IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 20 March 2024, accepted 28 March 2024, date of publication 10 April 2024, date of current version 14 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3387269

TOPICAL REVIEW

Exploring Methods to Optimize Gesture Elicitation Studies: A Systematic Literature Review

YUTING CHENG [1], ZHANWEI WU [1], AND RUOWEI XIAO [2], (Member, IEEE)
[1] School of Design, Shanghai Jiao Tong University, Shanghai 200241, China
[2] School of Design, Southern University of Science and Technology, Shenzhen 518055, China

Corresponding author: Ruowei Xiao (xiaorw@sustech.edu.cn)

ABSTRACT Gesture elicitation is a fundamental method for constructing gesture-based interactive systems. Existing research has shown that this method allows users' knowledge to serve as a design basis, minimizing biases that may occur from relying excessively on an expert's preference. However, how to conduct effective gesture elicitation studies remains an uncertain and open-ended issue. Through a thorough literature review, this study aims to identify important aspects that have significant impacts on the implementation of the elicitation study. The main findings of this paper are as follows: (1) Factors such as the choice of participants, referents, elicitation techniques, elicitation environments, and tasks can have a significant impact on the results of gesture elicitation; (2) Researchers have proposed various metrics for selecting gesture sets and evaluating their effects, but there is still debate over the sufficiency of these metrics; (3) Various methods have been proposed to improve gesture elicitation research, but there is still a lack of broad consensus on the best practices for gesture elicitation research, how to evaluate and report results, and how to compare results across studies. These findings indicate that gesture elicitation has become a valuable research and practical tool, but further study is needed to utilize it better.

INDEX TERMS Elicitation, gesture, human-computer interaction, participatory design, research method, systematic literature review.

I. INTRODUCTION

Gesture Elicitation Study (GES) is a user elicitation method specifically aimed at gesture design. Initially proposed and practiced by Wobbrock et al. [1], GES requires participants to suggest gestures related to specific referents (the effects of an action). This form of research has quickly become a popular tool, extensively used in the design of gesture interactions for mobile devices [2], smart homes [3], [4], autonomous driving [5], [6], and extended reality technologies [7], [8], [9].

The value of GES lies in its ability to enable researchers to gain a deep understanding of users' capabilities and preferences. By removing the execution gap from the system, it allows participants to interact with the system in any way they choose, meaning that the end users are directly involved in the design process [10]. Through such research, researchers can observe unrestricted interactions between users and emerging technologies, leading to the generation of a consensus set for targeted interaction behaviors from users [11].

Although GES has shown great potential in practice, current literature indicates that there are still research gaps remain unsolved, particularly the lack of a recognized best practice process. As a result, many researchers have proposed modifications and improvements based on their specific needs [12], including but not limited to: changes to the elicitation process [13], the display of referents [12], elicitation environment [7], evaluation metrics for elicitation [14], and assessment of elicitation effects [15]. This article attempts to summarize the latest research progress through a systematic literature review, to assist HCI researchers and designers in better designing and conducting gesture elicitation studies, or further developing this useful tool.

Upon conducting a literature search, we identified four review papers focused on gesture elicitation studies. Villarreal-Narvaez et al. [16] surveyed 216 studies up to

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

the year 2019 to address questions related to the application domains of gesture elicitation, such as: What are the main meta-data of GES? In what ways are two or more GES similar? Which areas require new or more GES? Subsequently, in 2024, Villarreal-Narvaez et al. [17] expanded the review scope and examined the corpus of 267 studies up to 2021 to provide descriptive, comparative and generative analysis about many aspects of GES research, such as: distribution, involved body parts, referents, gesture datasets, indicators, terminologies. Tsandilas [18] investigated studies up to 2018 (the number of studies is unspecified), primarily focusing on consensus calculation methods, identifying some issues with existing methods, and suggesting improvements. Vogiatzidakis and Koutsabasis [19] reviewed 47 studies up to 2018, covering aspects of gesture elicitation research such as application domains, technological maturity of the systems at hand, basic process, the dimensions of appropriateness, the profile of participants, gesture evaluation, and data analysis. Some conclusions of the above reviews are consistent with the findings of this article, such as: GES research and application show a growing trend; the most frequently reported measures of agreement in GES research is AR (agreement rate); the process of GES research is gradually standardized, but there is no universally accepted best process. However, the questions listed below have still not been clearly answered in previous reviews and therefore become the focus of this article:

RQ1: What is the typical process of Gesture Elicitation Studies?

RQ2: What are the methods, metrics, pros and cons of gesture consensus analysis?

RQ3: What factors influence the final outcomes of Gesture Elicitation Studies? Is there empirical evidence?

RQ4: How can the quality of gesture sets be evaluated?

The remaining sections of this article are organized as follows. Section two provides an overview of the paper selection process and criteria. Section three presents the findings and results of the investigation into the aforementioned research questions. Section four further discusses these research findings. Finally, conclusions are drawn in section five.

This article is an extension of the author's The 11th International Conference in Software Engineering Research and Innovation conference report "Factors Affecting the Results of Gesture Elicitation: A Review [20]". The conference report introduces the progress of GES research until 2023 and provides a preliminary discussion of RQ3. This article supplements the papers published in 2023-2024 that are not included in the conference report and then discusses three issues not covered in the conference report, namely: RQ1, RQ2, and RQ4, and provides a more complete analysis and explanation of RQ3.

II. METHOD

Systematic Literature Review (SLR) is a methodological research approach aimed at formally synthesizing the primary studies in a specific field through well-defined steps [21], usually including identification, screening, eligibility, and inclusion [22]. Fig. 1 illustrates the flowchart of the Literature Search and Evaluation for Inclusion process in this paper [23].

Step 1: Identification

Channels for Literature Search: We selected the following electronic databases for literature retrieval: (1) IEEE Xplore, (2) ACM Digital Library, (3) Springer Link, (4) Elsevier ScienceDirect, and (5) Engineering Village. These online digital libraries are primary sources for accessing literature in the field of human-computer interaction.

Keywords Used for the Search: We conducted the search using the following keywords: Q = "elicitation" AND "gesture", to retrieve more comprehensive and inclusive results. To facilitate the traceability of our literature search, and to periodically repeat the search in the same databases and sources to identify any new materials that may have emerged since the initial search, we set December 29, 2023, as the final day for literature retrieval. We have exported the reference list accordingly.

Refining Results With Additional Restrictions: We have chosen 2005 as the starting year, as it was the year when Wobbrock et al. popularized the elicitation method in Human-Computer Interaction (HCI) research. To avoid excessive and irrelevant search results, we conducted advanced searches limited to titles, keywords, and abstracts of all literature from 2005 to 2023. From the results obtained from these five databases, we identified a total of 1,283 references. Among them, 415 references were duplicates. We then further refined the remaining 868 papers based on the selection criteria described in steps 2 and 3.

Step 2: Screening

We filtered out 868 results based on the following criteria: (1) excluding studies from the abstract that did not involve elicitation research, did not focus on gestures, or were unrelated to interactive systems; (2) excluding papers that were not peer-reviewed research papers. As a result, we excluded 492 irrelevant references through the screening process, resulting in a total of 376 papers.

Step 3: Eligibility

We further screened the remaining literature through a quick review of abstracts and full-text content using the following criteria: (1) excluding papers without a clear research conclusion or lacking evidence to support the conclusion or with an unclear description of the research process; (2) excluding papers that were unrelated to the research question were excluded. Based on these criteria, we excluded 298 papers and selected n=78 final papers for this study review.

Step 4: Inclusion

Regarding the final 78 papers (see Appendix), we qualitatively classified each paper based on different analytical dimensions, such as studies evaluating the effects of gesture elicitation, studies investigating factors influencing gesture elicitation effects, and studies exploring methods to improve gesture elicitation effects. Details can be found in the next section.

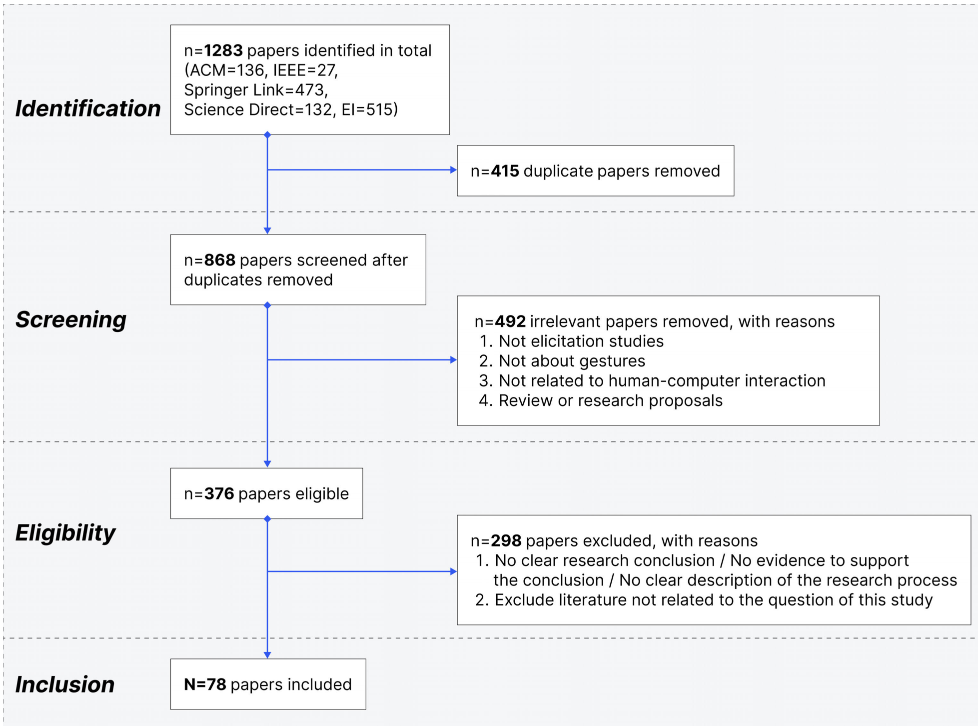


FIGURE 1. The four-phase flow diagram of SLR examination.

III. FINDINGS

In this section, we report on the trends, application domains, and basic characteristics of Gesture Elicitation Studies, as well as findings based on the research questions we reviewed.

A. DESCRIPTIVE ANALYSIS

1) TEMPORAL DISTRIBUTION

Research on Gesture Elicitation Methods is showing an upward trend, indicating an increasingly active field of study.

The majority of studies within this domain employ empirical research methods, which rely on actual observations, experiments, or survey data [24], showcasing the research community's preference for evidence-based approaches. In contrast, non-empirical research, which does not utilize quantitative data or scientific experimental designs, is less common in GES.

This survey encompasses papers published between 2005 and 2023, as indicated in Table 1, the first study proposing improvements to GES was published in 2013, followed by a general upward trend in the volume of related research publications. This developmental trend demonstrates the research community's ongoing pursuit in optimizing Gesture Elicitation Methods.

2) RESEARCH QUESTION

As an emerging area, the GES method encompasses a wide range of topics (Table 2), reflecting that there are still many aspects within this field that can be improved and further developed.

- *Research on Factors Affecting Gesture Elicitation Effects:* This type of research focuses on various factors in GES that may affect the quality and outcome of gestures as well as the efficiency of the experiment, such as the participants' gender, cultural background, knowledge level, task difficulty, and complexity.
- *Research on Methods for Improving Gesture Elicitation Effects:* How to design better gesture elicitation experiments to improve their effectiveness? For example, how to make participants understand and accept gesture elicitation more easily.
- *Research on Methods for Processing Gesture Data:* This area mainly focuses on how to process the collected gesture data to extract the best set of gestures.
- *Research on the Evaluation of Gesture Elicitation Outcomes:* This type of research focuses on the methods and metrics for evaluating the outcomes of gesture elicitation. For example, how to evaluate the differences in the effectiveness of gesture elicitation in different audiences, and whether gesture elicitation can help people better solve problems, learn, or remember.
- *Research on gesture elicitation toolkits:* This aspect mainly explores and designs tools and techniques that can effectively support gesture elicitation methods.

3) APPLICATION DOMAINS & RESEARCH OBJECT

As attention to improving the GES method continues to grow, its application scope and domains have become increasingly broad (as indicated in Table 3 & Table 4).

The primary application areas of gesture elicitation research have shifted from touch interactions like tablets and

TABLE 1. Number of papers examined within 2005–20231.

Year	Empirical research		Non-empirical research		Total (All)
	Total number of papers	Citation	Total number of papers	Citation	
2013	1	[S62]	0	/	1
2014	4	[S12][S31][S50][S77]	1	[S73]	5
2015	6	[S05][S08][S13][S46][S55][S56]	2	[S23][S42]	8
2016	2	[S45][S67]	1	[S10]	3
2017	6	[S11][S15][S18][S20][S48][S64]	0	/	6
2018	8	[S09][S14][S25][S34][S40][S43][S47][S75]	1	[S22]	9
2019	11	[S02][S04][S07][S17][S24][S27][S52][S53][S54][S65][S69]	0	/	11
2020	8	[S01][S19][S28][S37][S44][S57][S59][S70]	1	[S03]	9
2021	6	[S36][S63][S66][S72][S33][S78]	0	/	6
2022	4	[S38][S39][S41][S58]	4	[S29][S49][S51][S61]	8
2023	12	[S60][S71][S26][S06][S16][S32][S76][S74][S35][S30][S68][S21]	0	/	12

TABLE 2. Classification of research questions.

