Int. J. Human–Computer Studies 183 (2024) 103198

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Elicitation and Evaluation of Hand-based Interaction Language for 3D Conceptual Design in Mixed Reality

Lingyun Sun [a], Hongbo Zhang [a], Pei Chen [a,b,*], Zhaoqu Jiang [a], Xuelong Xie [a], Zihong Zhou [a], Xuanhui Liu [c], Xiaoyu Chen [c]

[a] College of Computer Science and Technology, Zhejiang University, Hangzhou, 310027, China
[b] Zhejiang-Singapore Innovation and AI Joint Research Lab, Hangzhou, 310027, China
[c] School of Art & Archaeology, Hangzhou City University, Hangzhou, 310015, China

A R T I C L E I N F O

Keywords:
hand-based interaction
conceptual design
mixed reality
embodied cognition

A B S T R A C T

Conceptual design is a fundamental stage in a design process. Traditional conceptual design tools impose limitations on the intuitive creation process of 3D objects, hindering their full potential. Applying hand-based interaction in mixed reality (MR) provides an immersive and intuitive creation mode without dimensionality transformations, facilitating conceiving ideas in 3D conceptual design. Although studies on hand-based languages have explored different gesture design methodologies and techniques from many individual perspectives, they lack holistic understanding and comprehensive suggestions for gesture application in conceptual design scenarios. This study aims to establish a set of hand-based languages that are available in cognitive, physical, and system-based aspects for 3D conceptual designs in the MR environment. A two-stage study was conducted in an MR environment, combining an elicitation design and a comprehensive evaluation experiment. Through the elicitation design, approximately 930 gesture actions, focusing on 31 targeted functionalities, were collected. By performing an evaluation experiment, a set of theoretically optimal hand-based interaction languages for 3D conceptual design was proposed, and the corresponding guidelines for hand-based interactions were clarified. Our results can further expand human-computer interactions in MR environments and inspire software builders to create novel hand-driven interaction modes or tools.

1. Introduction

Conceptual design is a fundamental stage in a product design process, where vague ideas are shaped into a product layout (Gomes et al., 2022) and basic decisions are made (Wang et al., 2002; Zou et al., 2023). Traditional sketching tools, including pencils, clay, and blocks, support the rapid three-dimensional (3D) sketches in conceptual design. However, these tools have some limitations, such as high technical requirements, a single perspective, and material limitations. With the recent technological advancements, mixed reality (MR), encompassing virtual reality (VR) and augmented reality (AR), has been implemented across all design stages, thereby revolutionizing the conceptual design process (Zou et al., 2023).

MR provides the intuitive human-computer interaction (HCI), overcoming the limitations of traditional tools (Ye and Campbell, 2006). For sketching, it offers benefits such as intuitive data entry, convenient perspective shifts, and material freedom. In terms of idea generation, MR enhances creativity and improves the emotional experience of design activities through an immersive virtual environment (Yang et al., 2018; Rieuf et al., 2017). Considering that its interactive and immersive nature is consistent with the designer's requirements in 3D creation, it is important to further explore MR's interaction modes and applications in 3D conceptual design. The hand-based interaction, which is a natural expression of mental concepts, is a common input method for MR-supported spaces (Memo and Zanuttigh, 2018). This form of interaction also aligns with conceptual design requirements because of its intuitiveness and expressive capabilities. Specifically, hand-based interaction provides a natural interface (Alkemade et al., 2017), supports creativity by linking designer perceptions and actions (Gibbs, 2006), avoids the restriction of design thinking flow through imprecise input (Van Dijk, 1995), reduces learning costs, and improves communication efficiency (Wang et al., 2002) in the conceptual design.

This study aims to establish a comprehensive set of hand-based languages that are widely available for 3D conceptual design in an MR

* Corresponding author: Pei Chen, College of Computer Science and Technology, Zhejiang University, Hangzhou, 310027, China
E-mail address: chenpei@zju.edu.cn (P. Chen).

https://doi.org/10.1016/j.ijhcs.2023.103198
Received 28 March 2023; Received in revised form 19 August 2023; Accepted 27 November 2023
Available online 30 November 2023
1071-5819/

environment. Although previous studies have extensively explored hand-based interactions, there are potential challenges that require further investigation and resolution. Specifically, previous studies have not focused on the specific application scenario of 3D conceptual design in MR to establish hand-based languages and propose guidelines. Some studies have proposed a mid-air gesture set for daily operations (Huang et al., 2021) or common scenarios (Lyu et al., 2023), such as Confirm and Drag. However, there is no universal hand-based language for complicated applications because hand-based vocabulary must be tailored for a specific task as well as match a specific program and user group (Nielsen et al., 2004). Besides, there exists insufficient holistic understanding of hand-based interactions and comprehensive evaluation of hand-based languages. Hand-based language methodologies are categorized into three groups: developer-defined, user-defined, and computationally defined. The first mainly focuses on the developers' expertise and experience (Burnett et al., 2013; Garzotto and Valoriani, 2013; Hinckley and Song, 2011); the second relies on the users' mental models (Wobbrock et al., 2009); and the third focuses on the identification of algorithms (Ashbrook and Starner, 2010; Dey et al., 2004; Lü and Li, 2013). However, previous studies have not holistically evaluated the amalgamation of diverse factors and proposed a comprehensive guideline.

In this study, we focused on a 3D conceptual design scenario and investigated hand-based languages. We conducted a two-stage experimental study and applied an assessment system that combines a user elicitation design and comprehensive measurements. Fig. 1 shows the workflow, methods, and outcomes of this study. First, 31 fundamental and non-overlapping atomic functionalities suitable for hand-based interactions in 3D conceptual design were summarized and defined through expert interviews. Subsequently, approximately 930 gesture actions were collected using an elicitation design with 12 participants. Finally, 65 high-frequency hand-based vocabulary items were extracted for further comparison and analysis. A comprehensive evaluation experiment was then conducted with an interdisciplinary group of 24 participants that included the assessment indices of discoverability, learnability, memorability, efficiency, and algorithm recognition. Our study encompassed expert knowledge, user preferences, and technical feasibility to ensure a holistic assessment of hand-based vocabularies in 3D conceptual design.

The contributions of this study are as follows. (1) We conducted a holistic experiment combining user elicitation design and a comprehensive measurement to evaluate hand-based interaction vocabularies. (2) We provided a set of theoretically optimal hand-based interaction languages for common functionalities in a 3D conceptual design. (3) We clarified the HCI guidelines for hand-based driven interactions in an MR environment for the conceptual design.

2. Related work

2.1. The 3D conceptual design in mixed-reality (MR) environment

In the conceptual design phase of 3D products, such as architectural, industrial, and game designs, designers strive to conceptualize spatial problems, externalizing their vague ideas via sketching (Alkemade et al., 2017). Currently, many designers employ various CAD software, such as AutoCAD, SolidWorks, and Alias, to improve the efficiency of 3D drawings. However, these tools require the compression of 3D reality into a 2D interface and subsequent decompression into a 3D virtual world, impeding an intuitive understanding of the spatial dimensions (Alkemade et al., 2017).

With advancements in input and display technologies, the MR technology has sidestepped this dimensionality transformation, offering novel interaction modes and visualization methods for 3D sketching. For example, CASSIE transposes sparse drawing strokes in 3D space into a 3D mesh (Yu et al., 2021), whereas SketchUp integrates drawing layouts and surface rendering into a virtual environment (Armin Rigo, 2020). In addition to drawing, MR-sculpting tools have been developed to assist users in creating shapes and deducing concepts in 3D virtual environments (Kodon, 2018; Oculus Medium, 2016; Shape Lab, 2018). Additionally, 3D sketches can be created by sweeping a virtual space along a path and then further manipulating it using free-form and parametric shaping and deformation (Vinayak et al., 2013; Huang et al., 2018).

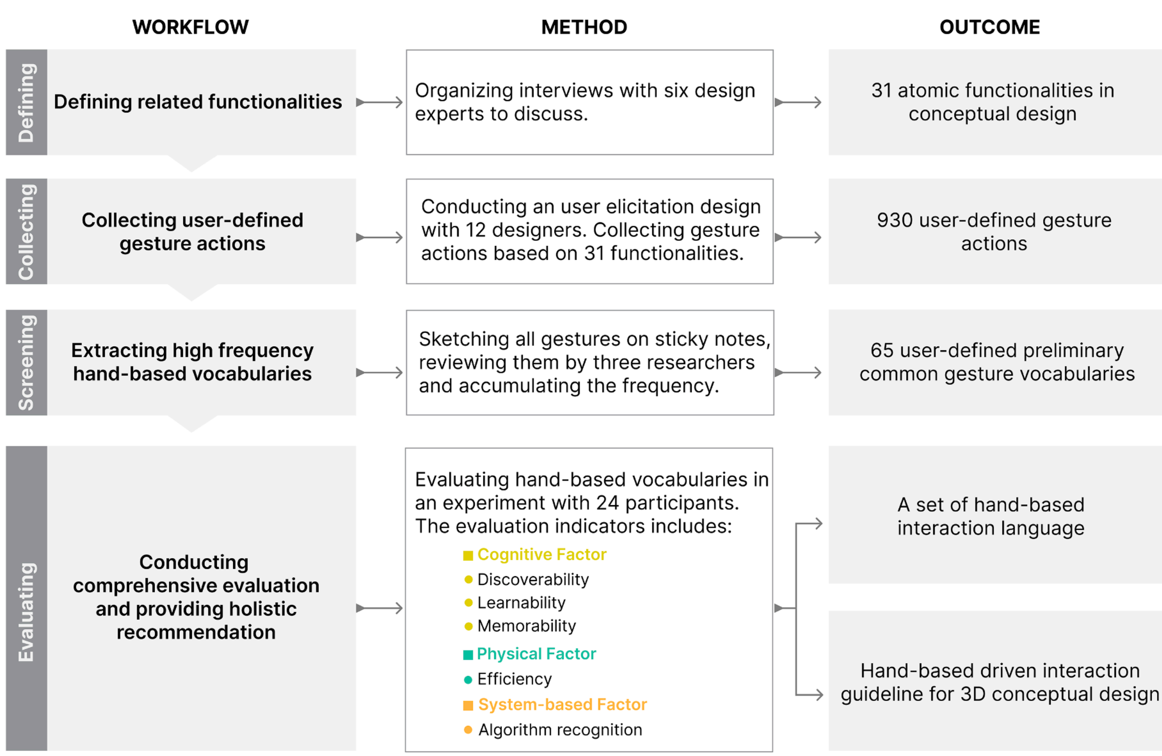


Fig. 1. Workflow, method, and outcome in this study.

Moreover, the signifier-based approach (Rentzen et al., 2019) realizes 3D conceptual modeling by directly manipulating the polygonal mesh vertices. These technological developments have increased the efficiency of idea externalization and sketch shaping in 3D design, bringing both opportunities and challenges to HCI. It is crucial to further explore 3D conceptual design interaction modes in an MR environment to reduce the cognitive load by providing users with a sense of direct interaction with the object itself (Hutchins et al., 1985).

2.2. Hand-based interface application in HCI

Hand-based interaction is a core mode of human expression that extends beyond interpersonal communication to manipulation of surrounding objects (Buchmann et al., 2004; Zimmerman et al., 1987). Over the past decade, gesture input has been increasingly integrated into various human-computer interfaces. Currently, hand-based interactions are employed in fields such as healthcare and automotive environments, enabling touchless interaction with medical imagery to minimize contamination risk (Lopes et al., 2017) and facilitating comfort function adjustments without diverting the gaze from the road (Riener et al., 2013; Vuletic et al., 2019), respectively. Numerous MR applications utilize hand-based interfaces to heighten immersion (Deller et al., 2006), especially in creative-oriented applications (Arora et al., 2017; Drey et al., 2020).

The application of gesture language is amplified when interaction involves the physicality of 3D objects, such as in 3D architectural urban planning (Buchmann et al., 2004), CAD design (Huang et al., 2018), and virtual pottery (Vinayak and Ramani, 2015). These areas rely on spatial visualization and 3D object perception. During 3D creation, freeform gestures are predominantly used for splines or surface generation (Arroyave et al., 2015). Previous studies emphasized that hand gestures enhance spatial concept expression during conceptual design (Vinayak et al., 2013). Similarly, Van Dijk (1995) suggested that hand movements complement the thought process in conceptual design, facilitating object shaping. Given the unique advantages of hand-based interfaces in 3D conceptual design, this study aims to extract and develop a set of hand-based interaction languages for targeted functions within the conceptual design process to improve the efficiency and usability of hand-based interactions.

2.3. Approach for creating hand-based interaction language

The main methodologies used to design hand-based languages are categorized into three groups: developer-defined, user-defined, and computationally defined approaches. The developer-defined methodology draws from experts' knowledge and experience (Aigner et al., 2012). Developers create hand-based languages using previously published hand-based vocabularies (Wobbrock et al., 2007), literature recommendations (Xia et al., 2016), observations of analog interactions (Hinckley et al., 2014), or design experience (Wigdor et al., 2011). Although this allows for rapid hand-based vocabulary creation, gesture language learnability and discoverability could suffer because of the discrepancy between the developer and end-user perspectives.

The user-defined methodology is based on eliciting gesture vocabularies and feedback from users. This approach, introduced by Wobbrock et al. (2009), has been widely adopted in various gesture interaction scenarios, such as freehand TV control (Morris, 2012; Vatavu, 2012), smart glasses (Tung et al., 2015), unmanned aerial vehicles (Peshkova et al., 2016), and deformable displays (Troiano et al., 2014). The variants of Wobbrock et al. (2009) were also proposed. The major difference among these variants is that studies differ in when users become engaged in the participatory design (Xia et al., 2022). For example, Löcken et al. (2012) invited users as part of the requirements analysis whereas Nielsen et al. (2004) began engagement after the requirements analysis was completed. Although the user-defined approach is one of the most popular and common methodologies in gesture creation, the gestures designed by users are the appropriations of existing gestures or are limited by the technological capabilities of their past experiences, which might neglect other important factors.

The computationally defined approach focuses on gesture recognition using computer systems. It uses pattern matching or learning algorithms to discern gestures based on template similarity or gesture input recognition (Xia et al., 2022). For example, Vatavu et al. (2014) utilized visualization techniques to analyze the gestures. Although this methodology can produce computer-friendly gestures, it may ignore some cognitive factors, such as gesture memorability, ease of learn, and logical functionality mapping.

3. Elicitation design

This experiment aimed to extract hand-based vocabularies by elicitation design and investigate gesture preferences during 3D conceptual design. This section comprises three parts: defining functionalities, collecting gesture actions, and extracting hand-based vocabularies.

3.1. Functionality defining

This study builds on prior work that asserts the task-specific nature of hand-based vocabulary (Nielsen et al., 2004) and tries to identify the functionalities required for 3D conceptual design. We adopted a top-down approach in which expert interviews were used to gather these functionality requirements, given that the experts' extensive experience and knowledge permit the effective identification of essential and distinct commands.

Six experts, each with over a decade of design experience, were invited to participate in the individual interviews. These experts identified the functionalities commonly applied in a major modeling software using sticky notes, including AutoCAD, Rhino, Blender, Alias, 3D Max, SolidWorks, and Creo Inventor. They were also asked to discuss the suitability of the functionalities for gesture modeling. Experts discussed based on the criteria for selecting a friendly functionality: whether it is commonly used in 3D creation scenarios and whether it is suitable to be conducted by gestures, which takes into consideration its prevalence within the realm of 3D conceptual design and its compatibility with hand-based interaction. Six experts were invited to participate because no additional valuable functionalities emerged beyond the fourth expert. This phenomenon is consistent with previous findings (Guest et al., 2006), suggesting that there might be a saturation point in qualitative research. In other words, the acquisition of new information becomes limited beyond a certain number of interviews.

After the expert interviews, the functional requirements were split into atomic physical acts to identify a basic set of functionalities (Aigner et al., 2012). Table 1 lists the 31 fundamental and non-overlapping atomic functionalities. We do not claim the list to be complete for 3D manipulation because the experts focused more on the 3D conceptual design process. Some commands for complex surface modeling, such as surface sweeping, were not considered.

