed and User-defined Gesture Sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1099–1108. DOI: http://dx.doi.org/10.1145/2470654.2466142

[74] Andrés Adolfo Navarro-Newball, Isidro Moreno, Edmond Prakash, Ali Arya, Victoria E. Contreras, Victor A. Quiceno, Santiago Lozano, Juan David Mejìa, and Diego Fernando Loaiza. 2016. Gesture based human motion and game principles to aid understanding of science and cultural practices. *Multimedia Tools and Applications* 75, 19 (Oct. 2016), 11699–11722. DOI: http://dx.doi.org/10.1007/s11042-015-2667-5

[75] Michael Nebeling, Alexander Huber, David Ott, and Moira C. Norrie. 2014. Web on the Wall Reloaded: Implementation, Replication and Refinement of User-Defined Interaction Sets. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14)*. ACM, New York, NY, USA, 15–24. DOI: http://dx.doi.org/10.1145/2669485.2669497

[76] Michael Nebeling, David Ott, and Moira C. Norrie. 2015. 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 (EICS '15)*. ACM, New York, NY, USA, 142–151. DOI: http://dx.doi.org/10.1145/2774225.2774846

[77] Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. 2004. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction*, Antonio Camurri and Gualtiero Volpe (Eds.). Springer, Berlin, 409–420. DOI: http://dx.doi.org/10.1007/978-3-540-24598-8_38

[78] F. R. Ortega, A. Galvan, K. Tarre, A. Barreto, N. Rishe, J. Bernal, R. Balcazar, and J. L. Thomas. 2017. Gesture elicitation for 3D travel via multi-touch and mid-Air systems for procedurally generated pseudo-universe. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI '17)*. 144–153. DOI: http://dx.doi.org/10.1109/3DUI.2017.7893331

[79] Mehdi Ousmer, Jean Vanderdonckt, and Sabin Buraga. 2019. An Ontology for Reasoning on Body-based Gestures. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '19)*. ACM, New York, NY, USA, 17:1–17:6. DOI: http://dx.doi.org/10.1145/3319499.3328238

[80] Tran Pham, Jo Vermeulen, Anthony Tang, and Lindsay MacDonald Vermeulen. 2018. Scale Impacts Elicited Gestures for Manipulating Holograms: Implications for AR Gesture Design. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 227–240. DOI: http://dx.doi.org/10.1145/3196709.3196719

[81] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013a. User-defined Gestures for Augmented Reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 955–960. DOI:http://dx.doi.org/10.1145/2468356.2468527

[82] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013b. User-Defined Gestures for Augmented Reality. In *Human-Computer Interaction - INTERACT 2013 (Lecture Notes in Computer Science)*. Springer, Berlin, Heidelberg, 282–299. DOI: http://dx.doi.org/10.1007/978-3-642-40480-1_18

[83] Marcus Pivato. 2015. Condorcet meets Bentham. *Journal of Mathematical Economics* 59 (2015), 58 – 65. DOI: http://dx.doi.org/10.1016/j.jmateco.2015.04.006

[84] F. Quek, D. McNeill, R. Ansari, Xin-Feng Ma, R. Bryll, S. Duncan, and K. E. McCullough. 1999. Gesture cues for conversational interaction in monocular video. In *International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings*. 119–126. DOI: http://dx.doi.org/10.1109/RATFG.1999.799234

[85] Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari. 2002. Multimodal Human Discourse: Gesture and Speech. *ACM Trans. Comput.-Hum. Interact.* 9, 3 (Sept. 2002), 171–193. DOI:http://dx.doi.org/10.1145/568513.568514

[86] Roman Rädle, Hans-Christian Jetter, Mario Schreiner, Zhihao Lu, Harald Reiterer, and Yvonne Rogers. 2015. Spatially-aware or Spatially-agnostic?: Elicitation and Evaluation of User-Defined Cross-Device Interactions *(CHI '15)*. ACM, New York, NY, USA, 3913–3922. DOI:http://dx.doi.org/10.1145/2702123.2702287