Research question	Number of Papers	%	Citation
Research on Factors Affecting Gesture Elicitation Effects	22	28.2%	[S04] [S07] [S08] [S15] [S16] [S34] [S36] [S37] [S38] [S40] [S41] [S45] [S47] [S50] [S53] [S54] [S57] [S58] [S59] [S60] [S64] [S74]
Research on Methods for Improving Gesture Elicitation Effects	27	34.6%	[S02] [S03] [S05] [S06] [S13] [S14] [S19] [S20] [S24] [S25] [S39] [S48] [S51] [S52] [S55] [S56] [S62] [S63] [S65] [S67] [S69] [S70] [S77] [S26] [S76] [S35] [S68]
Research on Methods for Processing Gesture Data	20	25.6%	[S01] [S09] [S10] [S11] [S12] [S17] [S18] [S22] [S23] [S28] [S29] [S43] [S49] [S61] [S66] [S73] [S75] [S71] [S30] [S21]
Research on the Evaluation of Gesture Elicitation Outcomes	3	3.8%	[S44] [S46] [S72]
Research on Gesture Elicitation Toolkits	6	7.7%	[S27] [S31] [S42] [S33] [S78] [S32]

TABLE 3. Application domains of GES.

Application	Number of Papers	%	Citation
Head-mounted device	11	14.1%	[S04] [S19] [S36] [S53] [S54] [S57] [S58] [S65] [S66] [S69] [S32]
Smart Environment device	24	30.8%	[S02] [S03] [S05] [S13] [S24] [S25] [S28] [S37] [S38] [S40] [S41] [S43] [S44] [S45] [S46] [S49] [S63] [S67] [S26] [S06] [S76] [S74] [S30] [S68]
Touchscreen	9	11.5%	[S08] [S12] [S20] [S27] [S47] [S59] [S60] [S64] [S71]
Hardware Devices	2	2.6%	[S11] [S18]
PC	2	2.6%	[S01] [S72]
Human-Robot/Drone Interaction	4	5.1%	[S07] [S15] [S39] [S48]
Cross-device Interaction	2	2.6%	[S14] [S56]
Wearable Device	2	2.6%	[S17] [S34]
General Purpose	22	28.2%	[S09] [S10] [S22] [S23] [S29] [S31] [S42] [S50] [S51] [S52] [S55] [S61] [S62] [S70] [S73] [S75] [S77] [S33] [S78] [S16] [S35] [S21]

smartphones to touchless or multi-modal interaction domains such as XR, smart environments, and human-robot interaction, illustrating the impact of technological and demand changes on research paradigms. Nearly half of the studies focus on mid-air gestures (38 out of 78, 48.7%), and the breadth and depth of research on mid-air gestures also reflect the growing demand and interest in novel human-computer interaction methods.

4) ELICITATION ENVIRONMENT

The elicitation environment refers to the experimental conditions and settings established by researchers (Fig. 2). Though primarily conducted in-lab settings (14 out of 78 studies, 17.9%), there has been a gradual increase in gesture elicitation studies conducted in-situ settings (6 out of 78, 7.7%) and virtual environments (9 out of 78, 11.5%) in recent years.

TABLE 4. Research object of GES.

Object	Number of Papers	%	Citation
Mid-air Gesture	38	48.7%	[S02] [S03] [S04] [S05] [S07] [S15] [S19] [S24] [S25] [S28] [S36] [S37] [S39] [S40] [S43] [S44] [S45] [S46] [S47] [S50] [S57] [S58] [S61] [S62] [S63] [S65] [S67] [S69] [S70] [S72] [S77] [S26] [S16] [S32] [S76] [S74] [S35] [S30]
Touch Gesture	9	11.5%	[S12] [S14] [S20] [S38] [S52] [S59] [S60] [S71] [S06]
Full-body Gesture	7	9.0%	[S01] [S13] [S41] [S53] [S54] [S55] [S66]
Hybrid Gesture	4	5.1%	[S34] [S56] [S64] [S75]
Lingual and palatal gestures	1	1.3%	[S21]
Microgesture	2	2.6%	[S11] [S17]
Handheld Gesture	2	2.6%	[S08] [S18]
Robot Gesture	1	1.3%	[S48]
Foot Gesture	2	2.6%	[S09] [S68]
/	12	15.4%	[S10] [S22] [S23] [S27] [S29] [S31] [S42] [S49] [S51] [S73] [S33] [S78]

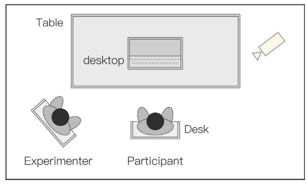
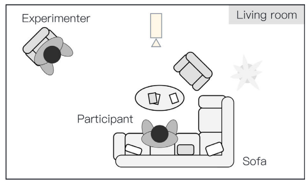
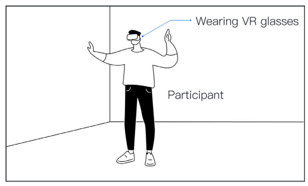
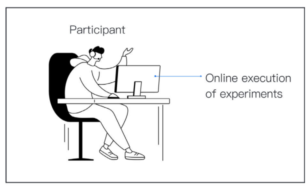
Experimental environment		Number of Papers	%	Illustration
Physical	In-lab [S01] [S12] [S25] [S37] [S41] [S47] [S52] [S62] [S67] [S72] [S16] [S30] [S68] [S21]	14	17.9%	
	In-situ [S17] [S24] [S38] [S45] [S70] [S06]	6	7.7%	
Virtual	VR [S04] [S18] [S63] [S65] [S69] [S32] [S76] [S74] [S35]	9	11.5%	
	Online [S36] [S60] [S26]	3	3.8%	
Not mentioned	[S02] [S03] [S05] [S07] [S08] [S09] [S11] [S13] [S14] [S15] [S19] [S20] [S28] [S34] [S39][S40] [S43] [S44] [S46] [S48] [S50] [S53] [S54] [S55] [S56] [S57] [S58] [S59] [S61][S64] [S66] [S75] [S77] [S71]	34	43.6%	/
/	[S10] [S22] [S23] [S27] [S29] [S31] [S42] [S49] [S51] [S73] [S33] [S78]	12	15.4%	/

FIGURE 2. Diagram of the GES experimental environment.

Besides, nearly half of the studies (34 out of 78, 43.6%) did not explicitly report the elicitation environment (missing or implied), which could hinder the replicability, reproducibility, and repeatability of these experiments [25].

5) REFERENT DISPLAY

A referent in the context of GES refers to the command/action that the input suggestion is intended to execute [10]. For instance, if the objective of a GES is to test how people use gestures to control music playback, then the referent could be "play music," which can be presented to the participants in the form of text, animations, or images. As a specific example (Fig. 3), an image of a "play" button could be displayed on the screen, or an animation of a musical note, to guide participants in making gesture controls. In this scenario, the referent acts as an instruction, guiding participants to perform the corresponding gesture control to accomplish the task of playing music.

In practical applications, the use and maturity of these presentation forms vary. We have categorized the presentation forms of referents according to their fidelity levels into four categories (as shown in Fig. 3). Although text-based referents still dominate, the use of referents with higher fidelity in experiments is gradually increasing. This trend reflects the research community's growing emphasis on enhancing the fidelity of experimental materials and the immersion of participants.

B. ANALYSIS ACCORDING TO RESEARCH QUESTIONS

GES, originating from the field of requirements engineering and participatory design, has been widely applied to assist designers in selecting the most suitable gesture sets for specific applications. As the GES method has not yet established a widely recognized process, this section aims to organize and summarize the currently common implementation processes, as well as the related methods and their effects.

1) RQ1: POPULAR RESEARCH PROTOCOL

In order to help systematize and structure gesture elicitation research and ensure the clarity and reproducibility in the experimental process, after summarizing the research processes of all literatures, we proposed a four-stage GES research protocol as follows (Fig. 4).

Stage 1: Determine command set, focusing on specific interaction requirements

(1) Design Tasks: Initially, it's crucial to determine the specific tasks to be executed via gestures. The Design of tasks is vital for ultimately determining the appropriate gesture set, directly impacting the actions users need to perform in a specific application.

Stage 2: Perform experiment, systematically collecting initial user reactions and gesture proposals

(2) Select Participants: After determining the tasks, the experimenter needs to select participants suitable to propose the gesture. These participants are typically end-users, as they have a unique understanding of the actual usage scenarios of the application.

(3) Group Collaboration Partners (Optional): Participants collaborate in groups to jointly propose gestures [26]. This method can facilitate the exchange of creative ideas and the collision of thoughts, making gesture proposals more diverse and innovative. However, not all studies require or are suited to the form of group collaboration.

(4) Set Up An Elicitation Environment (Optional): Creating a simulated or actual application scenario for gesture collection, including but not limited to setting up specific physical spaces and preparing necessary technical equipment. Although this step can greatly enhance the relevance of the research, it may be omitted in some projects due to resource and cost constraints.

(5) Priming or Framing (Optional): Before presenting the referents, participants are given visual [27] or kinesthetic priming [26], or are confined within a specific framework [26] for thinking. This step aims to optimize the gesture creation process through specific stimuli or constraints, enhancing the specificity and effectiveness of gesture design. Whether to adopt this step depends on the specific objectives of the research.

(6) Show Referents: Show the participants the results and processes related to the interaction tasks, which are known as referents [28]. This can be achieved by displaying application interfaces, simulating specific scenarios, or providing descriptions of the referents.

(7) Collect Gesture Proposals: When users participate in the gesture elicitation study, researchers need to record their gesture actions. Data can be collected in various ways, such as through video recording [29], [30], sensor tracking [31], [32], questionnaires [33], and other methods.

Stage 3: Determine the final gesture set, selecting the most suitable gestures based on user feedback and data analysis

(8) Gesture Data Processing: The collected data needs to be organized and encoded for subsequent analysis. Gestures are typically segmented into different categories, each assigned a unique identifier.

(9) Determine User Consensus: By comparing and analyzing gathered data, gestures with high consensus are identified. This helps determine which gestures will have widespread acceptance, thus providing a strong basis for the final gesture set.

(10) Select Gesture Set: Based on the results of user consensus and data analysis, the final gesture set is chosen. The final gesture set should meet design criteria such as memorability [33], ease [34], and goodness [29] to ensure that users can perform tasks easily and effectively in practical use.

Stage 4 (Optional): Test experiment results, verifying their effectiveness in actual applications

(11) Test Gesture Set (Optional): After identifying a set of potential gestures, these gestures undergo further validation and evaluation. This step helps ensure the practicality and user acceptance of the selected gestures, especially when they are designed for specific applications or systems. Its necessity depends on the purpose and context of the research.

2) RQ2: GESTURE CONSENSUS ANALYSIS & METRICS

The Agreement Rate is the most commonly used measurement for evaluating and selecting gesture sets, which refers to

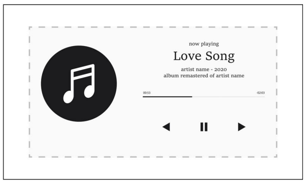
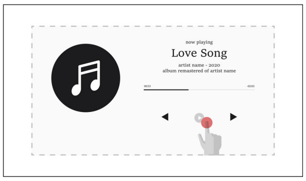
Referent display	Number of Papers	%	Illustration
Based on text or verbal description [S02] [S05] [S07] [S13] [S15] [S17] [S18] [S24] [S37] [S38] [S46] [S47] [S52] [S58] [S65][S69] [S70] [S72] [S71] [S06] [S74] [S21]	22	28.2%	Play music
Based on static images [S02] [S13] [S24] [S37] [S41] [S52] [S55] [S59] [S60] [S69] [S70] [S16]	12	15.4%	
Based on dynamic videos or animated GIFs [S09] [S14] [S18] [S36] [S37] [S45] [S47] [S52] [S56] [S58] [S69] [S70] [S72] [S16] [S74]	15	19.2%	
Based on interactive demos or real systems [S19] [S20] [S25] [S57] [S76] [S35]	6	7.7%	
Not mentioned [S01] [S04] [S08] [S11] [S12] [S28] [S34] [S39] [S40] [S43] [S44] [S48] [S50] [S53] [S54][S62] [S63] [S64] [S66] [S67] [S75] [S77] [S26] [S32] [S30] [S68]	26	33.3%	/
/ [S03] [S10] [S22] [S23] [S27] [S29] [S31] [S42] [S49] [S51] [S61] [S73] [S33] [S78]	14	17.9%	/
In some studies, the referent display is a combination of multiple types.			

FIGURE 3. **Referent display in GES.**

the degree of consensus among different users in understanding and using the same set of commands. In theory, a higher Agreement Rate indicates that a gesture set is more easily understood and used by most users.

The calculation formula for Agreement Rate was first proposed by Findlater et al. in 2012 [35] and later optimized by Vatavu and Wobbrock in 2015 [36]:

$$\mathrm{AR}\,(r) = \frac{\sum_{P_i \subseteq P} \frac{1}{2}|P_i|(|P_i|-1)}{\frac{1}{2}|P|(|P|-1)} \tag{1}$$

Subsequently, other researchers have proposed supplementary formulas, examples are: Vatavu and Wobbrock [37] proposed a Coagreement Rate, used to calculate the agreement rate between two different experimental groups; Morris [38] introduced two metrics: max-consensus and consensus-distinct ratio. The max-consensus metric represents the percent of participants who proposed the most popular interaction for a given referent. The consensus-distinct ratio metric represents the percent of the distinct interactions proposed for a given referent that achieved a given consensus threshold among participants. Huang et al. [39], considering that a participant might propose multiple gestures for each given referent, proposed a new method to calculate the consensus rate based on groups of gestures rather than individual gestures.

