HKUST SPD - INSTITUTIONAL REPOSITORY

Title	Tripartite Effects: Exploring Users' Mental Model of Mobile Gestures under the Influence of Operation, Handheld Posture, and Interaction Space
Authors	Zhu, Kening; Ma, Xiaojuan; Chen, Haoyuan; Liang, Miaoyin
Source	International Journal of Human-Computer Interaction, v. 33, (6), June 2017, p. 443-459
Version	Accepted Version
DOI	10.1080/10447318.2016.1275432
Publisher	Taylor & Francis Group
Copyright	© 2017 Taylor & Francis Group, LLC
License	CC BY-NC-ND

This version is available at HKUST SPD - Institutional Repository (<https://repository.ust.hk>)

If it is the author's pre-published version, changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published version.

Tripartite Effects: Exploring Users' Mental Model of Mobile Gestures Under the Influence of Operation, Handheld Posture, and Interaction Space

In this research, we conducted an elicitation study with 54 participants to evaluate user-defined mobile gestures given three types of operations, handheld postures, and interaction spaces, respectively, to investigate the effects of these factors on mental model. The results of the elicitation study revealed that each of the three factors significantly affected the users' mental models, which could be reflected by the characteristics (e.g., nature, number of steps, and spatial utility) of the user-defined mobile gestures. In addition, we found that users tended to assign different roles to mid-air and on-screen gestures based on their motor ability given a handheld gesture and their interpretation of the characteristics and requirements of an operation. We further derived a set of gestures commonly defined in the elicitation study for a list of representative operations, and share our insights into users' mental models and the implications on future mobile-gesture design in different contexts.

Keywords: mental model; interaction space; operation; handheld posture; mobile; gesture.

Introduction

As smartphones become more and more prevalent in our daily life, the technology of multi-touch sensing has been popularized, enabling users to control these mobile devices through gestures. Some common mobile gestures include pinching to zoom and swiping to navigate. A significant amount of research efforts have been focused on gesture-based interaction on mobile devices, and their results suggest the advantages of gesture over conventional GUI (Graphical User Interface)-based interaction (Fuccella, 2013). Thus, understanding user mental model on gesture-based interaction becomes essential for designers and researchers. We define the users' mental model as the ways in which users understand the process of certain operations in a system (Allen, 1997), which can be affected by their "*prior knowledge, individual abilities, and different*

interpretations about the purpose and functions.” In context of mobile text processing, the user’s mental model on the gesture interaction could be affected by their prior knowledge on different interaction spaces (on-screen and mid-air), their interactive abilities with different handheld postures (interacting using one thumb, interacting using one index finger, and interacting using two thumbs), and their interpretations about the functional operations (caret positioning, text editing, and input configuration). Therefore, we chose these three factors as the independent variables to investigate how these three factors (i.e. operations, handheld postures, and operations) affect user’s mental model on mobile gesture input for text processing.

While a variety of gestures can be mapped to distinct operations on a mobile phone, research (Wobbrock, 2009; Ruiz, 2011; Tung, 2015; Chan, 2016; Dim, 2016; Shimon, 2016; Arefin Shimon, 2016) has shown that the difference among the characteristics of various mobile operations could trigger different mental models or perceptions of users and further affect the design of gesture-based interaction for mobile platform. In addition, research (Wobbrock, 2008; Azenkot, 2012; Ng, 2014) showed that users performed differently with varied handheld postures, including one-handed and two-handed. An assortment of design implications have been developed by taking into account the usage of different handheld postures for different occasions. For example, vertical correction for the touch point should be added according to different handheld postures to minimize the error rate of typing (Azenkot, 2012). Thus, designers need to understand how users perceive the interactions under different contexts of task characteristics and holding postures in order to provide well-designed gestures for mobile interaction.

The recent development of new gesture-sensing technologies has expanded human-computer interaction to a new design space and paradigm: mid-air-gesture-based

interaction (Aigner, 2012), which moves the conventional surface-based 2D gesture space to the 3D space in the air, thus increasing the freedom and the expressiveness of the gestures. Mid-air gesture interaction has been widely applied in the context of desktop display, tabletop display, and large public display, and has caught the attention of the mobile industry. For example, Samsung has released its Galaxy S4 smartphone featuring palm-based Air Gesture for phone functions. The in-built motion sensors and cameras in smartphones support motion-based gestures and free-hand gestures around the phones. In addition, merging finger-sensing technology (e.g., Leap Motion (Leap)) makes fine-grained control possible with mid-air gestures, such as selecting small objects. There has been some research focusing on designing new mid-air interaction along with on-screen touch interaction for mobile phones through introducing external sensors (Chen, 2014) and wearable devices (Ni, 2011). In addition, the benefit of user-defined gestures on learnability and usability motivated researchers to investigate the user-defined mid-air gestures on mobile phones (Ruiz, 2011). While the existing elicitation studies analysed the variation of user-gestures in a fixed interaction space (either on screen or mid-air) for different operations, leaving handheld postures free for the user's choice, how these three factors affect users' mental model on gesture-based mobile interaction remains unexplored.

To address this problem, we conducted an elicitation study (Wobbrock, 2009) to investigate user-defined mobile gestures given three types of operations, handheld postures, and interaction spaces, respectively. As texting is one of the most essential activities in mobile phone usage, we were particularly interested in text-related operations, including caret positioning, changing input methods, content selection, clipboard control, etc. Three sub-categories (caret positioning, input configuration, and text editing) were chosen within the text-related tasks; one-thumb, two-thumbs, and

index-finger were chosen as the contexts for handheld postures; while the interaction spaces consisted of mid-air, on-screen, and the combination of mid-air and on-screen. We proposed a taxonomy to characterize the mobile gestures collected in the elicitation studies.

The elicitation study used a between-group design with 54 participants. That is, to investigate the effect of individual factors on mental model, we assigned 18 people to every condition in each dimension. The results revealed that each of the three factors significantly affected the nature, number of steps, and spatial utility of the user-defined mobile gestures. In addition, we found that users tended to assign different roles to mid-air and on-screen gestures based on their motor ability given a handheld gesture and their interpretation of the characteristics and requirements of an operation. Furthermore, we derived a set of gestures commonly defined in the elicitation study for a list of representative operations. We offer our insights into users' mental models and the implications on future mobile gesture design in different contexts.

The contribution of this research lies three fold: 1) quantitative and qualitative investigation on the potential impact of operation characteristics, hand posture capability, and the interaction space on users' perception and usage of mobile gestural interaction via a set of measurements including a taxonomy; 2) a list of user-defined gestures for a set of text-related tasks in different combinations of handheld postures and interaction spaces; 3) insights into users' mental model and design implications for mobile gesture interaction.

Related Work

Our work was inspired by previous researches on gesture-based interaction, with the particular interest in the unique form factor of the smartphone. We were also motivated by the existing works on investigating user performance with different handheld

postures and interaction spaces, and previous elicitation-based research on mobile platform.