3.2. Elicitation design

3.2.1. Participants

We recruited 12 participants (six males and six females) from different design majors (i.e., industrial design, art design, mechanical design, and interior design). Participants were 24.2 years old (SD = 1.85) on average and all were right-handed without musculoskeletal symptoms. Participants had an average of 4.5 years of design experience. They were familiar with the entire process of 3D product design and were skilled in using common 3D design software.

3.2.2. Experiment apparatus

The layout of the elicitation design environment is shown in Fig. 2. Participants were seated at a table and wore a head-mounted display

Table 1
Fundamental functionalities extracted from expert interview.

No.	Functionality	Command Description
F1	*Open Menu*	Open menus for accessing options and Operators.
F2	*Close Menu*	Close menus without activating any menu item.
F3	*Save*	Save files.
F4	*Confirm*	Confirm the upcoming action.
F5	*Cancel*	Cancel the upcoming action.
F6	*Undo*	Undo your last action.
F7	*Redo*	Redo your last action.
F8	*Zoom in*	Zoom the view in.
F9	*Zoom out*	Zoom the view out.
F10	*Rotate View*	Roll the view freely.
F11	*Select all*	Select all selectable objects.
F12	*Group*	Create a group for selected objects.
F13	*Ungroup*	Remove the group and places the individual objects into editor workspace.
F14	*Hide*	Hide all selected objects.
F15	*Show*	Reveal all hidden objects.
F16	*Lock*	Lock selected objects from being editable.
F17	*Unlock*	Unlock selected objects for editing.
F18	*Add Cube*	Interactively add a cube mesh object.
F19	*Add Sphere*	Interactively add a sphere mesh object.
F20	*Add Cylinder*	Interactively add a cylinder mesh object.
F21	*Delete*	Remove the selected object from the current scene.
F22	*Copy*	Reproduce selected object in another location.
F23	*Translation*	Move object or change location of objects.
F24	*Scale up*	Enlarge and change the scale of the object.
F25	*Scale down*	Shrink and change the scale of the object.
F26	*Rotate*	Change the orientation of elements.
F27	*Twist*	Bend or turn objects into a particular shape.
F28	*Extrude Region*	Extrude vertices and keep the new geometry connected with the original one.
F29	*Cut*	Interactively subdivide (cut up) geometry
F30	*Fillet Edge*	Smooth out edges and corners.
F31	*Loft*	Form a geometry between two mesh.

(HMD), HoloLens 2. The "before" and "after" situations of each functionality were programmed and played by the Unity software on a laptop computer. Then, they were presented to participants through Hololens 2. This experimental scene in MR was able to ensure that participants could see their virtual hand and simulate the design work in the MR environment as accurately as possible. On their chest, participants wore a portable camera (GoPro Hero 10) that focused on their hand gestures. Another camera (DJI Pocket 2) in front of participants was used to record the entire process and communication during the experiment. This study was approved by the Research Ethics Office of the university. Participants were apprised of and consented to the experimental video recordings prior to their involvement. Subsequent data processing was restricted to gesture actions, eschewing private data such as facial features or voices, to eliminate disclosure risks. This study has no other ethical or privacy impacts.

3.2.3. Procedure

During the gesture elicitation design, participants seated with their hands naturally resting on their thighs and performed the gesture tasks. Participants again interpreted detailed descriptions of the experiment, and the experimenter explicitly asked for confirmation before moving on. The "before" and "after" states of targeted functionalities were presented to participants (see Fig. 3 for an example and Appendix 1 for the full set of items). Participants were instructed to design three distinct gesture actions per functionality, thereby bridging the state transition. This strategy echoed that of Wu et al. (2019), wherein three gestures were derived in the initial phase to successfully mitigate the influence of end users' preconceived biases. After completing the gestures, participants clarified their design intentions. Participants were encouraged to create three different hand-based vocabularies for the same functionality, but fewer than three gestures were permitted if they thought it was difficult and impossible to complete. Initially, two extra functionalities were provided to participants to familiarize them with the experimental procedure before contacting the 31 targeted functionalities. The presentation order of the functionalities was counterbalanced to control the order effects. The experiment lasted for approximately 40 min.

3.3. Findings

3.3.1. Extracting the preliminary common hand-based vocabulary

The elicitation design yielded 930 gestures from 12 participants, some of whom exhibited similar characteristics. Three researchers counted the frequency of all gestures by coding the videos recorded using the two cameras. Each gesture was individually assessed, defined, and sketched for the analysis. Similar gestures were amalgamated to simplify frequency accumulation by employing regular discussions to deter bias and varied interpretations, as advised by Wu et al. (2019). For example, gesture actions that shared design intentions but differed in range or speed were grouped together. Thus, a preliminary set of common hand-based vocabulary was drawn based on frequency. Notably, the results of gesture frequency do not directly represent the recommendation order of the hand-based vocabulary. Preferences may change across elicitation experiments with the same participants (Choi et al., 2014), because the most frequent gesture is not always optimal (Huang et al., 2021). The aim is to capture common gesture preferences to extract representative hand-based vocabularies for further evaluation.

We attempted to extract the three most frequent gestures for all 31 target functionalities. However, after counting and screening by researchers, there were fewer than three hand-based vocabularies under targeted functionalities (i.e., *Add Sphere* (*F19*), *Translation* (*F23*), *Rotate* (*F26*), *Twist* (*F27*), *Extrude Region* (*F28*), *Fillet Edge* (*F30*), *Loft* (*F31*)).

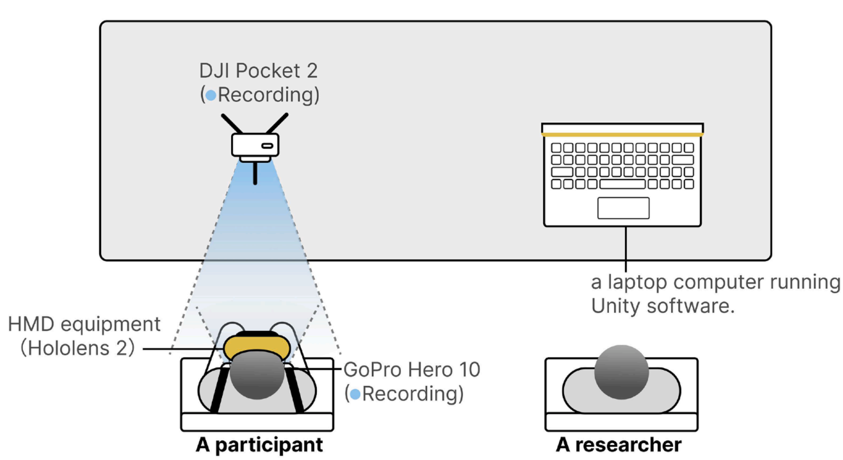


Fig. 2. Layout of the elicitation design.

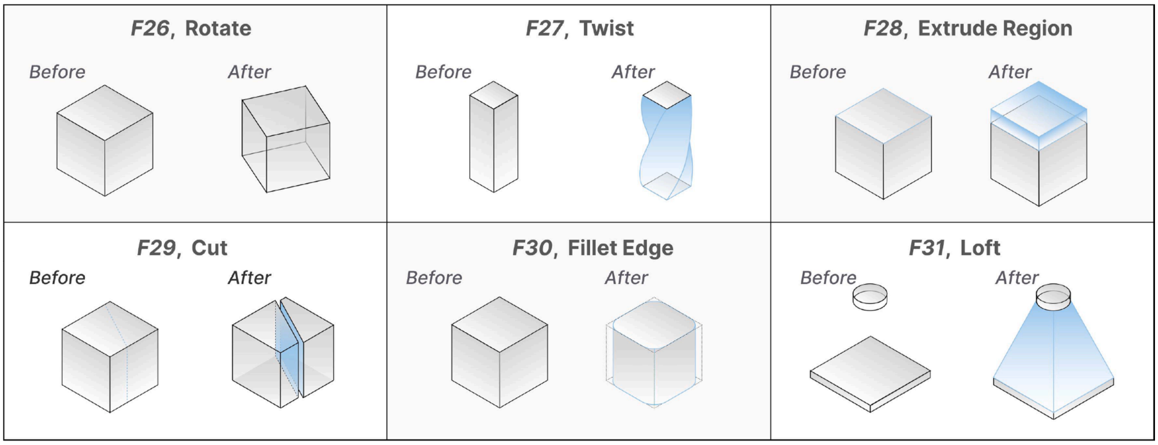


Fig. 3. Case examples of the "before" and "after" states for six targeted functionalities, with the definitions of states of all functionalities provided in Appendix 1.

Consequently, we extracted 86 distinct preliminary common vocabularies for 31 functionalities (Fig. 4). In the following sections, we comprehensively evaluate the preliminary common hand-based vocabularies. For the paired functionalities (e.g., *Open Menu & Close Menu, Undo & Redo*), we found that participants created the same gesture action but in the opposite order. Therefore, we considered only one of the paired functionalities for comprehensive evaluation.

3.3.2. Summarizing intention and thinking in gesture creation

Insights into participants' preferences in gesture creation were shown through dialogues during the elicitation design process and subsequent analysis of the experimental footage.

It was relatively easier for participants to create gesture actions for some functionalities that required spatial coordination or relationships with an object in the physical world, such as *Zoom in* (*F8*), *Translation* (*F23*), *Rotate* (*F26*), and *Twist* (*F27*). As these kinds of functionalities based on physical manipulation were intuitive, participants could interact in the MR environment by eye–hand coordination learned from the physical world. For example, moving the grabbing hand (*G48*) can move an object along with it in MR, and scaling can be performed by an interaction that mimics stretching of materials (Alkemade et al., 2017).

In contrast, it was challenging for participants to design gestures for functionalities lacking a corresponding relationship in the physical world and necessitating abstraction, such as *Save* (*F3*), *Confirm* (*F4*), *Hide* (*F14*) and *Lock* (*F16*). This is because they cannot borrow gesture actions directly from the physical world. In the elicitation design, participants were accustomed to creating gestures for this type of abstract functionality by combining their own cognitive experiences and linking gesture actions with functionality semantics. First, some "iconic gestures" were used to communicate information about objects or entities, such as specific sizes, shapes, and motion paths (Aigner et al., 2012). For example, participants formed an "O" with curved fingers and palms (*G37*), meaning "circle", to create hand-based vocabulary for *Add Sphere* (*F19*). Second, several "semaphoric gestures" were used to convey specific meanings (Quek et al., 2002). For example, the static gesture of pinching the forefinger and thumb (*G9*), meaning "Okay," was chosen to express *Confirm* (*F4*); and the dynamic gesture of flatting palm and shaking continuously (*G12*), meaning "refusal," was designed to convey *Cancel* (*F5*). Third, some "pantomimic gestures" were conducted to create hand-based vocabularies because participants preferred to perform or imitate a specific task related to functionality semantics. For example, participants simulated the action of turning the key (*G32*) to indicate *Lock* (*F16*) and covered their eyes with their palms (*G28, G29*) to illustrate *Hide* (*F14*). Fourth, participants combined their operating habits and experiences with traditional 2D software to create gesture actions for designing gestures in the MR environment. For example, participants wrote the path of "S" in mid-air (*G5*) to express *Save* (*F3*) because they were used to using the shortcut keys of "Command (⌘)" and "S" to express the same function when inputting by the keyboard in the traditional CAD software.

4. Comprehensive evaluation

The present experiment aimed to comprehensively assess and ascertain the optimal set of hand-based languages. Twenty-four participants were invited to complete the hand-based vocabulary evaluation using the HMD equipment in the MR environment. Behavioral feedback, operational performance, and motion data from the experiment were recorded for data analysis. Cognitive factors were evaluated using indicators of discoverability, learnability, and memorability. Physical and system-based factors were evaluated based on efficiency and algorithm recognition, respectively.

4.1. Participants

Twenty-four participants (10 males and 14 females) from different majors related to 3D conceptual design (P1–P18 from industrial design, P19–P21 from mechanical engineering design, P22–P23 from art design, and P24 from game design) were recruited through a university social network. This diverse selection was intentional owing to the collaborative and interdisciplinary nature of conceptual design. All participants were able-bodied and right-handed with an average age of 23.3 years (SD = 1.3). Table 2 presents the information on the anthropometric measures and design experiences of participants. The physiological data were measured in accordance with ISO 7250-1:2008. All participants were familiar with the 3D conceptual design process and related commands in the CAD software. This study was approved by the University Office of Research Ethics, and each participant confirmed the informed consent before beginning the experiment.

4.2. Experiment environment

Participants were asked to sit on a table with the HMD equipment (HoloLens 2). The preliminary common hand-based vocabularies extracted in Section 3.2 were filmed by researchers and presented to participants through the HMD to reduce the action differences among participants. A portable camera (GoPro Hero 10) was used to record videos of hand gestures. During the experiment, participants wore capture devices (Smartsuit Pro, Rokoko). Motion data were captured and recorded using smart suit sensors to measure the recognition rates of hand-based vocabularies. Fig. 5 illustrates the schematic of the experimental environment.

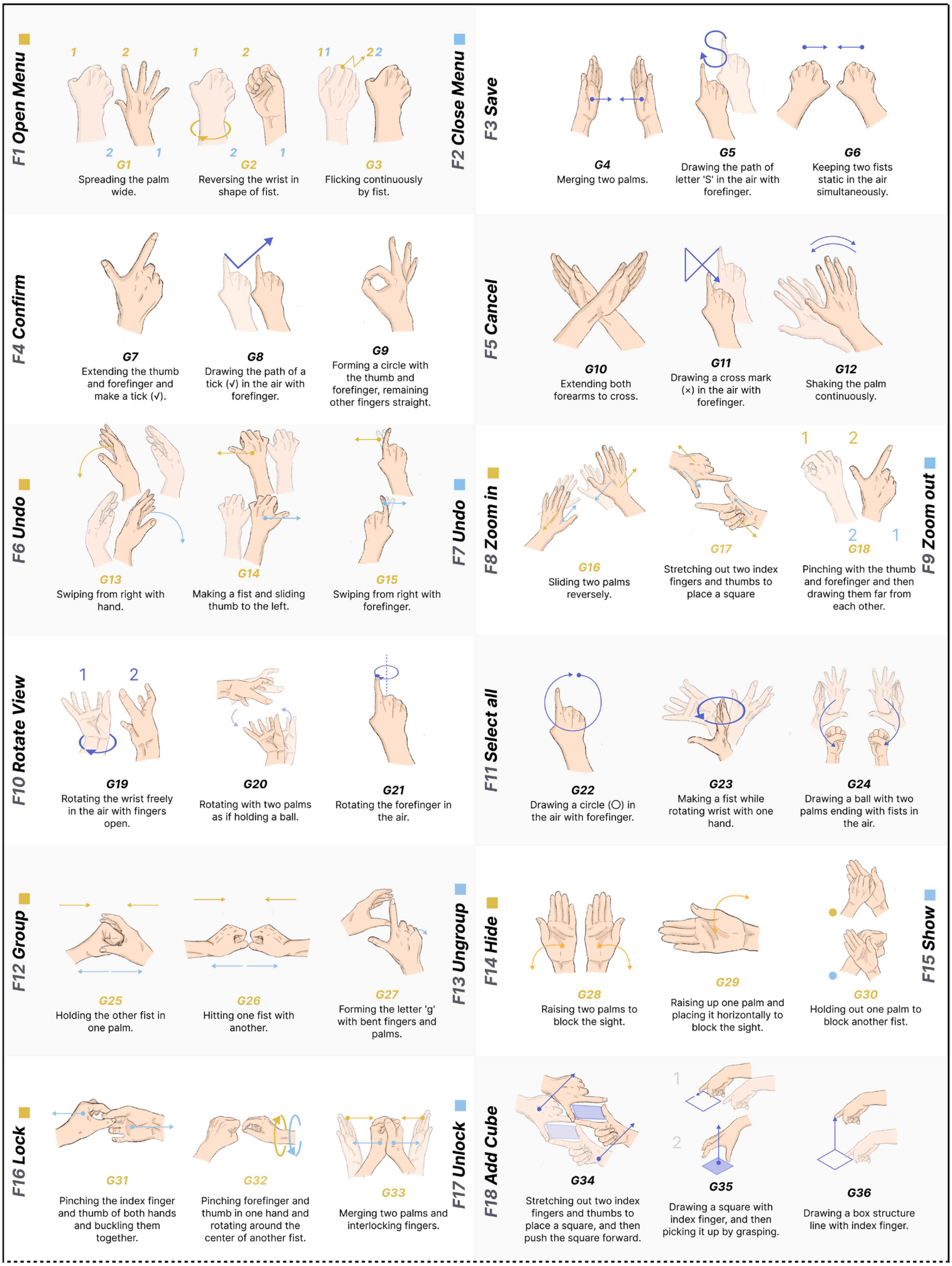


Fig. 4. Preliminary common hand-based vocabulary.