A key issue affecting the calculation of the agreement rate is the method for determining whether two gestures are equivalent. In most literature, researchers make a binary

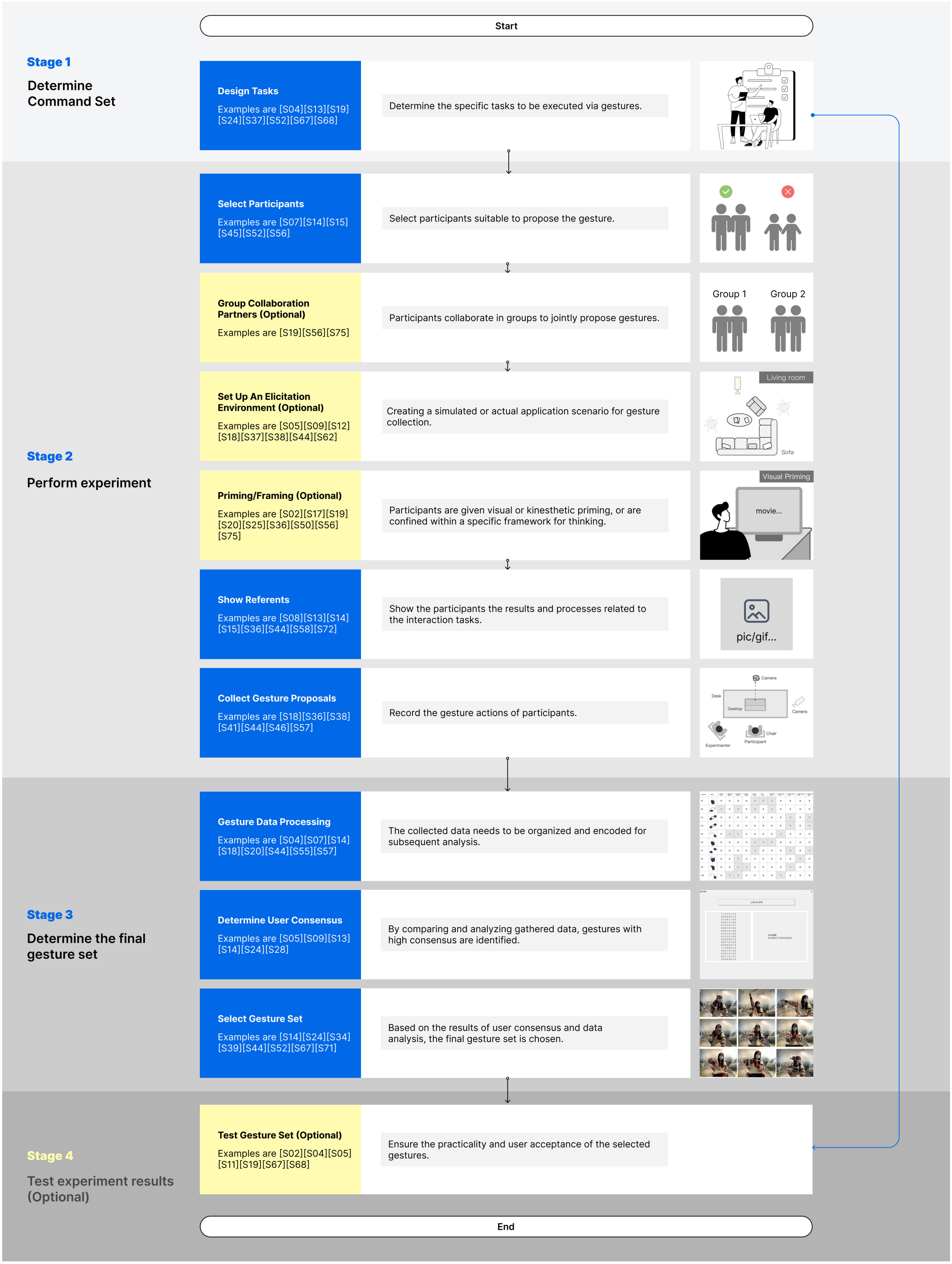


FIGURE 4. General steps of GES process flow.

judgment based on common sense. For example, two gestures are considered equivalent if they have the same shape or trajectory (value: 1), otherwise not equivalent (value: 0). Madapana et al. [40] argue that this ''hard classification'' method is not accurate enough and suggest replacing it with a ''soft classification'' where similarity can take any value between 0 and 1. Consequently, they proposed a method for calculating the soft agreement rate. i Apart from the agreement rate, researchers have also proposed evaluation metrics from other perspectives. For instance, Felberbaum and Lanir [29] proposed that the uniqueness (specification) of gestures may also be important in some circumstances, as a gesture that can ''be generalized and used for various referents'' is more likely to cause confusion and misoperation. To address this, they proposed the Specification Score:

$$S = \sum\nolimits_{r \in R_p} \left(\frac{|r|}{|R_p|} \right)^2 \tag{2}$$

Dingler et al. [41] proposed the transferability score:

$$T_{AB} = \frac{\sum_{p \in P} \sum_{c \in C} X_{c_p}}{N_P \times N_C \times w} \tag{3}$$

which is used to assess to what extent gestures suitable for Device A are also suitable for Device B. Not surprisingly, if a gesture is suitable for multiple devices, it can reduce the burden of having to relearn different gestures when switching between devices.

Finally, due to the increasing demand in the research community for calculating various metrics, some computation toolkits have also started to emerge. For instance, Vatavu and Wobbrock [36] have provided a toolkit called AGATe to assist in calculating agreement, disagreement, and coagreement rates. Other toolkits include but are not limited to: GestureAnalyzer [42], KinectAnalysis [43], GestAnalytics [44], Gelicit [45], CrowdSensus [46], and CrowdLicit [47], however, some of these are research-oriented and no longer updated.

3) RQ3: INFLUENCING FACTORS

The factors that have been found by empirical research to have an impact on the elicitation effect of gestures are listed below:

(1) Individual Characteristics of Participants

Six studies found participant selection is an important factor affecting the results of gesture elicitation experiments. Participants with different gender, cultural backgrounds, and levels of creativity have different responses to gesture elicitation, which could lead to differences in experimental results.

- *Gender differences:* Gesture preferences may differ between men and women. Vatavu and Wobbrock [37] calculated the level of agreement between male and female users and found that while overall they reached a similar consensus, there were significant differences in the agreement rates for seven referents. These findings suggest that females and males reach a consensus on gestures in different ways, depending on the nature of the referents and potentially related to cognitive preferences. However, the underlying mechanisms remain unclear.
- *Cultural background differences:* People from different cultural backgrounds may have different understandings and uses of gestures, which also affect the results of gesture elicitation. Dong et al. [33] found that participants from different cultural backgrounds showed different unconscious ideas and mental models of a communal mind, which led to different experimental results. For example, participants may use gestures in the opposite direction just because they have different cultural backgrounds. In some countries, putting a hand over the mouth may represent ''mute'', while in others it's completely different. Brito et al. [31] studied the influence of culture on gesture-based interactive systems and found that culture does not significantly affect the category of gestures used, but it affects the selection of specific gestures, especially when the function is related to the culture itself. Silpasuwanchai and Ren [48] found in a study on full-body game gestures that game experience affects gesture preference. Game players tend to define a more symbolic and effective unique gesture library, while inexperienced game players tend to define a more direct and physical gesture set. Wu et al. [49] found that the influence of cultural factors varies depending on the task and identified four types of tasks that are universally accepted by participants from different cultural backgrounds: tasks strongly associated with direction or order, tasks related to object manipulation, tasks dealing with objects that can be mapped to concrete objects in the real world, and tasks associated with symbols that are universally accepted. The two factors that influence cultural bias are the cultural norm in gesture expression and language.
- *Creativity:* Creativity is the ability of an individual in terms of originality, novelty, and usefulness [50]. Participants with higher levels of creativity may tend to adopt more unusual approaches to solve problems. Participants with lower levels of creativity may be more susceptible to the influence of habitual and conventional thinking, thus limiting their solution choices. Gheran et al. [51] found a negative correlation between participants' creativity and thinking time, indicating that higher creativity is associated with shorter thinking time in gesture elicitation. Although this finding did not reach statistical significance, it suggests a possible impact of creativity.

In addition, researchers have found that factors such as participant age (in 5 studies) and technical background (in 1 study) may also affect experimental results, but no clear conclusions have been drawn.

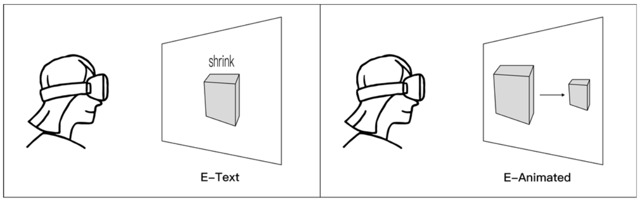


FIGURE 5. Description of E-Text and E-Animated.

(2) Referent

Four studies highlighted the impact of referent on experimental results, one of which explored the referent display, and the other two mentioned referent sequences. Additionally, one literature studied the effect of the physical size of the information carrier on gesture interaction.

In gesture elicitation, the display of the referent can be either static or dynamic, physical or digital. Examples are image versus video, on paper versus on screen. Different displays of stimulus may affect participants' understanding and cognition of the referent, thereby influencing the gestures they propose. Williams and Ortega [12] conducted a comparative study that discussed how different referent displays may affect the results of gesture and speech elicitation (Fig. 5). One experiment group used a textual referent, referred to as "E-Text" [52], and the other experiment group used an animated referent, referred to as "E-Animated" [53]. The results showed that the choice of referent display affected the results of gesture proposals and Agreement Rates. Specifically, it was observed that the consistency of gesture proposals in the E-Animated group was not as high as in the E-Text group. This may be due to differences in participants' understanding when using different displays of referents. For gesture proposals, the transition from textual to animated form had a greater impact when the meaning was ambiguous. Moreover, for all referents, user feedback indicated that the perceived workload in the E-Animated group was lower than in the E-Text group when generating gesture proposals, possibly because the animation provides more details about the movements.

In addition, presenting referents as a whole or individually to participants may also have an impact. Gheran et al. [51] randomly set the order of referents in the experiment to avoid bias and potential transfer effects from one referent to the next. However, by doing so, participants are found to have difficulty recognizing the relationship between the referents and proposing relevant gestures. Presenting all referents simultaneously or using a mixed method, i.e., randomly presenting referents to participants as a group, leads to different levels of agreement. Wu et al. [26], considering the integrity of the referent and the corresponding gestures, suggest presenting all target referents as a whole and allowing participants to change their initial gesture proposals if necessary.

The physical size of the referent carrier can influence people's perception of their surroundings, thereby changing the gesture proposal. Zhou and Bai [54] conducted experiments to compare gesture proposals under different map sizes. The results showed that the user-defined gestures for the same geographic information system (GIS) commands vary with map sizes. This variation manifests as follows: on smaller map sizes, users' gestures typically focus on finger or palm positions; on larger map sizes, users tend to use movements involving the entire arm or shoulder. Additionally, as the map size increases, the range of gesture movements also expands.

(3) Elicitation Techniques

In GES, two common elicitation techniques are the Priming Method and the Framing Method. Priming is a technique that influences users' cognition and behavior by providing them with preparatory information [10]. It helps users imagine a wider range of possibilities when generating gestures [55]. Priming can be done in various ways, including but not limited to sci-fi priming, kinesthetically priming, creative-mindset priming, and collaborative priming.

On the other hand, the Framing Method involves placing the design problem within a specific context to help participants think and design within a particular usage scenario. This approach makes gesture design more practically feasible and enhances participants' focus. Framing can be executed in various ways, including task framing, scenario framing, physical constraints, and choice-based framing, among others.

Silpasuwanchai and Ren [48] proposed and confirmed the benefits of a choice-based framing method, which provides a predefined list of possible gestures. Participants found this predefined list to be a useful reference when imagining new gestures, demonstrating the positive impact of this method on the quality of experimental results. Moreover, the choice-based framing method achieved higher consistency scores compared to past work [56]. Wu et al. [26] initiated participants in their elicitation experiments with a framework or scenario. The results of the experiment showed that when participants were required to propose gestures within the framework, they had a clear understanding of the tasks the target system should support and which gestures could or could not be used in a specific interaction environment. The framework also facilitated end-users in recommending gestures consistent with the scenario without increasing disagreements among end-users. Similar conclusions were also drawn in Vogiatzidakis et al.'s experiment [15], which was conducted within generic scenario as a framework. The results show participants achieved high consistency in gesture proposals.

Ali et al. [57] compared the learnability and memorability of the gestures proposed in three groups: sci-fi priming, creative-mindset priming, and no-priming. The results showed that users in the sci-fi group learned the gestures significantly faster and both primed gesture groups were more easily remembered. The author believes that the reasons for this phenomenon come from two aspects: first, the gestures triggered in the sci-fi environment are more attractive, easier to learn and remember; and second, the descriptions in sci-fi movies provide a sense of familiarity to the participants, leading to higher learnability and memorability.