Background of gesture-based interaction: from large display to mobile phone

Many research efforts have been devoted to characterizing gestures on interactive tabletop surfaces, both for collaborative and individual interaction. For example, Morris et al. examined the design space of cooperative gestures for art and photo manipulation on tabletop by multiple users to investigate the usability and the acceptability of cooperative gestures (Morris, 2006). Later, Wobbrock et al. (2009) introduced a taxonomy for the classification of more general surface gestures based on quantitative and qualitative data collected from an elicitation study. They had users, instead of designers, create gestures for different surface operations and proved that gestures generated from such processes were easier for end users to learn (Wobbrock, 2005).

While these early studies on user-defined gestures mainly focused large surfaces, the mobile phone has its own unique form factor which could affect users' perception. Specifically, the relatively small screen displays limited scale of contents and can afford fewer on-screen touch points than a large surface (Romano, 2014). However, the interaction space for mobile devices can be extended to the peripheral area around the devices (Freeman, 2014). In addition, users can hold the mobile devices in different postures which could affect the users' interaction capabilities. Also, mobile devices can vary in shapes, such as smartphones, wristbands, smartwatches, and smart glasses, etc. Elicitation-based research (Ruiz, 2011; Park, 2014; Tung, 2015; Chan, 2016; Dim, 2016; Shimon, 2016; Arefin Shimon, 2016) has been conducted for these devices demonstrating that user-defined gestures were significantly varied from those elicited for large touch surface.

The existing elicitation-based studies on mobile gesture interaction motivated us to examine how different combination of mobile operations, handheld postures, and interaction spaces affect users' perception on mobile interaction. In the following discussion, we will review the existing work on investigating the influence of different mobile operations, handheld postures, and interaction spaces on gesture-based mobile interaction.

Tasks affect users' interaction in mobile platform

There have been some studies indicating that user performance varied significantly among different operations on mobile platform due to the difference on interaction complexity, semantic meaning, and the form of the target object. Firstly, the effects of target size and shape in human-computer interaction have been extensively studied in light of Fitts' law (Mackenzie, 1992). In addition, the interaction complexity of the operations could affect the users' performance heavily. Fuccella et al. (2013) compared users' performance in a set of text-editing operations with gesture-based and widget-based inputs. While their study mainly focused on the comparison of the two different input methods in operation completion time, it can be inferred from the results that the completion time varies significantly across different operations within the same input method. More specifically, operations on inserting and deleting content required shorter time than caret movement and clipboard manipulation (e.g., copy, cut). Markussen et al. (2014) designed mid-air gestures for text-related actions (match, select, undo, and delete) in their mid-air word-gesture keyboard, Vulture. User studies suggested that the operations significantly affected the user performance in terms of correctness and speed.

As discussed earlier, the difference among mobile operations significantly affected the user intuition and agreement on gesture, which could be resulted from different mental models mapped to different operations. For example, users tended to

define gestures in opposite directions for semantically dichotomous operations, even if they did not appear consecutively. Users also often made simplifying assumptions for semantically similar operations, such as “enlarge” and “zoom in”, “pan” and “move”.

Handheld postures affect users’ interaction in mobile platform

Users have different preferences in hand posture for mobile interaction and may change hand posture depending on the use situation. Karlson, Bederson and Contreras-Vidal (2006) studied one-handed interaction by observing mobile usage in the airport to examine user behaviour with mobile devices. Their observations suggested that 60% of users engaged with their mobile phone in a one-handed posture when walking. Users were also seen to carry objects during input but encumbrance was not examined specifically in their study. More recently, an investigation by Hooper (2013) – who made over 1000 observations of users in different public settings, including bus stops and cafes – found that 49% of people used their mobile phone in the one-handed preferred thumb posture, while 36% held the device in the two-handed index finger position to input. The remaining 15% of users held the device in the two-handed both thumbs posture.

Other research has shown that users’ touch patterns and performances may vary according to hand postures. Azenkot and Zhai (2012) investigated touch-screen-based typing behaviors in different handheld postures (one thumb, index finger, and two thumbs), and their results showed that the handheld postures affected the speed and the accuracy of typing, significantly. Wobbrock et al. (2008) compared one-handed and two-handed interaction at the front and the back of a mobile device and found that users performed faster and more accurately in stroke drawing using the index finger in the two-handed posture.

Although several studies have investigated the comparison of different handheld postures affecting input performance, there is still no research on how the unique characteristics of each posture (e.g., accuracy, degree of freedom, reach-capability, etc.) may affect users' perception on defining mobile gestures. Understanding this component reveals the user's mental models of mobile gesture interaction under different postures.

Interaction spaces affect users' interaction

The development of mid-air-gesture sensing technology attracted researchers to study interactions occurring on and above the surface of the mobile screen. Marquardt et al. (2011) broadened the design space of tabletop interaction by unifying on- and above-surface gestures in a so-called continuous interaction space. Besides interactive tabletop, mid-air gestural controls have been widely used in large displays, such as interactive TV (Vatavu, 2012) and public billboard (Shoemaker, 2009). Bruder et al. (2013) compared 2D touch and 3D mid-air interactions in object selection in an augmented-reality tabletop system, and the results showed that 2D touch techniques were more efficient and accurate close to the screen but 3D mid-air gestures outperformed 2D touch for targets far away from the screen. Jakobsen et al. (2015) found that mid-air interaction was slower than touch for object acquisition for wall-based display. However, on-screen touch interaction often suffers from the fat-finger problem, while mid-air gesture can reduce the occlusion of the content.

Garcia-Lopez et al. (2015) found that users preferred zoom-in/out method in non-touch interaction. Yet, designing mid-air gesture remains difficult due to the lack of direct physical feedback, which could cause fatigue. Wacharamanotham et al. (2014) proposed a technique for desktop devices that reliably distinguishes mid-air clicking signal from homing movements based on the shapes drawn by the user. Chen et al.

(2014) proposed the combination of air and touch gestures through implementing the recognition of a set of user-defined combined gestures on mobile platform.

Although these studies have compared the usage of mid-air and on-screen interactions and some preliminary suggestions have been provided to combine these two, how users may perceive these two spaces for mobile gesture interaction remains unclear. Furthermore, there is a need for an in-depth investigation on how the interaction spaces, along with the operation variation and the handheld postures, may affect users' mental model. In our research, we conducted an elicitation study to investigate how mobile operations, handheld postures, and interaction spaces could affect user mental model of mobile gestural interaction and derived implications for future design of mobile gestural interaction.

Elicitation Study Design

In this study, we were particularly interested in investigating the research question on: How do tasks, spaces, and postures, synergistically affect the characteristics of the user-defined gesture?

For the design of the elicitation study, as summarized in Table 7, we focused on three independent factors (operation, handheld posture, and interaction space) and investigated their effects on the user-defined gestural interaction on mobile phones. We selected three different categories of text-related operations (caret positioning, input configuration, and text editing), three different handheld postures (one-hand holding with one-thumb operating, one-hand holding with one-index-finger operation, and two-hand holding with two-thumbs operating), and three different interaction spaces (on-screen, mid-air, and combined). Each workshop participant performed 30 gestures for 30 operations in one unique combination of interaction spaces and handheld postures. The following sub-sessions will describe the study design in detail, including the

selection and categorization of the operations, definitions of the interaction spaces, definitions of the handheld postures, the estimation of the sample size, the profiles of the participants, the study apparatus, the procedure, and the measured dependent variables.

