

Article

Mapping Directional Mid-Air Unistroke Gestures to Interaction Commands: A User Elicitation and Evaluation Study

Yiqi Xiao ^{1,*}, Ke Miao ¹ and Chenhan Jiang ²

¹ College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai 200093, China; miaoke@usst.edu.cn

² College of Architecture and Urban Planning, Tongji University, Shanghai 200092, China; 1510092@tongji.edu.cn

* Correspondence: yiqi.xiao@usst.edu.cn

Abstract: A stroke is the basic limb movement that both humans and animals naturally and repetitiously perform. Having been introduced into gestural interaction, mid-air stroke gestures saw a wide application range and quite intuitive use. In this paper, we present an approach for building command-to-gesture mapping that exploits the semantic association between interactive commands and the directions of mid-air unistroke gestures. Directional unistroke gestures make use of the symmetry of the semantics of commands, which makes a more systematic gesture set for users' cognition and reduces the number of gestures users need to learn. However, the learnability of the directional unistroke gestures is varying with different commands. Through a user elicitation study, a gesture set containing eight directional mid-air unistroke gestures was selected by subjective ratings of the direction in respect to its association degree with the corresponding command. We evaluated this gesture set in a following study to investigate the learnability issue, and the directional mid-air unistroke gestures and user-preferred freehand gestures were compared. Our findings can offer preliminary evidence that "return", "save", "turn-off" and "mute" are the interaction commands more applicable to using directional mid-air unistrokes, which may have implication for the design of mid-air gestures in human-computer interaction.

Keywords: directional unistroke gesture; user elicitation; spatial representation; semantic association; learnability



Citation: Xiao, Y.; Miao, K.; Jiang, C. Mapping Directional Mid-Air Unistroke Gestures to Interaction Commands: A User Elicitation and Evaluation Study. *Symmetry* **2021**, *13*, 1926. <https://doi.org/10.3390/sym13101926>

Academic Editor: Tomohiro Inagaki

Received: 14 September 2021

Accepted: 9 October 2021

Published: 13 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the field of human-computer interaction (HCI), gestures are known as a promising input in showing an interaction task or a command which semantically includes the action and the target object. Comparing with touch-based gesture, touchless mid-air gesture, if robust, demonstrate advantages as a style of natural HCI without the constraint of physical interfaces. This enables mid-air gesture to become an addition to voice-based or physical controls [1] or "a primary interaction style in many user scenarios such as interactive VR/AR applications, human-robot interactions, smart homes and IoT, interacting with autonomous devices in the wild" [2].

With a growing demand for various mid-air interaction scenarios, considerable attention has been devoted to the participatory design of gesture-based applications in recent years. For this purpose, user elicitation study became increasingly popular as a design approach to defining preferable gestures for a set of touchless interactions with remote displays or devices [2]. However, user elicitation studies have the nature of openness of its own. It is important to note that the user-preferred gestures for a certain command could vary across different elicitation studies, e.g., free-hand gestures for TV or large display [3–6]. Moreover, the taxonomy of user-elicited gestures indicates a diverse range of mental models [7] for constructing and understanding the command-to-gesture mappings. To our best

knowledge, few studies have been done with the concern for the standardization of such mappings aiming at universal gesture design.

In order to unify the way of mapping the command to a gesture, we concentrate on the directionality of unistroke gestures, attempting to link the command to gestural expression with spatial representations. That is, only the direction of in-air movement is utilized to denote the semantic meaning of command. Related works suggest that the activation of spatial orientation representations exists in the process of verb comprehension [8,9]. These representations are the motivation to construct orientational metaphors [10] which are more systematic descriptions than ontological and structural metaphors, and enable the linguistic elements in commands to be directly associated with the direction of unistrokes. In the 3D space there are six major directions: *upward*, *downward*, *leftward*, *rightward*, *backward*, and *forward*, meaning that there are mainly six directions for mid-air unistroke gesture that can be metaphorically used (Figure 1). As for the performance, mid-air stroke gesture has its advantages in the production time and the flexibility of movement length, thus can alleviate the perceived difficulty of gestures to some degree [11]. With less physical effort, this kind of gesture vocabulary is possible to lower the barriers for technology-naïve users to learning and performing.

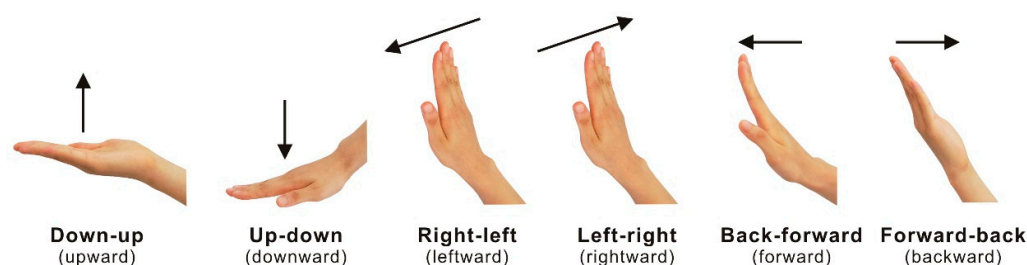


Figure 1. The six directional mid-air unistroke gestures.

A major feature of directional gesture is that it provides a symmetrical cognition to understand the semantics of its corresponding interactive command. As we know, in many cases the semantics of a command can be evidently understood as positive or negative, e.g., *open* vs. *close*. On the one hand, this trend helps in mapping the binary state option to a pair of symmetrical stroke gestures and thus to express meaning through the direction of mid-air unistroke. On the other hand, we can expect that there are multiple explanations for the association of the semantic meaning of a command with the moving direction of a mid-air unistroke gesture (e.g., we can map either an upward or a rightward motion to *volume up*). This could negatively affect the memorability of the mapping of a directional unistroke gesture to such a command.

The present study is conducted, considering the difference of verb-like commands in activating spatial representations, to identify the category of commands for which the directional mid-air unistroke gestures can work. To this end, a user elicitation study and an evaluation study is required. The rest of this paper is structured as follows. A brief overview of the cognitive strategies of directional unistroke gesture in existing studies is firstly offered, followed by discussing the theoretical background of mapping interaction commands to directionality of stroke gestures. In the next section, gestures are elicited by user rating of associating the moving direction of gesture with semantic meaning of the command. Finally, we gained a set of directional mid-air unistroke gestures with strong association to a particular direction. In the following section, we implement an in-lab test to evaluate the learnability of these mid-air unistroke gestures. The present study proposes a vocabulary of mapping the interaction commands to mid-air unistroke gestures based on the spatial representations in semantic understanding of commands. By comparing the user-elicited unistroke gestures with freehand gestures, the applicability of this vocabulary to specific commands is tested. Our method may offer implications for designing easy-to-learn and easy-to-perform mid-air gestures.

2. Related Work

2.1. Command-to-Gesture Mappings for Mid-Air Unistroke Gesture

In the broad sense, a stroke gesture for human–computer interaction can be performed by our upper and lower limbs [12], or even by the head or whole body [13]. As the basic body movements that humans naturally and repetitiously perform, stroke gestures are of a motor-intuitive nature [14] and so have been widely applied in touch-based interfaces for pen input [15–17], text-editing [18] smartphone use [19] and game [20]. Stroke gesture is also a common input for mid-air interaction, in particular for the control of cursor position [21] and distal pointing [22] on distant displays [23]. There are a number of studies focusing on the in-air limb movement for completing interactive tasks related to manipulating a 3D object [24], ubiquitous environment [25] and TV control [3]. In the present study we shift our focus to the unidirectional linear hand motions in a 3D space. More specifically, we direct our attention toward understanding the use of mid-air unistroke gestures in expressing semantic meanings of interactive commands. Very few studies, so far as we know, have explored similar topics except for Burnett et al. [26].

There has been a considerable number of studies reporting the mappings of interactive commands to freehand stroke gestures [5,6,27–30]. In most of these studies, user elicitation is adopted to obtain gesture proposals that can meet end-users' requirements and expectations. Without any constraint in gesture design criteria, participants tend to create unistroke gestures (i.e., move hand vertically or horizontally) for some specific commands, especially for *previous/next channel* and *volume up/down* [4,31,32]. The mental model of making gestures for the first two commands was interpreted in [33] as the use of two metaphors: moving the viewing window (as it happens with scrolling actions in classical GUI design) and moving the items themselves. However, for most of the tasks in abovementioned studies, the proportion of mid-air unistroke gesture in the elicited gesture set is not the highest. Aigner et al. [34] investigated the distribution ratio of gesture types for ten target tasks, found that *remove* and *cancel* are the only two for which more semaphoric strokes were chosen than the other types. The above results suggest that mid-air unistroke gestures are hardly the first choice for users when they are allowed to define gestures for some tasks.

In addition to user elicitation, mid-air unistroke gestures are presented in expert-level designs for gesture technology. The semantic associations of these gestures with task usually fall into two categories: iconic and metaphoric [35]. Iconic gestures are those in which the form of the gesture or its manner of execution embodies picturable aspects of semantic content [35]. For example, Vatavu [36] evaluated a smart-pocket gesture that can transfer information from a pocket device to the large display by raising the hand and then pointing forward. Loehmann's concept design [37] enabled drivers to exchange music or initiate calls with nearby cars through a *from-outside-to-inside* hand gesture depicting the process an object is grasped from outside the window. Similar studies on human-vehicle interactions, such as [38], reported tasks such as *scroll* and *return* expressed by mid-air swiping of the finger. Stroke gestures denoting semantic content of the command with the direction of hand movement are metaphoric [39]. One example of such a kind of stroke gesture was proposed by Bacim, Nabyouni and Bowman [40], using a chef's knife metaphor to map the left-to-right motion onto slicing the dataset of a three-dimensional image. Ackad, Kay and Tomitsch [41] defined a semaphoric gesture set for cyclists which uses "lift up hand" and "put down hand" to respectively denote the two commands: *more* and *back*. In summary, there are some cases successfully mapping mid-air unistroke gestures to relatively more abstract tasks. However, these designs varied in the metaphor they used.

2.2. Theories of Semantically Mapping a Command to the Direction

While directional unistroke gestures do not make up the main part of user-elicited gestures, other types of gesture can also exhibit direction-related characteristics, such as open the door—push forward with right hand [30], yes—thumbs-up hand pose [33] and turn

on—point up [32]. Such findings imply the role that orientational metaphors may play in the process of building task-to-gesture mappings. Lakoff and Johnson [42] coined the term “orientational metaphor” and defined it as a system using orientational concepts such as *up–down*, *forward–back*, *near–far* or *center–periphery* to construe more abstract concepts. Orientational metaphors structure concepts with the non-metaphorical linear orientations [42], e.g., the left side is perceived as *past* while the right as *future*. Orientational metaphors are often embodied in spatial-valence associations [43,44]. A downward movement could be used to indicate some “negative” intentions such as *decrease volume* [29]. However, as we know, this direction not always has a negative connotation. For example, in the case of falling fruit the downward movement represents the maturity [45] and bending down carries a meaning of modesty.