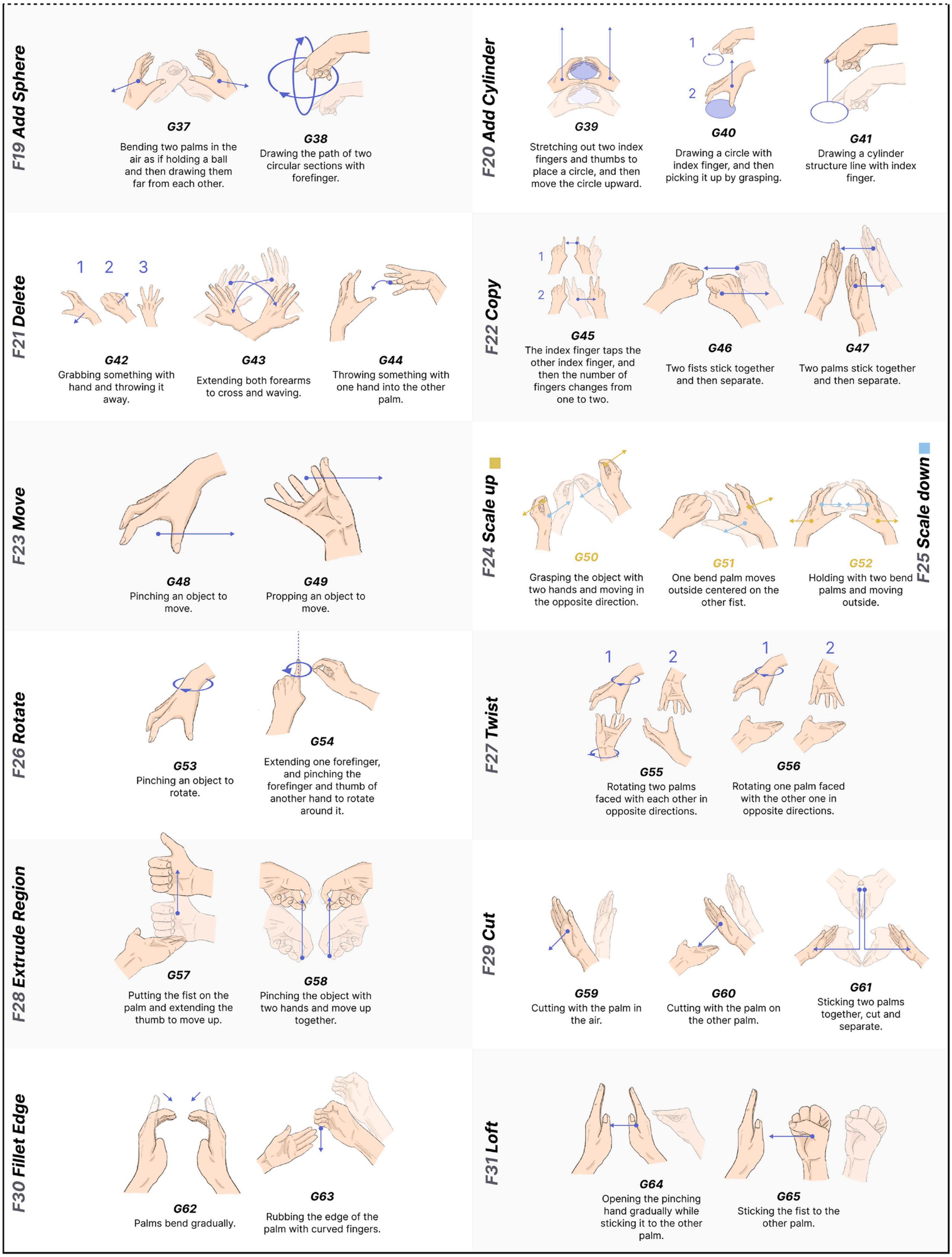


Fig. 4. (*continued*).

Table 2
Mean (SD) participant information.

	Age (yrs)	Hand length (mm)	Design experience(yrs)
Males (n = 10)	22.9 (1.4)	186.0 (8.3)	3.6 (1.0)
Females (n = 14)	23.6 (1.1)	173.0 (4.0)	3.8 (2.0)
All (n = 24)	23.3 (1.3)	178.4 (8.9)	3.7 (1.6)

4.3. Dependent variables

Cognitive, physical, and system-based factors were measured to comprehensively evaluate the hand-based vocabulary. All the evaluation factors are summarized in Table 3.

4.3.1. Cognitive factor

Cognitive factors are indispensable in hand-based language design. Previous studies indicated that the cognitive factors are important with user-defined, and occasionally with developer-defined methodologies (Annett and Bischof, 2013; Hinckley et al., 2010; Wobbrock et al., 2009). The discoverability, learnability, and memorability of hand-based vocabulary were measured in this study, which determined how intuitive it was to understand, how learnable it was to use, and how easy it was to remember.

4.3.1.1. Discoverability. The quality of gestures that enables a user to access intended referents is referred to as the discoverability of a gesture (Wobbrock et al., 2005), which is also called guessability, approachability, self-revealing, and metaphorically or iconically logical toward functionality (Xia et al., 2022). The discoverability was evaluated using the functionality-guessing approach proposed by Nielsen et al. (2004). Specifically, each hand-based vocabulary item was filmed as a video clip and presented to participants through a HMD. Participants were asked to guess and report the functionality represented by the gesture action after watching the video clip. Scoring was established based on the immediacy of correct identification after viewing. Functionalities that were identified instantly received three points, while successful identification after brief consideration yielded two or one points, based on whether it was within three or five view numbers. A score of zero was assigned to unsuccessful identification after five views. Researchers would not give any hints except "Yes" or "No" during test, and the video presentation order was randomized and balanced among participants.

4.3.1.2. Learnability. Learnability refers to the ease with which users can initiate effective interactions and achieve maximum performance through gestures (Dix et al., 2004). Learnability is one of the well-known factors in gesture language design and is often referred to as fast learning, recognition, and similarity. It fits well with its associated function, systematic chunking, mapping, uptake, or matching (Xia et al., 2022). Compared to discoverability, learnability emphasizes understanding a gesture rather than identifying its functional connections. The learnability scores were collected through a subjective evaluation using the Likert quintile scale (5 = Strongly High and 1 = Strongly Low). This was because the evaluation criteria for learnability were relatively subjective as they were built upon the users' collective past experiences and knowledge.

4.3.1.3. Memorability. Memorability refers to whether gestures can be effectively remembered (Nacenta et al., 2013). Despite its link to learnability, memorability is distinct and independent. Complicated or illogical gestures may be memorable because of their unique

Table 3
The summary of evaluation factors in hand-based vocabulary evaluation. ("■", "■", and "■" represent the cognitive, physical and system-based factor respectively.)

	Factor	Description	Evaluation
■	Discoverability	The quality of gestures that enables a user to understand and access intended functionality of those gestures.	Through guessing the corresponding functionality after experiencing the gesture action (Nielsen et al., 2004).
■	Learnability	The ease with which new users can begin effective interaction and achieve maximal performance with a gesture.	Through participants' subjective report via the Likert Quintile Scale (Xiao et al., 2021).
■	Memorability	The parameters of gestures that enables a user to effectively remember the hand-based vocabulary at a lower memory load.	Through measuring the number of gesture recalls after using a set of gesture languages (Nielsen et al., 2004).
■	Efficiency	The ease of performing or operating a gesture action with lower interaction cost and physical load.	Through measuring the average duration of gesture task completion time (Rubine, 1992).
■	Algorithm recognition	The performance of recognizing a gesture action correctly to its corresponding functionality by algorithm techniques.	Through setting up a gesture action data set and measuring the precision of recognition by the ResNet-152 (He et al., 2016).

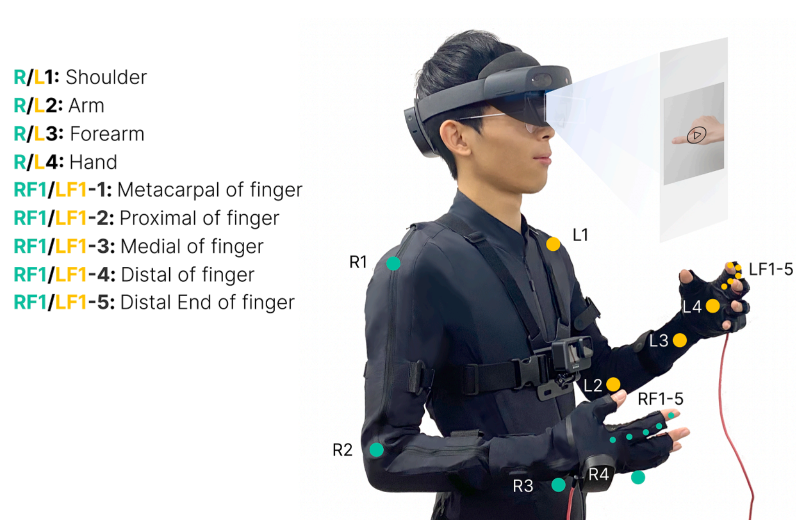


Fig. 5. Schematic diagram of comprehensive evaluation experiment and the joints for motion capture. There are five sensors on each finger, and only the sensors on the index finger (RF1and LF1) are marked in this figure for examples.

characteristics rather than ease of learning. In our study, following the methods proposed by Nielsen et al. (2004), participants were required to recall and perform gestures when the name of the function was displayed on the HMD. They were asked to recall as many gestures as possible under the same functionality. Memorability was quantified by scoring the order of vocabulary recall: the first hand-based vocabulary recalled earned three points, followed by the second and third recall earning two points and one point, respectively. Non-recalled or forgotten vocabulary received a score of zero.

4.3.2. Physical factor

Gesture language efficiency served as a measure of the physical factors in this study, indicating the complexity and efficiency of physical movements.

4.3.2.1. Efficiency. Efficiency is referred to as the difficulty, human performance, interaction cost, speed, duration, effort, ease of performance, or ease of operation of a gesture action. Therefore, increasing the input throughput and decreasing the physical effort required are necessary (Xia et al., 2022). In this study, gesture efficiency was assessed using camera-recorded task completion times (Rubine, 1992). Specifically, participants were asked to place their hands on their knees during rest periods and then lift them to perform a valid trial as naturally as possible during the task execution phase. Each hand-based vocabulary task was performed ten times. The first and last gesture movements were not included in the analysis because of their potential for behavioral adjustment when initiating or concluding a task (Gustafsson et al., 2018; Huang et al., 2022). The task completion time was defined as the time interval between moving from the knee to the complete execution of the gesture action in air. The average duration of eight middle attempts was used to evaluate efficiency.

4.3.3. System-based factor

The system-based factor relates to the performance of features and techniques contained within a system or device. These technologies are used to recognize the user's input and provide feedback to the user during and after the gesture has been performed (Xia et al., 2022). The system-based factor determines whether gestures can be effectively applied to the software or systems. The recognition rate served as a system-based factor in this study.

4.3.3.1. Algorithm recognition. The recognition of a gesture is the process of tracking it from its initial representation through its later conversion into a semantically meaningful command (Rautaray and Agrawal, 2015). Poorly or incorrectly recognized hand-based vocabulary frustrates users who have to repeat or exaggerate actions for correct recognition. Gesture recognition can be evaluated using gesture databases or real-time user performances (Vuletic et al., 2019).

A recognition classification algorithm based on deep neural methods was used in this study. First, in the data preprocessing stage, the relative coordinates were transformed into absolute spatial coordinates using recursive calculations. Second, a joint data screening was conducted. 74 joint motions were captured and recorded by the smartsuit, while only the joints related to gesture execution were screened and preserved. The joints reserved and used for the recognition calculation were the *Shoulder, Arm, Forearm, Hand,* and *the Metacarpal, Proximal, Medial, Distal,* and *Distal End of all fingers* (Fig. 5). Third, as the completion times for each gesture varied, the frame numbers for all the gesture data were standardized. Each gesture data was set to 400 frames, with 0 being used for pad data that were less than 400 frames, and data beyond 400 frames were deleted. Fourth, each data frame consisted of the relative coordinates of 56 joints from the previous frame, resulting in three coordinates per joint. In total, 400 data frames were combined to form a single data sample. For each of the 65 hand-based vocabularies, 190 effective data samples were collected, resulting in a total of 12350 data samples in the entire dataset.

We employed the ResNet-152 architecture to train the hand-based vocabulary classification model (He et al., 2016). It contains multiple convolutional layers to extract high-level features from data. To enhance the robustness of the model to noise, Gaussian noise with a mean and amplitude of 20 was added to all the joint coordinates in the dataset. We randomly chose 80% of the gestures in the dataset as the training set and used the other 20% of the data as the test set. Recognition precision is defined as the proportion of correct recognition of functionality through gesture motions.

4.4. Experimental protocol

This was a laboratory study with a within-subject design in which all participants performed 65 hand-based vocabulary tasks for 24 given functionalities. Fig. 6 shows the procedure for the evaluation experiment.

After the informed consent was confirmed, participants wore the action capture device and smart suit, and were instructed to sit with the HMD equipment. First, a discoverability test was conducted. A functionality introduction, including the name and diagram, was initially randomly presented to participants with an oral explanation to ensure that participants were already familiar with all the functions (see details in Appendix 1). Then, participants were required to guess the functionality name through watching the gesture action video clip. In the next stage, participants were asked to perform gestures after imitating and following an action video clip. After completing 10 times for each hand-based vocabulary item, the subjective learnability score was marked using the Likert quintile scale. The motion data and completion time of the executed gestures were recorded by the devices during the entire process. The experimental data collection consisted of 24 participants $\times$ 65 hand-based vocabularies $\times$ 10 repetitions $=$ 15600 gesture

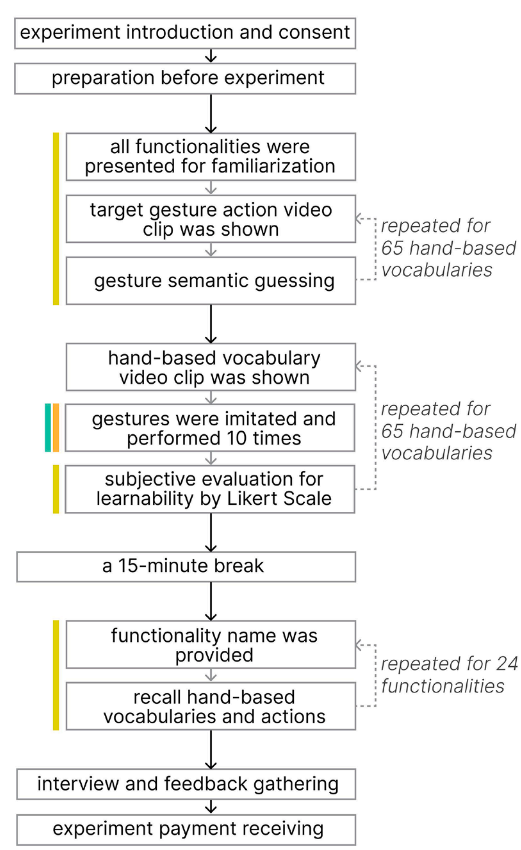
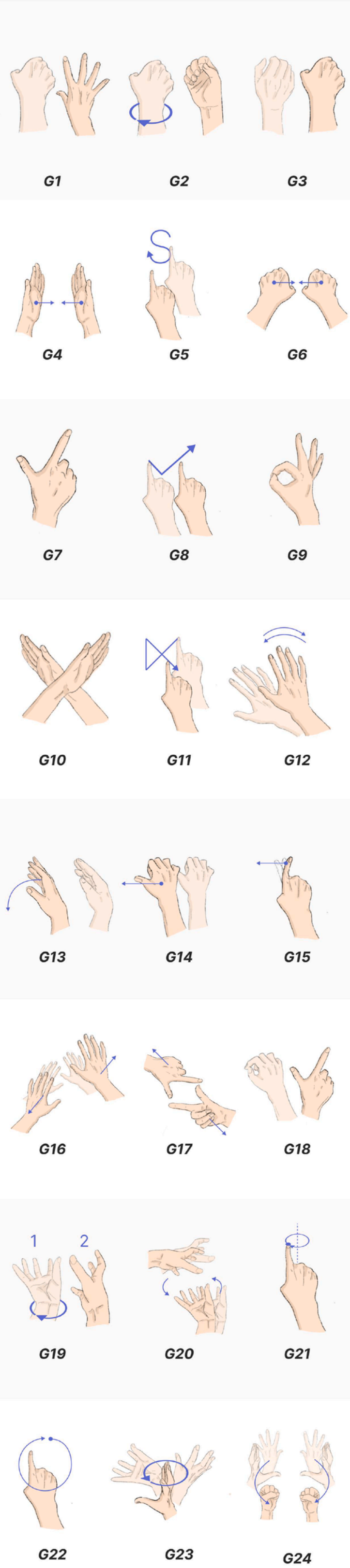


Fig. 6. Procedure of the evaluation experiment.