Selection and categorization of operations

Text processing (e.g., caret positioning, changing input methods, content selection, clipboard control, etc.) is one of the most essential and commonly used features on mobile phones, and it is involved in various operations, such as email, SMS, instant messaging, social network, etc. In our study, we modified and expanded the task list in Fuccella et al. (2013) and categorized mobile text-related activities into three task, based on the targets of control: caret positioning (controlling the position of the caret), text editing (controlling the text contents), and input configuration (controlling the input modes). We further specified 10 operations within each category via iterative discussions (Table 1): caret positioning included operations that specified the position of inputting text in the editor; input configuration included switching among different input modes (letter, emoticon, punctuation, shift, capitalization) and browsing the input options; and text editing included selecting content, deleting content, and clipboard control, such as copy, cut, and paste. We did not include the operation of typing characters because its performance could be easily affected by the layout of the keyboard, the input content, and the input method, thus it was out of the focus of this research. In addition, we excluded operations that were application-specific, such as adding an attachment in email, or combined operations that are often done automatically (e.g., capitalization is often automatically cancelled after clicking on a letter key). Table 1 shows the complete list of the 30 operations selected for the elicitation study.

As a user's mental model can be affected by the characteristics of the operations (e.g., the textual interpretation, the requirement, the objects to manipulate, etc.) (Allen, 1997), the three types of operations can be further categorized in terms of the conceptual complexity (Wobbrock, 2009), the semantic imageability (Cortese, 1997), and the form of the manipulated object. Table 2 provides detailed definitions of these characteristics.

Prior to the elicitation study, the conceptual complexity and the semantic imageability were rated by a group of users who did not participate the elicitation study but who had similar language ability. The form of the targeted object for each operation was estimated by the authors independently based on the existing literature (Roberts, 1980; Azenkot, 2012; Fuccella, 2013). These measured properties of the operations were used to interpret the significant statistical effect of the operations on the users' mental model in the elicitation study.

Semantic imageability

The semantic imageability (a.k.a. imagery) refers to the ease of evoking a non-verbal sensory representation based on a given word/phrase, involving "*the analysis of scenes and the generation of mental images*" (Paivio, 1986). Overall, imageability depicts the psychological effect of a linguistic unit and reflects the cognitive process of how meanings rely on information mediated by the senses (Paivio, Yuille and Madigan, 1968; Toggia and Battig, 1978; Gilhooly and Logie, 1980). For example, the word "apple" can easily evoke the image of the fruit, and the word "start" can evoke the images of flagging or the symbol of playing video; however, it takes longer time to evoke a mental image of words like "reason", "relevant", "decision", etc. The semantic imageability of each referent was rated on a 7-point Likert scale (1 = highly difficult to evoke a mental image, 7 = highly easy to evoke a mental image) by 21 mobile users

who did not participated the elicitation workshop but who had similar language background as the elicitation-study participants. In this test, we required the users to attempt to access mental images that were not related to the computing platform.

Figure 1 illustrates the average imageability rating for each operation, showing the potential relationship between mental images of the operations based on prior sensory experiences and how they are realized in the gesture space. ANOVA analysis was applied to test the relationship between operation and imageability rating ($F(29, 600) = 4.72, p < 0.05$). The results showed that text editing operations made it easier to evoke mental images than the other two group of operations (caret movement and input configuration), suggesting that operations with low imageability are either very abstract or they rarely occur in real life besides in computing usage. For example, all users put scissors as the mental image for the *Cut* operation, but the mental images evoked for the *Copy* operation included pen and paper, twins, and copy machine, etc.; and some users evoked a red cursor and a person embracing for selecting text content.

Conceptual complexity

Adopted from Wobbrock et al. (2009), *conceptual complexity* measures the magnitude to which a matter or concept is difficult to comprehend. Following the similar evaluation on imageability, we evaluated the conceptual complexity of each operation with another 20 participants of a similar background to the previous participants. The conceptual complexity of caret movement was significantly higher than text editing and input configuration (ANOVA: $F(29, 598) = 4.69, p < 0.05$). The average conceptual complexity of each operation is shown in Figure 2. The results suggest that how well users can conceptually process an operation may affect their abilities to express and execute it. Comments from the participants revealed the potential causes for the variation of conceptual complexity across different operations. One user in the rating

study commented, “*some are easy to understand, since (these functions) such as copy, cut, paste, normally have buttons and icons*”. Some other commented that they rated caret movement high because the touchscreen could not sense their fingers correctly. Thus, the conceptual complexity could be related to the user’s prior knowledge of the mobile interaction and the required accuracy of the operations.

The form of the manipulated object

While the form of caret is as small as one pixel, text editing involves the manipulation of a larger area in the editor, and the targets of input configuration are mostly imaginary/abstract concepts, such as input mode. We argue that smaller object requires higher accuracy of actions to trigger the operation, which is related to users’ ability in one particular handheld posture and interaction space.

Definition of interaction spaces

We defined interaction space as the area where users gesture to execute a certain operation. It may affect the user ability of performing gestures, including the degree of the freedom, the steadiness, and the endurance of the gesture (Wigdor, 2011). We specified three different interaction spaces for our study (Figure 3).

On-screen interaction (Figure 3, left) means restricting a user’s finger on the touchscreen. Therefore, the user can only perform translation and rotation on the X-Y plane, and translation in the Z-axis (i.e. pressing on the surface). On the other hand, on-screen interaction provides an anchor surface and tangible feedback, which can reduce user fatigue. Mid-air interaction (Figure 3, middle), in contrast, forbids any contact with the screen. It provides six degrees of freedom, but with low steadiness and endurance. The combination of surface and mid-air space (Figure 3, right) allows users to freely choose their preferred gesture space. Table 4 summarizes the comparison among the three different interaction spaces.

Definitions of handheld postures

The abilities and the constraints of hand action are also determined by the posture of holding the tool (Sturman, 1991). The tool refers to mobile phone in the scope of this paper. We investigated the three most common handheld postures of smartphones (Figure 4).

Suggested by Azenkot and Zhai's research (2012), one-thumb posture achieved the highest accuracy in text entry with lowest speed, while two-thumb posture attained highest entry speed but with the lowest accuracy, and the performance of index-figure posture was in between the two. However, Ng et al.'s study (2014) showed there were minor differences among the effects of encumbrance on the three postures in mobile interaction.

We further conducted a preliminary analysis of the capability of each handheld posture according to Sturman's criteria on hand actions capability (Sturman, 1991) and Hooper's analysis (2013) of different mobile handheld postures. One-thumb posture (Figure 4, left) is the most common way of holding and interacting with a mobile phone. However, it has limited range of motion, since only one thumb is free to move around a fixed point of contact with no coupling and spatial interference from the other hand. Two-thumb posture (Figure 4, middle) enables the coordination of two operating thumbs, and the range of motion is literally doubled. The downside, though, is that spatial interference possibly introduced. The third handheld posture involves holding the phone with one hand and interacting with the index finger of the other hand (Figure 4, right). This provides the highest freedom of motion without spatial interference, despite the lack of coupling with the other hand. Table 5 summarizes the comparison among the three different postures.