There are different explanations for the mechanism whereby a metaphor is constructed to connect a concept and orientation/direction. According to the neural theory of metaphor [46], orientational metaphor might be the result of repetitive activation of subjective and sensory-motor experience which finally leads to connection of neurons. For example, after observing the relationship between the height of water in a cup and its volume for numerous times, the respective neurons for these two concepts are activated simultaneously and a metaphor—more is up—is thus formulated [42]. A typical case of such metaphors, primary metaphor, automatically associates people’s subjective experience with the sensorimotor experience gained from repetitive interactions with the environment [47]. Therefore, it is reasonable to believe that different cultures [48], viewpoints [49,50] and embodied experiences [51] may contribute to the variety of the cognitive mappings of spatial orientation to some particular concepts.

When focusing on the sense of direction for verbs, theories besides conceptual metaphor theory have been developed as alternatives. Researchers claim that the spatial forms of representation are part of the metaphoric understanding that underlies much of human language, and are invoked in verb comprehension [8,52,53]. In the view of Narayanan [52], a reader understands a word or a sentence depicting an activity concurrently with simulating the motion by reconstituting the related human experience. Such a dynamic, perceptual simulation [54,55] triggers visual imagery [56,57] in our mind to which the manipulations in the reality are recurred. Zwaan [53] advocated the immersed experienter framework that also viewed the word comprehension as a simulation process. In this process, spatial representations are activated to reconstruct the experiential image of motion [8]. The semantic meaning of a verb can be construed with such spatial representations containing directional elements, semantic associations are therefore structured to some extent to form orientational metaphors.

Our study is an effort to assess how well the verb-like commands can be represented by directional unistroke gestures. The first experiment in our paper is conducted to reveal the degree that people associate interactive commands with certain spatial direction. This experiment was inspired by the work of Richardson et al. [58], where they claimed people reached a relatively higher agreement on associating specific verbs with a moving direction. Richardson et al. [58] used a square to represent the agent while a circle the patient, to form a subject–verb–object (SVO for short) sentence. Participants were then forced to choose one from four directions that best depicts the event described by such a sentence. It was found that participants tended to ascribe horizontal motion to specific verbs, while vertical to others. In an ensuing study [9], to assess the activation degree of spatial representations in real-time comprehension of verbs, a dual-task experiment was devised. It required the participants to respond to the stimuli which were vertically or horizontally displayed according to meanings of the recorded SVO sentences. Such an experimental design was seen in an earlier study conducted by Chatterjee, Southwood and Basilico [59] on the relationship between verb comprehension and the direction of its motion. Participants were told to draw a moving trajectory of the imaginary action immediately after they heard a simple active sentence. However, there has been some debate as to whether the verb can be independently mapped to spatial representation or not. Some works, for example, the

study of Bergen et al. [60], argued that spatial representations are the results of contextually understanding the whole sentence instead of merely lexical association. As suggested by Wu, Mo and Wang [61], the activation of spatial representation is not contingent on the language context, but rather only derived from the semantic interpretations of the verbs. The sentence, or more specifically, the constituents other than predicates, facilitate but not dominate the judgment of spatial representations. This would imply that such an activation process is automatic and not strategic. Considering the possible impacts of sentence context, we choose the SVO construction to describe the interaction commands.

3. Experiment 1: User Elicitation Study

This study examines the most agreed association of every command in a list with its semantic moving direction from user's perspective, thereby identifying the activation of spatial representation is whether direction-specific or not during the comprehension of commands.

3.1. Interaction Scenarios and Commands

To begin with, we chose two use scenarios as shown in Table 1, each consisting of eight commands. The reasons for these two options are (1) they are scenarios suitable for the application of freehand gesture technology (e.g., using a large display or interacting with the in-vehicle interface), and (2) they contain many commonly-used actions and are familiar to computer users.

Table 1. Descriptions of the 16 commands for two use scenarios.

Commands/Referents	Descriptions	Scenarios
Display	To awaken a device and images show on the screen.	Controlling a multi-media program or a device
Turn-on	To turn on a device or a machine to make it work.	
Turn-off	To turn off a device or a machine.	
Mute	To make the sound of a device quiet.	
Screen Capture	To capture an image displayed on a screen.	
Start	To turn on a device, an instrument or open an application.	
Stop	To turn off a device, an instrument or close an application.	
Zoom in	To show a close-up picture of items.	
Accept	To accept a request from system or other users.	Browsing web pages, using Apps or doing remote interactions
Delete	To discard a file or remove something.	
Hide	To make a program not show on the current page.	
Pop up	To make a hidden program or a window show on the current page.	
Reject	To reject a request from system or other users.	
Return	To go back to a previous stage of the program.	
Save	To save a file, picture or text.	
Send	To send out a digital file or signal.	

We referred to the previous user elicitation studies [4,25,30], extracting 125 commands to form a command list. The commands that can be obviously associated with direction, for instance, volume up/down and next/previous, were excluded from the list. Related functions and requirements were then selected and distributed to each scenario. To avoid similarity within and across applications, semantically similar commands (e.g., volume up and temperature up, enlarge object and zoom in, delete and remove) were merged into one verb or verb phrase as the referent [7]. To this end, we employed a qualitative analysis method to classify the mental simulations of the meaning of verbs. The first stage was inviting designers to diagram the commands, thereby finding out the similar ones according to the shared semantic structure. Next, these commands were grouped into one referent with a description specifying its intended result (Table 1).

3.2. Experimental Design

The experiment designed for the first stage of our study is clearly not a standard user elicitation because the unistroke gestures are predefined as directional. Participants just have to choose the most “appropriate” gesture among six options for each command according to their own feelings. This part of work is processed in a similar way to the choice-based elicitation method [3,30,31], consisting of two experimental sessions. We first adopted the quick response test employed by Richardson et al. [58] to investigate which spatial orientation representation is most likely to be activated for understanding a referent. In this test, participants were required to choose one direction that best depicts the command described by a SVO sentence by intuition. We collected the data so as to compute the agreement rating [62] of each command. The second session is user rating where participants were asked to rate each direction with a 3-point scale indicating different degrees of command-to-direction association: 3-fit for the command, 2-not very fit, 1-not fit. The reasons for this session are as follows: (1) The degree of activating spatial representations for specific verbs could be relatively low. For example, it could happen that twenty among thirty participants choose the upward movement as the best for a command, but they just choose the least bad option from many bad options. In this case, the high frequency of being selected by users does not necessarily mean the strong association of a moving direction with the corresponding command. (2) There could be some commands with which two or more directions can be equally associated, but participants are forced to choose one of them. With the rating scale, the strength of the association between a referent and a moving direction can be quantified, and thus the strengths of the different referent-direction associations can be compared.

3.3. Participants

Thirty student participants in the age range from 18 to 35 years ($M = 23.37$, $SD = 4.38$) took part in the elicitation study. Education background of participants is diversified, but none had studied linguistics and psychology. Moreover, all the participants are right-handers, and none of them had any experience with gesture design or gesture-based mid-air interaction techniques. Participants are volunteers and asked no reward for their participation.

3.4. Apparatus

A portable computer, ASUS 550JX4200, was deployed for the quick response test, showing participants the sentence and the image of directions with a video. The size of the display is approximately 34.5×19.4 cm with a resolution of 1920×1080 pixels. The experimental procedure was recorded with a camcorder.

3.5. Experimental Materials

In our experiment, we defined three sentence patterns for the description of commands: (i) subject-verb (e.g., the music stops), (ii) verb-object-preposition (e.g., save the content into...), and (iii) verb-object (e.g., start this program). The activation of spatial representation in verb comprehension, in some cases, depends on the body part that performs the action [63]. It also will be either promoted or inhibited in the context of using spatial languages. In order to avoid these effects, we chose the neutral, non-emotional nouns to construct the sentences and eliminated additional descriptive words which might have associations with specific directions.

3.6. Procedure

Prior to the experiment, participants read a brief introduction about the requirements of the whole test. The experiment began with the quick response test where participants were instructed to judge which direction of motion (represented as an arrow symbol) could best depict the sense of moving direction for a given verb and thus be most semantically associated with this verb-like command. Time limit for each judgment in this test was 10 s (a 10-s clip for each command). We showed participants the intended result of command

(e.g., start playing a video for *start*) with an animation in the following stage, telling them to confirm the choice made in the quick response test. Participants also have the right not to choose any one of the directions. More details of the result can be seen in Table 2. Participants then rated how well each of the six directions match the command. The experimenter suggested that participants carefully consider the scores. They were also allowed to give one or more other directions with the same rating.

Table 2. The frequency that participants choose one direction of mid-air motion as the best description of the sense of moving direction for a given command.

Commands/Referents	Upward	Downward	Leftward	Rightward	Backward	Forward
Accept	3	3	2	11	6	5
Delete	2	8	5	13	2	0
Display	8	2	0	3	2	12
Hide	2	14	0	0	12	2
Mute	0	16	4	0	7	1
Pop up	12	9	0	3	5	1
Reject	1	4	8	5	9	3
Return	3	4	14	2	6	1
Save	3	13	2	2	5	2
Screen-capture	3	11	4	7	1	4
Send	5	1	0	6	0	16
Start	13	2	0	5	0	10
Stop	2	12	0	1	2	0
Turn-off	0	17	1	7	0	5
Turn-on	20	0	0	4	0	6
Zoom-in	12	1	0	1	3	13

In addition to rating the directions, each command was presented with the textual description and the animation to participants to propose their own freehand gesture preference for it. This proposal should be a single-hand gesture. At the end of experiment, participants were asked to explain their scores and preferred gestures by verbally describing the semantic association of the command-to-direction mapping. The experiment time ranged from 45 to 60 min.

3.7. Results

We calculated agreement ratings (AR) of commands using the revised formula proposed by Vatavu and Wobbrock [62], see Formula 1, where P is the set of all proposed gestures for a referent r , $|P|$ the size of the set, and P_i subsets of identical proposals from P .

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

The results are shown in Figure 2. *Turn-on* has the highest score amongst 16 commands in terms of directional gesture, while *accept* gets the lowest score. Nine participants selected leftward direction as the best gesture for *reject*, and there were also nine votes for the backward direction. For most of the commands, the ARs of freehand gestures are not as high as those of directional unistroke gestures due to the unlimited number of potential options. *Zoom-in*, *screen-capture* and *accept* are the exceptions. For these three commands, participants have a stronger preference for certain freehand gestures based on their prior experience. Twenty-five of the thirty participants choose *finger splay* to represent *zoom-in*, fourteen choose *drawing a lasso* for *screen-capture*, and thirteen choose *thumbs up* for *accept*. In consideration of this, it seems not a better choice to represent these commands with directional unistroke gestures.