Functionality	Gestures	Discoverability	Learnability	Memorability	Efficiency	Comments
F1 Open Menu	*G1*, *G2*, *G3*	$F(2, 46) = 18.796$ $P < 0.001$ $\eta^2 = 0.450$ Post hoc: G1 > G2, G3	$F(2, 46) = 66.799$ $P < 0.001$ $\eta^2 = 0.744$ Post hoc: G1 > G2, G3	$F(2, 46) = 15.114$ $P < 0.001$ $\eta^2 = 0.397$ Post hoc: G1, G2 > G3	$F(2, 46) = 16.261$ $P < 0.001$ $\eta^2 = 0.079$ Post hoc: G1, G3 < G2	*1. G1 presents significantly higher discoverability and learnability scores than others.* *2. G1 and G2 present higher memorability score than G3.* *3. G1 and G3 present shorter task completion time than G2.*
F3 Save	*G4*, *G5*, *G6*	$F(2, 46) = 9.123$ $P = 0.002$ $\eta^2 = 0.284$ Post hoc: G4, G5 > G6	$F(2, 46) = 58.646$ $P < 0.001$ $\eta^2 = 0.718$ Post hoc: G5 > G4 > G6	$F(2, 46) = 51.856$ $P < 0.001$ $\eta^2 = 0.693$ Post hoc: G5 > G4 > G6	$F(2, 46) = 42.508$ $P < 0.001$ $\eta^2 = 0.184$ Post hoc: G4, G5 < G6	*1. G4 and G5 present higher discoverability score and shorter task completion time than G6.* *2. G5 presents the highest learnability and memorability scores.*
F4 Confirm	*G7*, *G8*, *G9*	$F(2, 46) = 13.800$ $P < 0.001$ $\eta^2 = 0.375$ Post hoc: G8, G9 > G7	$F(2, 46) = 1.473$ $P = 0.241$ $\eta^2 = 0.060$	$F(2, 46) = 46.924$ $P < 0.001$ $\eta^2 = 0.671$ Post hoc: G7, G8 > G9	$F(2, 46) = 9.129$ $P < 0.001$ $\eta^2 = 0.046$ Post hoc: G7, G8 < G9	*1. G8 and G9 present higher discoverability score than G7.* *2. G7 and G8 present higher memorability score and shorter task completion time than G9.*
F5 Cancel	*G10*, *G11*, *G12*	$F(2, 46) = 26.545$ $P < 0.001$ $\eta^2 = 0.536$ Post hoc: G10, G11 > G12	$F(2, 46) = 51.463$ $P < 0.001$ $\eta^2 = 0.691$ Post hoc: G10, G11 > G12	$F(2, 46) = 19.359$ $P < 0.001$ $\eta^2 = 0.457$ Post hoc: G10, G11 > G12	$F(2, 46) = 69.125$ $P < 0.001$ $\eta^2 = 0.268$ Post hoc: G10 < G11, G12	*1. G10 and G11 present higher cognitive scores than G12.* *2. G10 presents shorter task completion time than G11 and G12.*
F6 Undo	*G13*, *G14*, *G15*	$F(2, 46) = 2.106$ $P = 0.138$ $\eta^2 = 0.084$	$F(2, 46) = 44.698$ $P < 0.001$ $\eta^2 = 0.660$ Post hoc: G13, G15 > G14	$F(2, 46) = 2.670$ $P < 0.001$ $\eta^2 = 0.098$ Post hoc: G13, G15 > G14	$F(2, 46) = 55.678$ $P < 0.001$ $\eta^2 = 0.228$ Post hoc: G13, G15 < G14	*1. G13 and G15 present higher learnability and memorability scores and shorter task completion time than G14.*
F8 Zoom in	*G16*, *G17*, *G18*	$F(2, 46) = 7.881$ $P = 0.001$ $\eta^2 = 0.255$ Post hoc: G16, G18 > G17	$F(2, 46) = 8.011$ $P = 0.002$ $\eta^2 = 0.258$ Post hoc: G16, G17 > G18	$F(2, 46) = 21.030$ $P < 0.001$ $\eta^2 = 0.478$ Post hoc: G17 > G16 > G18	$F(2, 46) = 29.373$ $P < 0.001$ $\eta^2 = 0.135$ Post hoc: G16 < G18 < G17	*1. G16 and G18 present higher discoverability score than G17.* *2. G16 and G17 present higher learnability score than G18.* *3. G17 presents the highest memorability score.* *4. G16 presents the shortest task completion time.*
F10 Rotate View	*G19*, *G20*, *G21*	$F(2, 46) = 22.474$ $P < 0.001$ $\eta^2 = 0.494$ Post hoc: G19, G20 > G21	$F(2, 46) = 100.813$ $P < 0.001$ $\eta^2 = 0.814$ Post hoc: G19, G20 > G21	$F(2, 46) = 9.416$ $P = 0.004$ $\eta^2 = 0.290$ Post hoc: G19, G20 > G21	$F(2, 46) = 31.923$ $P < 0.001$ $\eta^2 = 0.144$ Post hoc: G21 < G19, G20	*1. G19 and G20 present higher cognitive scores than G21.* *2. G21 presents shorter task completion time than others.*
F11 Select all	*G22*, *G23*, *G24*	$F(2, 46) = 32.134$ $P < 0.001$ $\eta^2 = 0.583$ Post hoc: G24 > G22 > G23	$F(2, 46) = 96.288$ $P < 0.001$ $\eta^2 = 0.807$ Post hoc: G24 > G22 > G23	$F(2, 46) = 59.616$ $P < 0.001$ $\eta^2 = 0.722$ Post hoc: G24 > G22 > G23	$F(2, 46) = 28.471$ $P < 0.001$ $\eta^2 = 0.131$ Post hoc: G24 < G22, G23	*1. G24 presents the highest scores in all cognitive factors and the shortest task completion time.*

Fig. 7. Statistical results and corresponding comments on the cognitive and physical factors.

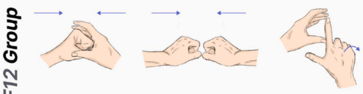
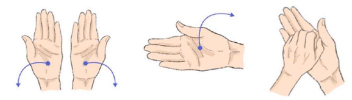
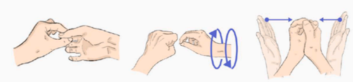
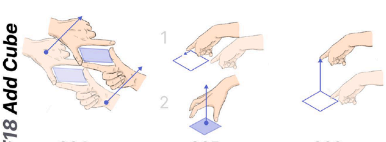
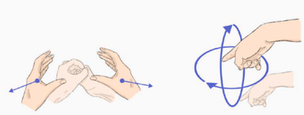
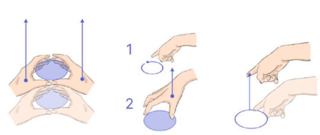
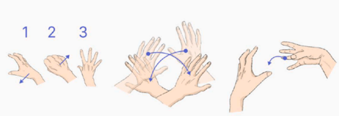
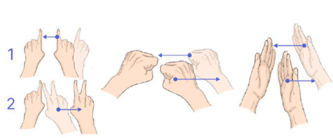
Functionality & Gestures	Discoverability	Learnability	Memorability	Efficiency	Comments
F12 Group  G25 G26 G27	$F(2, 46) = 10.195$ P < **0.001** $\eta^2 = 0.307$ Post hoc: G25, G26 > G27	$F(2, 46) = 13.632$ P < **0.001** $\eta^2 = 0.372$ Post hoc: G25, G26 > G27	$F(2, 46) = 2.642$ $P = 0.082$ $\eta^2 = 0.103$	$F(2, 46) = 4.715$ P = **0.01** $\eta^2 = 0.024$ Post hoc: G25, G26 < G27	*1. G25 and G26 present higher discoverability and learnability scores, and also shorter task completion time than G27.*
F14 Hide  G28 G29 G30	$F(2, 46) = 36.486$ P < **0.001** $\eta^2 = 0.613$ Post hoc: G28, G29 > G30	$F(2, 46) = 283.667$ P < **0.001** $\eta^2 = 0.925$ Post hoc: G29 > G28 > G30	$F(2, 46) = 84.356$ P < **0.001** $\eta^2 = 0.786$ Post hoc: G28, G29 >G30	$F(2, 46) = 4.074$ P = **0.02** $\eta^2 = 0.021$ Post hoc: G29 < G30	*1. G28 and G29 present higher discoverability and memorability scores than 30.* *2. G29 presents the highest learnability score and the shortest task completion time.*
F16 Lock  G31 G32 G33	$F(2, 46) = 13.912$ P < **0.001** $\eta^2 = 0.377$ Post hoc: G31 > G32 > G33	$F(2, 46) = 42.492$ P < **0.001** $\eta^2 = 0.649$ Post hoc: G31 > G32, G33	$F(2, 46) = 8.320$ P < **0.001** $\eta^2 = 0.279$ Post hoc: G31 > G32, G33	$F(2, 46) = 52.033$ P < **0.001** $\eta^2 = 0.216$ Post hoc: G31 < G33 < G32	*1. G31 presents the highest scores in all cognitive factors and the shortest task completion time.*
F18 Add Cube  G34 G35 G36	$F(2, 46) = 43.844$ P < **0.001** $\eta^2 = 0.656$ Post hoc: G36 > G35 > G34	$F(2, 46) = 82.702$ P < **0.001** $\eta^2 = 0.782$ Post hoc: G35, G36 > G34	$F(2, 46) = 16.045$ P < **0.001** $\eta^2 = 0.411$ Post hoc: G35, G36 > G34	$F(2, 46) = 477.527$ P < **0.001** $\eta^2 = 0.716$ Post hoc: G34 < G36 < G35	*1. G35 and G36 present higher learnability and memorability scores than G34.* *2. G36 presents the highest discoverability score.* *3. G34 presents the shortest task completion time.*
F19 Add Sphere  G37 G38	$F(1, 23) = 9.471$ P = **0.005** $\eta^2 = 0.282$ Post hoc: G38 > G37	$F(1, 23) = 0.107$ $P = 0.747$ $\eta^2 = 0.005$	$F(1, 23) = 0.523$ $P = 0.477$ $\eta^2 = 0.022$	$F(1, 23) = 14.000$ $P = 0.671$ $\eta^2 = 0.013$	*1. G38 presents higher discoverability score than G37.*
F20 Add Cylinder  G39 G40 G41	$F(2, 46) = 1.849$ $P = 0.175$ $\eta^2 = 0.074$	$F(2, 46) = 2.556$ $P = 0.098$ $\eta^2 = 0.100$	$F(2, 46) = 0.989$ $P = 0.379$ $\eta^2 = 0.041$	$F(2, 46) = 114.900$ P < **0.001** $\eta^2 = 0.378$ Post hoc: G39 < G41 < G40	*1. G39 presents the shortest task completion time.*
F21 Delete  G42 G43 G44	$F(2, 46) = 78.712$ P < **0.001** $\eta^2 = 0.774$ Post hoc: G42 > G43 > G44	$F(2, 46) = 312.719$ P < **0.001** $\eta^2 = 0.931$ Post hoc: G42 > G43 > G44	$F(2, 46) = 123.141$ P < **0.001** $\eta^2 = 0.843$ Post hoc: G42 > G43, G44	$F(2, 46) = 127.764$ P < **0.001** $\eta^2 = 0.403$ Post hoc: G42, G44 < G43	*1. G42 presents the highest scores in all cognitive factors.* *2. G42 and G44 present shorter task completion time than G43.*
F22 Copy  G45 G46 G47	$F(2, 46) = 29.982$ P < **0.001** $\eta^2 = 0.566$ Post hoc: G45, G46 > G47	$F(2, 46) = 47.520$ P < **0.001** $\eta^2 = 0.674$ Post hoc: G45 > G46 > G47	$F(2, 46) = 50.219$ P < **0.001** $\eta^2 = 0.686$ Post hoc: G45 > G46 > G47	$F(2, 46) = 4.079$ P < **0.001** $\eta^2 = 0.280$ Post hoc: G45 < G46, G47	*1. G45 presents the highest learnability and memorability scores, and also the shortest task completion time.* *2. G45 and G46 present higher discoverability score than G47.*

Fig. 7. (*continued*).

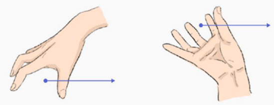
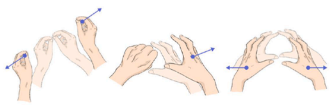
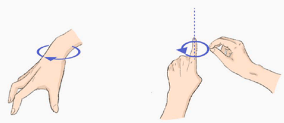
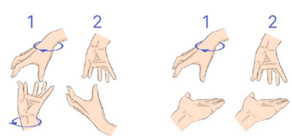
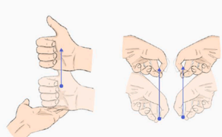
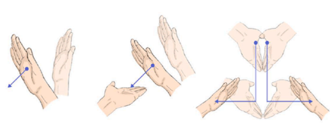
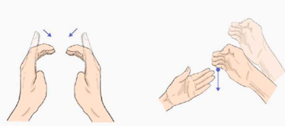
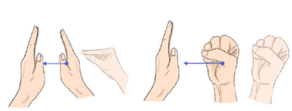
Functionality	Gestures	Discoverability	Learnability	Memorability	Efficiency	Comments
F23 Move	 G48 G49	F (1, 23) = 5.242 P = **0.032** η² = 0.186 Post hoc: G48 > G49	F (1, 23) = 0.923 P = 0.347 η² = 0.039	F (1, 23) = 40.114 P < **0.001** η² = 0.636 Post hoc: G48 > G49	F (1, 23) = 0.020 P = 0.889 η² = 0.000	1. G48 presents higher discoverability and memorability scores.
F24 Scale up	 G50 G51 G52	F (2, 46) = 3.719 P = **0.032** η² = 0.139 Post hoc: G50, G52 > G51	F (2, 46) = 13.800 P < **0.001** η² = 0.375 Post hoc: G50 > G51, G52	F (2, 46) = 14.109 P < **0.001** η² = 0.380 Post hoc: G50 > G51, G52	F (2, 46) = 31.224 P < **0.001** η² = 0.142 Post hoc: G50 < G51, G52	1. G50 presents the highest scores in learnability and memorability, and the shortest task completion time. 2. G50 and G52 present higher discoverability score than G51.
F26 Rotate	 G53 G54	F (1, 23) = 37.510 P < **0.001** η² = 0.620 Post hoc: G53 > G54	F (1, 23) = 66.860 P < **0.001** η² = 0.744 Post hoc: G53 > G54	F (1, 23) = 16.403 P < **0.001** η² = 0.416 Post hoc: G53 > G54	F (1, 23) = 51.284 P < **0.001** η² = 0.213 Post hoc: G53 < G54	1. G53 presents the highest scores in all cognitive factors and the shortest task completion time.
F27 Twist	 G55 G56	F (1, 23) = 25.274 P < **0.001** η² = 0.524 Post hoc: G55 > G56	F (1, 23) = 23.000 P < **0.001** η² = 0.500 Post hoc: G55 > G56	F (1, 23) = 54.746 P < **0.001** η² = 0.704 Post hoc: G55 > G56	F (1, 23) = 31.427 P < **0.001** η² = 0.143 Post hoc: G55 < G56	1. G55 presents the highest scores in all cognitive factors and the shortest task completion time.
F28 Extrude Region	 G57 G58	F (1, 23) = 8.364 P = **0.008** η² = 0.267 Post hoc: G58 > G57	F (1, 23) = 0.215 P = 0.647 η² = 0.009	F (1, 23) = 2.122 P = 0.159 η² = 0.084	F (1, 23) = 120.816 P < **0.001** η² = 0.390 Post hoc: G58 < G57	1. G58 presents higher discoverability score and shorter task completion time than G57.
F29 Cut	 G59 G60 G61	F (2, 46) = 13.477 P < **0.001** η² = 0.369 Post hoc: G59, G60 > G61	F (2, 46) = 79.222 P < **0.001** η² = 0.775 Post hoc: G59, G60 > G61	F (2, 46) = 86.448 P < **0.001** η² = 0.790 Post hoc: G59 > G60 > G61	F (2, 46) = 67.661 P < **0.001** η² = 0.264 Post hoc: G59 < G60 < G61	1. G59 presents the highest memorability score and the shortest task completion time. 2. G59 and G60 present higher discoverability and learnability score than G61.
F30 Fillet Edge	 G62 G63	F (1, 23) = 22.540 P < **0.001** η² = 0.495 Post hoc: G62 > G63	F (1, 23) = 71.875 P < **0.001** η² = 0.758 Post hoc: G62 > G63	F (1, 23) = 14.083 P < **0.001** η² = 0.503 Post hoc: G62 > G63	F (1, 23) = 20.089 P < **0.001** η² = 0.096 Post hoc: G62 < G63	1. G62 presents the highest scores in all cognitive factors and the shortest task completion time.
F31 Loft	 G64 G65	F (1, 23) = 9.000 P = **0.006** η² = 0.281 Post hoc: G64 > G65	F (1, 23) =0.193 P = 0.664 η² = 0.008	F (1, 23) = 6.419 P = **0.019** η² = 0.218 Post hoc: G64 > G65	F (1, 23) = 1.538 P = 0.216 η² = 0.008	1. G64 presents higher discoverability and memorability scores than G65.