Sample size estimation

Based on the existing results of user performance on mobile operations under different handheld modes/interaction spaces (Wobbrock, 2008; Azenkot, 2013; Markussen, 2014; Jakobsen, 2015), we conducted a power analysis suggesting a minimum sample size of 54. There would be 80% chance of detecting a gesture-related behavioral change (i.e. difference in nature, dimension usage, and number of steps) among different combination of tasks, handheld postures, and interaction spaces at the 0.05 significance level (Hertzog, 2008). This calculation was based on the assumption that the magnitude of the difference in performance among the groups would equal the standard deviation of the population, an assumption consistent with previous studies investigating the difference in user performances in different mobile tasks under different handheld postures (Wobbrock, 2008; Azenkot, 2013) and typing and selection task performance in different spaces (Markussen, 2014; Jakobsen, 2015).

Participants

Our elicitation study used a between-group design with 54 participants (28 male, age range 20–35) who were recruited through public posters at a local university campus. All participants had at least a high school education and had over one-year experience with smartphones, with 78% of them using smartphones for more than two years. Seventy-four percent of the participants usually interacted with their phone with one thumb; 26% with index fingers. Sixty-eight percent used smartphone mostly for messaging (including SMS and social messaging), 22% for phone calls, and 11% for social networking. To investigate the effect of and the interaction among individual factors on mental model, each participant was assigned to one particular combination of three factors, which was derived from our counter-balanced experimental order: 3 types of operations, 3 handheld postures, and 3 interaction spaces.

Apparatus

As shown in Figure 5, we used Huawei Ascend P1 as the model phone for the experiment, because this model was not mentioned in the pre-screening questionnaires to alleviate the possible effect of association with a familiar interface. We kept the mobile device off throughout the study to minimize potential biases caused by visual feedback (Ruiz, 2011). Participants were encouraged to treat the device as a piece of an “almighty” tool which is able to process any gesture input and explore design alternatives freely. We displayed the information on the specific combination of tasks, handheld posture, and interaction space, on a 22” monitor in front of the participants. We also videotaped hand movements of participants from two angles (top view and side view). The investigators took detailed notes on their observation of each session and marked any confusion or question encountered by the participants.

Procedure

Upon the arrival of a participant, the investigator introduced the purpose of the study and asked the participant to sign the consent form voluntarily. We then explained the flow of the experiment, introduced the definitions of each task, described each handheld posture and each interaction space, and presented the think-aloud protocol to encourage the participants to verbally describe their gestures. The study consisted of three parts, counter-balancing the three factors to remove possible order effect. During each part, the participant had to define gestures for randomized operations of a particular task, with a particular handheld posture, in a particular interaction space. We asked participants to repeat each gesture three times for data reliability. The study lasted for approximately 30 minutes.

Measures

To gain an in-depth understanding, we analyzed the user-defined gestures from different perspectives, including subjective ratings and feedback, gesture agreement, and a taxonomy of gesture design space.

Ease-to-Define, Comfort-of-Use, and Preferences

At the end of the study, participants rated how comfortable they were with surface, mid-air, and combined gestures, respectively, on a 7-point Likert scale. They also expressed and justified their preferences among these three types of interaction space.

Taxonomy of user-defined mobile gestures

We adapted the taxonomies of gestures defined in previous work (Wobbrock, 2009) and selected a set of parameters that may better capture the differences in gestures under different conditions. As listed in Table 6, the first three parameters are related to the physical properties of gestures: *number of steps* denotes whether the gesture consists of a single action or multiple actions; *degree of freedom* defines the space/dimensions in which the gesture is performed; and *context* reflects the possible use of additional GUI elements (e.g. button, menu, scrollbar, etc.). The value of these parameters were actually embedded in the gestures and assigned based on video analysis.

The parameter of nature, on the other hand, attempts to conceptualize the mapping between a gesture and objects or events in the real world, which could indicate the user's mental image evoked by the operation and further lead toward their mental model. In particular, physical gestures are direct manipulation of digital content and interface elements (e.g., clicking buttons or dragging). Metaphorical gestures are mimicking actions in everyday life, such as throwing or flipping a page. Symbolic gestures convey information via visually drawn symbols like a heart or a letter. Abstract

gestures are somewhat arbitrary with no explicit linkage to the intended operation or real-world stimuli.

Results of the Elicitation Study

In this section, we present findings from the data collected during the elicitation study regarding how interaction spaces, operations, and handheld postures may affect the perception and use of mobile-gestural interaction.

Level of agreement

All 54 participants provided gestures for each referent for different combinations of operations, handheld postures, and interaction spaces. For each combination, we grouped the gestures so that each group held one identical representative gesture which was clustered across all participants. We grouped the gestures through open coding and calculated the agreement scores for each operation with the methodology of Wobbrock et al. (2009). The overall agreement score across all operations was 0.26 (SD = 0.11), and one-way ANOVA revealed that the agreement rates were significantly affected by the interaction spaces ($F(2, 30) = 4.52, p < 0.01$). A post-hoc Turkey HSD test showed that the agreement rate of the user-defined mid-air gestures (0.27) was significantly higher than those on the surface ($p < 0.01$) or in the combined space ($p < 0.05$), indicating the user-defined gestures in mid-air were more diverse than the gestures defined in the other two spaces.

There was no significant difference between the diversity of surface gestures and combined gestures ($p = 0.09$). This could have been due to the high degree of freedom in the mid-air space, which could evoke more diverse gestures. Additionally, the agreement scores were significantly varied among different operations (one-way ANOVA: $F(2, N = 30) = 9.32, p < 0.05$). While the text-editing tasks obtained the highest semantic imageability rating (4.92/7) and the lowest conceptual

complexity(1.48/7), the gestures defined for text-editing operations were significantly more diverse (agreement rate: 0.19) than those for caret positioning (agreement rate: 0.23, post-hoc test: $p < 0.05$) and input configuration (agreement rate: 0.35, post-hoc test: $p < 0.01$). As well, the overall agreement for input configuration gestures was significantly higher than the caret-positioning gestures (post-hoc test: $p < 0.05$). One possible reason for these effects could be the significant difference among the semantic imageability of the three types of operations, meaning that users tended to design more types of gestures for text-editing operations which can easily evoke mental images.

Ease-to-define, Comfort-of-use, and Preference

One-way ANOVA suggested that the ease-to-define varied significantly across different operations ($F(2, 1620) = 2.721, p = 0.066$). The overall mean of ease-to-define score was 2.59 (SD = 1.53). From high to low, the ranking of the ease-to-define score was caret movement (mean = 2.69, SD = 1.56), text editing (mean = 2.65, SD = 1.52), and input configuration (mean = 2.44, SD = 1.50).