(4) Elicitation environment

In the reviewed literature, only two experimental studies were conducted on the differences in the GES elicitation environment. Due to the difficulty in conducting GES experiments during the pandemic, Perera et al. [7] conducted a study in a virtual environment and explored participants' preferences between self-guided remote VR GES and laboratory experiments. Subjective assessment results revealed that regardless of participants' previous experience in VR, participants were more inclined to participate in self-guided remote VR GES than in laboratory experiments. Similar conclusions were also reflected in the experiments of Chamunorwa et al. [58]. The experimental results showed that participants preferred conducting unsupervised studies but proposed fewer gesture sets in this condition. Between supervised and unsupervised studies, there were differences in the quantity and max-consensus of gestures. Specifically, under unsupervised study conditions, there were more gestures, and the max-consensus scores were higher. However, there were no significant differences in gesture Agreement Scores between the two groups.

(5) Task Design

The type and difficulty of the task can also influence the results of gesture elicitation experiments.

Zaiţi et al. [32] conducted a gesture elicitation study for television control and found that the type of task significantly affected agreement rates, with abstract tasks having the lowest. The experimental results were further confirmed by a Friedman test, which showed that task type had a significant impact on participants' self-reported goodness ratings and recall rates, with abstract tasks receiving the lowest ratings.

Ganapathi and Sorathia [59] found that in three experiments with different task difficulties, participants' subjective evaluations of gesture appropriateness, usability, user preferences, and effort differed. In experiments with lower task difficulty, the leaning gesture performed the best, while in experiments with higher task difficulty, the walking-in-place (WIP) gesture had the highest appropriateness. Participants believed that task difficulty would affect their perception of gestures.

Task difficulty also affects the time participants take to think about gestures. If the task is too difficult, participants may feel frustrated or incapable, which will also affect their time and energy spent thinking about gestures. Gheran et al. [51] found that the longer participants think, the lower the agreement rate will be. This may be because the more time participants allocate to the task, the more creative they want to be, and they therefore propose gestures that others are unlikely to propose. It may also be because some tasks go beyond their sensorimotor knowledge, and participants are unable to make similar gestures. Zaiţi et al. [32] also found a significant negative correlation between participants' agreement rate and thinking time.

Additionally, Tijana and Vuletic [60] compared the performance of participants in gesture elicitation under conditions with and without time constraints. The experimental results found that time limitations did not have a significant impact on the choice and nature of gestures.

4) RQ4: POST VALIDATION (TEST GESTURE SET)

In GES, evaluating the final set of gestures is typically considered the last stage of the research, also known as post-hoc testing or validation. This evaluation can be achieved through various metrics, including physiological data measurement, behavioral observation, scale assessment, user interviews, etc.

- *Physiological data measurement*: Physiological data measurement tools such as electroencephalography (EEG) and electromyography (EMG) can be used to record users' physiological reactions for the purpose of evaluating the gesture set. For example, Ruiz and Vogel [61] calculated the Consumed Endurance (CE) of each gesture to evaluate the effectiveness of soft constraints in reducing arm fatigue and proved that more diverse, non-traditional gestures could be generated by using different body parts and more subtle movements. The same technique is also used in Uva al. [62] study. Huang et al. [39] proposed to combine electromyography, electrogoniometry, subjective preference, and other indicators in GES evaluation.
- *Behavioral observation:* The effectiveness of the final gesture set can also be evaluated through observing behavioral performance during gesture interaction tasks, Vogiatzidakis and Koutsabasis [15] measured task completion time, errors (false positives and false negatives), and task success. Tan et al. [63] measured reaction time to examine the performance of individual microgestures.
- *Scale assessment:* Scale assessment can be done using tools such as SUS (System Usability Scale), UEQ (User Experience Questionnaire), etc., to allow users to evaluate the ease of use, effectiveness, satisfaction, and other aspects of the gesture set. Specifically, the evaluation can be done from the following four aspects: Learning level, Implementation level, Subjective evaluation level, and Objective matching level (Table 5). The learning level involves aspects such as the learnability and memorability of the gesture set. The implementation level includes indicators such as the ease of use and comfort of executing the gestures. The subjective evaluation level refers to the user's subjective feedback on the gesture set, such as their feelings and satisfaction. The objective matching level includes the degree of matching between gestures and specific referents.
- *User interview:* Structured or semi-structured interview can be deployed to understand the user's feelings and feedback on gestures, to further inform the design and use of gestures. Landay and Cauchard [30] conducted brief qualitative interviews with participants after each task, during which participants explained their interactions. Vogiatzidakis and Koutsabasis [28] used the think-aloud approach in their exploratory

TABLE 5. Scales in GES assess dimensions.

Evaluation dimension		Citation
Learning	memorability	[S46] [S05] [S01] [S19] [S37] [S45] [S67]
	learnability	[S20]
	discoverability	[S13]
	likeliness to remember	[S46]
Implementation	difficulty	[S01]
	effort or fatigue	[S04] [S13] [S66]
	ease	[S04] [S05] [S09] [S12] [S13] [S14] [S20] [S36] [S45] [S47] [S64] [S66]
	comfort	[S40] [S45] [S52] [S57]
	task load	[S58]
	legibility	[S65]
	performance	[S65]
Subjective evaluation	preference	[S04] [S05] [S19] [S54] [S64] [S65] [S66]
	goodness	[S09] [S14] [S34]
	social acceptability	[S14]
	suitability	[S12] [S17] [S45] [S46] [S66]
	appropriateness	[S04] [S66]
	willingness to use	[S20]
	enjoyment	[S37]
Objective matching	degree of matching	[S05] [S13] [S19] [S36] [S37] [S52] [S67]
	descriptiveness	[S01]
	goodness of mapping	[S47]
	naturalness	[S65] [S01]

experiment, which included discussions with participants. Chamunorwa et al. [64] used the "think-out-loud" protocol and post-hoc interviews in GES, revealing many qualitative results. Beşevli et al. [65] recorded participants' choices (or hesitation) between two alternative gestures and their selection criteria through a semi-structured interview.

C. OTHER FINDINGS

1) BIAS REDUCTION

In elicitation studies, legacy bias [66] and performance bias [61] are two common traps. The legacy bias represents participants' tendency to use a specific kind of gesture that they are accustomed to and less willingness to try other gestures, even if new gestures may be more effective or efficient. The legacy bias may affect users' acceptance and experience of new gesture operations. Performance bias refers to participants' lack of consideration of factors such as repetitive use and potential fatigue when designing gestures. Some gestures are fine when used sparingly, but may feel tiring, uncomfortable, or even unpleasant when used frequently.

To address legacy bias, Morris et al. [66] proposed three techniques: production, priming, and partners. The production technique involves asking users to generate multiple interaction proposals for each referent. The priming technique encourages participants to explore a broader range of interaction techniques by guiding them to think about new form factors or sensing technologies. The partners technique involves conducting elicitation studies in a group format rather than with individuals. Perera et al. [67] also integrated these three techniques into the Virtual Environment GES.

In addition, there are other methods that can be used to reduce legacy bias. Dong et al. [33] introduced a two-stage user survey to reduce participants' legacy bias: In the first stage, an open-ended questionnaire was used to gather a large set of gestures for the desired interaction commands. In the second stage, a multiple-choice questionnaire was used, with options for each question extracted from the answers in the preliminary survey. Through such a two-stage elicitation, the impact of users' personal experience on the gestures was reduced. Vogiatzidakis and Koutsabasis [28] prepared referents with minimal illustrations of devices (without handlers or controls) to minimize legacy bias or any other potential source of inspiration from a specific device. Other commonly used methods include not allowing any hints to be given to participants [49], random presentation of referents or tasks [68], and so on.

To address performance bias, Uva et al. [62] proposed a user-centered framework that considers human factors, memorability, and specific user needs tailored to application scenarios, to minimize the impact of legacy [66] and performance bias [61]. Similarly, Ruiz and Vogel [61] suggested using soft constraints to correct legacy and performance biases by penalizing physical movements. Li et al. [34] used a between-subject experimental design to avoid participant fatigue, dividing the participants equally into two posture groups: one-handed thumb-based posture and two-handed posture. They also used a game-theory-based method to design the study to prevent participants from randomly designing obscure gestures.

Although production, priming, and partners are reported as useful techniques for reducing legacy bias [13], [26], [66], [69], their actual impact often lacks systematic evaluation. Additionally, Hoff et al.'s [70] study suggests that the actual effectiveness of these techniques is usually not statistically significant. Similar conclusions were found in the studies of Ortega et al. [71] and Williams et al. [69], which argued that using production technique may reduce legacy bias only in a few cases, with a decrease in agreement rates. The overall legacy still seems to exist, so it is not entirely clear whether the production technique reduces legacy bias or not.

Meanwhile, researchers also suggest dialectically consider legacy bias, which in a sense can also provide simplification and convenience for gesture design [55]. This is because legacy-biased gestures are often widely accepted and familiar to users of a particular culture or region, and these users do not need to think or make much effort when using them. Using legacy-biased gestures may lead to systems that are more efficient, natural, and easy to learn and use. For example, Beşevli et al. [65] found that when gesture memory exists, participants tended to choose traditional gestures, even if they thought the non-legacy version was easier to perform.

2) GESTURE CLASSIFICATION

Classification is a useful extra step in gesture elicitation. It helps identify and distinguish different types of gestures and determines which gestures are most likely to be associated with specific problems or objectives. Before conducting a consensus analysis, categorizing gesture helps to better understand and organize gesture information, providing a better foundation for subsequent analysis. However, not all research follows clearly defined standards [9], [26], [49], [59], [72].

So far, researchers have proposed many different taxonomies for gesture classification (see Table 6), and this article summarizes the 4 most commonly used ones in Fig.6 (The selection criterion is that the number of papers using this taxonomy $\geqslant 5$):

Form (similar to the method by Wobbrock et al. [56]): The form dimension can be used to distinguish between static gestures and dynamic gestures [14]. Static gestures refer to a fixed gesture form that remains unchanged once formed. Dynamic gestures, on the other hand, refer to hand movements along specific paths, which can include curves, directions, and speeds.

Nature (similar to the approach by McNeill [80] and Wobbrock et al. [56]): The nature dimension includes four categories: physical, symbolic, metaphorical, and abstract [73]. Physical gestures are typically gestures that directly manipulate objects, such as rotating an object or dragging a UI element. Symbolic gestures are gestures based on visual symbols, for example, drawing a circle to indicate rotating an object. Metaphorical gestures are gestures that borrow from real-world actions or interaction styles, such as rotating a finger or palm to indicate the rotation of a selected object. Abstract gestures are gestures that do not fit into any of the aforementioned categories. For example, waving the palm may represent any referent, such as adjusting volume or refreshing a page.

Body parts (similar to the methods by Silpasuwanchai and Ren [48] and Wu et al. [14]): The body parts dimension refers not only to the number of hands used by the participant (one or two hands), but also can include other body parts (full body) [31].

Flow (similar to the method by Wobbrock et al. [56]): In the flow dimension, gestures can be classified into two types: discrete and continuous [28]. Discrete gestures consist of multiple atomic gestures that users need to perform consecutively to trigger a command. In contrast, continuous gestures respond to the execution of the user's gesture in a continuous manner.

IV. DISCUSSION

A. IMPLICATIONS FOR GESTURAL INTERFACE DESIGN AND IMPLEMENTATION

(1) With the continuous development of gesture interaction technology, GES is also experiencing a growth in application demands, design complexity, and technical intricacies.

In terms of application domains, we have witnessed a transition from touch systems [34], [81] to immersive systems [26], [59], [62], [67], a shift that not only broadens the usage scenarios of gesture interactions but also enhances the richness and depth of user experiences. At the design level, gestures are evolving from planar motions [72], [82] to three-dimensional spatial actions [83], and from simple movements [84] to complex composite actions [48] involving multiple movements, which demand higher creativity and a deeper understanding of user behavior from designers. Technologically, gesture elicitation is moving from a single modality [28] (gesture only) to multimodal approaches [52], [85] (combining gestures with voice, etc.), and from simple manual annotation classification techniques [29], [30] to integrated solutions that combine various complex technologies [27], [86]. The technology advancements not only improve the accuracy and reliability of gesture recognition but also pave the way for more complex and natural interactions. These changes reflect the dynamic trends of continued methodological innovation and active technological advancements within the field of gesture interaction.

(2) The design and development of gesture interaction systems are increasingly focused on better serving user needs, pursuing more natural and intuitive gesture interfaces, and designing gestures that can adapt to the habits of users from different cultural backgrounds. Especially after public health events such as the COVID-19 pandemic, the demand for contactless gestures has significantly increased, highlighting the potential of gesture interaction technology to enhance user experience.

(3) Users play an increasingly important role in the design process of gesture interaction systems, as they can provide crucial requirement information and actively participate in the co-creation of solutions.

Therefore, user-involved design methods, such as gesture elicitation, are receiving growing attention. The essence of this approach lies in integrating users' direct experiences and feedback into the design process, thereby fostering the development of interaction solutions that more closely align with users' actual needs and expectations. User-involved design not only deepens designers' understanding of user behaviors and preferences but also aids in uncovering innovative interaction methods, enhancing the practicality and user experience of gesture-based systems. Moreover, this method can stimulate users' creative potential, making them co-creators in the design process, which in turn increases the acceptance and satisfaction with the design solutions.