Fig. 7. (*continued*).

trials. After a 15-min break, the memorability test stage began. Participants were requested to recall and perform hand-based vocabulary and actions based on the provided functionality names.

Each stage included a training session to familiarize participants with the task procedure. Hand-based vocabularies or functionalities were randomly presented to participants at all experimental stages. Before the end of the experiment, each participant was invited to participate in a short interview to provide feedback on their feelings and suggestions related to the hand-based language they experienced. The entire experiment took approximately 80 min to complete.

4.5. Statistical analysis

Statistical analyses were performed using SPSS (V22.0, International Business Machines Corporation, Armonk, NY, USA). A Kolmogorov–Smirnov normality test was run at a significance level of 0.05 before the statistical analysis. The repeated-measures ANOVA was employed if the data followed a normal distribution, whereas Friedman tests were used if the data did not follow a normal distribution. Post-hoc multiple analyses were also performed, and Bonferroni correction was used to adjust for the *p*-value to identify significant differences among different hand-based vocabularies. The level of statistical significance for all of these analyzes was set at 0.05. Effect sizes were evaluated using partial ETA Squared. The results are reported as mean value (SD) and illustrated in Appendix 2.

5. Results of comprehensive evaluation

5.1. Cognitive factor and physical factor

Among the cognitive factors, 22, 18, and 21 functionalities had a significant effect on discoverability, learnability, and memorability, respectively. For the physical factor, 21 functionalities had a significant effect on task completion time. Fig. 7 provides a comprehensive presentation of these results with an accompanying commentary. For example, the first row compares three gestures (*G1, G2*, and *G3*) for the functionality *Open Menu* (*F1*). Significant effects were reported in discoverability [F (2, 46) = 18.796, $p < 0.001$, $\eta^2 = 0.450$], learnability [F (2, 46) = 66.799, $p < 0.001$, $\eta^2 = 0.744$], memorability [F (2, 46) = 15.114, $p < 0.001$, $\eta^2 = 0.397$], and efficiency [F (2, 46) = 16.261, $p < 0.001$, $\eta^2 = 0.079$]. The post-hoc test results indicated that the discoverability and learnability scores of *G1* were significantly higher than the discoverability and learnability scores of *G2* and *G3*, whereas the memorability scores of *G1* and *G2* were significantly higher than the memorability score of *G3*. In addition, the physical results showed that the task completion time of *G1* and *G3* were significantly shorter than the task completion time of *G2*. To interpret the statistical results, we added corresponding comments in the rightmost column "*(1) G1 presents significantly higher discoverability and learnability scores than others. (2) G1 and G2 present higher memorability scores than G3. (3) G1 and G3 present shorter task completion time than G2*".

5.2. System-based factor

Gesture motion data were recorded during the experiment, and the recognition rate was evaluated using the trained ResNet-152

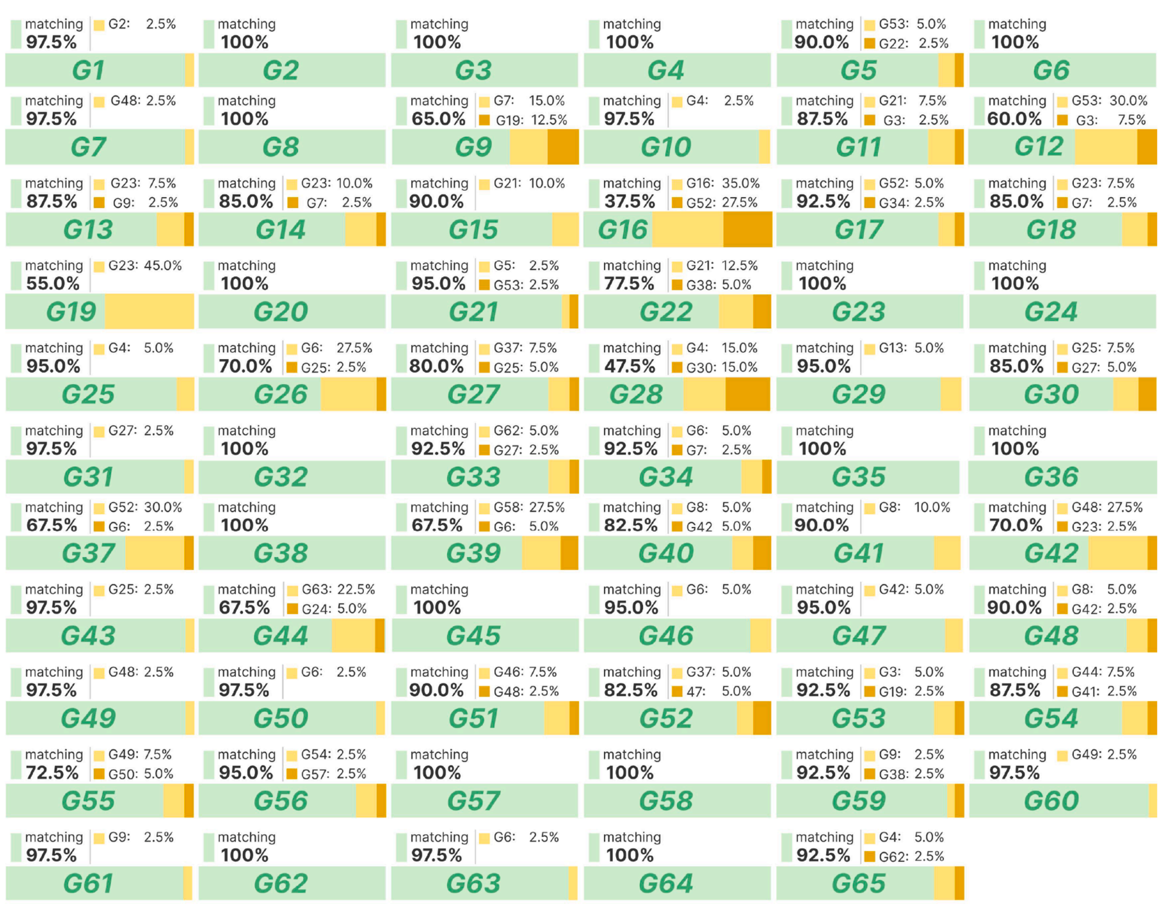


Fig. 8. Results of gesture recognition evaluation. The green segment denotes the probability of correct gesture recognition, termed "matching". The yellow and brown segment represent the likehood of misclassification, where one gesture is incorrectly identified as another.

classification model (He et al., 2016). On average, the model achieved a recognition accuracy of 88.73% among 65 distinct gestures. Fig. 8 shows the detailed results for each gesture.

6. Discussion

6.1. Summarized guidelines for hand-based interaction application in 3D conceptual design

This study collected and summarized participants' feedback toward the gesture input application in a 3D conceptual design through interviews after the gesture language evaluation experiment.

First, as an auxiliary interactive input method, hand-based input can make the conceptual design more intuitive and efficient. However, it may not be necessary for all functions during design. P1, P7, and P16 pointed that "*I think it is more efficient and convenient to operate with gestures for some complex functions, such as the F27 (Twist) and F30 (Extrude Edge). These operations with mouse and keyboard require multiple clicks and inputs, but they can be performed quickly by gesture input.*" P7, P10, and P20 also reported that "*For some commands of creating objects, such as the F18 (Add cube) and F19 (Add Sphere), it feels very intuitive to operate with gestures, just like objects fabricated by my hands.*" In addition,

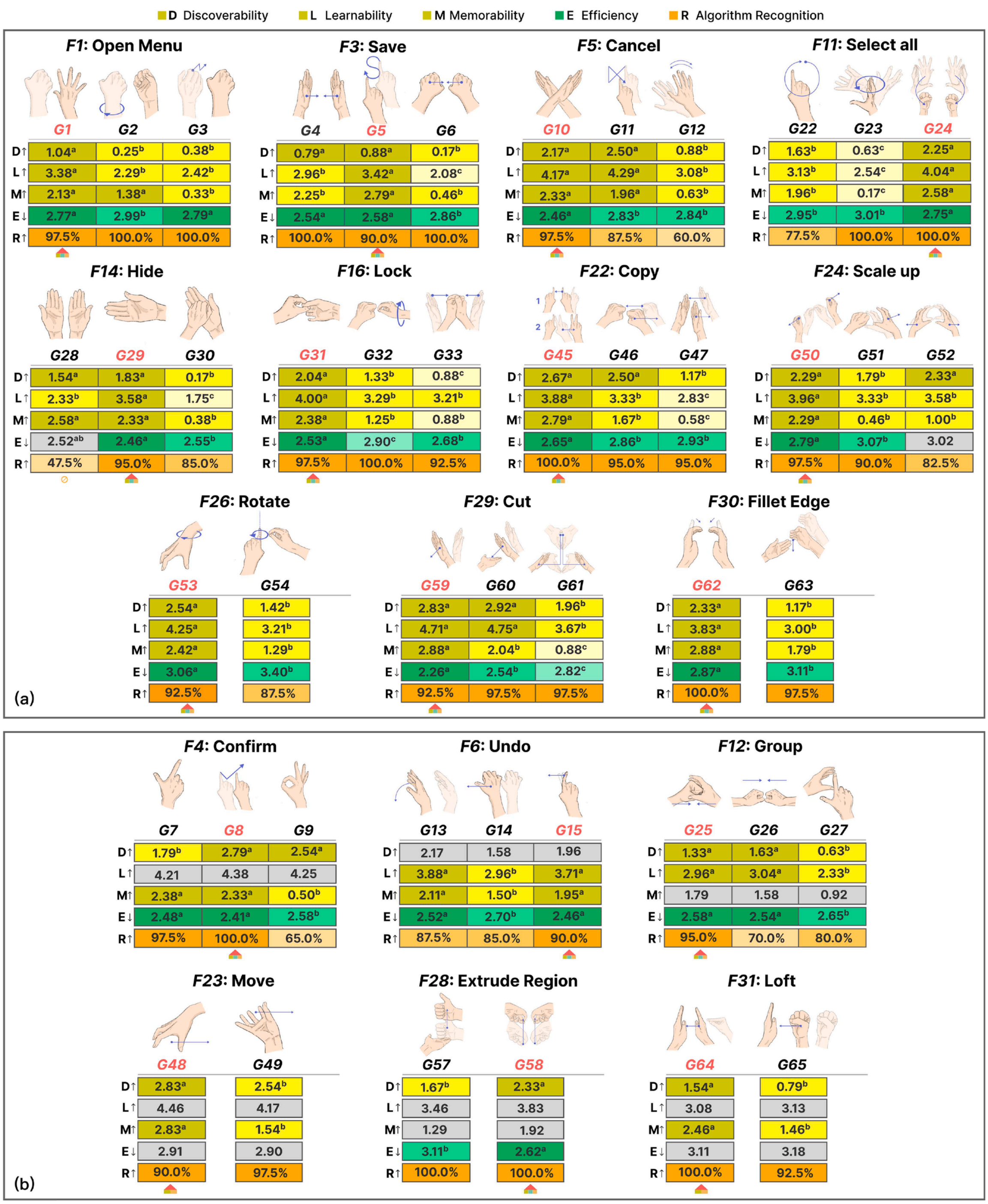


Fig. 9. Results of 17 functionalities having consistent results among factors. "↑" means a higher value is better, and vice versa. Different lowercase letters indicated significant differences among hand-based vocabularies. "▲" means the recommended gesture based on evaluation results of cognitive factor (■), physical factor (■), and system-based factors (■) comprehensively.

more than half of participants indicated that "*Gesture input can be applied to quickly perform some high-frequency operations in the process of creation, like Save (F3), Confirm (F4) and Undo (F6).*" However, not all functionalities require gesture inputs. Hand-based language is posited as an optional "shortcut key" style input form to be used as per individual users' preferences and habits. For example, the appearance of the mouse greatly improved the efficiency of HCI, however, it did not completely replace the keyboard. With the continuous development of technology, the gesture input can also be combined with other input forms, such as eye movement interaction and virtual keyboard input, to achieve embodied interactions in the MR environment (Gibbs, 2006).

Second, maintaining semantic unity in a hand-based language system can alleviate the cognitive load. P15 and P24 stated that "*Placing a square with two index fingers and thumbs was not only used in F8 (Zoom in) as G17, but also used in F18 (Add Cube) as G34. However, the same 'square' means a 'visual range' in G17 while it means an 'intersecting surface' in G34. If these two gestures appear in the same set of gesture language, it will make me more confused about the meaning of the gesture action.*" In contrast, P9 and P13 reported that "*As the thumb is used to indicate the direction (i.e., the thumb pointing to the left indicates the return) in the G14 under F6 (Undo), I can immediately realize that the thumb indicates the direction of extrusion when I meet the G58 under F28 (Extrude Region).*" Therefore, when designing a set of hand-based languages, it is necessary to consider whether the referential semantics of the same action are consistent in various hand-based vocabularies. For example, the gesture action style of creating a cylinder should be consistent with the gesture of creating a cube in the same set of gesture languages to reduce the cognitive load of users.

Third, hand-based languages should eliminate ambiguous referential relations because gestures have inherently vague and abstract semantics. Specifically, P6 and P19 elaborated that "*Pinching with the forefinger and thumb can mean 'OK', so that it can be used to refer to F4 (Confirm), and it can also refer to F3 (Save) in some cases in my mind.*" Similarly, P22 reported that "*In the discoverability test, I immediately realized that 'OK' was related to F4 (Confirm). However, in the later memorability test, I forgot whether the 'OK' gesture specifically refers to F3 (Save) or F4 (Confirm), but I was deeply impressed by the gesture of writing 'S' (G5) due to its distinction.*" Therefore, it is beneficial to scrutinize whether abstract semantic gestures can refer to distinct commands, while emphasizing the uniqueness of gesture actions, as pointed out by P19 that "*I would rather spend some costs to learn new gestures than use those gestures that seem easy to understand at first, but will always be confused later.*"

6.2. Gesture selection combining cognitive, physical, and system-based factors.

In this section, we present a comprehensive discussion of the candidate gestures associated with each functionality, drawing insights from cognitive, physical, and system-based findings. The goal in this section is to propose a recommended hand-based vocabulary for each functionality.