The think-out-aloud protocol suggested that most participants treated an operation as a challenging one to define an appropriate gesture for when he/she felt it was difficult to come up with a non-GUI-based gesture. This observation was further justified by participant comments stating that they could not comprehend GUI interaction as a type of gesture, although we told them it was acceptable to adopt existing GUI interaction, such as pressing buttons. While rating the ease-to-define for the *Enable Shift* operation, P8 commented, “*The first thing came up in my mind is to press the button, because there is always a shift button there.*”

The participants’ comments suggested the score of ease-to-define was also affected by their perception on the device sensibility. For instance, some participants rated it easier to define gestures for the operations of *Move caret to the beginning/end* of

the page than those of *Move caret left/right by one letter*, as they perceived it less accurate to sense gestures for smaller-scale movements like letter by letter. P9 commented, “*It is very hard for the phone to detect small movements, like moving one letter.*” Similarly, the ease-to-define score decreased as the size of to-select content increased in the text-editing-related tasks. Specifically, it was easier for participants to design gestures for *Select all* than *Select one word*.

Notably, participants often considered avoiding repeated gestures for different tasks although it was not a compulsory requirement. For instance, participants found it more difficult to define gestures for caret movements, as there were three tasks for moving left/right in different scales; P2, P11, and P30 commented, “*it is easy to run out of ideas for gestures to distinguish these tasks.*”

Participants’ perceived comfort-of-use varied significantly given different interaction space (One-way ANOVA: $F(2,81) = 21.27$, $p < 0.05$, $\eta^2 = 0.34$). A post-hoc Tukey HSD test suggested that participants were marginally more comfortable gesturing on the surface than in a combined interaction space ($p = 0.09$), while both interaction spaces were significantly more comfortable than mid-air-only gestures ($p < 0.05$). In addition, there were only 13% of gestures in the combined space that involved mid-air components. Thirty-eight out of the 54 participants selected surface gesture as their favorite, the main reasons being: 1) direct touch feedback ($n = 10$); 2) more controlled ($n = 8$); 3) more stable ($n = 8$); 4) more familiar ($n = 8$); and 5) more accurate ($n = 4$). The rest of the participants preferred combined gesture because they found it more fun ($n = 6$), more flexible ($n = 6$), and had more options ($n = 8$).

Taxonomic breakdown

Three independent coders assigned values of nature to every gesture performed by the participants. They first coded data from three participants and compared results to

remove possible misconception. Then they proceeded to code all the rest of the data. We performed an inter-rater reliability analysis using the Fleiss' Kappa statistic to determine the consistency among the three coders. The three coders achieved almost perfect agreement over 1620 gestures, with a Fleiss' Kappa score of 0.85 ($z=33.0$, $p<0.01$). A fourth coder resolved the disagreement among the three coders.

Major gestures were GUI-based physical gestures

Figures 7, 8, and 9 illustrate the breakdown of the 1620 gestures collected during the study according to the nature and the conditions (operations, handheld postures, and interaction spaces). We observed the user-defined gestures showing that 51% of the gestures elicited for the operations of input configuration and text editing (here caret positioning is excluded because most related gestures were physically manipulating the caret), and 72% of these physical gestures involved interaction with GUI elements (such as menu, button, etc.) regardless of the operations, the handheld postures, or the interaction spaces. One representative GUI-based physical gesture was "*click on the functional button.*" The Pearson correlation coefficient value of 0.43 suggested there was a moderate positive correlation between the number of physical-based gesture and the number of GUI-based gestures in each operation.

After being asked about the reasons for defining GUI-based gestures, 7 participants stated that they were the most familiar action for him/her (P1, P7, P10, P32, P36, P37, P45), and 12 participants commented that GUI-based gestures came up immediately to his/her mind while seeing the operations (P2, P3, P5, P7, P10, P16, P19, P25, P32, P33, P43, P54), which indicated the influence of their prior knowledge with mobile interaction. Interestingly, some participants perceived that GUI-based gestures cost less cognitive load to learn and memorize than non-GUI-based gestures (P1: "*I think buttons and menus are easier to understand and remember.*"), as a counter

argument for the existing research that suggested user-defined gestures can be easy to learn (Wobbrock, 2005).

While the user-defined gesture set in this elicitation study was dominated by physical and GUI-based gestures, statistical analysis revealed that different operations, handheld postures, and interaction spaces still had significant influence on users' choice of the nature, the number of steps, and the use dimension of the gestures.

Nature of gestures

Although physical gestures dominated the overall gesture set, multinomial logistic regression revealed that the nature of the user-defined gesture was significantly affected by the operations ($\chi^2(6, 1620) = 227.82, p < 0.05$), the interaction spaces ($\chi^2(6, 1620) = 47.95, p < 0.05$), and the handheld postures ($\chi^2(6, 1620) = 18.12, p < 0.05$). As expected, there were significantly more physical gestures for the operations involving caret positioning. The dominant gesture for caret positioning contains the movement of “swiping towards the corresponding position” (Figure 10), which we interpreted as the physical manipulation of moving the caret.

The multinomial logistic regression showed a significant interaction effect between handheld postures and the interaction spaces on the nature of gesture ($\chi^2(12, 1620) = 25.86, p < 0.05$). Besides the obvious fact that the physical gestures of manipulating caret dominated in the task of caret movement, different combination of handheld postures and interaction spaces strongly affected the nature of gesture for caret positioning. This could have been due to the difference in the range of movement of the handheld posture or the degree of freedom in the interaction space. Specifically, 27.5% of the user-defined gestures were symbolic (e.g., “drawing arrow”, “drawing half circle”, “drawing letters”, etc.) while using one thumb to move the caret on the screen, as shown in Figure 11. Participants commented that it was difficult to reach the targeted

position on the screen with one thumb, thus he/she tended to perform gestures as signals to trigger the operations (P10, P46, P47).

In addition, the multinomial logistic regression showed that operations and handheld postures had significant interaction effect on the nature of the user-defined gestures ($\chi^2(6, 1620) = 76.23, p < 0.01$). There were significantly more metaphorical (33.1%) and abstract (23.1%) gestures for the task of text editing with one-thumb posture than other combinations. This could have been due to the high imageability of text-editing operation and the low degree of freedom of the one-thumb posture. As shown in Figures 12–17, some of these gestures matched the descriptions of the mental images evoked by the operations, such as “throw to garbage bin” for *Delete*, “circle” for selecting content, and “click/press” as a metaphor for *Paste*, etc. While the freedom of movement was increased (e.g., index-finger posture), users tended to rely more on the menu manipulation to trigger the text-editing operations, which indicated a stronger reliance on prior knowledge when give more freedom. Some participants commented that he/she tended to place buttons in different positions on the screen and trigger the text-editing functions by direct button clicking, as “*the mobile screen is getting bigger*” (P3), and “*it is easy to reach different screen position with the index finger*” (P50).

Number of steps

The number of steps involved in a gesture was significantly influenced by various characteristics of the corresponding operation (multinomial logistic regression: $\chi^2(4, 1620) = 207.08, p < 0.05$). Figure 18 shows the distribution of the user-defined gestures according to the number of steps across different types of operations. For an operation with imaginary targeted concept (i.e. input configuration), non-physical gestures resulted in fewer number of actions. The Pearson correlation coefficient value of 0.87 suggests a significant positive correlation between the number of non-physical gestures

and the number of single-step gestures in the input-configuration-related operations. On the contrary, when the task was with more concrete but smaller manipulated object (e.g., caret positioning) and paired with the operation for content selection, participants took extra actions to ensure that the operation landed on the correct target and position. As participants (P3, P4, P6, P12, P17, P30, P41) pointed out, they preferred surface gesture for this type of task as it provided finer and more stable control with higher accuracy.