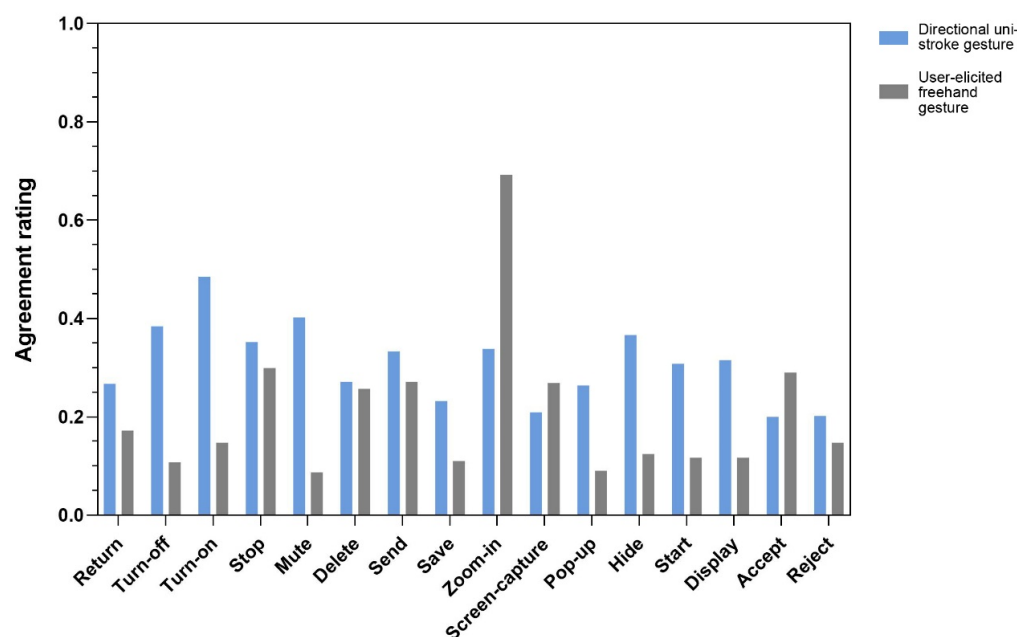


Figure 2. Agreement ratings of the 16 commands.

Since no normal distribution of the subjective scores was shown, we adopted a Wilcoxon's signed-rank test (confidence level = 95%) to compute the statistical difference in rating of directions for every command. Table 3 shows the two directions which received the highest and second-highest average score for each command. A highest-scoring direction was assigned as the one an unistroke gesture used for the command moves towards. The statistically significant difference between the highest and the second-highest score indicates a direction-specific activation of spatial representation. There are eight such mappings: *delete*—rightward, *mute*—downward, *return*—leftward, *save*—downward, *send*—forward, *stop*—downward, *turn-off*—downward and *turn-on*—upward.

Table 3. The differences between the highest and second-highest-scoring semantic association of referents with moving directions.

Commands	Mean Value		Difference		
	Highest Score	Second-Highest Score	Z	p-Value	r
Accept	1.900 (rightward)	1.733 (forward)	0.936	0.349	0.101
Delete	2.367 (rightward)	1.867 (downward)	2.368	0.018	0.317
Display	2.200 (forward)	2.000 (upward)	0.953	0.340	0.119
Hide	2.300 (backward)	2.167 (downward)	0.744	0.457	0.076
Mute	2.400 (downward)	1.833 (backward)	2.635	0.008	0.352
Pop up	2.267 (upward)	1.967 (downward)	1.455	0.146	0.168
Reject	1.967 (leftward)	1.900 (backward)	0.371	0.710	0.038
Return	2.467 (leftward)	2.033 (backward)	2.168	0.030	0.268
Save	2.267 (downward)	1.600 (backward)	2.880	0.004	0.405
Screen-capture	2.167 (downward)	1.900 (rightward)	1.291	0.197	0.167
Send	2.533 (forward)	2.067 (rightward)	2.004	0.045	0.296
Start	2.433 (upward)	2.233 (forward)	1.166	0.243	0.132
Stop	2.167 (downward)	1.567 (backward)	2.830	0.005	0.393
Turn-off	2.467 (downward)	1.933 (rightward)	2.537	0.011	0.280
Turn-on	2.600 (upward)	2.133 (forward)	2.311	0.021	0.317
Zoom-in	2.567 (upward)	2.500 (forward)	0.440	0.660	0.054

3.8. Discussion of Experiment 1

The results of experiment 1 revealed the difference between commands in the strength of association with direction. It is unsurprising that some commands such as *turn on* and *return* are those that can be more naturally mapped to one particular direction. Such commands are featured by a significant difference in the frequency between the highest-rated direction and the others. For some commands, we observed a slight difference in both the frequency and score between the highest-rated direction and others, such as *zoom in* and *hide*. This means that for participants, these commands can be semantically associated with specific directions, none of which has a significantly greater strength of semantic association. There are also some commands for which none of the average ratings of directions was high enough, such as *accept* and *reject*. Participants appeared to be forced, although finally a movement direction with the highest rating was obtained, to choose one from many unrelated directions to represent these commands. The relatively low rating of such a direction indicated that it was more arbitrarily mapped to the corresponding command. Accordingly, as the mappings of this kind are random in some sense, it seems not very much appropriate for the related commands to be semantically represented by the direction of mid-air unistroke gesture. The commands which were regarded as not fit for being mapped to the direction of gesture movement would not be incorporated in the next evaluation study.

4. Experiment 2: Evaluation Study

In the second experiment, we are interested in evaluating the idea of mapping directional mid-air unistroke gestures onto interaction commands. In order to test the difference of user-selected unistroke gestures in learnability for the commands which have a connection with one particular direction, we designed a two-phase experiment. We also compared the directional unistroke gestures and the freehand gestures elicited from end-users in the elicitation study. These two works helped to identify which command-to-gesture mappings would be more effective in facilitating users' learning and gesture recall.

4.1. Defining the Directional Unistroke Gestures and User-Elicited Gestures

Among the eight commands to which there are direction-specific associations, four were mapped to the same direction, i.e., the downward movement. In fact, this is a main problem for the defining of directional unistroke gestures, that is, six directions have to correspond to a variety of commands, which will lead to confusions and so the number of command-to-gesture mappings should be limited when applied to the real situation. In order to disambiguate the mappings to a certain extent, we specified different amplitudes for the same directional unistroke. For the commands with high frequency of use and low importance, the unistroke gesture is a palm movement performed with the wrist acting as the rotation axis. For the commands with low frequency of use and high importance, the whole arm is required to perform the unistroke gesture with the shoulder as the rotation axis. We here defined eight unistroke gestures as can be seen in Figure 3. The user-elicited freehand gestures with the highest frequency for these eight commands are also provided in Figure 3. All the sixteen gestures were encoded by capital letters.

4.2. Experimental Design

The first phase (i.e., learning phase) of the experiment followed a between-subjects design in which participants have to remember and learn gestures in a limited time. We recruited forty-eight participants, dividing them equally into three groups. We set three levels of time limit for learning: 160 s, 320 s and 480 s. Each of the three time-limits was assigned to one of the three groups of participants. The participants in each group were divided into half, one half first learned all the directional unistroke gestures, and the other freehand gestures. After the learning phase, participants immediately performed the corresponding gesture according to the command name, i.e., the commands. The order

of command names was randomized. In summary, the learning phase consisted of 16 participants \times 3 groups \times 8 gestures \times 2 types = 768 trials.

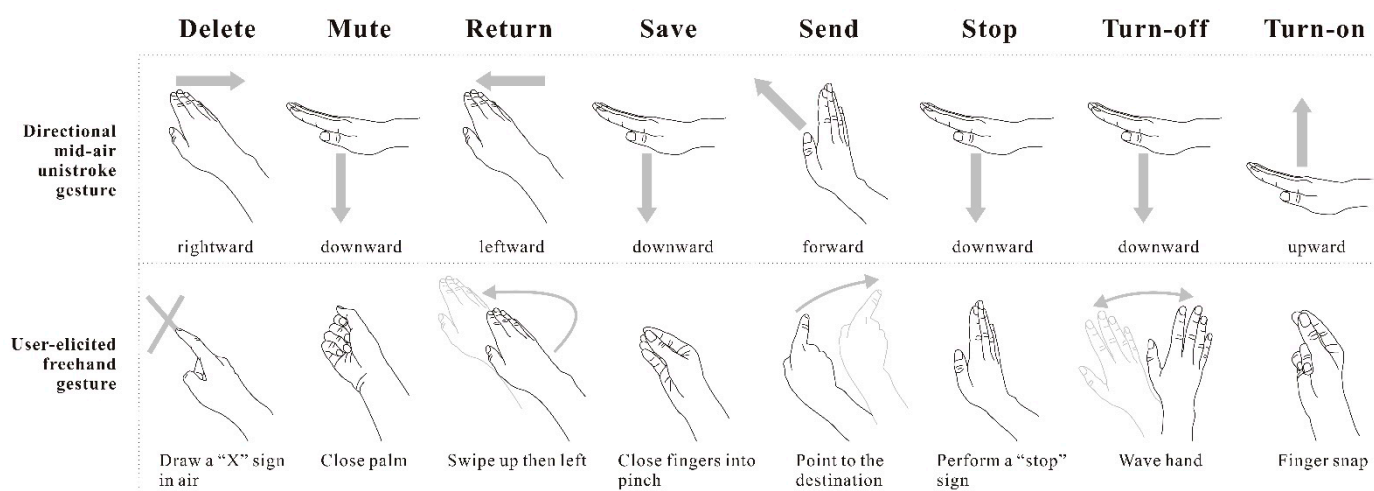


Figure 3. The two gesture sets: directional mid-air unistroke gestures and user-elicited freehand gestures for the eight commands.

The goal of the second phase (i.e., test phase) is to measure the difference in recall efficiency between the mid-air unistroke gestures. Sixteen participants who had learned the gestures for 480 s continued to participate in this phase. They recalled and performed a gesture in a very limited time according to the prompting word (i.e., the command name). The order of performing gestures was counterbalanced (Table 4), and the participant should repeat this order for four times. In summary, the test phase collected 16 participants \times 8 gestures \times 2 types \times 4 repetitions = 1042 trials.

Table 4. The order of trials in test phase. Each letter represents a gesture.

Participant Number	Order of Trials	Participant Number	Order of Trials
1	ABCDEFGHIJKLMNPO	9	IJKLPOMNBADC FEHG
2	EFGHABCDJILKNMPO	10	JILKOPNMF EHG BADC
3	BADC FEHGMNOPIJKL	11	MNOPLKIJC DABHGEF
4	FEHGBADCNMPOJILK	12	NMPOKLJIHG EFC DAB
5	CDAB IJKLEFGHPOMN	13	LKIJC DABOPNMGHFE
6	DCBAJILKPOMNEFGH	14	KLJIDCBAGHF EOPNM
7	HGEFMNOPLKIJABCD	15	POMNHGEFKLJIDCBA
8	GHFENMPOABCDLKIJ	16	OPNMGFHEDCBAKLJI

Two measurements were chosen for the objective evaluation of learnability: recall rate and response time (RT). Recall rate was measured by the number of correct recalls of gestures during each test phase. According to Mihajlov et al. [20], learnability is defined as “the ease with which a person learns to use an interactive system to achieve a goal”. We set different levels of time limit in the first phase as to compare the efficiency of learning gestures. If a gesture is easier to learn, it can be remembered in a shorter time, in comparison with the ones need more learning cost to improve the accuracy of memory. The time cost for responding to and performing a command and subjective user feedback are also effective learnability metrics [64]. Following the experimental setup of Nacenta et al. [65] for evaluating the memorability of user-defined gestures, we chose recall rate and response time as the two measures used in the evaluation experiment. Response time is computed as the period from the moment a prompting word appeared to the moment a correct

gestural response was made to that prompt. It is a metric to measure how well a gesture is remembered as a gesture retained in user's memory can be recalled for use quickly.

The experiment is similar to a Wizard-of-Oz test; and there are some differences. According to the Wizard-of-Oz method, the gestures performed by the participants will not really be recognized by a sensing device or system. Actually the experimenter activates the interaction effect in respond to a participant's behavior to simulate the use scenario. We only asked participants to reproduce the gestures after learning for a period of time, because we mainly focus on how memorable the gestures are rather than the degree to which gestures themselves are easy to be filtered, segmented and recognized.