[87] Julie Rico and Stephen Brewster. 2010. Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 887–896. DOI: http://dx.doi.org/10.1145/1753326.1753458

[88] P. Riehmann, M. Hanfler, and B. Froehlich. 2005. Interactive Sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* 233–240. DOI: http://dx.doi.org/10.1109/INFVIS.2005.1532152

[89] Isabel Benavente Rodriguez and Nicolai Marquardt. 2017. Gesture Elicitation Study on How to Opt-in & Opt-out from Interactions with Public Displays *(ISS '17)*. ACM, New York, NY, USA, 32–41. DOI: http://dx.doi.org/10.1145/3132272.3134118

[90] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined Motion Gestures for Mobile Interaction *(CHI '11)*. ACM, New York, NY, USA, 197–206. DOI: http://dx.doi.org/10.1145/1978942.1978971

[91] Jaime Ruiz and Daniel Vogel. 2015. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue *(CHI '15)*. ACM, New York, NY, USA, 3347–3350. DOI: http://dx.doi.org/10.1145/2702123.2702583

[92] Ovidia Schipor and Radu-Daniel Vatavu. 2018. Invisible, Inaudible, and Impalpable: Users' Preferences and Memory Performance for Digital Content in Thin Air. *IEEE Pervasive Computing* 17, 4 (Oct. 2018), 76–85. DOI: http://dx.doi.org/10.1109/MPRV.2018.2873856

[93] Marcos Serrano, Barrett M. Ens, and Pourang P. Irani. 2014. Exploring the Use of Hand-to-face Input for Interacting with Head-worn Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3181–3190. DOI: http://dx.doi.org/10.1145/2556288.2556984

[94] Matthias Seuter, Eduardo Rodriguez Macrillante, Gernot Bauer, and Christian Kray. 2018. Running with Drones: Desired Services and Control Gestures. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction (OzCHI '18)*. ACM, New York, NY, USA, 384–395. DOI: http://dx.doi.org/10.1145/3292147.3292156

[95] Teddy Seyed, Chris Burns, Mario Costa Sousa, Frank Maurer, and Anthony Tang. 2012. Eliciting Usable Gestures for Multi-display Environments. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 41–50. DOI: http://dx.doi.org/10.1145/2396636.2396643

[96] Alex Shaw and Lisa Anthony. 2016a. Analyzing the Articulation Features of Children's Touchscreen Gestures. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016)*. ACM, New York, NY, USA, 333–340. DOI:http://dx.doi.org/10.1145/2993148.2993179

[97] Alex Shaw and Lisa Anthony. 2016b. Toward a Systematic Understanding of Children's Touchscreen Gestures. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, 1752–1759. DOI: http://dx.doi.org/10.1145/2851581.2892425

[98] Chaklam Silpasuwanchai and Xiangshi Ren. 2014. Jump and Shoot!: Prioritizing Primary and Alternative Body Gestures for Intense Gameplay *(CHI '14)*. ACM, New York, NY, USA, 951–954. DOI: http://dx.doi.org/10.1145/2556288.2557107

[99] Chaklam Silpasuwanchai and Xiangshi Ren. 2015. Designing concurrent full-body gestures for intense gameplay. *International Journal of Human-Computer Studies* 80 (Aug. 2015), 1–13. DOI: http://dx.doi.org/10.1016/j.ijhcs.2015.02.010

[100] Tiffanie R. Smith and Juan E. Gilbert. 2018. Dancing to Design: A Gesture Elicitation Study. In *Proceedings of the 17th ACM Conference on Interaction Design and Children (IDC '18)*. ACM, New York, NY, USA, 638–643. DOI: http://dx.doi.org/10.1145/3202185.3210790

[101] Nikita Soni, Schuyler Gleaves, Hannah Neff, Sarah Morrison-Smith, Shaghayegh Esmaeili, Ian Mayne, Sayli Bapat, Carrie Schuman, Kathryn A. Stofer, and Lisa Anthony. 2019. Do User-defined Gestures for Flatscreens Generalize to Interactive Spherical Displays for Adults and Children?. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays (PerDis '19)*. ACM, New York, NY, USA, 24:1–24:7. DOI: http://dx.doi.org/10.1145/3321335.3324941