On the other hand, when participants were operating the phone with one thumb, they significantly favored gestures consisting of only one action (multinomial logistic regression: $\chi^2(4, 1620) = 16.71, p < 0.05$). In particular, when the range of motion was really narrow (i.e. with the one-thumb posture), people tended to use single-step gestures, such as clicking. Figure 19 shows 77.4% of the one-thumb gestures were 1-step, while it was 65.7% for index finger and 67.5% for two thumbs.

As suggested by Azenkot and Zhai's results (2012), one-thumb posture had the lowest speed in typing. Importantly, this posture has lowest range of motion (echoed by the participant comment "*one thumb is limited in terms of movement capability, especially in air*"), thus users preferred a gesture with a single step to increase the speed. In contrast, two-thumb and index-finger postures afforded higher speed and accuracy and thus accommodated more gestures with multiple steps.

Furthermore, the number of steps was also significantly affected by the interaction spaces (multinomial logistic regression: $\chi^2(4, 1620) = 12.39, p = 0.015$). As shown in Figure 20, there were more single-step gestures (74.7%) in the air, which could have been due to the fragileness of the interaction space, and the combined spaces evoked significantly more gestures with two steps (63.6%).

Usage of spaces

Since a large portion of the gestures were GUI-related, the z-axis was mostly involved in clicking actions. As there were significantly more GUI gestures for input configuration, most of the gestures for input configuration involved single-step button clicking along the z-axis, as shown in Figure 21.

The usage of different dimensions was significantly affected by handheld postures (multinomial logistic regression: $\chi^2(4, 1620) = 18.25, p < 0.05$). When performing gestures with one thumb, we observed that users tended to produce more gestures along the z-axis only and less 3D gestures using the x-y-z space, which could have been due to the limited range of movement. There were more x-y-z gestures (20.2%) with the two-thumbs posture, and most of these x-y-z gestures in two-thumbs posture involved the coupling of two fingers, meaning that one finger performed in the x-y plane and the other finger clicked along the z axis. Figure 22 shows the breakdown of the gestures according to their usage of space across different handheld postures. Forty-seven percent of one-thumb gestures used only the z-axis, compared with 39.4% for index and 40.8% for two thumbs.

The interaction spaces were also significantly affected the usage of different dimensions (multinomial logistic regression: $\chi^2(4, 1620) = 64.36, p < 0.05$). There were very few 3D gestures that fully leveraged the freedom of mid-air space, which instead evoked significantly more x-y gestures, as shown in Figure 23 (15.3% of the mid-air gestures were 3D gestures, and 50.2% were 2D). This suggested that with a large freedom of movement, users seldom included any x-y actions along with z-axis actions, which could have been due to the lack of resting surface and physical feedback in mid-air. When given both freedom and surface feedback in the combined spaces, users provided significantly more (25.5%) 3D gestures (mainly clicking/tapping, plus 2D

gestures, such as swiping). However, most of the 3D gestures contained the x-y actions and the z actions separated in sequence (e.g., press first and swipe) instead of 3D gestures using the x-y-z space simultaneously.

Many participants commented in the post-study interviews that, although they managed to come up with mid-air gestures for all operations as required, they would not recommend many of these gestures for real use. When asked the reasons, participants commented, *“in the air it is too easy to loss control; ”*(P27) *“when send message the accuracy is important, for game, fun is more important;”*(P1) *“in the air not very comfortable;”*(P45) *“it's very hard when your finger has to hold in air, until you spot the specific location, then touch – and I always tab the wrong spot.”* (P48)

Discussion

In this section, we share insights into gestural mobile interface design based on our observations on users' mental model in the elicitation study.

Surface and mid-air gestures play different roles in combined space

We observed that participants tended to assign different meanings to in-air gestures and on-screen gestures in the combined interaction space. We divided the gestures in the combined space into two types: signal and functional. We found that 34% of the combined gestures began with a signal action on the screen followed by a functional action, mostly in the air. Another 23% added an additional signal action at the end, while some 20% started with a mid-air functional action and ended with an on-screen signal. P40 mentioned, *“Mid-air gestures are more suitable for functionalities, such as copy, cut, undo, and gaming, while on-screen gestures are more for accurately locating the starting point.”*

In addition, participants tended to reuse the same gestures in different interaction spaces for different tasks. For instance, P25 defined “swipe up” on screen for *Cut*, and

the same swipe gesture in the air for *Delete*. He commented that swiping on screen means the temporary removal which could be followed by *Paste*, while swiping in the air means a permanent removal of the text content. P16 used the same directional swiping gestures for moving caret left/right letter by letter on screen and moving caret to the beginning/end of the row in the air. She explained that the large-scale movement of the caret requires less accuracy of the gesture, thus she can just use symbolic gestures to signal the system for triggering these tasks. We also observed that participants tended to adopt similar gestures in different parts of the screen to distinguish different operations. This suggests that it is necessary to develop a context-aware algorithm for gesture-based mobile interaction.

Furthermore, participants tended to assign different gestures to different fingers. Based on our observation, participants were more likely to make use of both thumbs when they could not reach certain part of the screen or when they needed to specify an area. For example, for the *Select one word* operation, several participants kept pressing on the beginning of the sentence with one thumb and swiped across the entire sentence with the other. However, in these cases, other than offloading the job to both hands, the core actions of the gestures were carried out by only one thumb.

Non-identical agreement for dichotomous operations

Echoing a previous study (Wobbrock, 2009), we observed that mirroring gestures occurred for dichotomous operations. Conversely, we observed a clear difference between the gesture agreement for *Previous page of options* (combined space = 0.59, mid-air = 0.20, on-screen = 0.31) and *Next page of options* (combined space: 0.27, mid-air: 0.26, on-screen: 0.40). This may indicate that users perceived the dichotomous operations differently. A T-test indicated that the conceptual complexity of these two

operations were significantly different ($t(10)=-2.67$, $p<0.05$), meaning users might perceived these two dichotomous operations differently.

Conclusion

In this paper, we presented two elicitation studies to investigate user mental model on gestural mobile interaction under the influence of different combinations of operations, handheld postures, and interaction spaces. Our findings suggest that users tended to define more symbolical and metaphorical single-step gestures while performing highly imageable operations (e.g., text editing) in the posture with low motion range (e.g., one thumb). Additionally, we observed that the 3D space in the mid-air was not fully leveraged by the users. In the future, we plan to run a larger-scale elicitation study with bias-removing techniques so that we can explore the possible interaction effect of operations and handheld posture more deeply. We also plan to study if user-related factors (e.g., experience with mobile devices, culture, age, etc.) may affect the conceptualization of mobile gestural interaction.