6.2.1. Functionalities presenting consistent results among factors

Eleven functionalities (*F1* (*Open Menu*), *F3* (*Save*), *F5* (*Cancel*), *F11* (*Select all*), *F14* (*Hide*), *F16* (*Lock*), *F22* (*Copy*), *F24* (*Scale up*), *F26* (*Rotate*), *F29* (*Cut*), and *F30* (*Fillet Edge*)) presented a candidate gesture which had consistent results for all evaluation factors. Hence, we identified 11 hand-based vocabularies in the corresponding functionalities (Fig. 9a). In addition, six functionalities (*F4* (*Confirm*), *F6* (*Undo*), *F12* (*Group*), *F23* (*Move*), *F28* (*Extrude Region*), and *F31* (*Loft*)) presented a candidate gesture which demonstrated partial consistency across the cognitive, physical, and system-based evaluations. Therefore, we endorsed hand-based vocabularies for these six functionalities because of their empirical support for lower cognitive load, higher efficiency, and superior recognition friendliness (Fig. 9b).

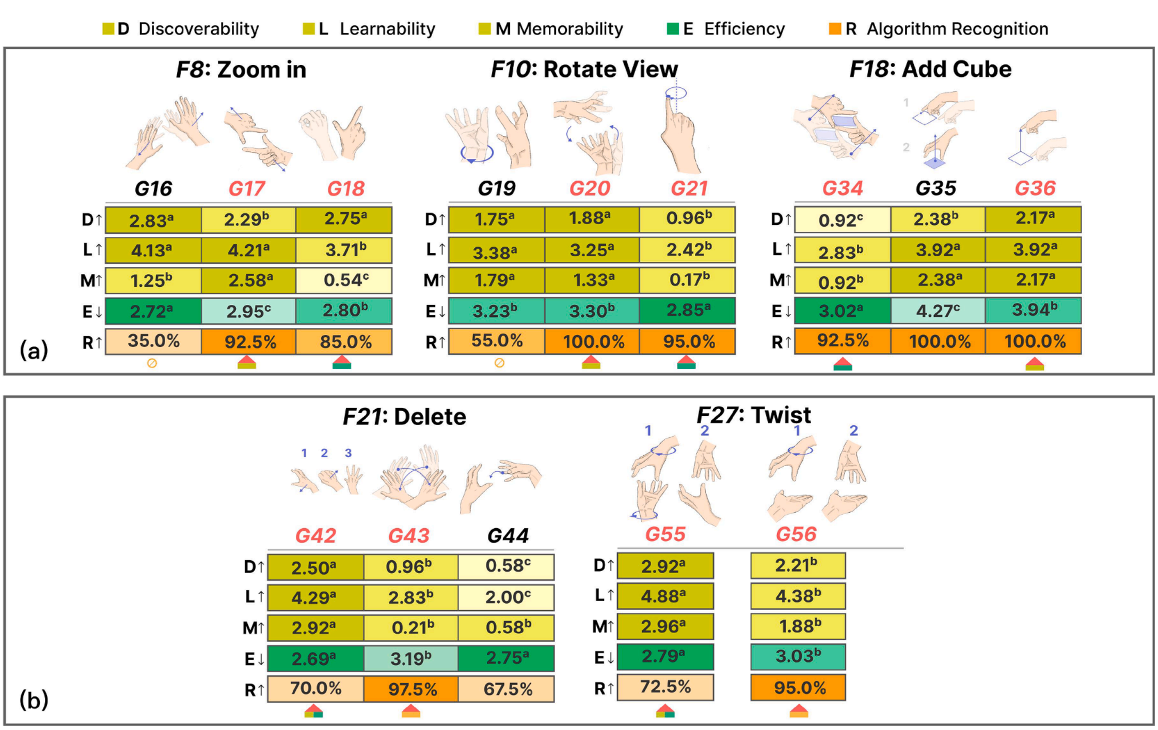


Fig. 10. Results of five functionalities having inconsistent results among factors. "↑" means a higher value is better, and vice versa. Different lowercase letters indicated significant differences among hand-based vocabularies. " ", " ", " " means the recommended gesture based on evaluation results of cognitive factor ( ), physical factor ( ), and system-based factors ( ) respectively.

6.2.2. Functionalities presenting inconsistent results among factors

Five functionalities presented different optimal gestures in different evaluation factors (Fig. 10). For example, discrepancies between cognitive and physical factors necessitate separate recommendations for cognition- and efficiency-oriented gestures that cater to distinct application requirements. Specifically, for some concept design scenarios that focused on 3D styling modeling, the targeted users were mainly design experts, and the usage frequency was high. Efficiency is a crucial factor because it increases the modeling throughput of expert designers and decreases the amount of physical effort required (Vuletic et al., 2019). Efficiency-oriented gestures are preferentially recommended for this type of application because users are fully familiar with these gesture actions through frequent use in daily work, even if some actions are not sufficiently intuitive. Conversely, for scenarios focused on presenting or discussing ideation, sketching, and rendering in the pre-design stage, which are interdisciplinary (Wang et al., 2002), cognition-oriented gestures play a pivotal role. They offer intuitive and comprehensible actions that enable users at all levels to swiftly learn and effectively employ the associated tools (Blackler et al., 2002). Therefore, based on our results, in *F8* (*Zoom in*), *F10* (*Rotate View*), and *F18* (*Add Cube*), three cognition-oriented gestures and three efficiency-oriented gestures were recommended (Fig. 10a). Additionally, vocabularies reflecting disparities between system-based factors and other factors have prompted the recommendation of both user- and computer-friendly gestures for different functions and scenarios. A previous study demonstrated that erroneous or unrecognizable gestures can impair user experience or deter gesture use (Benko and Wigdor, 2010). Therefore, in *F21* (*Delete*) and *F27* (*Twist*), two user-friendly and two computer-friendly gestures were recommended, as shown in Fig. 10b. Future efforts will further examine these inconsistent gestures through a more integral evaluation or enhancement of the technical recognition ability to facilitate a holistic understanding.

6.2.3. Functionalities presenting no significant result for more than half of the factors

For *F19* (*Add Sphere*) and *F20* (*Add Cylinder*), only one factor has significant results (Fig. 11), limiting our ability to draw conclusive findings from the gesture evaluation. Therefore, this study does not recommend a hand-based vocabulary for these two functionalities.

6.3. Potential application space and application barrier

6.3.1. Supplementing prior work.

While some previous studies have explored different gesture design methodologies and techniques, there is a lack of research that understands, identifies, or analyzes various factors (Xia et al., 2022). The contribution of this study is that it conducts a comprehensive experiment combining the indices of cognitive, physical, and system-based factors to determine a set of theoretically optimal user-defined gesture languages. Additionally, rather than adopting a universal mid-air gesture set (Huang et al., 2021), this study focused on creating an exactly matched hand-based language tailored to the 3D conceptual design scenario within a specific MR environment, as shown in Fig. 12, facilitating user comprehension and utilization of the gesture system (Nielsen et al., 2004).

6.3.2. Application possibilities for real-life use

The application possibilities of the results of this study can be summarized in two ways. On the one hand, a novel 3D conceptual design software can be designed and developed based on the proposed hand-based language and application guidelines. Previous studies indicated that hand-based interactions facilitate designers' conceptualization of information and complete cognitive processes (Hostetter, 2004). Correspondingly, participants' feedback in our study confirmed the increased intuitiveness and naturalness of creating 3D objects through gesture interaction in MR. Therefore, our hand-based language serves as a guiding framework for developers interested in devising novel conceptual design tools that embody immersive interactions. Moreover, our gesture application guidelines can inspire system and software developers to explore innovative hand-based modeling tools.

On the other hand, our hand-based language enhances auxiliary MR interactions in real-life applications. It can function as an additional input interaction, supplementing the mouse, keyboard, or operating handle of existing commercial 3D modeling software. With the development of MR technology, many traditional CAD software programs now support 3D creation and editing in MR environments using HMDs (e.g., Create VR for MAYA & Adobe Medium). Considering this, a selection or even an entire hand-based vocabulary can be applied to existing commercial software as an attached interaction mode, thereby improving the efficiency and experience of conceptual design work.

6.3.3. Application barriers in real-life use

In this study, a set of theoretically optimal hand-based languages was proposed based on a comprehensive experiment. However, a gap between theoretical guidelines and practical applications remains. We considered the application barriers in real-life use.

First, collecting gesture data is challenging, which predominantly involves visual or wearable sensing methodologies. Visual sensing employs plane or depth cameras to capture gesture imagery, which is expensive and has line-of-sight occlusion issues (Liu et al., 2020; Menolotto et al., 2020). Wearable sensing utilizes sensors on the user's hands to directly capture gesture data, but it encounters obstacles, such as sensor-induced data drift and susceptibility to electromagnetic interference (Zhou et al., 2022). Second, processing the time-series gesture data is challenging. Owing to their temporal nature, gesture data represent a sequential stream of continuous data (Lee & Bae, 2022). In the technical realm, the task of gesture localization, entailing the segmentation of a specific gesture vocabulary from a continuous stream,

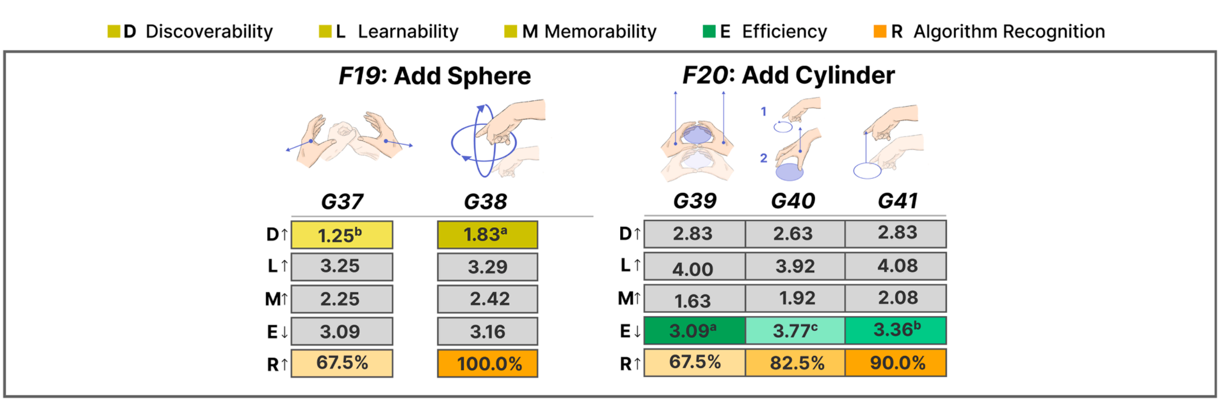


Fig. 11. Results for two functionalities that showed no significant outcome in over half of the evaluated factors.

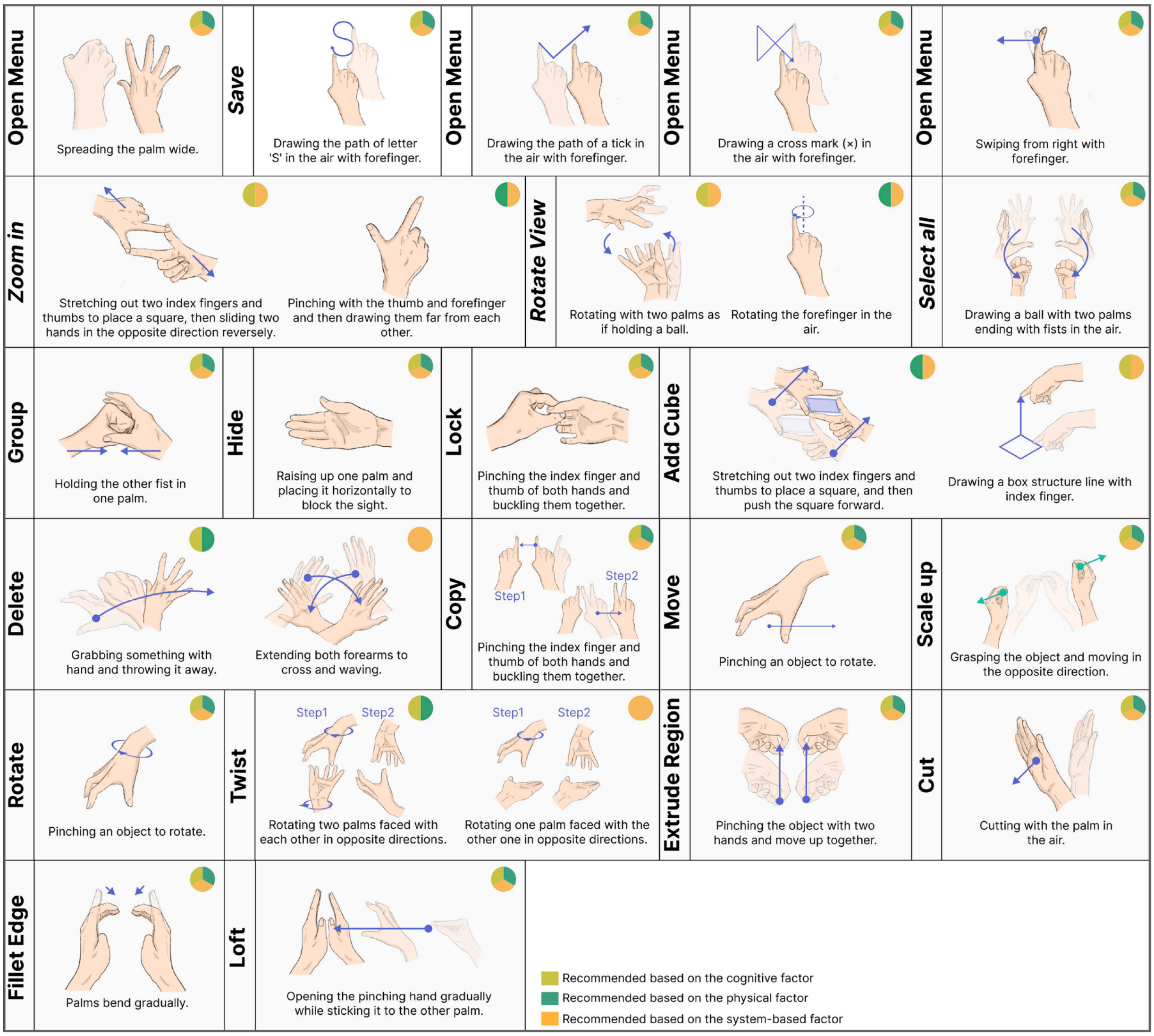


Fig. 12. Hand-based language recommendation for 3D conceptual design based on the comprehensive evaluation.

has long been recognized as a challenging endeavor. For instance, challenges arise when determining the precise initiation and termination times of gestures for practical applications. As different gestures may overlap in sequence, users may execute unintentional inertial actions during operations. Third, the design of gesture recognition algorithms faces challenges in terms of recognition accuracy, which is affected by gesture vocabulary, application scenarios, and user action preferences. Many gestures may share features similar to the complex gesture vocabulary designed in this study. In addition, different users may use the same gesture vocabulary in diverse manners, raising a bar for the recognition algorithm design.

In addition to dealing with technical challenges, the cognitive and physiological loads of users may differ from the results of theoretical experiments. The theoretical gesture language is often put forward by evaluating and testing each gesture vocabulary independently, while users often perform several gestures continuously and repeatedly to complete complex tasks in actual scenes. Consequently, enhancing the usability and user experience of gesture interaction goes beyond creating an intuitive and user-friendly gesture language. Additionally, it requires a demanding set of technical implementation criteria, verifying the theoretical gesture language and exploring the potential challenges in real-life tasks. We released our collected dataset and data processing software package at GitHub to facilitate relevant researchers to further explore real-life applications (see detail in Appendix 3).

6.4. Research limitations

The insights gained from this study can be used as reference for advancing the comprehension of gesture-based interactions or directly applying hand-based vocabulary in software. A structured list of limitations was provided based on the potential application space.

For researchers who wish to use our guidelines or methodology for reference, the following limitations may be considered:

- Problems of cultural dependence may exist in our elicitation design. The participants in this study were from the same country and had a similar cultural background, which might ignore worldwide cognitive variance during gesture elicitation experiments.
- Understanding gesture applications may not be sufficiently holistic. This can be enhanced by evaluating additional factors, such as transferability and social acceptability.
- The laboratory experiment involved a relatively small cohort of participants, and as a result, the insights gleaned might not be representative of all real cases.