4.3. Participants

All participants are unimpaired right-handed users (18 males and 30 females) with age range from 18 to 32 ($M = 22.19$, $SD = 3.38$). The age range of these participants is nearly of the same as that in elicitation study because we hope the deviations caused by the age difference can be reduced to a minimum level. Participants are student volunteers from the university. They had no experience of using mid-air interactive systems such as Leap Motion or Kinect and had not been given any prior training in gestural interaction.

4.4. Apparatus

We ran the experiment on the same portable computer as in the elicitation study, showing participants the learning material and the video for test phase on its display. During the experiment process, the ambient lighting and brightness of the display remained unchanged. There were other devices including a SONY high-speed camera utilized to capture the hand movements and a chair. Participants were told to sit on the chair with the distance between their chest and screen being adjusted to 60 ± 5 cm.

4.5. Experimental Materials

The learning material consisted of one video depicting the gesture motions and their mappings to the eight commands while presenting the intended results of the commands, and another one used for simulated interaction. Figure 4a gives an example of the first video for *mute*. In this case, the sound suddenly stopped and a short animation demonstrating the binary state of volume icon was presented to show participants what a visual effect of *mute* is. In the meantime, a video would start playing on the left side, showing how the gesture was performed. The duration of a video clip for each gesture is 4 s, and thus is 64 s in total for 8 gestures \times 2 types = 16 gestures. Figure 4b illustrates the second video which was prepared for participants to learn gestures by simulated interaction. The duration of such a video clip is 6 s, and there were 96 s in total. Each clip showed the command name, the arrow symbol indicating the moving direction of gesture and its intended result. For each participant, the order of presenting gestures in two videos was the same, however, it was randomized across participants within each group as mentioned in Section 4.2. These two videos would be played twice for the group with a learning time of 320 s, and three times for 480 s.

Prompting words used in test phase were also presented by video. A succession of dialogues in a question-and-answer style appeared on the screen as if an online chatting was ongoing, with the prompting word inserted in the answer sentence as the predicate (e.g., I will *mute* my phone) (Figure 5). Each answer described an interaction task that participants must imagine it as the one they wanted to achieve right now. The dialogues were almost identical in terms of the length of sentence. The height of the text was 5 mm. After a 2 s question appearing, the answer sentence lasted 2 s and then disappeared, as shown in Figure 5. A new message would be sent 6 s later, allowing participants to make gesture within this period.

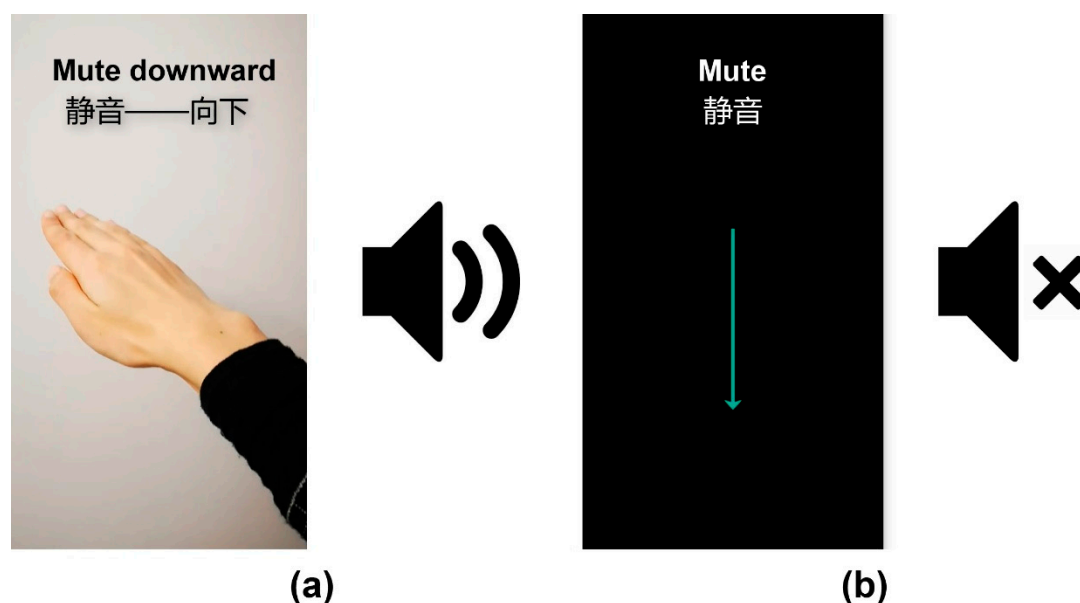


Figure 4. Two videos used for gesture learning: (a) the video depicting gesture motions; (b) the video prepared for simulated interaction.

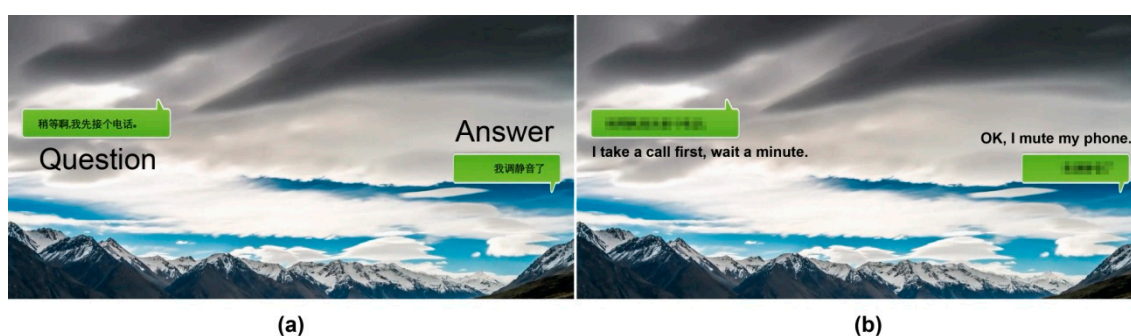


Figure 5. A screenshot of the video for evaluation test: (a) dialogues in a question-and-answer style in which the command name was inserted as the prompting word; (b) the English version of the dialogues.

4.6. Procedure

The main body of the learning phase was remembering the gesture set. At the very beginning, participants went through a video demo in which they experienced a pair of directional unistroke gestures for the commands not included in our research, such as *copy* and *paste*. This was a condensed version of the learning procedure that allowed the participants to understand the experimental requirements. The experimenters then explained all the commands to the participants. When participants confirmed their understanding of the commands, they began learning by following gesture motions shown in the first video. The experimenter would play the second video as soon as the participant performed the gesture in the direction of arrow. This simulation process was set up for them to better memorize the command-to-gesture mappings through body movements.

Participants were asked, in a “Show me the gesture” session, to reproduce gestures immediately after learning. In this session, participants put their both hands on the table with palms facing downward. They lifted the right hand up to reproduce the gesture and simultaneously articulate the word of direction (e.g., downward) upon hearing a referent word (e.g., stop) the experimenter uttered. During this process, participants were not informed of the correctness of gesture in order to avoid possible cross-gesture recall effects [65]. One experimenter sitting aside checked gestures with some noticeable difference and recorded them as incorrect.

After the learning phase, participants completed the subjective ratings of directional mid-air unistroke gestures, while stating how they understood the user-defined mappings observed in the elicitation study. The rating scale was the same as the one used in the elicitation study.

When the rating was finished, participants were encouraged to watch the videos again before entering the test phase. In this phase, they should perform the correct gesture for each prompting word as quickly as possible. An error was recorded when experimenters observed a stable hand or a wrong moving direction, but no feedback was provided for participants. Before performing a gesture, participants were requested to keep their hands on the table, and not to carelessly lift up the hand in advance unless they already had a definite answer. This setting was made to avoid unnecessary errors so that we could count the RT for each trial through extracting the key frames in video recordings. Additionally, note that in this case, it was a bit difficult to distinguish downward motions from upward ones because one needed to raise the hand before performing a downward stroke in this experiment. To solve this problem, we told participants in test phase to accelerate the downward motions with a palm thrust on the one hand, and two experimenters, on the other hand, would have to check the accuracy of every vertical hand motion with the inter-rater reliability no less than 0.61 ($0.61 \leq \kappa \leq 1$).

The process of executing 16 gestures lasted for about 160 s, with a time interval of 30 s between two repetitions. After the test phase, participants were shown the wrong gestures they made and interviewed in detail about the reasons for mistakes as well as their comments on the design of these gestures.

4.7. Data Analysis

The analytic methods applied in this study are non-parametric statistical methods. We employed the 2×2 chi-square test to compute the differences of gestures in recall rate which is a categorical variable (i.e., a gesture is correctly recalled or not). In terms of RT and subjective rating, a Kolmogorov–Smirnov test showed that a considerable part of the data is not normally distributed. Therefore, we used Wilcoxon's signed rank test to calculate the difference between two gestures in these two aspects.

4.8. Results

4.8.1. Recall Rate in the Learning Phase

In Figure 6, sixteen mid-air unistroke gestures were presented in random order of recall rate in the learning phase. For the group with the learning time of 160 s, there are significantly more trials (48 errors) recorded as errors ($\chi^2 = 10.137$, $p = 0.001$) than the group of 320 s (25 errors). Statistics also revealed the difference in recall rate ($\chi^2 = 8.683$, $p < 0.01$) between the group of learning for 320 s and 480 s (9 errors). Overall, a progressive decline of the number of incorrect gestures was observed with the increase of learning time. By comparison, recall rates of freehand gestures for all the groups are higher than unistroke gestures (160 s: 37 errors; 320 s: 15 errors; 480 s: 8 errors). No statistical difference was found between the group with the learning time of 320 s and 480 s ($\chi^2 = 2.341$, $p = 0.126$), suggesting that the user-elicited freehand gestures were well remembered with a less time.

Statistics indicated that the unistroke gestures differed greatly in correct recalls under the same level of learning time. *Mute—downward* and *turn-off—downward* reached a high recall rate only with a learning time of 160 s. By contrast, *send—forward* is the mapping with the lowest recall rate (3/16). Similarly, more than half of the participants failed to reproduce the correct gesture for *delete*. After a 320 seconds' learning, all the gestures had a recall rate more than 2/3, except for *delete—rightward* and *send—downward*. The number of correct gestures for *delete* was significantly lower than *turn-off*, *turn-on*, *mute* and *return* (*delete* vs. *turn-off*: $\chi^2 = 6.788$, $p < 0.01$; *delete* vs. *turn-on*: $\chi^2 = 9.309$, $p < 0.01$; *delete* vs. *mute*: $\chi^2 = 9.309$, $p < 0.01$; *delete* vs. *return*: $\chi^2 = 9.894$, $p < 0.01$). For the group with a 480 seconds' learning, the recall rate of *delete* remained lower than 75%. As for the freehand gestures, they were correctly reproduced with a higher rate than unistroke gestures for all

the commands except for *mute* ($\chi^2 = 9.309, p < 0.01$) *turn-off* and *save* after a 160-seconds' learning. For *delete* and *send*, the difference in recall rate is statistically significant (*delete*: $\chi^2 = 8.533, p < 0.01$; *send*: $\chi^2 = 10.165, p = 0.001$). When the learning time doubled, the number of correct unistroke gestures for *stop*, *delete* and *send* increased, but is still much smaller than that of freehand gestures.