References

- Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., ... & Koh, J. T. K. V. (2012). Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. Microsoft Research TechReport MSR-TR-2012-111.
- Allen, R. B. (1997). Mental models and user models. Handbook of human-computer interaction, 1, 49-63.
- Arefin Shimon, S. S., Lutton, C., Xu, Z., Morrison-Smith, S., Boucher, C., & Ruiz, J. (2016, May). Exploring Non-touchscreen Gestures for Smartwatches. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 3822-3833). ACM.
- Azenkot, S., & Zhai, S. (2012, September). Touch behavior with different postures on soft smartphone keyboards. In Proceedings of the 14th international conference

- on Human-computer interaction with mobile devices and services (pp. 251-260). ACM.
- Bruder, G., Steinicke, F., & Sturzlinger, W. (2013, July). To touch or not to touch?: comparing 2D touch and 3D mid-air interaction on stereoscopic tabletop surfaces. In *Proceedings of the 1st symposium on Spatial user interaction* (pp. 9-16). ACM.
- Chan, E., Seyed, T., Stuerzlinger, W., Yang, X. D., & Maurer, F. (2016, May). User Elicitation on Single-hand Microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3403-3414). ACM.
- Chen, X. A., Schwarz, J., Harrison, C., Mankoff, J., & Hudson, S. E. (2014, October). Air+ touch: interweaving touch & in-air gestures. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (pp. 519-525). ACM.
- Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review*, 4(2), 226-231.
- Dim, N. K., Silpasuwanchai, C., Sarcar, S., & Ren, X. (2016, June). Designing Mid-Air TV Gestures for Blind People Using User-and Choice-Based Elicitation Approaches. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (pp. 204-214). ACM.
- Freeman, E., Brewster, S., & Lantz, V. (2014, September). Towards usable and acceptable above-device interactions. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (pp. 459-464). ACM.
- Fuccella, V., Isokoski, P., & Martin, B. (2013, April). Gestures and widgets: performance in text editing on multi-touch capable mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2785-2794). ACM.
- Garcia-Lopez, E., de-Marcos, L., Garcia-Cabot, A., & Martinez-Herraiz, J. J. (2015). Comparing Zooming Methods in Mobile Devices: Effectiveness, Efficiency, and User Satisfaction in Touch and Nontouch Smartphones. *International Journal of Human-Computer Interaction*, 31(11), 777-789.

- Gilhooly, K.J. and Logie, R.H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behaviour Research Methods and Instrumentation*, 12, 395-427.
- Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in nursing & health*, 31(2), 180-191.
- Hoover, S. (2013). How do users really hold mobile devices. UXmatters, <http://www.uxmatters.com/mt/archives/2013/02/how-do-users-really-hold-mobile-devices.php>.
- Jakobsen, M. R., Jansen, Y., Boring, S., & Hornbæk, K. (2015). Should I stay or should I go? Selecting between touch and mid-air gestures for large-display interaction. In *Human-Computer Interaction–INTERACT 2015* (pp. 455-473). Springer International Publishing.
- Jonassen, D. H., & Henning, P. (1996, July). Mental models: Knowledge in the head and knowledge in the world. In *Proceedings of the 1996 international conference on Learning sciences* (pp. 433-438). International Society of the Learning Sciences.
- Karlson, A., Bederson, B., & Contreras-Vidal, J. (2006). Understanding single-handed mobile device interaction. *Handbook of research on user interface design and evaluation for mobile technology*, 86-101.
- Leap Motion. <https://www.leapmotion.com/>.
- MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction*, 7(1), 91-139.
- Markussen, A., Jakobsen, M. R., & Hornbæk, K. (2014, April). Vulture: a mid-air word-gesture keyboard. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1073-1082). ACM.
- Marquardt, N., Jota, R., Greenberg, S., & Jorge, J. A. (2011). The continuous interaction space: interaction techniques unifying touch and gesture on and above a digital surface. In *Human-Computer Interaction–INTERACT 2011* (pp. 461-476). Springer Berlin Heidelberg.
- Morris, M. R., Huang, A., Paepcke, A., & Winograd, T. (2006, April). Cooperative gestures: multi-user gestural interactions for co-located groupware. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1201-1210). ACM.

- Muller, K. (1989). Statistical power analysis for the behavioral sciences. *Technometrics*, 31(4), 499-500.
- Ng, A., Brewster, S. A., & Williamson, J. H. (2014, April). Investigating the effects of encumbrance on one-and two-handed interactions with mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1981-1990). ACM.
- Ni, T., Bowman, D., & North, C. (2011, May). AirStroke: bringing unistroke text entry to freehand gesture interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2473-2476). ACM.
- Paivio, A., Yuille, J.C. and Madigan, S.A. (1968). Concreteness, imagery and meaningfulness values for 925 words. *Journal of Experimental Psychology Monograph Supplement*, 76 (3, part 2).
- Park, W., & Han, S. H. (2014). An analytical approach to creating multitouch gesture vocabularies in mobile devices: A case study for mobile web browsing gestures. *International Journal of Human-Computer Interaction*, 30(2), 126-141.
- Roberts, T. L. 1980. Evaluation of Computer Text Editors. Ph.D. Dissertation. Stanford University, Stanford, CA, USA.
- Romano, M., Paolino, L., Tortora, G., & Vitiello, G. (2014). The tap and slide keyboard: a new interaction method for mobile device text entry. *International Journal of Human-Computer Interaction*, 30(12), 935-945.
- Ruiz, J., Li, Y., & Lank, E. (2011, May). User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 197-206). ACM.
- Samsung Galaxy S4. <http://www.samsung.com/global/microsite/galaxys4/>
- Shimon, S. S. A., Morrison-Smith, S., John, N., Fahimi, G., & Ruiz, J. (2015, August). Exploring User-Defined Back-Of-Device Gestures for Mobile Devices. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 227-232). ACM.
- Shoemaker, G., Findlater, L., Dawson, J. Q., & Booth, K. S. (2009, May). Mid-air text input techniques for very large wall displays. In *Proceedings of Graphics interface 2009* (pp. 231-238). Canadian Information Processing Society.
- Sturman, D. J. (1991). Whole-hand input (Doctoral dissertation, Massachusetts Institute of Technology).

- Toglia, M.P. and Battig, W.F. (1978). *Handbook of Semantic Word Norms*. New York: Erlbaum.
- Tung, Y. C., Hsu, C. Y., Wang, H. Y., Chyou, S., Lin, J. W., Wu, P. J., ... & Chen, M. Y. (2015, April). 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). ACM.
- Vatavu, R. D. (2012, July). User-defined gestures for free-hand TV control. In *Proceedings of the 10th European conference on Interactive tv and video* (pp. 45-48). ACM.
- Wacharamanotham, C., Todi, K., Pye, M., & Borchers, J. (2014, April). Understanding finger input above desktop devices. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1083-1092). ACM.
- Wigdor, D., & Wixon, D. (2011). *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier.
- Wobbrock, J. O., Aung, H. H., Rothrock, B., & Myers, B. A. (2005, April). Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems* (pp. 1869-1872). ACM.
- Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009, April). User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1083-1092). ACM.
- Wobbrock, J. O., Myers, B. A., & Aung, H. H. (2008). The performance of hand postures in front-and back-of-device interaction for mobile computing. *International Journal of Human-Computer Studies*, 66(12), 857-875.