[102] Maximilian Speicher and Michael Nebeling. 2018. GestureWiz: A Human-Powered Gesture Design Environment for User Interface Prototypes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 107, 11 pages. DOI: http://dx.doi.org/10.1145/3173574.3173681

[103] Poorna Talkad Sukumar, Anqing Liu, and Ronald Metoyer. 2018. Replicating User-defined Gestures for Text Editing. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces (ISS '18)*. ACM, New York, NY, USA, 97–106. DOI:http://dx.doi.org/10.1145/3279778.3279793

[104] Yanke Tan, Sang Ho Yoon, and Karthik Ramani. 2017. BikeGesture: User Elicitation and Performance of Micro Hand Gesture As Input for Cycling *(CHI EA '17)*. ACM, New York, NY, USA, 2147–2154. DOI: http://dx.doi.org/10.1145/3027063.3053075

[105] Florent Taralle, Alexis Paljic, Sotiris Manitsaris, Jordane Grenier, and Christophe Guettier. 2015. A Consensual and Non-ambiguous Set of Gestures to Interact with UAV in Infantrymen *(CHI EA '15)*. ACM, New York, NY, USA, 797–803. DOI: http://dx.doi.org/10.1145/2702613.2702971

[106] Piotr Tarka. 2018. An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & Quantity* 52, 1 (01 Jan 2018), 313–354. DOI: http://dx.doi.org/10.1007/s11135-017-0469-8

[107] Vuletic Tijana, Alex Duffy, Laura Hay, Chris McTeague, Gerard Campbell, and Madeleine Grealy. 2019. Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies* 129 (2019), 74 – 94. DOI: http://dx.doi.org/10.1016/j.ijhcs.2019.03.011

[108] Deborah J. Tippins, Sophia (Sun Kyung) Jeong, Purita P. Bilbao, and Lourdes N. Morano. 2018. *Participant Empowerment Through Photoelicitation in Ethnographic Education Research*. Springer International Publishing, Cham, Switzerland, 93–113.

[109] Theophanis Tsandilas. 2018. Fallacies of Agreement: A Critical Review of Consensus Assessment Methods for Gesture Elicitation. *ACM Trans. Comput.-Hum. Interact.* 25, 3, Article 18 (June 2018), 49 pages. DOI: http://dx.doi.org/10.1145/3182168

[110] Jean Vanderdonckt, Nathan Magrofuoco, Suzanne Kieffer, Jorge Pérez, Ysabelle Rase, Paolo Roselli, and Santiago Villarreal. 2019. Head and Shoulders Gestures: Exploring User-Defined Gestures with Upper Body. In *Design, User Experience, and Usability. User Experience in Advanced Technological Environments (Lecture Notes in Computer Science)*, Aaron Marcus and Wentao Wang (Eds.). Springer International Publishing, 192–213. DOI: http://dx.doi.org/10.1007/978-3-030-23541-3_15

[111] Radu-Daniel Vatavu. 2012. User-defined Gestures for Free-hand TV Control *(EuroITV '12)*. ACM, New York, NY, USA, 45–48. DOI: http://dx.doi.org/10.1145/2325616.2325626

[112] Radu-Daniel Vatavu. 2013. The Impact of Motion Dimensionality and Bit Cardinality on the Design of 3D Gesture Recognizers. *International Journal of Human-Computer Studies* 71, 4 (2013), 387–409. DOI: http://dx.doi.org/10.1016/j.ijhcs.2012.11.005

[113] Radu-Daniel Vatavu. 2019. The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 224:1–224:13. DOI: http://dx.doi.org/10.1145/3290605.3300454 event-place: Glasgow, Scotland Uk.