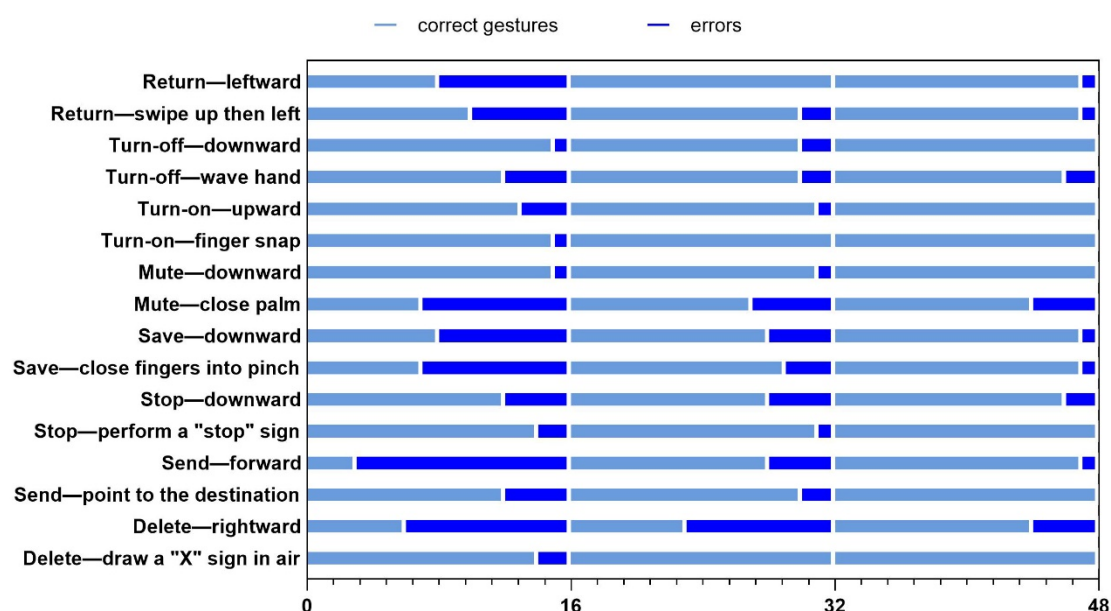


Figure 6. Number of correctly recalled gestures for different levels of learning time (from left to right: 160 s, 320 s, 480 s).

The differences in recall rate between three levels of time limit suggested the learning efficiency of unistroke gestures. A certain period of time was required for some gestures to be remembered, especially for *return—leftward* and *send—forward*. The recall rate of these two gestures after 320 seconds' learning is significantly higher than after 160 s (*return*: $\chi^2 = 8.167, p < 0.01$; *send*: $\chi^2 = 10.165, p = 0.001$). By comparison, the recall rate of *delete—rightward* and *stop—downward* moderately increased with the learning time having been tripled, respectively to 75% and 87.5%. However, 87.5% of the trials of user-elicited freehand gestures for *delete* and *stop* are correct for the group with the learning time of 160 sec. Overall speaking, the recall rate of the user-elicited freehand-gestures for *stop* ($\chi^2 = 4.360, p < 0.05$), *delete* ($\chi^2 = 23.851, p < 0.001$) and *send* ($\chi^2 = 8.000, p < 0.01$) is significantly higher than that of directional unistroke gestures.

We first collected users' statements to explain the significant increase in recall rate for some gestures. *Return-to-leftward* is a metaphor mainly based on the design of timeline, or left arrow. However, mapping the rightward movement to *return* is also reasonable for some participants who are used to scroll pages with multi-touch gestures. This is because the previous page will reappear only when the current window is dragged rightward. In this sense, *return—leftward* is a gesture requiring a certain amount of time to be fully understood by some smartphone users.

Participants in the group of 160 s frequently performed the gesture for *send* as a rightward or upward swipe, as they claimed the file "should be moved out of the device, not pushed into it". Some of them failed to reproduce a gesture. When more learning time was provided, more participants seemed to be able to recognize the link between *send* and a forward stroke gesture. They explained this link using some knowledge of action, such as "stretching one's hand forward to give something to someone", etc.

Most of the participants agreed that *save* enables the association of its semantic meaning with a downward movement. However, this association was also not naturally perceived. For many participants, a click gesture or a long press is more familiar to them for

save. Such an action was executed as a downward press, which may contribute to user's understanding of the *save—downward* mapping.

4.8.2. Recall Accuracy in the Test Phase

In the test phase, participants correctly performed the directional gestures for *mute* and *turn-off* for every trial, as presented in Figure 7. A chi-square test was employed for pairwise comparisons of the correct number of gesture recall between any two of the 16 commands. It demonstrated that *mute—downward*, *turn-on—upward*, *turn-off—downward*, *stop—downward*, *return—leftward* and *save—downward* are not significantly different in recall accuracy with each other (*mute* vs. *send*: $\chi^2 = 12.034$, $p = 0.001$). This result implied a similar level of recall accuracy for these six command-to-gesture mappings. *Send—forward* is at a median level of recall accuracy as indicated by statistical data. *Delete—rightward* is the only mapping with the number of errors ≥ 20 , resulting in a significant difference as compared with *save* ($\chi^2 = 11.184$, $p = 0.001$). The effect of commands on recall accuracy of unistroke gesture is statistically significant (Pearson's chi-square test: $\chi^2 = 73.150$, $df = 7$, $p < 0.001$). This suggested a wide variation of directional mid-air gestures in recall accuracy when the time for reproduction of gesture was limited.

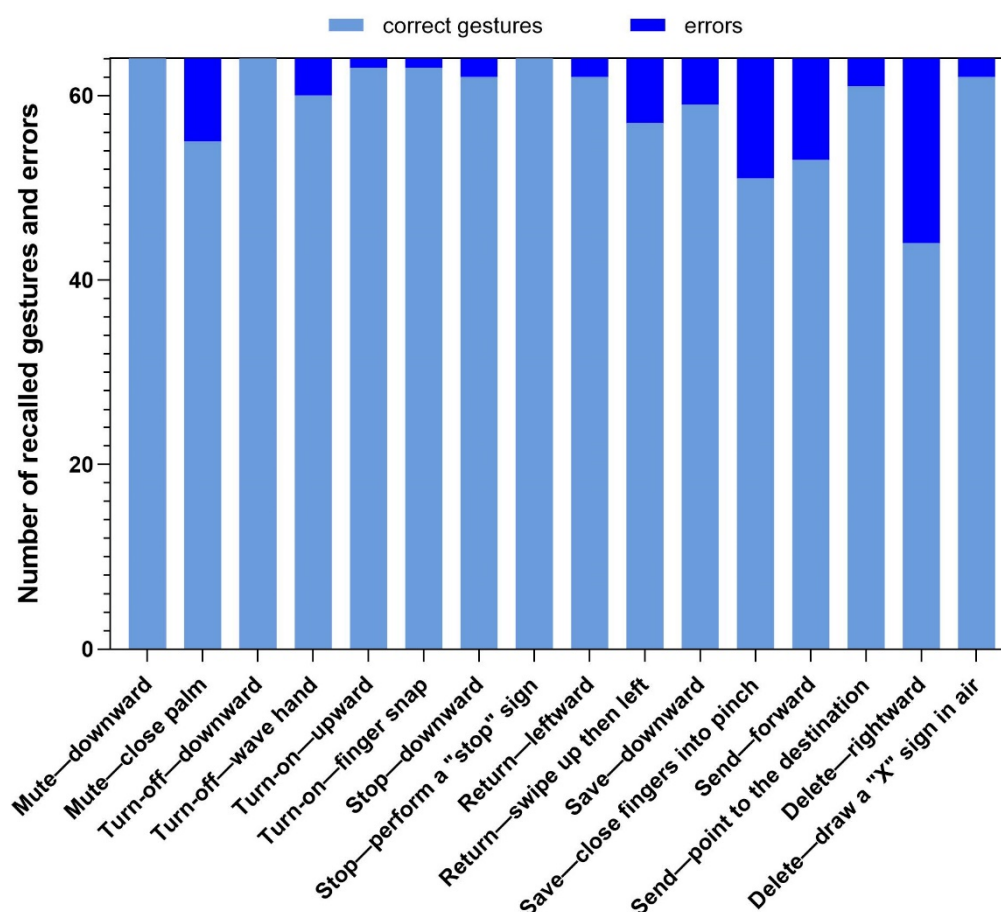


Figure 7. Number of recalled gestures and errors for each command.

In total, the accuracy rate of user-elicited freehand gestures (37 errors) is slightly higher than that of directional unistroke gestures (39 errors). By statistical analysis, a significant higher accuracy rate of the unistroke gestures for *mute* ($\chi^2 = 7.649$, $p < 0.01$) and *save* ($\chi^2 = 4.137$, $p < 0.05$) was detected, and a higher accuracy rate of unistroke gestures for *return* and *turn-off* was also found. For *delete—rightward* and *send—forward*, the accuracy rate is significantly lower than that of *drawing a "X" sign* ($\chi^2 = 17.784$, $p < 0.001$) and *pointing to the destination* ($\chi^2 = 5.133$, $p < 0.05$), respectively.

We reviewed the problems of specific gestures in learnability. *Delete—rightward* was among the lowest recall rate in the gesture set. Some participants complained about the ambiguity of this command, as one participant stated: “The directional gesture for *Delete* depends on what is going to be deleted. If I drop files to the recycle bin, a downward gesture can be a better choice. Only when, for example, I need to remove something like pictures or texts, I agree the appropriateness of a rightward stroke gesture because it resembles the behavior of drawing a horizontal line from left to right to cross out items in paper writing.” Some other participants gave their own reasons for mapping *delete* to leftward or upward gesture.

4.8.3. Response Time

The average response times of the directional unistroke gestures are shown in Figure 8. We found that the average RT for *turn-off—downward* ($M = 1.124$ s, $SD = 0.338$) is the shortest among all the gestures, but not significantly shorter than *mute—downward* ($Z = -0.822$, $p = 0.411$), *stop—downward* ($Z = -0.829$, $p = 0.407$) and *turn-on—upward* ($Z = -0.920$, $p = 0.358$). For *return—leftward*, the average RT is 1.226 s, followed by *save—downward* ($M = 1.284$ s, $SD = 0.385$), *send—forward* ($M = 1.456$ s, $SD = 0.491$) and *delete—rightward* ($M = 1.580$ s, $SD = 0.587$). *Send—forward* is of significantly longer RT than *save—downward* ($Z = -2.234$, $p < 0.05$).