B. IMPLICATIONS FOR BETTER CONDUCT GES RESEARCH

Considering the evidence of existing studies, gestures elicited from participants (end-users) as opposed to those provided by designers are more diverse, more popular with target users, and easier to remember. However, there is ample empirical evidence indicating that their effectiveness is influenced by various factors, including:

(1) Characteristics of the participants, particularly their gender, cultural background, creativity, age, and technical background.

TABLE 6. **The literature on gesture classification is mentioned in GES.**

Authors	Study	Taxonomy
1. V. Brito *et al.*[31]	Analysis of Cross-Cultural Effect on Gesture-Based Human-Robot Interaction	Form, Body parts, Nature
2. Wang *et al.*[73]	Defining and Analyzing a Gesture Set for Interactive TV Remote on Touchscreen Phones	Form, Nature, Binding, Flow, Attention
3. Dingler *et al.* [41]	Designing consistent gestures across device types: Eliciting RSVP controls for phone, watch, and glasses	Nature, Flow, Dimension, Complexity, Interaction, Location
4. MAGROFUOCO *et al.* [27]	Eliciting Contact-Based and Contactless Gestures with Radar-Based Sensors	Dimensionality, Range of motion, Body part, Gesture nature， Laterality, Gesture form, Scale invariance
5. Vogiatzidakis *et al.*[28]	Frame-based elicitation of mid-air gestures for a smart home device ecosystem	Nature, Form, Body Part, Flow, Spatial
6. Gheran *et al.*[51]	Gestures for smart rings: Empirical results, insights, and design implications	Nature, Structure, Complexity, Symmetry, Locale
7. Aslan *et al.*[68]	Pen + mid-air gestures: Eliciting contextual gestures	Form, Nature, Binding, Flow
8. S. Williams *et al.*[69]	The cost of production in elicitation studies and the legacy bias-consensus trade off	Flow, Nature, Fingers used, Palm shape, The number of hands used
9. S. Williams *et al.*[12]	The Impacts of Referent Display on Gesture and Speech Elicitation	Fingers used, Hands used, The shape of the hand, Motion of the gesture
10. Zhu *et al.*[74]	Tripartite Effects: Exploring Users' Mental Model of Mobile Gestures under the Influence of Operation, Handheld Posture, and Interaction Space	Number of Steps, Degree of Freedom, Context, Nature
11. Wu *et al.*[14]	Understanding freehand gestures: a study of freehand gestural interaction for immersive VR shopping applications	Nature, Body parts, Form, Viewpoint
12. Huang *et al.*[39]	User-Defined Gestures for Mid-Air Interaction: A Comparison of Upper Limb Muscle Activity, Wrist Kinematics, and Subjective Preference	Joint movement, Finger number
13. McAweeney *et al.*[75]	User-driven design principles for gesture representations	Body context, Environmental context, Perspective, Frame, Color, Gesture elements
14. Emmanuele Uva *et al.*[62]	A User-Centered Framework for Designing Midair Gesture Interfaces	Number of hands, Mode of hands, Direction of motion, Amplitude of the motion, Height of execution, Hand shape
15. Tan *et al.*[63]	Bikegesture: User elicitation and performance of micro hand gesture as input for cycling	Contact, Bending, Tilting, Multimodal Physical Sensing, Function Mapping
16. Bhowmick *et al.*[76]	Understanding Gesture Performance for Object Selection in VR: Classification and Taxonomy of Gestures in HCI	Hand dominance, Multi-body part movement gestures
17. Wei *et al.* [77]	User-defined gestures for mediated social touch on touchscreens	Nature, Cardinality, Movement forms
18. Zhou *et al.* [54]	User-defined mid-air gestures for multiscale GIS interface interaction	Motion state, Physical feature, Semantic feature
19. Chen *et al.* [78]	User-Defined Foot Gestures for Eyes-Free Interaction in Smart Shower Rooms	Form, Number of the foot (feet), Complexity, Flow, Nature, Symmetry
20. Villarreal-Narvaez *et al.* [79]	Exploring user-defined gestures for lingual and palatal interaction	Number of dimensions, Nature, Complexity, Score, Reflection time
21. Silpasuwanchai *et al.* [48]	Designing concurrent full-body gestures for intense gameplay	Body parts, Nature

- Many gestures have cultural meanings and influence the choices of users from different cultural backgrounds.
- Men and women have different emphases in understanding gestures.
- Users with high creativity generate gestures more efficiently.

(2) The display of referents, particularly the fidelity and the sequence of presenting referents.

- Using animated descriptions makes it easier for participants to understand than text descriptions.
- Presenting all referents at once makes it easier for participants to understand the overall requirements, but presenting them separately helps to avoid bias and potential transfer effects.
- In spatial operations (such as maps), larger referents induce higher consensus rates in gesture elicitation outcomes.

(3) Elicitation techniques, mainly Priming and Framing techniques.

- Appropriate use of Priming techniques can make the elicited gestures easier to learn and remember.
- Appropriate use of Framing techniques can lead to a higher consensus rate in the gesture elicitation outcomes.

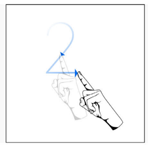
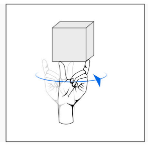
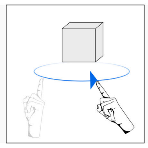
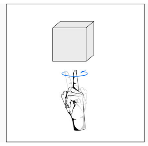
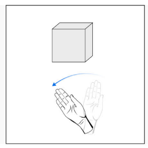
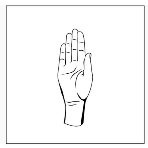
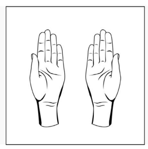
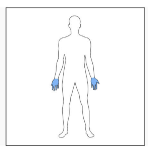
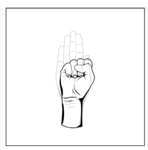
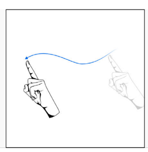
Taxonomy		Description	Illustration
Form [S07],[S12],[S17],[S24],[S47],[S57],[S65],[S68]	Static	Fixed gesture patterns.	
	Dynamic	The hand moves in a specific path.	
Nature [S07],[S12],[S14],[S17],[S24],[S34],[S47],[S57],[S64],[S65],[S71],[S68]	Physical	Gestures that act directly on an object.	
	Symbolic	Gestures depict a symbol.	
	Metaphorical	Gestures that draw on real-world movements or interactions.	
	Abstract	Arbitrary gesture-command mapping.	
Body parts [S07],[S13],[S17],[S24],[S65]	One hand	Execute with one hand.	
	Two hands	Execute with two hands.	
	Full body	A gesture is combined with another part of the body.	
Flow [S12],[S14],[S24],[S47],[S57],[S68]	Discrete	Command is triggered when a specific gesture is completed.	
	Continuous	Command is triggered during the gesture execution.	

FIGURE 6. **Four common gesture classification dimensions.**

(4) Elicitation environment.

- Compared to being observed in a laboratory setting, participants prefer to propose gestures alone in an immersive virtual environment, and the gestures produced by participants alone (without an experimenter observing) are more diverse and have a higher consensus rate.

(5) Task design, especially task difficulty, and type.

- The higher the task difficulty, the longer the participants think, resulting in a lower consensus rate in gesture elicitation outcomes.
- Concrete tasks, as opposed to abstract tasks, elicit a higher consensus rate in gestures.

The above conclusions indicate that when designing gesture elicitation study protocols, it is necessary to properly design the experimental plan based on the research objectives and user characteristics, such as whether the goal is to obtain a gesture set with a higher consensus rate or be more innovative; whether the target users have significant age and cultural characteristics; and whether the experimental tasks can be designed to be less difficult or more specific.

C. IMPLICATIONS FOR FUTURE RESEARCH OF GES METHODS

According to the perspectives and suggestions of researchers in existing literature, although the gesture elicitation method has become an effective tool in the research and practice of interaction systems, some important issues still lack universally recognized answers or best practice processes, including:

(1) What is the best workflow for gesture elicitation studies?

Section III of this paper summarizes the most common linear workflow, but in practice, some studies have adopted non-linear workflows, such as those including partial iterations or branches [75]. However, researchers have not explained why this approach was taken or whether it yields better results. In future studies, the research community can encourage controlled experiments to compare the impact of different workflows on gesture elicitation outcomes and summarize best practice processes under different conditions based on empirical evidence.

(2) How to evaluate the execution quality of a particular work step and its relationship with the final outcome?

Current gesture elicitation studies have proposed various evaluation metrics for the final gesture set, such as consensus rate, but there is still a lack of agreement on how to select participants, design referents and tasks, deploy elicitation environments, and apply elicitation techniques. As mentioned earlier, these issues significantly affect the final results, but the mechanisms of their impact are still unclear.

Introducing theories and models from related research fields and considering different contexts may help to narrow this research gap, for example: assessing the intelligibility of gestures based on communication theories, or understanding gesture interpretation based on cross-cultural studies. It is worth mentioning that, outside of GES community, broader gesture research has also proposed different metrics, e.g. Inter-Rater Agreement [87] (Metric: Cohen's kappa, Fleiss' kappa, etc.), Consensus Percentage [88] (Metric: Percentage of participants or raters who agree on the interpretation of a gesture), Confusion gesture Matrices [89] (Metric: Matrix showing the frequency of different interpretations for each gesture), Semantic Differential Scales [90] (Metric: Participants rate gestures on semantic scales).

(3) How to better design gestures for actual systems based on gesture elicitation results?

The gesture elicitation method can effectively bring out user preferences, but this does not mean that allowing users to decide what gestures the final system should use is always the best approach, especially when the consensus rate of the gesture set provided by users is low. A new approach proposed in current research is not to directly use the specific gestures obtained from gesture elicitation studies, but to use feature encoding or semantic encoding to summarize the knowledge obtained from gesture elicitation to inform gesture design. However, the specific standards for encoding, how to design gestures based on encoded knowledge, and how to evaluate their effects still require further research.

It is worth noting that the results of gesture elicitation studies are influenced by many factors that may affect the data collected and the interpretations made. Although there is empirical evidence in the literature, it is important to note that the field is broad and specific studies may focus on different aspects. In the future, research can be further expanded to include other criteria, such as: Cultural Differences, Context and Task Specificity, Individual Differences, Gesture Type and Form, Participant Engagement and Comfort, Technology and Methodology, Task Complexity, Cross-Disciplinary Collaboration, and Feedback Mechanisms.

(4) How to handle biases introduced in gesture elicitation studies?

Current research has identified two types of biases introduced by the gesture elicitation method: legacy bias, which refers to the influence of users' past experience with gestures on their choices for new products; and performance bias, which refers to fatigue, discomfort, and other phenomena caused by participants frequently using gestures during the elicitation process. Researchers have proposed some methods to reduce bias, including production, priming, partners, and the use of soft constraints, but their effectiveness remains to be confirmed. Additionally, some studies suggest that bias is not always negative; for example, gestures with legacy bias might actually be in line with user habits and easy to understand and remember.

(5) How to conduct gesture elicitation research in emerging immersive systems like VR?

Theoretically, VR technology can provide a more immersive, flexible, and customizable experimental environment and stimuli for gesture elicitation studies, thus can be conducted remotely without experimenter intervention. However, its practical application is still rare, and its methods and

TABLE 7. List of papers included for review.