Author Biographies

Kening Zhu is an Assistant Professor at the School of Creative Media. He received his PhD from the National University of Singapore, and his bachelor degree on Computer Science from Huazhong University of Science and Technology, China. His research interests include mobile interaction design, tangible user interfaces and rapid prototyping.

Xiaojuan Ma is an assistant professor of Human-Computer Interaction at the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. Her research interests include multimodal interaction, data visualization, gamification, crowdsourcing, and computational linguistic.

Haoyuan Chen is a student research assistant in the School of Creative Media, City University of Hong Kong, under the supervision of Dr. Kening Zhu. He is now finishing his bachelor degree on Computer Science in CityU. He has participated in various HCI projects, including mobile interface and child-computer interaction.

Miaoyin Liang is a research assistant in the JC School of Public Health and Primary Care, CUHK. She received her master degree on public health from University of Hong Kong, and her bachelor of science from Sun Yat Sen University, China. Her expertise lies in quantitative study on human behaviour.

Figure 1. Imageability ratings of operations.

Figure 2. Conceptual-complexity ratings of operations.

Figure 3. Three types of interaction space: surface (left), mid-air (middle), and combined (right).

Figure 4. Three types of handheld posture: one thumb (left), two thumbs (middle), and index finger (right).

Figure 5. Setup of the elicitation study.

Figure 6. Agreement rates for user-defined mobile text-related gestures.

Figure 7. Breakdown of gestures according to the nature across different operations.

Figure 8. Breakdown of gestures according to the nature across different handheld postures.

Figure 9. Breakdown of gestures according to the nature across different interaction spaces.

Figure 10. Dominant gestures for caret movement. (The circles in “Index finger in the combined space” and “Two thumbs in all spaces” mean single click.)

Figure 11. Symbolic gestures for caret movement with one thumb on screen.

Figure 12. Non-physical gestures for text editing with one thumb in the air (The circles indicate tapping in the air).

Figure 13. Non-physical gestures for text editing with one thumb on the screen (The circles indicate tapping on the screen, and two circles indicate double clicking, and the black circle mean holding for a long time).

Figure 14. Non-physical gestures for text editing with one thumb in the combined space
(The circles indicate tapping on the screen, and two circles indicate double clicking, and the black circle mean holding for a long time).

Copy

Cut, Paste

Delete

Delete all

Select all

Select one
sentence

Select one
word

Redo

Undo

Figure 15. Non-physical gestures for text editing with the index finger in the air (The circles indicate tapping on the screen, and two circles indicate double clicking, and the black circle mean holding for a long time).

Select one word/sentence

Select all₁

Select all₂

Select all₃

Redo

Undo

Figure 16. Non-physical gestures for text editing with the index finger on the screen.

Figure 17. Non-physical gestures for text editing with the index finger in the combined space.

Figure 18. Breakdown of gestures according to the number of steps across different operations.

Figure 19. Breakdown of gestures according to the number of steps across different handheld postures.

Figure 20. Breakdown of gestures according to the number of steps across different interaction spaces

Figure 21. Breakdown of gestures according to the usage of space across different operations.

Figure 22. Breakdown of gestures according to the usage of space across different handheld postures.

Figure 23. Breakdown of gestures according to the usage of space across different interaction spaces

Table 1. The list of 30 text-related operations divided into 3 categories.

Caret Movement	Input Configuration	Text editing
Move Caret left by one letter	Switch to letter mode	Select one word
Move Caret right by one letter	Switch to number mode	Select one sentence
Move Caret left by one word	Switch to punctuation mode	Select all
Move Caret right by one word	Switch to emoticon mode	Copy
Move Caret up by one row	Enable/Disable capitalization	Cut
Move Caret down by one row	Enable/Disable shift	Paste
Move Caret to the beginning of row	Show more options	Delete
Move Caret to the end of row	Next page of options	Delete all
Move Caret to the beginning of page	Previous page of options	Undo
Move Caret to the end of page	Select from options	Redo

Table 2. Definitions of the analytical characteristics of mobile operation that could affect user's mental model.

Analytical characteristics	Definition
Conceptual Complexity	Conceptual complexity refers to how straightforward it is to understand the textual meaning of the phrase.
Semantic Imageability	Semantic imageability refers to how easy it is to relate the phrase to a mental image. (Pavio, 1986)
Form of targeted object	The targeted objects of text-related operations include caret, text content, and input setting.

Table 3. The comparison among the three types of text-related mobile operations.

	Caret movement	Input Configuration	Text editing
Conceptual Complexity	High	Medium	Low
Semantic Imageability	Low	Medium	High
The form of the targeted object	single pixel- based position	Imaginary concept	A large area of text content

Table 4. The comparison among the three types of interaction spaces.

	Degree of freedom	Steadiness	Edurance
On-screen	Low	High	High
Mid-air	High	Low	Low
Combined	High	Medium	Medium

Table 5. The comparison among the three types of handheld postures.

	Accuracy	Speed	Range of motion	Hand Coupling	Spatial Interference
One thumb	High	Low	Low	No.	No.
Two thumbs	Low	High	Medium	Yes	Yes
Index	Medium	Medium	High	No	No

Table 6. Taxonomy of user-defined gestures.

Taxonomy of Mobile Gestures		
Number of Steps	Single step	Containing one action
	Two steps	Containing two actions
	Three or more steps	Containing three or more actions
Degree of Freedom	X-Y	On the 2D plane
	Z	Along the Z axis
	X-Y-Z	In 3D space
Context	With context	Requiring additional GUI elements
	Without context	Not requiring additional GUI elements
Nature	Physical	Manipulating interfaces
	Metaphorical	Depicting a metaphor
	Symbolic	Depicting a symbol
	Abstract	Acting arbitrary

Table 7. Experiment design.

Independent variables	<ul style="list-style-type: none"> • Operation {Caret positioning, Input configuration, Text editing}, • Interaction Space {On-surface, Mid-air, Combined} • Handheld Posture {One hand holding and one thumb operating, One hand holding and one index finger operating, two hands holding and two thumbs operating}
Dependent variables	<ul style="list-style-type: none"> • Agreement rates of the user-defined gestures • Taxonomic properties of the user-defined gestures (nature, number of steps, usage of space, context)
Number of Participant	54
Duration per participant	30 minutes
Location and language	The study was conducted in City University of Hong Kong, Hong Kong SAR. The written language in the study is English. In oral communication, Cantonese (35/54), Mandarin (10/54) and English (9/54) were used according to the participants' preference.
Major Task of the experiment	Holding the mock-up phone in three different postures, each participant needed to design 30 gestures to achieve 30 mobile operations categorized in three types in three different spaces.
Procedure	<p>Step 1: Briefing of study purpose & consent acquisition.</p> <p>Step 2: Introduction of three independent factors and the three particular combinations for the participant.</p> <p>Step 3: Gesture design.</p> <p>Step 4: Debriefing, post-study questionnaire, and reward collection.</p>