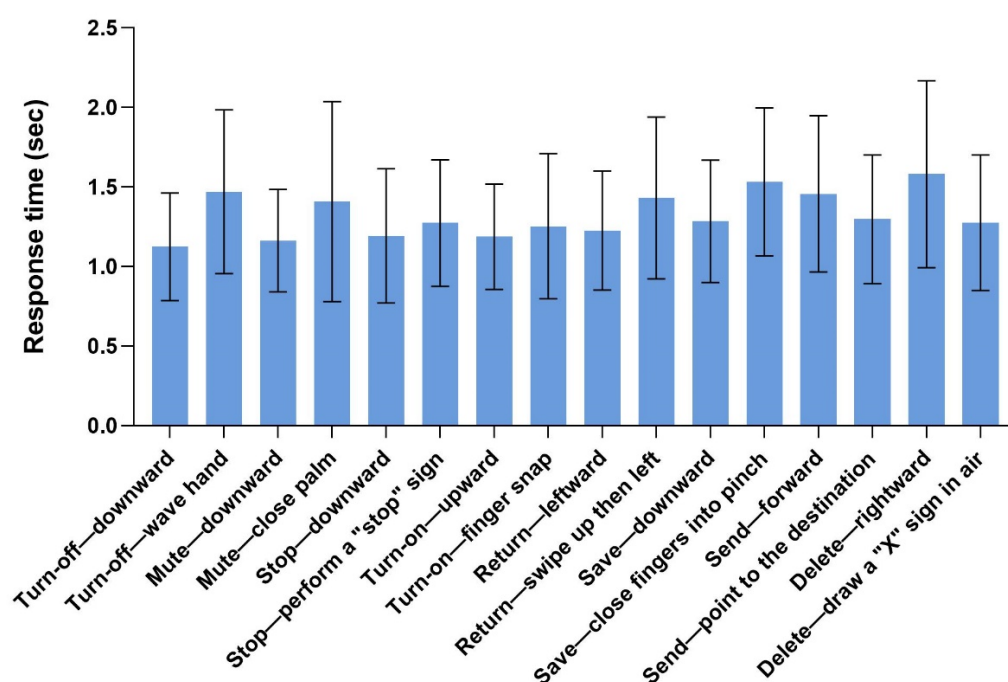


Figure 8. The average response times of command-to-gesture mappings.

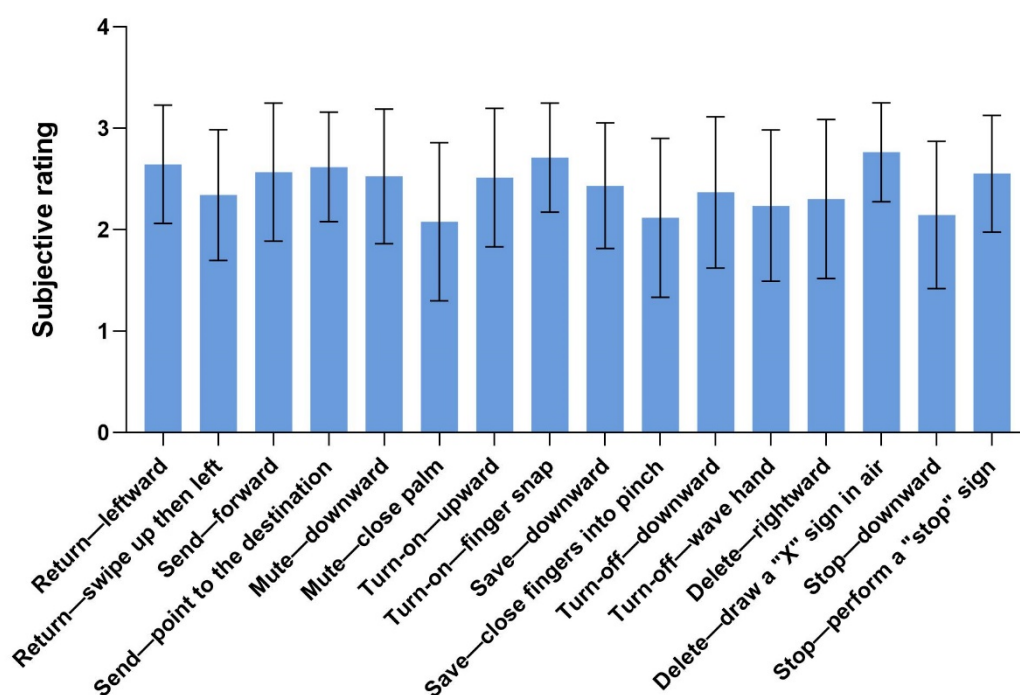
Regarding the RT, there is a statistically significant difference between directional unistroke gestures and freehand gestures for some commands (Table 5). Interestingly, the RT of unistroke gestures for *return*, *save*, *mute*, *stop*, *turn-on* and *turn-off* is shorter than that of freehand gestures. For *delete*, the rightward stroke took significantly more time to recall than performing a “X” sign with one hand. Similarly, the RT for performing *extending one hand forward* for *Send* is significantly longer than *pointing to the destination*. The failure to recall *delete—rightward* and *send—forward* with a less time is also indicative of their lack of agreement on semantic association.

Table 5. Response time of directional unistroke gestures and freehand gestures for all the commands.

Commands	Z	Sig.
Turn-on	−1.262	0.207
Turn-off	−4.150	0.000
Mute	−2.999	0.003
Stop	−1.383	0.167
Return	−3.026	0.002
Save	−3.033	0.002
Send	−2.057	0.040
Delete	−3.294	0.001

4.8.4. Subjective Rating

We found a wide variation ($\chi^2 = 33.758$, $df = 7$, $p < 0.001$) of directional unistroke gestures in terms of the subjective rating with a Friedman's test. Figure 9 presents the average rating scores of the commands. We performed a Wilcoxon's signed rank test in order to compare the commands with each other, and the results are as follows. *Return—leftward* scored higher than all other commands, with no significant differences between it and *send—forward* ($Z = -0.740$, $p = 0.459$), *mute—downward* ($Z = -1.208$, $p = 0.227$) and *turn-on—upward* ($Z = -1.424$, $p = 0.154$). *Stop—downward* received significantly lower ratings than *save—downward* ($Z = -2.639$, $p < 0.01$). Statistics also showed that the subjective rating of the user-elicited freehand gesture for *stop* is significantly higher than that of the directional mid-air unistroke gesture, and for *delete* and *turn-on*, too. By contrast, the directional mid-air unistroke gesture for *mute*, *return* and *save* were significantly more rated as “fit for the command” as compared with the freehand gesture (Table 6).

**Figure 9.** The average subjective ratings of command-to-gesture mappings.

We noticed the inconsistency between measurements of learnability and subjective ratings for *stop—downward*. *Stop—downward* seemed to be an easy-to-learn gesture since it could be correctly remembered and quickly recalled (see Figures 7 and 8), but it was not regarded by participants as a good example of symbolizing *stop*. Two participants supported the mapping of downward motion to *stop*, arguing that this association is explicable because the symbol for the pause key is usually two vertical lines. However,

more participants (19/48) thought that *stop* should not be represented with a directional movement but rather a static posture such as the T gesture for basketball game. Participants also preferred the user-elicited freehand gesture. Nevertheless, some of the participants agreed that a downward movement can denote the meaning of *stop* better than the other directions with a sense of pacifying; and they could easily differentiate it from the other five which have no relation to *stop* at all.

Table 6. Subjective ratings of directional unistroke gestures and freehand gestures for all the commands.

Commands	Z	Sig.
Turn-on	−2.072	0.038
Turn-off	−1.167	0.243
Mute	−3.513	0.000
Stop	−3.475	0.001
Return	−2.891	0.004
Save	−2.699	0.007
Send	−0.546	0.585
Delete	−3.900	0.000

4.9. Discussion of Experiment 2

In the evaluation study, we further investigated the applicability of directional unistroke gestures by assessing the learnability of the selected command-to-gesture mappings. As demonstrated in previous studies [9,58] and our elicitation study, for some verbs there is a tendency for their semantic meaning to be inherently associated with a certain moving direction. However, this does not necessarily mean that using the directional mid-air unistrokes as interaction inputs to express these verbs is a preferable idea. The gestures that are borrowed from object manipulation and universally-accepted symbols [30] could be more intuitive for users than directional movements. As shown in the experimental results, the recall rate of directional unistroke gestures after a 160-seconds' learning time is much lower than that of freehand gestures as a whole. In addition, the elicited freehand gestures for *delete*, *stop* and *send* were performed with a higher accuracy rate than their directional unistroke gestures in the test phase. However, it should be noted that there are significantly fewer errors in performing the unistroke gesture than the freehand gesture for *mute* and *return*. As the freehand gestures are more complex, and some are multi-step, there are more errors for them caused by the non-standard actions. Moreover, six among the eight unistroke gestures can be performed with a less response time over the freehand gestures. Presumably this is because the six unidirectional mid-air gestures are known to users as they are very simple and common movements. Users can recall such a gesture more by "choosing" one from a list of options. These results are the evidences for the claim that directional unistroke gestures are also desirable gestural proposals for some certain commands to be issued.

For the commands with association to a specific direction, the learnability of their corresponding mid-air unistroke gestures greatly varies. Our results showed that it is more effective to express *mute*, *turn-off*, *save* and *return* with directional mid-air unistroke gestures, as they have command-to-gesture mappings which consume less time to learn well, and can be recalled with less time and higher accuracy rate when compared to the other commands.

While many of the participants approved the mapping of *stop* to the downward stroke gesture, they did not think that this mapping was appropriate enough. The frequency that participants rated the downward direction as "fit for" representing *stop* was relatively lower as compared to many other commands in the elicitation study (see Table 3). This means that this command was not considered as being able to evoke a feeling of direction as much as the other ones. However, *Stop—downward* and *Mute—downward* were proved to be learnable gestures. We summarized three experiences that participants used to assign a downward gesture to *mute* from their explanations as below: (1) dragging the volume

bar downwards to the bottom, (2) some actions such as sitting down or bending the body to settle down, and (3) covering up an object which is making sound. There are sign languages and symbolic representations in other user elicitation studies for *mute*, as it is a binary toggle, such as cover the mouth [3], draw a letter “M” or make a clenched fist [6]. To build the mapping for *mute* to a moving direction, users were more likely to activate the experience of taking actions to make something quiet. Although there are differences in the posture, these actions are performed as a from-up-to-down motion.

5. General Discussion

5.1. Suitable Commands for the Mapping to Direction of Unistroke Gesture

Traditionally multiple cognitive strategies, or rather, mental models, were employed in both user-elicited and expert-level design methods to link the mid-air gestures to interaction tasks. We were inspired by previous studies in the cognitive linguistics domain to pay close attention to the phenomenon whereby the activation of spatial representations accompanies verb comprehension. This activation is a process where people retrieve their embodied experience in spatial movements to facilitate the semantic processing of verbs. Consequently, the directionality of spatial movements plays a part more or less in the semantic structuring of verbs and many other concepts. Drawing on this theory, the present study made an attempt at integrating the mental models for interactive gestures by mapping the directions of mid-air movement to the commands. The mappings yield gestures which are represented merely as directional unistrokes.

From the interviews with participants, we concluded two decisive factors for the command to better associate it with a specific directional unistroke gesture. First, the command should permit a manipulative interaction which is visible for users to activate particular spatial experience strongly related to that command. For example, most of the participants think that *return* and *save* are the two commands with stable mapping respectively to leftward and downward movement. Some participants comment that, “to return to the main page, we are used to click the left arrow key, for once or more. Mobile apps and some computer software are designed in this manner” and “when something is saved, it is usually put into a container, or if it is a file or something digital, for me it will be usually moved to the menu bar at the bottom of the screen”. Previous studies [9,58] reported differences in the activation degree of spatial representation between concrete and abstract verbs. This implies that the association of concrete verbs with directional movement are inherently more detectable than abstract ones. There are several abstract verbs in this study which, when used as the interactive command, are usually presented as a button on traditional user interfaces with the working progress invisible to users, for example, *accept*, *reject*, *start* and *screen-capture*. For such commands, in particular *accept* and *start*, the related spatial experience is rather individualized and diversified [6].