For developers who want to apply our hand-based vocabulary to specific software, the following limitations may be considered:

- Biomechanical or ergonomic evaluations were not performed. For long-term and high-intensity gesture applications, the impact of fatigue on interaction should be further considered.
- Gesture recognition accuracy may be influenced by the recognition technology used. Contact hardware sensors were used to capture and record gesture movements. Other gesture recognition methods, such as camera capture or leap motion capture, may affect recognition rates.

7. Conclusion and future work

This paper presents a two-stage experimental study to establish a set of widely available hand-based languages for 3D conceptual design in an MR environment. A novel experiment combining a user-elicitation design and comprehensive evaluation was conducted. Cognitive (i.e., discoverability, learnability, and memorability), physical (i.e., efficiency), and system-based (i.e., algorithm recognition) factors were used to evaluate the performance of the gestures and extract the hand-based interaction language. We provided a set of theoretically optimal hand-based interaction languages for common functionalities in 3D conceptual design and clarified the related application guidelines. Our findings can further expand HCI in MR environments and inspire software developers to create novel hand-driven interaction modes or tools.

Some of the future studies are as follows. First, the real-life use of the proposed hand-based language need be explored to clarify its potential in practical applications. For example, a modeling software that interacts with gesture inputs can be developed. Notably, it is meaningful to invite designers to engage in real 3D conceptual design projects for a long duration and re-evaluate the cognitive and physical factors based on the proposed hand-based language. Second, more factors can be adopted to evaluate hand-based language for a holistic understanding. For example, in addition to the efficiency measured in this study, other physical factors, such as biomechanics and ergonomics, can be explored. Physiological action data can be collected and fitted to a fatigue model to further demonstrate the relationship between actions and fatigue. Third, multimodal interaction modes in an MR environment can be further explored. In addition to hand-based interaction, the potential of other forms of body interactions, such as near-eye and micro-gesture interactions, can be further investigated. The amalgamation of multimodal interaction modes might hold tremendous promise in fostering unparalleled chances of achieving natural HCI.

Declaration of generative AI in scientific writing

The authors declare that they have not used the generative AI and AI-assisted technologies in the writing process.

Funding

This work was supported by the National Key Research and Development Program of China (2022YFB3303300), the National Natural Science Foundation of China (Grant No. 62207024), and the Ng Teng Fong Charitable Foundation in the form of ZJU-SUTD IDEA Grant.

CRediT authorship contribution statement

Lingyun Sun: Funding acquisition, Resources, Writing – review & editing. **Hongbo Zhang:** Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Pei Chen:** Conceptualization, Supervision, Writing – review & editing, Resources. **Zhaoqu Jiang:** Conceptualization, Formal analysis, Investigation, Writing – review & editing. **Xuelong Xie:** Investigation, Formal analysis, Writing – review & editing. **Zihong Zhou:** Investigation, Formal analysis, Writing – review & editing. **Xuanhui Liu:** Supervision, Writing – review & editing. **Xiaoyu Chen:** Supervision, Resources.

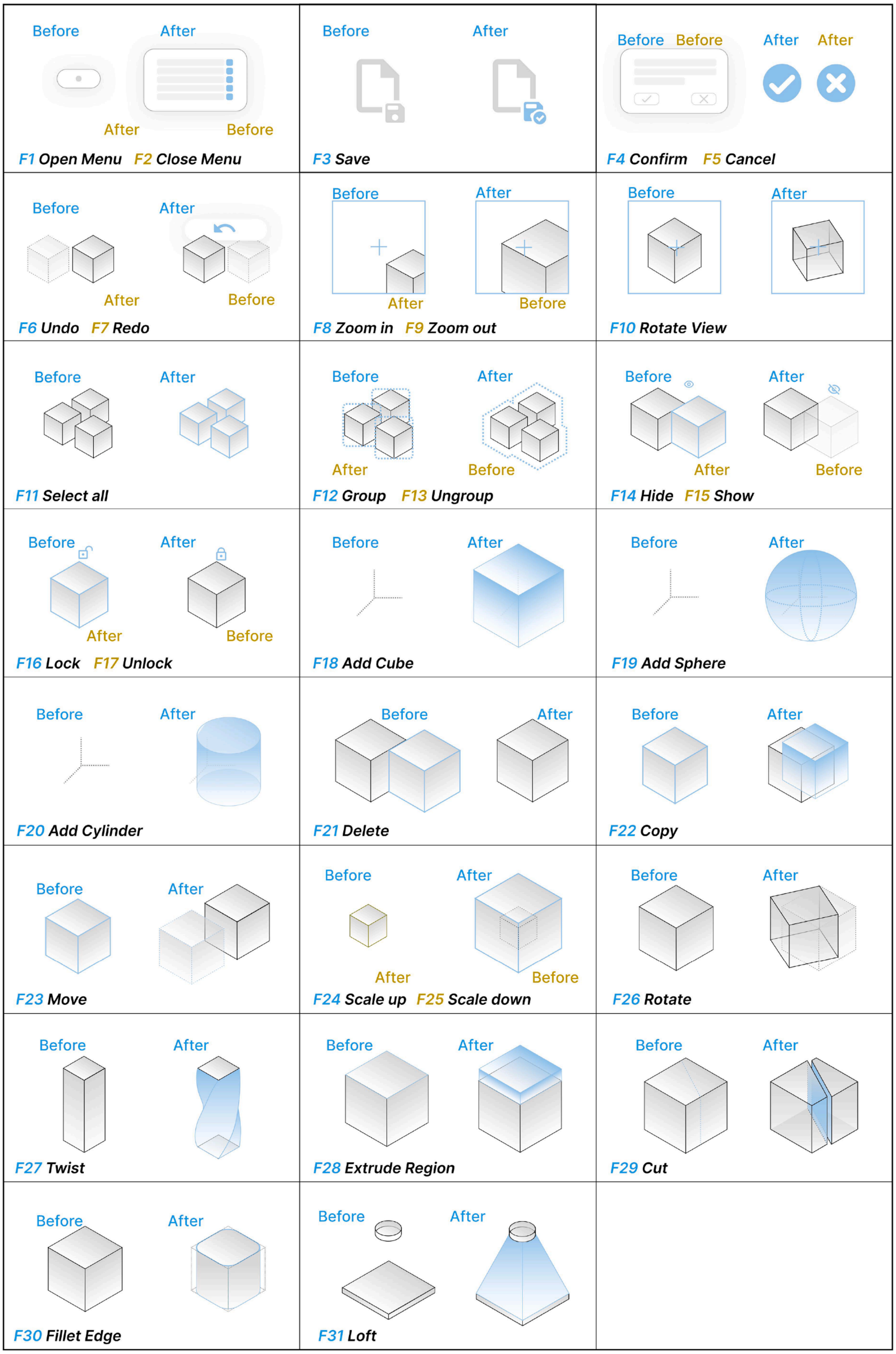
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix 1. The “before” and “after” states of targeted functionality in elicitation design



* During the elicitation design, the description and explanation of targeted functionality were provided to participants. The "before" and "after" states of targeted functionality were also presented to participants through the HMD equipment. Participants were informed to design gesture actions to make the transition between those two states.

Appendix 2. The statistic results of discoverability, learnability, memorability and efficiency score.

D: discoverability (points) L: learnability (points) M: memorability (points) E: efficiency (seconds)

Functionality	Gesture	D *	L *	M *	E*
F1	***G1***	1.04 (0.55)	3.38 (0.65)	2.13 (1.19)	2.77 (0.51)
(Open Menu)	***G2***	0.25 (0.44)	2.29 (0.75)	1.38 (1.31)	2.99 (0.62)
	G3	0.38 (0.49)	2.42 (0.83)	0.33 (0.64)	2.79 (0.54)
		D *	**L** *	**M** *	**E***
F3	***G4***	0.79 (0.59)	2.96 (0.75)	2.25 (0.85)	2.54 (0.51)
(Save)	***G5***	0.88 (0.90)	3.42 (0.83)	2.79 (0.51)	2.58 (0.42)
	G6	0.17 (0.38)	2.08 (0.58)	0.46 (0.93)	2.86 (0.62)
		D *	**L**	**M** *	**E***
F4	***G7***	1.79 (0.98)	4.21 (0.59)	2.38 (0.71)	2.48 (0.35)
(Confirm)	***G8***	2.79 (0.51)	4.38 (0.49)	2.33 (0.76)	2.41 (0.50)
	G9	2.54 (0.59)	4.25 (0.79)	0.50 (0.59)	2.58 (0.43)
		D *	**L** *	**M** *	**E***
F5	***G10***	2.17 (0.87)	4.17 (0.38)	2.33 (1.05)	2.46 (0.03)
(Cancel)	***G11***	2.50 (0.66)	4.29 (0.55)	1.96 (0.95)	2.83 (0.03)
	G12	0.88 (0.90)	3.08 (0.93)	0.63 (0.82)	2.84 (0.04)
		D	**L** *	**M** *	**E***
F6	***G13***	2.17 (0.96)	3.88 (0.90)	2.21 (1.02)	2.52 (0.44)
(Undo)	***G14***	1.58 (0.93)	2.96 (0.75)	1.50 (1.14)	2.70 (0.45)
	G15	1.96 (0.91)	3.71 (0.75)	1.95 (1.14)	2.46 (0.39)
		D *	**L** *	**M** *	**E***
F8	***G16***	2.83 (0.38)	4.13 (0.34)	1.25 (1.26)	2.72 (0.54)
(Zoom in)	***G17***	2.29 (0.69)	4.21 (0.78)	2.58 (0.88)	2.95 (0.45)
	G18	2.75 (0.44)	3.71 (0.86)	0.54 (0.78)	2.80 (0.49)
		D *	**L** *	**M** *	**E***
F10	***G19***	1.75 (0.44)	3.38 (0.71)	1.79 (1.44)	3.23 (0.80)
(Rotate View)	***G20***	1.88 (0.61)	3.25 (0.79)	1.33 (1.40)	3.30 (0.94)
	G21	0.96 (0.36)	2.42 (0.83)	0.17 (0.38)	2.85 (0.94)
		D *	**L** *	**M** *	**E***
F11	***G22***	1.63 (0.92)	3.13 (0.74)	1.96 (1.00)	2.95 (0.55)
(Select All)	***G23***	0.63 (0.49)	2.54 (0.88)	0.17 (0.38)	3.01 (0.49)
	G24	2.25 (0.85)	4.04 (0.46)	2.58 (0.78)	2.75 (0.49)
		D *	**L** *	**M**	**E***
F12	***G25***	1.33 (0.70)	2.96 (0.75)	1.79 (1.22)	2.58 (0.58)
(Group)	***G26***	1.63 (1.01)	3.04 (0.95)	1.58 (1.25)	2.54 (0.55)
	G27	0.63 (0.77)	2.33 (1.17)	0.92 (1.28)	2.65 (0.68)
		D *	**L** *	**M** *	**E***
F14	***G28***	1.54 (1.22)	2.33 (1.17)	2.58 (0.50)	2.52 (0.55)
(Hide)	***G29***	1.83 (0.56)	3.58 (0.72)	2.33 (0.70)	2.46 (0.44)
	G30	0.17 (0.38)	1.75 (0.68)	0.38 (0.49)	2.55 (0.51)
		D *	**L** *	**M** *	**E***
F16	***G31***	2.04 (1.00)	4.00 (0.66)	2.38 (1.13)	2.53 (0.03)
(Lock)	***G32***	1.33 (0.56)	3.29 (0.69)	1.25 (1.15)	2.90 (0.05)
	G33	0.88 (1.12)	3.21 (0.88)	0.88 (1.12)	2.68 (0.05)
F18	***G34***	0.92 (0.88)	2.83 (0.76)	0.92 (0.88)	3.02 (0.67)
(Add Cube)	***G35***	2.38 (0.82)	3.92 (0.50)	2.38 (0.82)	4.27 (0.52)
	G36	2.17 (0.76)	3.92 (0.41)	2.17 (0.76)	3.94 (0.62)
		D *	**L**	**M**	**E**
F19	***G37***	1.25 (0.68)	3.25 (0.79)	2.25 (0.85)	3.09 (0.59)
(Add Sphere)	***G38***	1.83 (0.76)	3.29 (0.91)	2.42 (0.72)	3.16 (0.57)
		D	**L**	**M**	**E***
F20	***G39***	2.83 (0.38)	4.00 (0.66)	1.63 (1.01)	3.09 (0.65)
(Cylinder)	***G40***	2.63 (0.58)	3.92 (0.58)	1.92 (1.10)	3.77 (0.63)
	G41	2.83 (0.38)	4.08 (0.50)	2.08 (0.88)	3.36 (0.69)
		D *	**L** *	**M** *	**E***
F21	***G42***	2.50 (0.83)	4.29 (0.75)	2.92 (0.28)	2.69 (0.03)
(Delete)	***G43***	0.96 (0.20)	2.83 (0.64)	0.21 (0.66)	3.19 (0.04)
	G44	0.58 (0.58)	2.00 (0.78)	0.58 (0.93)	2.75 (0.03)
		D *	**L** *	**M** *	**E***
F22	***G45***	2.67 (0.64)	3.88 (0.45)	2.79 (0.41)	2.65 (0.03)
(Copy)	***G46***	2.50 (0.72)	3.33 (0.64)	1.67 (1.05)	2.86 (0.03)
	G47	1.17 (0.96)	2.83 (0.56)	0.58 (0.58)	2.93 (0.04)
		D *	**L**	**M** *	**E**
F23	***G48***	2.83 (0.38)	4.46 (0.59)	2.83 (0.64)	2.91 (0.52)
(Translation)	***G49***	2.54 (0.51)	4.17 (0.76)	1.54 (0.93)	2.90 (0.60)
		D *	**L** *	**M** *	**E***
F24	***G50***	2.29 (0.62)	3.96 (0.69)	2.29 (1.23)	2.79 (0.04)
(Scale up)	***G51***	1.79 (0.93)	3.33 (0.92)	0.46 (0.88)	3.07 (0.04)
	G52	2.33 (0.70)	3.58 (0.58)	1.00 (1.25)	3.02 (0.04)
		D *	**L** *	**M** *	**E***
F26	***G53***	2.54 (0.66)	4.25 (0.53)	2.42 (1.02)	3.06 (0.62)
(Rotate)	***G54***	1.42 (0.58)	3.21 (0.98)	1.29 (1.27)	3.40 (0.58)
		D *	**L** *	**M** *	**E***
F27	***G55***	2.92 (0.41)	4.88 (0.34)	2.96 (0.20)	2.79 (0.04)
(Twist)	***G56***	2.21 (0.72)	4.38 (0.71)	1.88 (0.61)	3.03 (0.05)
		D *	**L**	**M**	**E***

(*continued on next page*)

(continued)

D: discoverability (points) L: learnability (points) M: memorability (points) E: efficiency (seconds)					
Functionality	Gesture	D *	L *	M *	E*
F28	***G57***	1.67 (0.96)	3.46 (0.66)	1.29 (1.30)	3.11 (0.04)
(Extrude Region)	***G58***	2.33 (0.70)	3.83 (0.70)	1.92 (1.41)	2.62 (0.05)
		D *	**L** *	**M***	**E***
F29	***G59***	2.83 (0.38)	4.71 (0.46)	2.88 (0.34)	2.26 (0.03)
(Cut)	***G60***	2.92 (0.28)	4.75 (0.44)	2.04 (0.55)	2.54 (0.05)
	G61	1.96 (1.04)	3.67 (0.70)	0.88 (0.45)	2.82 (0.04)
		D *	**L** *	**M** *	**E***
F30	***G62***	2.33 (0.70)	3.83 (0.76)	2.88 (0.34)	2.87 (0.05)
(Fillet Edge)	***G63***	1.17 (0.92)	3.00 (0.72)	1.79 (0.88)	3.11 (0.04)
		D *	**L**	**M** *	**E**
F31	***G64***	1.54 (0.93)	3.08 (0.93)	2.46 (1.14)	3.11 (0.78)
(Loft)	***G65***	0.79 (0.66)	3.13 (0.70)	1.46 (1.10)	3.18 (0.77)

* There was a significant main effect for the factor.

Appendix 3. The gesture motion data and data processing file in this study.

In order to facilitate further development of alternative gesture language, we have uploaded the dataset obtained from the experiment and the processed software packages to https://github.com/xiexueshuang/Gesture-Recognition.