S-ID	References
S01	Céspedes-Hernández, David, et al. "A grammar for specifying full-body gestures elicited for abstract tasks." Journal of Intelligent & Fuzzy Systems 39.2 (2020): 2433-2444.
S02	Uva, Antonio Emmanuele, et al. "A user-centered framework for designing midair gesture interfaces." IEEE Transactions on Human-Machine Systems 49.5 (2019): 421-429.
S03	Madapana, Naveen, et al. "Agreement study using gesture description analysis." IEEE transactions on human-machine systems 50.5 (2020): 434-443.
S04	Ganapathi, Priya, and Keyur Sorathia. "Elicitation study of body gestures for locomotion in HMD-VR interfaces in a sitting-position." Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games. 2019.
S05	Dong, Haiwei, et al. "An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions." IEEE access 3 (2015): 543-555.
S06	Chamunorwa, Michael, et al. "An Empirical Comparison of Moderated and Unmoderated Gesture Elicitation Studies on Soft Surfaces and Objects for Smart Home Control." Proceedings of the ACM on Human-Computer Interaction 7.MHCI (2023): 1-24.
S07	Brito, Icaro V., et al. "Analysis of cross-cultural effect on gesture-based human-robot interaction." International Journal of Mechanical Engineering and Robotics Research 8.6 (2019): 18-29.
S08	Köpsel, Anne, and Nikola Bubalo. "Benefiting from legacy bias." interactions 22.5 (2015): 44-47.
S09	Felberbaum, Yasmin, and Joel Lanir. "Better understanding of foot gestures: An elicitation study." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.
S10	Vatavu, Radu-Daniel, and Jacob O. Wobbrock. "Between-subjects elicitation studies: Formalization and tool support." Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016.
S11	Tan, Yanke, Sang Ho Yoon, and Karthik Ramani. "BikeGesture: user elicitation and performance of micro hand gesture as input for cycling." Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2017.
S12	Wang, Yuntao, et al. "Defining and Analyzing a Gesture Set for Interactive TV Remote on Touchscreen Phones." 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops. IEEE, 2014.
S13	Silpasuwanchai, Chaklam, and Xiangshi Ren. "Designing concurrent full-body gestures for intense gameplay." International Journal of Human-Computer Studies 80 (2015): 1-13.
S14	Dingler, Tilman, et al. "Designing consistent gestures across device types: Eliciting RSVP controls for phone, watch, and glasses." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.
S15	Jane, L. E., et al. "Drone & Wo: Cultural Influences on Human-Drone Interaction Techniques." CHI. Vol. 17. 2017.
S16	Vuletic, Tijana, et al. "Effects of activity time limitation on gesture elicitation for form creation." Journal of Engineering Design 34.11 (2023): 963-985.
S17	Magrofuoco, Nathan, et al. "Eliciting contact-based and contactless gestures with radar-based sensors." IEEE Access 7 (2019): 176982-176997.
S18	Navas Medrano, Samuel, Max Pfeiffer, and Christian Kray. "Enabling remote deictic communication with mobile devices: An elicitation study." Proceedings of the 19th international conference on human-computer interaction with mobile devices and services. 2017.
S19	Wu, Huiyue, et al. "Exploring frame-based gesture design for immersive VR shopping environments." Behaviour & Information Technology 41.1 (2022): 96-117.
S20	Li, Wing Ho Andy, Kening Zhu, and Hongbo Fu. "Exploring the design space of bezel-initiated gestures for mobile interaction." International Journal of Mobile Human Computer Interaction (IJMHCI) 9.1 (2017): 16-29.
S21	Villarreal-Narvaez, Santiago, Jorge Luis Perez-Medina, and Jean Vanderdonckt. "Exploring user-defined gestures for lingual and palatal interaction." Journal on Multimodal User Interfaces 17.3 (2023): 167-185.
S22	Tsandilas, Theophanis. "Fallacies of agreement: A critical review of consensus assessment methods for gesture elicitation." ACM Transactions on Computer-Human Interaction (TOCHI) 25.3 (2018): 1-49.
S23	Vatavu, Radu-Daniel, and Jacob O. Wobbrock. "Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit." Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015.
S24	Vogiatzidakis, Panagiotis, and Panayiotis Koutsabasis. "Frame-based elicitation of mid-air gestures for a smart home device ecosystem." Informatics. Vol. 6. No. 2. MDPI, 2019.
S25	Cafaro, Francesco, Leilah Lyons, and Alissa N. Antle. "Framed guessability: Improving the discoverability of gestures and body movements for full-body interaction." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.
S26	Friedman, Adina, and Francesco Cafaro. "From Thoughts to Interaction: Designing Controls for Video Playback Gestures with Embodied Schemata." Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. 2023.
S27	Magrofuoco, Nathan, and Jean Vanderdonckt. "Gelicit: a cloud platform for distributed gesture elicitation studies." Proceedings of the ACM on Human-Computer Interaction 3.EICS (2019): 1-41.

TABLE 7. ***(Continued.)*** **List of papers included for review.**

S28	Madapana, Naveen, Glebys Gonzalez, and Juan Wachs. "Gesture agreement assessment using description vectors." 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). IEEE, 2020.
S29	Tsandilas, Theophanis, and Pierre Dragicevic. "Gesture Elicitation as a Computational Optimization Problem." Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022.
S30	Alyamani, Hasan J. "Gesture Vocabularies for Hand Gestures for Controlling Air Conditioners in Home and Vehicle Environments." Electronics 12.7 (2023): 1513.
S31	Jang, Sujin, Niklas Elmqvist, and Karthik Ramani. "GestureAnalyzer: visual analytics for pattern analysis of mid-air hand gestures." Proceedings of the 2nd ACM symposium on Spatial user interaction. 2014.
S32	Li, Ang, et al. "GestureExplorer: Immersive Visualisation and Exploration of Gesture Data." Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023.
S33	Dang, Hai, and Daniel Buschek. "GestureMap: Supporting Visual Analytics and Quantitative Analysis of Motion Elicitation Data by Learning 2D Embeddings." Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021.
S34	Gheran, Bogdan-Florin, Jean Vanderdonckt, and Radu-Daniel Vatavu. "Gestures for smart rings: Empirical results, insights, and design implications." Proceedings of the 2018 Designing Interactive Systems Conference. 2018.
S35	Arendttorp, Emilie Maria Nybo, et al. "Grab It, While You Can: A VR Gesture Evaluation of a Co-Designed Traditional Narrative by Indigenous People." Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023.
S36	Ali, Abdullah, Meredith Ringel Morris, and Jacob O. Wobbrock. "“I Am Iron Man” Priming Improves the Learnability and Memorability of User-Elicited Gestures." Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021.
S37	Wu, Huiyue, et al. "Influence of cultural factors on freehand gesture design." International Journal of Human-Computer Studies 143 (2020): 102502.
S38	Chamunorwa, Michael, et al. "Interacting with rigid and soft surfaces for smart-home control." Proceedings of the ACM on Human-Computer Interaction 6.MHCI (2022): 1-22.
S39	Canuto, Clebeson, et al. "Intuitiveness Level: Frustration-Based Methodology for Human–Robot Interaction Gesture Elicitation." IEEE Access 10 (2022): 17145-17154.
S40	Beşevli, Ceylan, et al. "Investigating the effects of legacy bias: User elicited gestures from the end users perspective." Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems. 2018.
S41	Danielescu, Andreea, and David Piorkowski. "Iterative design of gestures during elicitation: Understanding the role of increased production." Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022.
S42	Nebeling, Michael, David Ott, and Moira C. Norrie. "Kinect analysis: a system for recording, analysing and sharing multimodal interaction elicitation studies." Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems. 2015.
S43	Gonzalez, Glebys, et al. "Looking beyond the gesture: Vocabulary acceptability criteria for gesture elicitation studies." Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 62. No. 1. Sage CA: Los Angeles, CA: SAGE Publications, 2018.
S44	Vogiatzidakis, Panagiotis, and Panayiotis Koutsabasis. "Mid-air gesture control of multiple home devices in spatial augmented reality prototype." Multimodal Technologies and Interaction 4.3 (2020): 61.
S45	Hoff, Lynn, Eva Hornecker, and Sven Bertel. "Modifying gesture elicitation: Do kinaesthetic priming and increased production reduce legacy bias?." Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction. 2016.
S46	Zaiţi, Ionuţ-Alexandru, Ştefan-Gheorghe Pentiuc, and Radu-Daniel Vatavu. "On free-hand TV control: experimental results on user-elicited gestures with Leap Motion." Personal and Ubiquitous Computing 19 (2015): 821-838.
S47	Aslan, Ilhan, et al. "Pen+ mid-air gestures: Eliciting contextual gestures." Proceedings of the 20th ACM International Conference on Multimodal Interaction. 2018.
S48	Björnfot, Patrik, and Victor Kaptelinin. "Probing the design space of a telepresence robot gesture arm with low fidelity prototypes." Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. 2017.
S49	Jurewicz, Katherina A., and David M. Neyens. "Redefining the human factors approach to 3D gestural HCI by exploring the usability-accuracy tradeoff in gestural computer systems." Applied Ergonomics 105 (2022): 103833.
S50	Morris, Meredith Ringel, et al. "Reducing legacy bias in gesture elicitation studies." interactions 21.3 (2014): 40-45.
S51	Gheran, Bogdan-Florin, et al. "RepliGES and GEStory: visual tools for systematizing and consolidating knowledge on user-defined gestures." Proceedings of the 2022 International Conference on Advanced Visual Interfaces. 2022.
S52	Wu, Huiyue, et al. "Seeking common ground while reserving differences in gesture elicitation studies." Multimedia Tools and Applications 78 (2019): 14989-15010.
S53	Ortega, Francisco R., et al. "Selection and manipulation whole-body gesture elicitation study in virtual reality." 2019 ieee conference on virtual reality and 3d user interfaces (vr). IEEE, 2019.
S54	Ortega, Francisco R., et al. "Selection and Manipulation Whole-Body Gesture Elicitation Study in Virtual Reality."

TABLE 7. ***(Continued.)*** **List of papers included for review.**

S55	Ruiz, Jaime, and Daniel Vogel. "Soft-constraints to reduce legacy and performance bias to elicit whole-body gestures with low arm fatigue." Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015.
S56	Rädle, Roman, et al. "Spatially-aware or spatially-agnostic? Elicitation and evaluation of user-defined cross-device interactions." Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015.
S57	Williams, Adam S., et al. "The cost of production in elicitation studies and the legacy bias-consensus trade off." Multimodal Technologies and Interaction 4.4 (2020): 88.
S58	Williams, Adam S., and Francisco R. Ortega. "The Impacts of Referent Display on Gesture and Speech Elicitation." IEEE Transactions on Visualization and Computer Graphics 28.11 (2022): 3885-3895.
S59	Kim, Sangyeon, and Sangwon Lee. "Touch digitality: Affordance effects of visual properties on gesture selection." Extended abstracts of the 2020 CHI conference on human factors in computing systems. 2020.
S60	Kim, Sangyeon, and Sangwon Lee. "Touchable pixels: Examining the affordance effect between an on-screen object and a user-elicited gesture on the touchscreen." Computers in Human Behavior 140 (2023): 107588.
S61	Zhou, Xiaoyan, Adam S. Williams, and Francisco R. Ortega. "Towards Establishing Consistent Proposal Binning Methods for Unimodal and Multimodal Interaction Elicitation Studies." International Conference on Human-Computer Interaction. Cham: Springer International Publishing, 2022.
S62	Lee, Sang-Su, et al. "Towards more natural digital content manipulation via user freehand gestural interaction in a living room." Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. 2013.
S63	Perera, Madhawa, et al. "Towards Self-Guided Remote User Studies-Feasibility of Gesture Elicitation using Immersive Virtual Reality." 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2021.
S64	Zhu, Kening, et al. "Tripartite effects: exploring users' mental model of mobile gestures under the influence of operation, handheld posture, and interaction space." International Journal of Human–Computer Interaction 33.6 (2017): 443-459.
S65	Wu, Huiyue, et al. "Understanding freehand gestures: a study of freehand gestural interaction for immersive VR shopping applications." Human-centric Computing and Information Sciences 9 (2019): 1-26.
S66	Bhowmick, Shimmila, Keyur Sorathia, and Pratul Chandra Kalita. "Understanding Gesture Performance for Object Selection in VR: Classification and Taxonomy of Gestures in HCI." Proceedings of the 12th Indian Conference on Human-Computer Interaction. 2021.
S67	Wu, Huiyue, Jianmin Wang, and Xiaolong Zhang. "User-centered gesture development in TV viewing environment." Multimedia Tools and Applications 75 (2016): 733-760.
S68	Chen, Zhanming, Huawei Tu, and Huiyue Wu. "User-defined foot gestures for eyes-free interaction in smart shower rooms." International Journal of Human–Computer Interaction 39.20 (2023): 4139-4161.
S69	Wu, Huiyue, et al. "User-defined gesture interaction for immersive VR shopping applications." Behaviour & Information Technology 38.7 (2019): 726-741.
S70	Wu, Huiyue, et al. "User-defined gesture interaction for in-vehicle information systems." Multimedia Tools and Applications 79 (2020): 263-288.
S71	Wei, Qianhui, Jun Hu, and Min Li. "User-defined gestures for mediated social touch on touchscreens." Personal and Ubiquitous Computing 27.2 (2023): 271-286.
S72	Huang, Jinghua, et al. "User-defined gestures for mid-air interaction: a comparison of upper limb muscle activity, wrist kinematics, and subjective preference." International Journal of Human–Computer Interaction 37.16 (2021): 1516-1537.
S73	Grijincu, Daniela, Miguel A. Nacenta, and Per Ola Kristensson. "User-defined interface gestures: Dataset and analysis." Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces. 2014.
S74	Zhou, Xiaozhou, and Ruidong Bai. "User-defined mid-air gestures for multiscale GIS interface interaction." Cartography and Geographic Information Science (2023): 1-14.
S75	McAweeney, Erin, Haihua Zhang, and Michael Nebeling. "User-driven design principles for gesture representations." Proceedings of the 2018 chi conference on human factors in computing systems. 2018.
S76	Perera, Madhawa, et al. "Using Virtual Reality to Overcome Legacy Bias in Remote Gesture Elicitation Studies." International Conference on Human-Computer Interaction. Cham: Springer Nature Switzerland, 2023.
S77	Nebeling, Michael, et al. "Web on the wall reloaded: Implementation, replication and refinement of user-defined interaction sets." Proceedings of the ninth acm international conference on interactive tabletops and surfaces. 2014.
S78	Bellucci, Andrea, et al. "Welicit: A wizard of oz tool for vr elicitation studies." IFIP Conference on Human-Computer Interaction. Cham: Springer International Publishing, 2021.

effectiveness require further research. One study that can be carried out immediately is to conduct in-depth comparison of the user's feelings, behaviors and gesture elicitation results under VR environment vs traditional environment.

(6) How to establish a more complete theoretical framework and evaluation system to enhance the validity and generalizability of gesture-based interaction research and design?

Since individual studies may focus only on specific factors or contexts, future work will need to take into account the cumulative body of research in gesture research, compare and comprehensively analyze the results of gesture elicitation research with broader gesture research results and empirical evidences.

V. CONCLUSION

Gesture elicitation methods have received widespread academic attention in the process of continuous improvement and have been empirically applied in multiple fields, including but not limited to VR, smart home, touchscreen, operating room, human-robot/drone interaction, and wearable devices. This literature review found that the research community is continuously optimizing and improving gesture elicitation methods based on empirical evidence and user feedback to enhance their applicability and effectiveness in various fields. Practical guidance has also been proposed to provide more theoretical and practical support for improving gesture elicitation methods.