Second, in absence of the common prior knowledge of interaction, there should be one widely-accepted semantic interpretation for the command drawing upon the sensorimotor level of knowledge [66] in daily life, such as *stop*. As a counter example, *zoom in* was more accepted to be associated with two moving directions: upward and forward. We debriefed the participants, finding that some deeply ingrained experiences have influences on the semantic relatedness judgment of this command. Participants used the ego-moving schema [49] to imagine a process that the vision is drawing closer to an object as the effect of performing a *splay* gesture on touch-screen. Using this schema, *zoom in* was more likely to be reflected by a from-back-to-forward motion. From an observer viewpoint [35], an increase in the size of digital image goes with the *zoom in* gesture. This increase is pictorially represented as an upward movement which can be explained by the “MORE IS UP” metaphor. In addition to *zoom in*, the high error rate and long RT of *delete* and *start* may be directly caused by the multi interpretations for directional unistroke gesture. To understand the directionality of *start*, participants retrieved spatial knowledge mainly in culture (e.g., the upward direction can be associated to launching in Chinese) or expertise level (e.g., a car moves forward when the driver starts it). The mapping of a command

to directional unistroke gesture can be more learnable only when it meets any of the two conditions.

Although some commands have the semantic meaning that can be associated with one particular moving direction more strongly than the other directions, it does not mean that a mid-air stroke gesture moving in this direction is the best option for learning. For some commands, there are freehand gestures more suitable for memory. These gestures draw on people's "pre-established associations to their personal memories" [65] or legacy bias to express the semantic meaning of the commands with more familiar symbols or sign language to users, such as *delete*, *send* and *stop*. As a result, the pre-defined directional unistroke gestures for these commands require a relatively higher learning cost than the user-preferred gestures, even if these commands are regarded as direction-specific in the semantic meaning. We recommend the future designers to consider this, and not give priority to using directional unistroke gestures for these commands in the case that the situation and recognition system allow the use of complex but more intuitive gestures.

Through the evaluation test a notable difference in gesture learnability was found, indicating that using the direction of mid-air unistroke gesture as a method to construct command-to-gesture mappings is task-dependent [67], and is more suitable for some commands. According to the experimental results, *return—leftward*, *mute—downward*, *turn-off—downward* and *save—downward* are highly acceptable to users and also more learnable and intuitive when compared with the other mappings and the corresponding user-elicited gestures. These mappings have a greater potential to be the interaction primitives for a wide range of human–computer systems.

5.2. Ambiguity of Directional Mid-Air Unistroke Gestures

An inevitable problem will emerge with the application of directional gestures, that is, the same directional in-air movement may correspond to multiple interaction commands as there are only six directions used for signifying many more commands. This problem violates the basic principle of gesture design to some extent, and it is easy to cause the misrecognitions of interactive system. Moreover, some of the results could be partly influenced by these 'double-mappings', as participants would perform better with commands that were mapped to a direction which was only used for one certain command than with commands that were mapped to a direction that was used for multiple commands. The solution to this problem in our study is using the amplitude of motion to distinguish multiple commands associated to the same unistroke gesture. More empirical evidence is needed in the future to reveal the influences of different amplitudes of mid-air unistroke gestures with the same moving direction on the learning of these gestures.

Two other methods are also suggested to disambiguate the unistroke gestures with the same movement amplitude. One is to assign the same mid-air unistroke gesture to the commands that will not be triggered in the same situation. For example, *turn-off* and *save* are used for multimedia control and file management respectively, so they can be both mapped to the long downward stroke. However, for some commands such as *turn-off* and *mute*, which could be used in the same scenario (i.e., playing a multimedia file), it is advisable to specify the object to which a command will be issued before the gesture is performed. In this case, a directional unistroke gesture requires indeed a more complex operation process than using freehand gestures.

5.3. Limitations

We highlight several limitations of the present research invalidating the results. Generally, in order to obtain the user-elicited gesture vocabulary with higher reliability, there should be a better sampling method to cover as many user groups as possible. In this regard, the current research has the following deficiencies. First, a set of unistroke gestures have been examined against the learning capacity of younger technology-naïve users; and yet open questions remain on how elderly people will evaluate these gestures for practical use. Simply focusing on the student participants may underplay the difficulties

which older users will face in natural interaction. Meanwhile, a new concern will be raised as we notice the differences between languages in the way of generating orientational metaphors. Previous studies have demonstrated that the user-defined gesture for certain interactive commands will vary greatly with different cultural backgrounds [30,67]. In this study, however, all the participants are native speakers of Chinese and the commands were described in Chinese. Future studies will seek to compare the cognition and user experience of mid-air gestures between users of different ages and languages. Moreover, a full appreciation of the effect of sample size on the reliability of user-defined gestures is far from complete. According to Choi et al. [68], the inconsistency between two user groups in user-elicited gestures for the same command will decrease as the number of subjects increases. The sample size for our study aim is not large enough to delve further into the minor differences that may exist between user groups. Therefore, further work has to be done on validating the experiment results by incorporating different user communities. At the current stage, the existing findings are what future work can build upon to explore more research issues of mid-air unistroke gestures.

6. Conclusions

We introduced an idea of mapping commonly seen interaction commands to the directional mid-air unistroke gestures as an attempt to avoid the using of different gestures for the same command for different interactive systems. Such a mapping is grounded in the spatial representation that is most popularly and obviously activated in language comprehension. To elicit users' preferences for the mappings, we designed a study to investigate how much users would associate the commands with moving directions of the unistroke gestures. The direction with the highest rating score for a command was of course defined as its stroke gesture. This elicitation study helped in identifying the interaction commands that users more agreed their semantic connection with a particular direction. With this result, the feasibility of our research idea was preliminarily tested. We conducted an evaluation study which provided insights into differences in the learnability of the command-to-gesture mappings, and between the directional mid-air unistroke gestures and user-preferred freehand gestures. Finally, we identified the commands for which the directional mid-air unistroke gestures can work better than user-defined gestures. These commands are "return", "save", "turn-off" and "mute". Our findings offer the possibility of unifying users' cognitions of 3D gestures through a set of systematic and symmetric gesture vocabulary. In future research, the directional mid-air unistroke gesture for more other commands and users' preferences to the related command-to-gesture mappings will be further explored.

Author Contributions: Conceptualization, Y.X.; methodology, Y.X.; software, Y.X.; validation, Y.X., K.M. and C.J.; investigation, Y.X.; data curation, Y.X.; writing—original draft preparation, Y.X.; writing—review and editing, C.J.; visualization, K.M.; supervision, K.M.; funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shanghai Municipal Education Commission [grant numbers 1020309801].

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: The authors would like to thank all participants in this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Coskun, A.; Kaner, G.; Bostan, I. Is smart home a necessity or a fantasy for the mainstream user? A study on users expectations of smart household appliances. *Int. J. Des.* **2017**, *12*, 7–20.
2. Koutsabasis, P.; Vogiatzidakis, P.J.B. Empirical research in mid-air interaction: A systematic review. *Int. J. Hum.-Comput. Interact.* **2019**, *35*, 1–22. [\[CrossRef\]](#)
3. Dong, H.; Danish, A.; Figueroa, N.; El Saddik, A. An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions. *IEEE Access* **2015**, *3*, 543–555. [\[CrossRef\]](#)
4. Zaiți, I.A.; Pentiuc, Ș.; Vatavu, R.D. On free-hand TV control: Experimental results on user-elicited gestures with leap motion. *Pers. Ubiquitous Comput.* **2015**, *19*, 821–838. [\[CrossRef\]](#)
5. Wu, H.Y.; Wang, J.M.; Zhang, X.L. User-centered gesture development in TV viewing environment. *Multimed. Tools Appl.* **2015**, *75*, 733–760. [\[CrossRef\]](#)
6. Chen, Z.; Ma, X.; Peng, Z.; Zhou, Y.; Yao, M.; Ma, Z.; Wang, C.; Gao, Z.; Shen, M. User-defined gestures for gestural interaction: Extending from hands to other body parts. *Int. J. Hum.-Comput. Interact.* **2018**, *34*, 238–250. [\[CrossRef\]](#)
7. Wobbrock, J.O.; Morris, M.R.; Wilson, A.D. User-defined gestures for surface computing. In Proceedings of the CHI'09, Boston, MA, USA, 4 April 2009; pp. 1083–1092. [\[CrossRef\]](#)
8. Zwaan, R.A. Spatial iconicity affects semantic relatedness judgments. *Psychon. Bull. Rev.* **2003**, *10*, 954–958. [\[CrossRef\]](#)
9. Richardson, D.C.; Spivey, M.J.; Barsalou, L.W.; McRae, K. Spatial representations activated during real-time comprehension of verbs. *Cogn. Sci.* **2003**, *27*, 767–780. [\[CrossRef\]](#)
10. Lakoff, G.; Johnson, M. The metaphorical structure of the human conceptual system. *Cogn. Sci.* **1980**, *4*, 195–208. [\[CrossRef\]](#)
11. Vatavu, R.D.; Vogel, D.; Casiez, G.; Grisoni, L. Estimating the perceived difficulty of pen gestures. In Proceeding of the INTERACT'11, Lisbon, Portugal, 5 September 2011; pp. 89–106. [\[CrossRef\]](#)
12. Han, T.; Alexander, J.; Karnik, A.; Irani, P.; Subramanian, S. Kick: Investigating the use of kick gestures for mobile interactions. In Proceedings of the MobileHCI, Stockholm, Sweden, 30 August 2011. [\[CrossRef\]](#)
13. Probst, K.; Lindlbauer, D.; Haller, M.; Schwartz, B.; Schrempf, A. A chair as ubiquitous input device: Exploring semaphoric chair gestures for focused and peripheral interaction. In Proceedings of the CHI'14, Toronto, ON, Canada, 26 April 2014; pp. 4097–4106. [\[CrossRef\]](#)
14. Chattopadhyay, D.; Bolchini, D. Motor-intuitive interactions based on image schemas: Aligning touchless interaction primitives with human sensorimotor abilities. *Interact. Comput.* **2014**, *27*, 327–343. [\[CrossRef\]](#)
15. Hong, J.I.; Landay, J.A. SATIN: A toolkit for informal ink-based applications. In Proceedings of the UIST, San Diego, CA, USA, 5 August 2000. [\[CrossRef\]](#)
16. Appert, C.; Zhai, S. Using strokes as command shortcuts: Cognitive benefits and toolkit support. In Proceedings of the CHI, Boston, MA, USA, 4 April 2009; pp. 2289–2298.
17. Tu, H.; Yang, Q.; Liu, X.; Yuan, J.; Ren, X.; Tian, F. Differences and similarities between dominant and non-dominant thumbs for pointing and gesturing tasks with bimanual tablet gripping interaction. *Interact. Comput.* **2018**, *30*, 243–257. [\[CrossRef\]](#)
18. Leiva, L.A.; Alabau, V.; Romero, V.; Toselli, A.H.; Vidal, E. Context-aware gestures for mixed-initiative text editing UIs. *Interact. Comput.* **2015**, *27*, 675–696. [\[CrossRef\]](#)
19. Bevan, C.; Fraser, D.S. Different strokes for different folks? Revealing the physical characteristics of smartphone users from their swipe gestures. *Int. J. Hum.-Comput. Stud.* **2016**, *88*, 51–61. [\[CrossRef\]](#)
20. Mihajlov, M.; Law, E.L.C.; Springett, M. Intuitive learnability of touch gestures for technology-naïve older adults. *Interact. Comput.* **2015**, *27*, 344–356. [\[CrossRef\]](#)
21. Sziladi, G.; Ujbanyi, T.; Katona, J.; Kovari, A. The analysis of hand gesture based cursor position control during solve an IT related task. In Proceedings of the 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary, 11 September 2017.
22. Chen, C.; Wang, J. The semantic meaning of hand shapes and Z-dimension movements of freehand distal pointing on large displays. *Symmetry* **2020**, *12*, 329. [\[CrossRef\]](#)
23. Katona, J.; Peter, D.; Ujbanyi, T.; Kovari, A. Control of incoming calls by a Windows Phone based brain computer interface. In Proceedings of the 15th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary, 19 November 2014; pp. 121–125.
24. Guy, E.; Punpongsonon, P.; Iwai, D.; Sato, K.; Boubekeur, T. LazyNav: 3D ground navigation with non-critical body parts. In Proceedings of the 3D User Interfaces (3DUI), 2015 IEEE Symposium, Arles, France, 23–24 March 2015; pp. 43–50.
25. Chan, E.; Seyed, T.; Stuerzlinger, W.; Yang, X.D.; Maurer, F. User-elicitation on single-hand microgestures. In Proceedings of the SIGCHI Conference on Human Factors in Computer Systems, San Jose, CA, USA, 7–12 May 2016; pp. 3403–3414.
26. Burnett, G.; Crundall, E.; Large, D.; Lawson, G.; Skrypchuk, L. A study of unidirectional swipe gestures on in-vehicle touch screens. In Proceedings of the AutomotiveUI'13, Eindhoven, The Netherlands, 28 October 2013; pp. 22–29. [\[CrossRef\]](#)
27. Jahani-Fariman, H. Developing a user-defined interface for in-vehicle mid-air gestural interactions. In Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion (IUI'17 Companion), New York, NY, USA, 7 March 2017; pp. 165–168.
28. May, K.; Gable, T.M.; Walker, B.N. Designing an In-Vehicle Air Gesture Set Using Elicitation Methods. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; pp. 74–83.