References

Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindlbauer, D., Ion, A., Zhao, S., 2012. Understanding mid-air hand gestures: a study of human preferences in usage of gesture types for HCI. Microsoft Research Technical Report MSR-TR-2012-111.

Alkemade, R., Verbeek, F.J., Lukosch, S.G., 2017. On the efficiency of a VR hand gesture-based interface for 3D object manipulations in conceptual design. Int. J. Hum.–Comput. Interact. 33 (11), 882–901. https://doi.org/10.1080/10447318.2017.1296074.

Annett, M., Bischof, W.F., 2013. Your left hand can do it too!: investigating intermanual, symmetric gesture transfer on touchscreens. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1119–1128. https://doi.org/10.1145/2470654.2466144.

Arora, R., Kazi, R.H., Anderson, F., Grossman, T., Singh, K., Fitzmaurice, G., 2017. Experimental evaluation of sketching on surfaces in VR. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 5643–5654. https://doi.org/10.1145/3025453.3025474.

Arroyave-Tobón, S., Osorio-Gómez, G., Cardona-McCormick, J.F., 2015. Air-modelling: a tool for gesture-based solid modelling in context during early design stages in AR environments. Comput. Ind. 66, 73–81. https://doi.org/10.1016/j.compind.2014.10.007.

Ashbrook, D., Starner, T., 2010. MAGIC: a motion gesture design tool. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2159–2168. https://doi.org/10.1145/1753326.1753653.

Bærentzen, A., Frisvad, J.R., Singh, K., 2019. Signifier-based immersive and interactive 3D modeling. In: 25th ACM Symposium on Virtual Reality Software and Technology, pp. 1–5. https://doi.org/10.1145/3359996.3364257.

Blackler, A.L., Popovic, V., Mahar, D.P., 2002. Intuitive use of products. In: Common Ground. Design Research Society International Conference, 2002, pp. 120–134. http://eprints.qut.edu.au/1879/.

Buchmann, V., Violich, S., Billinghurst, M., Cockburn, A., 2004. FingARtips: gesture based direct manipulation in augmented reality. In: Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, pp. 212–221. https://doi.org/10.1145/988834.988871.

Burnett, G., Crundall, E., Large, D., Lawson, G., Skrypchuk, L., 2013. A study of unidirectional swipe gestures on in-vehicle touch screens. In: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 22–29. https://doi.org/10.1145/2516540.2516545.

Choi, E., Kwon, S., Lee, D., Lee, H., Chung, M.K., 2014. Towards successful user interaction with systems: focusing on user-derived gestures for smart home systems. Appl. Ergon. 45 (4), 1196–1207. https://doi.org/10.1016/j.apergo.2014.02.010.

Dean, H.R., 1992. The Automatic Recognition of Gestures. In: The Automatic Recognition of Gestures, 1992. UMI.

Deller, M., Ebert, A., Bender, M., Hagen, H., 2006. Flexible gesture recognition for immersive virtual environments. In: Tenth International Conference on Information Visualisation (IV'06), pp. 563–568. https://doi.org/10.1109/IV.2006.55.

Dey, A.K., Hamid, R., Beckmann, C., Li, I., Hsu, D., 2004. a CAPpella: programming by demonstration of context-aware applications. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 33–40. https://doi.org/10.1145/985692.985697.

Dix, A., Finlay, J., Abowd, G., Beale, R., 2004. Evaluation techniques. Hum-Comput. Interact. 318–364.

Drey, T., Gugenheimer, J., Karlbauer, J., Milo, M., Rukzio, E., 2020. VRSketchIn: exploring the design space of pen and tablet interaction for 3D sketching in virtual reality. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14. https://doi.org/10.1145/3313831.3376628.

Duncan, F.A.R., Fijałkowski, M., 2020. Version 16.0.0. Baroque software. https://vrsketch.eu/.

Garzotto, F., Valoriani, M., 2013. Touchless gestural interaction with small displays: a case study. In: Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI, pp. 1–10. https://doi.org/10.1145/2499149.2499154.

Gibbs, R.W.J., 2006. Embodiment and Cognitive Science. Cambridge University Press, Cambridge, UK.

Gomes, M.G., Ogliari, A., Fernandes, R.B., Marques, K.O., 2022. Evaluation of physical models as creative stimuli in conceptual design of products. Des. Stud. 81, 101–119. https://doi.org/10.1016/j.destud.2022.101119.

Guest, G., Bunce, A., Johnson, L., 2006. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 18 (1), 59–82. https://doi.org/10.1177/1525822X05279903.

Gustafsson, E., Coenen, P., Campbell, A., Straker, L., 2018. Texting with touchscreen and keypad phones—a comparison of thumb kinematics, upper limb muscle activity, exertion, discomfort, and performance. Appl. Ergon. 70, 232–239. https://doi.org/10.1016/j.apergo.2018.03.003.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Hinckley, K., Pahud, M., Benko, H., Irani, P., Guimbretière, F., Gavriliu, M., Chen, X., Anthony, Matulic, F., Buxton, W., Wilson, A., 2014. Sensing techniques for tablet+stylus interaction. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, pp. 605–614. https://doi.org/10.1145/2642918.2647379.

Hinckley, K., Song, H., 2011. Sensor synaesthesia: touch in motion, and motion in touch. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 801–810. https://doi.org/10.1145/1978942.1979059.

Hinckley, K., Yatani, K., Pahud, M., Coddington, N., Rodenhouse, J., Wilson, A., Benko, H., Buxton, B., 2010. Pen + touch = new tools. In: Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology, pp. 27–36. https://doi.org/10.1145/1866029.1866036.

Hostetter, A.B., 2004. On the tip of the mind: gesture as a key to conceptualization. In: Proceedings of the Annual Meeting of the Cognitive Science Society.

Hrvoje, B., Wigdor, D., 2010. Imprecision, inaccuracy, and frustration: the tale of touch input. In: Proceedings of the Tabletops - Horizontal Interactive Displays. Springer.

Huang, J., Jaiswal, P., Rai, R., 2018. Gesture-based system for next generation natural and intuitive interfaces. Artif. Intell. Eng. Des., Anal. Manuf.

Huang, J., Qi, M., Mao, L., An, M., Ji, T., Han, R., 2021. User-defined gestures for mid-air interaction: a comparison of upper limb muscle activity, wrist kinematics, and subjective preference. Int. J. Hum.–Comput. Interact. 37 (16), 1516–1537. https://doi.org/10.1080/10447318.2021.1898825.

Huang, J., Zhang, H., Mao, L., Zhang, D., Li, J., Ji, T., Han, R., 2022. The effect of tablet computer configurations and touchscreen gestures on human biomechanics, performance, and subjective assessment. Int. J. Hum.–Comput. Interact. 1–17. https://doi.org/10.1080/10447318.2022.2111051.

Hutchins, E.L., Hollan, J.D., Norman, D.A., 1985. Direct manipulation interfaces. Hum.–Comput. Interact. 1 (4), 311–338. https://doi.org/10.1207/s15327051hci0104_2.

Kodon, 2018. TenkLabs Kodon. https://www.tenklabs.com/kodon.

Lee, M., Bae, J., 2022. Real-time gesture recognition in the view of repeating characteristics of sign languages. IEEE Trans. Ind. Inf. 18 (12), 8818–8828. https://doi.org/10.1109/TII.2022.3152214.

Liu, S., Zhang, J., Zhang, Y., Zhu, R., 2020. A wearable motion capture device able to detect dynamic motion of human limbs. Nat. Commun. 11 (1), 5615. https://doi.org/10.1038/s41467-020-19424-2.

Löcken, A., Hesselmann, T., Pielot, M., Henze, N., Boll, S., 2012. User-centred process for the definition of free-hand gestures applied to controlling music playback. Multimedia Syst. 18 (1), 15–31. https://doi.org/10.1007/s00530-011-0240-2.

Lopes, D.S., Parreira, P.D., de, F., Paulo, S.F., Nunes, V., Rego, P.A., Neves, M.C., Rodrigues, P.S., Jorge, J.A., 2017. On the utility of 3D hand cursors to explore medical volume datasets with a touchless interface. J. Biomed. Inform. 72, 140–149. https://doi.org/10.1016/j.jbi.2017.07.009.

Lü, H., Li, Y., 2013. Gesture studio: authoring multi-touch interactions through demonstration and declaration. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 257–266. https://doi.org/10.1145/2470654.2470690.

Lyu, F., Li, H., Fu, Q., Liu, Y., Huang, J., Deng, Z., 2023. Effects of spatial constraints and ages on children's upper limb performance in mid-air gesture interaction. Int. J. Hum. Comput. Stud. 170, 102952 https://doi.org/10.1016/j.ijhcs.2022.102952.

Memo, A., Zanuttigh, P., 2018. Head-mounted gesture controlled interface for human-computer interaction. Multimed. Tools Appl. 77, 27–53.

Menolotto, M., Komaris, D.-S., Tedesco, S., O'Flynn, B., Walsh, M., 2020. Motion capture technology in industrial applications: a systematic review. Sensors 20 (19), 5687. https://doi.org/10.3390/s20195687.

Morris, M.R., 2012. Web on the wall: insights from a multimodal interaction elicitation study. In: Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, pp. 95–104. https://doi.org/10.1145/2396636.2396651.

Nielsen, M., Storring, M., Moeslund, T.B., Granum, E., 2004. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. Gesture-Based Commun. Hum.-Comput. Interact. 2915, 409–420. https://doi.org/10.1007/978-3-540-24598-8_38.

OculusMedium. 2016. OculusMedium. https://oculus.com/medium/.

Peshkova, E., Hitz, M., Ahlström, D., 2016. Exploring user-defined gestures and voice commands to control an unmanned aerial vehicle. In: Proceedings of the International Conference on Intelligent Technologies for Interactive Entertainment, pp. 47–62.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., McCullough, K.E., Ansari, R., 2002. Multimodal human discourse: gesture and speech. ACM Trans. Comput.-Hum. Interact. 9 (3), 171–193. https://doi.org/10.1145/568513.568514.

Rautaray, S.S., Agrawal, A., 2015. Vision based hand gesture recognition for human computer interaction: a survey. Artif. Intell. Rev. 43 (1), 1–54. https://doi.org/10.1007/s10462-012-9356-9.

Riener, A., Ferscha, A., Bachmair, F., Hagmüller, P., Lemme, A., Muttenthaler, D., Pühringer, D., Rogner, H., Tappe, A., Weger, F., 2013. Standardization of the in-car gesture interaction space. In: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 14–21. https://doi.org/10.1145/2516540.2516544.

Rieuf, V., Bouchard, C., Meyrueis, V., Omhover, J.-F., 2017. Emotional activity in early immersive design: sketches and moodboards in virtual reality. Des. Stud. 48, 43–75. https://doi.org/10.1016/j.destud.2016.11.001.

ShapeLab, 2018. ShapeLab. https://store.steampowered.com.

Troiano, G.M., Pedersen, E.W., Hornbæk, K., 2014. User-defined gestures for elastic, deformable displays. In: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces, pp. 1–8. https://doi.org/10.1145/2598153.2598184.

Tung, Y.-C., Hsu, C.-Y., Wang, H.-Y., Chyou, S., Lin, J.-W., Wu, P.-J., Valstar, A., Chen, M. Y., 2015. User-defined game input for smart glasses in public space. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3327–3336. https://doi.org/10.1145/2702123.2702214.

van Dijk, C.G.C., 1995. New insights in computer-aided conceptual design. Des. Stud. 16 (1), 62–80. https://doi.org/10.1016/0142-694X(95)90647-X.

Vatavu, R.-D., 2012. User-defined gestures for free-hand TV control. In: Proceedings of the 10th European Conference on Interactive TV and Video, pp. 45–48. https://doi.org/10.1145/2325616.2325626.

Vatavu, R.-D., Anthony, L., Wobbrock, J.O., 2014. Gesture heatmaps: understanding gesture performance with colorful visualizations. In: Proceedings of the 16th International Conference on Multimodal Interaction, pp. 172–179. https://doi.org/10.1145/2663204.2663256.

Vinayak, Murugappan, S., Liu, H., Ramani, K., 2013. Shape-it-up: hand gesture based creative expression of 3D shapes using intelligent generalized cylinders. Comput. Aided Des. 45 (2), 277–287. https://doi.org/10.1016/j.cad.2012.10.011.

Vinayak, Ramani, K., 2015. A gesture-free geometric approach for mid-air expression of design intent in 3D virtual pottery. Comput. Aided Des. 69, 11–24. https://doi.org/10.1016/j.cad.2015.06.006.

Vuletic, T., Duffy, A., Hay, L., McTeague, C., Campbell, G., Grealy, M., 2019. Systematic literature review of hand gestures used in human computer interaction interfaces. Int. J. Hum. Comput. Stud. 129, 74–94. https://doi.org/10.1016/j.ijhcs.2019.03.011.

Wang, L., Shen, W., Xie, H., Neelamkavil, J., Pardasani, A., 2002. Collaborative conceptual designÐstate of the art and future trends. Comput. Aided Des.

Wigdor, D., Benko, H., Pella, J., Lombardo, J., Williams, S., 2011. Rock & rails: extending multi-touch interactions with shape gestures to enable precise spatial manipulations. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1581–1590. https://doi.org/10.1145/1978942.1979173.

Wobbrock, J.O., Aung, H.H., Rothrock, B., Myers, B.A., 2005. Maximizing the guessability of symbolic input. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems, pp. 1869–1872. https://doi.org/10.1145/1056808.1057043.

Wobbrock, J.O., Morris, M.R., Wilson, A.D., 2009. User-defined gestures for surface computing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1083–1092. https://doi.org/10.1145/1518701.1518866.

Wobbrock, J.O., Wilson, A.D., Li, Y., 2007. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, pp. 159–168. https://doi.org/10.1145/1294211.1294238.

Wu, H., Wang, Y., Qiu, J., Liu, J., Zhang, X., Luke, 2019. User-defined gesture interaction for immersive VR shopping applications. Behav. Inf. Technol. 38 (7), 726–741. https://doi.org/10.1080/0144929X.2018.1552313.

Xia, H., Araujo, B., Grossman, T., Wigdor, D., 2016. Object-oriented drawing. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4610–4621. https://doi.org/10.1145/2858036.2858075.

Xia, H., Glueck, M., Annett, M., Wang, M., Wigdor, D., 2022. Iteratively designing gesture vocabularies: a survey and analysis of best practices in the HCI literature. ACM Trans. Comput.-Hum. Interact. 29 (4), 1–54. https://doi.org/10.1145/3503537.

Xiao, Y., Miao, K., Jiang, C., 2021. Mapping directional mid-air unistroke gestures to interaction commands: a user elicitation and evaluation study. Symmetry 13 (10), 1926. https://doi.org/10.3390/sym13101926.

Yang, X., Lin, L., Cheng, P.-Y., Yang, X., Ren, Y., Huang, Y.-M., 2018. Examining creativity through a virtual reality support system. Educ. Technol. Res. Dev. 66 (5), 1231–1254. https://doi.org/10.1007/s11423-018-9604-z.

Ye, J., Campbell, R.I., 2006. Supporting conceptual design with multiple VR based interfaces. Virtual Phys. Prototyp. 1 (3), 171–181. https://doi.org/10.1080/174527506010171.

Yu, E., Arora, R., Stanko, T., Bærentzen, J.A., Singh, K., Bousseau, A., 2021. CASSIE: curve and surface sketching in immersive environments. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–14. https://doi.org/10.1145/3411764.3445158.

Zhou, Z., Chen, P., Lu, Y., Cui, Q., Pan, D., Liu, Y., Li, J., Zhang, Y., Tao, Y., Liu, X., Sun, L., Wang, G., 2022. 3D deformation capture via a configurable self-sensing IMU sensor network. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 7, pp. 1–24. https://doi.org/10.1145/3580874.

Zimmerman, T.G., Lanier, J., Blanchard, C., Bryson, S., Harvill, Y., 1987. A hand gesture interface device. Spec. Interest Group Comput.-Hum. Interact. 18 (4) https://doi.org/10.1145/1165387.275628.

Zou, N., Gong, Q., Chai, Q., Chai, C., 2023. The role of virtual reality technology in conceptual design: positioning, applications, and value. Dig. Creativity 1–25. https://doi.org/10.1080/14626268.2023.2166080.