This article, through a systematic review of the literature, summarizes the research progress in gesture elicitation methods and discusses factors that affect experimental results, including referents, participant selection, elicitation techniques, experimental environments, and tasks, as well as how these factors impact the results and effectiveness of gesture design. The discussions provide new perspectives and considerations for future research in gesture elicitation methods, aiming to promote theoretical and practical advancements in the field. Our goal is to provide researchers and practitioners with a comprehensive reference framework to help them conduct gesture elicitation studies more effectively and contribute to the future development of gesture interaction design.

APPENDIX
LIST OF PAPERS INCLUDED FOR REVIEW

See Table 7.

REFERENCES

[1] J. O. Wobbrock, H. H. Aung, B. Rothrock, and B. A. Myers, "Maximizing the guessability of symbolic input," in *Proc. CHI Extended Abstr. Hum. Factors Comput. Syst.*, Apr. 2005, pp. 1869–1872.

[2] Z. Dong, J. Zhang, X. Bai, A. Clark, R. W. Lindeman, W. He, and T. Piumsomboon, "Touch-move-release: Studies of surface and motion gestures for mobile augmented reality," *Frontiers Virtual Reality*, vol. 3, Aug. 2022, Art. no. 927258, doi: 10.3389/frvir.2022.927258.

[3] N. Samimi, S. von der Au, F. Weidner, and W. Broll, "AR in TV: Design and evaluation of mid-air gestures for moderators to control augmented reality applications in TV," in *Proc. 20th Int. Conf. Mobile Ubiquitous Multimedia*, May 2021, pp. 137–147, doi: 10.1145/3490632.3490668.

[4] P. Vogiatzidakis and P. Koutsabasis, "'Address and command': Two-handed mid-air interactions with multiple home devices," *Int. J. Hum.-Comput. Stud.*, vol. 159, Mar. 2022, Art. no. 102755, doi: 10.1016/j.ijhcs.2021.102755.

[5] N. A. N. Ch, D. Tosca, T. Crump, A. Ansah, A. Kun, and O. Shaer, "Gesture and voice commands to interact with AR windshield display in automated vehicle: A remote elicitation study," in *Proc. 14th Int. Conf. Automot. User Interface Interact. Veh. Appl.*, Sep. 2022, pp. 171–182, doi: 10.1145/3543174.3545257.

[6] M. Kazhura, "Exploring new depths: How could passengers interact with future in-car holographic 3D displays?" in *Proc. Int. Conf. Hum.-Comput. Interact.*, in Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2022, pp. 35–61, doi: 10.1007/978-3-031-04987-3_3.

[7] M. Perera, T. Gedeon, M. Adcock, and A. Haller, "Towards self-guided remote user studies–feasibility of gesture elicitation using immersive virtual reality," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2021, pp. 2576–2583, doi: 10.1109/SMC52423.2021.9658673.

[8] T. Chen, L. Xu, X. Xu, and K. Zhu, "GestOnHMD: Enabling gesture-based interaction on low-cost VR head-mounted display," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 5, pp. 2597–2607, May 2021, doi: 10.1109/TVCG.2021.3067689.

[9] H. Wu, Y. Wang, J. Qiu, J. Liu, and X. Zhang, "User-defined gesture interaction for immersive VR shopping applications," *Behav. Inf. Technol.*, vol. 38, no. 7, pp. 726–741, Jul. 2019, doi: 10.1080/0144929x.2018.1552313.

[10] A. S. Williams and F. R. Ortega, "A concise guide to elicitation methodology," 2021, *arXiv:2105.12865*.

[11] S. Buchanan, B. Floyd, W. Holderness, and J. J. LaViola, "Towards user-defined multi-touch gestures for 3D objects," in *Proc. ACM Int. Conf. Interact. tabletops Surf.*, Oct. 2013, pp. 231–240.

[12] A. S. Williams and F. R. Ortega, "The impacts of referent display on gesture and speech elicitation," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 11, pp. 3885–3895, Nov. 2022, doi: 10.1109/TVCG.2022.3203090.

[13] F. Cafaro, L. Lyons, and A. N. Antle, "Framed guessability: Improving the discoverability of gestures and body movements for full-body interaction," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–12.

[14] H. Wu, W. Luo, N. Pan, S. Nan, Y. Deng, S. Fu, and L. Yang, "Understanding freehand gestures: A study of freehand gestural interaction for immersive VR shopping applications," *Hum.-Centric Comput. Inf. Sci.*, vol. 9, no. 1, p. 43, Dec. 2019, doi: 10.1186/s13673-019-0204-7.

[15] P. Vogiatzidakis and P. Koutsabasis, "Mid-air gesture control of multiple home devices in spatial augmented reality prototype," *Multimodal Technol. Interact.*, vol. 4, no. 3, p. 61, Aug. 2020, doi: 10.3390/mti4030061.

[16] S. Villarreal-Narvaez, J. Vanderdonckt, R.-D. Vatavu, and J. O. Wobbrock, "A systematic review of gesture elicitation studies: What can we learn from 216 studies?" in *Proc. ACM Designing Interact. Syst. Conf.*, Jul. 2020, pp. 855–872.

[17] S. Villarreal-Narvaez, A. Sluÿters, J. Vanderdonckt, and R.-D. Vatavu, "Brave new GES world: A systematic literature review of gestures and referents in gesture elicitation studies," *ACM Comput. Surv.*, vol. 56, no. 5, pp. 1–55, Jan. 2024, doi: 10.1145/3636458.

[18] T. Tsandilas, "Fallacies of agreement: A critical review of consensus assessment methods for gesture elicitation," *ACM Trans. Comput.-Hum. Interact.*, vol. 25, no. 3, pp. 1–49, Jun. 2018, doi: 10.1145/3182168.

[19] P. Vogiatzidakis and P. Koutsabasis, "Gesture elicitation studies for mid-air interaction: A review," *Multimodal Technol. Interact.*, vol. 2, no. 4, p. 65, Sep. 2018, doi: 10.3390/mti2040065.

[20] Y. Cheng, Z. Wu, and R. Xiao, "Factors affecting the results of gesture elicitation: A review," in *Proc. 11th Int. Conf. Softw. Eng. Res. Innov.*, 2023, pp. 169–176.

[21] C. Okoli, "A guide to conducting a standalone systematic literature review," *Commun. Assoc. Inf. Syst.*, vol. 37, no. 43, pp. 879–910, Nov. 2015.

[22] Y. Xiao and M. Watson, "Guidance on conducting a systematic literature review," *J. Planning Educ. Res.*, vol. 39, no. 1, pp. 93–112, Mar. 2019, doi: 10.1177/0739456x17723971.

[23] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, and D. Moher, "Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement," *J. Clin. Epidemiol.*, vol. 134, pp. 103–112, Jun. 2021, doi: 10.1016/j.jclinepi.2021.02.003.

[24] V. Dan, "Empirical and nonempirical methods," in *The International Encyclopedia of Communication Research Methods*, 1st ed., J. Matthes, C. S. Davis, and R. F. Potter, Eds. Hoboken, NJ, USA: Wiley, 2017, pp. 1–3, doi: 10.1002/9781118901731.iecrm0083.

[25] B.-F. Gheran, S. Villarreal-Narvaez, R.-D. Vatavu, and J. Vanderdonckt, ''RepliGES and GEStory: Visual tools for systematizing and consolidating knowledge on user-defined gestures,'' in *Proc. Int. Conf. Adv. Vis. Interface*, Jun. 2022, pp. 1–9, doi: 10.1145/3531073.3531112.

[26] H. Wu, S. Fu, L. Yang, and X. Zhang, ''Exploring frame-based gesture design for immersive VR shopping environments,'' *Behav. Inf. Technol.*, vol. 41, no. 1, pp. 96–117, Jan. 2022, doi: 10.1080/0144929x.2020.1795261.

[27] N. Magrofuoco, J.-L. Pérez-Medina, P. Roselli, J. Vanderdonckt, and S. Villarreal, ''Eliciting contact-based and contactless gestures with radar-based sensors,'' *IEEE Access*, vol. 7, pp. 176982–176997, 2019, doi: 10.1109/ACCESS.2019.2951349.

[28] P. Vogiatzidakis and P. Koutsabasis, ''Frame-based elicitation of mid-air gestures for a smart home device ecosystem,'' *Informatics*, vol. 6, no. 2, p. 23, Jun. 2019.

[29] Y. Felberbaum and J. Lanir, ''Better understanding of foot gestures: An elicitation study,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–12, doi: 10.1145/3173574.3173908.

[30] A. Landay and J. R. Cauchard, ''Drone & Wo: Cultural influences on human-drone interaction techniques,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2017, pp. 6794–6799, doi: 10.1145/3025453.3025755.

[31] I. V. Brito, E. O. Freire, E. A. N. Carvalho, and L. Molina, ''Analysis of cross-cultural effect on gesture-based human–robot interaction,'' *Int. J. Mech. Eng. Robot. Res.*, vol. 8, no. 6, pp. 852–859, 2019.

[32] I.-A. Zaiţi, Ş.-G. Pentiuc, and R.-D. Vatavu, ''On free-hand TV control: Experimental results on user-elicited gestures with leap motion,'' *Pers. Ubiquitous Comput.*, vol. 19, nos. 5–6, pp. 821–838, Aug. 2015, doi: 10.1007/s00779-015-0863-y.

[33] H. Dong, A. Danesh, N. Figueroa, and A. E. Saddik, ''An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions,'' *IEEE Access*, vol. 3, pp. 543–555, 2015, doi: 10.1109/ACCESS.2015.2432679.

[34] W. H. A. Li, K. Zhu, and H. Fu, ''Exploring the design space of bezel-initiated gestures for mobile interaction,'' *Int. J. Mobile Hum. Comput. Interact.*, vol. 9, no. 1, pp. 16–29, Jan. 2017, doi: 10.4018/ijmhci.2017010102.

[35] L. Findlater, B. Lee, and J. Wobbrock, ''Beyond QWERTY: Augmenting touch screen keyboards with multi-touch gestures for non-alphanumeric input,'' in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, May 2012, pp. 2679–2682, doi: 10.1145/2207676.2208660.

[36] R.-D. Vatavu and J. O. Wobbrock, ''Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit,'' in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, Apr. 2015, pp. 1325–1334, doi: 10.1145/2702123.2702223.

[37] R.-D. Vatavu and J. O. Wobbrock, ''Between-subjects elicitation studies: Formalization and tool support,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2016, pp. 3390–3402, doi: 10.1145/2858036.2858228.

[38] M. R. Morris, ''Web on the wall: Insights from a multimodal interaction elicitation study,'' in *Proc. ACM Int. Conf. Interact. Tabletops Surf.*, Nov. 2012, pp. 95–104, doi: 10.1145/2396636.2396651.

[39] J. Huang, M. Qi, L. Mao, M. An, T. Ji, and R. Han, ''User-defined gestures for mid-air interaction: A comparison of upper limb muscle activity, wrist kinematics, and subjective preference,'' *Int. J. Hum.-Comput. Interact.*, vol. 37, no. 16, pp. 1516–1537, Oct. 2021, doi: 10.1080/10447318.2021.1898825.

[40] N. Madapana, G. Gonzalez, and J. Wachs, ''Gesture agreement assessment using description vectors,'' in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Nov. 2020, pp. 40–44, doi: 10.1109/FG47880.2020.00043.

[41] T. Dingler, R. Rzayev, A. S. Shirazi, and N. Henze, ''Designing consistent gestures across device types: Eliciting RSVP controls for phone, watch, and glasses,'' in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2018, pp. 1–12, doi: 10.1145/3173574.3173993.

[42] S. Jang, N. Elmqvist, and K. Ramani, ''GestureAnalyzer: Visual analytics for pattern analysis of mid-air hand gestures,'' in *Proc. 2nd ACM Symp. Spatial user Interact.*, Oct. 2014, pp. 30–39, doi: 10.1145/2659766.2659772.

[43] M. Nebeling, D. Ott, and M. C. Norrie, ''Kinect analysis: A system for recording, analysing and sharing multimodal interaction elicitation studies,'' in *Proc. 7th ACM SIGCHI Symp. Eng. Interact. Comput. Syst.*, Jun. 2015, pp. 142–151, doi: 10.1145/2774225.2774846.

[44] O. T. Buruk and O. Özcan, ''GestAnalytics: Experiment and analysis tool for gesture-elicitation studies,'' in *Proc. ACM Conf. Companion Publication Designing Interact. Syst.*, Jun. 2017, pp. 34–38, doi: 10.1145/3064857.3079114.

[45] N. Magrofuoco and J. Vanderdonckt, ''Gelicit: A cloud platform for distributed gesture elicitation studies,'' *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. EICS, pp. 1–41, Jun. 2019, doi: 10.1145/3331148.

[46] A. X. Ali, M. R. Morris, and J. O. Wobbrock, ''Crowdsourcing similarity judgments for agreement analysis in end-user elicitation studies,'' in *Proc. 31st Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2018, pp. 177–188, doi: 10.1145/3242587.3242621.

[47] A. X. Ali, M. R. Morris, and J. O. Wobbrock, ''Crowdlicit: A system for conducting distributed end-user elicitation and identification studies,'' in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2019, pp. 1–12, doi: 10.1145/3290605.3300485.