29. Bostan, I.; Buruk, O.T.; Canat, M.; Tezcan, M.O.; Yurdakul, C.; Göksun, T.; Özcan, O. Hands as a controller: User preferences for hand specific On-Skin gestures. In Proceedings of the 2017 Conference on Designing Interactive Systems, Edinburgh, UK, 10–14 June 2017; pp. 1123–1134.
30. Wu, H.; Gai, J.; Wang, Y.; Liu, J.; Qiu, J.; Wang, J.; Zhang, X.L. Influence of cultural factors on freehand gesture design. *Int. J. Hum.-Comput. Stud.* **2020**, *143*, 102502. [\[CrossRef\]](#)
31. Dim, N.K.; Silpasuwanchai, C.; Sarcar, S.; Ren, X. Designing mid-air TV gestures for blind people using user- and choice-based elicitation approaches. In Proceedings of the DIS'ed16, New York, NY, USA, 4 June 2016; pp. 204–214. [\[CrossRef\]](#)
32. Vogiatzidakis, P.; Koutsabasis, P. Frame-based elicitation of mid-air Gestures for a smart home device ecosystem. In *Informatics; Multidisciplinary Digital Publishing Institute*: Basel, Switzerland, 2019; Volume 6.
33. Vatavu, R.D.; Zaiti, I.A. Leap gestures for TV: Insights from an elicitation study. In Proceedings of the TVX'14, New York, NY, USA, 25 June 2014; pp. 131–138.
34. Aigner, R.; Haller, M.; Lindlbauer, D.; Ion, A.; Zhao, S.; Koh, J. Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for HCI. *Microsoft Res. Tech. Rep. MSR-TR-2012-111* **2012**, *2*, 30.
35. McNeill, D. *Gesture and Thought*; University of Chicago Press: Chicago, IL, USA, 2008; pp. 34–41. [\[CrossRef\]](#)
36. Vatavu, R.D. Smart-Pockets: Body-deictic gestures for fast access to personal data during ambient interactions. *Int. J. Hum.-Comput. Stud.* **2017**, *103*, 1–21. [\[CrossRef\]](#)
37. Loehmann, S. Experience Prototyping for Automotive Applications. Ph.D. Dissertation, LMU München, Fakultät für Mathematik, Informatik und Statistik, 2015.
38. Xiao, Y.; He, R. The intuitive grasp interface: Design and evaluation of micro-gestures on the steering wheel for driving scenario. *Univers. Access Inf. Soc.* **2020**, *19*, 433–450. [\[CrossRef\]](#)
39. Cienki, A.; Müller, C. Metaphor, gesture and thought. In *Cambridge Handbook of Metaphor and Thought*; Gibbs, R.W., Ed.; APA: Worcester, MN, USA, 2008; pp. 483–501. [\[CrossRef\]](#)
40. Bacim, F.; Nabiyouni, M.; Bowman, D.A. Slice-n-Swipe: A free-hand gesture user interface for 3D point cloud annotation. In Proceedings of the 2014 IEEE Symposium on 3D User Interfaces, Minnesota, MN, USA, 29–30 March 2014; pp. 185–186. [\[CrossRef\]](#)
41. Ackad, C.; Kay, J.; Tomitsch, M. Towards learnable gestures for exploring hierarchical information spaces at a large public display. In Proceedings of the Gesture-based Interaction Design: Communication and Cognition, Toronto, ON, Canada, 26 April 2014; Volume 49, pp. 16–19.
42. Lakoff, G.; Johnson, M. *Philosophy in the Flesh—The Embodied Mind and Its Challenge to Western Thought*; Basic Books: New York, NY, USA, 1999.
43. Hurtienne, J.; Stöfel, C.; Sturm, C.; Maus, A.; Rötting, M.; Langdon, P.; Clarkson, J. Physical gestures for abstract concepts: Inclusive design with primary metaphors. *Interact. Comput.* **2010**, *22*, 475–484. [\[CrossRef\]](#)
44. Huang, Y.; Tse, C.S. Re-examining the automaticity and directionality of the activation of the spatial-valence Good is Up metaphoric association. *PLoS ONE* **2015**, *10*, e0123371. [\[CrossRef\]](#)
45. Kóczy, J.B. Orientational metaphors. In *Nature, Metaphor, Culture: Cultural Conceptualizations in Hungarian Folksongs*; Springer: Singapore, 2018; pp. 115–132.
46. Lakoff, G. The neural theory of metaphor. In *The Metaphor Handbook*; Cambridge University Press: Cambridge, UK, 2009.
47. Grady, R. Foundations of Meaning: Primary Metaphors and Primary Scenes. Ph.D. Dissertation, University of California at Berkeley, Berkeley, CA, USA, 1997.
48. Gu, Y.; Mol, L.; Hoetjes, M.; Swerts, M. Conceptual and lexical effects on gestures: The case of vertical spatial metaphors for time in Chinese. *Lang. Cogn. Neurosci.* **2017**, *32*, 1048–1063. [\[CrossRef\]](#)
49. Boroditsky, L. Metaphoric structuring: Understanding time through spatial metaphors. *Cognition* **2000**, *75*, 1–28. [\[CrossRef\]](#)
50. Boroditsky, L.; Ramscar, M. The roles of body and mind in abstract thought. *Psychol. Sci.* **2002**, *3*, 185–189. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Casasanto, D.; Bottini, R. Spatial language and abstract concepts. *WIREs Cogn. Sci.* **2014**, *5*, 139–149. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Narayanan, S. Embodiment in Language Understanding: Modeling The Semantics Of Causal Narratives. Ph.D. Dissertation, University of California at Berkeley, Berkeley, CA, USA, 1997.
53. Zwaan, R.A. The immersed experiencer: Toward an embodied theory of language comprehension. *Psychol. Learn. Motiv.* **2004**, *44*, 35–62. [\[CrossRef\]](#)
54. Barsalou, L.W. Language comprehension: Archival memory or preparation for situated action? *Discourse Process.* **1999**, *28*, 61–80. [\[CrossRef\]](#)
55. Zwaan, R.A.; Madden, C.J.; Yaxley, R.H.; Aveyard, M.E. Moving words: Dynamic representations in language comprehension. *Cogn. Sci.* **2004**, *28*, 611–619. [\[CrossRef\]](#)
56. Hayward, W.G.; Tarr, M.J. Spatial language and spatial representation. *Cognition* **1995**, *55*, 39–84. [\[CrossRef\]](#)
57. Stanfield, R.A.; Zwaan, R.A. The effect of implied orientation derived from verbal context on picture recognition. *Psychol. Sci.* **2001**, *12*, 153–156. [\[CrossRef\]](#)
58. Richardson, D.C.; Spivey, M.J.; Edelman, S.; Naples, A.J. Language is spatial: Experimental evidence for image schemas of concrete and abstract verbs. In Proceedings of the 23rd Annual Meeting of the Cognitive Science Society, Mahwah, NJ, USA, 26–29 July 2001; pp. 873–878.
59. Chatterjee, A.; Southwood, M.H.; Basilico, D. Verbs, events and spatial representations. *Neuropsychologia* **1999**, *37*, 395–402. [\[CrossRef\]](#)
60. Bergen, B.K.; Lindsay, S.; Matlock, T.; Narayanan, S. Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cogn. Sci.* **2007**, *31*, 733–764. [\[CrossRef\]](#) [\[PubMed\]](#)

-
61. Wu, L.M.; Mo, L.; Wang, R.M. The activation process of spatial representations during real-time comprehension of verbs. *Acta Psychol. Sin.* **2006**, *38*, 663–671.
 62. Vatavu, R.D.; Wobbrock, J.O. Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit. In Proceedings of the CHI' 15, Seoul, Korea, 18 April 2015; pp. 1325–1334.
 63. Maouene, J.; Hidaka, S.; Smith, L.B. Body parts and early-learned verbs. *Cogn. Sci.* **2008**, *32*, 1200–1216. [[CrossRef](#)]
 64. Grossman, T.; Fitzmaurice, G.W.; Attar, R. A survey of software learnability: Metrics, Methodologies and Guidelines. In Proceedings of the CHI'09, Boston, MA, USA, 4 April 2009; pp. 649–658.
 65. Nacenta, M.A.; Kamber, Y.; Qiang, Y.; Kristensson, P.O. Memorability of pre-designed & user-defined gesture sets. In Proceedings of the CHI'13, Paris, France, 27 April 2013; pp. 1099–1108.
 66. Blackler, A.L.; Hurtienne, J. Towards a unified view of intuitive interaction: Definitions, models and tools across the world. *MMI-Interakt.* **2007**, *13*, 36–54.
 67. Wu, H.; Zhang, S.; Liu, J.; Qiu, J.; Zhang, X. The gesture disagreement problem in free-hand gesture interaction. *Int. J. Hum.-Comput. Interact.* **2018**, *35*, 1102–1104. [[CrossRef](#)]
 68. Choi, E.; Kwon, S.; Lee, D.; Lee, H.; Chung, M.K. Can user-derived gesture be considered as the best gesture for a command? Focusing on the commands for smart home system. In Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting, Boston, MA, USA, 1 September 2012; pp. 1253–1257.