[114] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1325–1334. DOI: http://dx.doi.org/10.1145/2702123.2702223

[115] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2016. Between-Subjects Elicitation Studies: Formalization and Tool Support *(CHI '16)*. ACM, New York, NY, USA, 3390–3402. DOI: http://dx.doi.org/10.1145/2858036.2858228

[116] Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. 2014. Leap Gestures for TV: Insights from an Elicitation Study *(TVX '14)*. ACM, New York, NY, USA, 131–138. DOI:http://dx.doi.org/10.1145/2602299.2602316

[117] Dong-Bach Vo, Eric Lecolinet, and Yves Guiard. 2014. Belly Gestures: Body Centric Gestures on the Abdomen. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*. ACM, New York, NY, USA, 687–696. DOI: http://dx.doi.org/10.1145/2639189.2639210

[118] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. 2018. Gesture Elicitation Studies for Mid-Air Interaction: A Review. *Multimodal Technologies and Interaction* 2, 4 (2018), 65–. DOI: http://dx.doi.org/10.3390/mti2040065

[119] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. 2019. Frame-Based Elicitation of Mid-Air Gestures for a Smart Home Device Ecosystem. *Informatics* 6, 2 (June 2019), 23. DOI: http://dx.doi.org/10.3390/informatics6020023

[120] Oleg Špakov and Päivi Majaranta. 2012. Enhanced Gaze Interaction Using Simple Head Gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. ACM, New York, NY, USA, 705–710. DOI: http://dx.doi.org/10.1145/2370216.2370369

[121] Markus L. Wittorf and Mikkel R. Jakobsen. 2016. Eliciting Mid-Air Gestures for Wall-Display Interaction *(NordiCHI '16)*. ACM, New York, NY, USA, 3:1–3:4. DOI: http://dx.doi.org/10.1145/2971485.2971503

[122] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1869–1872. DOI:http://dx.doi.org/10.1145/1056808.1057043

[123] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing *(CHI '09)*. ACM, New York, NY, USA, 1083–1092. DOI: http://dx.doi.org/10.1145/1518701.1518866

[124] Jacob O. Wobbrock, Brad A. Myers, and John A. Kembel. 2003. EdgeWrite: A Stylus-based Text Entry Method Designed for High Accuracy and Stability of Motion. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)*. ACM, New York, NY, USA, 61–70. DOI: http://dx.doi.org/10.1145/964696.964703

[125] Huiyue Wu, Yu Wang, Jiayi Liu, Jiali Qiu, and Xiaolong (Luke) Zhang. 2019a. User-defined gesture interaction for in-vehicle information systems. *Multimedia Tools and Applications* (Aug. 2019). DOI: http://dx.doi.org/10.1007/s11042-019-08075-1

[126] Huiyue Wu, Liuqingqing Yang, Shengqian Fu, and Xiaolong (Luke) Zhang. 2019b. Beyond remote control: Exploring natural gesture inputs for smart TV systems. *Journal of Ambient Intelligence and Smart Environments* 11, 4 (Jan. 2019), 335–354. DOI: http://dx.doi.org/10.3233/AIS-190528

[127] Ying Yin and Randall Davis. 2013. Gesture Spotting and Recognition Using Salience Detection and Concatenated Hidden Markov Models. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI '13)*. ACM, New York, NY, USA, 489–494. DOI: http://dx.doi.org/10.1145/2522848.2532588

[128] Ionuţ-Alexandru Zaiţi, Ştefan-Gheorghe Pentiuc, and Radu-Daniel Vatavu. 2015. On free-hand TV control: experimental results on user-elicited gestures with Leap Motion. *Personal and Ubiquitous Computing* 19, 5-6 (Aug. 2015), 821–838. DOI: http://dx.doi.org/10.1007/s00779-015-0863-y

[129] Xiaojie Zha and Marie-Luce Bourguet. 2016. Experimental Study to Elicit Effective Multimodal Behaviour in Pedagogical Agents *(DAA '16)*. ACM, New York, NY, USA, 1:1–1:6. DOI: http://dx.doi.org/10.1145/3005338.3005339