[48] C. Silpasuwanchai and X. Ren, ''Designing concurrent full-body gestures for intense gameplay,'' *Int. J. Hum.-Comput. Stud.*, vol. 80, pp. 1–13, Aug. 2015, doi: 10.1016/j.ijhcs.2015.02.010.

[49] H. Wu, J. Gai, Y. Wang, J. Liu, J. Qiu, J. Wang, and X. Zhang, ''Influence of cultural factors on freehand gesture design,'' *Int. J. Hum.-Comput. Stud.*, vol. 143, Nov. 2020, Art. no. 102502, doi: 10.1016/j.ijhcs.2020.102502.

[50] P. Childs, J. Han, L. Chen, P. Jiang, P. Wang, D. Park, Y. Yin, E. Dieckmann, and I. Vilanova, ''The creativity diamond—A framework to aid creativity,'' *J. Intell.*, vol. 10, no. 4, p. 73, Sep. 2022, doi: 10.3390/jintelligence10040073.

[51] B.-F. Gheran, J. Vanderdonckt, and R.-D. Vatavu, ''Gestures for smart rings: Empirical results, insights, and design implications,'' in *Proc. Designing Interact. Syst. Conf.*, Jun. 2018, pp. 623–635, doi: 10.1145/3196709.3196741.

[52] A. S. Williams, J. Garcia, and F. Ortega, ''Understanding multimodal user gesture and speech behavior for object manipulation in augmented reality using elicitation,'' *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 12, pp. 3479–3489, Dec. 2020, doi: 10.1109/TVCG.2020.3023566.

[53] A. S. Williams and F. R. Ortega, ''Understanding gesture and speech multimodal interactions for manipulation tasks in augmented reality using unconstrained elicitation,'' *Proc. ACM Hum.-Comput. Interact.*, vol. 4, pp. 1–21, Nov. 2020, doi: 10.1145/3427330.

[54] X. Zhou and R. Bai, ''User-defined mid-air gestures for multiscale GIS interface interaction,'' *Cartography Geograp. Inf. Sci.*, vol. 50, no. 5, pp. 481–494, Sep. 2023, doi: 10.1080/15230406.2023.2183900.

[55] A. Köpsel and N. Bubalo, ''Benefiting from legacy bias,'' *Interactions*, vol. 22, no. 5, pp. 44–47, Aug. 2015.

[56] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, ''User-defined gestures for surface computing,'' in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2009, pp. 1083–1092.

[57] A. Ali, M. Ringel Morris, and J. O. Wobbrock, '''I am iron man': Priming improves the learnability and memorability of user-elicited gestures,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–14, doi: 10.1145/3411764.3445758.

[58] M. Chamunorwa, M. P. Wozniak, S. Krämer, H. Müller, and S. Boll, ''An empirical comparison of moderated and unmoderated gesture elicitation studies on soft surfaces and objects for smart home control,'' *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. MHCI, pp. 1–24, Sep. 2023, doi: 10.1145/3604245.

[59] P. Ganapathi and K. Sorathia, ''Elicitation study of body gestures for locomotion in HMD-VR interfaces in a sitting-position,'' in *Proc. Motion, Interact. Games*, Oct. 2019, pp. 1–10, doi: 10.1145/3359566.3360059.

[60] T. Vuletic, C. McTeague, G. Campbell, L. Hay, and M. Grealy, ''Effects of activity time limitation on gesture elicitation for form creation,'' *J. Eng. Des.*, vol. 34, no. 11, pp. 963–985, Nov. 2023, doi: 10.1080/09544828.2023.2271773.

[61] J. Ruiz and D. Vogel, ''Soft-constraints to reduce legacy and performance bias to elicit whole-body gestures with low arm fatigue,'' in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Apr. 2015, pp. 3347–3350.

[62] A. E. Uva, M. Fiorentino, V. M. Manghisi, A. Boccaccio, S. Debernardis, M. Gattullo, and G. Monno, ''A user-centered framework for designing midair gesture interfaces,'' *IEEE Trans. Hum.-Mach. Syst.*, vol. 49, no. 5, pp. 421–429, Oct. 2019, doi: 10.1109/THMS.2019.2919719.

[63] Y. Tan, S. H. Yoon, and K. Ramani, ''BikeGesture: User elicitation and performance of micro hand gesture as input for cycling,'' in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, May 2017, pp. 2147–2154, doi: 10.1145/3027063.3053075.

[64] M. Chamunorwa, M. P. Wozniak, S. Vöge, H. Müller, and S. C. J. Boll, ''Interacting with rigid and soft surfaces for smart-home control,'' *Proc. ACM Hum.-Comput. Interact.*, vol. 6, pp. 1–22, Sep. 2022, doi: 10.1145/3546746.

[65] C. BeŞevli, O. T. Buruk, M. Erkaya, and O. Özcan, ''Investigating the effects of legacy bias: User elicited gestures from the end users perspective,'' in *Proc. ACM Conf. Companion Publication Designing Interact. Syst.*, May 2018, pp. 277–281, doi: 10.1145/3197391.3205449.

[66] M. R. Morris, A. Danielescu, S. Drucker, D. Fisher, B. Lee, M. C. Schraefel, and J. O. Wobbrock, ''Reducing legacy bias in gesture elicitation studies,'' *Interactions*, vol. 21, no. 3, pp. 40–45, May 2014.

[67] M. Perera, T. Gedeon, A. Haller, and M. Adcock, ''Using virtual reality to overcome legacy bias in remote gesture elicitation studies,'' in *Human-Computer Interaction* (Lecture Notes in Computer Science), M. Kurosu and A. Hashizume, Eds. Cham, Switzerland: Springer, 2023, pp. 200–225, doi: 10.1007/978-3-031-35596-7_14.

[68] I. Aslan, T. Schmidt, J. Woehrle, L. Vogel, and E. André, ''Pen + mid-air gestures,'' in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 135–144, doi: 10.1145/3242969.3242979.

[69] A. S. Williams, J. Garcia, F. De Zayas, F. Hernandez, J. Sharp, and F. R. Ortega, ''The cost of production in elicitation studies and the legacy bias-consensus trade off,'' *Multimodal Technol. Interact.*, vol. 4, no. 4, p. 88, Dec. 2020, doi: 10.3390/mti4040088.

[70] L. Hoff, E. Hornecker, and S. Bertel, ''Modifying gesture elicitation: Do kinaesthetic priming and increased production reduce legacy bias?'' in *Proc. 10th Int. Conf. Tangible, Embedded, Embodied Interact.*, Feb. 2016, pp. 86–91, doi: 10.1145/2839462.2839472.

[71] F. R. Ortega, K. Tarre, M. Kress, A. S. Williams, A. B. Barreto, and N. D. Rishe, ''Selection and manipulation whole-body gesture elicitation study in virtual reality,'' in *Proc. IEEE Conf. Virtual Reality 3D User Interface (VR)*, Mar. 2019, pp. 1723–1728.

[72] H. Wu, J. Liu, J. Qiu, and X. Zhang, ''Seeking common ground while reserving differences in gesture elicitation studies,'' *Multimedia Tools Appl.*, vol. 78, no. 11, pp. 14989–15010, Nov. 2018, doi: 10.1007/s11042-018-6853-0.

[73] Y. Wang, C. Yu, Y. Zhao, J. Huang, and Y. Shi, ''Defining and analyzing a gesture set for interactive TV remote on touchscreen phones,'' in *Proc. IEEE 11th Int. Conf. Ubiquitous Intell. Comput. IEEE 11th Int. Conf. Autonomic Trusted Comput. IEEE 14th Int. Conf. Scalable Comput. Commun. Associated Workshops*, Dec. 2014, pp. 362–365, doi: 10.1109/UIC-ATC-ScalCom.2014.84.

[74] K. Zhu, X. Ma, H. Chen, and M. Liang, ''Tripartite effects: Exploring users' mental model of mobile gestures under the influence of operation, handheld posture, and interaction space,'' *Int. J. Hum.-Comput. Interact.*, vol. 33, no. 6, pp. 443–459, Jun. 2017, doi: 10.1080/10447318.2016.1275432.

[75] E. McAweeney, H. Zhang, and M. Nebeling, ''User-driven design principles for gesture representations,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–13, doi: 10.1145/3173574.3174121.

[76] S. Bhowmick, K. Sorathia, and P. Chandra Kalita, ''Understanding gesture performance for object selection in VR: Classification and taxonomy of gestures in HCI,'' in *Proc. India HCI*, Feb. 2022, pp. 90–93, doi: 10.1145/3506469.3506483.

[77] Q. Wei, J. Hu, and M. Li, ''User-defined gestures for mediated social touch on touchscreens,'' *Pers. Ubiquitous Comput.*, vol. 27, no. 2, pp. 271–286, Apr. 2023, doi: 10.1007/s00779-021-01663-9.

[78] Z. Chen, H. Tu, and H. Wu, ''User-defined foot gestures for eyes-free interaction in smart shower rooms,'' *Int. J. Hum.-Comput. Interact.*, vol. 39, no. 20, pp. 4139–4161, Dec. 2023, doi: 10.1080/10447318.2022.2109260.

[79] S. Villarreal-Narvaez, J. L. Perez-Medina, and J. Vanderdonckt, ''Exploring user-defined gestures for lingual and palatal interaction,'' *J. Multimodal User Interface*, vol. 17, no. 3, pp. 167–185, Sep. 2023, doi: 10.1007/s12193-023-00408-7.

[80] R. Arnheim and D. McNeill, ''Hand and mind: What gestures reveal about thought,'' *Leonardo*, vol. 27, no. 4, p. 358, 1994.

[81] S. Kim and S. Lee, ''Touchable pixels: Examining the affordance effect between an on-screen object and a user-elicited gesture on the touchscreen,'' *Comput. Hum. Behav.*, vol. 140, Mar. 2023, Art. no. 107588, doi: 10.1016/j.chb.2022.107588.

[82] S. Kim and S. Lee, ''Touch digitality: Affordance effects of visual properties on gesture selection,'' in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–8, doi: 10.1145/3334480.3382914.

[83] R. Rädle, H.-C. Jetter, M. Schreiner, Z. Lu, H. Reiterer, and Y. Rogers, ''Spatially-aware or spatially-agnostic? Elicitation and evaluation of user-defined cross-device interactions,'' in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Apr. 2015, pp. 3913–3922, doi: 10.1145/2702123.2702287.

[84] M. R. Morris, J. O. Wobbrock, and A. D. Wilson, ''Understanding users' preferences for surface gestures,'' in *Proceedings of Graphics Interface 2010*. Toronto, ON, Canada: Canadian Information Processing Society, 2010, pp. 261–268.

[85] X. Zhou, A. S. Williams, and F. R. Ortega, ''Eliciting multimodal Gesture+Speech interactions in a multi-object augmented reality environment,'' in *Proc. 28th ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2022, pp. 1–10, doi: 10.1145/3562939.3565637.

[86] S. Navas Medrano, M. Pfeiffer, and C. Kray, ''Enabling remote deictic communication with mobile devices: An elicitation study,'' in *Proc. 19th Int. Conf. Hum.-Comput. Interact. with Mobile Devices Services*, Sep. 2017, pp. 1–13.

[87] E. L. Hill, D. V. M. Bishop, and I. Nimmo-Smith, ''Representational gestures in developmental coordination disorder and specific language impairment: Error-types and the reliability of ratings,'' *Hum. Movement Sci.*, vol. 17, nos. 4–5, pp. 655–678, Aug. 1998, doi: 10.1016/s0167-9457(98)00017-7.

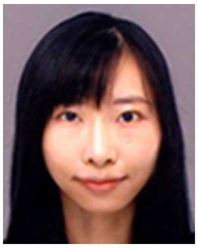
[88] M. Hosseini, T. Ihmels, Z. Chen, M. Koelle, H. Müller, and S. Boll, ''Towards a consensus gesture set: A survey of mid-air gestures in HCI for maximized agreement across domains,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2023, pp. 1–24, doi: 10.1145/3544548.3581420.

[89] S. Kallio, J. Kela, and J. Mantyjarvi, ''Online gesture recognition system for mobile interaction,'' in *Proc. IEEE Int. Conf. Syst., Man Cybern., Conf. Theme-Syst. Secur. Assurance*, Oct. 2003, pp. 2070–2076, doi: 10.1109/ICSMC.2003.1244189.

[90] A. Lascarides and M. Stone, ''A formal semantic analysis of gesture,'' *J. Semantics*, vol. 26, no. 4, pp. 393–449, Nov. 2009, doi: 10.1093/jos/ffp004.

YUTING CHENG received the Bachelor of Engineering degree from the South China University of Technology, in 2021. She is currently pursuing the master's degree in design with Shanghai Jiao Tong University. Her research interest includes gesture elicitation design in intelligent environment interactions.

ZHANWEI WU received the Ph.D. degree from Shanghai Jiao Tong University, China. He is currently an Associate Professor with Shanghai Jiao Tong University and the Director of Digital China and Innovation Design Studio. His research interests include human-centered digital innovation, experience technology, and design.



RUOWEI XIAO (Member, IEEE) was a Postdoctoral Researcher with the Gamification Group, Tampere University, Finland, for three years. She is currently an Assistant Professor with the School of Design, Southern University of Science and Technology. Her research work has encompassed multidisciplinary contributions in the areas of ubiquitous computing, gamification, interactive experience and design, applied variously to smart city, entertainment, and education.